Data analytic approaches to detect stress from physiological data

by Hieu Vu Tran

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of
Prof. Paul J. Kennedy
Prof. Carolyn McGregor AM

University of Technology Sydney
Faculty of of Engineering and IT

January 2025

Certificate of Original Authorship

I, Hieu Vu Tran, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy in the School of Computer Science, Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research was supported by an Australian Government Research Training Program (RTP) Scholarship doi.org/10.82133/C42F-K220.

Production Note:

SIGNATURE: Signature removed prior to publication.

[Hieu Vu Tran]

DATE: 8th January, 2025

PLACE: Sydney, Australia

Acknowledgements

I would like to thank my supervisors, Professor Paul Kennedy and Professor Carolyn McGregor, for all the supervision and guidance during the four years of my PhD. Professor Paul Kennedy, with a supportive supervisory approach and numerous discussions, helped me significantly in overcoming the difficulties I faced during my PhD experiments, including those involving topological data analysis. His detailed comments on my thesis and publications have helped me become a much better academic writer. Professor Carolyn McGregor, with extensive knowledge of physiology, helps me make my research much more complete in its physiological aspects. With their funds, I can attend academic conferences, which greatly help me stay connected to the research community.

I also want to thank my family members, especially my parents, for their unlimited support. I owe a great deal of gratitude to my best friend, Khanh. He is a great listener who understands my PhD challenges and supports me when I need it.

I also want to say thank you to my colleagues at the University of Technology Sydney. Thank you, Matthew Gaston, for answering my questions regarding the use of the high-performance computers at UTS. Thank Tobias and Nico, my laboratory mates, for making my PhD journey much more fun.

I am grateful for the scholarship provided by Vingroup Science and Technology Scholarship Program.

Contents

List of Figures

List of Tables

List of Publications

Listed below are the publications and other outputs associated with the research presented in this thesis.

Tran, H. V., McGregor, C., Kennedy, P. J., 2023, 'Detecting stress from multivariate time series data using topological data analysis', *Australasian Joint Conference on Artificial Intelligence*, Springer, pp. 341–353

This publication is associated with Chapter 5.

Tran, H. V., Cheong, S.A., McGregor, C., Kennedy, P. J., Complex network approach to classify stress from wearable devices, IEEE Access 2024 (Submitted)

This publication is associated with Chapter 6.

List of Abbreviations and Symbols

Abbreviation	Description
ANS	Autonomic nervous system
ARIMA	Auto Regressive Integrated Moving Average
BVP	Blood volume pulse
CNN	Convolutional Neural Networks
DBN	Deep belief network
EEG	Electroencephalogram
EDA or GSR	Skin conductivity or Galvanic skin response
EKG or ECG	Electrocardiogram
EMA	Ecological momentary assessment
EMG	Electromyogram
GAN	Generative adversarial network
HRV	Heart rate variability
IAPS	Cold pressor test and International Affective Picture System
kNN	k nearest neighbors
LDA	Linear discriminant analysis
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MIST	Montreal Imaging Stress Task
NASA-TLX	NASA Task Load Index
PANAS	Positive and Negative Affect Schedule
PCA	Principal component analysis
POMS	Profile of Mood States
PSS	Perceived Stress Scale
RMSE	Root Mean Square Error
RSME	Rating Scale Mental Effort
SAM	Self-Assessment Manikin
SCWT	Stroop color-word inference test
SRI	Stress Response Inventory
SSRS	Stress Self-Rating Scale
SSSQ	Short Stress State Questionnaire
SSST	Sing-a-Song Stress Test
STAI	State-Trait Anxiety Inventory
SVM	Support Vector Machine
SWELL_KW	Smart Reasoning for Well-being at Home and at Work Knowledge Work
TDA	Topological Data Analysis
TSST	Trier Social Stress Test
ROC	Receiver Operating Characteristic
RNN	Recurrent Neural Networks
VAR	Vector Autoregression
WESAD	Wearable Stress and Affect Detection

Abstract

Stress has become one of the most critical dangers to human health. Recently, with the development of wearable devices that can measure physiological features more accurately and collect more data, stress detection using the data collected from wearable devices has become an emerging research topic. There are currently many studies on detecting and predicting stress. However, these studies are not comprehensive enough because (i) they provide limited comparison with other methods, (ii) they simply use statistics to show correlation between physiological features, (iii) they use mainly classical methods or advanced methods with raw physiological data, or (iv) they do not compare to other methods using benchmark datasets. New data analytic approaches, such as topological data analysis (TDA), which is robust to noise and can analyse high-dimensional data, can also create more opportunities for research about stress. Using complex networks representations of the physiological data and associated algorithms to detect stress is a promising solution. Since complex networks have the potential to capture phenomena of real-world systems.

This thesis presents novel approaches to present the relationship between stress and physiological time series features using topological data analysis and complex networks. Benchmark datasets named Driver Stress Data, WESAD and SWELL_KW are used to deploy experiments. This thesis has three contributions to knowledge. Contribution one is a series of methods to detect stress from conventional machine learning and statistical methods, as well as deep learning methods with the outer product of time series data or image-based data as input. These methods were validated on benchmark datasets. This chapter presents one of the first experiments to compare classical machine learning and statistical methods in the context of physiological data. This chapter also describes two novel deep learning inputs, which are the outer product of the time series signal and image-based data. In the future, the latter feature can be improved to utilise the current strong development of computer vision methods to develop a comprehensive stress detection method. Contribution two is a novel topological data analysis method for detecting stress in long multivariate physiological time series data. This method can overcome the high computational expenses of TDA, especially for multivariate time series data. It can predict stress with higher performance than state-of-the-art time series classification methods and other stress detection methods on a benchmark dataset. Contribution three is a novel complex networks approach to classify stress. This is one of the first complex networks combined with deep learning approach to handle long multivariate time series, and also has the ability to classify each time stamp of a time series. This complex network method achieved excellent results on stress detection benchmark datasets such as WESAD and SWELL-KW. The results of this method are also better than those of state-of-the-art time series classification methods and other stress detection methods. Two feature extraction processes using TDA and complex networks, together with the corresponding deep learning models for stress detection, are detailed in this report. These two new approaches were validated on stress benchmark datasets, compared with other stress detection methods on the same datasets, and evaluated against state-of-the-art time series classification methods.

Chapter 1

Introduction

1.1 Background

Stress nowadays has become a critical problem for human beings. Long-term stress contributes to significant neurological disorders, heart issues, stomach ulcers, asthma, diabetes, headaches, accelerated aging, and early death in humans, and 70% of dangerous diseases can be caused by stress (Rasheed 2016). Stress is harmful to human health. Stress can cause permanent emotional, physiological, or behavioural changes in humans, increasing the risk of having diseases (Cohen et al. 2007; Davis et al. 2017). Stress is also a primary reason that causes an abundance of mental health disorders, such as depressive disorder, bipolar disorder, and posttraumatic stress disorder (Davis et al. 2017). According to research in 2019 by the World Health Organization, 970 million people (1/8 of the world population) were suffering from at least one type of mental health disorder (Kestel et al. 2022). The rate of people who suffer from mental health disorders remained unchanged at 1/8 of the world population from 2000 to 2019 (Kestel et al. 2022).

Stress is defined as the "non-specific response of the body to any demand upon it" (Kirschbaum et al. 1993). Selye described the word "stress" as the consequence of any significant threat to homeostasis (Selye 1978). The term "homeostasis" was first defined by Walter Bradford Cannon in 1926 and refers to the maintenance of a constant internal physical and chemical state by living systems (Cannon 1926; Neligan and Baranov 2013). Based on this definition, Claude Bernard stated that the ability to maintain a balanced life depends heavily on the capacity to keep the human body's internal environment stable despite changes in the external environment (Bernard 1957).

Detecting stress is urgent, as it is vital for improving human health. Humans have become increasingly aware of their stress levels and are still seeking the best way to detect stress. Classical stress detection methods use a blood or saliva sample to measure cortisol levels or specialised medical devices to monitor physiological signals (Jiménez-Mijangos et al. 2023; Ramírez et al. 2023). The most common classical methods use questionnaires, interviews, or self-reports (Jiménez-Mijangos et al. 2023; Ramírez et al. 2023). All these classical methods require participants to either join a laboratory environment with specialised medical equipment or answer questions (Jiménez-Mijangos et al. 2023). Meanwhile, evaluating questions is time-consuming, so not all stress conditions are recorded promptly. Because of these disadvantages, there is substantial demand for stress detection methods to enable earlier interventions (Gedam and Paul 2021; Hickey et al. 2021). Stress detection using physiological data collected from wearable devices is the most suitable way for two reasons (Babu et al. 2024; Panicker and Gayathri 2019).

First, physiological signals are among the most common and reliable sources for stress detection (Panicker and Gayathri 2019). There are numerous advantages of physiological signals over other sources for detecting stress, including high accuracy, high reliability, cost-effectiveness, non-invasiveness, and continuous collection (Goshvarpour et al. 2017; Panicker and Gayathri 2019; Yoo et al. 2018). There are several other advantages. Using physiological signals can also simplify the entire stress detection process, from data collection to interpretation of results. These can also make it easy for participants to join the stress detection experiment. Using physiological signals increases trustworthiness because participants cannot hide changes in their physiological signals. Finally, physiological signals cause fewer potential sensitive issues due to cultural differences. The high accuracy of physiological data collected from wearable devices has also been reported in other studies. Ollander et al. (2016) deployed research to detect stress from a wearable device (Empatica E4 wristband) and stationary laboratory sensors and devices. Konstantinou et al. (2020) also compared the stress detection capabilities of a wearable device (Microsoft Band 2) and standard stationary devices, such as Biopac MP150. They both stated that wearable technology is equivalent to stationary technology in its capacity to assess stress. Betti et al. (2017) confirmed that physiological data collected from wearable devices can have a similar ability to detect stress as well as salivary cortisol.

Second, wearable devices can continuously monitor physiological data and are promising tools for measuring stress (Babu et al. 2024). The rapid development of technology can enable wearable devices to measure physiological signals more precisely, making them more affordable (Babu et al. 2024). These reasons make wearable devices more popular (492.1 million devices were sold worldwide in 2022) and can be more common tools for detecting stress (Babu et al. 2024). The most commonly used wearable devices in academic research on stress detection are the Empatica E4, Apple Watch, Spire Stone Doppel, and Stila (Ramírez et al. 2023).

Stress detection research is still in the early stages of development. Numerous stress detection methods used classical data analytic approaches. These classical approaches used original time series data, time-domain and frequency-domain features as inputs and machine learning models such as support vector machines, random forest, and k-nearest neighbour to classify stress (Babu et al. 2024; Hag et al. 2021; Mentis et al. 2023). State-of-the-art data analytic approaches, combined with the rapid development of wearable devices, offer great potential for developing excellent stress detection methods. This PhD project aims to develop stress detection methods using novel data analytics and physiological time-series data collected via wearable devices. The two state-of-the-art data analytic approaches used in this PhD thesis are topological data analysis (TDA) and complex networks. These can provide insights from complex, large-scale data, a characteristic that traditional data analytic approaches cannot offer (Amaral 2022; Skaf and Laubenbacher 2022). Topological data analysis can capture the "shape" of the data (Skaf and Laubenbacher 2022), and complex networks can build connections between discrete and unstructured data to detect patterns and predict behaviours (Amaral 2022).

1.1.1 Definitions of stress

Stress is a widespread aspect of daily human life. One of the most common experiences of humans nowadays is facing stressful situations (Anisman and Merali 1999). However, there is still no official definition of stress. According to Mason (1975), there are three approaches to define stress. It can be defined as an organism's physical, emotional, and cellular reactions (the strain), external events that cause these reactions (the stressor), or events that occur when a person reacts to the surrounding environment. Mason (1975) called these three definitions:

response, stimulus and stimulus-response interactions, respectively. With the first definition, stress can be easily recognised when unusual reactions occur, such as the human body becoming sweaty, the human heartbeat becoming faster, the human face becoming pale, or trembling in the human's hands. Regarding the second definition, external events primarily relate to a physical or socio-cultural context (Aldwin 2007). Physical context can arise when certain physical factors, either immediately or later, cause harmful effects on the human body. Examples of immediate results include fire, catastrophe, and tsunami, and examples of the last-effect case include pollution and noise (Aldwin 2007). Socio-cultural contexts are associated with a person's well-being, status and role. Lack of well-being can be a source of stress. With the third type of definition, stress is related to the demands of the surrounding environment and personal resources; stress occurs when personal resources cannot meet these demands.

Hans Selye was the first researcher to describe biological stress and coined the term 'response'. He made the rudimentary definition of stress as "the nonspecific neuroendocrine response to the body" (Selye 1936). He first used the term 'stress' in 1950 in his paper (Selye 1950). He also called the factor that causes stress the stressor, for example, fear, anxiety, cold, heat or pain (Rochette and Vergely 2017). Stress can be defined as a human reaction to demand or threat (Can et al. 2019a). Other researchers used the combinations of 'stimulus' and 'stimulus-response interactions' approaches to define stress since they mentioned relationships between the external environment and incidents caused by interaction between human beings and stressors. According to Fink, stress occurs when an environment's demands exceed a person's awareness and endurance (Fink 2016). Beehr and Newman defined workplace stress by surveying various studies; they consider any work-related aspect that causes abnormal changes in workers' psychological and physiological states to be stress (Schuler 1980).

People experience stress at least once in their lives, yet nobody clearly defines or describes it (Fink 2016). The reason is that stress is highly subjective (Can et al. 2019a). Stress can vary among individuals, depending on personal susceptibility and resilience (Fink 2016). An acceptable level of stress is beneficial in daily life, helping to improve productivity, focus, and emotional state (Can et al. 2019a). However, if the stress reaches a dangerous level, it will harm the human's well-being, affecting the emotional state, quality of life and productivity (Fink 2016;

Umematsu et al. 2019b). Stress is nowadays a critical mental health problem. As the World Health Organization reported, stress affects more than 450 million people around the world (World Health Organization 2021). For this reason, early stress detection is essential.

1.1.2 Significance

Stress is one of the most dangerous threats to human well-being. Early detection of negative stress responses becomes more practical with the rapid development of wearable devices. This research area is at an early stage of development. Many researchers create benchmark datasets, stress induction tests, and methods for stress detection and prediction. However, there is still a lack of a standard for negative stress detection methods. This research will provide some approaches to this research topic.

First, it will provide a comprehensive survey of the relationship between physiological features and the stress response from several perspectives, such as machine learning, deep learning, complex networks, and algebraic topology. As a result, more unique insights about these relationships can be discovered. Second, it will provide new stress detection methods using novel data representations. These results can improve human health, as the effect can be detected earlier from physiological characteristics.

1.2 Research challenges

Stress detection is an emerging research topic right now. This thesis focuses on detecting stress in multivariate physiological time series data. The data are physiological signals such as brain signals (electroencephalogram), skin conductance, muscle signals, blood volume pulses, skin temperature, and respiration data, collected from wearable devices. A wearable device is a small electronic device that can be worn on the body or attached to clothing and enables noninvasive, long-term monitoring of essential physiological and biochemical parameters (Babu et al. 2024). Wearable devices are smartwatches, rings, wristbands, and smart patches (Babu et al. 2024). Examples of wearable devices are ActiGraph devices, Fitbit devices, Garmin devices, Apple watches, Empatica devices, Zephyr BioHarness, Astroskin and Hexoskin (Cosoli et al. 2024; Inibhunu et al. 2021; MacQuarrie et al.

2024; McGregor 2021; Miyakoshi and Ito 2024). The data format is a time series of various physiological measures sampled at a particular rate; each time series of physiological measures contains multiple timestamps, and each timestamp is associated with a stress value.

However, there are still various limitations to current research approaches. First, numerous researchers use simple statistical or machine learning techniques to analyse stress, so the insights are unclear, or the designed methods can not detect stress effectively. As a result, these research approaches are not helpful in real-world contexts. Some researchers used statistical techniques such as ANOVA and Pearson's correlation to examine correlations between signals that can cause stress and stress levels (Sano et al. 2018; Wang et al. 2014). However, the statistical approach cannot provide a proper and complete stress detection system, as these studies merely present correlations between input features and stress rather than a complete method for detecting stress from input features. Some other researchers applied classical machine learning models, such as logistic regression, Support Vector Machines, Random Forest, and k-nearest neighbour (Babu et al. 2024; Mentis et al. 2023) to detect stress. Logistic regression typically works well only for detecting linear relationships in the data and cannot capture complex patterns (Schober and Vetter 2021; Stoltzfus 2011; Tu 1996). Support Vector Machines incur high computational costs when trained on large datasets and are unsuitable for unbalanced datasets (Bhavsar and Ganatra 2012; Cervantes et al. 2020). Random Forest tends to incur high computational expense when working with large datasets and is also sensitive to noise (Hengl et al. 2018). K-nearest neighbours have poor runtime performance with huge datasets (Cunningham and Delany 2021; Dhanabal and Chandramathi 2011; Triguero et al. 2016). But nowadays, with the rapid development of technology, wearable devices have become common and can collect complex, lengthy data that contains multiple signals. So, these research studies using classical methods are not very applicable.

Deep learning overcomes the disadvantages of classical machine learning models such as logistic regression, Support Vector Machines, Random Forest, and k-nearest neighbour. Feature selection is necessary for all classical machine learning models to make these models become effective and achieve good classification results; however, it is time-consuming and requires expert knowledge (Badr et al. 2024; Craik et al. 2019). Meanwhile, deep learning models can still perform well with direct input data and even without feature selection, since these models can perform classification tasks alongside feature selection (Aloysius and Geetha 2017; Badr et al. 2024; LeCun et al. 2015). Deep learning can handle vast amounts of data compared to conventional machine learning (Badr et al. 2024). It is an important advantage because of the rapid development of wearable devices, which enables the collection of more and more data.

Secondly, researchers have mainly used simple handcrafted features that apply some time-series transformation methods, so these features may not provide complete information about the physiological data (Li and Liu 2020). The calculation of these handcrafted features also adds burden to the data preprocessing process, since each set of features must be manually generated from the corresponding physiological signals (Li and Liu 2020). The most common features used to detect stress are heart rate, heart rate variability, skin temperature, galvanic skin response, blood pressure, respiration rate and their classical time-domain and frequency-domain indices (Gedam and Paul 2021; Li and Liu 2020). The limit of classical frequency-domain indices is the ability to handle non-stationary data (Al-Fahoum and Al-Fraihat 2014; Mohammed and Rantatalo 2020). The classical time-domain indices also have drawbacks in complexity and the possible loss of temporal precision (Cohen 2004).

Some data transformations may yield better input features for stress detection experiments; for example, instead of using the original time series data, the outer product of the time series data, or images generated from the time series data, can be used. More advanced and novel data analytics approaches, such as topological data analysis (TDA) and complex networks, can provide more comprehensive inputs to deep learning models for stress detection. Topological data analysis is able to capture the “shape” of the data and preserve the global information of the data (Pun et al. 2022; Skaf and Laubenbacher 2022). Complex networks have the power to analyse complex systems, from natural systems to social systems (Ding et al. 2024; Gao et al. 2021; Rubinov and Sporns 2010). Complex networks have the ability to analyse multivariate, noisy and non-stationary physiological data, which can be a promising approach for classifying responses to stress. However, there are no suitable roadmaps for applying new techniques to real-world applications, such as stress detection. Stress detection involves physiological data in the form of long, multivariate time series. A TDA experiment with a proper feature extraction method and acceptable computational cost for long time series data remains lacking. Finding a suitable graph topology for representing physiological time series and extracting suitable complex network features from multivariate time series data is still a research topic that must be addressed.

Thirdly, numerous studies on stress detection did not provide comprehensive comparisons with other research, making it difficult to assess the method's effectiveness. Most researchers use their own collected data (Airij et al. 2018; Asif et al. 2019; Cheema and Singh 2019; Lanius et al. 2024; Mandal et al. 2023; Phutela et al. 2022; Ray et al. 2023; Tian 2022; Vila et al. 2018; Xia et al. 2019). The unique characteristics of these datasets can lead to better classification results in the associated research, which may not occur when validated on other datasets. Some researchers just showed their results and concluded without complete comparisons with different stress classification methods (Keshan et al. 2015; Li et al. 2019; Rizwan et al. 2019). There are only a few comparisons using benchmark datasets (Mishra et al. 2020).

This PhD thesis provides methods that integrate deep learning with TDA and also complex networks features to detect stress from long multivariate time series data. The results of these methods and associated experiments were validated on common stress detection benchmark datasets, which makes it easier to compare the results with other methods.

1.3 Research questions

The central theme of this project is to detect stress and analyse its manifestation through changes in physiological data. This research will concentrate on answering these research questions:

Research question 1 : How can we detect stress using conventional machine learning and statistical methods, deep learning models with the outer product of physiological time series as inputs or deep learning models with image-based inputs?

Research question 2 : Is it a promising approach to detect negative stress responses in physiological time series data using topological data analysis (TDA)?

Research question 3 : Is it a promising approach to detect negative stress responses in physiological time series data using complex networks?

These research questions lead to the following objectives.

Objective 1 : Develop and validate methods that use conventional machine learning and statistical approaches, or the outer product of physiological time series data or image-based data, generated from physiological time series data, to detect stress.

Success condition: deploy methods to detect stress from conventional machine learning and statistical methods, stress detection methods where the inputs are in novel formats, such as the outer product of physiological time series or image-based data. Stress detection methods based on conventional machine learning and statistical methods are among the earliest approaches to physiological multivariate time series data. The latter stress detection methods utilised novel 2D data formats similar to images.

Objective 2 : Develop and validate topological data analysis methods to detect negative stress responses.

Success condition: deploy a novel topological data analysis method that identifies stress, validate this method on benchmark datasets, compare with other stress detection methods on the same datasets and compare with time series classification methods.

Objective 3 : Develop and validate complex networks methods to detect negative stress responses

Success condition: deploy a novel complex networks method to detect negative stress response, validate this method on benchmark datasets, compare with other stress detection methods on the same datasets and compare with time series classification methods.

1.4 Contributions to knowledge

This research will contribute novel approaches to analysing physiological features and time series analysis. These approaches also align with the declared research objectives.

Contribution 1: Validated stress detection methods that use conventional machine learning and statistical methods, and stress detection methods with inputs that are the outer product of physiological time series data or image-based data.

This contribution includes a complete and one of the first comparisons between conventional machine learning methods and a conventional statistical method to detect stress in a physiological context. This contribution also provides novel input for stress detection from time series data. The new image-based input can be combined with advanced computer vision methods to develop a comprehensive stress detection method from time series data.

Contribution 2: A novel topological data analysis approach to detect stress. This method is validated on benchmark datasets, compared with stress detection methods on the same datasets and compared with time series classification methods.

Contribution 3: A novel complex networks method to detect stress. This method is validated on benchmark datasets and compared with stress detection methods on the same datasets and time series classification methods.

1.5 Thesis structure

The remainder of this thesis is structured as follows:

Chapter 2 reviews the literature on stress detection using wearable devices. This chapter presents a comprehensive literature review of physiological feature analysis across several topics. First, this chapter presents physiological and derived statistical features for detecting stress. This chapter then presents the current challenges in collecting data for stress detection research. This chapter also includes tests to induce stress and an overview of the most common stress benchmark datasets. Methods for analysing physiological data and for stress detection research using classical machine learning, deep learning, and forecasting methods were also provided in this chapter. Finally, research gaps are presented.

Chapter 3 describes three benchmark datasets named Driver Stress Data, Wearable Stress and Affect Detection (WESAD), and Smart Reasoning for Well-being at Home and at Work Knowledge Work (SWELL-KW) and two data analytic techniques named topological data analysis and complex networks used in this thesis. The section on topological data analysis provides information on the core concepts of TDA, a literature review of state-of-the-art research on TDA and machine learning, and time series and change point detection. A section on complex networks presents how to use time series segmentation methods and complex network approaches to convert time series data into networks, current complex network approaches to analyse time series data, and a literature review of state-of-the-art research on complex networks and time series.

Chapter 4 presents methods to detect stress using the outer product of physiological time series data and image-based data. This chapter also contains an experiment using the conventional machine learning methods (LSTM, RNN) and statistical method (VAR) to classify stress from multivariate physiological time series.

Chapter 5 presents a method for detecting stress in physiological time series data using topological data analysis. The results of the classification methods are validated on a common benchmark dataset, Wearable Stress and Affect Detection (WESAD). The results are also compared with other time series classification and stress detection methods on the same datasets.

Chapter 6 presents a method for detecting stress in physiological time series data using complex networks. The results of the classification methods are validated using common benchmark datasets such as Wearable Stress and Affect Detection (WESAD) and Smart Reasoning for Well-being at Home and at Work Knowledge Work (SWELL-KW). The results are also compared with those of other time series classification and stress detection methods on the same datasets.

Chapter 7 concludes the thesis by reviewing the findings' implications and potential future research directions.

Chapter 2

Literature review

This chapter critically reviews the literature related to this research project. It provides information on stress detection using data collected from wearable devices. The topics include a list of standard physiological features usually used to detect stress, current data collection challenges for stress detection research, laboratory tests used to simulate stressful situations, and benchmark datasets to detect stress. Classical approaches to detect stress and their limitations are presented, followed by a critical review of more recent approaches. This chapter ends with a summary of the research gaps in the area.

2.1 Physiological features to recognise stress

The stress signal will be sent to the hypothalamus when the human body faces physical or psychological stressors such as strenuous physical activities or challenging exams. The hypothalamus mobilises the sympathetic nervous system (SNS). The sympathetic nervous system awakens the body using several hormones such as cortisol and epinephrine. As a result, the value of several physiological features of the human body will rise, such as heart rate, blood pressure, speed of respiration, blood sugar and the amount of oxygen in the lung and brain (see figure 2.1) (Healey and Picard 2005; Ollander et al. 2016; Schuler 1980). These changes in physiological measurements can be detected using wearable devices. These transformations will make the body stronger and boost focus and reaction. When the indicators of stress disappear, the autonomic nervous system (ANS) will dismiss all these above stress processes, activate rest, and bring the human body to a normal state. The value of all these physiological features will return to baseline (Can et al. 2019a; Fink 2016; Gjoreski et al. 2016). This process is called the fight or flight mechanism (Fink 2016; Gjoreski et al. 2016).

[Production note: This figure is not included in this digital copy due to copyright restrictions.]

Figure 2.1: Stress and physiological features (Ledoyt 2019)

There are several ways to measure stress: estimating biological markers such as human cortisol level and amylase, analysing answers to subjective self-evaluation questions, and evaluating the physiological measurements of stress (Kim et al. 2018; Muaremi et al. 2014). These are also common ways to label the stress levels of stress detection datasets.

Commonly used subjective self-evaluation questions are the Perceived Stress Scale (PSS), Stress Response Inventory (SRI), Stress Self-Rating Scale (SSRS), State-Trait Anxiety Inventory (STAI), Self-Assessment Manikin (SAM) and Positive and Negative Affect Schedule (PANAS), Short Stress State Questionnaire (SSSQ) and Rating Scale Mental Effort (RSME) (Alberdi et al. 2016; Can et al. 2019a).

An emerging trend to detect stress is using wearable devices due to their mobility and accuracy compared to stationary devices when detecting stress-related features (Ollander et al. 2016). In a survey, a wearable device (Empatica E4 wristband) was compared with stationary laboratory sensors and devices to estimate physiological features and detect stress (Ollander et al. 2016). Ollander et al. (2016) confirmed that wearable devices have the same ability to measure heart rate and stress as stationary devices. This unique ability has the potential to open a promising research area of using wearable devices to detect stress. Some types of physiological measurements can be collected to identify stress, such as heart signal, brain signal, skin conductivity, muscle signal, blood volume pulse, skin temperature and respiration data (measured by chest cavity expansion) (Alberdi et al. 2016; Gjoreski et al. 2016; Healey and Picard 2005; Hernandez et al. 2011; Hovsepian et al. 2015; Muaremi et al. 2014; Sano and Picard 2013). These are reviewed below.

2.1.1 Heart signal

The electrocardiogram (ECG) evaluates the heart's electrical activity, and the ECG is measured by electrodes attached to the human body (Can et al. 2019a). There are four components of heart behaviour as captured by an electrocardiogram: the baseline, P wave, QRS wave and T wave (Baldoumas et al. 2018). The heart signal is one of the most important features for measuring human stress, and it includes several measures, namely the ECG, heart rate variability (HRV), and RR intervals, which are the times between heartbeats (see figure 2.2) (Gjoreski et al. 2016). RR intervals are also known as inter-beat intervals (IBI). When stress occurs, heart rate variability decreases, and RR intervals become shorter. The normal-to-normal (NN) intervals are another form of RR interval, but normal-to-normal intervals only measure the time between what algorithms determine are normal heartbeats (McCraty and Shaffer 2015). The ECG data was used in experiments of this thesis.

There are several statistical formulas to calculate time and frequency domain features from the normal-to-normal interval (see Table 2.1 and 2.2). The two most common ones are the root mean square of consecutive interbeat interval differences (RMSSD) and the standard deviation of the beat-to-beat interval (SDNN) (Kemper et al. 2007; Kleiger et al. 2005). However, there are still limitations with both time and frequency domain features for real-world stress detection experiments. Real-world stress detection experiments handle long physiological time series data. Generating time domain features are high computational expense processes (Al-Fahoum and Al-Fraihat 2014; Mohammed and Rantatalo 2020). This process can become harder when handling long time series. Frequency domain features can only handle stationary data (Cohen 2004). Meanwhile, almost all physiological time series data is non-stationary (Weber et al. 1992).

[Production note: This figure is not included in this digital copy due to copyright restrictions.]

Figure 2.2: Normal ECG signal and RR interval (Baldoumas et al. 2018)

Table 2.1: Heart rate variability and main statistical time-domain features (Shaffer and Ginsberg 2017)

Time-domain features	**Unit**	**Descriptions**
AVNN	ms	Average of all normal-to-normal intervals
SDNN	ms	Standard deviation of all normal-to-normal intervals
RMSSD	ms	Root mean square of successive RR interval differences
pNN50	%	Percentage of differences between successive normal-to-normal intervals that are greater than 50 ms

2.1.2 Brain signal

The brain signal is measured by the electroencephalogram (EEG). There are four wavebands of EEG: Alpha (8 to 13 Hz), Beta (13 to 30 Hz), Delta (0.1 to 4

Table 2.2: Heart rate variability and main statistical frequency-domain features (Shaffer and Ginsberg 2017)

Frequency-domain features	**Unit**	**Description**
ULF	ms^2	Total spectral power of all normal-to-normal intervals less than 0.003 Hz
VLF	ms^2	Total spectral power of all normal-to-normal intervals between 0.0033 and 0.04 Hz
LF	ms^2	Total spectral power of all normal-to-normal intervals between 0.04 and 0.15 Hz
HF	ms^2	Total spectral power of all normal-to-normal intervals between 0.15 and 0.4 Hz
LF-HF	%	Ratio of low to high-frequency power

Hz) and Theta (4 to 8 Hz) (Alberdi et al. 2016; Can et al. 2019a). The Alpha waveband presents a normal state when the mind is calm and balanced, and the Beta waveband presents emotional states. The Alpha activity will decrease, and the Beta activity will increase during a stressful time (Alberdi et al. 2016; Can et al. 2019a).

Seo et al. (2010) describe an experiment with 33 healthy participants, aged from 30 to 40 years old. Their research used Pearson correlation to find the relationship between EEG, electrocardiogram and salivary cortisol level (an indicator of stress). They observed a significant negative correlation between salivary cortisol level and ECG as measured with the SDNN and between the high-frequency band Beta of EEG and ECG as measured with the SDNN. They also observed a significant positive correlation between salivary cortisol levels and the high-frequency Beta band of EEG. They also found that the group of participants experiencing stress had the highest value of the Beta band. That experiment demonstrated that the Beta band EEG is a strong indicator of stress.

Zhang et al. (2014) explored the situation when humans must face cognitive workload and rest conditions. They combined galvanic skin response (GSR), heart rate (HR), and EEG with their proposed filtering technique to classify stress levels from a dataset with 16 participants. The State-Trait Anxiety Inventory (STAI) measured the stress statuses. There are two kinds of stress status in this work. One status happens when participants face danger or threat, and another one when faced with worry or discomfort. They concluded that EEG achieved higher accuracy (87.5%) in predicting stress than GSR and HR, respectively (75%, 62.5%). They also found that the high Beta band of EEG contributes more than the middle Beta and Alpha bands to classification. The conclusions from the paper can be more convincing if there is information about how to access the paper's dataset. Other researchers can do experiments with the same dataset and use other machine learning and deep learning methods to validate the stress detection powers of the physiological features from the data.

Several calculated characteristics of EEG have already been applied to detect stress, and the most frequently used ones are mean power ratio, mean amplitude of EEG, mean amplitude of Event-Related Potentials, Alpha, Beta and Theta wavebands (Alberdi et al. 2016; Can et al. 2019a). The EEG data was not used in the experiments of this thesis.

2.1.3 Skin conductance

Skin conductance gauges the skin's electrical characteristics (Figner and Murphy 2011). It is also known as electrodermal activity (EDA), galvanic skin response (GSR), skin conductance response (SCR), skin conductance activity (SCA), electrodermal level (EDL) and the electrodermal response (EDR). When stress happens, sweat glands are activated as a result, and there is more sweat at the surface of the skin, which increases skin conductance.

When a human body faces arousal events, high cognitive workload, or some strenuous physical activities, the human brain will change the electrical characteristics of the skin (Peternel et al. 2012) by reducing resistance and raising conductance (Dawson et al. 2017). EDA is typically measured using the trapezoid muscle (Greene et al. 2016). EDA is also one of the first features to detect stress and is a strong indicator of stress (Healey and Picard 2005). Some researchers even used EDA as the labelled data for their model (Peternel et al. 2012). The most popular features of EDA are maximum and minimum value, root mean square, mean amplitude and the standard deviation of the amplitude (Alberdi et al. 2016). The EDA data was used in the experiments of this thesis.

2.1.4 Muscle signal

Electromyogram (EMG) is the muscle's electrical activity. Some researchers observed a relationship between EMG and stress. Wijsman et al. (2013), in their research, pointed out that the amplitude of the EMG is higher when stress happens, and stress has a high correlation with signals from the upper trapezius muscle. They also confirmed that low-frequency characteristics of EMG also rise when stress appears. The popular statistical features of EMG are mean, median, standard deviation and root mean square of the amplitude (Alberdi et al. 2016). The EMG data was used in the experiments of this thesis.

2.1.5 Blood volume pulse

The Blood volume pulse (BVP) is the amount of blood that goes through the photoplethysmographic (PPG) sensor in each pulse (Peper et al. 2007). PPG can be put in various body parts to measure BVP; the most prevalent place is the finger. Amplitude of the blood volume pulse reduces when stress happens. BVP is one of the first physiological features used to detect stress in the experiment of Zhai and Barreto (2006). In that experiment, two authors stated that BVP has the same detection power as EDA and skin temperature. Heart rate can also be calculated approximately from BVP (Peper et al. 2007). The BVP data was not used in the experiments of this thesis.

2.1.6 Skin temperature

Skin temperature can also measure stress since skin temperature will change when humans experience stressful situations. In normal conditions, human skin temperature is generally between 32°C and 35°C. Skin temperature can change for many reasons such as fever, starvation, strenuous physical activities, climate and physiological changes (Quazi et al. 2012). Only physiological changes can vary skin temperature when the other factors do not happen. The cause of physiological changes is localised changes in blood movement originating from vascular resistance or arterial blood pressure, and these two characteristics are controlled by the autonomic nervous system (Maaoui et al. 2008). The skin temperature change is caused by a change in blood flow that the autonomic nervous system manages. The skin temperature data was used in the experiments of this thesis.

2.1.7 Respiration

Respiration data, as measured by chest cavity expansion, can also be considered as a physiological measurement of stress. Grossman (1983) states that stress can cause phenomena such as a boost in respiration rate and irregular breathing. Nevertheless, in the study of Healey and Picard (2005), they realised that respiration data is not a strong indicator of stress such as HR and EDA. The respiration data was used in the experiments of this thesis.

2.2 Data collection challenges

To detect stress, researchers collect physiological data from candidates. Then they use reports or questions to recognise the stress level of candidates. They can also estimate stress conditions using biological markers such as human cortisol level and amylase. Another practical approach to collect ground truth is ecological momentary assessment (EMA), in which symptoms of affective level can be managed in real-time contexts (Asselbergs et al. 2016; Moskowitz and Young 2006). Traditional EMA collection methods include diaries or paper-and-pencil questionnaires; new EMA collection methods integrate with mobile applications (Asselbergs et al. 2016). Mobile EMA can gather data at a specific time, multiple times a day or right after an event happens (Asselbergs et al. 2016).

Data quality is a significant factor in the effectiveness of stress detection. Excellent quality data must be correct, complete, appropriate, up to the minute, meticulous, relevantly described and contain sufficient information to aid decision making (Wyatt and Liu 2002). Notable topics for setting up high-quality data to achieve the requirements of Wyatt and Liu (2002) are the standard data collection processes, how to extract useful information from vast amounts of data, how to choose suitable artifact detection techniques, suitable strategies to label massive amounts of data, and the types of devices with appropriate battery life.

2.2.1 Quality of collected data

There are several requirements to ensure the quality of data. The devices that collect physiological features must achieve the standard to provide excellent data quality, which Wyatt and Liu defined (Wyatt and Liu 2002). The apparatus must also be placed in a suitable place and follow standards to collect physiological data (Alberdi et al. 2016; Can et al. 2019a). An appropriate sampling frequency should simulate the signal. Khusainov et al. (2013) stated that for monitoring daily life activities, the frequency sampling of 20Hz is enough for almost all cases, and audio, speech and physiological features need to use a higher frequency with a maximum value of 40kHz. Kwon et al. (2018) declared that the minimum frequency for ECG for adults should be 250Hz, and the acceptable minimum frequency in situations not requiring domain frequency features should be 100Hz. This thesis used three benchmark datasets to test and then validate the algorithms that were developed in this thesis. All these datasets contain ECG. One small dataset has a frequency of 15.5 Hz which is less than 20Hz. So, this small dataset is not suitable for analysing physiological data in general as stated by Khusainov et al. (2013). The frequency of this small dataset is also less than 250Hz (15.5 Hz) as a result this small dataset is not suitable for analysing ECG as stated by Kwon et al. (2018). However, this dataset was only used for testing algorithms. Since the algorithms have high computation expenses, these need to be tested with a small dataset before applying on a larger scale. The two modern and more enormous datasets used to validate the algorithms have frequencies much higher than 20Hz and 250Hz (700Hz and 2048 Hz). One modern and more enormous dataset was downsampled to 128Hz to avoid huge computational expenses. However, this frequency is still suitable since the experiments do not use any domain frequency feature. So, the two datasets fulfil the frequency requirement to validate stress detection algorithms.

After collecting data, pre-processing data steps need to be applied. Data quality can be affected by magnetic, random, instrumentation or electric noise (Khusainov et al. 2013). Standard methods to remove noise are filters such as Kalman, median, Butterworth low-pass, wavelet decomposition or Wiener (Alberdi et al. 2016; Can et al. 2019a; Khusainov et al. 2013). Suitable features will be collected for the specific type of signal. Appropriate filters will be chosen for specific kinds of noise(Khusainov et al. 2013). Principal component analysis (PCA), least mean square, regression analysis or independent component analysis (ICA) can be applied for removing artifact (Alberdi et al. 2016; Can et al. 2019a; Khusainov et al. 2013). However, a data system or model with all these requirements may not have high-quality data. This thesis develops novel approaches using new data analytics methods that can be robust to noise and artifacts, so these pre-processing data processes were not used in the experiments of this thesis.

Stress data can be gathered from multiple wearable devices and sensors. There are some issues when managing data from multiple sources; for example, data with different physical units, dimensions, and data can be missed or inconsistent (Alberdi et al. 2016; Can et al. 2019a). Data fusion frameworks and voting systems were proposed to handle missing and inconsistent data (Alberdi et al. 2016). This thesis uses benchmark datasets collected using the same wearable devices and sensors, so data fusion frameworks and voting systems were not used in the experiments of this thesis.

2.2.2 Large amounts of data

Another challenge with a continuous collection process is the amount of data. A massive amount of data from an ongoing collection process can affect the storage of devices, battery life of devices, and complexity of approaches to handle data (Alberdi et al. 2016; Can et al. 2019a). This challenge can arise from a large number of features and records. Issues with massive numbers of records can be handled using onboard signal processing methods, which can reduce the computational costs of transfer and save storage (Alberdi et al. 2016; Can et al. 2019a). Suitable sliding window methods can be used to handle an appropriate number of records rather than many records simultaneously. Large numbers of features can be handled using feature extraction algorithms. Feature extraction algorithms such as PCA, waveform information, power spectral analysis, wavelet coefficients, non-linear analysis, and linear discriminant analysis (LDA) can be used to reduce the amount of data needed to be stored but still keep the necessary information (Alberdi et al. 2016; Chen et al. 2017; Wang et al. 2013). Researchers can apply feature selection methods to reduce the complexity of approaches, choosing the optimal subset of features from the whole collection of features.

There are several approaches for feature selection, such as filter, wrapper and embedded. Filter methods use statistics to rank the best features for classification (Tang et al. 2014). The filter methods are also simple and low-cost, but these ignore the correlation between features and labels. The wrapper methods used classification algorithms' accuracy or error rate to evaluate the performance of features (Tang et al. 2014). With this unique characteristic, the performance of wrapper methods will be much better than filter methods. However, wrapper approaches also are expensive to run. The embedded approach is a combination of filter and wrapper methods. It first uses statistical criteria to select several best subsets of features and then applies classification accuracy to select the best subset (Tang et al. 2014). The embedded ones have better performance accuracy than the filter and run faster than the wrapper. Examples of filter methods are Mutual Information (MI), Linear Discriminant Analysis (LDA) and Pearson correlation. Sequential Forward Search (SFS) and Sequential Backward Search (SBS) are examples of wrapper methods. L1 regularisation and Random Forest are examples of embedded methods.

2.2.3 Labelled data

One of the most critical data collection concerns is how to accurately label the data. To be labelled, the data must follow strict pre-defined regulations in the laboratory environment. Still, this process becomes more arduous in a real-world environment since it requires additional human resources. There are two main approaches to labelling the data: using self-reports or questions and using mobile applications to collect user stress conditions (Can et al. 2019a). This thesis used three benchmark datasets and all used the self-reports or questions to label the data.

2.2.4 Types of devices and battery life

One factor relating to the quality of stress datasets is the set of features of the measurement devices. Many devices are stationary laboratory ones that are not easy to deploy on a real-world experiment to collect data (Konstantinou et al. 2020). Some are obtrusive, requiring to attach to the human body and making stress detection experiments more artificial and also can change the stress states of participants (Gaskin et al. 2017).

Devices used for stress detection processes should be unobtrusive and non-invasive. Non-invasive devices are accessories that do not break the skin or physically enter the body. Unobtrusive and non-invasive devices can support collecting physiological data seamlessly and pervasively without interaction between participants and the system (Postolache et al. 2010; Zhang et al. 2014). These devices can be accessories, stick-on electronic tattoos, or attachments to clothes or daily living environment (Zhang et al. 2014). With the fast development of mobile phones and smartwatches, wearable devices have revolutionised data collection in an unobtrusive and non-invasive way (Cosoli et al. 2023; Kazanskiy et al. 2024).

2.3 Simulated events to cause stress

Researchers proposed simulated events, tasks or tests to create stressful conditions. Artificial stressors include performing arithmetic or calculation tasks, memory tasks, logical tasks, public speaking, car-driving simulations, games or viewing unpleasant images (Alberdi et al. 2016). The most common tests from the literature are the Trier Social Stress Test (TSST), Montreal Imaging Stress Task (MIST), Stroop color-word inference test (SCWT), Sing-a-Song Stress Test (SSST), Cold pressor test and International Affective Picture System (IAPS) (Brouwer and Hogervorst 2014; Can et al. 2019a; Dedovic et al. 2005; Kirschbaum et al. 1993; Lowery et al. 2003; Stroop 1935).

The Trier Social Stress Test (TSST) contains three steps totalling 15 minutes (Kirschbaum et al. 1993). The TSST environment is a room with three judges and one participant, a video camera and a tape recorder. Researchers collect participants' blood, saliva, and heart rate data during the test. In the first step, the participant prepares a talk for a job interview in 10 minutes. The participants must convince the judges that he or she is suitable for this position by using the talk. In the second step, participants present the speech to three judges. The judges ask the participants to continue if the talk length is less than 5 minutes. In the last step, the participants must repeatedly subtract 13 starting from 1,022 as quickly and accurately as possible. If the participant makes a mistake, then he or she will have to start from the beginning. At the end of the test, an investigator explains the test's purpose to the participant. After that, the participant will rest from 30 to 70 minutes (Kirschbaum et al. 1993).

The Montreal Imaging Stress Talk contains three steps: rest, control and experimental (Dedovic et al. 2005). The test environment has a computer screen. In the rest step, the participant only looks at the screen. In the control step, the participant sees arithmetic questions via the computer screen and submits answers. No information about the time limit, progress bar or user's performance is displayed on the screen. In the experimental step, the complexity of the questions increases, and the time for finishing the task decreases. Stress will happen because of the user's performance, the time limit is shown on the computer screen, and the reduction of the time limit and increase of complexity of questions (Dedovic et al. 2005).

The Stroop colour-word inference test was developed by Stroop (1935), in which participants need to read three lists that contain the name of colours; the two first lists are 'congruent', and the last one is 'incongruent' (Scarpina and Tagini 2017; Stroop 1935). The first list contains the names of colours in black ink (Word - W), and the second has colourful squares (Colour - C). The last table is 'incongruent' because colour-words (CW) are in clashing colours (for example, the word 'blue' is in red ink) (Scarpina and Tagini 2017; Stroop 1935).

Participants can see regular messages on a screen for one minute in the Sing-a-Song Stress Test. Finally, the screen displays the name of a song, and then participants need to sing the song aloud. Participants need to stand up during the experiment. Brouwer and Hogervorst (2014) stated this is a simple and easy way to cause stress, and they recognised that heart rate and GSR are higher during singing time than during watching time (Brouwer and Hogervorst 2014).

The cold pressor test induces stress by demanding participants place their hands in cool water for one minute (Lowery et al. 2003). The test also collects heart rate and blood pressure. They can put their hand out of the water if it feels too painful. The International Affective Picture System contains a wide range of images from normal ones (for example, a house) to ones that can cause unpleasant feelings (Can et al. 2019a).

These stress induction tests were often employed during stress detection experiments. These can be the standard to simulate stress for future research. This thesis used three benchmark datasets; one used the Trier Social Stress Test to create stressful conditions. The other two datasets created stressful environments using actual stressful events (one dataset has driving events, and another has working events).

2.4 Benchmark data sets

There are several standard benchmark datasets for stress detection described in the literature. Stress detection datasets can be categorised by environments such as the laboratory, strictly controlled environments, less controlled environments and everyday life. In the laboratory, the whole environment is controlled. Automobile and office are strictly controlled environments where the camera and sensors observe all actions and changes in physiological characteristics. Movement is limited in these environment. The university is an example of an environment with less control since activities are mostly only viewed in the classroom, hall, or particular university locations. The last type of environment is everyday life.

Three benchmark datasets, namely Wearable Stress and Affect Detection (WESAD), Smart Reasoning for Well-being at Home and at Work Knowledge Work (SWELL-KW), and A Database for Emotion Analysis using Physiological Signals (DEAP) are laboratory datasets. The Wearable Stress and Affect Detection (WESAD) data was introduced by Schmidt et al. (2018b). The Smart Reasoning for Well-being at Home and at Work Knowledge Work (SWELL-KW) dataset was created to understand stress in a working environment (Koldijk et al. 2014). The DEAP database measured stress conditions when participants listen to music videos (Koelstra et al. 2011). Driver Stress Data and Affective Road Data were built in a strictly controlled environment (automobile). The Driver Stress Data and AffectiveRoad were introduced in 2000 and 2018 (Haouij 2018; Healey 2000). In these real driving environments, the driving roads of participants were fixed before experiments, and drivers were observed by observers or through recorded videos. StudentLife and SNAPSHOT are datasets from the less controlled environment (university) (Sano et al. 2018; Wang et al. 2014).

The laboratory is the most common environment to set up stress data. WESAD, SWELL-KW, and DEAP are examples of benchmark datasets set up in the laboratory. WESAD contains data from 15 participants in the laboratory environment. All participants were graduate students at the same faculty. They were not pregnant and did not smoke heavily. None of these students had any mental issues or cardiovascular problems. Their ages were 27.5 ± 2.4 years old. There were twelve male participants and three female participants. The physiological data included GSR, BVP, ECG, EMG, respiration data, skin temperature, and a 3-axis accelerometer (Schmidt et al. 2018b). The stress condition was simulated using Trier Social Stress Test, and emotional data was captured using Positive and Negative Affect Schedule, State-Trait Anxiety Inventory, and Short Stress State Questionnaire.

The experiment of SWELL-KW dataset contains two stressors: time to finish the deadline and disturbance from unrelated emails. Workers must work under normal conditions and two stressors to gain data. That dataset contains information on computer interactions (with keyboard, mouse and applications), facial expressions, body stands and physical features (heart rate, EDA) (Koldijk et al. 2014). The affective data was gathered using the NASA Task Load Index, Rating Scale Mental Effort, Self Assessment Manikin Scale and the author's self-defined scale log to represent stress levels.

The DEAP database collected EEG, EDA, BVP, respiration data, and skin temperature of 32 participants when they watched 40-minute music videos (Koelstra et al. 2011). Self-assessment manikins were used as an indicator of ground truth.

Driving is also an exciting environment for researchers to set up datasets. The Driver Stress Data dataset gathered GSR, ECG, respiration data, and EMG data for 9 drivers (Healey 2000). The drivers joined a real driving experiment when they faced several specific conditions of the roads which associated with several emotional states.

Another real driving dataset is AffectiveRoad. The state of the environment, the car's speed, and physiological data (EDA, HR delivered from BVP, skin temperature) were collected (Haouij 2018). Both datasets categorised parking, driving on the freeway, and driving in the city as no stress, moderate stress, and immense stress, respectively (Haouij 2018; Healey 2000).

StudentLife and SNAPSHOT were collected from everyday life environments. The Student Life is the dataset collected from 48 students from Dartmouth College during ten weeks (Wang et al. 2014). Among 48 students were 30 undergraduate students and 18 graduate students; 38 were males, 10 were females, 23 were Caucasians, 23 were Asians, and 2 were African Americans (Wang et al. 2014). The researchers gathered activity data (stationary, walking, running, driving, cycling), conversation data (the amount and duration of conversations), sleep data (time go to bed and get up, duration of sleep) and location data (where did the students stay and for how long). They also collected emotional data using mobile EMA and surveys from SurveyMonkey's website. The emotional data include depression, stress, loneliness, and flourishing status (Wang et al. 2014).

Another real-life data set is SNAPSHOT which collected skin conductance, acceleration data, and skin temperature from 66 undergraduate students (Sano et al. 2018). Among these students, 47 were males and 19 were females. Their ages were 20.1 ± 1.5 years. They also added personality types, phone usage data, and sleep data. The participant needed to fill out questionnaires and write diary entries to recognise emotional level (Sano et al. 2018).

This PhD thesis used the Driver Stress Data, WESAD and SWELL-KW as benchmark datasets. The Driver Stress Data is one of the first stress detection datasets with physiological data. That dataset is small, so it is suitable for testing the developed algorithms of this thesis and avoiding huge computational expenses. The WESAD and SWELL-KW as benchmark datasets are newer with significant amounts of timestamps that can be used to validate the algorithms.

2.5 Simple classification methods for classifying stress from physiological datasets

Scientists use several ways to analyse the stress detection data sets. They can apply statistical approaches, for example, statistical tests to identify significant or maybe potentially meaningful correlations between physiological features and the ground truth of stress or machine learning classification methods to detect stress. The three most common machine learning methods used for stress detection described in the literature are Support Vector Machine (SVM), Random Forest and k nearest neighbors (kNN).

The first practical and realisable approach to detect stress was proposed by Healey and Picard (2005). They suggest an approach to recognise the stress status in a driving environment and collect ECG, EMG, GSR and respiration data from 24 drivers with at least fifty-minute duration per driver. Before that early research on stress detection using wearable devices and stress detection processes were complicated since they required special and expensive tools. These tools had the potential to distract drivers. There were three levels of stress defined in that experiment: high level when drivers go to a city, medium level when they drive along a highway and low level when they rest. Twenty-two features were calculated using a five-minute interval. A five-minute interval was applied because this is the standard to measure HRV with the spectrogram (Healey and Picard 2005). They applied both machine learning and statistical methods and presented two conclusions: that they can recognise the three-class stress level with an accuracy of 97% and that skin conductivity, and heart rate features that have the highest correlation with driving stress (Healey and Picard 2005). Since then, many further attempts to detect stress have been carried out.

That initial research still has limitations regarding the selected classification algorithm and calculated correlation process. Two experiments were deployed in that research; the first tries to spot stress using Linear Discriminant Analysis (LDA) and the second tries to find how well attributes reflect stress. Choosing LDA for the classification task may not be a strong point of that research, even though the research reached an accuracy of 97%. LDA performs well when the decision boundaries are linear, but when the data has nonlinear decision boundaries, kNN may perform better (James et al. 2013). LDA also does not perform well on an unbalanced dataset (Xing et al. 2019). However, natural-world and medical datasets such as stress datasets have unbalanced characteristics and nonlinear decision boundaries (Abdulsadig and Rodriguez-Villegas 2024; Haidri 2023). Models trained by LDA may overfit and not be practical in a real-world scenario (Liu and Gillies 2016).

The method the authors applied to detect the relationship between stress and physiological traits also has limitations. That experiment used a stress metric with data deduced from a recorded video. That stress metric was measured every second using a rule defined by Healey and Picard (2005). For example, if the driver showed no stress signal, then the value of the stress metric at this second was 0; if the driver drove straight and only turned around, then this value became 1, and if the driver turned the steering wheel, adjusted their gaze and turned his/her body then the value was 3. The authors built a correlation coefficient to show the relationship between physiological features and the stress index. However, that approach has two limitations. The first is that many things may affect humans' reactions when driving on open roads, and then that stress metric does not perfectly reflect stress level. The second is about synchronisation; the stress metric was measured every second while skin conductivity and EMG were measured every 1.4 seconds and 30 million seconds (Healey and Picard 2005). Without the synchronised process before the calculated process of correlation, then a time difference issue will happen.

2.5.1 Statistical approaches

Wang et al. (2014) and Sano et al. (2018) applied statistical approaches to find relationships between some features of stress data with stress. Using the Pearson correlation and Student Life dataset, Wang et al. (2014) declared that students with more conversation in the afternoon have less stress than other students. In a similar approach, Sano et al. (2018) found that EDA and skin temperature correlate with the value of some stress-indicated surveys. That approach provides excellent insight into detecting relationships between stress and features. However, these two pieces of researches are not comprehensive stress detection researches.

2.5.2 Hidden Markov Model

Vildjiounaite et al. (2018) made a stress detection model using the Hidden Markov Model in two experiments for 28 people over a few days for each person and two people over 200 days. Among 28 people, there were 4 males and 24 females. Their ages ranged from 20 to 47 years old. None of them have any mental issues or serious diseases. Collected data is the type of mobile phone applications users interact with and the duration of the interactions. The authors used a time-window technique on the data and applied a Hidden Markov Model to classify the windows into output types. In this unsupervised model, observations are inferred features from time window data, and hidden states are the model's output classes. That model's accuracy was not very high (just 59% and 70% for the first and second experiments, respectively). Still, that research had a special advantage because the data did not need to be labelled. This advantage can contribute more to handling labelling data when the collected data is massive. However, that research had a considerable disadvantage: since mobile phone usage data was used, no stress data was collected when people did not use the phone.

2.5.3 Logistic regression

Logistic regression is a straightforward machine learning algorithm, but it is still useful for detecting stress in some cases. Zubair et al. (2015) produced IoT wearable devices that record EDA from 12 participants aged 23 to 56. They also used an accelerometer to recognise movements and improve the quality of the stress detection process; they utilised logistic regression to predict stress to get an accuracy of 91.66% (Zubair et al. 2015).

That approach's limitation is that logistic regression only works well when there is a linear separation between groups of classification labels (James et al. 2013). A high accuracy classification value did not mean this was a good approach since the authors did not provide detailed information on wearable devices and the data-collecting process. Furthermore, they did not describe how long they had collected data from people or if there had been any pre-processing process.

2.5.4 Bayesian network

Hong et al. (2012) collected EDA, respiratory data, skin temperature, and accelerometer data from 12 people. They developed methods to detect stress for three physical activities: sitting, walking, and cycling. With Bayesian network, they achieved an accuracy of 82%. The authors also declared that their Bayesian network is a simple approach, and they need to deploy more complicated algorithms such as SVM to make their research more practical (Hong et al. 2012).

2.5.5 Fuzzy logic

Sierra et al. (2011) collected only two physiological features, EDA and HR, of 80 females. They collected features for two scenarios, which are 3-5 minutes duration and 10 minutes duration, applied fuzzy logic and achieved the accuracy of 95% and 99.5% respectively (Sierra et al. 2011). The authors declared using hyperventilation and talk preparation to provoke stress. There is no clear evidence that these two methods can effectively induce stress. The authors also confirmed that their method may be applicable in some specific scenarios.

2.5.6 K nearest neighbors

Wang et al. (2013) used a database from an experiment by Healey and Picard (2005). They gathered HRV and then applied trend-based and parameter-based mechanisms. Features were inferred from long HRV and 5-minute HRV in these two methods, respectively (Wang et al. 2013). In that approach, the feature selection method was kernel-based class separability (KBCS), and the feature reduction methods were principal component analysis (PCA) and linear discriminant analysis (LDA) (Wang et al. 2013). They applied k nearest neighbors (kNN) for their approach and reached an accuracy of 97%. In a real-life experiment on stress, there may be more data that can be labelled as no stress than data that can be labelled as stress. In this case of the unbalanced dataset, kNN may not perform well (García et al. 2008). The high accuracy in that experiment may not happen again in other experiments, and kNN may not be the most suitable stress detection algorithm.

2.5.7 Random Forest

Gjoreski et al. (2016) used an accelerometer to detect participants' movement type to classify stress or physiological reactions caused by movement. They used a device that gathers BVP, EDA, HR information, RR intervals, skin temperature, and acceleration data and calculated 63 features from these data (Gjoreski et al. 2016). They deployed stress detection tests with Random Forest in the laboratory and real-life environments. They labelled the stress using STAI in the first environment and EMA combined with mobile phones in the second environment (Gjoreski et al. 2016). In the laboratory, two class classifications (no stress and stress) and three class classifications (no stress, moderate stress and immense stress) were performed with accuracies of 83% and 72%, respectively.

Gjoreski et al. (2016) set up an accelerometer to detect the type of movement to perceive physiological measurements change because of physical activities or stress, and this unique feature provides context to the approach. This context information differentiated between changes in physiological features caused by stress or other factors such as eating, exercise, and hot weather (Gjoreski et al. 2016). Contextual information was an exciting point of that research since it made it more practical to real life. Contextual information has the potential to boost the accuracy of the approach. In a real-life environment, they gained accuracies of 76% and 92% for the no context and context cases, respectively (Gjoreski et al. 2016). That Random Forest approach achieved quite good performance. However, Random Forest's approaches must also be carefully considered since the running time is high, sometimes making the approaches impractical for real-time classification.

2.5.8 Support Vector Machine

Support Vector Machine (SVM) is one of the most common algorithms for detecting stress. The accuracy of this approach sometimes reaches a very high value. In research of Castaldo et al. (2017), authors collected only HRV from 42 students from University Frederico II on two days. The first day is verbal exam day, which is a stressful day, and the second day is a rest day. They investigated the stress detection ability of HRV features shorter than 5 minutes; they experimented with SVM to see stress and achieved the highest accuracy of 64%. They concluded that the short version of HRV (less than 5 minutes as reported in the research) could still investigate stress in real life experiments, however the accuracy is not high.

Hernandez et al. (2011) showed a solution that uses skin conductance and SVM to detect the stress condition of nine call centre employees in one week; during that time, the employees answered 1500 calls. The authors achieved an accuracy of 78%. Hovsepian et al. (2015) suggested a model named cStress, which uses ECG and respiration data combined with SVM to detect stress. Training data was collected from 21 people, and testing data was collected from 26 people in the laboratory and 20 in a week-long field study. They applied a 1-minute window of ECG and respiration data. This short time frame and the accelerometer recognises no physical activity. This can be used as input of their approach (Hovsepian et al. 2015). In the laboratory, they achieved a recall of 89%, and in a field study, they achieved an accuracy of 72% (Hovsepian et al. 2015). In an experiment, five students' PPG and EDA were collected, then joined a TSST test to cause stress (Sandulescu et al. 2015). SVM was applied to predict stress, and their accuracy for five students ranged from 73.26% to 83.05% (Sandulescu et al. 2015).

Stress during driving is one of the topics that has gained much attention from researchers. In this subfield, SVM achieves huge success with really high-accuracy implementations. Lee et al. (2016) made wearable gloves to collect PPG and a few other sensors, such as an accelerometer, gyroscope, and magnetometer. They set up a driving environment in their laboratory to collect data from 28 people. The driving environment contains several sessions with three landmarks: city, highway, urban areas and weather conditions like rainy, cloudy and sunny. A driving behaviour survey was used for each participant before and after driving sessions to gather stress information. They recognised stress with SVM and obtained an accuracy of 95%. With the driving data set of Healey and Picard (2005) which contains ECG, EDA and respiration data, Chen et al. (2017) applied PCA to extract features and Sparse Bayesian Learning to select the best combination of features. They used SVM and got the most successful result with an accuracy of 99%.

SVM consistently underperforms when the data is noisy (Ray 2019), and the researcher needs to consider this case. Experiments of the three research of Castaldo et al. (2017), Hernandez et al. (2011) and Hovsepian et al. (2015) were deployed in a real-life environment where many factors can easily interrupt physiological data. In an experiment of Castaldo et al. (2017), data was collected from students, and even in one phase, it was collected during a verbal exam. Hernandez et al. (2011) collected data during working hours, and Hovsepian et al. (2015) gained data from 7 days of people in their natural field environment. Data from their three environments can easily be noisy. So this may be the main reason for the reduced accuracy of the three SVM's research of Castaldo et al. (2017), Hernandez et al. (2011) and Hovsepian et al. (2015). Meanwhile, the last two research studies of Lee et al. (2016) and Healey and Picard (2005) were deployed in the laboratory and a controlled driving environment, improving accuracy significantly.

2.5.9 Multiple approaches

When researchers build stress detection systems, they do not know the best algorithms to apply. The solution to this practical issue is to deploy several algorithms and choose the one with the highest classification results.

Muaremi et al. (2014) used physical and physiological features during sleeping time to estimate the stress levels of 10 participants. There are three stress levels in this experiment which are low, moderate and high. The physical features are the number of times a participant is awake, the number of times a participant changes sleep positions, upper body activity level, arm activity level and sleep duration (Muaremi et al. 2014). The physiological features are ECG, respiration data, HRV, body temperature, and EDA. All participants undertook Hajj pilgrimage; 7 were males, and 3 were females. Their average age was 41. The authors applied logistic regression, SVM, kNN, Random Forest, and neural networks to classify data and recognised that SVM has the highest accuracy of 71% (Muaremi et al. 2014). That research presents an exciting topic of detecting stress using nightly sleep patterns. This topic can improve the quality of health care. However, there are some limitations to that research. The authors collected whole-night data and considered it as sleeping data, but people are sometimes awake at night. So, in this case, collected night data does not always mean providing sleep patterns to detect stress. The values of physiological features of waking time and sleeping time are different, which may affect the research results. Besides that, the experiment is only for a small number of participants, and they are all participants who undertook Hajj pilgrimage. If the authors want to make it more complete, they need to increase the number and diversity of the participant population.

Sierra et al. (2010) presents a scheme to recognise stress using heart rate data and GSR. The scheme contains four stages: relaxation, hyperventilation, speech, and relaxation. They defined two new parameters to determine the stress detection process: true stress detection rate and true non-detection stress (Sierra et al. 2010). These two parameters demonstrated an accuracy of correctly recognising stress and non-stress conditions. In their approach, the results of two machine learning mechanisms, kNN and Fisher discriminant analysis using two assessment parameters, were compared (Sierra et al. 2010). That approach also considered having higher numbers of participants (80) to be tested than other systems. The authors also confirmed that the effectiveness of their two new parameters needs to be validated using public datasets.

Ramos et al. (2014) extended their previous research in 2012 to detect stress (Hong et al. 2012; Ramos et al. 2014). Researchers gathered the HR, skin temperature, breathing rate, acceleration, and GSR of 20 participants over 160 hours. There were 10 male and 10 female participants, ages 18 to 38. Among them, 10 were Caucasians, 7 were Asians, 2 were African Americans and 1 was White/Hispanic. Their height ranged from 5’4" to 6’3", their weight ranged from 90 to 175 pounds, and their body mass index ranged from 16 to 25. None of the participants had any issues with cerebrovascular, cardiac or pulmonary arrest (Ramos et al. 2014). They presented a clustering and classification system that effectively detects stress and does not require expensive tools in real-life conditions. In the first process, participants joined a data recording process when they faced some stressors and joined some physical activities. The stressors included random noises, cold water, and mathematical tests. In the second step, logistic regression and Naive Bayes were used to classify stress (Ramos et al. 2014). The authors stated that a combination of activity recognition and stress recognition is a good approach to stress detection. They declared that labelling the activities using supervised learning methods is unsuitable for real-world environments, and they trained an unsupervised learning algorithm to differentiate between high-level and low-level activities. The accuracy of that method is 65%, but it can be practical in daily life environments.

Plarre et al. (2011) built a two-stage stress-detection process. In the first stage, they collected ECG and respiration data from 21 people, simulated stressful conditions using public speaking, testing arithmetic and cold pressure, and gathered data from five questions to label the stress data (Plarre et al. 2011). SVM, J48 decision tree and J48 with Adaboost were applied to detect stress, and the accuracy was 89.17%, 87.67% and 90.17% respectively (Plarre et al. 2011). The authors also stated that the normalised data would improve the quality of classification approaches. In the second stage, they used the Hidden Markov Model to observe the output of the first step and predict the probability of stress at particular moments (Plarre et al. 2011).

In recent research, Can et al. (2019b) collected heart signal, EDA and accelerometer signals from 21 people using smartwatches under normal daily conditions. Among 21 participants, 18 were males and 3 were males aged around 20. All participants need to join a nine-day algorithmic programming contest camp. Each day, the participants needed to join a lecturer and then take a problem-solving test. On the ninth day of the contest, participants needed to join a contest, and they solved challenging questions in a limited time. The score of the participants was also displayed on the screen so that it would cause more stress for the participants. NASA-TLX questionnaires of all participants were collected in all tests and contests to measure their stress levels. The authors designed suitable removal artifacts and feature selection for real-life and continuous conditions. Can et al. (2019b) applied many standard classification methods in the literature (PCA and LDA, PCA and SVM, kNN, Logistic Regression, Random Forest, and Multilayer Perception) to the dataset. They gained the highest accuracy of 97.92% with Random Forest.

Tiwari et al. (2023) used feature selection methods such as ANOVA F-test, Mutual Information (MI), Recursive Feature Elimination (REF), and Correlation coefficient. They tested many machine learning models with feature collections to detect stress. The machine learning models were K-Nearest Neighbour, Linear Discriminant Analysis, Random Forest, Logistic Regression, Gradient Boosting, and Support Vector Machine. They declared that Random Forest was the best classifier due to the model's performance. If the authors can provide the running time, especially of the feature selection process, then the research can be much more convincing. Feature selection is time-consuming and can significantly affect the model's effectiveness.

The two pieces of research of Plarre et al. (2011) and Can et al. (2019b) already provided practical stress detection solutions in daily life environments using non-invasive and unobtrusive wearable devices combined with complete data pre-processing progress. These devices can be easily worn in daily life environments. Plarre et al. (2011) used AutoSense wearable suite with six sensors that can continuously transfer data via a wireless internet connection for 72 hours. In the case of the experiment of Can et al. (2019b), they used smart wrist-worn wearables. Two researches also have detailed pre-processing processes. Plarre et al. (2011) concentrated on normalizing the data. Meanwhile, Can et al. (2019b) investigated removing artifacts and noise caused by hot weather or physical activities. Abnormal values of hot weather or physical activities can greatly affect recorded values of electrodermal activity and, as a result, can change the accuracy of the stress detection process (Can et al. 2019b).

2.6 New methods to analyse physiological features

With the rapid development of mobile devices, researchers can collect more ground truth data on stress via ecological momentary assessment (EMA) (Wang et al. 2014). With more labelled data collected, more advanced machine learning approaches such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) can be used to detect stress. These deep learning approaches can turn massive amounts of data into insights (Waring et al. 2020).

According to Woodward et al. (2020), advanced machine learning approaches can help recognise intricate patterns in the data, a revolution in affective computing and stress recognition. Feature extraction and dimension reduction are always essential to improving machine learning methods. With traditional machine learning approaches, feature extraction and dimension reduction are computationally expensive (Bota et al. 2019; Zhang et al. 2020a). The deep learning methods also have an excellent ability to learn hierarchical feature representations that can decrease the computational expense of the feature extractions and dimension reduction compared with standard machine learning methods, especially in emotion recognition (Bota et al. 2019; Zhang et al. 2020a).

2.6.1 Deep learning

Several deep learning approaches to recognise stress and emotion, such as deep belief network (DBN), Convolutional Neural Networks (CNN) and Long short-term memory (LSTM), are in the literature.

A deep belief network is a generative deep learning model that consists of several Restricted Boltzmann Machines. Hassan et al. (2019) used an unsupervised deep belief network (DBN) and Fine Gaussian Support Vector Machine (FGSVM) to detect emotion levels and get an accuracy of 89.53%. The DBN produced deep-level features from EDA, PPG and Zygomaticus Electromyography (zEMG) sensors (Hassan et al. 2019). Then inferred features from EDA, PPG and zEMG were united with these deep-level features to build a feature fusion vector as the input to FGSVM to recognise emotion (Hassan et al. 2019). They used the DEAP dataset to test their model. The authors showed that their SVM combined with generated features performed better than the standard SVM or Naïve Bayes approaches (Hassan et al. 2019). This method has the advantage of applying a generative model to create more features. In the case of stress and emotion detection, this advantage becomes more important. Generative deep learning models can generate more extensive data sets with various features from the original small benchmark datasets with few physiological features. However, the authors did not present why they chose the Gaussian kernel for their approach.

Convolutional Neural Networks is a deep neural network with many successes in computer vision and was recently introduced in stress and emotion recognition. Song et al. (2018) used a combination of CNN and spectral theory called graph convolutional network. They considered EEG channels as nodes and connections between these EEG channels as edges of the graph (Song et al. 2018). They showed that the methods perform better than many other machine learning algorithms, such as SVM and deep belief networks and their method's accuracy was 90.4%.

The authors deployed their experiment on SEED and DREAMER datasets, customarily utilising EEG for emotion recognition. Using Graph Convolutional Neural Networks (GCNN) can also show which EEG channels can contribute the most to the recognition process (Song et al. 2018). However, these two datasets only have minimal features. According to the authors, the range of EEG of two datasets is still tiny, so the deep learning algorithm is less effective.

Granados et al. (2018) applied a deep convolutional neural network in their research. That network is a specialised design of Pyakillya et al. (2017) and can gather data from physiological measurements. The authors used a fully connected neural network to detect mood from ECG and GSR signals. The authors apply the Pan–Tompkins algorithm from the ECG to detect QRS complexes. Then, from inter-beat intervals, statistical features were calculated and used as input for the deep convolutional network. The authors made several comparisons to show the effectiveness of their model. Nevertheless, there are currently some limitations to that deep model. Their first comparison is not persuasive since they only compared their model with classical classification models such as Random Forest, Ada Boost and linear discriminant analysis (LDA) and with just one multilayer perceptron, which is not described in detail. In another comparison, the authors demonstrated that their model achieved better accuracy than other deep learning models; however, these models were deployed on different datasets. This model of Granados et al. (2018) was based on a CNN model of Pyakillya et al. (2017) without any modification. However, even the authors of the original CNN model are still concerned that their design is incomplete. Pyakillya et al. (2017) thought that their model would be much better if they used LSTM or RNN instead of CNN, since their model works with time series data. The authors also thought generating more GAN data could help their model perform better. The basic model of Pyakillya et al. (2017) only achieved the best accuracy of 86%, and this led to low accuracy of the extended model of Granados et al. (2018). The model of Granados et al. (2018) only has accuracy for detecting arousal and valence at just 76% and 75%, respectively. Huynh et al. (2021), Alshamrani (2021), and Gupta et al. (2021) classified stress from the WESAD dataset using CNN models. They all gained promising results. The research of Huynh et al. (2021) utilised a 60-second time window with a dataset frequency of 700Hz. Thus, each time window can contain many timestamps associated with physiological states. In the WESAD dataset, each timestamp also has a value for one type of emotional state. So if the window was large, there were many types of emotional and physiological states within it. This can make it very hard to capture the pattern of what physiological states cause a specific emotional state. And as a result, it can reduce the model's performance. Alshamrani (2021) and Gupta et al. (2021) did not provide clear descriptions of the CNN models they used.

Li and Liu (2020) trained a 1-dimensional CNN and a deep neural network to detect emotion levels. The authors used two datasets collected from 15 participants using chest-worn and wrist-worn devices. The chest-worn sensors gathered ECG, GSR, EMG, skin temperature, respiratory rate and data from the 3-D axis accelerometer; meanwhile, wrist-worn data sensors collected BVP, GSR, skin temperature and data from the 3-D axis accelerometer. They applied the 1-dimensional CNN model with the first dataset and the deep neural network with the second one. The authors demonstrated that their deep learning approaches could overcome the disadvantage of needing handcrafted features of traditional machine learning methods. They showed that both 1-dimensional CNN and the deep neural network perform better than traditional machine learning methods (LDA, AdaBoost, Random Forest) in both datasets. 1-dimensional Convolutional Neural Networks and the deep neural network attained an accuracy of 99.80% and 99.65% when detecting stress; meanwhile, the two best performances of traditional machine learning models were only 93.12% and 88.33% (LDA and Random Forest), respectively (Li and Liu 2020). They also confirmed that the 1-dimensional CNN performed better than the deep neural network. However, in their experiment, the 1-dimensional CNN handles more intensive chest-worn sensor input data than the deep neural network's wrist-worn data. The chest-worn sensors also gather data at a much higher frequency, and as a result, the data is higher quality. According to the authors, the limitation of that research is the lack of data. Lisowska et al. (2021) also applied a 1-dimensional CNN to detect stress from the WESAD dataset. The model's performance was not very impressive. The author used raw physiological features rather than handcrafted features. The raw physiological data typically contain substantial noise and provide little information for use as handcrafted features in a classification model. They also utilised a relatively large time window (60 seconds) with a high-frequency dataset (700 Hz). This leads to fewer samples being used for training and to a high computational cost. This also led to cases in which the model did not capture the short-term pattern of physiological data. For stress detection, physiological features can change rapidly, and if the model cannot capture these rapid changes, it cannot be a good classification model. The raw data and the large time window may be the reason for this model's not high performance.

Some authors used hybrid models for stress and emotion recognition. The best combination is CNN and LSTM, which achieved a very high result and was confirmed by both Kanjo et al. (2019) and Rastgoo et al. (2019) to have great potential in handling physiological features. However, the combination of LSTM and the linear model of Xing et al. (2019) can not get good performance.

Kanjo et al. (2019) applied a hybrid approach that combined LSTM and CNN to process the EnvBodySens dataset, which is a dataset with heart rate, EDA, body temperature, movement, and other location data such as noise level, UV level and air pressure. Data were collected from 40 female participants (average age of 28) who walked for less than 45 minutes in almost the same conditions (11 am and 20 °C). The authors confirmed that deep learning is more effective in stress and emotion recognition than common machine learning approaches, and the combination of CNN and LSTM performed better than CNN. The authors stated that the prediction accuracy of the hybrid approach is high, with an average value of 94.7%. Meanwhile, the accuracy of CNN and typical neural network were 78.6% and 72.9% respectively (Kanjo et al. 2019). However, the dataset for that research is small, according to the authors (Kanjo et al. 2019). Large-scale research with a more comprehensive dataset is necessary to confirm the research results.

Rastgoo et al. (2019) set up the same hybrid approach on their own driving dataset and achieved an average accuracy of 92.8%. Xing et al. (2019) suggested a system of emotional recognition using Stack Auto Encoder (SAE) and LSTM. The SAE is a linear mixing model for gathering EEG signals from the EEG source (DEAP dataset in that experiment). They also used hybrid LSTM-RNN to forecast emotion mode. They achieved a precision of 81.10% and 74.38% for valence and arousal, respectively (Xing et al. 2019). Both the research of Rastgoo et al. (2019) and Xing et al. (2019) use only one physiological feature: ECG in the first one and EEG in the second one. There will be many effects, such as movements or problems of data collection devices, that will seriously affect these two pieces of research.

2.6.2 Stress detection using forecasting methods

Machine learning and statistics are the most common tools for predicting the future. These techniques also can be used to classify stress. The most common machine learning forecasting methods are RNN and LSTM. The most common statistical ones are ARIMA and VAR. These machine learning and statistical models are the most commonly used stress prediction models.

Auto-Regressive Integrated Moving Average (ARIMA) is an autoregression model for predicting time series data. An autoregression model predicts the current value using linear combinations of past values (Hyndman and Athanasopoulos 2018). The Integrated part utilized differencing to make the data stationary. Differencing subtracts the current value from the previous observation value (Yamak et al. 2019). The Auto Regressive part of the model is calculated by sum of all multiplication of previous observation values of time lags with corresponding coefficients, which are estimated by the model (see equation 2.1) (Yamak et al. 2019). Auto-regressive time series has values that only depend on the past values. The Moving Average part is calculated by the sum of all multiplication of the previous error of forecasted time lags with a coefficient of these lags (see equation 2.2) (Yamak et al. 2019). Moving Average time series will have value only depending on the forecasted time lag's error. Formulating a time series with both the Auto Regressive part and the Moving Average part is equation 2.3. Three parameters d, q, p of the ARIMA model represent the number of differences that need to be applied to convert from a non-stationary to a stationary time series, the number of lags of the Auto-Regressive formula, and number of elements in the Moving Average window (Yamak et al. 2019).

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_1 \tag{2.1}$$

$$Y_t = \alpha + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_2 \epsilon_{t-2} + \theta_q \epsilon_{t-q} \tag{2.2}$$

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_2 \epsilon_{t-2} + \theta_q \epsilon_{t-q} \tag{2.3}$$

where:

Y_t : is the value at time lag t

β_i : is the coefficient of time lag $t - i$

θ_i : is the coefficient of lagged forecast $t - i$

ϵ_i : is lagged forecast error i

α : is the intercept term

Like ARIMA, Vector Autoregression (VAR) is also an autoregression model. ARIMA is a model of univariate time series, and VAR is a model of multivariate time series (Hyndman and Athanasopoulos 2018). VAR also needs differencing to make the data stationary. This multivariate forecasting model has only one parameter, p (lag order); the value of p is the optimal value which can be chosen by one of four methods: least Bayesian information criterion, Hannan-Quinn criterion or Final Prediction Error, Akaike's Information Criterion (Assaad and El-adaway 2021; Hyndman and Athanasopoulos 2018). Bayesian information criterion is a commonly used method for VAR. VAR can present relationships between K multivariate time series with p lag orders using (2.4) (Assaad and El-adaway 2021).

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + e_t \tag{2.4}$$

where:

Y_t : $(y_1, y_2, \ldots, y_K)$ value of K time lags with index t of K time series

A_i : is the coefficient matrix

e_t : is the error term

Recurrent Neural Networks is a neural network that can remember past data and predict future data. Long Short-Term Memory is an RNN extension that helps handle the vanishing gradient issue of RNN (Yamak et al. 2019). Each LSTM's cell has an input gate, a forget gate and an output gate, which can have some actions with information such as forget, remember, output or pay attention (Yamak et al. 2019). Gates update the cell states and hidden states store values for the following step (Yamak et al. 2019).

Suhara et al. (2017) were pioneers in this research field when they tried to use LSTM and SVM to classify the depressed mode of self-reported data from the past few days. Researchers deployed that experiment to answer two research questions. The first one is whether a depressed mood can be detected. The second is how many days of the past data are required at a minimum to ensure correct classifications. The researchers combined time-series feature values to one feature vector so a SVM could handle the temporal data (Suhara et al. 2017). Their experiment had data of 2,382 people for a total of 345,158 days. Suhara et al. (2017) predicted the depressed mood using LSTM and SVM with high accuracy; their experiment had the area under the curve of Receiver Operating Characteristic (ROC) of 0.886 and 0.837 for LSTM and SVM, respectively (Suhara et al. 2017).

The investigation also found that 14 days of previous data is enough to detect depression.

That research can be an inspiration for emotion forecasting research with massive data and detailed analysis. The most significant limitation of that research lies in the source of data. That research utilised self-reported data, so participants must provide their mood levels daily to predict the future. According to Suhara et al. (2017), previous mood data is also the best feature to predict future mood levels. This particular characteristic can cause huge drawbacks for that system. If only one day's data was missed, the model could not predict mood for two weeks (Taylor et al. 2017). The model can not achieve good results even though it only handles a simple task: past mood features to forecast future mood levels. It is simpler than other tasks, such as using many physiological features to predict mood levels. Self-reported data cannot be as accurate as data collected using wearable devices since it relies on personal perspectives. So, it is critical to deploy future stress prediction research that utilises wearable devices.

Both Taylor et al. (2017) and Yu et al. (2019) utilized Multitask Learning deep neural networks and the SNAPSHOT dataset to forecast stress. However, the two groups achieved limited results, with the highest accuracy being 82% and the highest F1 score being 0.74. Taylor et al. (2017) forecasted the next day's stress and mood condition based on today's data. Stress is objective and differs between people. Taylor et al. (2017) stated that their approach can build a personalised model for each person but still use a generalised model for many other people. The authors used k-means clustering to group users by age and personality; the authors considered the prediction of each group as a single task of Multitask Learning. The method's highest accuracy was 82%. Yu et al. (2019) also applied to multitask learning with several deep learning models to detect mood and stress. They also had two schemes to build tasks: either considering each participant as a task or grouping participants by gender and personality types and considering each group as a task. The authors declared that multitask neural network is their best model with an F1 score of 0.71, 0.74, and 0.66 for mood, health, and stress.

There are some limitations to both pieces of research. Taylor et al. (2017) can only predict stress for one day, and the authors did not mention whether the accuracy could improve if more past data were used. Yu et al. (2019) can predict emotional levels up to 7 days in the future; however, the way they consider predicting the stress of each participant as a task seems unrealistic when a dataset has many participants. Both groups did not provide a detailed explanation of how they grouped the participants. This information is essential in building a good design of tasks for the models. Poor design when creating tasks may be the reason for the low performance of both pieces of research.

Traditional machine learning models were also applied to forecast stress. In an experiment, Umematsu et al. (2020) used Random Forest to predict stress and compared results on a dataset of 201 college students and 39 workers from a high-tech company. The authors admitted that that experiment does not generalise to the whole population (Umematsu et al. 2020). However, Umematsu et al. (2020) did not clearly explain why and how they use Random Forest, a classification method, to forecast stress.

LSTM is a dominant algorithm in stress forecasting research areas since it has been used in many research pieces. According to both experimental results of two pieces of research Umematsu et al. (2019b) and Suhara et al. (2017), LSTM has better prediction power than standard classification algorithms such as SVM. These two groups modified SVM for forecasting purposes. In the research of Acikmese and Alptekin (2019), the authors utilised LSTM, CNN, and CNN-LSTM to predict stress in the Student Life data set. However, the best accuracy is not high, 62.83% with the LSTM model. Umematsu et al. (2019b) collected EDA, skin temperature and accelerometer data from 142 people with a total of 2276 days of SNAPSHOT dataset to predict the stress. The authors applied logistic regression, SVM and LSTM for data from 7 previous days, and the time-series analysis algorithm outperformed the first two approaches with an accuracy of 83.6% (Umematsu et al. 2019b). In another study, these researchers applied the same methods but with only data from morning to afternoon (10 am to 5 pm). It achieved a high accuracy result of 86% (Umematsu et al. 2019a).

In stress classification research using prediction methods, the current state-of-the-art methods are RNN and LSTM. Currently, these classical and old deep learning methods are not state-of-the-art time-series-analysis methods. RNN and LSTM were introduced in 1986 and 1996. Recently, many efforts have been made to propose new forecasting algorithms and surveys to combine statistical and machine learning models to choose the best forecasting approach. This trend can create new methods for stress forecasting.

Makridakis et al. (2018b) compared eight statistical algorithms with 18 machine learning algorithms (including LSTM and RNN) with the same benchmark dataset. The result is that the seven methods with top performance are statistical ones. The last statistical approach can still be ranked in the twelfth position. The current state-of-the-art usually combines machine learning and statistical perspective. Many researchers pointed out the limitations of all current machine learning methods during several versions of a famous time-series analysis competition named M. This competition's winners and successful methods are only statistical ones or combined machine learning techniques and statistics. Makridakis et al. (2018a) confirmed that the combination methods outperform other types of methods in the M4 contest, with 12 of the 17 methods with the best performance being combined methods, and the rank of machine learning methods being low compared to statistical and combined methods. Nevertheless, the authors also believed that these machine learning algorithms are standard, and with modifications from the standard ones, the performance will improve greatly. Oreshkin et al. (2019) proposed a new model adopting backward, forward residual links associated with a very deep neural network. They demonstrated that their method raises the accuracy by 3% to compare with the winner of the M4 competition.

2.7 Research gaps

There is considerable research on the use of wearable devices to detect stress. However, there are research gaps that researchers need to consider. All of the approaches reviewed in this thesis have limitations. Some of these approaches use machine learning classification algorithms to detect stress without a complete comparison with other classification methods. Some of these approaches use statistical techniques such as ANOVA and Pearson's correlation to assess the relationship between features and stress. Some of these approaches use traditional machine learning methods combined with raw physiological data or handcrafted time series features to detect stress. Many researchers also did not utilise benchmark datasets for their research.

Typically, authors only apply one to three classification methods to detect stress. They declare that the methods with the highest accuracy are the best. Nevertheless, many other classification methods, which may have greater accuracy, have never been tested before. According to the No Free Lunch Theorem of

Wolpert (1996), if researchers make no assumptions about the data, there is no guarantee that a model works better than another. Researchers must apply many techniques to choose the best methods for particular data. They also need to provide comparisons between the methods.

Some stress detection approaches, such as Wang et al. (2014) and Sano et al. (2018), only concentrate on the relationship between each dataset feature and stress. The two groups of authors are concerned only with the relationship between each dataset's feature and stress. Statistical tests are good tools for understanding the data. The actual purpose of these two papers is to introduce new datasets; they do not constitute comprehensive research on stress detection.

Current research on stress mainly applies traditional machine learning methods to classical inputs. The methods for stress detection have already been presented in the literature review, and most are classical data-analytic approaches that apply traditional machine learning methods. Classical machine learning methods tend to perform better on small and medium-sized datasets. Meanwhile, for large physiological datasets, the effectiveness of these classical models can be reduced. Deep learning can gain better insights into data since deep learning models have the potential to detect complex patterns from the data. Classical inputs for stress detection methods are raw physiological data or simple handcrafted time series features derived from the raw data. However, these standard inputs do not provide the model with comprehensive information. The reason is that physiological data collected with wearable devices are typically heterogeneous, noisy, and incomplete. Raw and derived data can be noisy. Stress detection models trained on these standard inputs can become unreliable. Novel data analytics with strong capabilities for handling noise in physiological data can be a promising approach to improve classical inputs.

There are also more disadvantages to all these current approaches. Many research studies have not used any benchmark dataset, so it is challenging to compare the results. The research results will be more convincing if authors test their models with benchmark datasets from other research studies. However, there are few efforts like that (Mishra et al. 2020). This PhD research will deploy novel methods using new data analytic approaches on benchmark datasets to compare effectiveness with traditional stress detection methods.

This PhD project will address the above research gaps. First, it will propose new approaches to detect stress using the outer product of physiological time series and image-based data which is generated from physiological time series data. This project also provides novel TDA topological data analysis and complex networks methods to detect stress. Each method is compared with other state-of-the-art stress detection methods and time series classification methods. So, this project addresses the first research gap which is about lacking comparison of stress detection methods.

This research with new data analytic approaches such as topological data analysis and complex networks has the ability to get deeper insights than traditional machine learning methods, traditional statistical tests and out-of-date machine learning models, even deep learning methods with raw physiological data as inputs. So, this project addresses the second and third research gaps that are related to the limitations of current stress detection methods. There are several reasons that can explain this ability. Topological data analysis is a powerful tool for capturing the shape of data. It is robust to noise, which often happens when wearable devices collect physiological data. Complex networks have huge advantages since this method can provide insight into real systems, such as systems with stress and non-stress states, and can do high-level classification with data.

This research scheme also has the potential to overcome the disadvantage of other stress detection research and the last research gap, which did not use benchmark datasets. This project will provide comprehensive stress detection methods since it proposes new novel stress detection algorithms with new data format and advanced deep learning models to apply to benchmark datasets, compared with other state-of-the-art detection methods on the same datasets.

Chapter 3

Data and methodology

This chapter introduces all three benchmark physiological time series datasets used in this thesis: Driver Stress Data, Wearable Stress and Affect Detection (WESAD) and Smart Reasoning for Well-being at Home and at Work Knowledge Work (SWELL-KW). The first dataset (Driver Stress Data) is among the earliest collected for stress detection. This dataset, with a few thousand timestamps per participant, was used in initial experiments to select suitable algorithms for the thesis. The two other datasets are WESAD and SWELL-KW, which are more recent and more complete for evaluating stress detection experiments. These two data sets, each with a few million records, were used to confirm the effectiveness of the algorithms presented in this thesis. The environments of these two datasets are also closely aligned with real-world environments.

New and more advanced data analytic approaches, such as topological data analysis and complex networks, which are better suited to physiological time series analysis, will also be introduced. Topological data analysis has a unique tool called persistent homology, which can capture and summarise multi-scale relational information (Tauzin et al. 2021). Topological data analysis, which can extract information from a wide range of input data and generate many types of persistent homology, can be a suitable tool for handling multivariate time series data (Tauzin et al. 2021). Complex networks representing real-world systems are also a promising tool for analysing time series since complex networks topological features can capture the natural characteristics of the time series (Li et al. 2022). However, research on these two data analytic approaches for long multivariate time series data remains limited due to their high computational complexity.

3.1 Datasets

Three benchmark datasets are used to evaluate the algorithms in this project. The first one is Driver Stress Data (Healey 2000), the second one is WESAD (Schmidt et al. 2018b), and the third one is SWELL-KW (Koldijk et al. 2014). The first dataset is a classical dataset for stress detection. WESAD and SWELL-KW are newer datasets with sufficient records to validate advanced data analytics approaches and stress detection models. See Table 3.1 to understand information about all these stress benchmark datasets.

3.1.1 Driver Stress Data

This dataset was compiled in 2000 (Healey 2000). The dataset gathered nine drivers' GSR, ECG, respiration data, and EMG data. There are several types of roads that drivers must navigate. When the drivers were resting, they did not face stress; when they drove through an uninterrupted highway, they faced a low level of stress; when they went through the city, they faced a higher level of stress; and when they encountered the exits for the turnaround or the two-lane merge, then they faced the highest level of stress (Healey 2000). In the car, there is a computer running the DOS operating system, and this laptop was remotely managed by a laptop, a microphone, a tape recorder, five physiological sensors, and three video cameras (Healey 2000). The EMG and GSR sensors were attached to the hands of the drivers, another GSR sensor was attached to the leg, the ECG sensor was at the torso, beneath the driver's clothes, and the respiration sensor was around the driver's diaphragm (See figure 3.1) (Healey 2000). To define driver stress levels, the authors used a task design based on stress levels associated with road types, questionnaire analysis, and recorded videos (Healey 2000). The physiological data were sampled at 15.5 Hz.

Nine drivers joined the experiment. Three drivers are undergraduates, and little information was provided about the other six drivers. The first three drivers participated in the driving test multiple times, and each of the remaining six participated only once. The first three drivers were two males and then a female undergraduate with three, four and eight years of driving experience, respectively (Healey 2000). The first driver did not drive during the last three years, and the second did not drive for one month before the experiment. This was also the first time he drove in the city where the experiment took place (Healey 2000). During

Table 3.1: Information of stress benchmark datasets include Driver Stress Data, Wearable Stress and Affect Detection (WESAD) and Smart Reasoning for Well-being at Home and at Work Knowledge Work (SWELL-KW). The sign '-' in the table presents the corresponding information is not provided clearly.

Stress benchmark datasets	Driver Stress Data	WESAD	SWELL-KW
Publication year	2000	2018	2014
Number of participants	9	15	25
Gender of participants	-	12 men and 3 women	17 men and 8 women
Age range of participants	-	27.5 ± 2.4	25 ± 3.25
Environment	Driving environment	Laboratory environment	Working environment
Wearable devices	The GSR, ECG, respiration and EMG sensor	RespiBAN Professional	Body sensors
Emotional states	Low stress, Medium stress, High stress	Baseline, stress, amusement	Neutral, stressed from time pressure, and stressed from interruption
Physiological features	GSR, ECG, respiration data, and EMG data	GSR, ECG, EMG, respiration data, skin temperature, 3-axis accelerometer data	GSR, ECG and heart rate
Frequency of data	15.5 Hz	700 Hz	2048 Hz

the data collection process, sometimes problems happen (Healey 2000).

3.1.2 Wearable Stress and Affect Detection

Schmidt et al. (Schmidt et al. 2018b) introduced the Wearable Stress and Affect Detection (WESAD) dataset, which measured physiological data from 15 partici-

[Production note: This figure is not included in this digital copy due to copyright restrictions.]

Figure 3.1: In the experiment of detecting stress when driving, the drivers wore five physiological sensors: two skin conductivity sensors (SC), one on the left foot and one on the left hand; an electrocardiogram (EKG) on the chest; an electromyogram (EMG) on the left shoulder; and a chest cavity expansion respiration sensor around the diaphragm (Healey 2000). A computer at the back of the car had sensors connected to it.

pants under stress, baseline, amusement, and meditation conditions over two hours. Graduate students in the same faculty are the participants. They did not smoke heavily and were not pregnant. None of these students had any heart or mental health difficulties. Their age was 27.5 ± 2.4 years. Twelve men and three women participated in the study. The physiological data were collected using RespiBAN Professional, and Figure 3.2 illustrates how the devices were connected to the participants' bodies (Schmidt et al. 2018b). RespiBAN Professional contains an accelerometer biosensor, biosignal acquiring hardware and a respiration biometric sensor (Pereira et al. 2023). Physiological measures are galvanic skin response (GSR) measured in microsiemens (μS), electrocardiogram (ECG) measured in millivolts (mV), electromyogram (EMG) measured in millivolts, respiration rate data measured in %, skin temperature measured °C, and a 3-axis accelerometer measured g ($1\ g = 9.80665\ m/s^2$). All data were sampled at 700 Hz.

The 15 participants are indexed from 2 to 17; data for participants 1 and 12

[Production note: This figure is not included in this digital copy due to copyright restrictions.]

Figure 3.2: The RespiBAN and the ECG, GSR, EMG, and respiration rate sensors are placed on the participants' bodies. (Schmidt et al. 2018b)

are not saved because errors were made in recording their progress. The data include eight label types, of which only three are relevant to this work: baseline, stress state, and amusement state. Three were selected to facilitate comparison to state-of-the-art methods. See Table 3.2 to understand the description of labels of the WESAD dataset.

3.1.3 Smart Reasoning for Well-being at Home and at Work Knowledge Work

The Smart Reasoning for Well-being at Home and at Work Knowledge Work (SWELL-KW) dataset was created to understand stress in a working environment (Koldijk et al. 2014). The experiment lasted 107 minutes and contained two stressors: time to finish a deadline and disturbance from unrelated emails. Workers must work under normal conditions and two stressors to gain data. That dataset contains information on computer interactions (with keyboard, mouse and applications), facial expressions, body posture, and physiological features (heart rate, electrocardiogram (ECG), electrodermal activity (EDA) (Koldijk et al. 2014). Data was collected from 25 students: 8 female and 17 male. Their age was 25 ± 3.25 years. The students had experience handling large volumes of work and frequently used email on computers. They did not smoke or drink alcohol three hours before the experiment. The frequency of all physiological signals is 2048 Hz.

Table 3.2: Description of labels of the WESAD dataset. The corresponding rows are the participant identifier, the length of the time series for each participant (number of timestamps), the number of 3 types of emotional labels for stress classification, the percentage of baseline states, the percentage of stress states, and the percentage of amusement states.

Participant identifier	Length time series	Emotional labels	Percentage of baseline	Percentage of stress	Percentage of amusement
2	4,255,300	1,484,700	53.94	29	17.06
3	4,494,000	1,508,500	52.9	29.7	17.4
4	4,460,401	1,515,501	53.49	29.33	17.18
5	4,330,199	1,551,900	54.04	29.09	16.87
6	4,909,100	1,541,400	53.59	29.52	16.89
7	3,631,600	1,538,601	53.96	29.12	16.92
8	3,791,900	1,546,299	52.92	30.33	16.75
9	3,614,100	1,537,900	53.71	29.36	16.93
10	3,811,500	1,593,900	51.82	31.84	16.34
11	3,628,100	1,559,600	52.96	30.52	16.52
13	3,841,600	1,558,201	53.01	29.83	17.16
14	3,839,500	1,558,901	52.99	30.31	16.7
15	3,641,400	1,563,100	52.62	30.72	16.66
16	3,903,200	1,554,701	53.13	30.3	16.57
17	4,255,300	1,484,700	53.94	29	17.07
Average	4,027,146	1,539,860	53.27	29.86	16.87

The SWELL-KW data set contains three labels: neutral, stressed from time pressure, and stressed from interruption. Data from eighteen of the participants were used in this experiment. See Table 3.3 to understand the description of labels of the SWELL-KW dataset. The IDs of these participants range from 1 to 25. Data from other participants with IDs 8, 11, 15, 20, 22, 23, and 25 were not used in this paper's experiments since they had one label or their participant's heart rate, and EDA data were missing or unusable. In these experiments, only physiological data were used. Fourteen participants have all three labels. Participant 3 does

Table 3.3: Description of labels of SWELL-KW dataset. The corresponding rows are the participant identifier, the length of the time series for each participant (number of timestamps), the number of 3 types of emotional labels for stress classification, the percentage of neutral states, the percentage of stressed states from time pressure, and the percentage of stressed states from interruption.

Participant identifier	Length time series	Percentage of interruption stress	Percentage of neutral	Percentage of time pressure stress
1	$10,690,560$	28.74	44.83	26.43
2	12,042,240	33.67	42.86	23.47
3	2,949,120	50	0	50
4	1,843,200	0	73.33	26.67
5	5,406,720	43.18	52.27	4.55
6	8,478,720	39.13	33.33	27.54
7	10,444,800	48.24	23.53	28.23
9	3,072,000	52	20	28
10	5,160,960	0	61.9	38.1
12	8,970,240	35.62	56.16	8.22
13	6,144,000	38	12	50
14	8,847,360	54.17	43.06	2.77
16	12,656,640	39.81	34.95	25.24
17	11,673,600	43.16	29.47	27.37
18	9,216,000	48	29.33	22.67
19	5,529,600	91.11	2.22	6.67
21	8,970,240	13.7	53.42	32.88
24	3,809,280	0	16.13	83.87
Average	7,550,293	36.585	34.933	28.482

not have the label "neutral" and three participants, 4, 10, and 24, do not have the label "stress from interruption". The data was down-sampled to 128Hz by applying a preprocessing code of Quadrini et al. (2022) and from the Python code [1].

[1] https://github.com/ggerardlatek/STREDWES-SWELL

3.1.4 Ethics statement

The three datasets used in this thesis are publicly available. The ethical applications of these three datasets were approved at UTS with the names ETH24-10092, Detecting stress in driving environment (for the dataset Driver Stress Data), and ETH24-9602, Stress detection using wearable devices (for the two datasets WESAD and SWELL-KW).

3.2 Methodology

This section presents the methodology for the data analysis part of the thesis. The first approach described is topological data analysis (TDA). The TDA sub-section provides an overview of TDA and its advantages, core concepts of TDA, state-of-the-art research on TDA with machine learning and its limitations, state-of-the-art research on TDA with time series data and its limitations, and state-of-the-art research on TDA with change point detection. The second approach described is complex networks. The complex networks subsection provides an overview of complex networks, time series segmentation methods for converting time series into networks, state-of-the-art complex networks approaches for converting time series into networks, current data analysis approaches based on complex networks, state-of-the-art complex networks and time series analysis, and limitations.

3.2.1 Topological data analysis

Topological data analysis (TDA) is a relatively new approach for analysing high-dimensional data. TDA is a geometric method with the ability to capture the shape and patterns of data (Lum et al. 2013; Munch 2017). Figure 3.3 shows how TDA captures the shape of a hand (Lum et al. 2013). The input is a hand represented by many points. TDA used a distance metric and filter functions with values associated with the locations of data points to convert the shape of the hand to a network with nodes and edges. Persistent homology is the first and essential concept of TDA (Hofer et al. 2017). TDA uses a powerful tool named persistent homology to construct persistence diagrams and extract topological characteristics of the data (Carrière et al. 2020; Chazal and Michel 2021; Fugacci et al. 2016).

Persistent homology is a graphical tool, but instead of just using data points to represent real-world systems, persistent homology presents the systems using collections of points, edges, triangles, tetrahedrons and higher-dimensional polytopes and summaries and observes the topological information changes of systems by counting the number of connected components, holes, voids (Aktas et al. 2019). Persistent homology can provide a multi-scale summarisation of the data instead of specific views of the data, but still keep a summary data at a suitable size for computation analysis purposes (Aktas et al. 2019; Pun et al. 2022). This can help features generated from persistent homology to be useful as deep learning model's features (Aktas et al. 2019; Tauzin et al. 2021) since deep learning features need to keep global data underlying characteristics of the systems and be simple enough in size and dimension that can be used as deep learning model's inputs (Pun et al. 2022). Persistent homology features can support deep learning in handling high-dimensional data and multivariate time series data (Pun et al. 2022; Tauzin et al. 2021). There are still several research gaps in the use of topological data analysis features as inputs to deep learning models (Pun et al. 2022; Tauzin et al. 2021). First, how to find a way to convert the data into a suitable persistent homology. Secondly, how to extract vectorised features from the persistent homology. Third, extracting features from persistent homology with a reasonable computational expense. Finally, how to find useful features from persistent homology. However, current research cannot bridge the gap between deep learning and topological data analysis.

Many types of persistent homology have been introduced in the literature. Frosini (1990) introduced the 0-dimensional persistent homology. Edelsbrunner et al. (2000) mentioned the whole concept of persistent homology and built a graphical presentation for persistent homology known as the persistence diagram. Carlsson et al. (2005) defined the concept of persistent homology again and introduced the persistence barcode, an equivalent tool to the persistence diagram.

TDA has unique advantages. First, TDA can analyse and get data insights from a coordinate-free point of view. This unique characteristic helps the TDA approach to rely only on the distance between objects (Lum et al. 2013; Musa et al. 2021). Secondly, results from TDA approaches will remain invariant if small changes happen to the input, which means TDA is robust to noise (Hofer et al. 2017; Lum et al. 2013; Musa et al. 2021). Thirdly, TDA has the power to convert

[Production note: This figure is not included in this digital copy due to copyright restrictions.]

Figure 3.3: The process of using TDA to capture the shape and pattern of a hand (Lum et al. 2013). In picture A, there is a point cloud of a 3D hand. In pictures B and C, a filter function is used, and the data points are binned into several groups. Each group has a specific colour. In picture D, each group is clustered. A graph is constructed that captures the shape of the hand.

data to a compressed form to more effectively analyse the data (Lum et al. 2013). Lastly, input data for TDA can take many forms, such as images, time series, and graphs (Hofer et al. 2017).

With all these advantages, TDA can be a useful approach to detecting stress. Modern stress detection research typically requires handling long, multivariate physiological time series data. Time series data can be inputted into TDA. TDA even has the potential to compress long multivariate time series data. Physiological time series data also typically contain substantial noise, and TDA can effectively handle such data.

3.2.1.1 Core concepts of topological data analysis

Simplicial complexes In TDA, data is stored in a form called simplicial complex. TDA can use filtration to transform raw data (point clouds, images, and networks) into a persistence diagram (Myers et al. 2019; Wu and Hargreaves 2021). The persistence diagram can be used to calculate many TDA features, such as Betti curves, landscape distance, heat, persistent image, bottleneck distance, Wasserstein distance, and silhouette. These features can present the topological characteristics of data. In this part, some core concepts of TDA, such as simplicial complexes, homology, and persistent homology, will be presented.

A simplicial complex is a generalised form of a graph (Myers et al. 2019; Wu and Hargreaves 2021). A simplex σ is a vertex set and $\sigma \subseteq V$, where V is a collection of vertices (Myers et al. 2019). The dimension of simplex σ can be measured using a formula $dim(\sigma) = \mid \sigma \mid -1$. A face τ contains a simplex σ if $\sigma \subseteq \tau$ and can be expressed by formula $\sigma \preceq \tau$.

A simplicial complex K can be defined as a set of simplices σ that meets a condition (Myers et al. 2019; Wu and Hargreaves 2021). The condition is if $\sigma \in K$ and $\sigma \subseteq \tau$ then $\tau \in K$. The simplicial complex K can also be defined as a set of simplices where the simplices are closed and form faces of objects (Myers et al. 2019; Wu and Hargreaves 2021). The largest dimension of all simplices σ is the dimension of simplicial complex K, and the formulation of the dimension of K is $dim(K) = \max_{\sigma \in K} dim(\sigma)$. The collection of all simplices σ which have the highest dimension of d can be defined as a d-skeleton of simplicial complex K, $K^{(d)} = \sigma \in K | dim(\sigma) < d$. A filtration is nested set of d-skeletons, $K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_{n-1} \subseteq K_n$. The homology measures the number of

[Production note: This figure is not included in this digital copy due to copyright restrictions.]

Figure 3.4: Simplex, from left to right is 0-simplex (a point or a vertex), 1-simplex (a line or an edge), 2-simplex (a triangle) and 3-simplex (tetrahedron) (Wu and Hargreaves 2021).

structures of a specific dimension in a simplicial complex (Myers et al. 2019). The structures can be connected components in the case of zero dimension, loops in the case of one dimension, or voids in the case of two dimensions (Myers et al. 2019). In summary, simplicial complexes can be collections of points, edges, triangles (Aktas et al. 2019).

Persistent homology The primary tool of TDA is persistent homology, and this tool can analyse data by observing homology changes under the effect of a filtration parameter ϵ. Persistent homology can capture the states of topological change when data are surveyed with many filtration parameters (Hofer et al. 2017; Yen and Cheong 2021). When the filtration parameter changes, then topological features of data (connected components in case of 0-dimension, loops in case of 1-dimension, or voids in case of 2-dimensions) will appear or disappear. The events associated with the appearance and disappearance of the topological features are birth and death, respectively. The whole collection of sequences with birth and death can be represented as a persistence diagram or persistence barcode (Hofer et al. 2017).

The simplest way to explain persistent homology is by imagining a ball attached to every single data point of the raw data, as in Figure 3.5 (Chazal and Michel 2021). When the radius of the ball (filtration parameter ϵ) is zero, then each data point is a 0-simplex, and an event named birth will appear with all balls (see figure 3.5 a) (Chazal and Michel 2021; Yen and Cheong 2021). Now, imagine that the radius of the ball (filtration parameter ϵ) grows gradually; then, a ball will overlap other balls. When the ball of data point index i overlaps with a ball of data point index j, then persistent homology will connect two balls using a line, and now we have a connected component i, j. The overlap of the balls will create another event named death. The persistence barcode now has a complete sequence with the starting point (birth) and the ending point (death) (see figure 3.5 b) (Chazal and Michel 2021). The filtration parameter will continue to grow, and if it reaches a particular value, then some 1-dimensional features (loops) will appear. In the figure 3.5 c, two blue 1-dimensional features appear, and two new blue corresponding sequences appear at the persistence barcode. Then the circle grows more prominent, and one of the 1-dimensional features will be filled. As a result, the corresponding sequences will end (see figure 3.5 d). The two 1-dimensional features are both filled in picture 3.5 f. The persistence barcode can also be presented as a persistence diagram where the birth and death value of each sequence corresponds with the x-coordinate and y-coordinate of a point from the diagram (see figure 3.6) (Aktas et al. 2019; Chazal and Michel 2021).

Betti curves and persistent landscape Topological signatures are a summary representation of topological features (Hofer et al. 2017). Topological features can be computed using persistent homology, an important tool in TDA (Aktas et al. 2019; Kemme and Agyingi 2025). Topological signatures can be extracted from the persistence diagram using some statistical summary functions (Bubenik 2015; Munch 2017). The two most common functions are Betti curves and persistent landscape. Suppose a persistence diagram D and a point p (birth, death) belong to D. Consider there are two piecewise functions (3.1) and (3.3). The value of x is from 0 to the maximum value of the maximum filtration parameter. The formula of the Betti curve can be calculated by equation (3.2). The formula of landscape distance is formula (3.4).

$$f_p(x) = \begin{cases} 1 & \text{if } birth_p <= x < death_p \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

$$Betti_D(x) = \sum_{p \in D} f_p(x) \tag{3.2}$$

$$g_p(x) = \begin{cases} x - birth_p & \text{if } birth_p <= x <= (birth_p + death_p)/2 \\ death_p - x & \text{if } (birth_p + death_p)/2 < x <= death_p \\ 0 & \text{if other cases} \end{cases} \tag{3.3}$$

[Production note: This figure is not included in this digital copy due to copyright restrictions.]

Figure 3.5: This figure present how persistent homology and filtration parameter ϵ build the persistence barcode and persistence diagram. Subfigures are described in the text. Graphs a-f x-axis is ϵ, y = data point. For the persistence diagram, x = births and y = deaths. Figure from (Chazal and Michel 2021).

[Production note: This figure is not included in this digital copy due to copyright restrictions.]

Figure 3.6: Persistence barcode (a) and persistence diagram (b) (Aktas et al. 2019). The persistence barcode in this example has five 0-dimensional features with birth and data values (0,1), (0,1), (0,2), (0,2), and (0,8). Since two pairs of sequences have the same coordinates (0,1) and (0,2), the persistence diagram has three corresponding blue points. The persistence barcode has three 1-dimensional features (2,7), (3,6), and (4,5), so there are three corresponding red points in the persistence diagram.

$$Landscape_D(x) = \max_{p \in D} g_p(x) \tag{3.4}$$

Persistent image Adams et al. (2017) proposed another method for extracting features from the persistence diagram, called persistence images. From the persistence diagram, a differentiable probability distribution and a weighting function are used to construct a persistence surface via coordinate transformation and normalisation (Adams et al. 2017; Zhou et al. 2022). Then, the persistent image is generated by discretising a pertinent subdomain and integrating the surface across each area in the discretised domain (Adams et al. 2017; Zhou et al. 2022). The limitations of persistence images are quadratic time and storage complexity, and the requirements of parameters such as the weight functions, the probability distribution and the resolution of the grid (Hensel et al. 2021).

3.2.1.2 Topological data analysis and Machine learning

Topological data analysis has already been successful in several fields, including chemistry, materials science, cellular data, and drug design (Carrière et al. 2020; Wu and Hargreaves 2021). There are some promising benefits that TDA can help to boost current power in machine learning (Kim et al. 2020). Using TDA techniques, machine learning models can have new specialised input features. Finally, even data in complex formats can be processed with proper filtration and may not require any preprocessing, unlike standard machine learning models.

Recently, many researchers have proposed new TDA approaches for machine learning. From TDA methods, topological summary (signatures) of data will be captured, and these signatures typically are unstructured (for example, a persistence barcode with many sequences, and each sequence has a birth and a death) (Carrière et al. 2020; Hofer et al. 2017). These unstructured data are not suitable inputs to machine learning models because many operations, such as addition and scalar multiplication, are not well-defined (Carrière et al. 2020; Hofer et al. 2017). Many researchers apply mathematical formulas to transform these data into vectors and then use them as input to machine learning models. However, this research is still in its early stages.

Hofer et al. (2017) proposed a topological signature for machine learning models. This is the first study to attach a topological layer to a machine learning model. The authors built a signature using a logarithmic transformation. Hofer and his colleagues tested the topological layer using SVM on four datasets and achieved good results. The authors also show that their signature can be applied to a wide range of inputs. However, that research has limitations: the approach works well for some specific tasks but does not preserve all essential topological information of the persistence diagram (Hofer et al. 2017; Kim et al. 2020). So that approach may not be suitable for all tasks.

Zaheer et al. (2017) introduced Deep Sets, which enable the use of sets as inputs to machine learning models. Both Hofer et al. (2019) and Carrière et al. (2020) proposed TDA machine learning structure based on Deep Sets. Machine learning models normally use vectors as input rather than sets. To use sets as input for machine learning models, the models need to have permutation invariant characteristics, which means the models will get the same output even if the input order is changed (Wagstaff et al. 2019). Permutation invariant characteristic was first mentioned and applied by Qi et al. (2017) and was fully proposed via Deep Sets by Zaheer et al. (2017). Qi et al. (2017) presents a deep learning model that uses a set of point clouds as input instead of vectors. That research tries to keep the properties of point sets. The names of these properties are unordered, interactions between points and invariant under transformations (Qi et al. 2017). The first property is unordered. The meaning of the first property is that there is no order of the points that belong to the point cloud. The meaning of the second property (interactions between points) is that the model must preserve the local structure of adjacent points. The meaning of the third property (invariant under transformations) is that the input points set for machine learning models should be invariant under some transformations. The authors proposed a model with a max-pooling layer that acts as a symmetric function to capture local and global information from the point cloud. They show that the model preserves all properties of point sets by conducting experiments on 3D point sets. That model can utilise the sum of latent presentations to acquire permutation invariant characteristics.

Hofer et al. (2019) use three functions to simulate the idea of Deep Sets, and they call them structural elements. Three structural elements are the exponential structure, rational structure, and rational hat structural elements. They are equivalent to centred Gaussian, peak, and cone functions of the persistence diagram (Hofer et al. 2019). The authors show that these three structural elements will have stable performance. Nevertheless, the authors still confirm that maybe there are better structure elements, and they recommend that these can be improved with more linear parameter gradients (Hofer et al. 2019).

Another approach that utilises Deep Sets is the research of Carrière et al. (2020). That proposes a deep TDA model based on permutation invariant operations and a representation function. The latent representation of Carrière et al. (2020) can convert from many TDA data structures into vectors and then use these vectors as input to the designed machine learning model. The authors claimed that their models could be applied to all common TDA data structures; however, several significant structures, such as the Betti curve and persistent images, are still missing from the paper. That model can also generate only one TDA feature for the machine learning model.

There is still a problem with the Deep Sets structure (Wagstaff et al. 2019).

According to Zaheer et al. (2017), their model performs well on both countable and uncountable domain sets. However, Wagstaff et al. (2019) proved that continuity of uncountable domain sets is necessary for the permutation invariant of Deep Sets to work. Both the research of Hofer et al. (2019) and Carrière et al. (2020), which rely on the Deep Sets structure, require a comprehensive investigation before deployment in real-world applications.

In recent research, many authors have sought to integrate TDA into machine learning to simulate the working process of current machine learning models. Rieck et al. (2019) proposed a new measure to calculate the complexity of deep learning models. Gabrielsson et al. (2020) deployed several experiments to demonstrate the TDA's ability to enhance machine learning. The authors propose a new TDA layer for deep learning, deploy machine learning's regularisation technique using TDA, build a deep topological generative model, and set up a topological adversarial attack (Gabrielsson et al. 2020). However, that research primarily presents the ideals of applying TDA to multiple aspects of machine learning rather than conducting comprehensive research on a specific area. The authors, including Gunner Carlsson, one of the founders of TDA, confirmed that the research is an initial investigation of possible directions for TDA and machine learning.

Kim et al. (2020) listed three main weaknesses of current TDA and machine learning models, presented a TDA classification approach and applied it to two datasets. The first limitation is that each TDA and machine learning model is based on one specific filtration to convert data into a persistence diagram. The second weakness is that the input data type will change the stability of the result of the current models. The third weakness is that the current TDA layer cannot be put in the middle-hidden layer of deep learning models. The authors developed a topological layer based on a weighted persistent landscape and demonstrated that it could overcome all three weaknesses. One of the most novel aspects of their research is the combination of distance-to-measure (DTM) with Vietoris-Rips to make model results stable across input data types. DTM is a distance function for a compact set proposed by Chazal et al. (2011). They also proved the differentiability of the topological layer, allowing it to be placed anywhere in the models. However, the authors did not address the first weakness, as they still use standard filtration (Vietoris-Rips). The results of the experiments would be more persuasive if models were deployed to handle more complicated tasks.

They tested their models on simple tasks, including handwriting detection on the MNIST dataset and object detection on a synthetic dataset.

Reinauer and his colleagues introduced Performer, a neural network that uses Transformer architecture (Reinauer et al. 2021). The authors tested the classification results of their network and outperformed those of both Carrière et al. (2020) and Kim et al. (2020) by 3.5% and 4%, respectively (Reinauer et al. 2021). Three benchmark datasets were ORBIT5k, ORBIT100k and MUTAG, with 1000, 100,000 objects and 188 graphs. According to the authors, their methods include a TDA tool named Saliency Map, which can indicate the importance of each point in the persistent diagram with respect to the associated classification task. meanwhile, Carrière et al. (2020) and Kim et al. (2020) used handcrafted TDA features from persistent diagram (Reinauer et al. 2021). Saliency Map that can interpret TDA features, identify important TDA features for specific tasks, and also reduce data dimensions (Reinauer et al. 2021).

Zhou and his colleagues proposed the idea of directly extracting topological representation from point clouds using their new neural network named TopologyNet (Zhou et al. 2022). This topological representation can be used for many deep learning tasks, such as generating point clouds. The authors confirmed that their tools could take less time than traditional ways to extract topological representation from point clouds (Zhou et al. 2022). The authors also tested TopologyNet in classification tasks with three datasets and compared their results with other TDA machine learning methods. Their tool's performance was comparable to that of different methods, and the time to complete the experiments was much shorter, at just a few seconds rather than several tens of minutes for other TDA methods (Zhou et al. 2022). However, the accuracy was not very high, as it achieved only 64.7% and 74.7% on two datasets. The effectiveness of this method with deep learning still needs to be validated with several experiments, such as classification tasks.

There are several limitations of research on TDA and machine learning (Zia et al. 2024). First, applying TDA incurs a high computational cost; many TDA computations become impractical in practice, suggesting that current deep learning and TDA research remain successful only with small data (Zhou et al. 2022; Zia et al. 2024). Even the state-of-the-art experiments such as experiment of Reinauer et al. (2021) are still for a few hundred or few thousand objects in small datasets. To process and handle just 5000 3D point clouds using TDA, the total memory required is 19.4 GB (Zhou et al. 2022). The time complexity using TDA with persistent homology is from quasi-linear complexity to cubic complexity when handling real-world problems, to the much harder cases (Zhou et al. 2022). Calculating Betti numbers from persistent homology with a massive number of complexes is practically impossible in computational terms (Zia et al. 2024). These challenges can cause TDA approaches to consume excessive time or overflow memory (Zhou et al. 2022). The selection of appropriate methods to convert input data into persistent homology and to compute TDA features can help overcome this challenge. The second challenge is to design a universal framework for vectorising TDA features for deep learning inputs (Zia et al. 2024). There is currently no widely accepted universal mechanism to convert persistent homology to vector features (Hofer et al. 2017; Moor et al. 2020). The reasons are because of theoretical and computational challenges (Zia et al. 2024). Researchers must analyse the problem they aim to solve to identify better ways to convert. The third limitation is the challenge of properly extracting TDA features from high-dimensional inputs (Zia et al. 2024). Types and properties of TDA features will be extracted from high-dimensional data, which remains an open research question.

3.2.1.3 Topological data analysis and time series analysis

TDA has succeeded in several fields, such as material science, cellular data and drug design (Carrière et al. 2020; Wu and Hargreaves 2021). It shows promise for enhancing machine learning (Kim et al. 2020) with feature extraction or reducing the need for preprocessing.

The use of TDA in time series is at an early stage and has significant limitations. Current researches on TDA and time series classification are based on subsampling, but their effectiveness on large datasets remains limited. Two groups generated TDA features from time series (Dirafzoon et al. 2016; Emrani et al. 2014). Subsampling methods for both groups were based on K-nearest-neighbour density, random, and max-min algorithms. However, Dirafzoon et al. (2016) omitted to mention the data size and Emrani et al. (2014) used short breathing sound signals containing only several thousand point clouds. Other methods were ineffective for classifying time series using TDA. Mostly distance-based methods such as k-nearest neighbour (kNN) are used (Marchese and Maroulas 2018; Seversky et al. 2016). Both Chung et al. (2020) and BenjamTorr (2021) validated these time series classification methods on univariate UCR time series data. This is a univariate dataset with 128 short time series for classification tasks (Chen et al. 2015). Chung et al. (2020) proposed a classification method based on a persistent curve, which gained good results on UCR time series data. BenjamTorr (2021) used Betti numbers with UCR univariate time series data and got good results too. Whilst these are appropriate with small data sets, the effort of calculating distances between persistence diagrams to apply these methods is too expensive for larger datasets (Karan and Kaygun 2021). Recent methods modify using time windows, but they remain somewhat limited. Majumdar and Laha (2020) showed a TDA method to classify univariate time series data. Tymochko et al. (2021) used non-overlapping windows to detect sleep states and extract persistent images from the windows, but their model was overfit.

In summary, current TDA approaches for time series require modification to be suitable for longer or multivariate time series. Their computational cost hinders the use of time series, even data with a few thousand data points (Sanderson et al. 2017).

3.2.1.4 Topological data analysis and change point detection

Change points can divide a time series into many segments. Detecting change points is essential to time series analysis (van den Burg and Williams 2020). Detecting change points can find sudden and momentous changes in the data, and early change point detection is very critical since this can signal future disasters (van den Burg and Williams 2020). In physiological datasets, change point detection is also essential, as it can help identify early, drastic harm to the human body. There is currently a lot of researches on change-point detection. However, TDA is still a new approach for detecting change points since little research exists in this area. There is also no research on detecting change points in physiological time series using TDA.

Gidea (2017) converted time series stock data into time-varying weighted networks and observed changes in these networks over time using the persistence diagram. They showed that critical topological changes in the networks are indicative of change points. Gidea and Katz (2018) used a different method called time windows to measure critical changes in financial time series data. The authors measured the changes using a persistent landscape, a variation of a persistence diagram, combined with the L^p norm. They found that the value of L^p norm correlates with the value of a stock. The value of L^p will grow to a high value when the value of the stock gets to its peak and reduces fast when the value of the stock crashes. Yen and Cheong (2021) assume that topological signatures will change when the stock price changes. Yen et al. (2021) developed four toy models of fusions using TDA to explore the financial market. The authors observed that far before the crash, the number of simplices declined gradually with the rise of the filtration parameter, and near the time of the crash, the number of simplices declined quickly with the rise of the filtration parameter.

3.2.2 Complex networks

Complex networks analysis uses networks as a representation for complex systems (Gao et al. 2021; Rubinov and Sporns 2010). Complex networks are excellent tools to capture data patterns (Colliri et al. 2018). Research on complex networks has been advancing rapidly recently and has already made significant contributions to management, biology, engineering, economics, and especially time series analysis (Gao et al. 2021; Jiang et al. 2017). Using complex networks to transform time series data into graphs can preserve the time series's topology, including nonlinearity and chaotic characteristics, and the outputs can be used to predict trends or as features for input to machine learning methods, including deep learning models (Zheng et al. 2021).

The traditional complex networks approach to analysing time series data contains two steps. The first step is to convert time series data into a network. The second step analyses and classifies data based on the type of network or network statistical attributes.

3.2.2.1 Complex networks

A complex network consists of nodes and edges: the nodes represent individual objects in the system that the complex network presents, and the edges represent relationships between the objects (Zhang et al. 2010). Complex networks with node and edge topologies can describe many systems around us, such as transport systems, collaboration networks among researchers, the World Wide Web, food webs, electrical power systems, and telephone systems (Latora and Marchiori 2002; Strogatz 2001; Tran et al. 2019; Zhang et al. 2022).

There are many natural network systems, and these networks can have some similar structures and topologies. There are three main types of network structure: random networks, small-world networks and scale-free networks (Albert and Barabási 2002; Barabási 2013; Boccaletti et al. 2006; da Fontoura Costa et al. 2011). The terminology of random networks, small-world networks, and scale-free networks was introduced by Erdos and Rényi (1959), Watts and Strogatz (1998) and Barabási and Albert (1999), respectively. The degree distributions of random and small-world networks follow the Poisson distribution and exhibit variation, whereas those of scale-free networks follow a power law (da F. Costa et al. 2007). Different network structures have different, distinct characteristics. In random networks, nodes have uniform connection probabilities; in small-world networks, the distance between nodes is close, while hubs (nodes with many connections) only appear on scale-free networks (Perera et al. 2017; Sporns et al. 2004). Many real-world systems have the form of scale-free networks (Albert and Barabási 2002; da F. Costa et al. 2007). These structures can support particular tasks with the networks, such as classification. Park and Barabási (2007) had experimented using two properties with a protein interaction network and a mobile phone network. They showed that the properties of network nodes have a relationship with the underlying global structures of this network, and based on their relationships, nodes can be classified. Scale-free networks can present many real-world systems, and the hubs of these networks can also be used to detect special or important objects in real-world systems. However, there is still very little research about complex networks and long multivariate physiological time series.

Many statistical attributes can be used to measure and research complex networks (Barabási 2013; Curiskis et al. 2015; da F. Costa et al. 2007). These features have the form of vectors and can be used to classify objects of graphs (see figure 3.8). The complex network features can present the characteristics associated with the distance between nodes, number of clusters of graphs, degree distributions, and centrality measurements (da F. Costa et al. 2007). These features can also be used for time series classification (Silva et al. 2021, 2022).

[Production note: This figure is not included in this digital copy due to copyright restrictions.]

Figure 3.7: Topology of a random, small-world, and scale-free network, each has 24 nodes and 86 edges (Sporns et al. 2004). On the left is a random network; each node has the same connection probabilities. In the middle is a small-world network; almost all edges are around the circle, except a few that connect far-away nodes; these edges are the shortcuts that make the distance between all nodes small. On the right is a scale-free network; most nodes (red nodes) are only connected to a few other nodes, but some (black nodes) connect with more than 12 nodes (Sporns et al. 2004). These black nodes are hubs of the network.

[Production note: This figure is not included in this digital copy due to copyright restrictions.]

Figure 3.8: The complex networks can be mapped into feature vectors (da F. Costa et al. 2007). Sometimes, these vectors cannot be mapped back to the original networks, but in other cases, remapping processes convert these features to the original networks; in these cases, these features are complete representations of the networks. (da F. Costa et al. 2007).

3.2.2.2 Time series segmentation methods to convert from time series to networks

This part presents time series segmentation methods such as Hidden Markov Models (HMM), Fast Lowcost Online Semantic Segmentation (FLOSS), Entropy and ShaPe awaRe timE-Series SegmentatiOn (EXPRESSO) and Classification Score Profile (ClaSP), which can be used to convert from time series to networks. With these methods, nodes and edges of networks are defined as physiological states and correlations between these physiological states, respectively.

Hidden Markov Models are stochastic state-space models. Hidden Markov Models assume the existence of unobservable "hidden" or "latent" states that can be used to segment the time series. The values of the time series that can be observed or observations are impacted by these concealed states. Determining the current "latent" state from the collection of available observed data is one of the objectives of the model. In an HMM, a latent state corresponds to a collection of continuous data, and this collection follows a specific distribution (Oates et al. 1999; Yuan and Mitra 2016).

Fast Lowcost Online Semantic Segmentation is a temporal time series segmentation (Gharghabi et al. 2017). It adds a custom arc curve to a time series. The arc curve at a specific time series location represents the probability of a change occurring at that location. Entropy and ShaPe awaRe timE-Series SegmentatiOn (ESPRESSO) is an expanded version of FLOSS (Deldari et al. 2020), with different ways to calculate and build the arc curve. The authors of ESPRESSO confirm that this library can handle repeated data, which remains a challenge with FLOSS.

Classification Score Profile is a time series segmentation method that utilises machine learning and sliding windows (Schäfer et al. 2021). The time series will be divided into many overlapping windows, and the characteristics of all windows will be extracted. A change point will be assumed to happen at time lag i. All windows from the left of this time lag are labelled 0, and all windows from the right are labelled 1. A binary classifier is built, and if the cross-validation score is high for this change point, it is considered a real change point in the time series.

Based on these time series segmentation methods, the time series can be partitioned into groups that serve as graph nodes. The limitation of all current methods is their computational expense, which makes them suitable only for short time series data. Both state-of-the-art time series segmentation methods, FLOSS and ClaSP, as well as HMM, have quadratic complexity (Gharghabi et al. 2017; Schäfer et al. 2021). In an experiment by Schäfer et al. (2021), the total time ClaSP requires to segment time series with 5K points and 10K points is under 10s and 100s, respectively. The time complexity of this experiment is quadratic in practice. In another experiment of Gharghabi et al. (2017), FLOSS handles 270,000 data points in 45 minutes. The complexity of the algorithms and the time required to handle short time series suggest that these state-of-the-art methods still require substantial improvements to process long time series data.

3.2.2.3 Current complex networks approaches to convert from time series to networks

With a suitable scheme, time series data can be converted into a network using time series segmentation methods or complex networks approaches. This part surveys current state-of-the-art complex networks approaches to convert from time series to networks.

Traditional complex networks approaches for time series only concentrated on converting univariate data to a network. The three types of complex networks approaches for univariate time series utilise different principles, namely proximity, visibility, and transition, as described in Silva et al. (2021). The proximity graph method measures the similarity between data time windows, then represents states of the data as nodes and measurements of similarity, with or without a threshold, as edges (Silva et al. 2021). The transition graph method maps symbols to specific patterns in time series data and then uses these symbols and transitions between symbols as edges and nodes of a network, respectively (Silva et al. 2021). Both proximity and proximity graph methods require several parameters to construct the networks, and these parameters can alter the network topology and the time series analysis results (Silva et al. 2021). The visibility graph method represents time stamps as nodes and geometric relationships between values of time stamps as edges (Silva et al. 2021). The visibility graph method preserves the global properties of the time series when converting it to a graph (Lacasa et al. 2008; Silva et al. 2021). The visibility graph method can also store complete characteristics of individual time stamps (Stephen et al. 2015). Another advantage of this method is that it also does not require preprocessing the time series and is parameter-free. As a result, it can reduce the complexity of converting time series data (Silva et al. 2021). These advantages can make the visibility graph method a promising choice for multivariate time series classification tasks, including classifying the states at each timestamp, which typically incurs high computational costs.

Examples of proximity networks methods are the cycles networks of Zhang and Small (2006) and the recurrence network of Marwan et al. (2009). An example of a visibility network method is the visibility graph of Lacasa et al. (2008). An example of a transition network method is ordinal partition transition networks of McCullough et al. (2015) and Small (2013). The first approach to handle time series using complex networks was proposed by Zhang and Small (2006). They divided the time series into many cycles, treated each cycle as a node of the graph, and considered an edge between two nodes based on the phase-space distance between their cycles (Zhang and Small 2006). That first method serves as a model for many subsequent methods. A simple approach to handle time series using complex networks is the visibility graph method proposed by Lacasa et al. (2008). The authors considered each time series timestamp as a graph node and visualised each time-lag value as a vertical bar. They defined the terminology 'visibility' between two nodes as the ability to see the top of one node's vertical bar from the top of the other node's vertical bar. There will be an edge between two nodes if a visibility relationship exists. Another critical method is the recurrence network, in which the recurrence matrix of the time series is mapped onto an adjacency network of complex networks (Marwan et al. 2009). Ordinal partition transition networks considered nodes as continuous patterns of time series and edges as the frequency of transitions between two continuous patterns (McCullough et al. 2015; Small 2013).

However, with the development of data collection and storage devices, multivariate time series data has become more popular recently. This leads to additional requirements of complex networks methods for multivariate time series data. The most notable methods for multivariate time series are an extension of ordinal partition transition networks for multivariate time series (Zhang et al. 2017), an extension of visibility networks, which is multiplex visibility networks (Lacasa et al. 2015), two extensions of recurrence networks, which are multiplex recurrence networks (Eroglu et al. 2018) and inter-system recurrence networks (Feldhoff et al. 2012, 2013), and causal effect networks (Runge et al. 2015). The main concept of many methods is to build a single network from multivariate time series data; this concept was implemented using several preprocessing steps and several techniques to identify relationships between time series and construct edges (Silva et al. 2021). These steps and techniques can increase computational costs when handling long, multi-dimensional time series.

3.2.2.4 Current data analysis approaches based on complex networks

Data in network form can be analysed based on the type of complex networks or the characteristics of complex networks.

There are several types of complex networks. Networks derived from real-world systems are typically scale-free networks. This type of network has degree distributions that follow a power law (See equation (3.5) and figure 3.9) (Barabási and Bonabeau 2003; Loscalzo 2012). Scale-free networks also have a unique characteristic: some nodes have many more connections than the rest of the nodes (Barabási and Bonabeau 2003). These particular nodes are 'hubs' or networks.

$$P(k) \sim k^{-\gamma} \tag{3.5}$$

where:

$P(k)$: fraction of nodes with k connections

k: number of connections

γ: a parameter that has a value between 2 and 3

There is also another way to analyse complex networks named high-level classification, which was suggested by Silva and Zhao (2012). Low-level classification techniques can only use physical characteristics such as distance, similarity, and data distribution to recognise differences between data. In contrast, high-level classification techniques use the data's organisational structure (Carneiro and Zhao 2013; Colliri et al. 2018). The authors applied a high-level classification technique that utilises three characteristics of networks: the average degree, clustering coefficient and assortativity combined with SVM to recognise handwritten digits (Silva and Zhao 2012). Carneiro and Zhao (2013); Colliri et al. (2018); Vilca and Zhao (2020) applied high-level classification with three of the same characteristics of networks, as in the research of Silva and Zhao (2012), six characteristics of networks, and betweenness centrality, respectively. The six complex networks features which were used by Colliri et al. (2018) are average degree, assortativity,

Figure 3.9: Two pictures above show a random network and a scale-free network. The scale-free network has hubs with many connections. The two pictures below show the Poisson degree distribution of the random network and the log-log graph of the power law degree distribution of the scale-free network (Loscalzo 2012). The slope of the line is γ

average local clustering coefficient, transitivity, average shortest path length, and second moment of degree distribution.

3.2.2.5 Complex networks and time series analysis

Complex networks have already achieved success in some areas of time series research. Ávila et al. (2013) converted from physiological time series data (beat-to-beat intervals and systolic and diastolic blood pressures) to networks and classified between healthy women and preeclamptic patients using cardiovascular data. Ruan et al. (2019) applied ordinal partition transition networks to find coupling direction and infer causality between two real-world climate time series. Subramaniyam and Hyttinen (2015) used inter-system recurrence networks and several global complex network features to classify EEG signals as focal or non-focal in patients with epilepsy. Runge et al. (2019) presented methods that used graphs for time series to find causal relationships between time series. Those researchers applied similar complex network approaches to compare time series data, which are only applicable to short time series where the computational complexity is feasible.

The requirements for using complex networks with time series have led to rapid developments in this research area in recent years. However, many researchers in this area focus on clustering the whole time series rather than classifying each timestamp of the time series, or only applying to very short time series. Gao et al. (2019a) utilised multiplex visibility networks to cluster EEG signals of normal and fatigue states. That research only used a short time series, so the converted network has only 62 nodes. Gao et al. (2019b) used multiplex recurrence networks to detect drivers' fatigue. In that paper's experiment, the authors used only 5 minutes of data at 200Hz to train and test the method (the total number of time stamps is 1000). Those two experiments used only very short time series data, which does not scale well to longer time series.

Chen et al. (2018) proposed a method to classify EEG data using complex network features. However, since the author applied a functional brain network method named synchronisation likelihood. According to Rosales et al. (2015), synchronisation likelihood is very computationally expensive. Thus, the approach of Chen et al. (2018) is unsuitable for real-world long time series data. The authors claimed high accuracy for a balanced classification task. However, there is no comparison with other methods.

Supriya et al. (2016) applied machine learning models (support vector machine and K-nearest neighbour) combined with global complex networks features (modularity and average weighted length) to cluster EEG time series data. That approach is appropriate only for univariate time series, and the authors classify the entire time series as epileptic or not. That method contrasts with classifying each point of the time series, a task critical for stress detection. There are some very similar approaches Diykh et al. (2017); Wang et al. (2016) which are similar to the approach of Supriya et al. (2016) research.

In summary, current approaches to complex networks research with time series data have three disadvantages. First, current methods are more suitable for univariate or short-time series data. However, with the rapid development of high-quality wearable devices, much longer, multivariate time series data are now available. Secondly, many methods contrast with segmenting the individual time stamps. Classifying individual timestamps in time series data is necessary for stress detection. The last one is that current complex networks methods combined with deep learning did not provide a comparison with state-of-the-art time series classification methods.

3.3 Conclusions

This chapter presents benchmark datasets and all novel methodologies used in this PhD thesis. This thesis's experiments used all three datasets: Driver Stress Data, WESAD, and SWELL-KW. This chapter also reviews topological data analysis and complex networks, two novel data analytic approaches.

This thesis experiments classify stress from long multivariate time series data. These experiments also apply topological data analysis and complex networks. Since TDA experiments incur substantial computational costs, especially when generating TDA features from time series, using a time window can be a useful technique to filter a reasonable subset of data for feature extraction. Betti numbers are also a simple feature that can be applied to experiments in this thesis. In complex network experiments, the visibility graph method can be a suitable way to convert time series data into networks, as it offers many advantages. The advantages of the visibility graph method are that it preserves the complete characteristics of each timestamp in the time series, does not require preprocessing, and is parameter-free. Suitable complex network features are also chosen to extract from the time series data.

Chapter 4

Stress detection using conventional methods and image-based data

4.1 Introduction

This chapter introduces several conventional methods and one image-based method for detecting stress in physiological time series data. This chapter contains three experiments. The first experiment detects stress in Driver Stress Data using two conventional machine learning methods and one statistical method. The two machine learning methods are Long short-term memory (LSTM) and recurrent neural networks (RNN) models, respectively. The statistical method is Vector Autoregressive (VAR). The second experiment uses a 2D CNN model to detect stress from data generated by the outer product of physiological time series data. The third experiment detects stress from image-based data. This experiment uses the recurrent plot method to generate image-based data from physiological time series data. The contributions of this chapter are the outcomes of these three experiments. The outcome of the first experiment is a complete comparison between machine learning and statistical models in the context of classification for physiological data. The outcomes of the second and third experiments serve as novel inputs to deep learning models. The outcome of the third experiment is a novel image input that can be utilised using computer vision methods to improve the effectiveness of the stress detection process.

4.2 Stress detection using conventional machine learning and statistical methods

Whether machine learning and statistical models will perform better is an important research question in the time series research community. Makridakis et al. (2018a,b) compared many machine learning and statistical approaches in several time series competitions. The authors aim to determine which method performs better. This subsection addresses the above question in a physiological context by comparing the performance of machine learning and statistics on multivariate time series, using LSTM, RNN and VAR as examples. LSTM, RNN and VAR are all common methods for time series classification (Gopali et al. 2024). The advantage of this experiment is that it provides a detailed description of the implementation rather than a general explanation, unlike other similar research. This is also the first comparison of machine learning and statistics for multivariate physiological time series; meanwhile, other research has focused only on univariate time series, which is simpler and less practical nowadays. After setting up the VAR model, this experiment also checks the Durbin-Watson Statistic to identify any data patterns that the VAR model still misses when making classification. This test will make this experiment more complete than other research using VAR.

Recently, some studies have attempted to answer the question of comparing the performance of statistical and machine-learning classification models on real datasets. These researches mostly utilise LSTM and ARIMA, the two most common machine learning and statistical methods. Namini et al. (2018); Tang et al. (2020); Temur et al. (2019); Yamak et al. (2019) and Weytjens et al. (2021) made comparisons of the performance of LSTM and ARIMA on multiple stock datasets, a bitcoin price dataset, a house price dataset, an ionospheric dataset and a cash flow dataset, respectively. Namini et al. (2018); Tang et al. (2020) and Weytjens et al. (2021) showed that LSTM perform better than ARIMA, meanwhile Yamak et al. (2019) and Temur et al. (2019) showed the better performance of ARIMA compare to LSTM. None of the authors specified how to choose ARIMA parameters, how to make the data stationary before applying ARIMA, or the detailed structure of the LSTM model used. Only Temur et al. (2019) mentioned the implementation shortly when they used more than one combination of three parameters p, q, and d or ARIMA and tested the model. However, the authors chose one combination without confirming that it would perform better than the others. The only group of authors who provided detailed parameters and hyperparameters for their LSTM model is Tang et al. (2020). However, they neglect details on how to convert data to a stationary form for ARIMA. Weytjens et al. (2021) does not even mention the dataset size, the train and the test set.

All these researches concentrate only on univariate time series. Meanwhile, many practical tasks, such as analysing physiological data, have multivariate time series. There is also a difference between the physiological data and the data used here, because the physiological data is noisy. As a result, it is usually hard to extract meaningful patterns from physiological data(Guo et al. 2020).

This experiment uses a version of the first dataset (Driver Stress Data) (Healey 2000). This dataset contained GSR, ECG, respiration data and EMG of 9 drivers (see some features from Figure 4.1). In addition to these features, several statistical features were calculated from heart rate variability. These features belong to two types: time-domain (average of all normal-to-normal intervals (AVNN), standard deviation of all normal-to-normal intervals (SDNN), root mean square of successive RR interval differences (RMSSD), percentage of differences between successive normal-to-normal intervals that are greater than 50 ms (pNN50)) and frequency-domain (total spectral power of all normal-to-normal intervals less than 0.003 Hz (ULF), total spectral power of all normal-to-normal intervals between 0.0033 and 0.04 Hz (VLF), total spectral power of all normal-to-normal intervals between 0.04 and 0.15 Hz (LF), total spectral power of all normal-to-normal intervals between 0.15 and 0.4 Hz (HF), ratio of low to high-frequency power (LF-HF).

The sampling frequency of all features is 15.5 samples per second. The correlation of features was calculated and shown in the Figure 4.2. As we can see, stress shows a strong positive correlation with heart rate, a significant positive correlation with respiration, and a strong negative correlation with AVNN. The training, validation, and testing data contain 60%, 20%, and 20% of the physiological data, respectively.

Mean Absolute Error and Root Mean Square Error can be used to compare and measure the performance of classification models. These two metrics measured the differences between real and classified values of all-time lags (Hyndman and Athanasopoulos 2018; Namini et al. 2018).

$$MAE = \left(\frac{1}{n}\right) \sum_{i=1}^{n} \left| Y_i - \hat{Y}_i \right| \tag{4.1}$$

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2} \tag{4.2}$$

where n: is the total number of observations

Y_i : is the real value

$\hat{Y}_i$: is the forecasted value

An LSTM model, an RNN model and a VAR model are used to detect stress. An LSTM model is configured with two LSTM layers and two dropout layers: the first LSTM layer has 64 nodes, and the second has 32 nodes. The output is passed through a single fully connected layer, followed by an activation function. The activation functions of the two LSTM layers are sigmoid, and the activation function of the dense layer is tanh. Adam optimizer was used with a small learning rate (0.0001). This LSTM model used nineteen previous values to classify a future value. The values of all physiological features were normalised to the range from 0 to 1. Then, the result of LSTM with the first driver with the number of epochs is 100, which can be seen in Figure 4.3. Since the number of epochs increased from 1 to 100, the loss values for both the training and validation sets approached 0.25 and 0.24, respectively. Thus, this model performs well on physiological time series data. An RNN model with a similar structure, two RNN layers, two dropout layers, and one dense layer was also used to examine the data. Adam optimizer with a small learning rate (0.0001) was also used with the RNN model. The RNN model also yields results similar to those of the LSTM model. The result of the RNN model can be seen in Figure 4.4. For the VAR model, the Augmented Dickey-Fuller Test is applied to all features to test for stationarity. The EMG and AVNN data for the first driver are not stationary, so first-order differencing is applied to the time series data. The lag order p of the VAR model will be set to seven, so both the LSTM and the VAR will use nineteen past values to make a classification. Coefficients of all time lags of equation (2.4) are shown in Table 4.1. After building the VAR model, some observed data patterns may not be explained by the model. Durbin Watson Statistic is used to check if this condition happens; the statistical result ranges from 0 to 4, and if the statistical value is close to 2, there is no remaining pattern. As we can see in Table 4.2, the result of the Durbin

Watson Statistic shows that this VAR model works well for the physiological data since all the values are close to 2.

The final result shows a similar performance of LTSM and RNN, and also better performance of LSTM and RNN compared with VAR, since the Mean Absolute Error and Root Mean Square Error of LSTM on the test set are 0.39 and 0.49, RNN on the test set are 0.4 and 0.5, while the values of VAR on the test set are 0.53 and 0.72, respectively. The LSTM has stable training and loss curves. The rapid drop and subsequent plateau in the RNN's validation loss may be due to the vanishing gradient problem. The performance of LTSM and RNN is better than that of VAR for classifying stress for the first driver. VAR requires stationary inputs and typically works best with linear data. The input physiological data were already normalised to improve VAR performance. However, because the data are physiological, they exhibit complex patterns that are difficult to detect with a linear model such as VAR. This is why machine learning models perform better than VAR in this case. These early results can inform a comprehensive study of stress classification using LTSM, RNN and VAR.

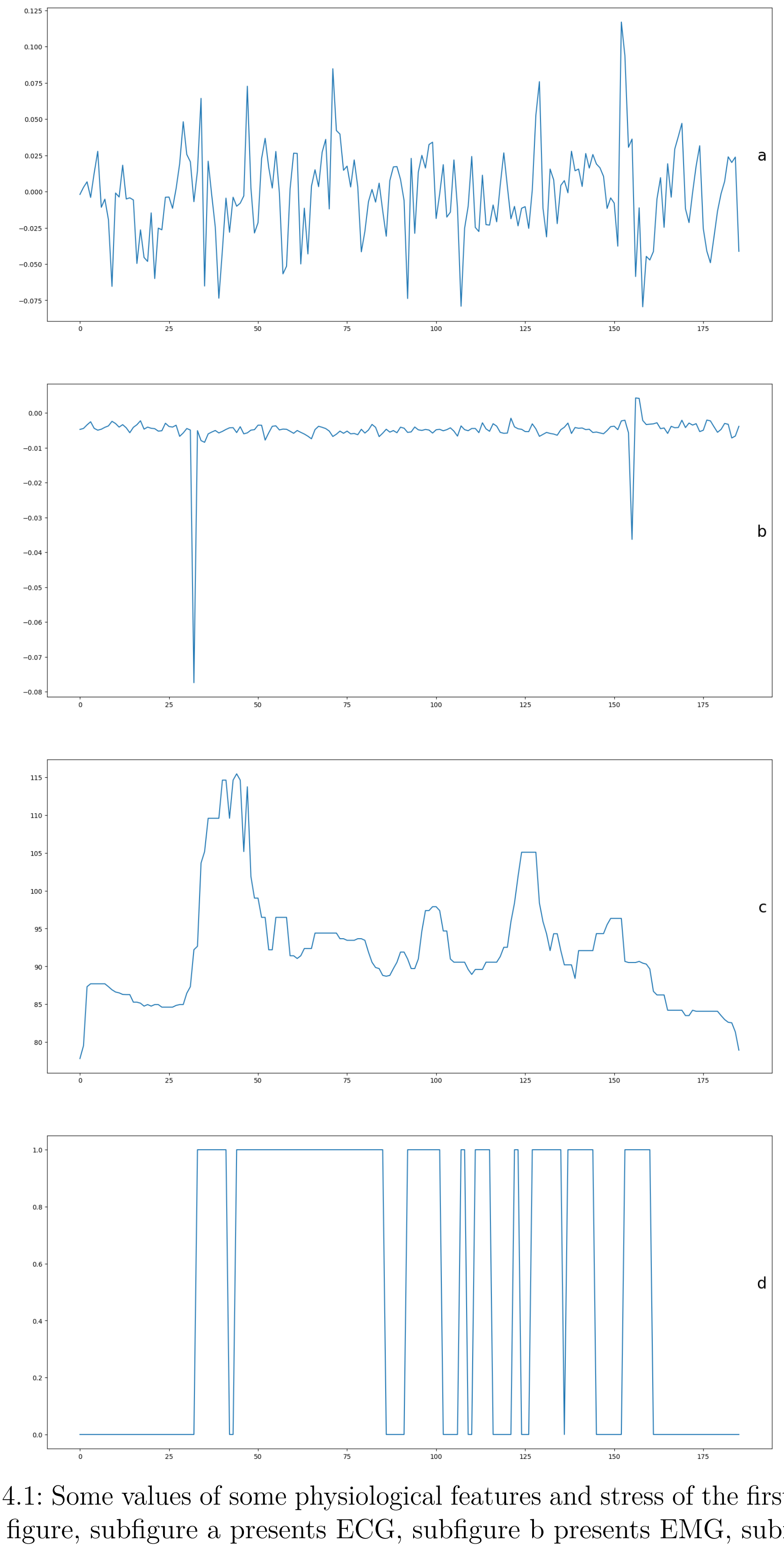


Figure 4.1: Some values of some physiological features and stress of the first driver. In this figure, subfigure a presents ECG, subfigure b presents EMG, subfigure c presents HR, and subfigure d presents stress value.

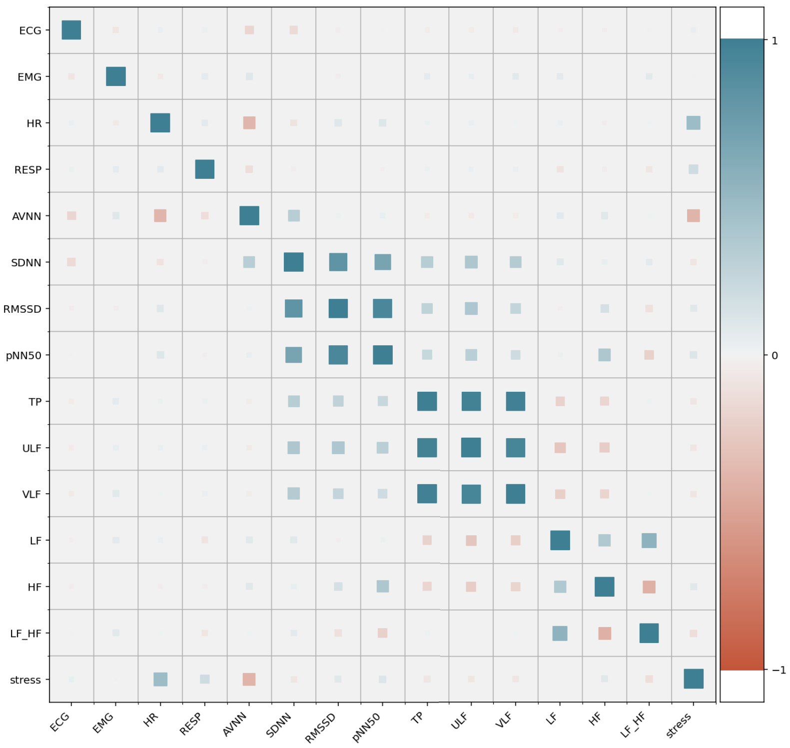


Figure 4.2: Correlations of physiological features of the first driver. The cyan and red present positive correlations and negative correlations, respectively. The size of the square and the colour intensity present the magnitude of the correlations.

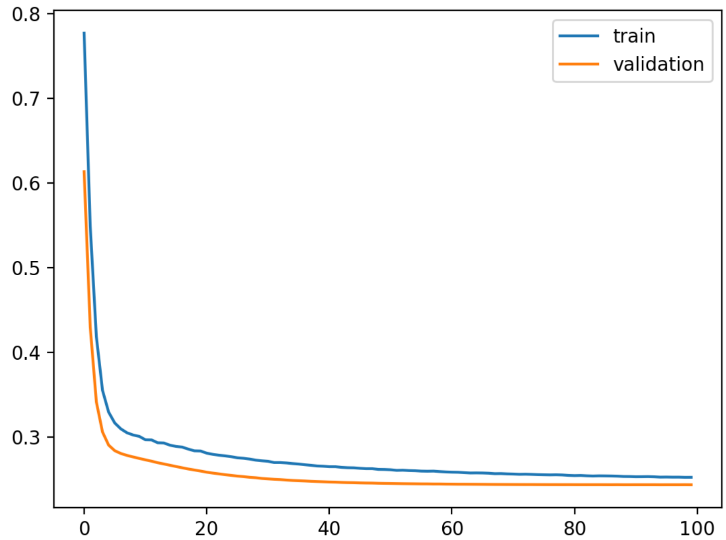


Figure 4.3: Train and validation loss of LSTM model with the number of epochs is 100

Table 4.1: Coefficients of VAR model

Coefficients	**L1**	**L2**	**L3**	**L4**
ECG	-0.181	2.13	0.157	-1.879
EMG	-12.517	1.373	6.536	2.926
HR	0.021	-0.054	0.001	0.032
RESP	0.006	0.135	-0.111	0.06
AVNN	-11.544	0.225	3.027	-3.685
SDNN	38.486	19.98	-18.507	0.94
RMSSD	40.408	-20.399	70.276	20.847
pNN50	-5.013	8.193	-5.14	2.781
TP	7626.459	2155.913	1174.333	2419.105
ULF	-7514.88	-2593.563	-1777.832	-2514.032
VLF	-8008.472	-1797.6300	-531.051	-2294.666
LF	-8144.436	-3039.483	-1114.706	-1494.326
HF	-6084.988	-3673.796	-1774.476	-6381.777
LF_HF	0.045	0.04	0.037	-0.069
stress	0.635	-0.097	0.068	-0.028

Table 4.2: Value of Durbin Watson Statistic

Features	**Value**
ECG	1.96
EMG	2.09
HR	1.97
RESP	2.2
AVNN	1.93
SDNN	2.03
RMSSD	1.97
pNN50	1.94
TP	2.1
ULF	2.09
VLF	2.11
LF	1.99
HF	2.0
LF_HF	1.98
stress	2.05

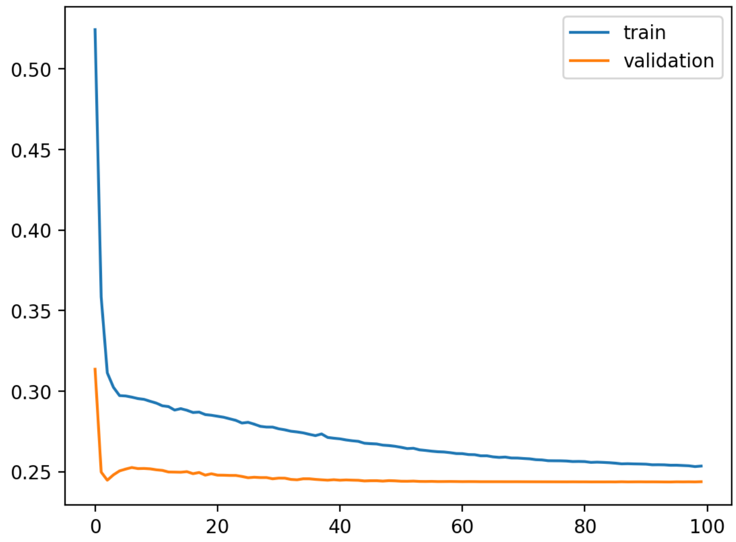


Figure 4.4: Train and validation loss of RNN model with the number of epochs is 100

4.3 Stress detection using outer product of physiological time series data

This experiment uses the WESAD dataset; the sampling frequency is 700 Hz. This method extracts time series data using a window. Each window has a fixed length (700 time stamps) with an 8-time series (galvanic skin response, electrocardiogram, electromyogram, respiration rate, skin temperature, and a 3-axis accelerometer data). The outer product will be applied to each pair of time series to produce a 2-dimensional matrix. The collections of 2D matrices formed from all pairs of time series are used as input to a 2D CNN model. The 2D CNN model in this research contains three 2D convolutional layers, two max pooling layers, one flattened layer, and two dense layers (Fig. 4.5). The model will classify each participant's data from the WESAD dataset using three classes (stress, baseline, and amusement). The training, validation, and testing data contain 60%, 20%, and 20% of each participant's data, respectively. Testing the model on the first participant of the WESAD dataset, the baseline F1 score is 70%, and the two other classes are 0%.

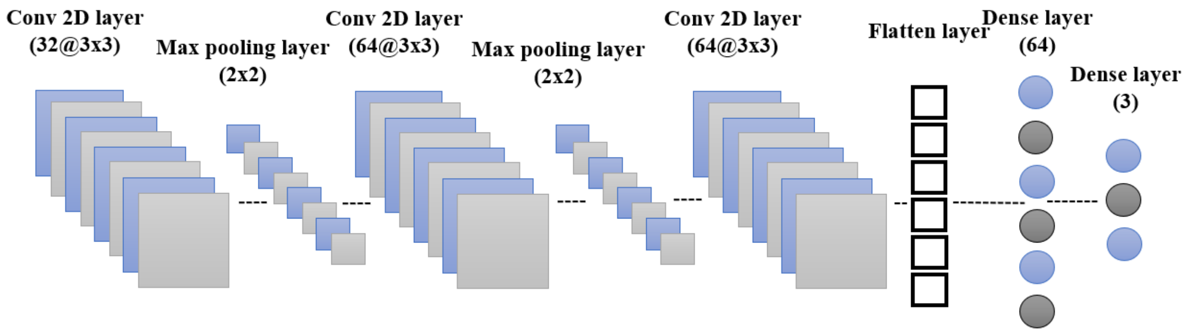


Figure 4.5: The 2D CNN model

4.4 Stress detection using recurrence plot

This method converts from time series to image-based data using the recurrence plot method. The recurrence plot presents the recurrence states of a natural system (Eckmann et al. 1995; Iwanski and Bradley 1998). Given a univariate time series, with two data values (x_i, y_i) and (x_j, y_j). The colour of the recurrence plot at (i, j) is based on the distance between two data points. If the distance exceeds a specific threshold, the colour is black. Otherwise, the colour is white.

This experiment uses the WESAD dataset. Each time series with a window size of 700 was converted into eight recurrence plot images of size 700*700. Examples of recurrence plots of the first time window are shown in Fig. 4.6. The collections of recurrence plots were used as input for the 2D CNN model (Fig. 4.5). The algorithm uses stress, baseline, and amusement classes to categorise each participant's data from the WESAD dataset. 60%, 20%, and 20% of each participant's data are used for training, validation, and testing, respectively. The final classification results for three classes, baseline, stress, and amusement, are shown in Table 4.3. The classification results are not high. This may be because the recurrence plot is sensitive to noise (He and Yu 2023). The physiological data are a complex system that contains noise (Scarciglia et al. 2023).

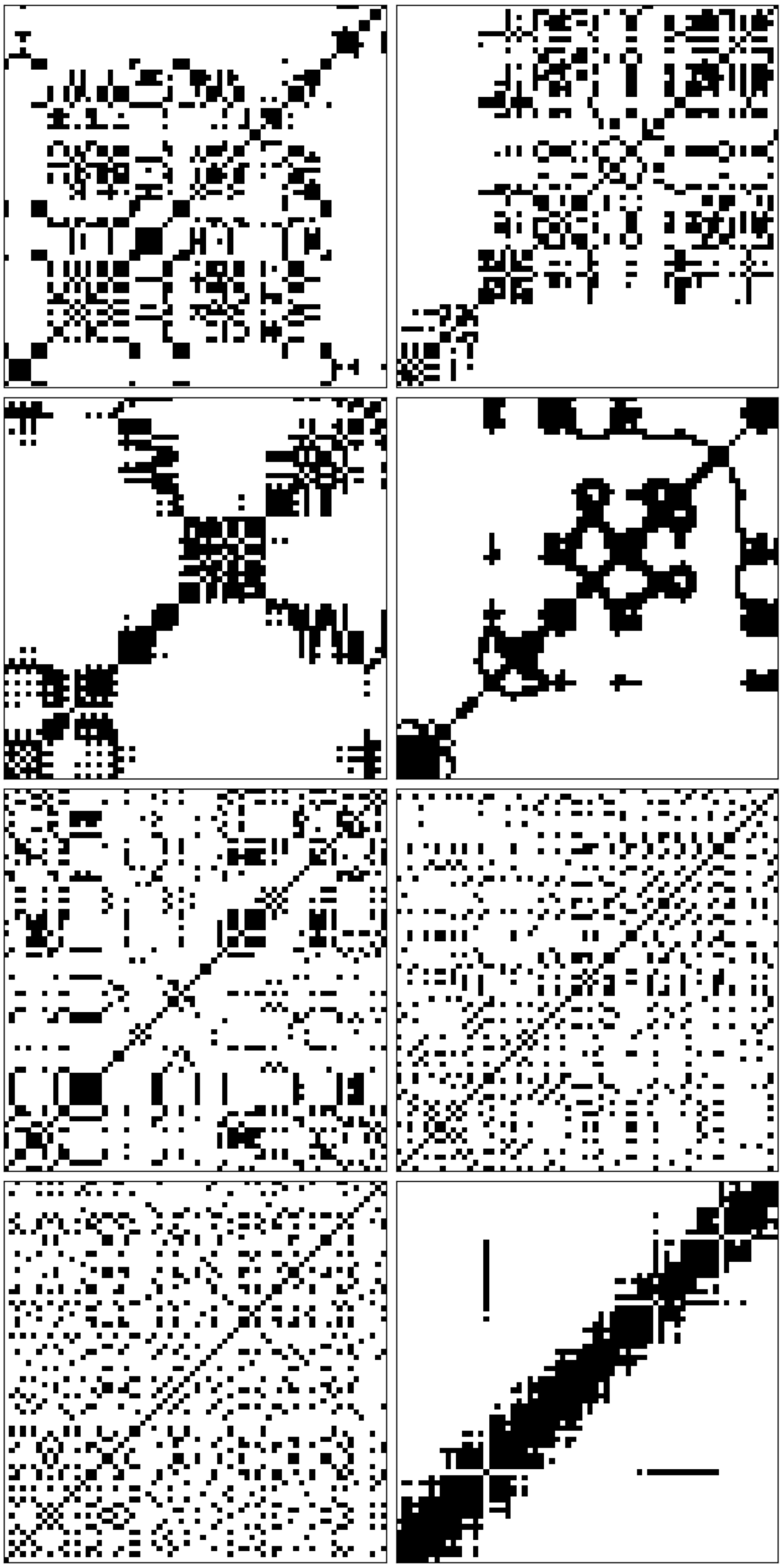


Figure 4.6: Example of recurrence plot of an 8-time series of the first-time window of participant 2.

Table 4.3: Test results when applying a deep learning model with recurrence plot data on each participant.

Participants id	F1 baseline	F1 stress	F1 amusement	Balanced accurancy
2	0.72	0.18	0	0.57
3	0.92	0.77	0.02	0.79
4	0.84	0.48	0.76	0.74
5	0.8	0.8	0	0.73
6	0.71	0.18	0.06	0.57
7	0.71	0.18	0.06	0.57
8	0.7	0.04	0	0.54
9	0.82	0.85	0.04	0.76
10	0.7	0.03	0.29	0.53
11	0.66	0.61	0.42	0.6
13	0.7	0.42	0.5	0.61
14	0.7	0.03	0.04	0.5
15	0.73	0.37	0.04	0.58
16	0.74	0.59	0.19	0.65
17	0.69	0.55	0.05	0.58

4.5 Conclusions

This chapter presents three experiments to detect stress in multivariate physiological time series data. The first uses traditional machine learning and statistical methods on the Driver Stress Data dataset. The conventional machine learning methods (LSTM and RNN) outperformed the statistical model (VAR). This experiment may be among the first to compare traditional machine learning and statistical methods with multivariate physiological time series data. The second and third methods use the outer product of the time series signal and image-based data as input to a deep learning model to detect stress in the WESAD dataset. However, the results are not good, with low F1 scores for each emotional state. The data generated by conventional transformation methods may not provide sufficiently informative features for deep learning classification tasks.

This chapter's contributions are the outcomes of three experiments. The outcome of the first experiment is novel in the context of physiological data, providing a comparison of machine learning methods (LSTM and RNN) and a statistical method (VAR) for stress classification. Machine learning models (LSTM and RNN) already outperformed statistical models (VAR) in this special case. The research scheme comparing machine learning and statistical methods for classifying physiological data is not suitable for further use. The reason is that physiological data are always nonstationary and nonlinear. The statistical models only work well with stationary or linear data. In the case of physiological data, machine learning models tend to perform better than statistical models. The LSTM model has the best performance in this experiment. But LSTM lacks a strong ability to retain long sequential information (Kong et al. 2025). Meanwhile, physiological data always contains long sequential information. Thus, the first experiment is the first encounter with the benchmark datasets used in this thesis. The first experiment used only a small subset of the Driver Stress Data, an old and small physiological dataset for stress classification. The outcome of the second experiment provides novel input for the deep learning models. However, even using GPU to increase the speed, it still took more than 44 hours (161,179 seconds) to run. The second experiment also achieved a zero F1 score for stress detection. Due to the very high running time and low score, the second experiment was applied only to the first participant of the WESAD dataset. The F1 score for stress detection increased in the third experiment compared to the second. The outcome of the third experiment provides novel input in the form of images. So it can apply the advantages of deep learning methods for computer vision. Two data transformations in the second and third experiments, using the outer product of the time-series signal and image-based data, still do not provide sufficient inputs for stress detection methods.

Chapter 5

Detecting stress from physiological time series data using topological data analysis

5.1 Introduction

This chapter presents two experiments: the first uses topological data analysis (TDA) to detect change points, and the second uses TDA to detect stress from long multivariate physiological time series data. Finally, conclusions are given.

The main contribution of this chapter is a novel topological data analysis approach to detect stress. There are currently many researchers trying to detect stress, mainly from data collected from wearable devices. This is because wearable devices can measure changes in physiological features caused by stressors. Machine learning or statistics is often used to detect stress by classifying data into stress and non-stress states. However, physiological time series data collected from wearable devices are noisy. Researchers currently preprocess data to successfully detect stress. This requires a deep understanding of both computer science and clinical aspects. Topological data analysis is a novel approach with the potential to help with noise. Topological data analysis is a combination of computational, statistical and topological methods that find the “shape” of the data (Gidea and Katz 2018) and is robust to noise. By “shape” it means the topological features present in the data. However, TDA is considered computationally expensive, and many researchers believe it is effective only for handling small amounts of data (Savle et al. 2019; Skaf and Laubenbacher 2022). Consequently, TDA research on long, multivariate time series data remains underdeveloped.

The contribution of this chapter is a TDA approach to classify long multivariate physiological time series data, one of the first to apply TDA to such data. It overcomes the problem of long processing times and high memory requirements. Extension from the univariate to the multivariate case is non-trivial. Compared with state-of-the-art multivariate time series classification methods and stress detection methods, we achieve strong results on the Wearable Stress and Affect Detection (WESAD) dataset. Our proposed method uses a time window and suitable parameters to transform data from time series to TDA and a 1D CNN model to achieve good results. We evaluate our method using the WESAD benchmark dataset (Schmidt et al. 2018b). WESAD is the most recent benchmark stress dataset containing about 60 million records from 15 participants and recording Galvanic Skin Response (GSR), Electrocardiogram (ECG), Electromyogram (EMG), respiration data, skin temperature, and 3-axis accelerometer data.

5.2 Topological data analysis and change point detection

A change point can be used to detect the onset and offset of notable events in physiological data. For example, change points can identify unstable respiratory patterns that can lead to panic attacks in the human body (Rosenfield et al. 2010). This section introduced a topological data analysis method for detecting change points in time series data. This experiment used a version of the Driver Stress Data that was introduced in section 3.1.1. The frequency of all data signals is 15.5 samples per second. This experiment utilised a partition of data of the first driver's heart rate (see figure 5.1). Then it detected change points in the data using two TDA features (see figure 5.2 and figure 5.3). The approach detects change points by using sliding windows. Change points can be detected by observing changes in topological features, such as the Betti curve and the landscape distance, over time. A fixed-size window runs through the whole time series, and the size of each time window is 3-time lags (3 samples) with a stride of 1-time lag. Data from each window is converted to a persistence diagram using persistent homology. With many windows, there are many persistence diagrams. The topological differences of these persistence diagrams are calculated using the pairwise distance of the Betti curve and landscape distance (two TDA features which are introduced in section Betti curves and persistent landscape). Each kind of difference is then normalised to get a value between 0 and 1. The normalised topological differences are assumed to be probabilities that change points happen. If the value of topological differences at a specific time lag of the whole time series overcomes a certain threshold (in this example, the threshold is set as 0.3), this is considered a change point (see figure 5.2 and figure 5.3).

In this PhD project, TDA was used to develop a method to detect change points. The technique can detect specific events in physiological time series data. It has the ability to detect stress because stress is also associated with special events. However, TDA is computationally intensive. The TDA-based change point detection method will have very high computational expenses when handling real-world stress detection benchmark datasets. In this experiment, the TDA-based change-point detection method was tested only with the first driver. Additional experiments with other drivers from the Driver Stress Data were not conducted due to the ineffectiveness in identifying change points.

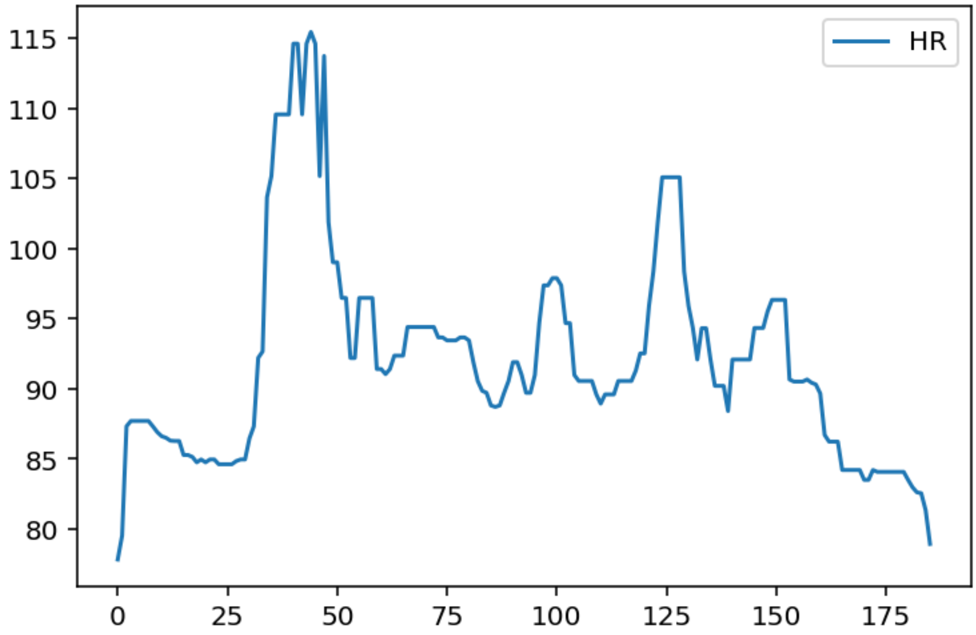


Figure 5.1: Heart rate value of the first driver. The x-axis presents time in minutes, the y-axis presents the number of beats per minute

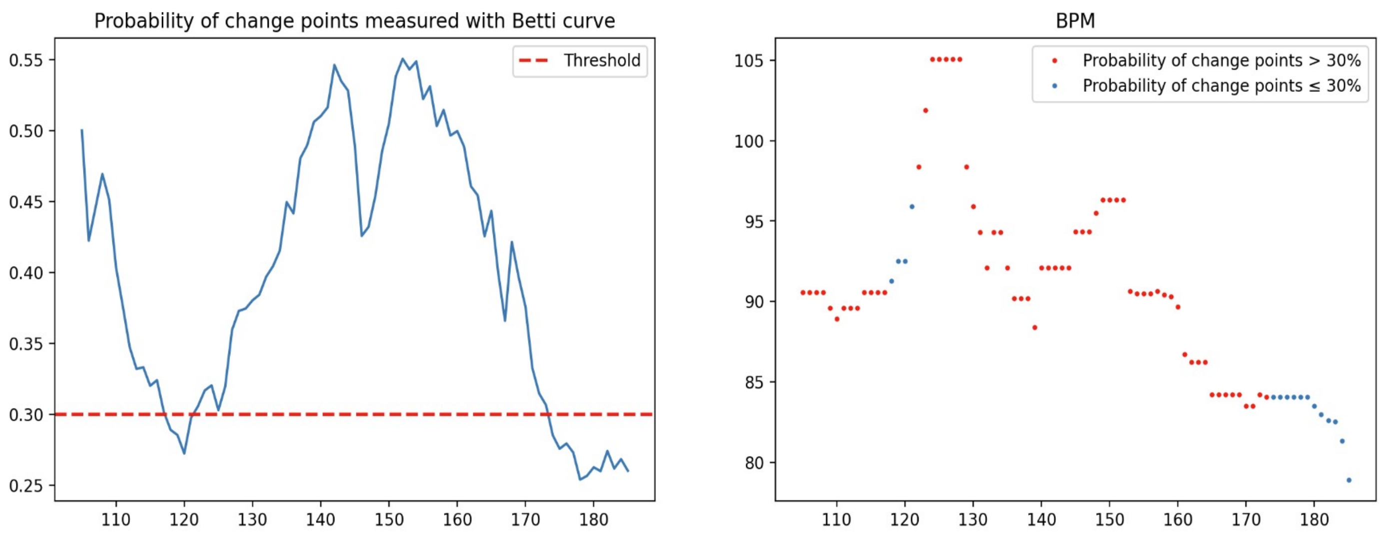


Figure 5.2: Detect change point of the heart rate of the first driver (see figure 5.1) using Betti curve. In the left figure, the x-axis presents time, and the y-axis presents normalised topological differences based on the Betti curve. In the right figure, the x-axis presents time; the y-axis presents the number of beats per minute.

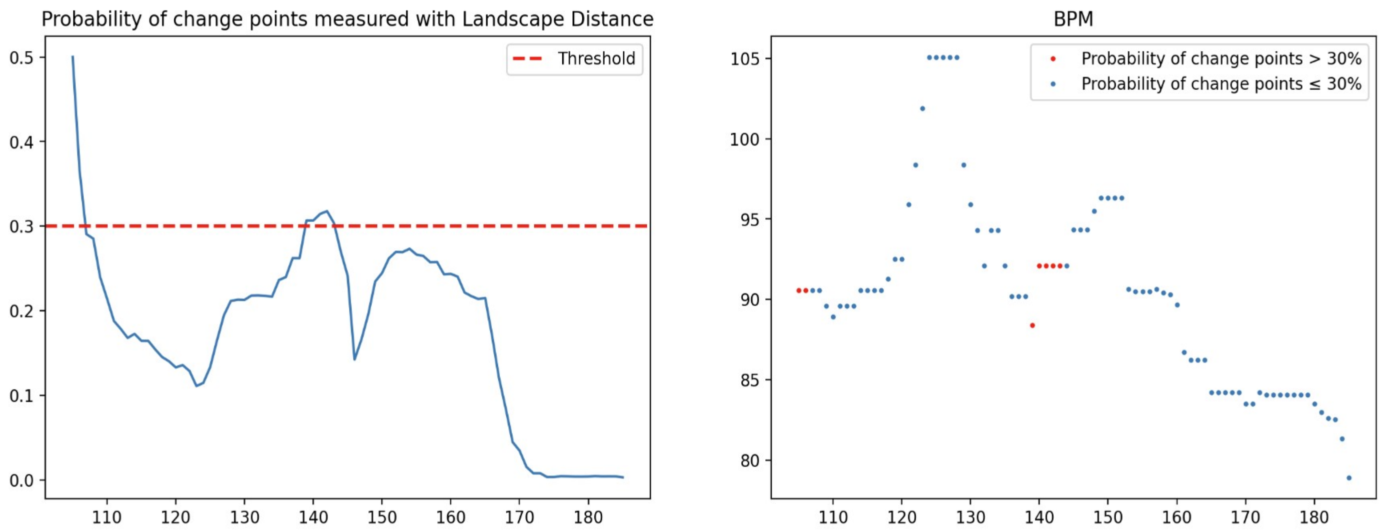


Figure 5.3: Detect change point of the heart rate of the first driver (see figure 5.1) using landscape distance. In the left figure, the x-axis presents time, and the y-axis presents normalised topological differences based on landscape distance. In the right figure, the x-axis presents time; the y-axis presents the number of beats per minute.

5.3 Stress detection using using topological data analysis

This section presents a novel stress detection method using TDA. The technique can already handle the high computational expenses of other TDA approaches. It was also validated on a benchmark dataset named Wearable Stress and Affect Detection (WESAD). It was compared with other stress detection methods and state-of-the-art time series classification methods on the same datasets.

5.3.1 Methodology

This experiment classifies stress based on Betti numbers using a convolutional neural network. Our methodology applies four steps as shown in Fig. 5.4. Step 1 transforms a windowed time series into point clouds using an adapted Takens embedding. We do this because a point cloud is a standard input to TDA. The existing Takens embedding converts from a univariate time series to a point cloud. We have adapted this to the multivariate case. Step 2 transforms the point cloud into a persistence diagram. Step 3 extracts the Betti numbers and combines

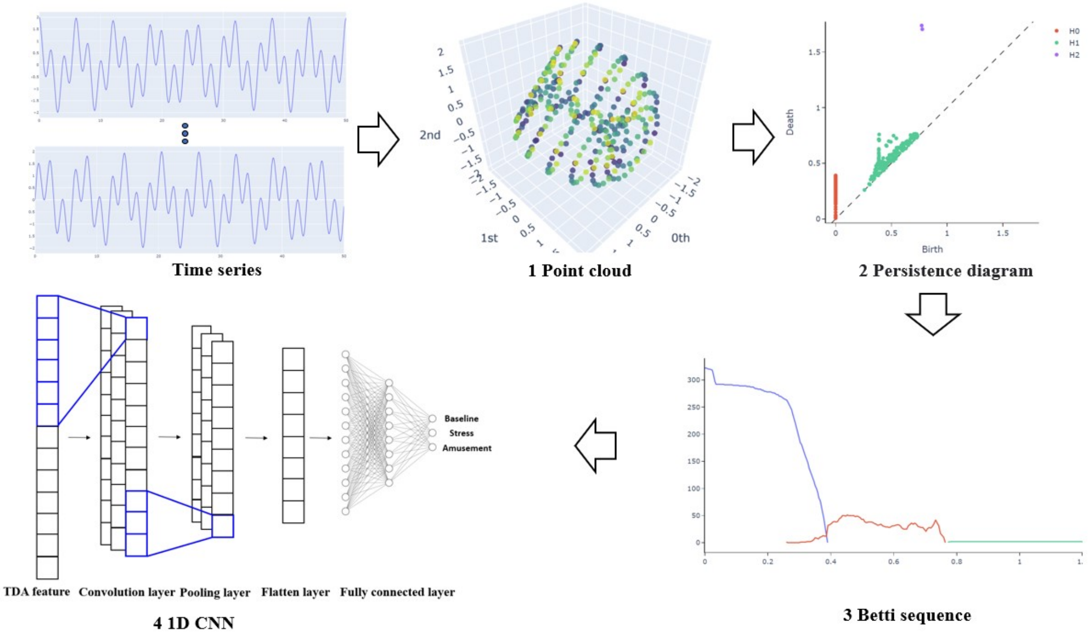


Figure 5.4: Schematic of our experimental methodology consisting of four steps (shown as arrows). Step 1 transforms a time series into a point cloud representation. Step 2 builds the persistence diagram. Step 3 extracts the Betti sequence. Step 4 classifies data into three classes with a 1D CNN model.

them into a single sequence. Step 4 feeds the sequence into a 1D convolutional neural network (CNN) to detect stress into three labels: "baseline", "stress", and "amusement". Only records with labels corresponding to these three classes are kept for this experiment. In summary, 37,601,196 of the total 60,807,600 records were removed to make our experiment comparable to other WESAD experiments. The `giotto-tda` library (Tauzin et al. 2021) was used for TDA operations in steps 2 and 3.

5.3.1.1 Step 1: convert multivariate time series to point cloud

First, the multivariate time series data is divided into equal-length time windows with a length of 1.1 seconds, each converted into a point cloud. The duration of the time windows was chosen based on the experiment presented in Table 5.4. Several time-window lengths were tested to determine the optimal value. Several values of length of the time window were tested to choose the optimal one. These values are 140 observations (0.2 seconds), 280 observations (0.4 seconds), 420 observations (0.6 seconds), 560 observations (0.8 seconds), 700 observations (1 seconds) and 770 observations (1.1 seconds). The model achieved the best average results at window size of 770 observations, and this is also the optimal window size. It was not computationally feasible to explore longer windows. Note that a label is associated with each time window, not the entire participant. To this end, we adapt the Takens embedding (Takens 1981) to the multivariate setting. Starting from a multivariate time series $f(t_1, t_2, ..., t_k)$ with k the number of variables and t_i relating to univariate time series i, each of length l. Given an embedding dimension d and time delay parameter τ, the sequence of vectors is defined as $(f(t_{1j}),\ f(t_{1j}+\tau),\ f(t_{1j}+2\tau),\ \ldots,\ f(t_{1j}+(d-1)\tau),\ \ldots,\ f(t_{kj}),\ f(t_{kj}+\tau),\ f(t_{kj}+2\tau),\ \ldots,\ f(t_{kj}+(d-1)\tau))$. The index j ranges from 1 to $l-d+1$ and $f(t_{ij})$ is value of time lag j, univariate time series i. For our experiments, we choose $d = 3$ and $\tau = 1$. The Takens embedding is applied to this sequence, resulting in a single point cloud.

5.3.1.2 Step 2: generate the persistence diagram

In the next step, the point cloud data is used to build a persistence diagram. As the value of the filtration parameters increases, the process can become computationally intensive. Consequently, we apply a multi-threaded approach. The TDA edge-removal technique (Boissonnat and Pritam 2020) is also used to retain only critical connections between points when applying persistent homology, whilst maintaining consistent topological characteristics. Only topological features from levels 0 to 2 are generated. These are connected components, loops and voids, respectively.

Determining the persistent homology requires generating topological features with increasing value of the filtration parameter (Pereira and de Mello 2015; Skaf and Laubenbacher 2022). For a specific point cloud, the filtration parameter can range from 0 to the maximum distance between any two points in the point cloud. Calculation of this maximum distance is computationally expensive, so we make the assumption that there is one point in the cloud that has all maximum values in all dimensions, and there is another point that has all minimum values in all dimensions. The distance between these two points is the maximum possible distance between any two points and defines the maximum possible value of the filtration parameter. When the filtration parameter exceeds this maximum possible value, no further topological features are generated so that it can be used as the maximum possible value of the filtration parameter, with a consequent reduction in computational expense.

Based on our proposed method for transforming a multivariate time series from the WESAD dataset to a point cloud, using $d = 3$ and $\tau = 1$, then one point in the Takens embedding space has $8 \times 3 = 24$ dimensions: 3 dimensions for each of the 8 features: Galvanic skin response (GSR), Electrocardiogram (ECG), Electromyogram (EMG), respiration data, skin temperature, and a 3-axis accelerometer (x, y, z). Let $\boldsymbol{p}_i$ be defined as the vector $(min_{xi}$, min_{yi}, min_{zi}, min_{GSRi}, min_{ECGi}, min_{EMGi}, min_{rdi}, $min_{sti})$ and $\boldsymbol{q}_i$ be defined as $(max_{xi}$, max_{yi}, max_{zi}, max_{GSRi}, max_{ECGi}, max_{EMGi}, max_{rdi}, $max_{sti})$, which are the maximum and minimum values of the respective features of participant i. The resulting function to calculate the maximum possible filtration parameter $F_{\max}$ of participant i is defined as

$$F_{\max}(i) = \sqrt{\tau(\boldsymbol{q}_i - \boldsymbol{p}_i)^2} \tag{5.1}$$

Based on the minimum and maximum values of each participant (Table 5.1) and the maximum possible filtration parameter of each participant (Table 5.2), the maximum possible filtration parameter of all participants is for participant 9 and takes the value $F_{\max} = 80$.

5.3.1.3 Step 3: extract the Betti numbers and combine into one sequence

The next step calculates (β_0, β_1, β_2), the vector of the number of 0-dimension, 1-dimension, and 2-dimension topological features. We use a vector (β_0, β_1, β_2) of Betti numbers to partly describe a dataset as input to a deep neural network. Given the maximum filtration parameter value of 80 each with three β values, we result in $80 \times 3 = 240$ input values for each time window input into the deep learning model.

5.3.1.4 Step 4: 1D CNN model

The 1D CNN has already achieved some success in handling time series data. 1D CNNs have the advantage when processing time series data over 2D CNNs, such as being easier to train and still achieving state-of-the-art performance (Kiranyaz et al. 2021). In addition, 1D CNN has lower computational requirements than 2D CNN.

Table 5.1: The minimum and maximum value of eight physiological features of all 15 participants. The values of the first and twelfth participants were removed by the author of WESAD because of errors that happened when recording data. The data contains galvanic skin response (GSR), Electrocardiogram (ECG), Electromyogram (EMG), respiration data, skin temperature, and a 3-axis accelerometer (x, y, z)

Participant	**x**	**y**	**z**	**ECG**	**EMG**	**GSR**	**temperature**	**respiration**
min of 2	0.21	-0.66	-1.13	-1.49	-0.41	0.26	28.04	-27.90
max of 2	2.02	0.53	1.24	1.49	0.30	7.57	34.37	27.37
min of 3	-6.6	-6.6	-6.6	-1.5	-1.5	0	30.6	-50
max of 3	2.98	1.32	4.50	1.49	1.46	16.36	34.52	19.68
min of 4	-0.0098	-0.62	-1.44	-1.49	-0.19	0.91	32.34	-20.94
max of 4	1.69	0.46	1.94	1.49	0.36	5.50	34.09	21.27
min of 5	-1.0086	-1.38	-4.78	-1.49	-0.73	3.4	33.85	-25.99
max of 5	2.66	1.6	4.44	1.49	0.33	12.73	35.46	33.42
min of 6	0.24	-0.67	-3.46	-0.71	-0.44	5.17	32.87	-31.52
max of 6	1.84	0.55	2.28	1.49	0.47	12.25	35.22	38.8
mĩn of 7	0.37	-0.62	-3.43	-1.16	-0.53	10.98	33.84	-40.3
max of 7	1.796	0.48	2.5	1.49	0.55	22.41	35.37	33.09
min of 8	-0.47	-0.77	-1.79	-1.44	-1.25	3.29	33.63	-24.35
max of 8	2.32	0.64	2.26	1.49	1.03	5.71	35.77	22.61
min of 9	0.32	-0.37	-1.83	-1.49	-0.47	1.32	33.59	-45.82
max of 9	1.96	0.33	1.08	1.49	0.35	4.24	35.23	32.97
min of 10	0.48	-0.43	-1.85	-1.46	-0.15	0.41	33.65	-26.93
max of 10	1.47	0.48	1.53	1.47	0.14	2.25	35.77	31.89
min of 11	0.29	-0.82	-1.54	-0.87	-0.35	5.62	33.43	-36.19
max of 11	2.03	0.49	2.61	1.49	0.33	8.16	35.23	30.08
min of 13	0.25	-0.54	-0.89	-1.05	-0.28	2.02	32.53	-30.08
max of 13	2.84	0.51	0.91	1.49	0.46	6.63	34.86	37.88
min of 14	0.24	-2.45	-2.37	-1.49	-0.45	2.46	32.75	-31.62
max of 14	1.97	1.24	1.54	1.49	0.35	3.26	34.98	35.91
min of 15	0.25	-0.62	-2.26	-1.25	-0.37	1.25	33.01	-22.19
max of 15	2.10	0.71	3.7	1.49	0.50	4.31	34.9	28.75
min of 16	0.29	-0.47	-3.43	-1.4	-0.22	0.63	33.28	-25.28
max of 16	1.35	0.40	2.35	1.49	0.18	5.5	34.71	15.92
min of 17	0.18	-0.66	-1.32	-1.25	-0.9	5	32.1	-28.02
max of 17	1.85	0.54	1.47	1.49	1.13	8.21	34.75	28.68

Table 5.2: The maximum possible distance of each participant

Participant	Maximum filtration parameter
2	56.29
3	73.72
4	42.79
5	61.14
6	71.02
7	74.61
8	47.5
9	80
10	59.07
11	66.56
13	68.29
14	67.89
15	51.54
16	42.04
17	57.06

Following some experimentation, we trained a deep neural model (shown in Fig. 5.5) containing two 1D convolutional layers, two dropout layers, one max pooling layer, one flatten layer, and three dense layers. The first two dense layers utilised L2 regularization. Dropout and L2 regularization are applied to reduce overfitting. The pooling layer reduces the number of features extracted, allowing the model to focus on essential features. Constructed features are transformed into 1-dimensional data using flattening. The model is optimised using Adam with a categorical cross-entropy loss function. Model performance is evaluated using balanced accuracy and F1, since the dataset is unbalanced with respect to the label.

Experiments ran on four machines of UTS with the configuration 2x AMD EPYC 7532 2.40GHz 32 cores 256M L3 Cache (Max Turbo Freq. 3.33GHz),1024GB 3200MHz ECC DDR4-RAM (Eight Channel), 2x 1.92TB SSD SAS Drives (Raid 1) and 2x 3.84TB SSD SAS Drives (Raid 0), NVIDIA 16GB Graphics Card.

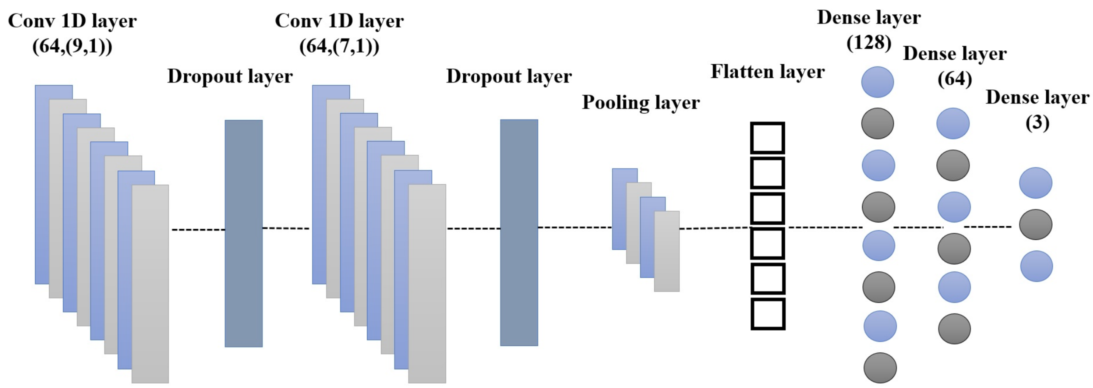


Figure 5.5: Schematic of our 1D CNN model giving sizes of each layer.

5.3.2 Results

5.3.2.1 Results and comparisons

In the first experiment, we train a model for each participant separately, using a 60/20/20 split for training, validation, and test, respectively. Results, including F1 score for each label, balanced accuracy, and elapsed time, are presented in Table 5.3. Results are averaged over three runs with random initial weights, but other parameter values are kept the same, including dropout rate 0.5, learning rate 5×10^{-4}, and L2 0.01. The last row of the table shows averages for all participants. The size of the time window is 1.1 seconds (which equates to 770-time lags), and the stride is 1/700 second (which is a one-time lag).

In a second experiment, we explored different window sizes. Table 5.4 presents results for 140 (equivalent to 0.2 seconds), 280 (0.4s), 420 (0.6s), 560 (0.8s), 700 (1s) and 770 (1.1s) and shows that a window size of 770 gives the best results for both the F1 score of stress and also balanced accuracy compared to other window sizes. The values in this table are the average results for participants.

In a third experiment, we compare our approach to state-of-the-art methods for multivariate time series classification. Methods compared are Time Convolutional Neural Network (Time CNN) and Fully Connected Neural Network (FCN) (Zhao et al. 2017), Multi-Layer Perceptron Network (MLP), ResNet (Wang et al. 2017), LSTMFCN (Karim et al. 2019), Time series attentional prototype network (TapNet) (Zhang et al. 2020b), and InceptionTime (Fawaz et al. 2020). Default parameter settings from `sktime` were used. Average results over three runs are shown in Table 5.5. Compared with the state-of-the-art multivariate time series

Table 5.3: Test results when applying deep learning model on each participant.

Participants id	F1 baseline	F1 stress	F1 amusement	Balanced accurancy	Time
2	1	0.993	0.987	0.993	9h25m
3	0.993	0.997	0.987	0.99	10h23m
4	1	1	0.997	1	10h46m
5	0.993	0.993	0.977	0.99	12h55m
6	1	1	1	1	10h29m
7	0.987	0.993	0.977	0.987	10h42m
8	0.99	0.987	0.977	0.987	11h14m
9	1	1	1	1	10h56m
10	0.92	0.92	0.75	0.893	11h44m
11	1	1	1	1	10h40m
13	1	0.997	0.99	1	11h22m
14	1	1	0.993	1	10h31m
15	0.993	0.993	0.99	0.993	10h50m
16	1	1	0.997	1	10h52m
17	0.99	0.997	0.977	0.99	11h53m
Average	0.991	0.991	0.973	0.988	11h58m48s

Table 5.4: Average test results when training with different window sizes.

Window size	F1 baseline	F1 stress	F1 amusement	Balanced accuracy
140	0.85	0.839	0.723	0.823
280	0.931	0.905	0.861	0.913
420	0.955	0.941	0.875	0.937
560	0.951	0.944	0.859	0.935
700	0.951	0.952	0.841	0.933
770	0.991	0.991	0.973	0.988

Table 5.5: Comparisons with the state-of-the-art time series classification methods. The best result is bold.

Methods	F1 baseline	F1 stress	F1 amusement	Balanced accuracy
Time CNN	0.657	0.174	0.093	0.418
Full CNN	**0.998**	0.99	0.971	0.991
MLP	0.693	0	0	0.379
LSTMFCN	0.968	0.969	0.923	0.961
InceptionTime	0.942	0.926	0.893	0.929
TapNet	**0.998**	0.983	**0.982**	**0.992**
ResNet	0.996	0.987	0.944	0.984
Proposed method	0.991	**0.991**	0.9673	0.988

classification methods, our proposed method achieves the best F1 score for stress detection and the third-best balanced accuracy. TapNet, a recent deep learning method, performs slightly better than our approach.

Finally, in Table 5.6, we compare our results against published results for competitor methods on the WESAD dataset. The average F1 score for stress and the average balanced accuracy score of all participants are presented. Lisowska and colleagues used a 1D CNN to classify stress from the raw WESAD data (Lisowska et al. 2021). Huynh and colleagues applied a customized deep neural network model (Huynh et al. 2021). Two research teams utilised convolutional neural networks (Alshamrani 2021; Gupta et al. 2021). Tiwari et al used Random Forest (Tiwari et al. 2023). Our proposed method has the highest F1 score and accuracy. Our process, with a short and optimal time window (1.1 seconds) and a short stride (1/700 second), can easily detect patterns in which combinations of physiological features are associated with which emotional states and rapid changes of physiological systems. TDA data format, with the ability to analyse noisy data, can provide better inputs for our 1D CNN model than noisy raw physiological features or simple handcrafted features. The TDA approach also does not require preprocessing steps, which are common in many other state-of-the-art methods, such as that of Tiwari et al. (2023). All these reasons make our proposed methods achieve excellent performance compared with other state-of-the-art stress detection methods.

Table 5.6: Comparison with other methods, some other methods only provide accuracy or F1 score. The '-' sign in the table represents values that were not provided.

Methods	Average F1 score	Average accurancy
Lisowska et al. 2021 (Lisowska et al. 2021)	0.705	-
Huynh et al. 2021 (Huynh et al. 2021)	-	0.835
Alshamrani 2021 (Alshamrani 2021)	0.84	0.85
Gupta et al 2021 (Gupta et al. 2021)	-	0.854
Tiwari et al. 2023 (Tiwari et al. 2023)	0.98	0.98
Proposed method	**0.991**	**0.988**

5.4 Conclusions

This chapter presents an experiment on change point detection for univariate time series. However, this experiment is not practical for long multivariate time series data due to the high computational cost of implementation, whereas it is practical for short, univariate time series data.

The contribution of this chapter is a novel topological data analysis approach to detect stress. We implemented a Takens embedding for multivariate time series to convert the time series to point clouds and then extract TDA features (Betti numbers) at various values of a filtration parameter to represent the persistent homology of these data. These extracted features were used to predict stress labels using a 1D CNN model. Our experiments confirmed that our method performs well against state-of-the-art multivariate time series classification methods and competing state-of-the-art methods on the WESAD data. This TDA method also overcomes the enormous computational cost of classical TDA methods. Initially, this TDA method mapped high-dimensional multivariate time series data to a lower dimension using a bespoke Taken embedding. This makes the process of storing, processing, and analysing TDA data quicker and simpler. Second, this method provided a way to determine the optimal filtering parameter for each participant's data. It is better to set a threshold rather than the maximum value for the filtering parameter. Large, intricate, and completely integrated complexes can be avoided. This optimal filtration parameter can reduce the computational cost of TDA processes.

The limitation of this method is that it requires computers with a tremendous amount of RAM to run. These TDA method experiments were run on a high-performance computer with 1024 GB of RAM. The method also takes a lot of time to run. The shortest time to classify a participant is 9 hours and 25 minutes, and the longest is 12 hours and 55 minutes. On average, the method took 11 hours, 58 minutes, and 48 seconds to classify one participant.

Chapter 6

Detecting stress from physiological time series data using complex networks

6.1 Introduction

This chapter presents two experiments that use complex networks to detect stress in multivariate physiological time series data. The first one is based on the network characteristics of scale-free networks, which were already presented in Section 3.2.2.4. The second one utilises machine learning and complex networks features to detect stress.

The main contribution of this chapter is a novel complex networks method to detect stress. This method combines complex networks with deep learning to detect stress from multivariate time series data. Various data analytics techniques, including statistics, machine learning, and many others, have previously been used to identify stress responses from physiological time series data. However, physiological data is characterised by high noise levels, which can prevent the effectiveness of stress response detection (Längkvist et al. 2014; Ma et al. 2023; Trirat et al. 2024). Complex network is a data representation that geometrically depicts the interconnections among the components of a real-world system (Iacobello et al. 2021; Sousa and Barata 2021). Applying complex networks to time series, especially multivariate ones, is a promising research area and can be a viable solution for stress response detection (Gao et al. 2021; Li et al. 2022). One of the critical questions in this research area is how to use complex network learning representations to capture information from time series data. This question contains two parts. The first part of the question involves applying suitable complex networks methods to convert from time series data to networks. Complex networks approaches can convert from time series data to networks. These networks have nodes and edges. Nodes can be individual time lags or groups of time lags. Edges connect these nodes, which have some relationships. The second part is choosing an appropriate topology of graphs and corresponding features from hundreds of potential statistically complex networks features to analyse time series. These features can show global, intermediate, nodes and edges' level characteristics of the networks (Silva et al. 2021). Global properties are properties of the whole elements of networks. Intermediate properties can present properties of subgraphs. Nodes' and edges' level properties exhibit individual characteristics of nodes and edges. These complex network features are represented as vectors and can capture specific characteristics of time series data by summarising the underlying dynamic process (Silva et al. 2021). These complex network features can overcome the limitations of classical time series features, which often suffer from computational issues when handling long time series (Silva et al. 2021). Complex network features in vector format can be used as input to deep learning models or complex networks approaches to analyse time series data.

Our proposed method addresses the question using the visibility graph method, weighted networks, and specific local complex networks features. The answer can be discovered using public long physiological multivariate time series data experiments. This chapter uses the visibility graph method of Lacasa et al. (2008) to convert from the multivariate time series to networks. The visibility graph method considers each time stamp as a node in a graph and visualises the value of each timestamp using vertical bars. A terminology named 'visibility' between two nodes is declared if the top of the vertical bar of one node can be seen from the top of the vertical bar of other nodes. There will be an edge between the two nodes if a visibility relationship exists. This method is suitable for noisy and non-stationary multivariate time series data (Sannino et al. 2017). The visibility graph is also parameter-free and does not require additional preprocessing, so its computational expense will be lower than that of other methods (Silva et al. 2021). Another reason to choose the visibility graph method is that it converts each timestamp into a node, preserving all labels of the time series data. With more labels, the effectiveness of this proposed classification method can be found clearly instead of using other complex networks conversion methods that down-sample the time series data, which combines several time stamps to one node (Stephen et al. 2015).

The visibility graph method creates edges based on timestamp heights, so weighted edges with values that correlate with height can distinguish more from less critical nodes. The weighted network can also yield more robust results than binary edges (Subramaniyam and Hyttinen 2015).

Four complex network features, namely degree centrality, clustering coefficient, closeness centrality, and betweenness centrality, were extracted from the built network. Degree centrality and the clustering coefficient can identify more essential nodes by detecting nodes with more connected edges and those with a denser connected network of neighbours. Closeness centrality and betweenness centrality see the more important nodes by recognising nodes with shorter distances to other nodes in the network, and how often a node is bridged between the two shortest distances in the graph. The method was evaluated using the Wearable Stress and Affect Detection (WESAD) and The Smart Reasoning for Well-being at Home and at Work Knowledge Work (SWELL-KW) data sets. The WESAD and SWELL-KW datasets contain long multivariate physiological time series data from 15 and 25 participants, respectively. Signals from the WESAD dataset are Electromyogram (EMG), Electrocardiogram (ECG), Galvanic Skin Response (GSR), respiration data, skin temperature, and 3-axis accelerometer data. The signals from the SWELL-KW dataset are Electrocardiogram (ECG), Electrodermal activity (EDA) and heart rate data.

This method has two advantages: (1) To the best of our knowledge, this provides one of the first complete complex networks methods to classify each time stamp of a time series. This thesis also uses a collection of complex networks features that can present the multivariate time series combined with a novel weighted graph method. This collection of features differs from previous collections. (2) This method also gains excellent results compared with state-of-the-art time series classification methods and other stress classification methods on the WESAD and SWELL-KW datasets.

6.2 Experiment 1: time series segmentation and scale-free network

There are two steps to analyzing the physiological time series data. Firstly, the physiological time series data are converted into a network using current state-of-the-art methods. These methods are time series segmentation and complex networks approaches. There is already an application in the health domain using this method (Inibhunu 2020). Physiological states are clustered into nodes, and correlations between these physiological states are transformed into edges. Secondly, the network's characteristics will be surveyed.

This experiment is based on two research hypotheses. The first is based on the fact that many real-world networks are scale-free networks (Barabási and Bonabeau 2003; Broido and Clauset 2019), so there is an expectation that the stress dataset network will also be scale-free. The second hypothesis concerns that much more data can be labelled as non-stress states than data that can be labelled as stress states. It represents the stress conditions of human beings in real life. So, after converting from time series to network, nodes of stress states have more critical roles than nodes of non-stress states in network topology. In scale-free networks, important nodes are called hubs (Barabási and Bonabeau 2003). In the network representing the stress dataset, nodes representing stress states will be hubs, and nodes representing non-stress states will be peripheral nodes. Then the hubs and peripheral nodes can be detected since hubs always have more edges. The second research hypothesis states that stress states can be detected as hub nodes in the network and non-stress states as non-hub nodes.

6.2.1 Methods to convert from time series to graph

This part analyses methods for converting time series into graphs. These methods include both time series segmentation and complex networks methods.

Time series segmentation methods To achieve the aim of clustering physiological states, some state-of-the-art time series segmentation methods such as HMM, FLOSS, EXPRESSO and ClaSP were surveyed (see section 3.2.2.2). However, all these above methods are only suitable when the time series length is about a thousand-time lags, and it is not computationally reasonable to apply to the WESAD dataset with a million-time lags. A more suitable approach, the visibility graph, was applied to this experiment. The multivariate time series is transformed into a multilayer network, and then network characteristics are analysed.

Visibility graph Lacasa et al. (2008) introduced the visibility graph method, an approach to transform from time series to graphs. This algorithm considers every time lag as a node of the graph. Consider a vertical column at each time lag of the time series, with height equal to the value at that time lag. If there is a direct line between the two highest points of vertical columns of two time lags (nodes) without any intersection of any between vertical columns of other time lags, there will be an edge between the two corresponding nodes. This direct line is called a 'visibility line' and does not intersect any data height (Lacasa et al. 2008). Assume that there is a time series, t_i is the time lag with index i, and y_i is the value of the time series at this time lag. The visibility condition can be expressed in the equation (6.1). There is a visibility line between two data values (t_a, y_a) and (t_b, y_b) if with any data value (t_c, y_c) between the two data values then

$$y_c < y_b + (y_a - y_b)\frac{t_b - t_c}{t_b - t_a} \tag{6.1}$$

Figure 6.1 shows an example with a series of electrodermal activity values. The electrodermal activity time series, time series with visibility lines, and the converted graph from the time series are shown on the left, middle, and right. The Python library ts2vg [1] was used to visualise the figure.

6.2.2 Results

Time series data from the WESAD dataset are converted into networks using the visibility graph algorithm. Each participant has eight physiological signals, so each participant's data is converted to a multilayer network with eight layers. The weight of the edge between two points is calculated using the slope of these two points. For example the weight between two points (t_a, y_a) and (t_b, y_b) is calculated using equation (6.2).

$$w_{ab} = \frac{y_b - y_a}{t_b - t_a} \tag{6.2}$$

[1] https://github.com/CarlosBergillos/ts2vg

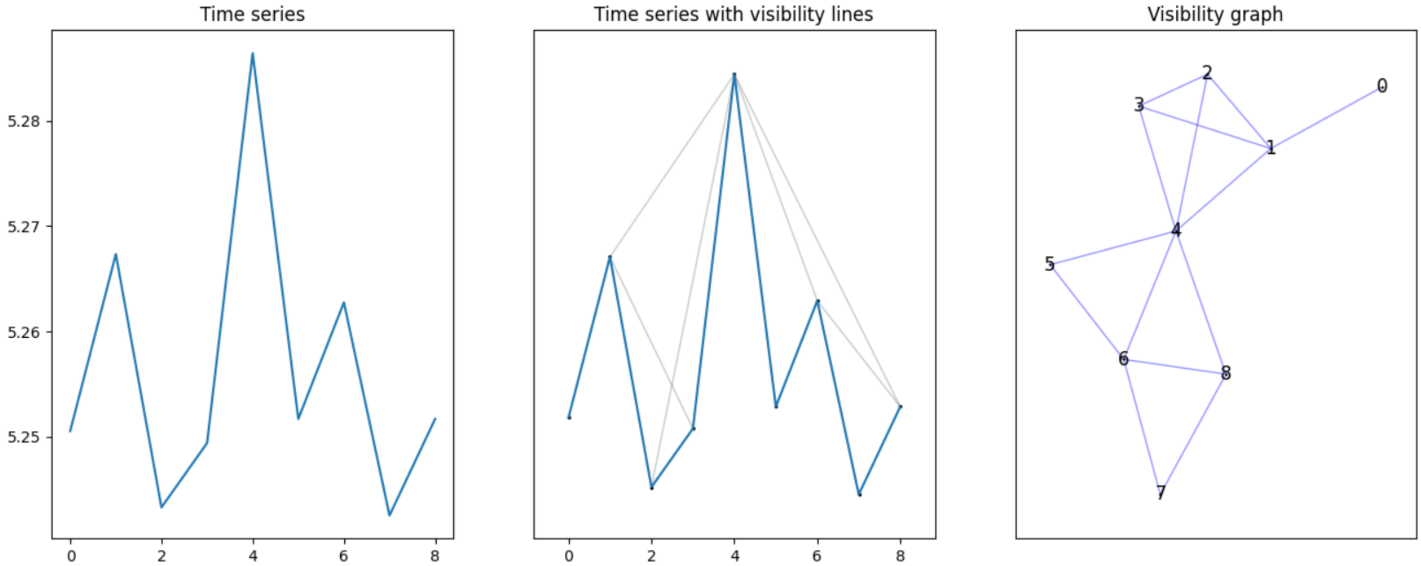


Figure 6.1: Example of a visibility graph. The left figure is the original time series with 9 time stamps. The middle figure shows the time series with visibility lines connecting the timestamps. The right figure is a graph with 9 nodes and all possible edges that connect the nodes. The grey line in Figure 6.1 presents the line in Equations 6.1 and the weight of the line is presented in Equations 6.2

Adjacency matrices can represent all layers of the multilayer network. Each participant has eight physiological features, so each participant has eight adjacency matrices. Then, all layers of the multilayer network are combined into a single network by summing up all layers' weights. Each participant has one combined network. The aim of combining all layers is to create a single network and then analyse it to confirm whether it is scale-free. If the combined network is scale-free, then hubs can be detected. The scale-free networks have log-log graphs with the shape of straight lines as described in figure 3.9 of Section 3.2.2.4. The degree distribution presents the frequency of nodes' degree values. In the degree distribution diagram, the x-axis is the node degree values, and the y-axis is the frequency of the node degree values. The log-log graphs contain logarithmic scale on both the x and y axes of the degree distribution diagram.

After the combination, the final network of participant 2 exhibits a log-log graph that is not a straight line, indicating that this network is not scale-free (see figure 6.2). The reason is six signals, x coordinate (ax), y coordinate (ay), z coordinate (az), EDA, EMG, and body temperature, can be presented as log-log graphs of scale-free networks, but the respiration data and ECG can not be (see figure 6.3 and 6.4). The combined networks of other participants are also not scale-free networks. Other participants' combined networks can be found in the Appendix A. The time series, degree distribution and log-log graphs of other participants can be found in the Appendix B. With other participants, six signals, x coordinate (ax), y coordinate (ay), z coordinate (az), EDA, EMG, and body temperature, can also represent scale-free networks. Still, other participants' respiration data and ECG can not be. Since no combined network of participants is a scale-free network, the assumption of the scale-free network (nodes of stress states will be the hubs and nodes of non-stress states will be peripheral nodes) can not be applied.

This research experiment is not a comprehensive solution to stress detection. The reason is that the combined networks are not scale-free, and stress states cannot be detected as hubs in scale-free networks within the combined networks. Some modifications can be tested to find a better solution. First, this approach can apply more sophisticated methods to combine networks rather than summing weights. The new combined networks, formed using a novel combination method, can be scale-free. In case the degree distribution of combined networks still does not follow a power law, the combined networks still can be considered scale-free networks if the degree distribution of combined networks follows some other version of the power law function (Chattopadhyay et al. 2020). This process requires calculated values of many functions. Second, the combined networks can be small-world networks. Scale-free and small-world networks both have low average shortest-path lengths (Perera et al. 2017). Differences between scale-free and small-world networks are that the degree distribution of the former is bell-shaped, whereas that of the latter follows the power law (Perera et al. 2017). Small-world networks are representations of physiological systems, such as the brain (Han et al. 2025). The small-world networks can be detected using the small-worldness index (Hosseini and Kesler 2013). Hubs can also be found in the small-world networks. However, this approach may have disadvantages. Combining the layers of the networks, then trying other versions of power law functions or calculating the small-worldness index, can lead to high computational expenses. Another solution is to extract features from each layer of the network and then analyse the network. These features are complex network features that depend on weights. These features also encode temporal dependencies, meaning that values at a specific time lag are affected by values at previous time lags. This characteristic can improve the ability to classify stress.

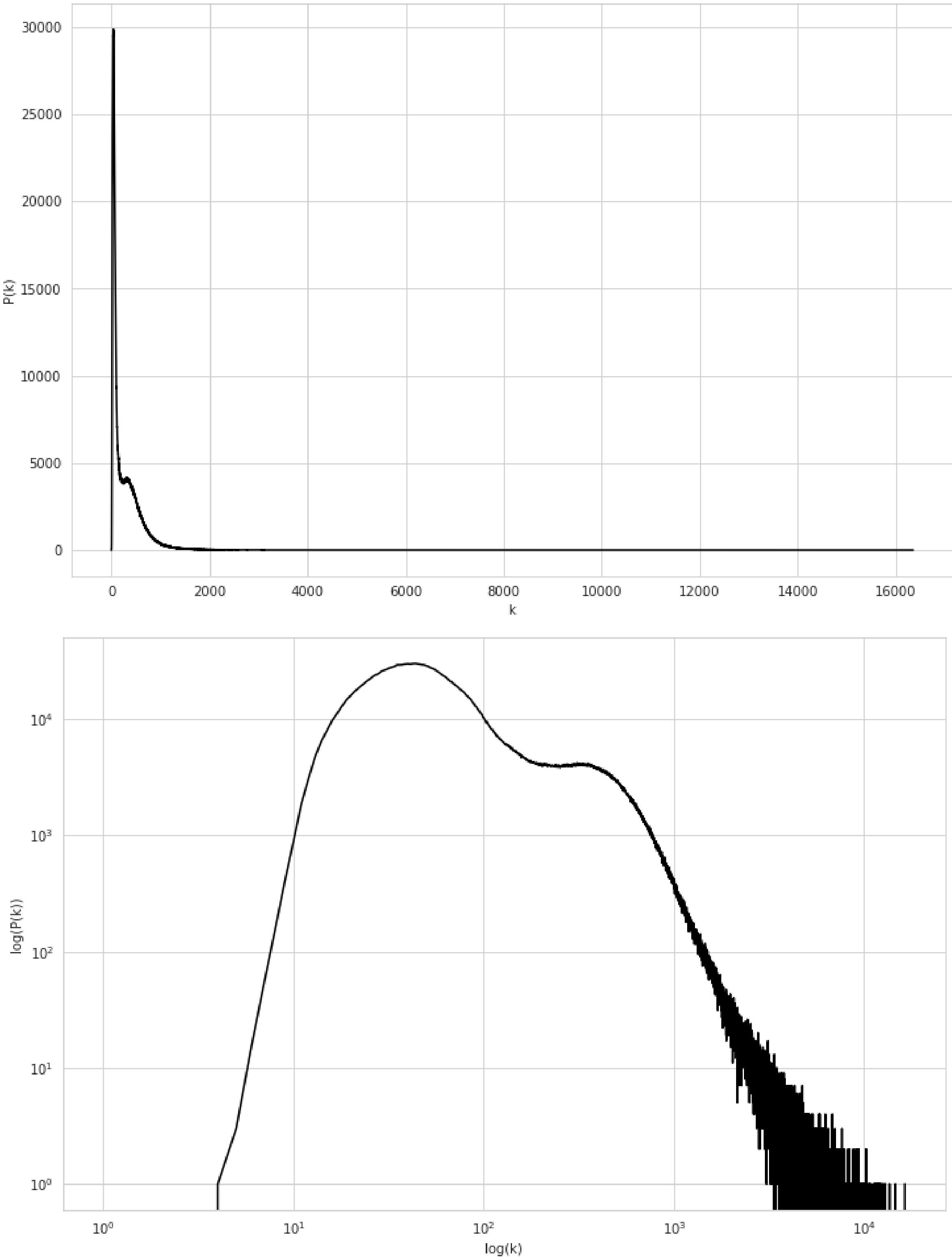


Figure 6.2: Degree distribution and the log-log graph of the combined network of the participant 2

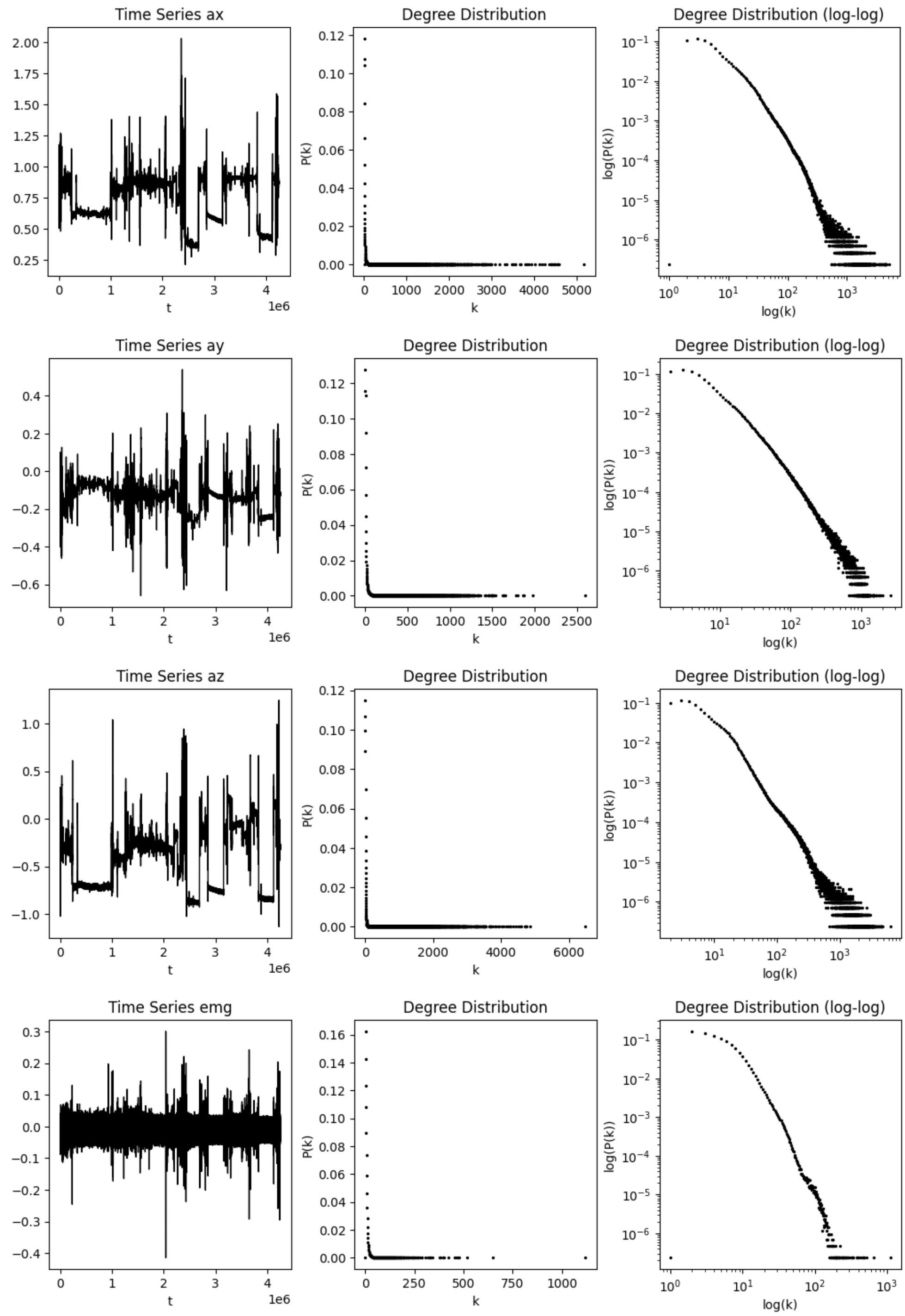


Figure 6.3: Degree distribution of networks which present physiological time series data of the first participant (participant identifier number 2) from the WESAD dataset. The data are 3-axis accelerometer data (x, y, z) and EMG. From left to right, the signal, the degree distribution, and the log-log graph

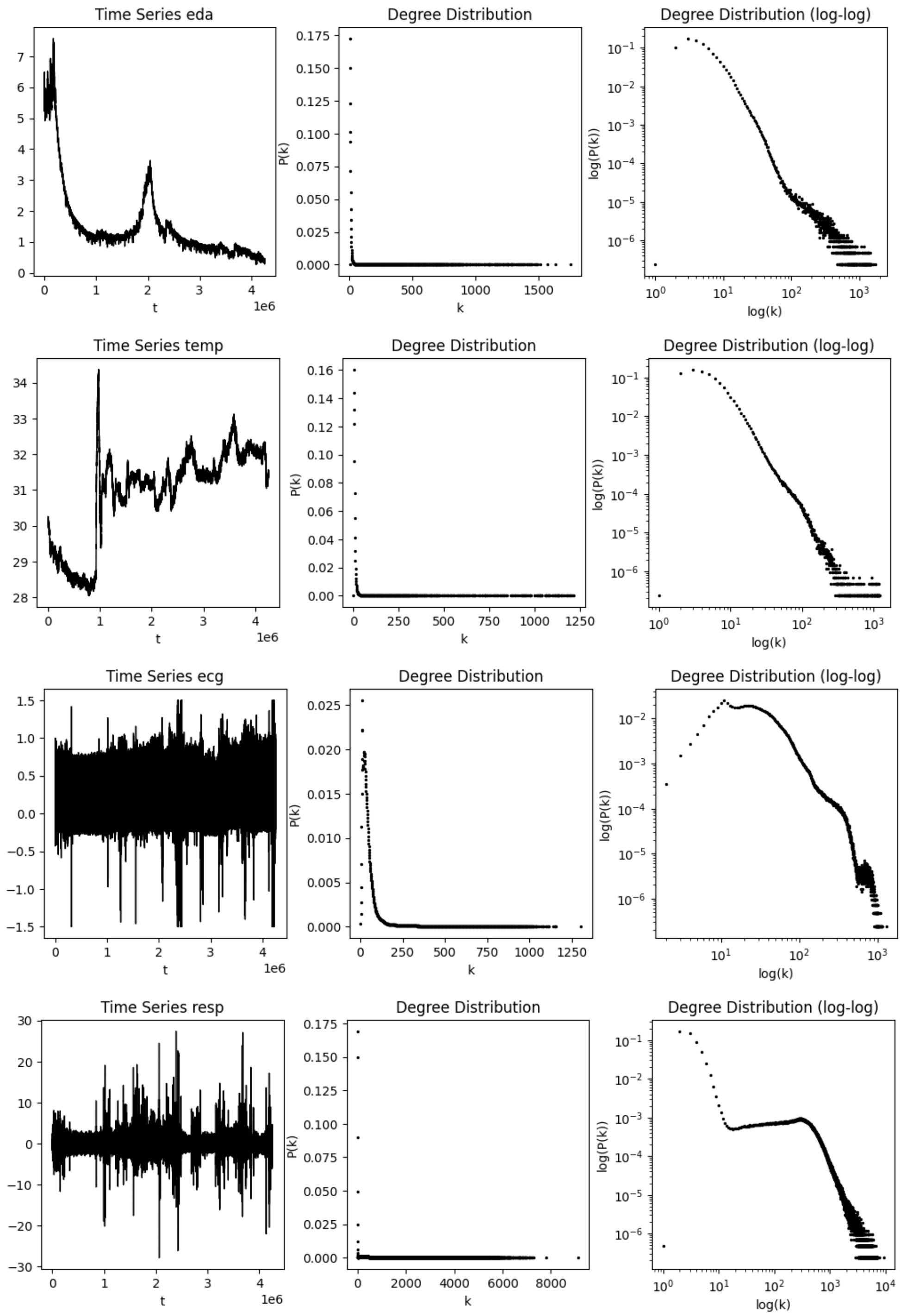


Figure 6.4: Degree distribution of networks which present physiological time series data of the first participant (participant identifier number 2) from the WESAD dataset. The data is EDA, temperature data, ECG, and respiration data. From left to right, the signal, the degree distribution, and the log-log graph

6.3 Experiment 2: complex networks features and neural networks

6.3.1 Methodology

This experiment contains four steps, as illustrated in Figure 6.5. The multivariate time series data from WESAD and SWELL-KW were converted to multi-layer networks using the visibility graph method in the first step. The second step applied a proposed weight to the multi-layer networks. In the third step, four local complex network features were extracted from each time window and used as input for the machine learning and deep learning models. Finally, in the fourth step, the models learned features from the graphs and performed stress detection.

6.3.1.1 Step 1: convert multivariate time series to visibility graphs

Time series data from the WESAD and SWELL-KW datasets were converted into networks using the visibility graph algorithm. Each participant of the WESAD and SWELL-KW datasets has eight and three physiological signals, respectively. WESAD contains data on galvanic skin response (GSR), electrocardiogram (ECG), electromyogram (EMG), respiration rate, skin temperature, and a 3-axis accelerometer. SWELL-KW has heart rate, electrocardiogram (ECG) and electrodermal activity (EDA). Each participant's data was converted to two multi-layer complex networks with eight and three layers, respectively. The number of layers in these two multi-layer networks equals the number of physiological signals in the WESAD and SWELL-KW datasets. The network of each layer contains the same collections of nodes, and each node corresponds to a timestamp of the time series. Each network from different layers of the multi-layer networks has different edges, and the visibility graph algorithm determines the existence of edges. To reduce computational time, a divide-and-conquer method was used to construct the visibility graphs (Lan et al. 2015).

6.3.1.2 Step 2: build weighted networks

The weight of the edge between two points is calculated using the slope between these two points. For example the weight between two points (t_a, y_a) and (t_b, y_b) is calculated using equation (6.3).

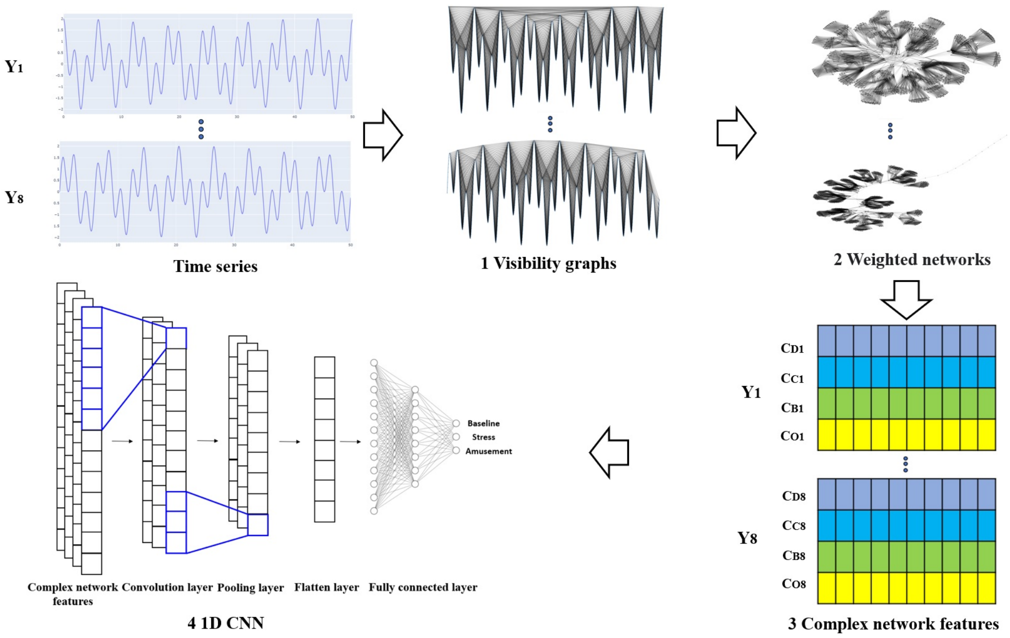


Figure 6.5: Four steps of this proposed method. Step 1 converts multivariate time series data to networks. Step 2: create weighted networks. Each network node corresponds to a physiological time series timestamp. The weight of an edge is the angle between two connected nodes that make up this edge. Step 3 extracts four complex network features: degree centrality (C_D), closeness centrality (C_C), betweenness centrality (C_B), and clustering coefficient (C_O). These features are defined from the equation 6.4 through 6.7. Step 4 classifies emotional states.

$$w_{ab} = \frac{y_b - y_a}{t_b - t_a} \tag{6.3}$$

From the networks, only time stamps with one of the three labels (baseline, stress, amusement for WESAD and neutral, stressed from time pressure, stressed from interruption for SWELL-KW) were used to extract features in the next step.

6.3.1.3 Step 3: extract complex networks features

This step applied a fixed-length time window to the time series. This experiment extracted four complex networks features: degree centrality, closeness centrality, betweenness centrality, and clustering coefficient from each time window (Barrat et al. 2008; da F. Costa et al. 2007). Let $G(V, E)$ be a graph, and V, E be the collection of nodes and edges, respectively. $|V|$ is the number of nodes.

The *degree centrality* of a node is the number of edges that connect to this node (Boudin 2013). This feature can be normalised using degree centrality as input to the deep learning model for stress detection. Let $N(V_i)$ be the number of nodes that connect to node i. The degree centrality value is normalised by dividing by the maximum possible number of edges of one node, using

$$C_D(V_i) = \frac{N(V_i)}{|V| - 1} \tag{6.4}$$

The *closeness centrality* of a node is the inverse of the average length of all shortest paths between this node and all other nodes in the graph (Boudin 2013). Let $distance(V_i, V_j)$ be the shortest distance between two nodes i and j. The formula of closeness centrality is

$$C_C(V_i) = \frac{|V| - 1}{\sum_{V_j \in V} distance(V_i, V_j)} \tag{6.5}$$

Betweenness centrality is the number of times a node becomes a bridge of the shortest path between two other nodes (Boudin 2013). Let $\sigma(V_j, V_k)$ be the number of shortest paths between two nodes j and k, and $\sigma(V_j, V_k|V_i)$ is the number of shortest paths between two nodes j and k that also go through node i. The formula for the betweenness centrality of node i is

$$C_B(V_i) = \frac{\sum_{V_i \neq V_j \neq V_k \in V} \frac{\sigma(V_j, V_k|V_i)}{\sigma(V_j, V_k)}}{(|V| - 1)(|V| - 2)/2} \tag{6.6}$$

The *clustering coefficient* of a node shows how all connected nodes to this node and all related edges of these connected nodes become a complete graph (Saramäki et al. 2007). For example, node i of the graph connects with k_i nodes. Let $G_i(k_i, E_i)$ be the sub graph that contain all k_i nodes and $|k_i|$ is number of nodes of this sub graph. If G_i is a complete graph then E_i equal to $(|k_i|)(|k_i| - 1)/2$. The formula of the clustering coefficient of the node i is (6.7) :

$$C_O(V_i) = \frac{2E_i}{(|k_i|)(|k_i| - 1)/2} \tag{6.7}$$

In this part, an example of the four complex networks features is given using a specific network topology (Pavlopoulos et al. 2011) with seven nodes and six edges (see Fig. 6.6). Node V_1 will have degree centrality $C_D(V_1) = 4/6$ since there are four nodes connected with node V1 and the total number of nodes are 7. The clustering coefficient of node V_1 $C_0(V_1) = 0/6$ since with the subgraph with four nodes V_2, V_5, V_6, V_7 does not connect. Meanwhile, the total number of possible

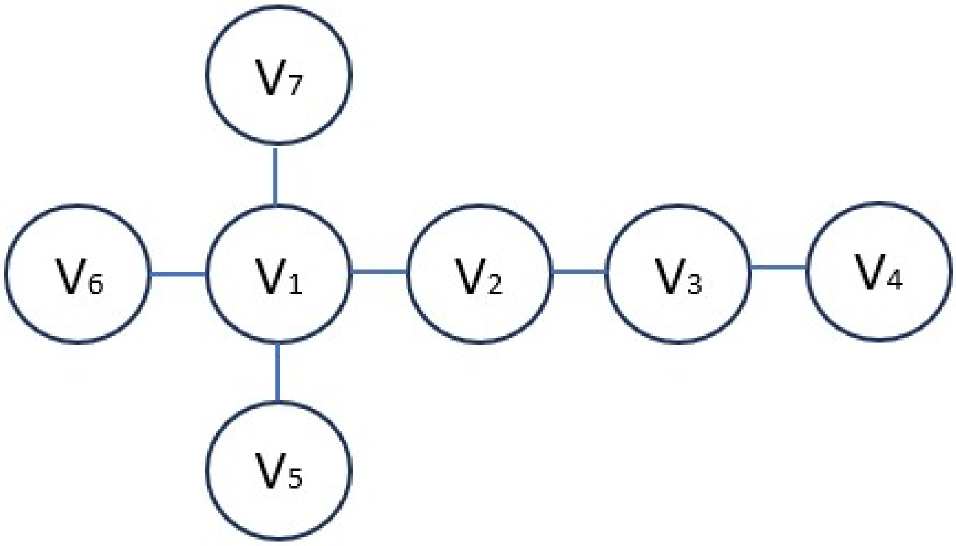


Figure 6.6: Example of a network.

edges is 4*3/2. The closeness centrality is equal to 6/(4*1+1*2+1*3). Since the shortest distance from node V_1 to four nodes V_2, V_5, V_6, V_7 is 1, to node V_2, V_5 is 2 and 3 respectively. The betweenness centrality is equal to 12/25. Twelve shortest paths go through node V_1, and there are only the shortest paths between corresponding nodes. The list of twelve shortest paths is V_2 - V_5, V_2 - V_6, V_2 - V_7, V_3 - V_5, V_3 - V_6, V_3 - V_7, V_4 - V_5, V_4 - V_6, V_4 - V_7, V_5 - V_6, V_6 - V_7, V_5 - V_7.

To reduce computational expense and reduce running time, betweenness centrality and clustering coefficient were calculated using approximation methods of Cohen et al. (2014) and Riondato and Kornaropoulos (2014), respectively.

6.3.1.4 Step 4: machine learning and deep learning models

With each time series, four features are extracted: degree centrality, closeness centrality, betweenness centrality, and clustering coefficient. Two machine learning models are set up for the experiment: one Random Forest and one neural network. Two models were validated on the WESAD dataset. Each model was trained separately on each participant's time series data, using a 60/20/20 split for training, validation, and testing, respectively.

Random Forest and a simple neural network were used, with the features as inputs, to classify stress. The first one is a Random Forest. With the aim of improving the result of Random Forest, grid search and Bayes optimization are

Table 6.1: The parameters of Grid search

Parameters	Ranges of value	Best value
number of trees	500, 1000, 2000	2000
number of features to consider when looking for the best split	auto, sqrt, log2	auto
maximum depth of the tree	10, 20, 50, 100	50
function to measure the quality of a split	gini, entropy	entropy

Table 6.2: The parameters of Bayes optimization

Parameters	Ranges of value	Best value
number of trees	500, 1000, 2000	2000
number of features to consider when looking for the best split	auto, sqrt, log2	sqrt
maximum depth of the tree	from 1 to 100	92
function to measure the quality of a split	gini, entropy	gini

applied for hyperparameter optimisation. The parameters of the two optimization methods of participant 2 are listed in Table 6.1 and Table 6.2. The second one is a neural network with four fully connected layers. The results of all methods of participant 2 are listed in Table 6.3. The results showed that traditional machine learning models can not achieve very good results on this time series data. In terms of F1 of stress detection for participant 2, the method achieved the highest result of 0.4594. Meanwhile, current methods of (Alshamrani 2021; Lisowska et al. 2021; Tiwari et al. 2023) have already achieved F1 scores of 0.732, 0.88, and 0.95, as reported in the respective papers. Deep learning models specialised for time series, such as 1D CNN, need to be evaluated to achieve better results. Due to the long-running time of this experiment (14 hours for just participant 2) and the F1 score was not high, this experiment was only tested with one participant. This experiment assumes that the data of participant 1 are representative of the dataset.

The deep learning model in this research contains three 1D convolutional

Table 6.3: The results of machine learning models using complex networks features

Models	F1 baseline	F1 stress	F1 amusement
Random Forest	0.8040	0.4561	0.7119
Random Forest Grid Search	0.7523	0.4594	0.4618
Random Forest Bayes optimization	0.7524	0.347	0.4627
Neural network	0.7687	0.3247	0.6270

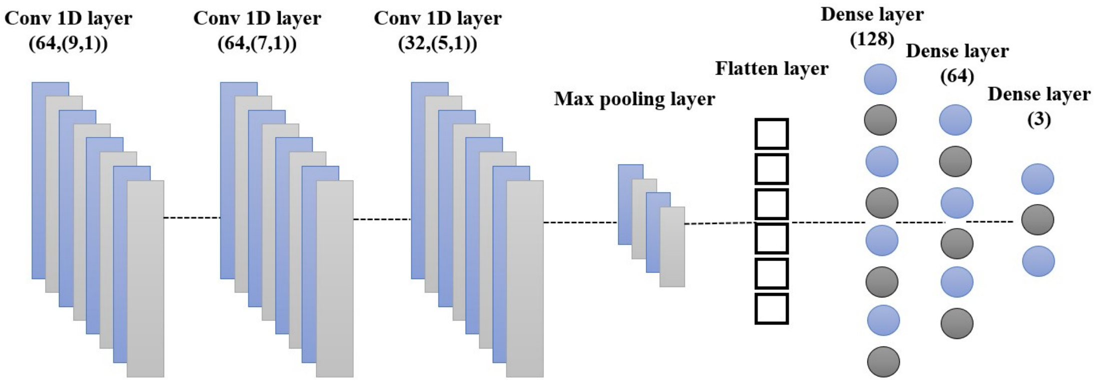


Figure 6.7: Topology of 1D CNN model.

layers, one max pooling layer, one flattened layer, and three dense layers (Fig. 6.7). The pooling layer reduces the number of built features, allowing the model to focus on essential parts of features. Constructed features will be transformed into 1-dimensional data using the flattening technique, and one-dimensional data will be used as input for two dense layers, then passed to the output layer. The deep learning model will apply the Adam optimiser with the value of the learning rate of 5×10^{-4}. The model uses the loss function of categorical cross-entropy to handle multi-class classification. The number of epochs is 20. This deep learning model classifies into the three emotional states for both WESAD and SWELL-KW. The output data is transformed into a category with a one-hot encoding. The model's performance was assessed using balanced accuracy and F1 score. The dataset was unbalanced, so balanced accuracy and F1 score are suitable to determine the ability of the deep learning model to detect stress.

6.3.2 Results

There were three experiments run. The first experiment was run to find the time window that can provide optimal results for the 1D CNN model. The second experiment applied complex network features to the 1D CNN model for each of the 15 WESAD participants and the 18 SWELL-KW participants. The third experiment compared the model results with state-of-the-art multivariate time series classifiers.

In the first experiment, 1D CNN models and complex networks features were trained separately with each participant's data. Each participant's training, validation, and test sets are 60%, 20%, and 20% of each participant's data, without shuffling or overlap. The models ran three times for each participant, and the final results were the average results of 15 participants (Table. 6.4) of the WESAD dataset and 18 participants of the SWELL-KW dataset (Table. 6.5). Several time window sizes were evaluated to identify the optimal configuration for the deep learning model. The value of the stride of the time window remains at 1. With WESAD, the collection of time window sizes contains 35, 70, 140, 210, 280, and 350 observations, equivalent to 0.05, 0.1, 0.2, 0.3, 0.4, and 0.5 seconds, respectively. The table shows optimal results for a time window size of 350 observations (0.5 seconds). With SWELL-KW, the collection of time windows is 32, 64, 128, and 192, corresponding to 0.25, 0.5, 1, and 1.5 seconds, respectively. The optimal time window for SWELL-KW is 192 observations (1.5 seconds). The optimal time windows of the two datasets are the largest windows. It was not computationally possible to explore longer time windows.

With the second experiment, the 1D CNN models were trained with the optimal time window size of 350 observations equivalent to 0.5 seconds for WESAD (Table. 6.6) and 192 observations equivalent to 1.5 seconds for SWELL-KW (Table. 6.7), and the stride length was 1. The last row of each of the tables shows the average results for all participants. The participants' results are pretty consistent.

With the third experiment, the average results were compared with the average results of other state-of-the-art time series classification methods, namely Time Convolutional Neural Network (Time CNN) and Fully Connected Neural Network (FCN) (Zhao et al. 2017), Multi-Layer Perceptron Network (MLP) (Wang et al. 2017), LSTMFCN (Karim et al. 2019), Time series attentional prototype

Table 6.4: Average emotional classification results when training the model 3 times using various window sizes of WESAD

Window size	F1 baseline	F1 stress	F1 amusement	Balanced accuracy
35	0.841	0.702	0.523	0.746
70	0.885	0.79	0.689	0.823
140	0.951	0.911	0.874	0.927
210	0.983	0.968	0.954	0.973
280	0.993	0.988	0.982	0.989
350	**0.997**	**0.991**	**0.989**	**0.994**

Table 6.5: Average emotional classification results when training the model 3 times using various window sizes of SWELL-KW

Window size	F1 neutral	F1 time pressure stress	F1 interruption stress	Balanced accuracy
32	0.701	0.583	0.7	0.701
64	0.857	0.775	0.837	0.857
128	0.968	0.95	0.961	0.968
192	**0.989**	**0.982**	**0.985**	**0.989**

network (TapNet) (Zhang et al. 2020b), InceptionTime (Fawaz et al. 2020) and ResNet (Wang et al. 2017). All the state-of-the-art methods were deployed using the library `sktime`. Default parameter settings from `sktime` were used. The results for WESAD, SWELL-KW are shown in Tables 6.8 and 6.9. Our proposed methods achieved the best results in both stress response and amusement classification, and the second-best in baseline classification, and balanced accuracy on the WESAD dataset. The scores for baseline classification and balanced accuracy are almost the same as those of the best methods, TapNet and ResNet, respectively. Using good combinations of input features, our 1D CNN model with simple topology can already perform as well as very complicated deep learning models such as Tapnet and Resnet (with 35 layers and 50 layers, respectively). In the case of SWELL-KW, our proposed method gained the best results regarding both F1 scores for classifying stress because of time pressure and interruption. The methods also achieved the best results in terms of balanced accuracy.

The average F1 score for stress classification and balanced accuracy of our

Table 6.6: The emotional classification results for each participant of WESAD.

Participant id	F1 baseline	F1 stress	F1 amusement	Balanced accuracy	Time
2	1	0.993	1	0.993	3h2m
3	1	0.993	1	1	3h45m
4	0.993	0.993	0.98	0.993	3h2m
5	1	0.997	0.99	0.997	3h39m
6	1	0.99	1	1	3h10m
7	0.993	0.983	0.967	0.987	3h15m
8	0.997	0.993	0.987	0.997	3h16m
9	0.997	0.987	1	0.99	3h21m
10	0.99	0.987	0.973	0.987	3h14m
11	0.993	0.987	0.997	0.993	3h8m
13	1	0.997	0.987	0.997	3h49m
14	1	0.993	1	1	3h14m
15	0.997	0.977	0.963	0.983	3h 4m
16	1	0.99	1	0.997	3h52m
17	1	1	0.99	1	3h26m
Average	0.997	0.991	0.989	0.994	3h21m50s

proposed methods were also compared with other published WESAD stress classification methods. On these methods, four used deep learning models (Alshamrani 2021; Gupta et al. 2021; Huynh et al. 2021; Lisowska et al. 2021) and one used classical machine learning model (Tiwari et al. 2023). Our methods gain the best F1 score and balanced accuracy compared with other methods.

Table 6.7: The emotional classification results for each participant of SWELL-KW. The '-' sign in the table indicates that the corresponding labels are not given for this participant.

Participants id	F1 neutral	F1 time pressure stress	F1 interruption stress	Balanced accuracy
1	0.99	0.99	0.9867	0.9867
2	0.9967	1	1	1
3	-	0.9933	0.9933	0.9933
4	1	1	-	1
5	1	0.9867	0.9967	0.9967
6	0.99	0.98	0.9867	0.9867
7	0.9867	0.96	0.9833	0.9767
9	1	0.99	0.9967	0.99
10	1	1	-	1
12	1	0.99	0.99	0.99
13	0.99	0.99	0.9867	0.99
14	0.9867	0.95	0.9833	0.9833
16	0.9933	0.9933	1	0.9967
17	0.99	0.9767	0.9933	0.9867
18	0.9533	0.9467	0.9533	0.9533
19	0.97	0.9633	1	0.9933
21	0.9733	0.9633	0.9233	0.9633
24	0.99	0.9967	-	0.9967
Average	0.989	0.982	0.985	0.989

Table 6.8: The classification results of state-of-the-art multivariate time series classification methods and this proposed method on WESAD dataset

Methods	F1 baseline	F1 stress	F1 amusement	Balanced accuracy
Time CNN	0.907 ± 0.039	0.71 ± 0.122	0.63 ± 0.176	0.825 ± 0.077
Full CNN	0.926 ± 0.045	0.753 ± 0.161	0.716 ± 0.187	0.864 ± 0.097
MLP	0.696 ± 0.006	0.022 ± 0.039	0 ± 0	0.538 ± 0.011
LSTMFCN	0.973 ± 0.035	0.972 ± 0.014	0.931 ± 0.033	0.968 ± 0.024
InceptionTime	0.997 ± 0.005	0.973 ± 0.005	0.914 ± 0.028	0.979 ± 0.006
TapNet	0.996 ± 0.006	0.98 ± 0.018	0.984 ± 0	$\mathbf{0.995 \pm 0.001}$
ResNet	$\mathbf{0.998 \pm 0}$	0.972 ± 0.002	0.935 ± 0.031	0.982 ± 0.004
Proposed method	0.997 ± 0.001	$\mathbf{0.991 \pm 0.002}$	$\mathbf{0.989 \pm 0.002}$	0.994 ± 0.002

Table 6.9: The classification results of state-of-the-art multivariate time series classification methods and this proposed method on SWELL-KW dataset

Methods	F1 neutral	F1 time pressure stress	F1 interruption stress	Balanced accuracy
Time CNN	0.483 ± 0.094	0.298 ± 0.05	0.527 ± 0.059	0.647 ± 0.043
Full CNN	0.848 ± 0.029	0.768 ± 0.021	0.821 ± 0.036	0.865 ± 0.018
MLP	0.29 ± 0	0.125 ± 0	0.357 ± 0	0.546 ± 0
LSTMFCN	0.859 ± 0.02	0.752 ± 0.067	0.85 ± 0.007	0.874 ± 0.018
InceptionTime	0.944 ± 0.009	0.888 ± 0.029	0.952 ± 0.012	0.954 ± 0.01
TapNet	0.741 ± 0.018	0.577 ± 0.061	0.752 ± 0.016	0.768 ± 0.023
ResNet	0.888 ± 0.02	0.811 ± 0.043	0.843 ± 0.059	0.893 ± 0.028
Proposed method	$\mathbf{0.989 \pm 0.003}$	$\mathbf{0.982 \pm 0.002}$	$\mathbf{0.985 \pm 0.001}$	$\mathbf{0.989 \pm 0}$

Table 6.10: The classification results of other publications and this proposed method on the WESAD dataset. The sign '-' in the table presents the corresponding values that are not given.

Methods	Average F1 score	Average accuracy
Lisowska et al. 2021 Lisowska et al. (2021)	0.705	-
Huynh et al. 2021 Huynh et al. (2021)	-	0.835
Alshamrani 2021 Alshamrani (2021)	0.84	0.85
Gupta et al 2021 Gupta et al. (2021)	-	0.854
Tiwari et al. 2023 Tiwari et al. (2023)	0.98	0.98
Proposed method	**0.991**	**0.994**

6.4 Conclusions

This chapter presents complex network approaches for detecting stress responses in long, multivariate physiological time series data. The first experiment uses network topology and complex network features, combined with machine learning, to detect stress. The main contribution of this chapter is the second experiment, which presents a novel method for complex networks to detect stress. This method chose the suitable complex networks representation and features to capture information from the time series data, and also had a suitable 1D CNN model to learn from these features. The optimal time window size for the complex network is shorter than that of the TDA method (0.5 seconds compared to 1,1 seconds, respectively). This complex networks method has an average running time of just 30% compared with the TDA stress detection method on the WESAD data set (3 hours 21 minutes 50 seconds compared to 11 hours 58 minutes 48 seconds). The F1 scores of the two methods are almost identical; the average balance accuracy of the TDA method is slightly better than that of the complex networks method. This proposed method has already achieved excellent results compared with state-of-the-art multivariate time series classification methods and other stress classification methods on the WESAD and SWELL-KW datasets. Our method is a personalised stress response detection model. According to Koldijk et al. (2016); Schmidt et al. (2018a,b), different individuals exhibit different physiological responses to stress, so stress detection methods should be personalised. The limitation of this method is that it is not a general stress detection method that uses participants' physiological data to infer the stress of other participants.

Chapter 7

Conclusion

Stress is a dangerous threat to people. Indicators of a negative stress response can be detected from physiological data collected using wearable devices. There are many stress detection methods for physiological time series. However, these methods still have many limitations. Many methods use traditional machine learning or statistical techniques to detect or find relationships between features and stress. More advanced methods, such as deep learning, can provide better insights into emotional states. Many studies apply only simple handcrafted features to physiological time series data. State-of-the-art data analytic methods, such as topological data analysis and complex networks, which capture the “shape” of data and analyse complex systems, are promising approaches for generating more complete features for detecting stress. Many stress detection methods also do not provide comprehensive comparisons with other methods on benchmark datasets. This can not validate the effectiveness of the methods. There is an urgent demand for developing novel stress detection methods that can overcome the current limitations of current stress detection methods.

This PhD thesis has three research questions. The experiments in chapters 4, 5, and 6 address the three research questions and limitations of previous stress detection research.

Research question 1 : How can we detect stress using conventional machine learning and statistical methods, deep learning models with the outer product of physiological time series as inputs or deep learning models with image-based inputs?

Answer: Chapter 4 provides three stress detection approaches using conventional machine learning and statistical techniques, a deep learning model with the outer product of physiological time series data as input, and a deep learning model with image-based data as input. The first approach uses conventional machine learning and statistical methods to classify stress states. The second and third approaches utilise the outer product of the time series signal and image-based data as input to a 2D CNN model to detect stress. The classification results are not good since conventional transformation methods may not provide good features for deep learning models to perform classification tasks.

Research question 2 : Is it a promising approach to detect negative stress responses in physiological time series data using topological data analysis (TDA)?

Answer : Chapter 5 provides a comprehensive approach to detect stress from long multivariate physiological time series data by extracting features using TDA, which were then input into a 1D CNN. The TDA approach to classify stress has already overcome some of the limitations of TDA that prevented its widespread adoption for handling massive amounts of data, such as long multivariate time series. First, this TDA approach used a custom Taken embedding to map high-dimensional multivariate time series data to a lower dimension. This allows the approach to become faster and easier when saving, processing and analysing TDA data. Finally, this approach provided a formulation for finding the optimal filtration parameter for each participant's data. Setting a threshold on the filtration parameter is preferable to using the maximum value. It can avoid handling large, complicated and fully connected complexes. This approach achieved better classification performance than other stress detection methods and state-of-the-art time series classification methods on the same benchmark dataset.

Research question 3 : Is it a promising approach to detect negative stress responses in physiological time series data using complex networks?

Answer : Chapter 6 provides a comprehensive complex networks approach to detect stress from time series data. The current state-of-the-art in complex network research primarily applies to short or univariate time series data.

This approach provides a complete complex networks method to handle long multivariate physiological time series data. The reasons are the difficulty of mapping multivariate time series to networks while preserving information, the high dimensionality, and the computational complexity. The method developed a complex networks representation of multivariate physiological data and then extracted complex network features for input into a 1D CNN model. This method has several advantages when classifying multivariate time series data. Firstly, it provides network representation that can preserve the information of multivariate data. Secondly, it addressed the problem of high dimensionality and computational complexity through feature extraction. It converts from a multivariate time series to a low-dimensional set of complex network features. It also reduces computational complexity, since deep learning models need only work with a set of features rather than the entire time series. Thirdly, this method has the ability to provide good stress classification results. It classifies the emotional states associated with each node of the time series, which is also an individual node of the network. The complex networks features of this method are degree centrality, closeness centrality, betweenness centrality, and clustering coefficient. These features represent important levels of the individual nodes of the graph and, as a result, can improve classification power. The first three features are centrality features, which contain information about a node: how many connections this node has, how often a node lies on the shortest path between other nodes, and how close a node is to all other nodes. The local clustering coefficient feature contains information about how interconnected the neighbours of this node are. The classification results of this complex networks method are good when validated on benchmark datasets and when compared with other stress detection methods and state-of-the-art time series classification methods on the same benchmark datasets.

The research in this thesis also contributes to the knowledge, corresponding to the contributions stated in chapter 1.

Contribution 1: Validated stress detection methods that use conventional machine learning and statistical methods, and stress detection methods with inputs that are the outer product of physiological time series data or image-based data.

This contribution of this PhD thesis presents three stress detection approaches. The first approach uses conventional machine learning and statistical approaches. The second and third approaches utilise a deep learning model with inputs that are the outer product of physiological time series data or image-based data. All these approaches are validated on benchmark datasets. The experiments of this contribution are reported in chapter 4. This chapter first comprehensively compares the use of LSTM, RNN, and VAR for detecting stress. This experiment is the first comparison of machine learning and statistics in multivariate physiological time series. This chapter presents an experiment to detect stress using a 2D CNN model trained on the outer product of physiological time series. The classification result is not suitable when validated on a benchmark dataset. This chapter continues to introduce a method for detecting stress using a 2D CNN model and image-based data. Image-based data are generated from physiological time series using the recurrent plot method. However, because the recurrent plot method is susceptible to noise, which often occurs in physiological time-series data, its classification performance is also low. Due to the complexity and noisy nature of physiological data, standard machine learning or statistical methods aren't suitable; therefore, there's a need for data transformation. Simple transformations of the data, using the outer product of the time series signal and image-based data, are also insufficient for stress detection. More advanced methods for data transformation, such as topological data analysis and complex networks, can yield better stress classification results.

Contribution 2: A novel topological data analysis approach to detect stress. This method is validated on benchmark datasets, compared with stress detection methods on the same datasets and compared with time series classification methods.

Statistics, machine learning, and novel data analytics approaches have been used to detect stress in physiological time series data. However, such data is noisy, which can limit the effectiveness of algorithms. Topological data analysis (TDA) is a novel approach that can handle noisy data and may be promising for analysing physiological time series data. However, TDA is currently in the early stages of development, with researchers still grappling with the problem of feature extraction from TDA signatures for machine learning. The current state-of-the-art in TDA handles only small-scale computer vision or univariate time series data analysis due to its computational expense. This contribution presents a TDA method for stress detection and validates it on the public Wearable Stress and Affect Detection (WESAD) dataset. The method achieved an F1 score of 0.991 for stress on the WESAD dataset. This contribution is a complete TDA method to classify long multivariate physiological time series data that overcomes some of the computational expenses and demonstrates its effectiveness compared to state-of-the-art methods. The current limitation of this method is that it is a personalised stress detection method, which can detect stress for each participant. This can be improved by identifying similarities among participants in age, gender, and occupation, and applying a single model to a specific group of participants. Another limitation is the high running time. This can be improved by reducing the size of the point cloud and the computational cost of computing persistent homology from the point cloud. The experiment of this contribution is reported in chapter 5.

Contribution 3: A novel complex networks method to detect stress. This method is validated on benchmark datasets and compared with stress detection methods on the same datasets and time series classification methods.

Currently, state-of-the-art applications of complex networks to time series data are limited to clustering or classifying short time series. This contribution is a technique for utilising complex networks to identify stress from physiological time series data. It is validated on two publicly available datasets: Wearable Stress and Affect Detection (WESAD) and The Smart Reasoning for Well-being at Home and at Work Knowledge Work (SWELL-KW). WESAD and SWELL-KW both contain long and multivariate time series data. This contribution has two accomplishments. Firstly, this contribution proposes a novel representation that can capture the information of long multivariate time series data. This representation uses the visibility graph method, and weighted networks combined with local complex networks features. The second accomplishment is achieving an F1 score of 0.991 ± 0.002 for stress response detection in the WESAD dataset, and 0.982 ± 0.002 and 0.985 ± 0.001 for stress responses caused by time pressure and interruption, respectively, in the SWELL-KW dataset. This method also achieved the best stress-response detection score among state-of-the-art multivariate time series classification methods. This method has a limitation: it is a personalised stress-detection method. The experiment of this contribution is reported in chapter 6.

7.1 Future research directions

With increased access to wearable devices and improvements in the quality of their collected physiological data, stress detection from wearables is an important research direction for the future. This PhD thesis provides novel approaches to detect stress from physiological signals collected using wearable devices. The TDA approach and the complex networks, combined with the deep learning approach, both achieve better classification results than deep learning models with the outer product of physiological time series as inputs and deep learning models with image-based inputs. Compared to the TDA approach, the ideal time window size for the complex network is smaller (0.5 seconds against 1.1 seconds, respectively). When compared to the TDA stress detection approach on the WESAD data set, this complex networks method has an average running duration of only 30% (3 hours 21 minutes 50 seconds against 11 hours 58 minutes 48 seconds). Although the two approaches' F1 scores are nearly identical, the TDA method's average balance accuracy is marginally higher than the complex networks method's. However, some new research directions can still address these methods' limitations.

Firstly, novel image-based input generated from the recurrence plot can be used in future stress detection methods. A 2D CNN with a simple design was used with this input in the experiment of this thesis. Applying a more complex 2D CNN structure or a more advanced computer vision model can improve performance.

Secondly, the computational cost of the TDA method is high, with an average of almost 12 hours per participant on the WESAD dataset. The shortest and longest times to classify a participant are 9 hours and 25 minutes and 12 hours and 55 minutes, respectively. Future work can reduce the running time of the TDA method. The running time can be reduced by decreasing the computational expenses of the TDA process. The input to the TDA process, the point cloud, can be applied to principal component analysis (PCA) or to Mapper to reduce the input data size. Mapper with the ability to detect and retain global mathematical features (Cyranka et al. 2019) may be a better choice for reducing the size of input features. Subsampling methods can then be used to reduce computational cost when building persistent homology from the point clouds. Subsampling methods can work with fractions of a point cloud, preserve all topological aspects, and reduce computational expense.

Thirdly, both the TDA approach and the complex networks yielded promising results, but these approaches are individual stress detection methods. The thesis shows that different classifiers need to be trained for each individual: a single classifier performs poorly across individuals. The reason is that stress responses vary across individuals, so individual stress detection methods are more appropriate (Koldijk et al. 2016; Schmidt et al. 2018a,b). Future work can determine the similarity between participants and, based on this, develop a more generalised machine learning stress detection method. Some factors, such as age, sex, occupation, and physical condition, can be used to group participants in stress detection experiments. Stress detection methods will be developed for each participant group to improve the quality of stress detection and generalisation.

Fourthly, the combined network approaches did not work since the combined ones are not scale-free. Instead of adding weights, this method can combine networks using more complex techniques. Using a revolutionary combination technique, the new merged networks can be scale-free. Even if the degree distribution of the merged networks does not follow a power law, they can still be regarded as scale-free networks if it exhibits a power-law-like form. Small-world networks may be a suitable representation of the combined networks instead of scale-free networks. Small-world networks are also similar to scale-free networks. The average shortest-path length is modest for both small-world and scale-free networks. Scale-free networks have a power-law degree distribution, whereas small-world networks have a bell-shaped degree distribution. The small-worldness index can be used to identify small-world networks. The small-world networks also contain hubs. Hubs can be stress states in the stress system's networks.

Fifthly, more research can be conducted to explore the physiological benchmark datasets and better understand the signals. This can indicate which physiological signals are more important than others for stress detection. Statistical tests such as Granger Causality and Mutual Information (MI) can determine which physiological features have more influence on stress. Gradient-based importance approaches (such as Saliency Maps, Integrated Gradients (IG)), or a Transformer, can be used to assess feature importance. During training, the gradient of the output can be computed from the gradient of the input. This can help determine which fraction of the input matrix has the most significant effect on specific tasks. It is a mechanism for deciding feature importance. Transformer with Attention Weights, which can be associated with a fraction of the time series input, indicating relevance with the current processing step. Then the Transformer can identify which parts of the input data have a more critical impact on the final classification results. Posthoc approaches such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) can also be helpful in this case.

Finally, the complex network stress detection method can be further developed into a real-time stress classification method, as it has an acceptable level of complexity. The reason is that the algorithm used short-time windows (0.5 and 1.1 seconds) for two datasets and applied the visibility graph method, which is parameter-free and does not require additional preprocessing.

7.2 Summary

This PhD thesis makes three contributions to knowledge. The first contribution is several simple methods for stress detection. The first method uses LSTM, RNN and VAR to predict stress. It can be a comprehensive comparison of machine learning and statistical methods for predicting stress. The second method uses a 2D CNN and a time window approach to detect stress from the outer product of physiological data. The third one uses a 2D CNN and image data generated from time series data. However, neither approach has achieved good classification results. This led us to continue seeking more suitable methods for stress detection using advanced data analytic approaches, such as topological data analysis and complex networks. The second contribution is a novel topological data analysis approach for detecting stress in long multivariate physiological time series data. This method can handle the high computational expenses of the traditional TDA approach. The third contribution is a novel complex networks method to detect stress. This method proposed a suitable scheme to convert long multivariate physiological time series data to graphs and extract the suitable complex network features for using as inputs of the deep learning model. Both the second and third contributions were validated on stress benchmark datasets and compared with state-of-the-art time series classification methods. The results of both two contributions are excellent.

Given the increasing use of wearable devices and the preponderance of stress in the modern world, the results of this thesis demonstrate the feasibility and challenges of detecting stress. There will be great opportunities for future research in this important area.

Appendix A

Combined networks of other participants of WESAD dataset

This appendix shows the combined graphs of all 15 participants of the WESAD dataset. The algorithm to combine graphs is presented in section 6.2.

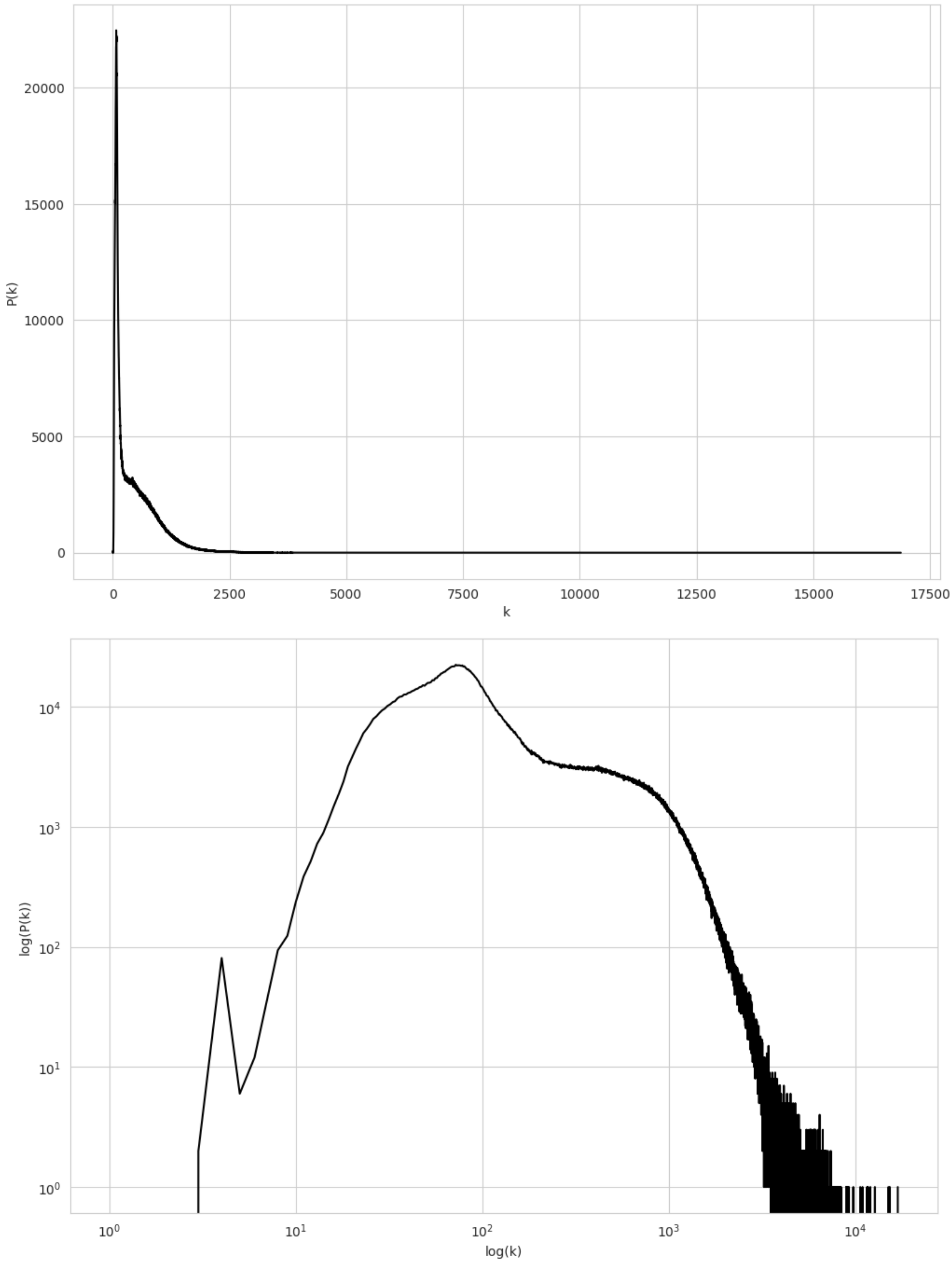


Figure A.1: Degree distribution and the log-log graph of the combined network of the participant 3

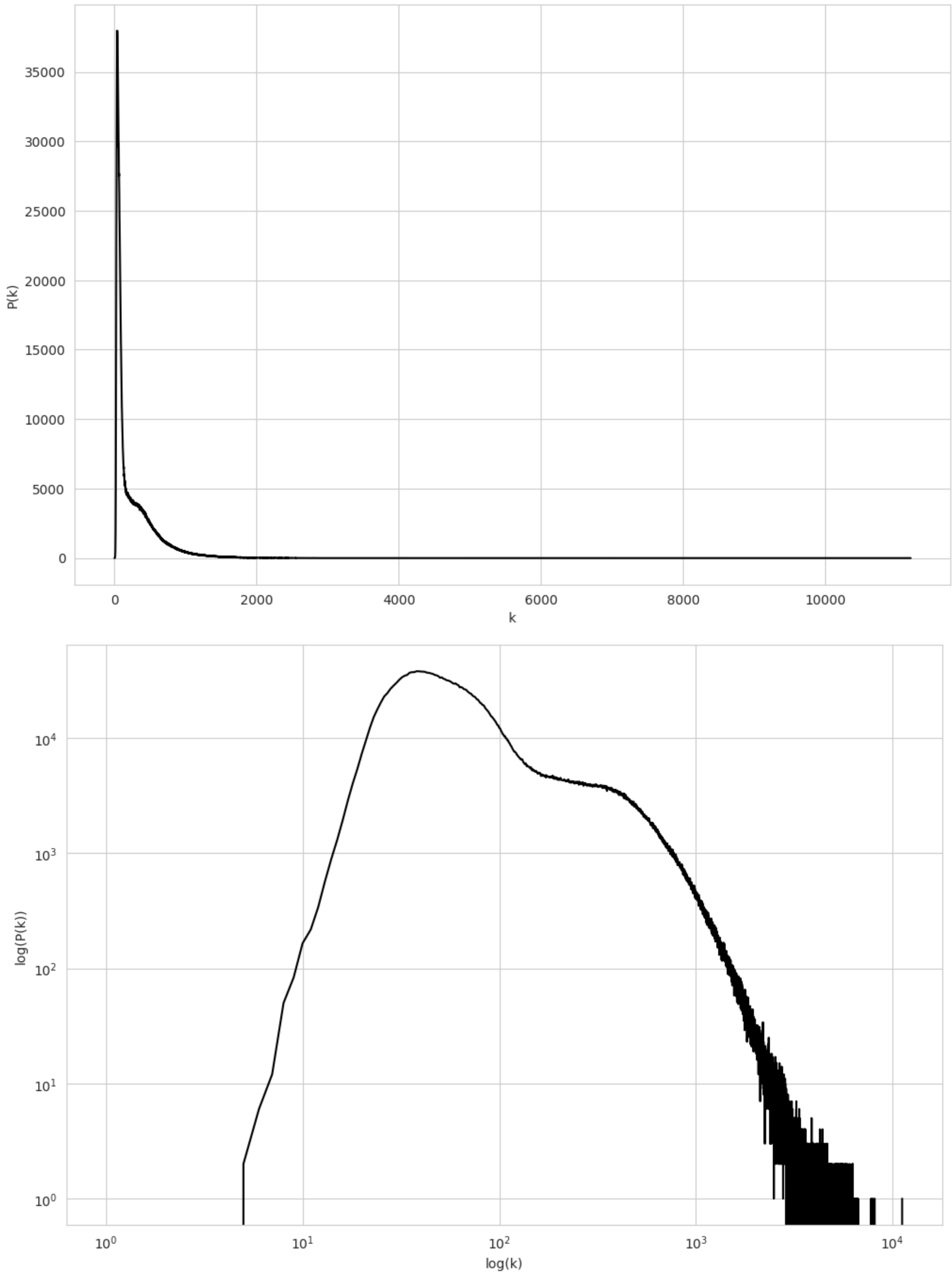


Figure A.2: Degree distribution and the log-log graph of the combined network of the participant 4

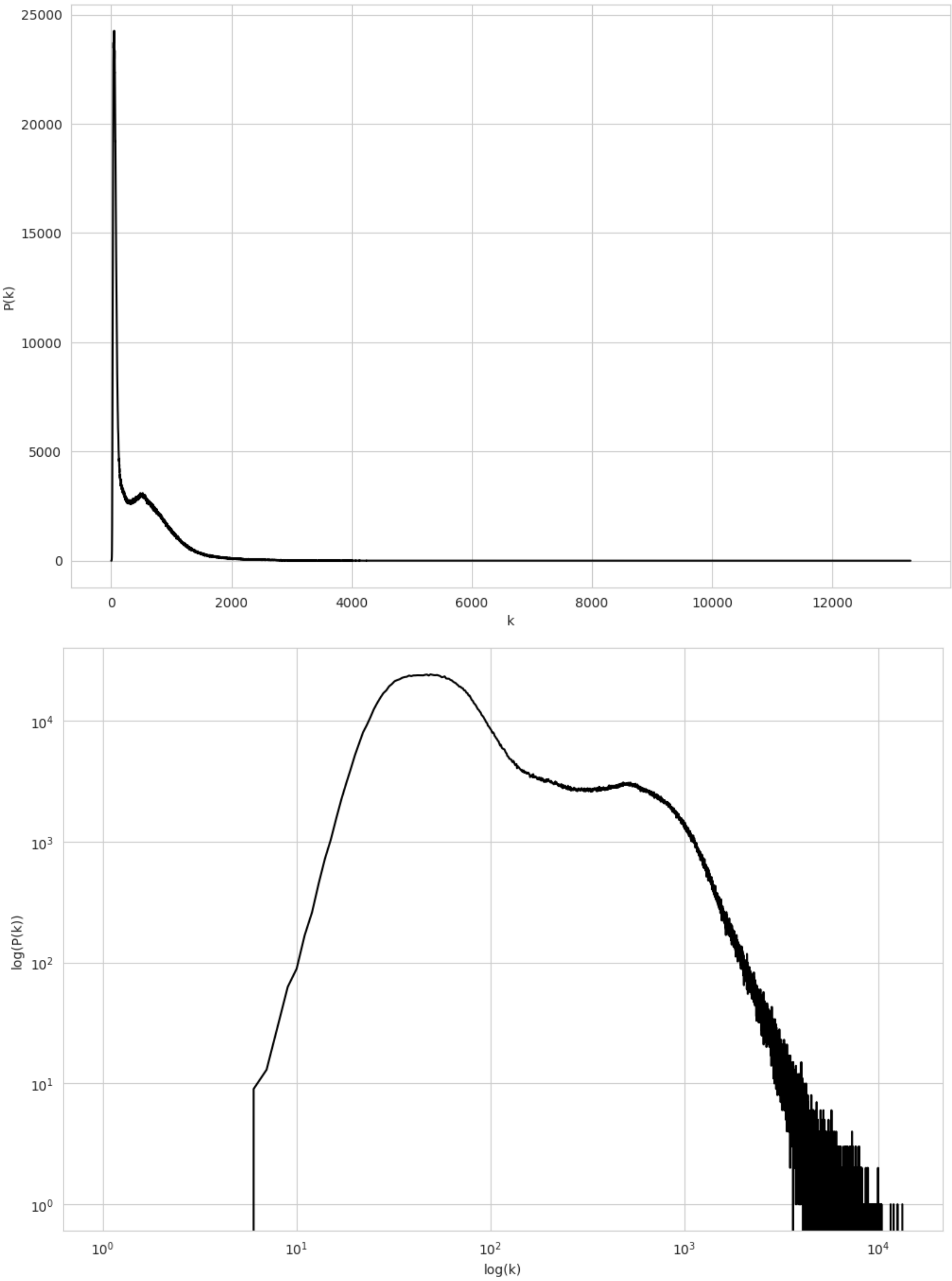


Figure A.3: Degree distribution and the log-log graph of the combined network of the participant 5

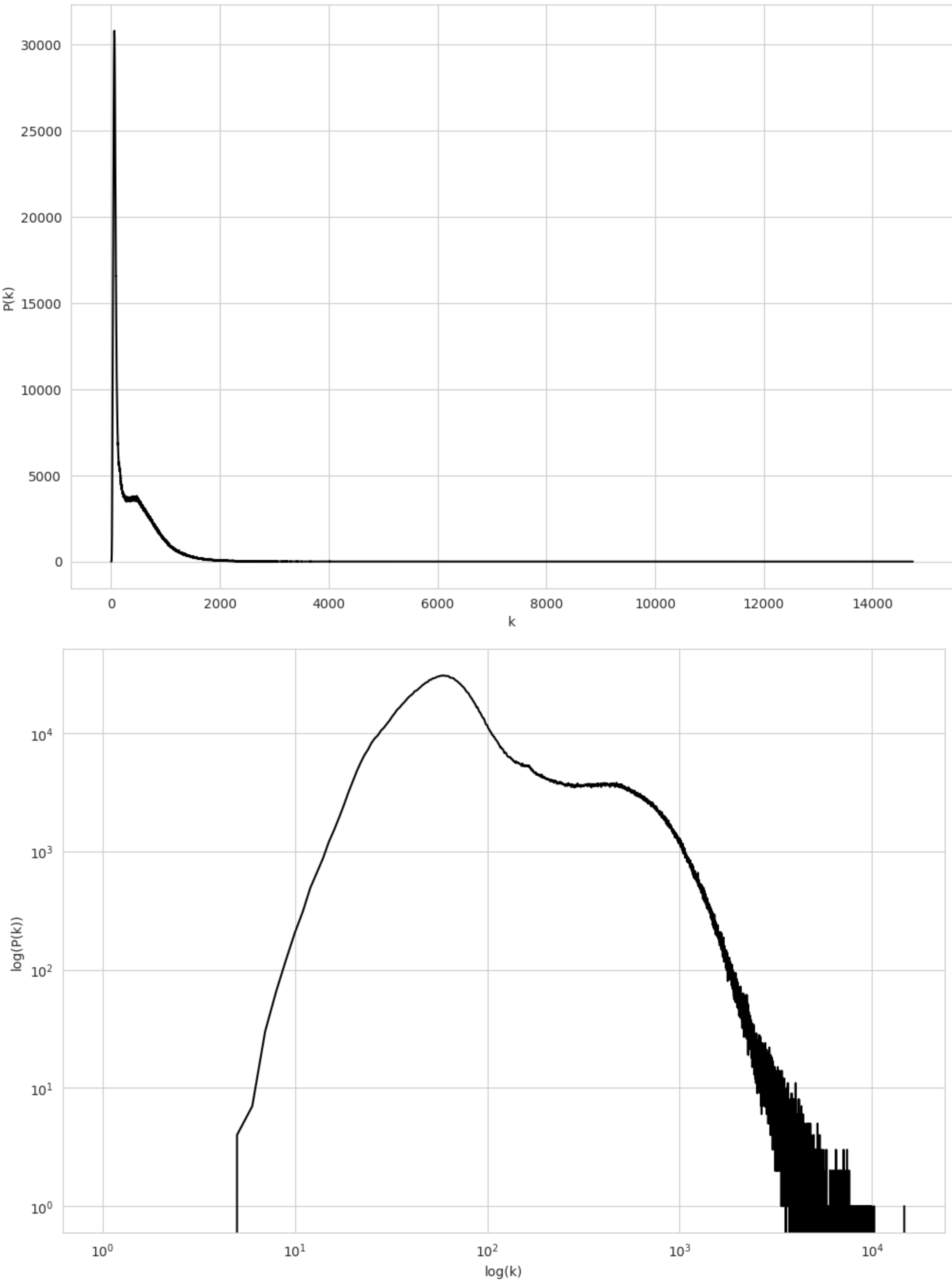


Figure A.4: Degree distribution and the log-log graph of the combined network of the participant 6

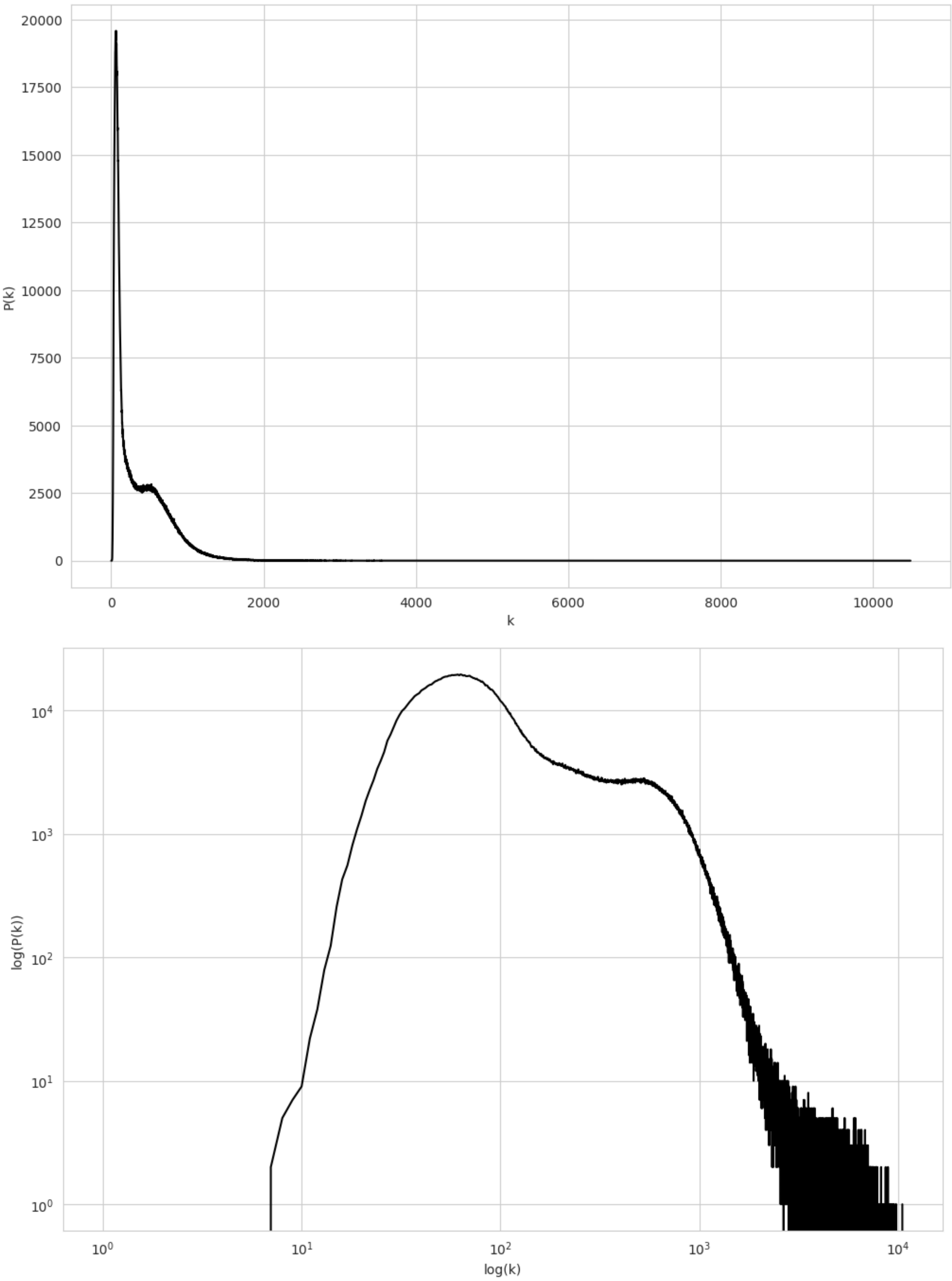
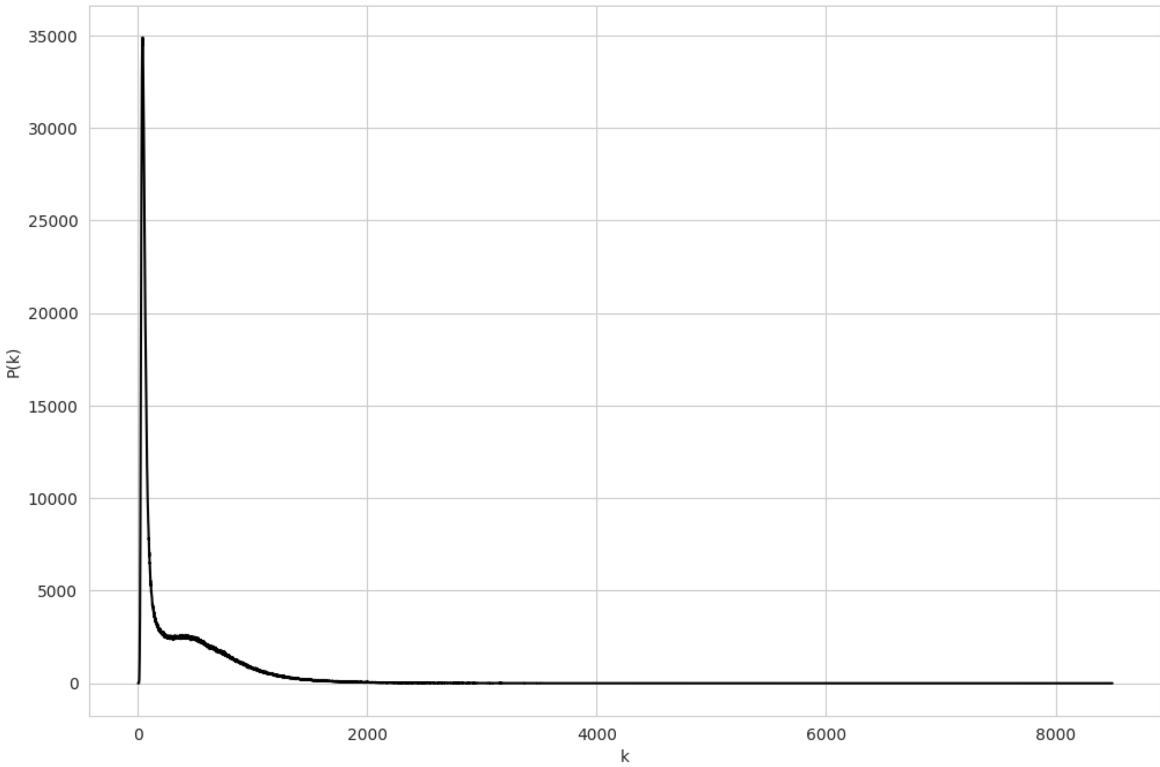


Figure A.5: Degree distribution and the log-log graph of the combined network of the participant 7



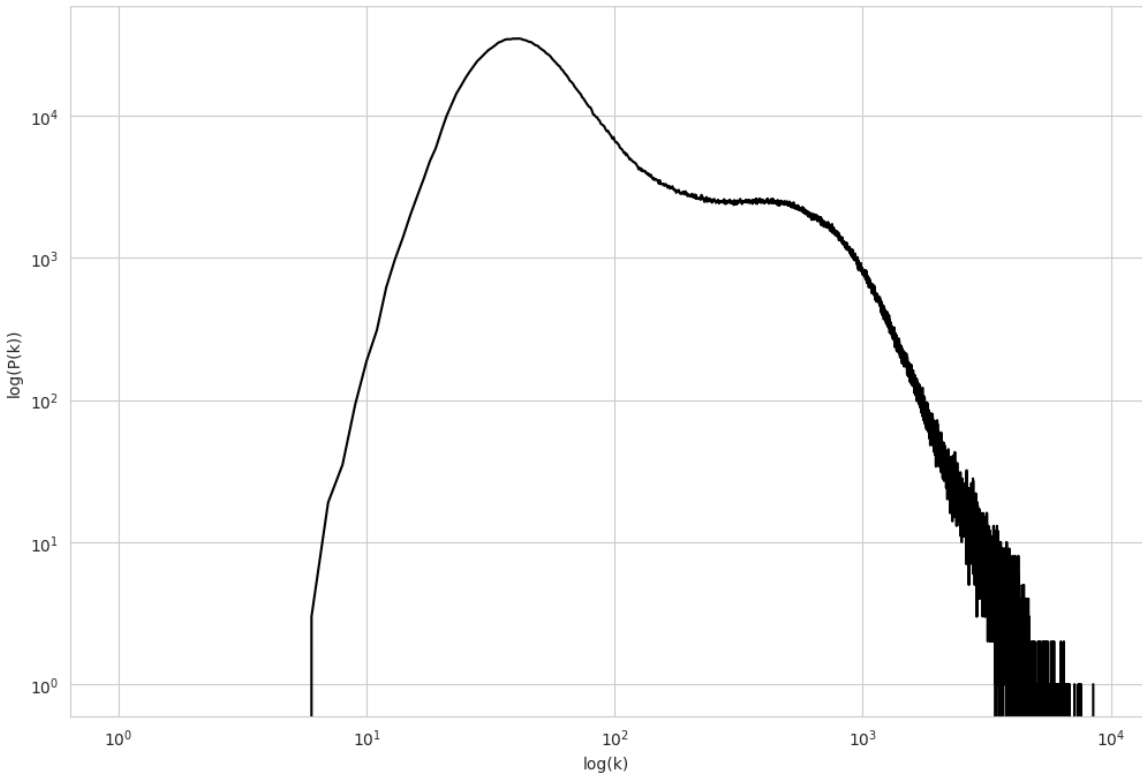


Figure A.6: Degree distribution and the log-log graph of the combined network of the participant 8

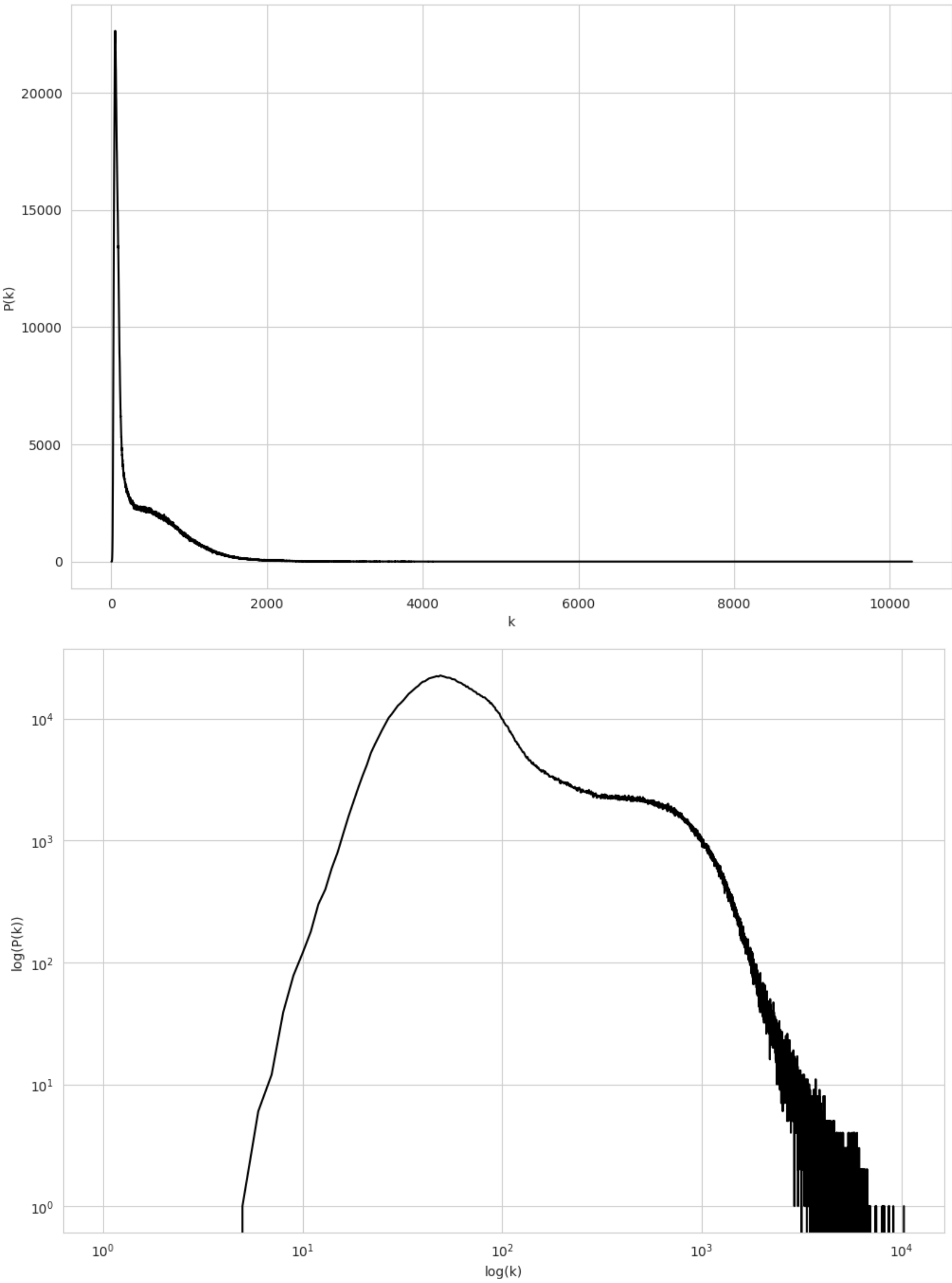


Figure A.7: Degree distribution and the log-log graph of the combined network of the participant 9

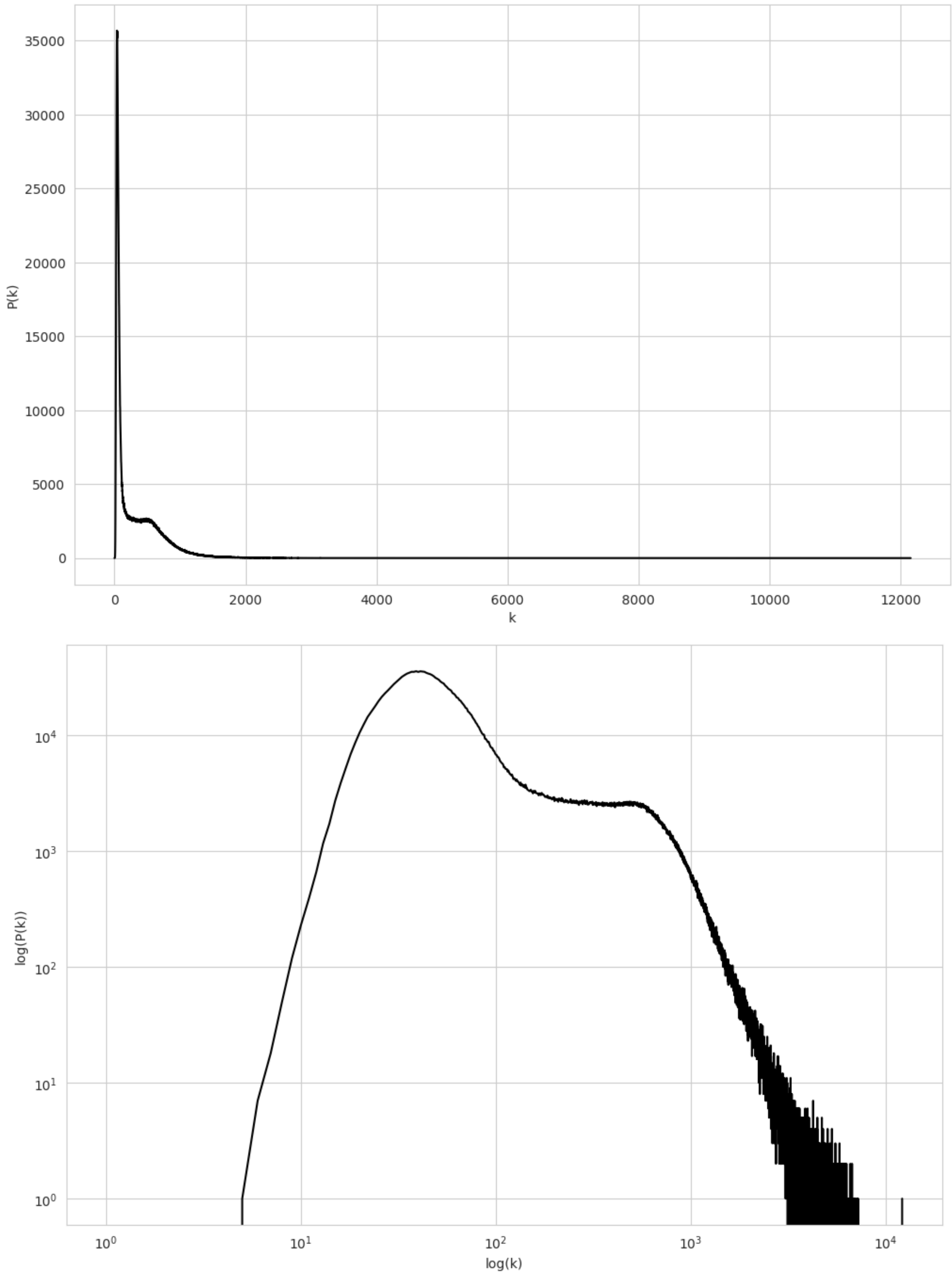
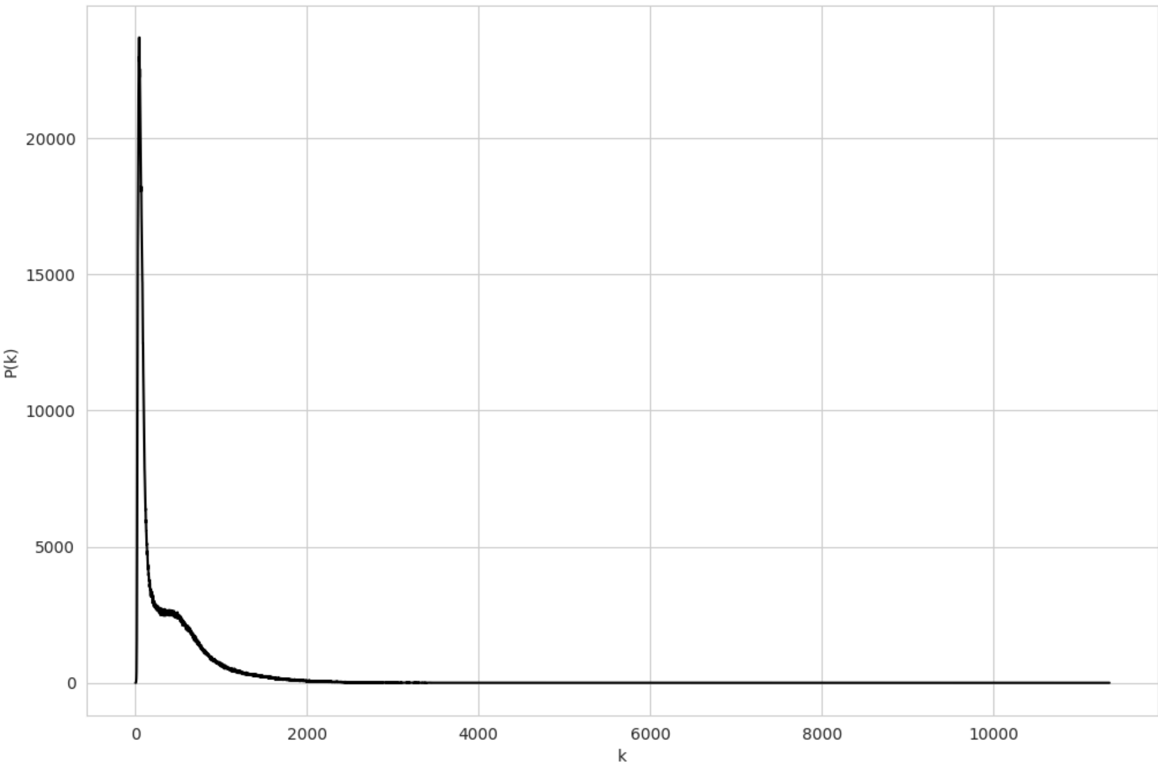


Figure A.8: Degree distribution and the log-log graph of the combined network of the participant 10



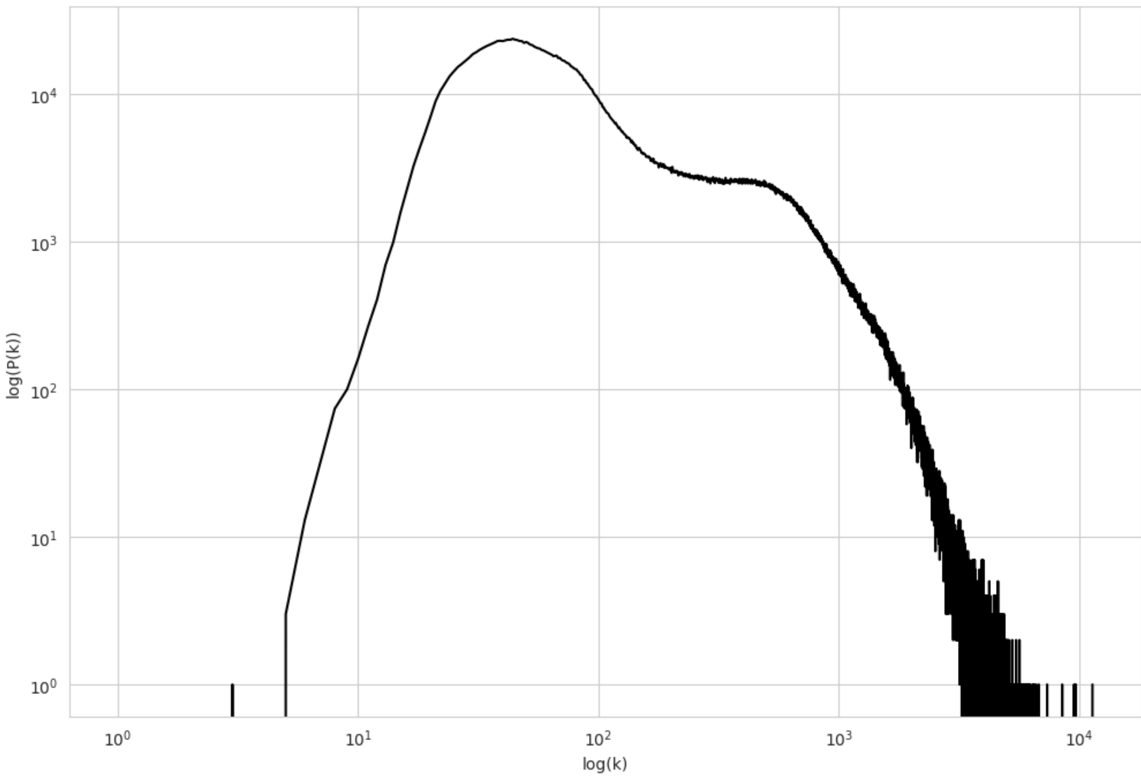


Figure A.9: Degree distribution and the log-log graph of the combined network of the participant 11

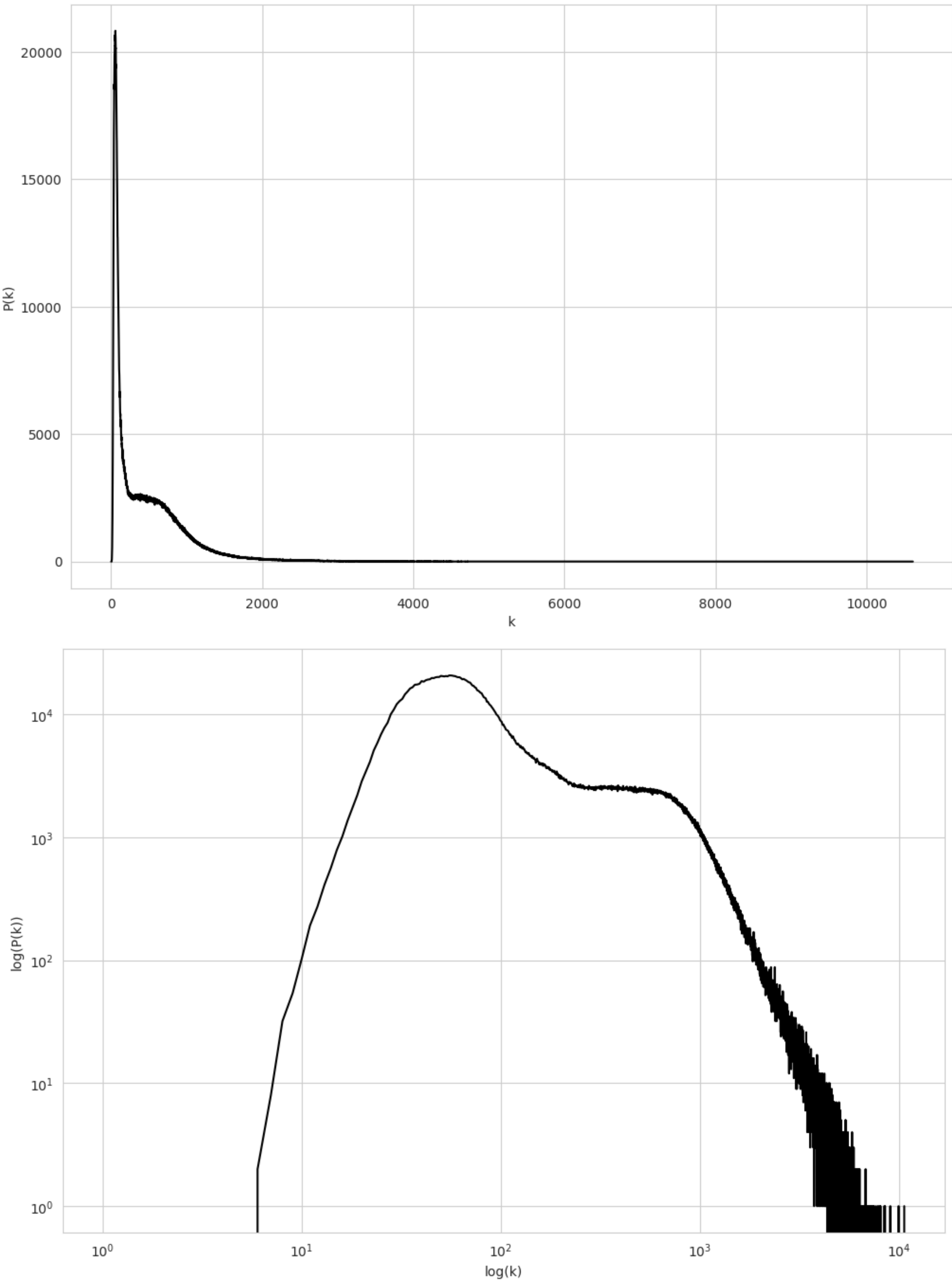


Figure A.10: Degree distribution and the log-log graph of the combined network of the participant 13

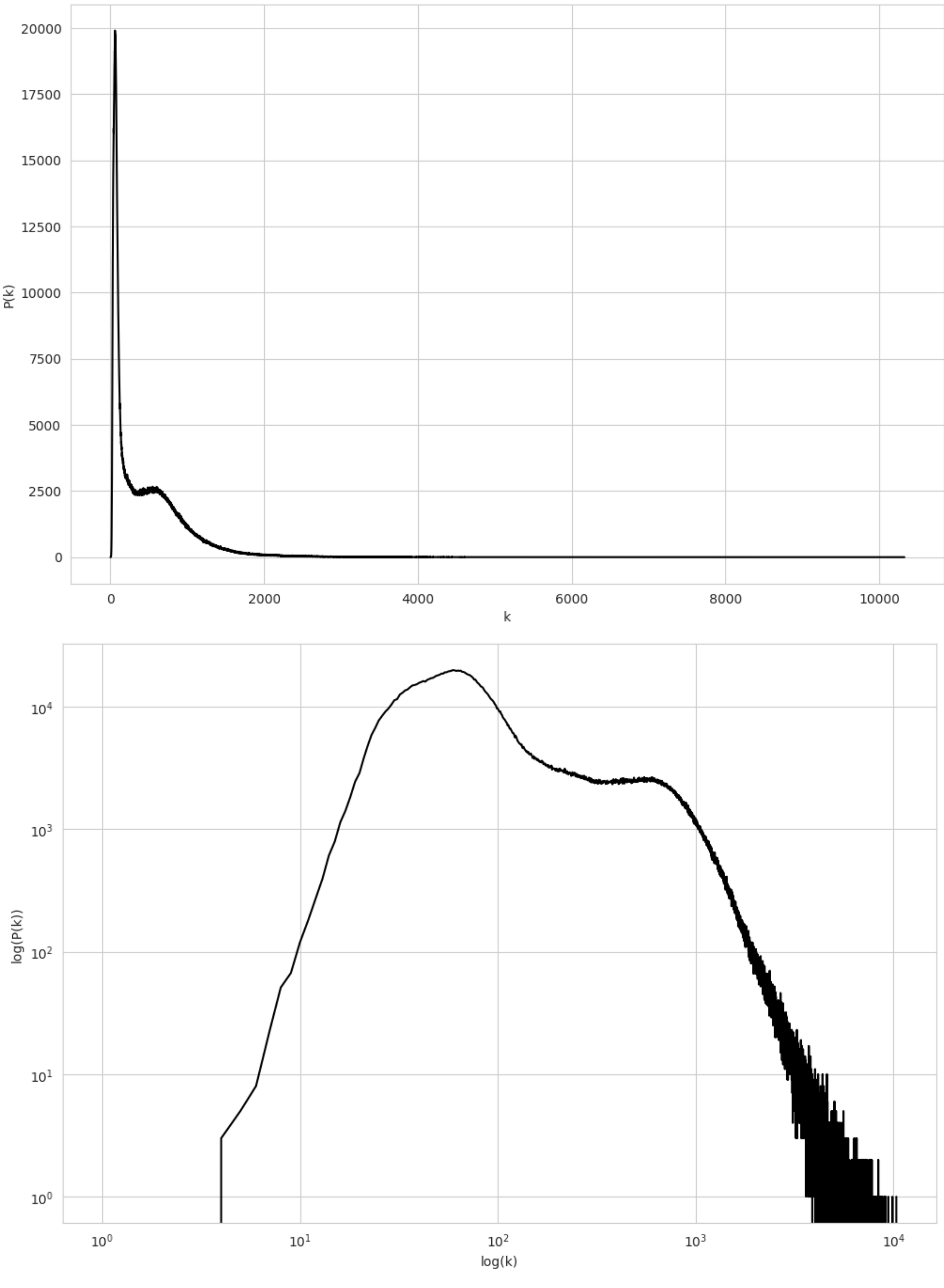


Figure A.11: Degree distribution and the log-log graph of the combined network of the participant 14

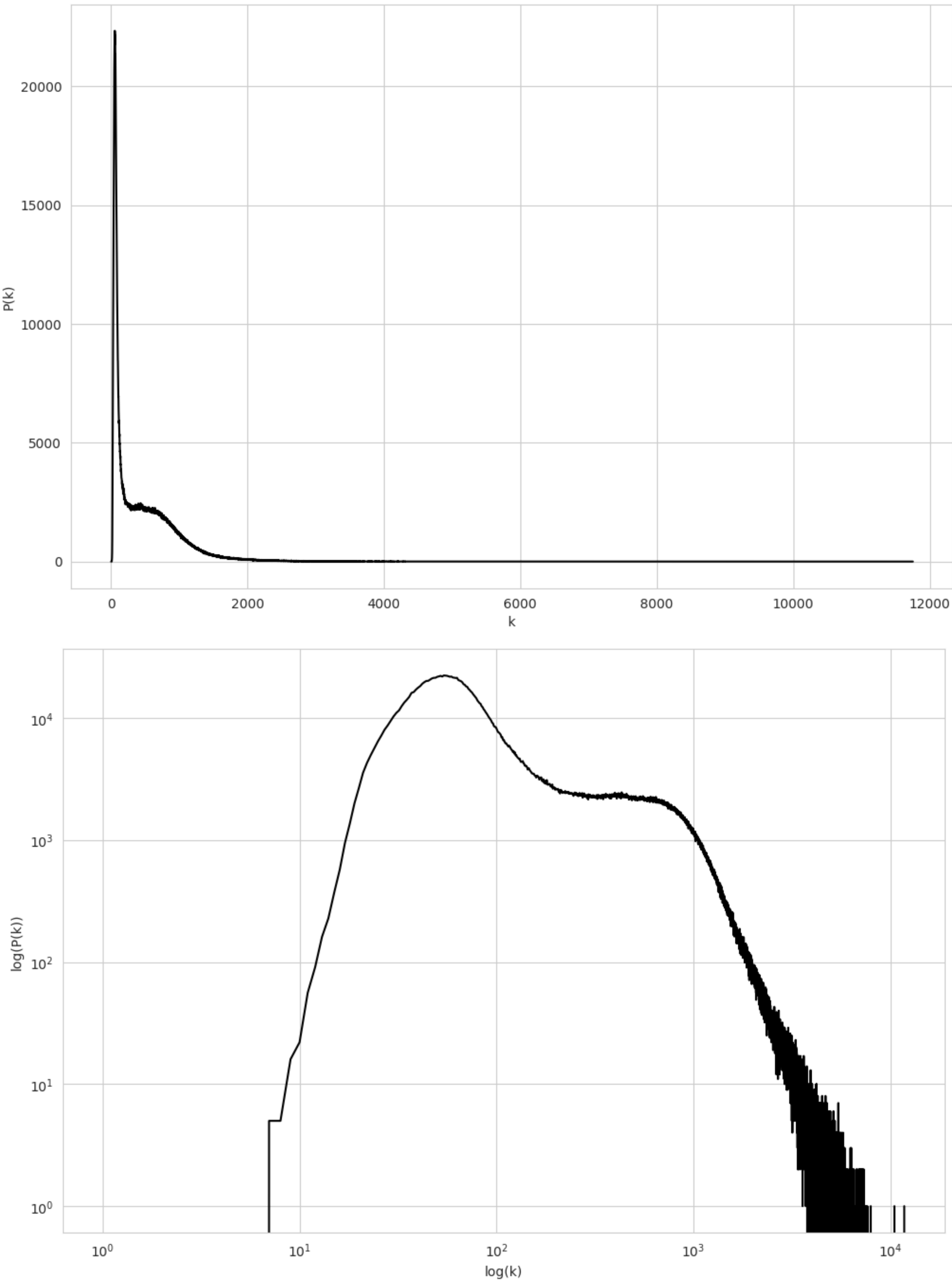


Figure A.12: Degree distribution and the log-log graph of the combined network of the participant 15

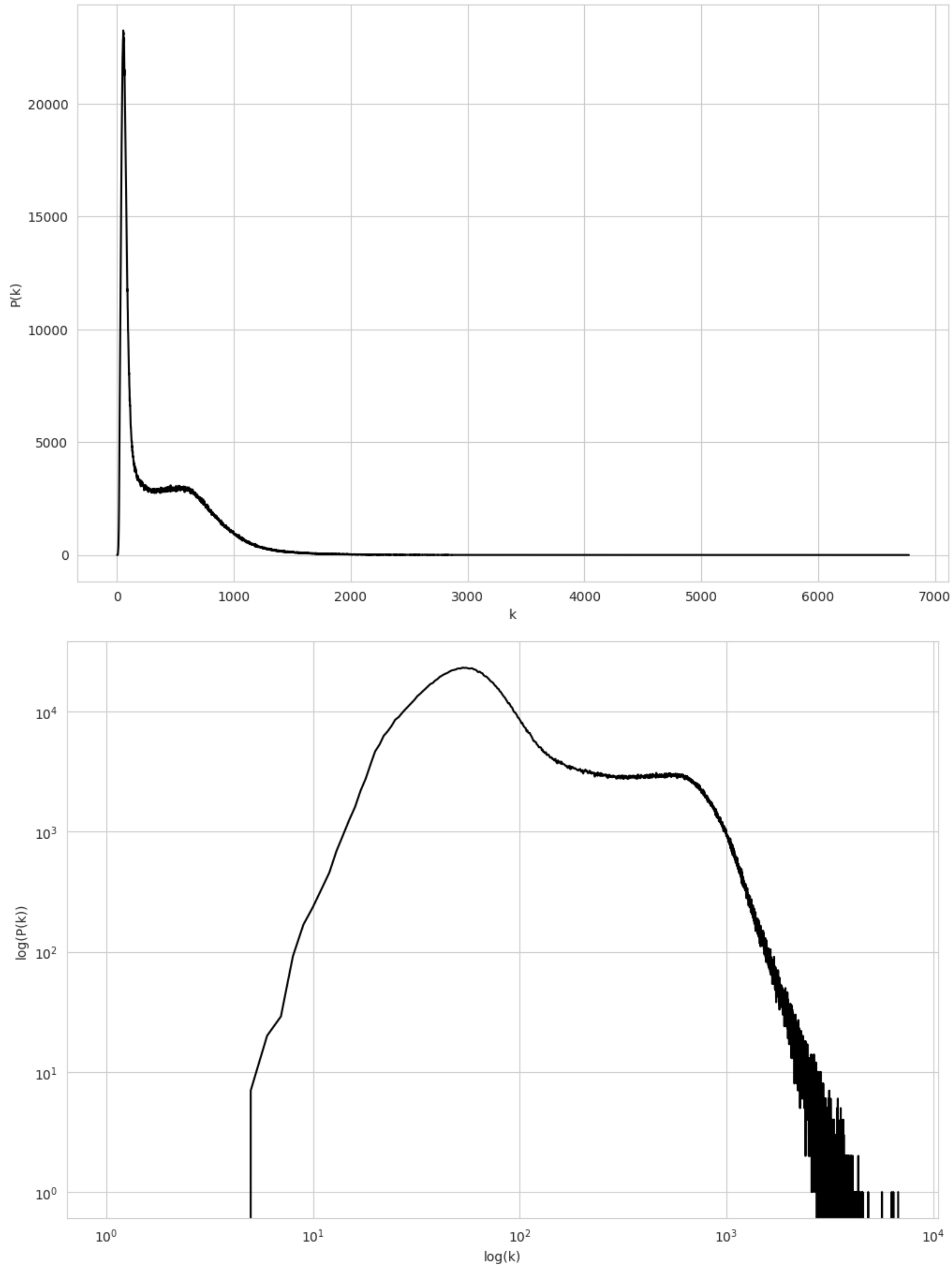


Figure A.13: Degree distribution and the log-log graph of the combined network of the participant 16

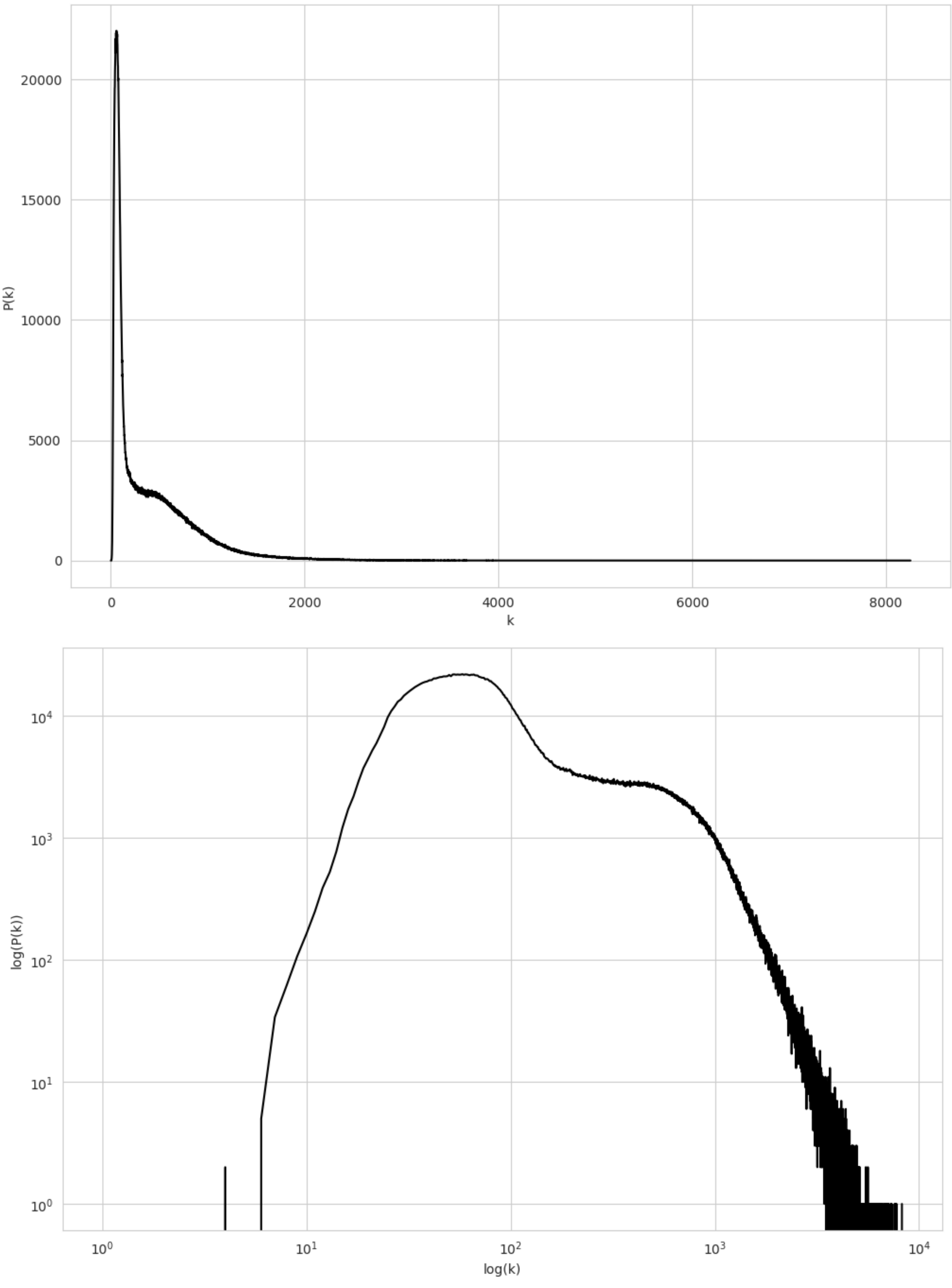


Figure A.14: Degree distribution and the log-log graph of the combined network of the participant 17

Appendix B

WESAD physiological time series data and corresponding log-log graphs

This appendix shows the time series, degree distribution and log-log graphs of all 15 participants of the WESAD dataset. The algorithm to combine graphs is presented in section 6.2.

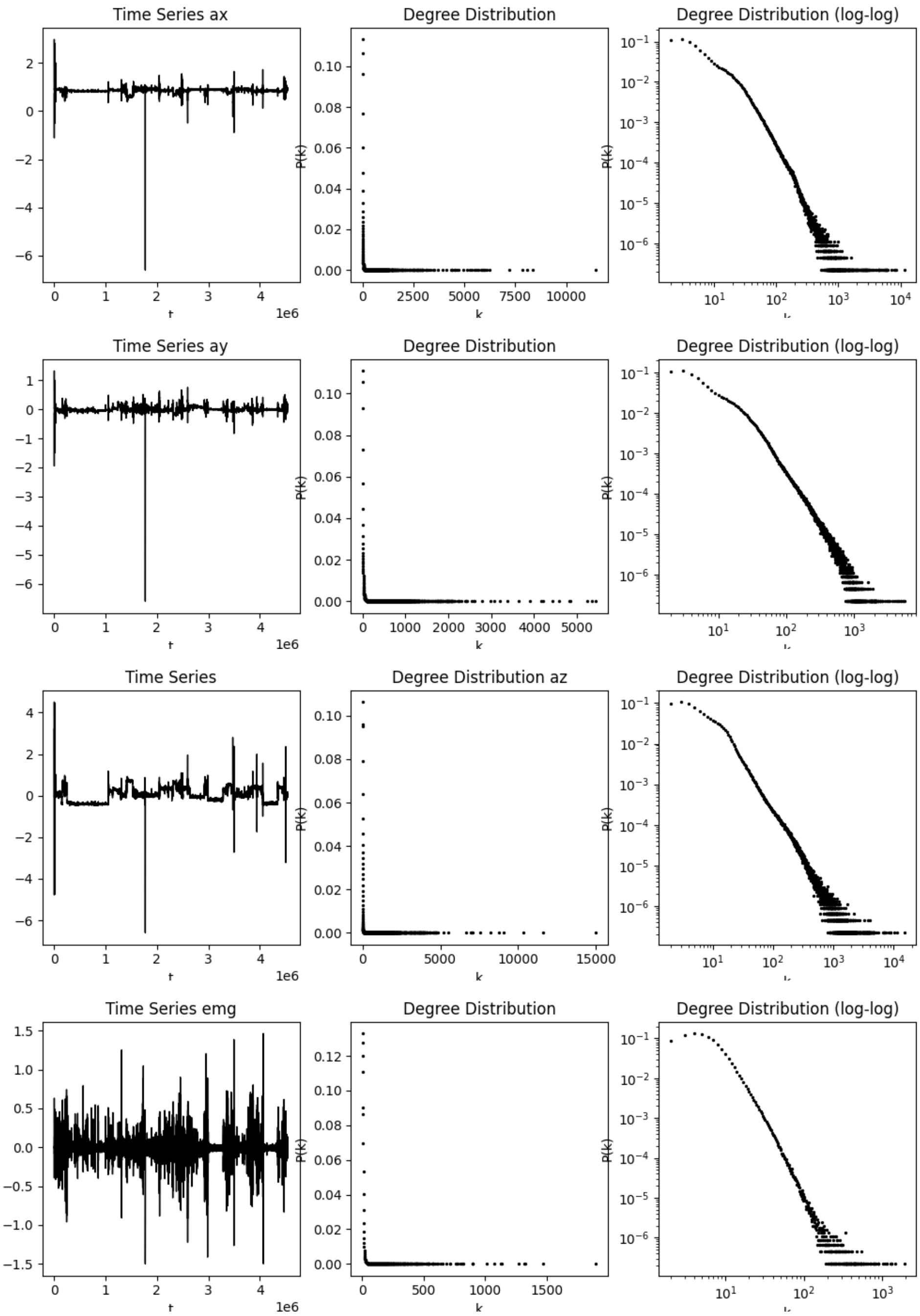


Figure B.1: Degree distribution of networks which present physiological time series data of the participant with identifier number 3 from the WESAD dataset. The data are 3-axis acceleromete data (x, y, z) and EMG. From left to right, the signal, the degree distribution, and the log-log graph

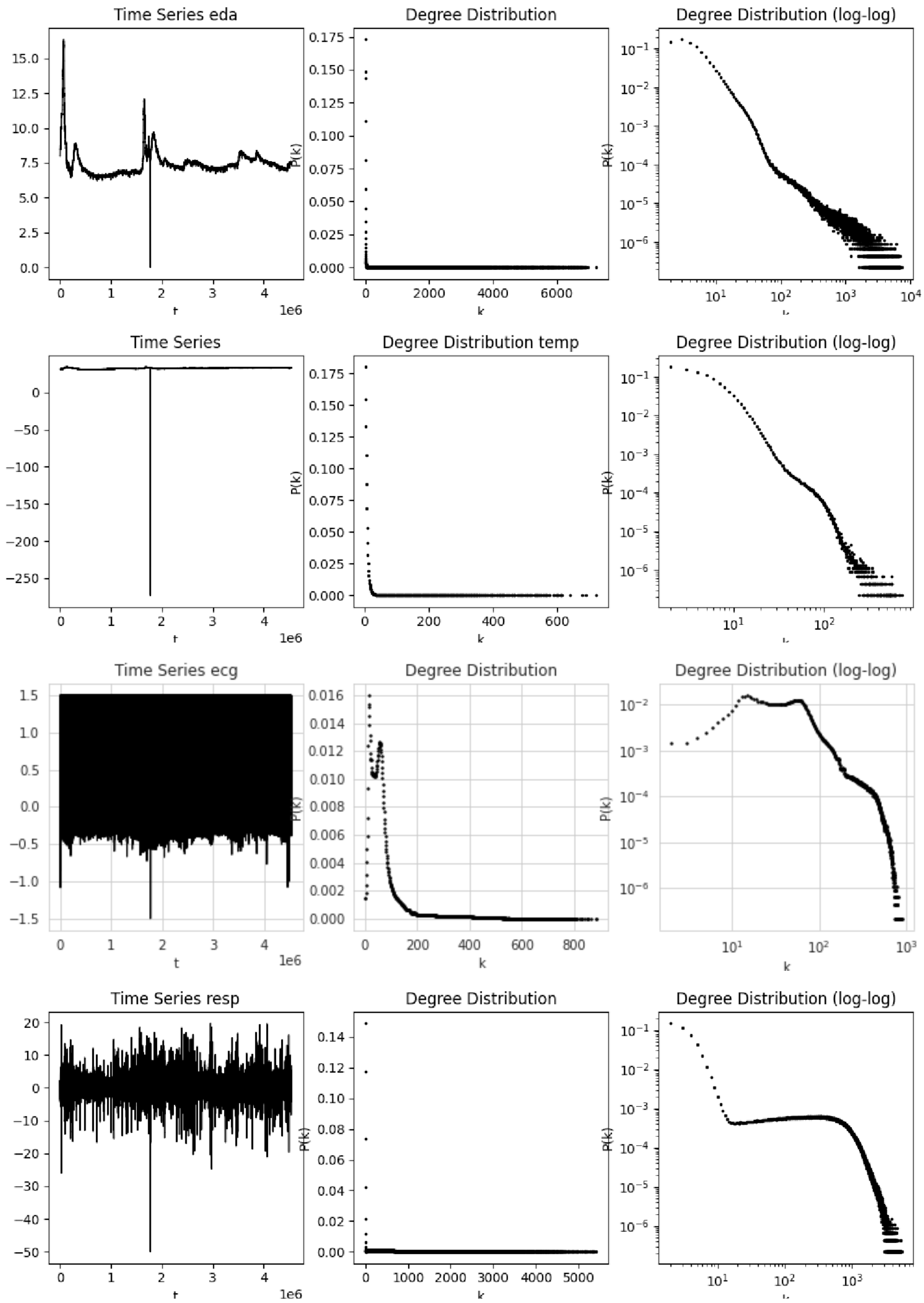


Figure B.2: Degree distribution of networks which present physiological time series data of the participant with identifier number 3 from the WESAD dataset. The data is EDA, temperature data, ECG, and respiration data. From left to right, the signal, the degree distribution, and the log-log graph

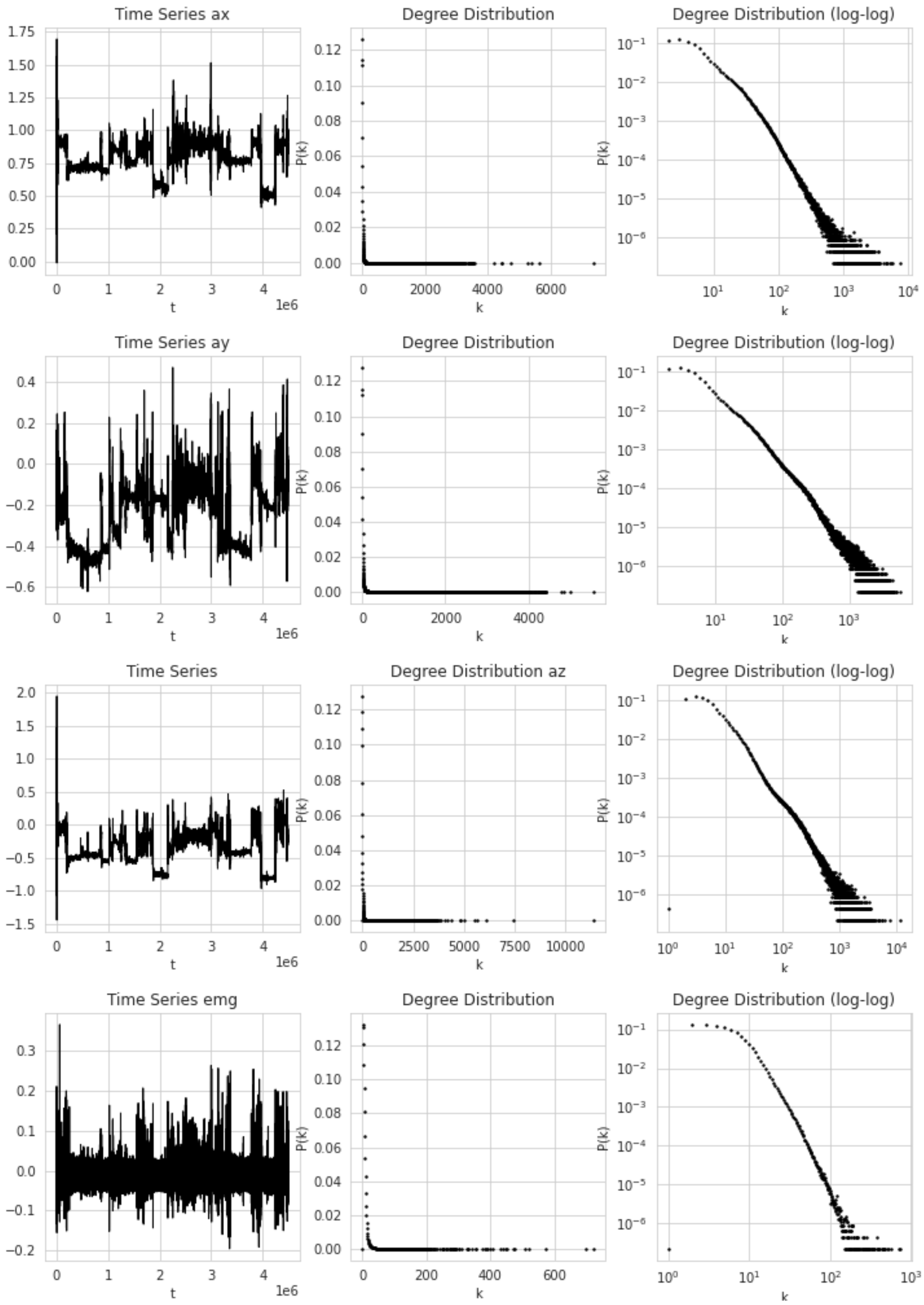


Figure B.3: Degree distribution of networks which present physiological time series data of the participant with identifier number 4 from the WESAD dataset. The data are 3-axis acceleromete data (x, y, z) and EMG. From left to right, the signal, the degree distribution, and the log-log graph

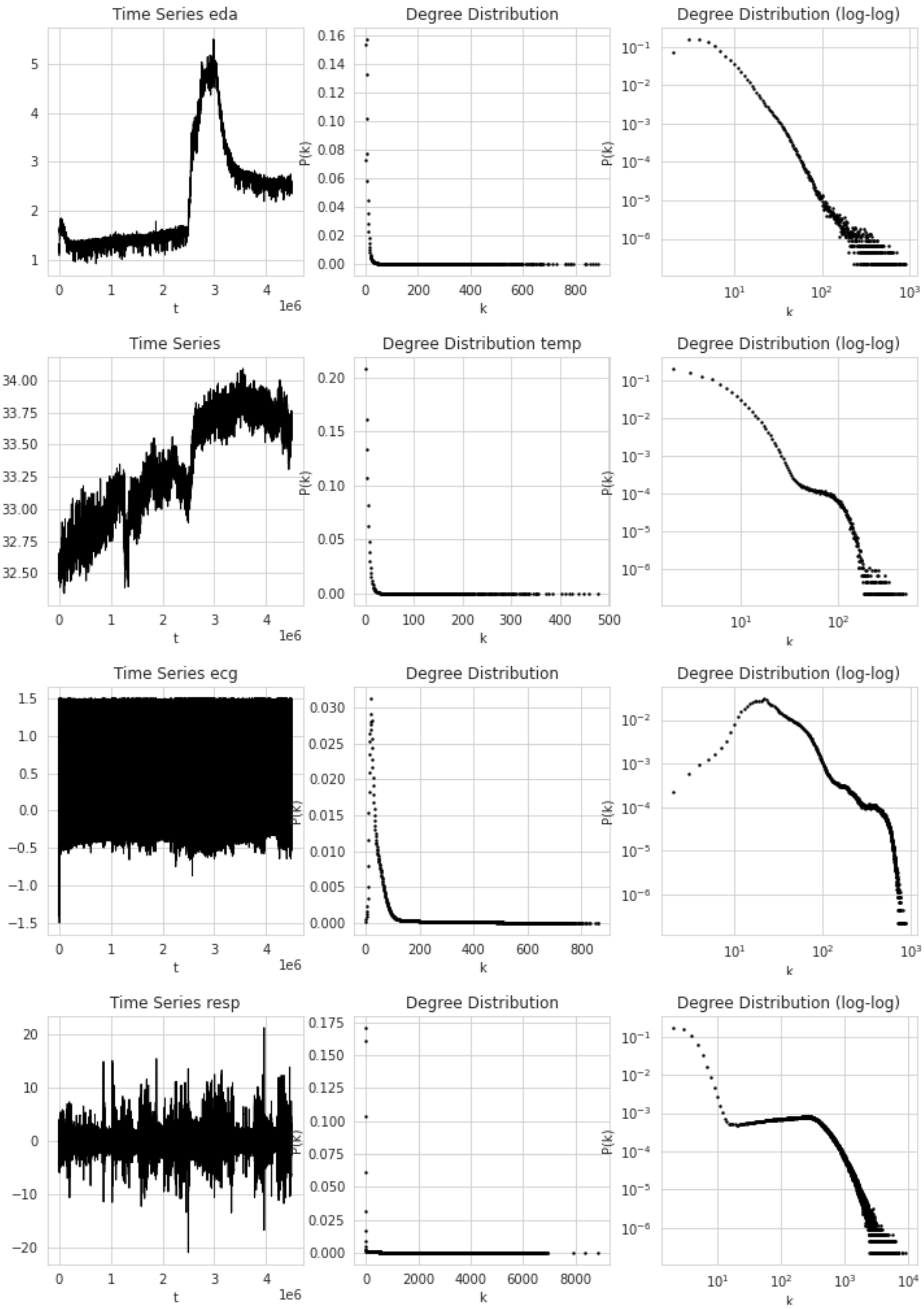


Figure B.4: Degree distribution of networks which present physiological time series data of the participant with identifier number 4 from the WESAD dataset. The data is EDA, temperature data, ECG, and respiration data. From left to right, the signal, the degree distribution, and the log-log graph

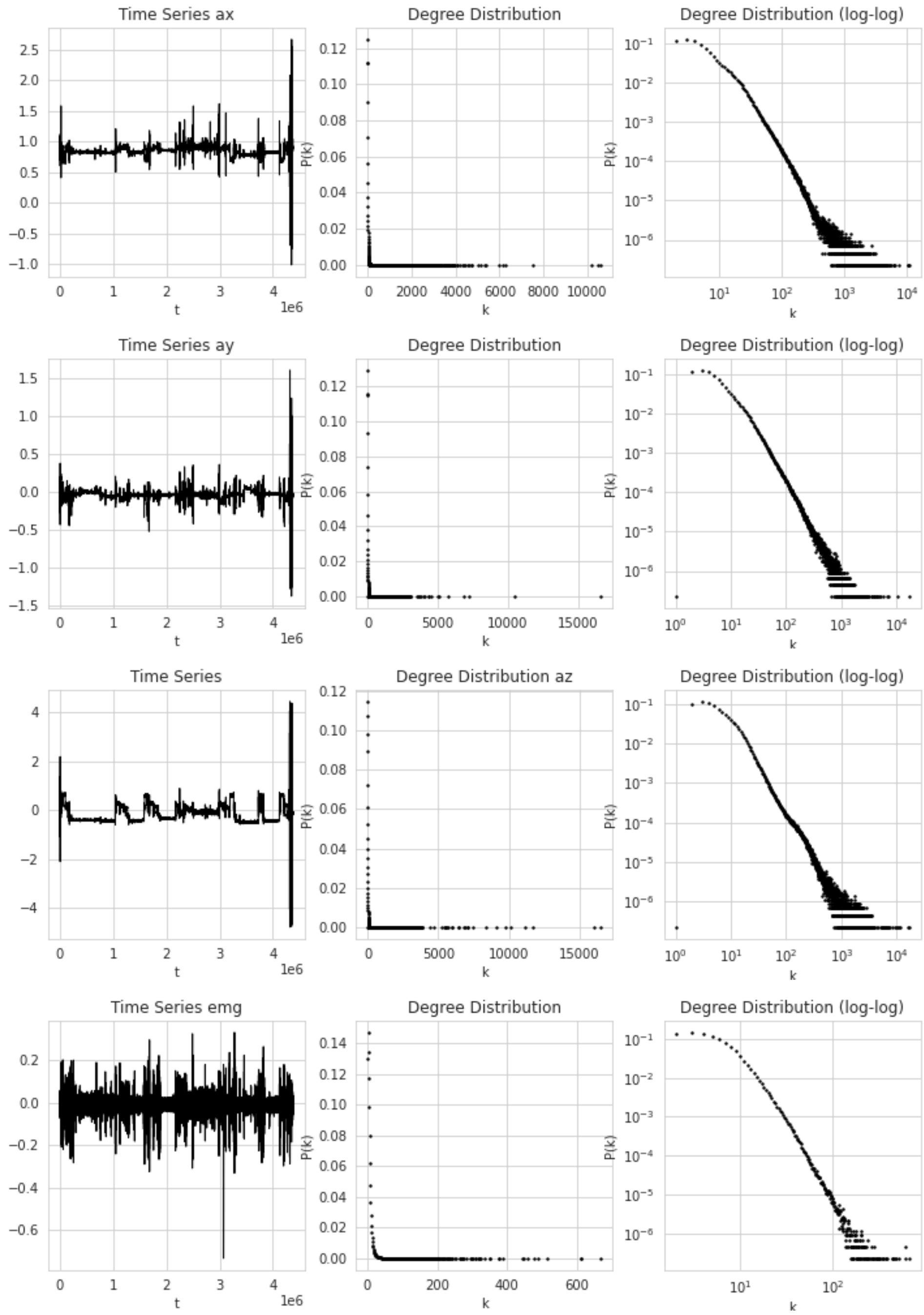


Figure B.5: Degree distribution of networks which present physiological time series data of the participant with identifier number 5 from the WESAD dataset. The data are 3-axis acceleromete data (x, y, z) and EMG. From left to right, the signal, the degree distribution, and the log-log graph

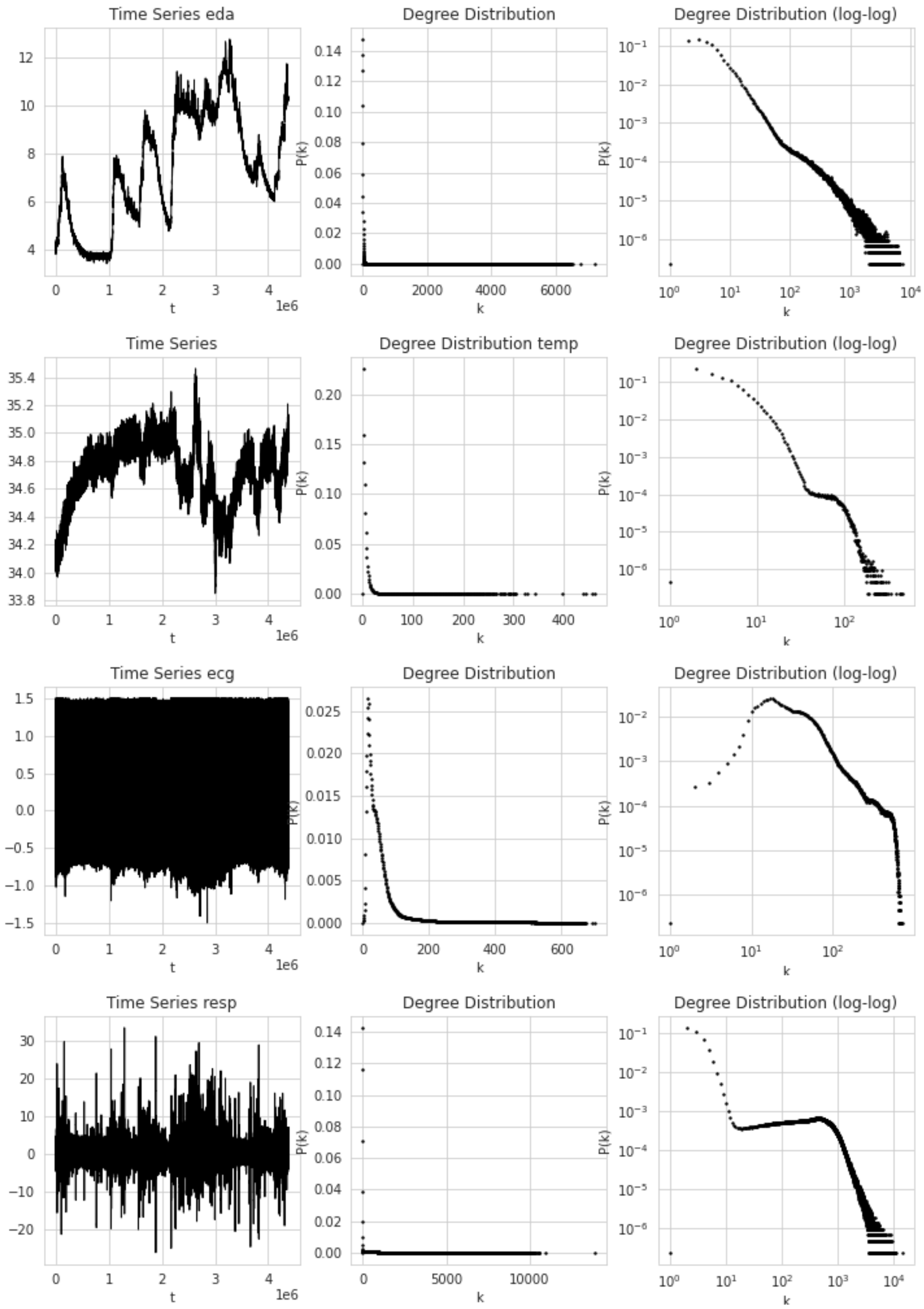


Figure B.6: Degree distribution of networks which present physiological time series data of the participant with identifier number 5 from the WESAD dataset. The data is EDA, temperature data, ECG, and respiration data. From left to right, the signal, the degree distribution, and the log-log graph

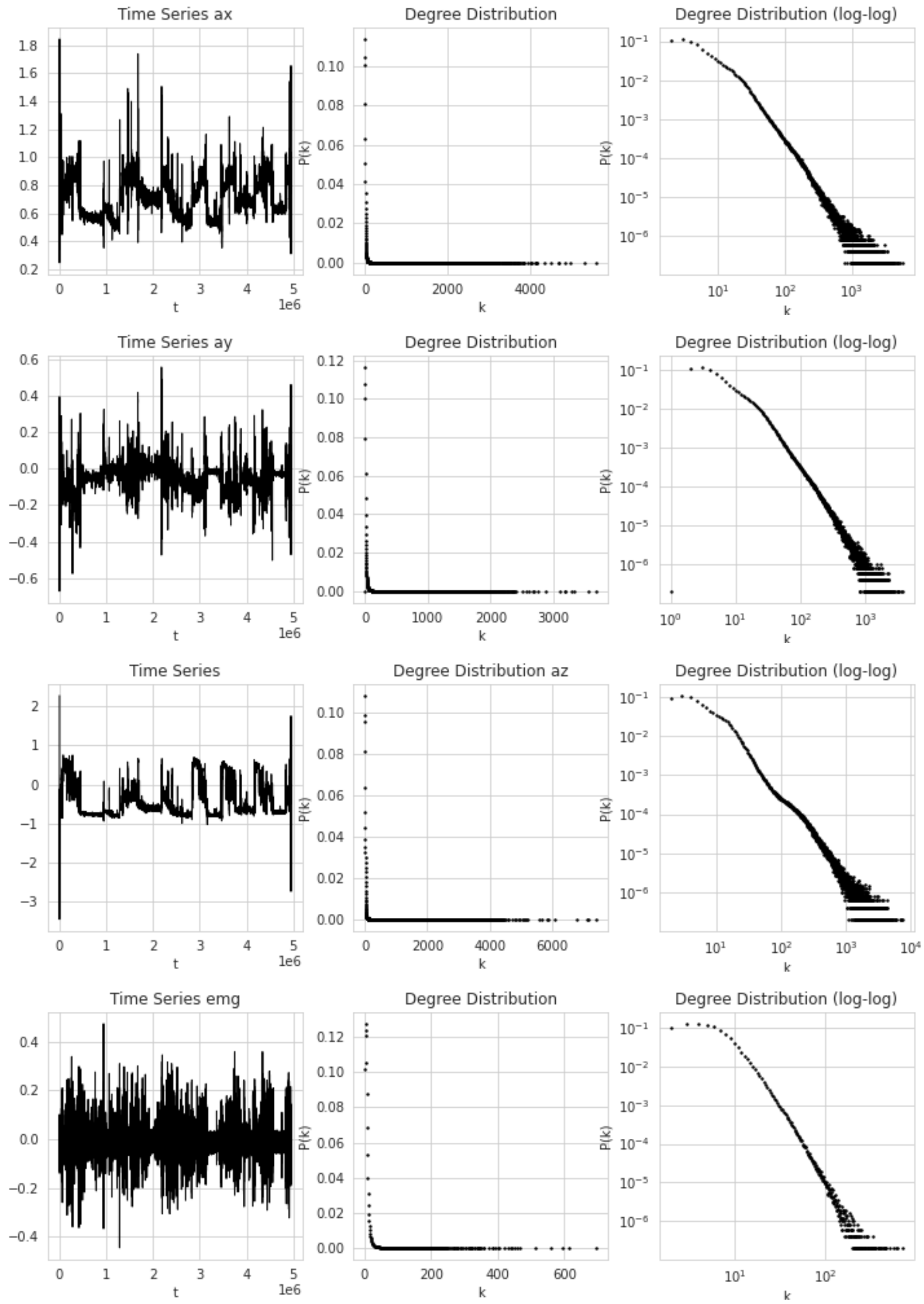


Figure B.7: Degree distribution of networks which present physiological time series data of the participant with identifier number 6 from the WESAD dataset. The data are 3-axis acceleromete data (x, y, z) and EMG. From left to right, the signal, the degree distribution, and the log-log graph

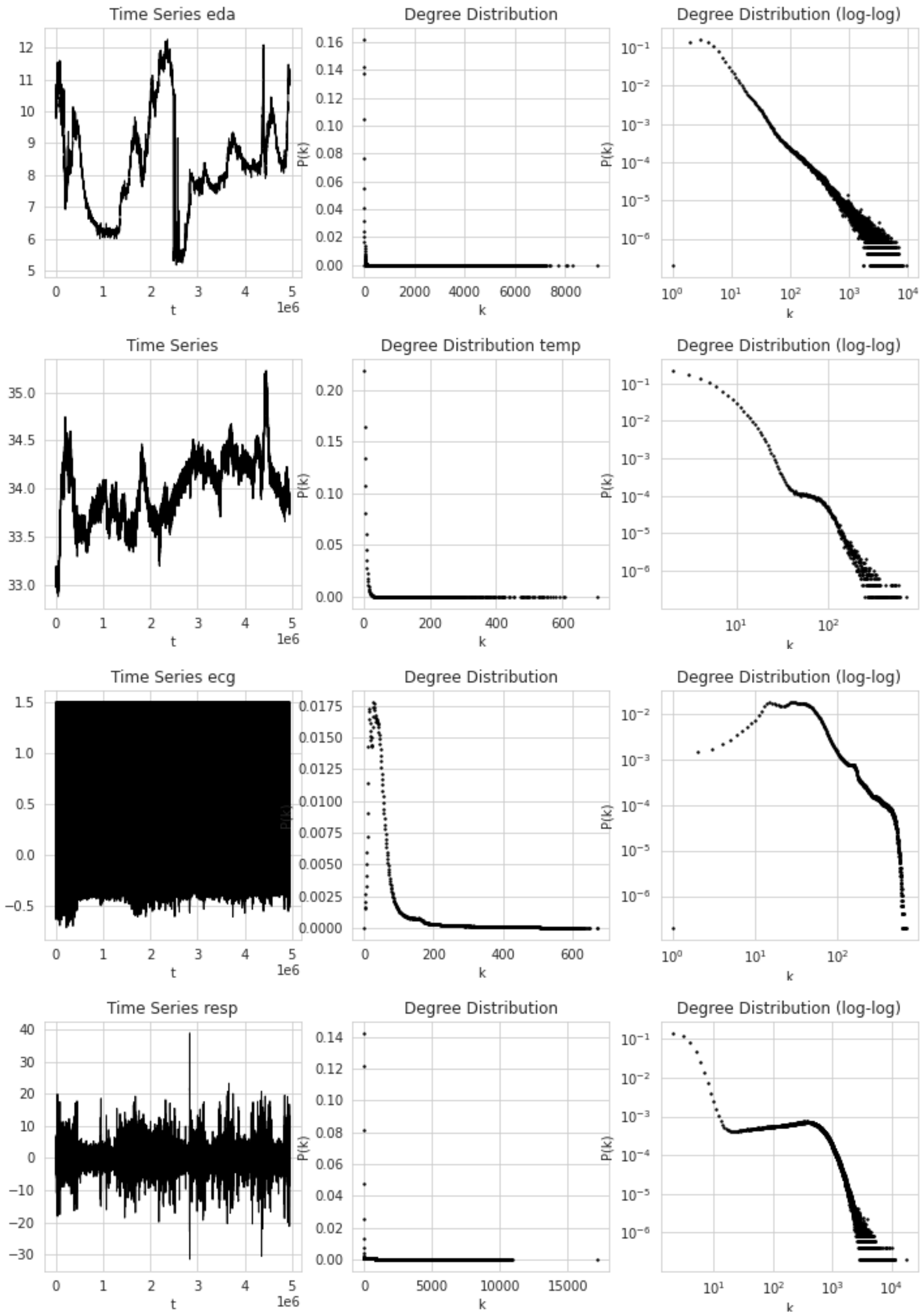


Figure B.8: Degree distribution of networks which present physiological time series data of the participant with identifier number 6 from the WESAD dataset. The data is EDA, temperature data, ECG, and respiration data. From left to right, the signal, the degree distribution, and the log-log graph

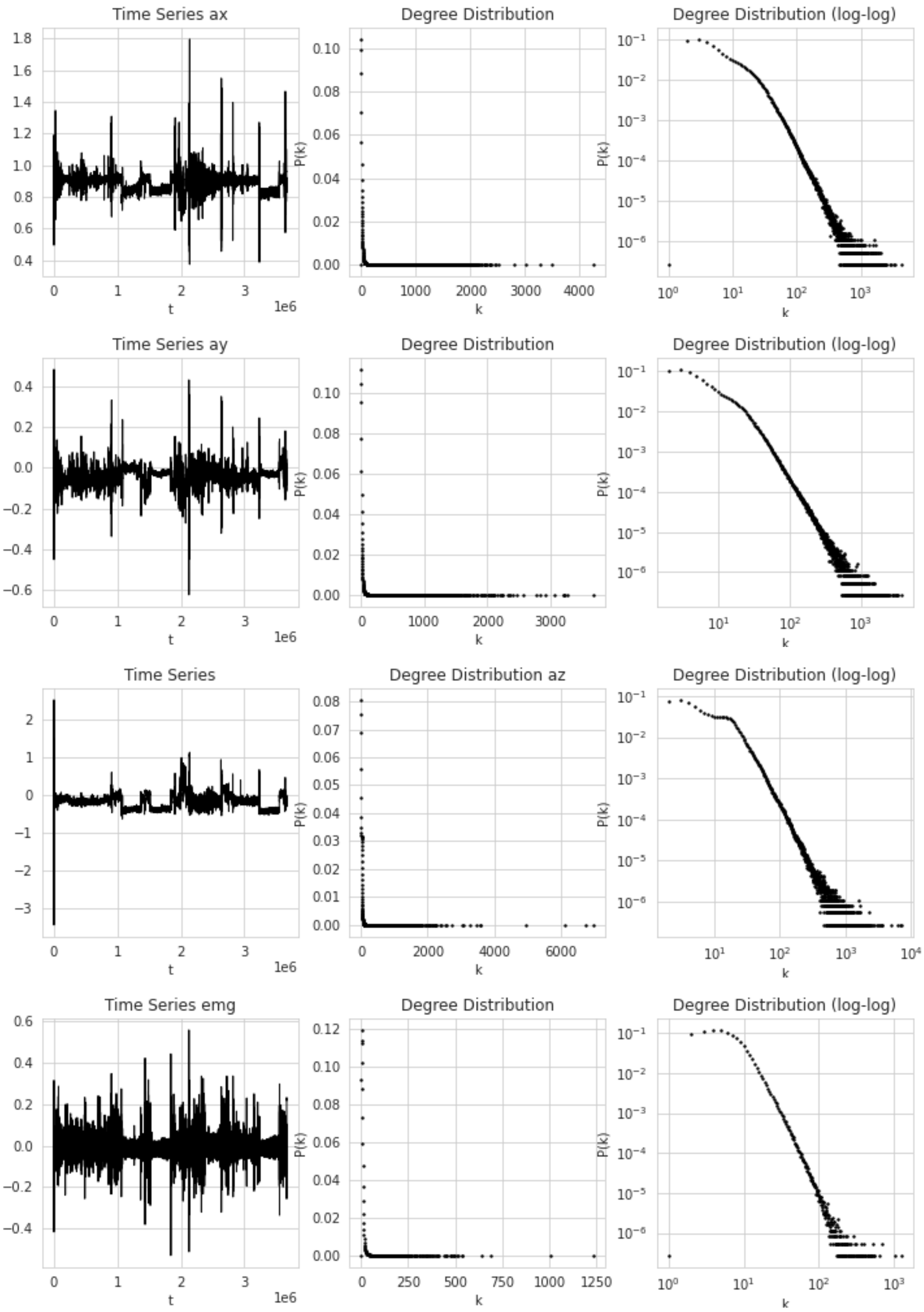


Figure B.9: Degree distribution of networks which present physiological time series data of the participant with identifier number 7 from the WESAD dataset. The data are 3-axis acceleromete data (x, y, z) and EMG. From left to right, the signal, the degree distribution, and the log-log graph

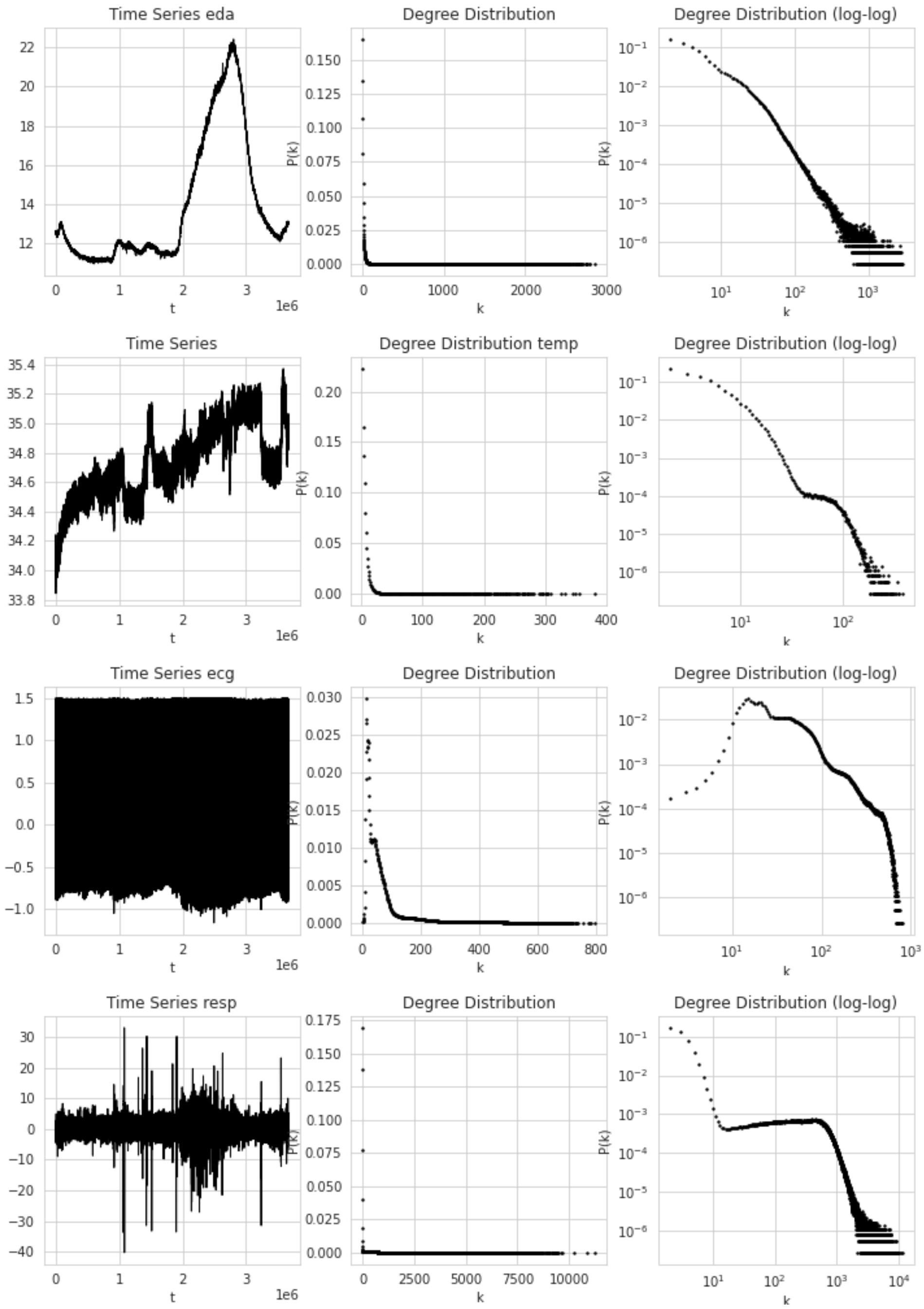


Figure B.10: Degree distribution of networks which present physiological time series data of the participant with identifier number 7 from the WESAD dataset. The data is EDA, temperature data, ECG, and respiration data. From left to right, the signal, the degree distribution, and the log-log graph

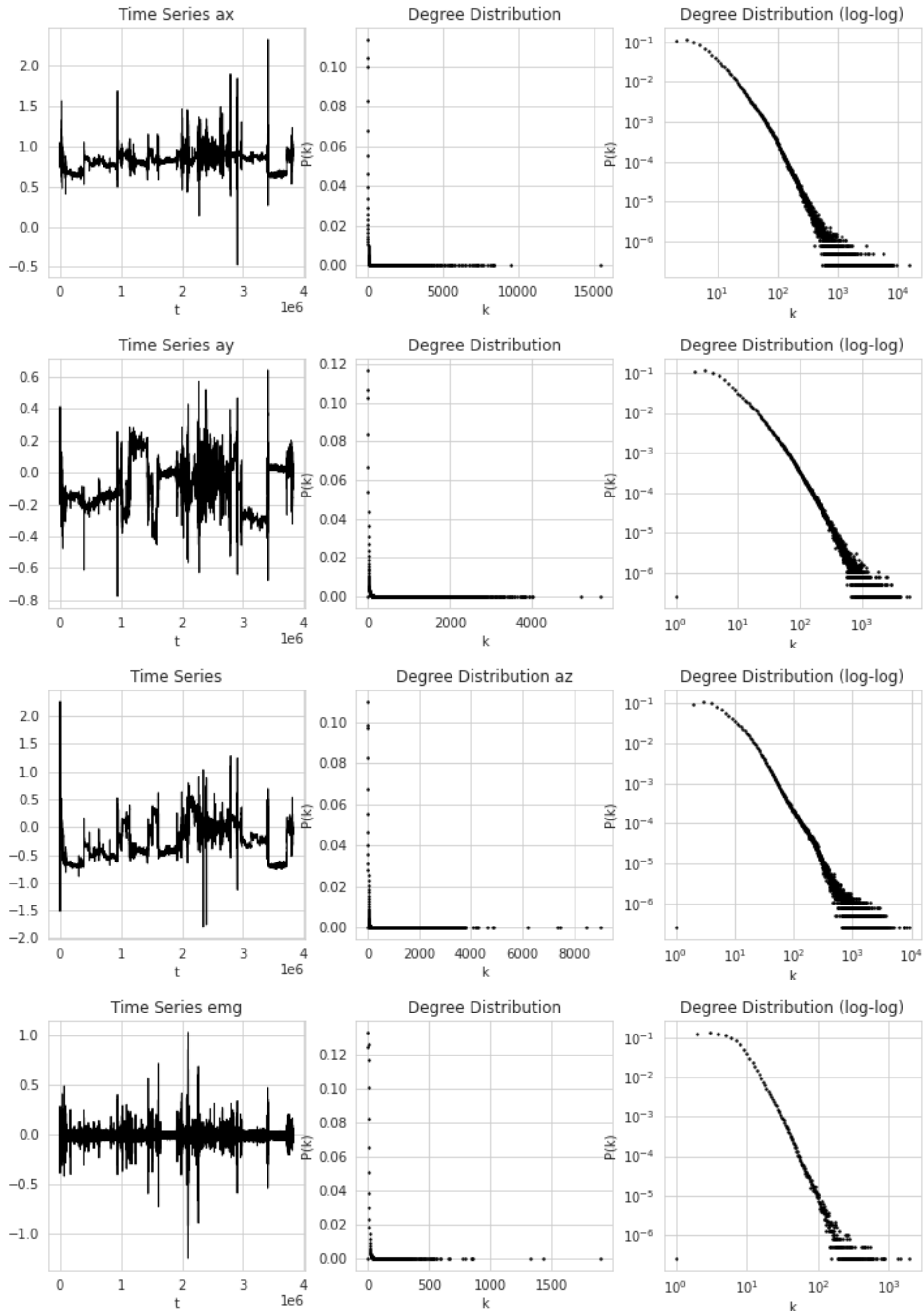


Figure B.11: Degree distribution of networks which present physiological time series data of the participant with identifier number 8 from the WESAD dataset. The data are 3-axis acceleromete data (x, y, z) and EMG. From left to right, the signal, the degree distribution, and the log-log graph

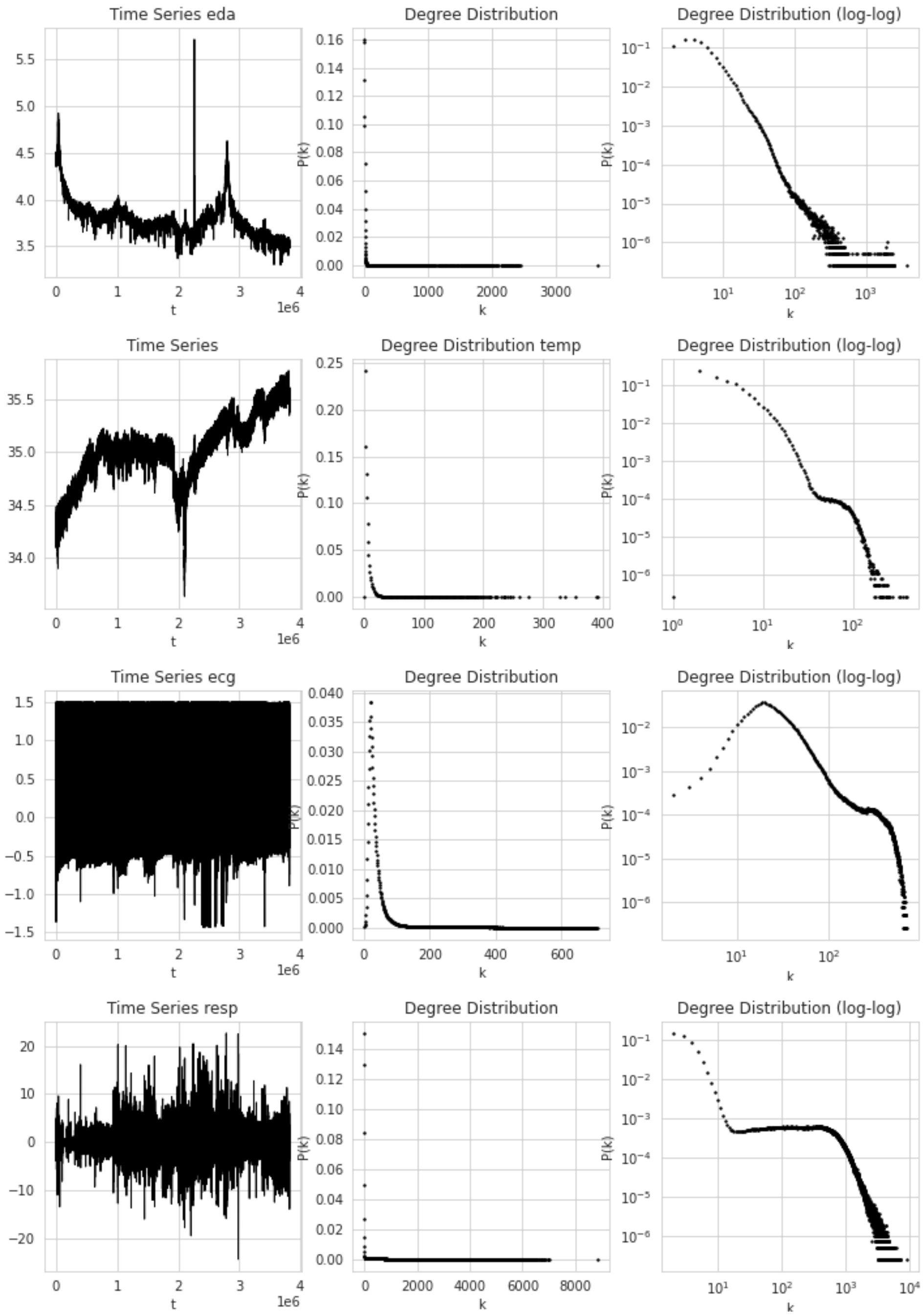


Figure B.12: Degree distribution of networks which present physiological time series data of the participant with identifier number 8 from the WESAD dataset. The data is EDA, temperature data, ECG, and respiration data. From left to right, the signal, the degree distribution, and the log-log graph

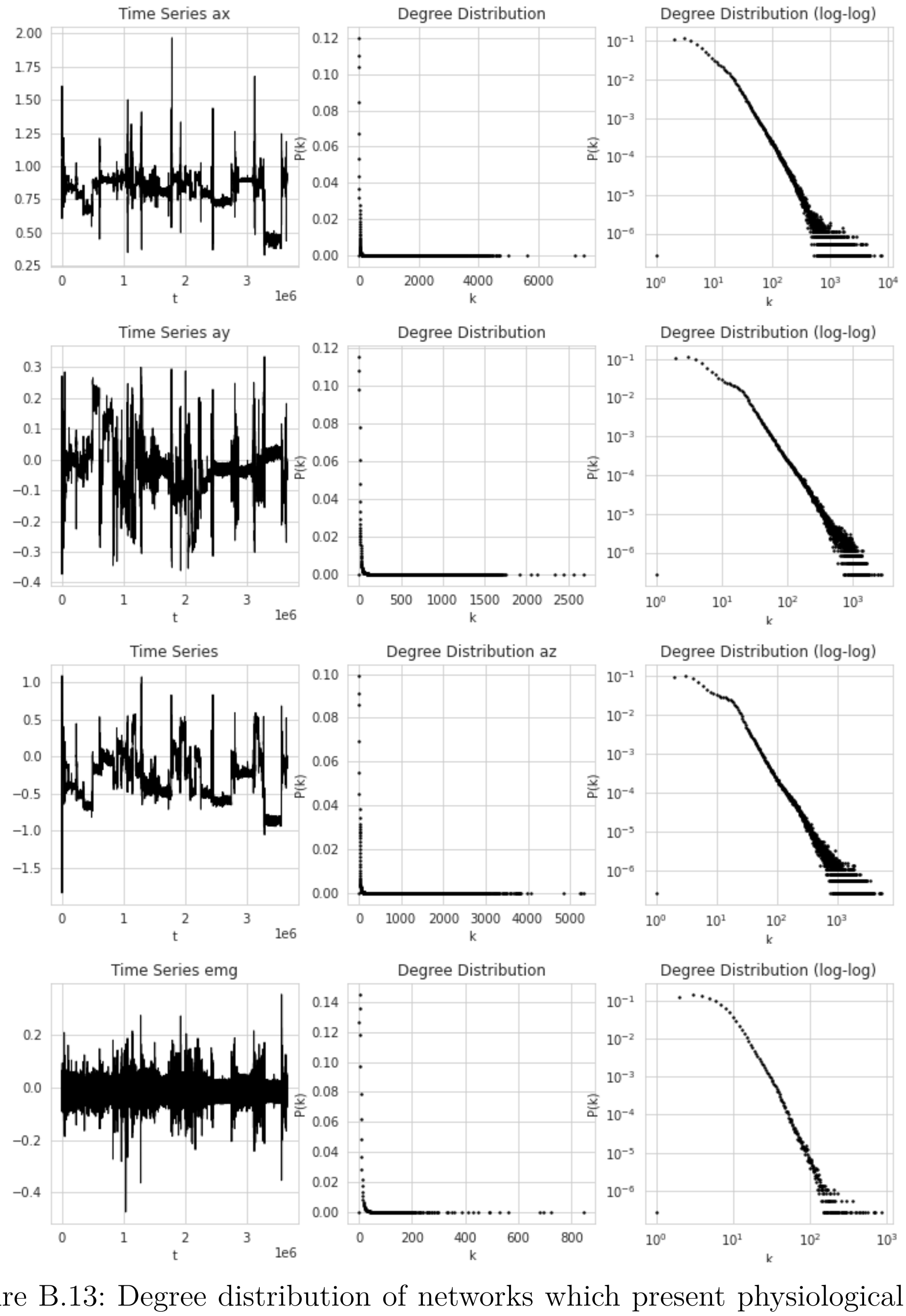


Figure B.13: Degree distribution of networks which present physiological time series data of the participant with identifier number 9 from the WESAD dataset. The data are 3-axis acceleromete data (x, y, z) and EMG. From left to right, the signal, the degree distribution, and the log-log graph

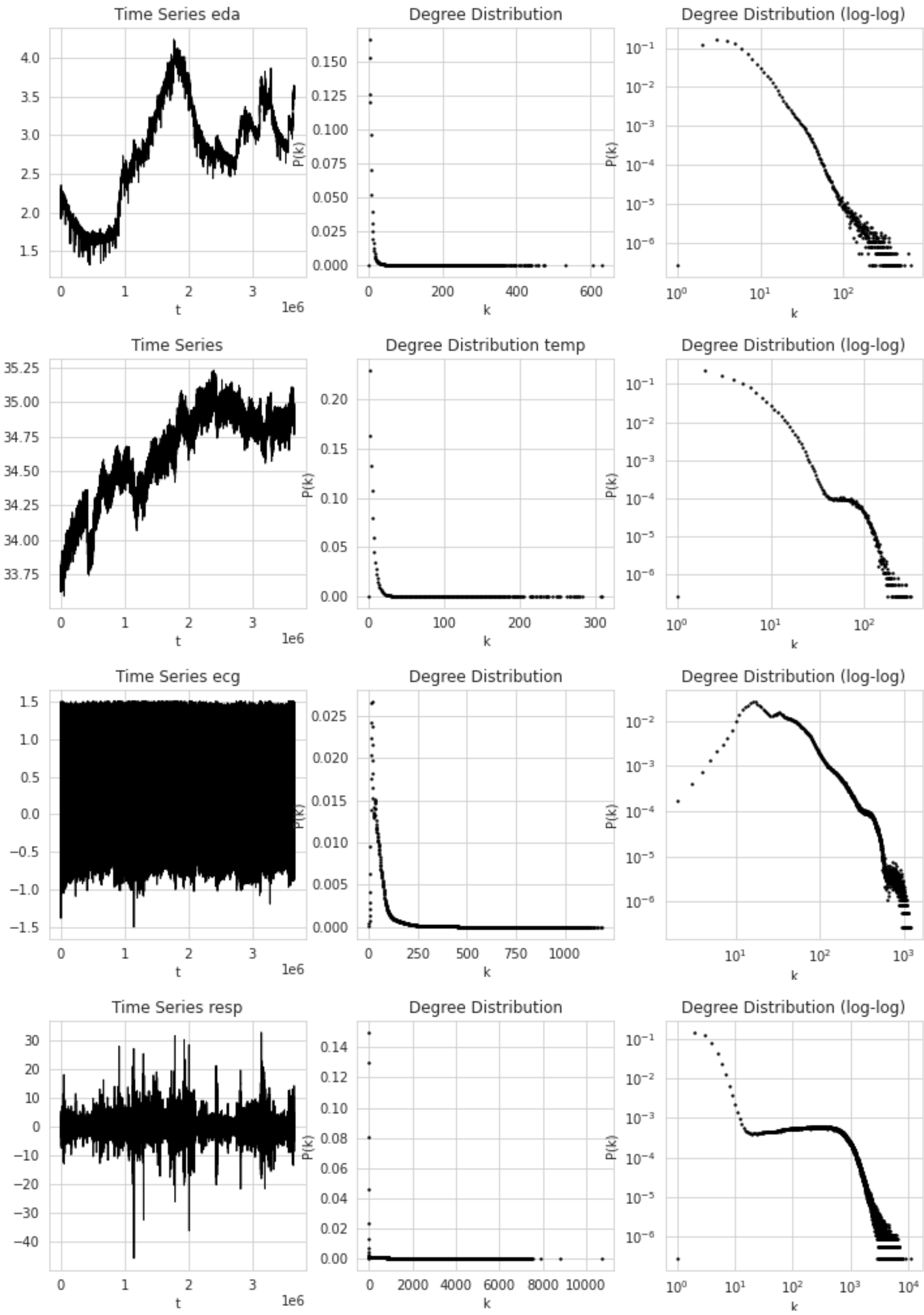


Figure B.14: Degree distribution of networks which present physiological time series data of the participant with identifier number 9 from the WESAD dataset. The data is EDA, temperature data, ECG, and respiration data. From left to right, the signal, the degree distribution, and the log-log graph

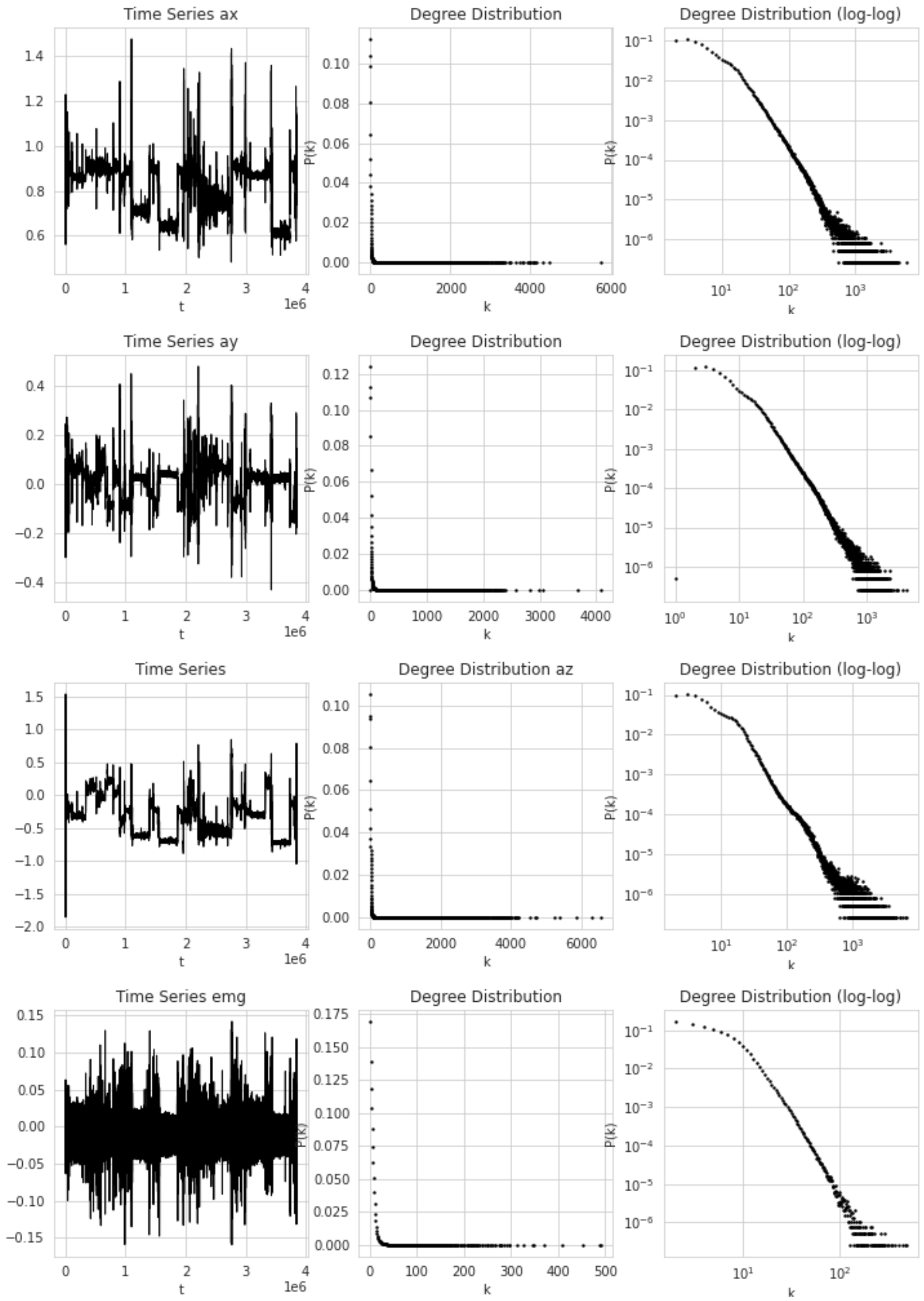


Figure B.15: Degree distribution of networks which present physiological time series data of the participant with identifier number 10 from the WESAD dataset. The data are 3-axis acceleromete data (x, y, z) and EMG. From left to right, the signal, the degree distribution, and the log-log graph

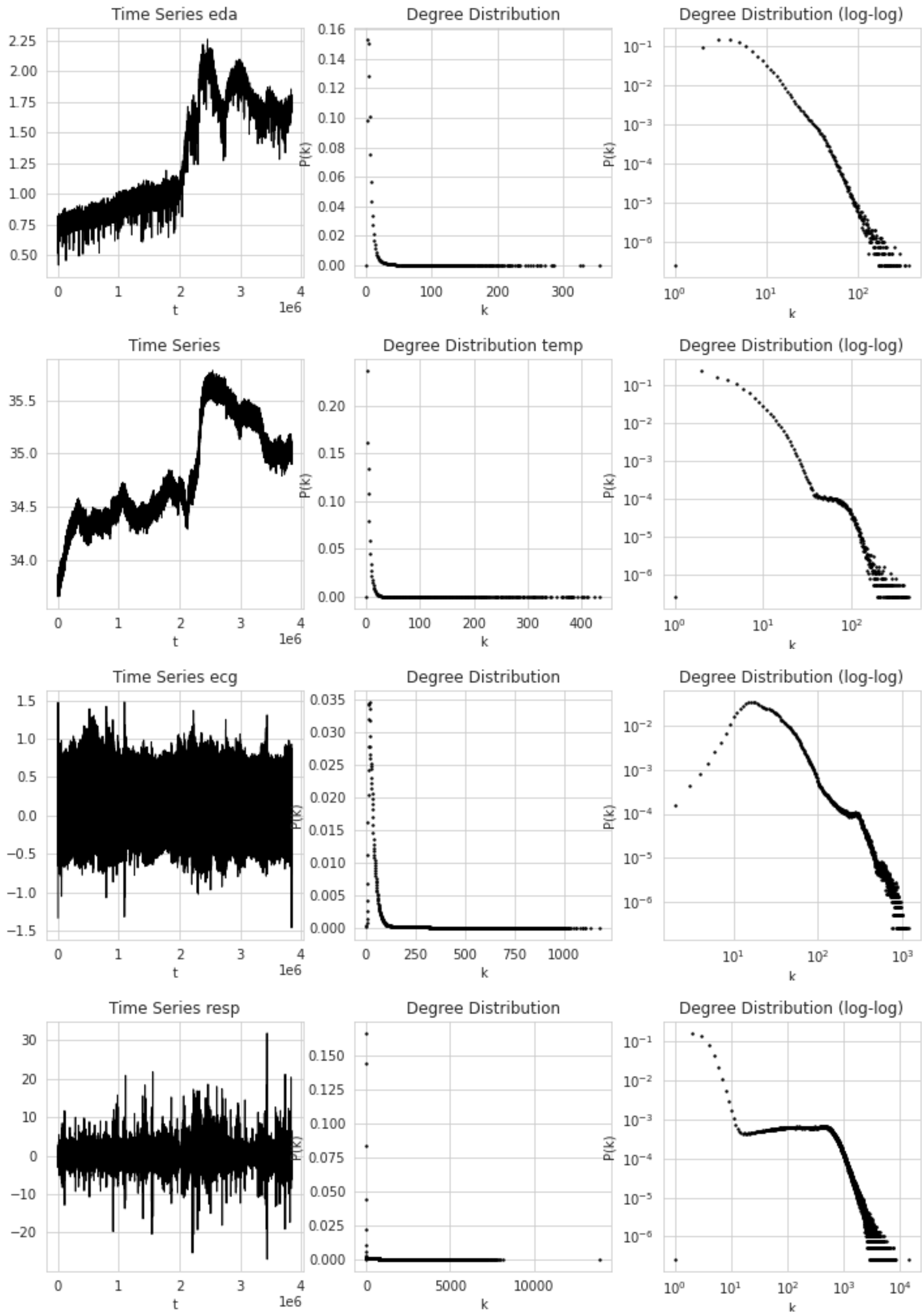


Figure B.16: Degree distribution of networks which present physiological time series data of the participant with identifier number 10 from the WESAD dataset. The data is EDA, temperature data, ECG, and respiration data. From left to right, the signal, the degree distribution, and the log-log graph

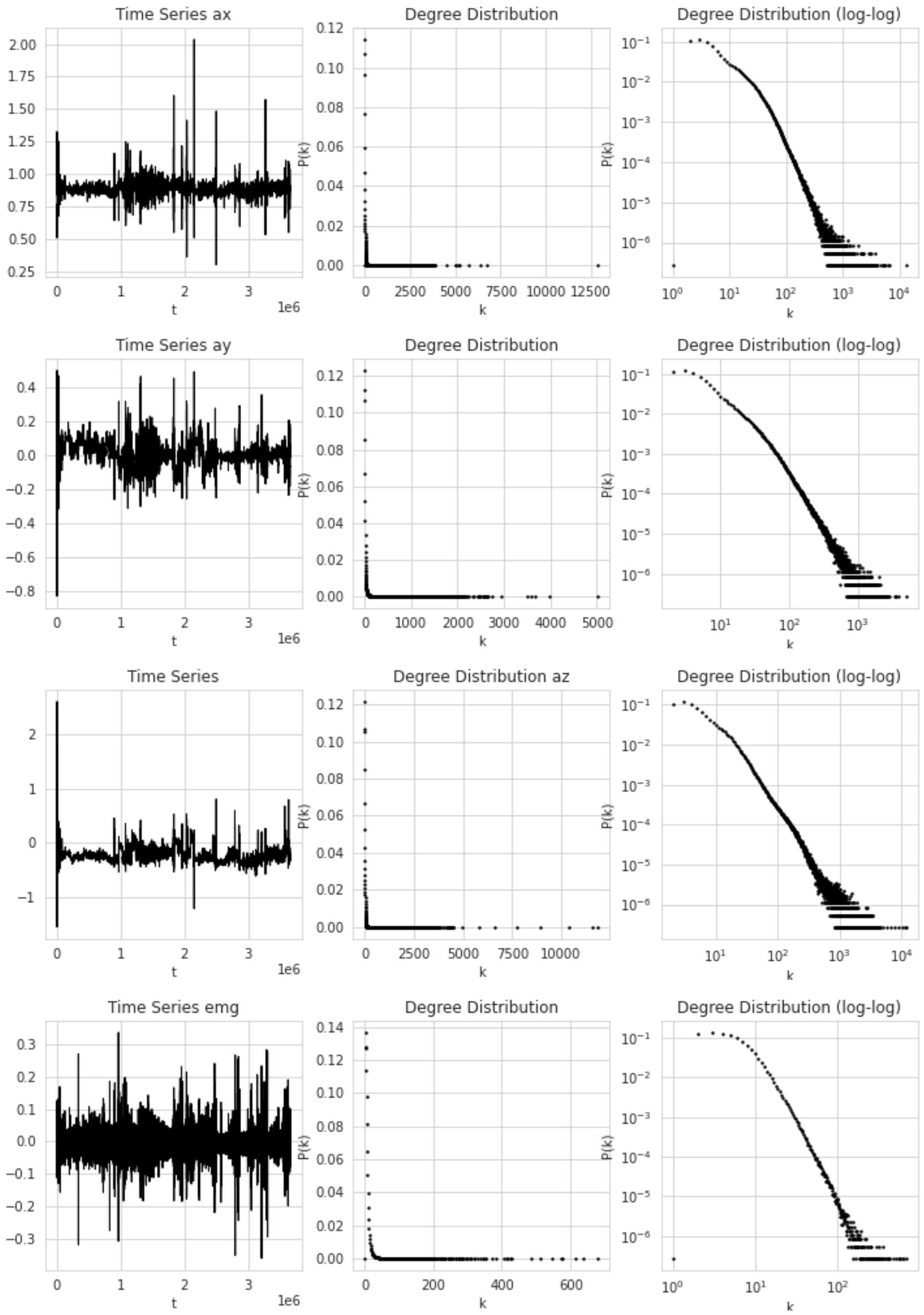


Figure B.17: Degree distribution of networks which present physiological time series data of the participant with identifier number 11 from the WESAD dataset. The data are 3-axis acceleromete data (x, y, z) and EMG. From left to right, the signal, the degree distribution, and the log-log graph

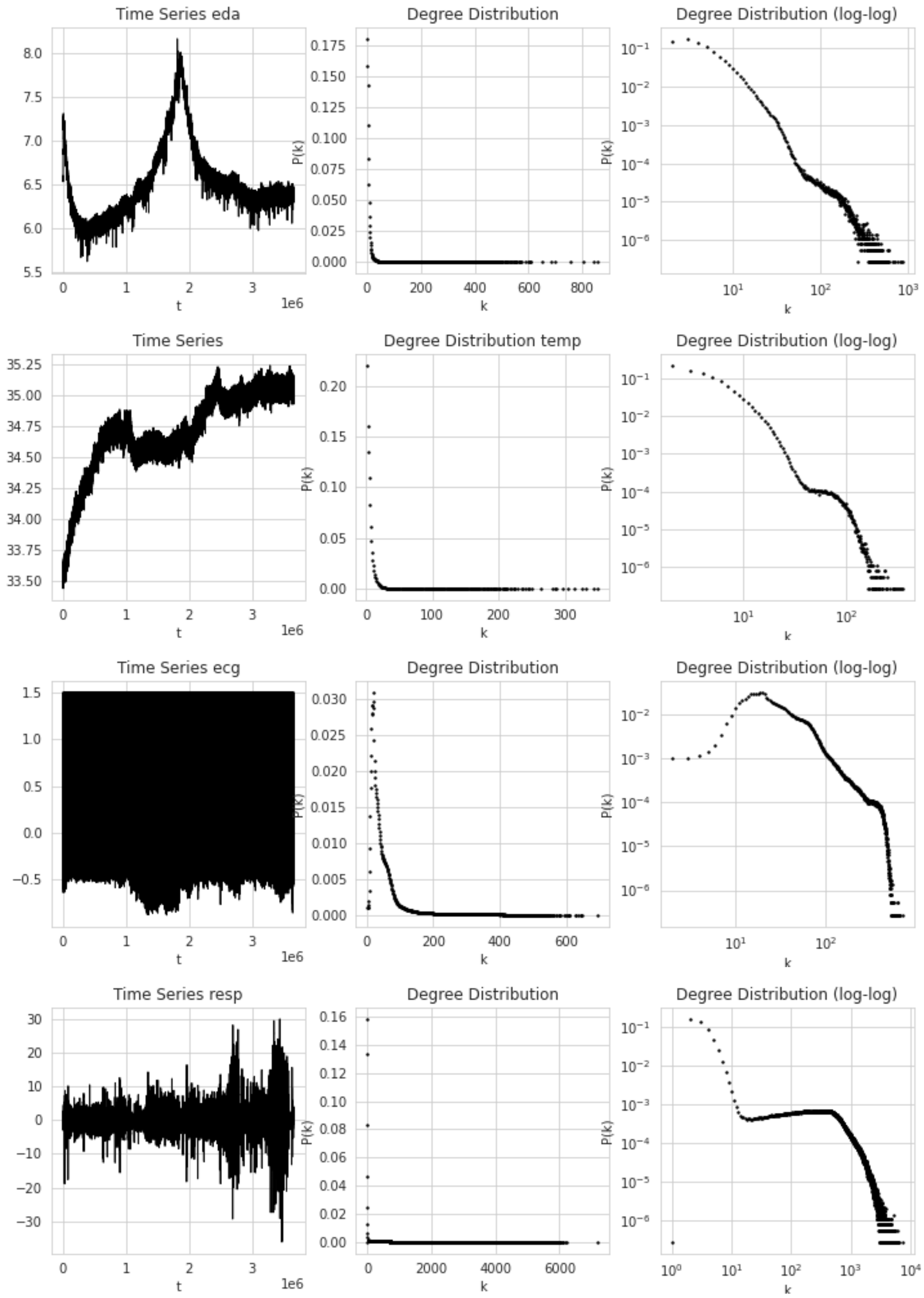


Figure B.18: Degree distribution of networks which present physiological time series data of the participant with identifier number 11 from the WESAD dataset. The data is EDA, temperature data, ECG, and respiration data. From left to right, the signal, the degree distribution, and the log-log graph

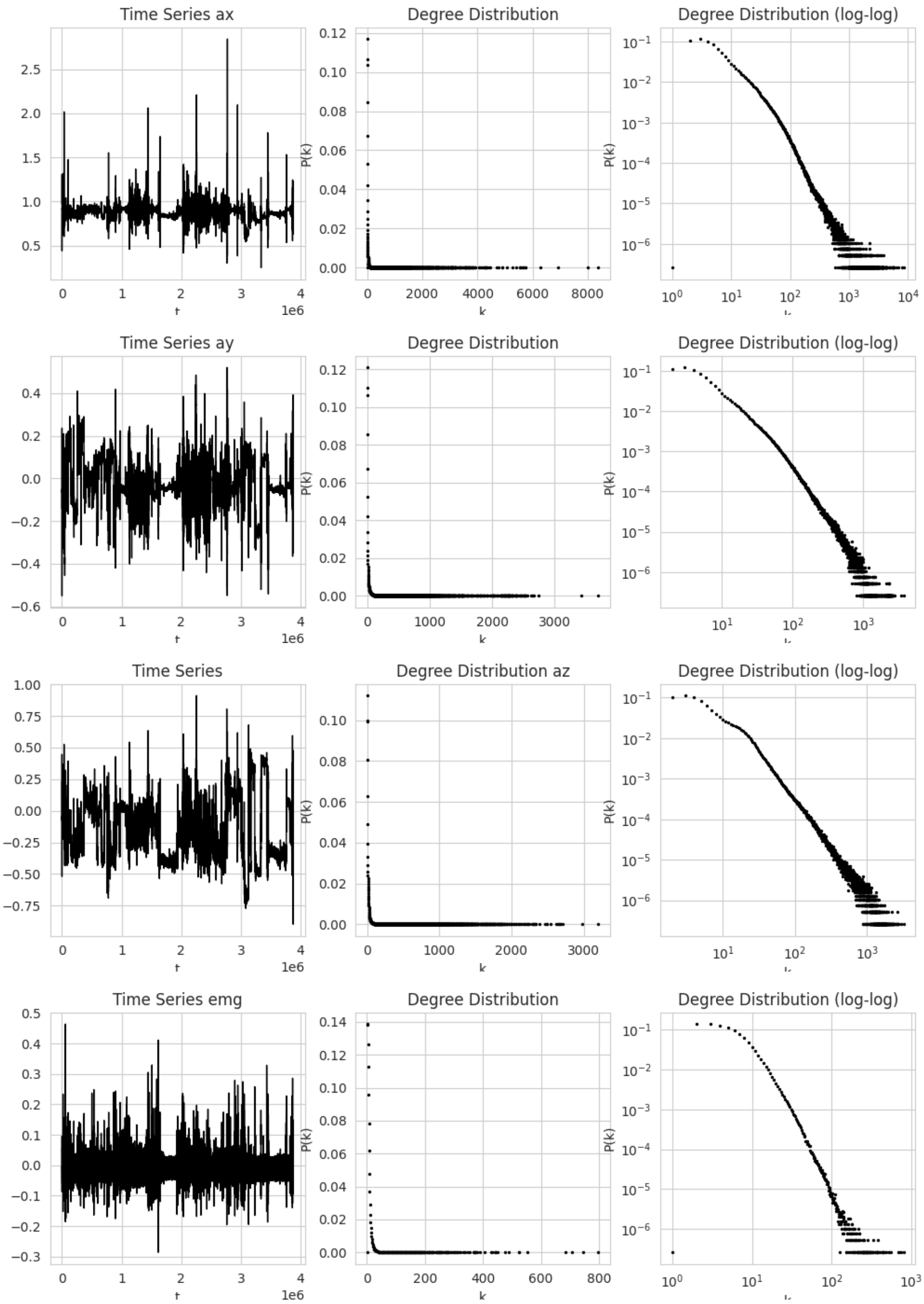


Figure B.19: Degree distribution of networks which present physiological time series data of the participant with identifier number 13 from the WESAD dataset. The data are 3-axis acceleromete data (x, y, z) and EMG. From left to right, the signal, the degree distribution, and the log-log graph

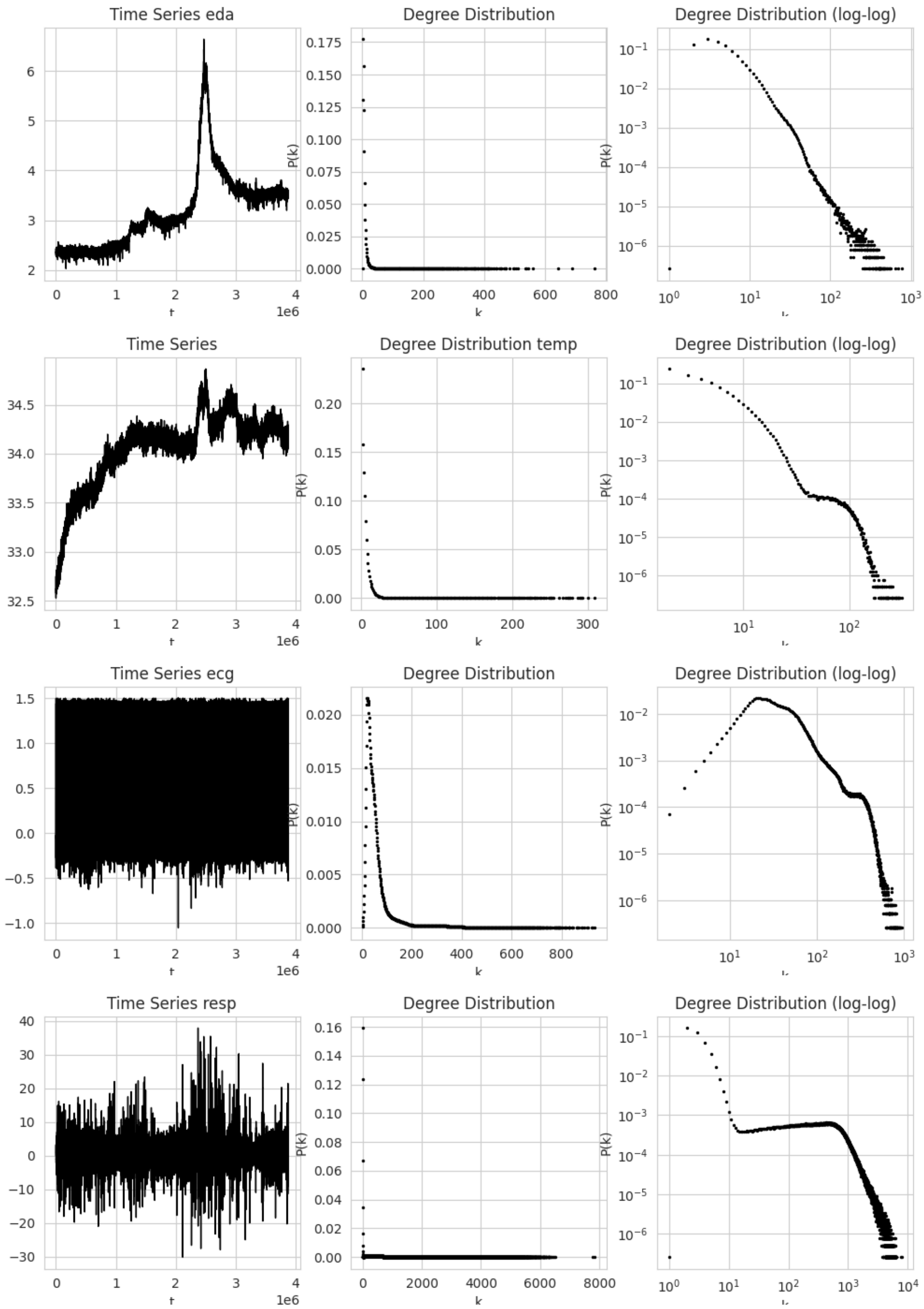


Figure B.20: Degree distribution of networks which present physiological time series data of the participant with identifier number 13 from the WESAD dataset. The data is EDA, temperature data, ECG, and respiration data. From left to right, the signal, the degree distribution, and the log-log graph

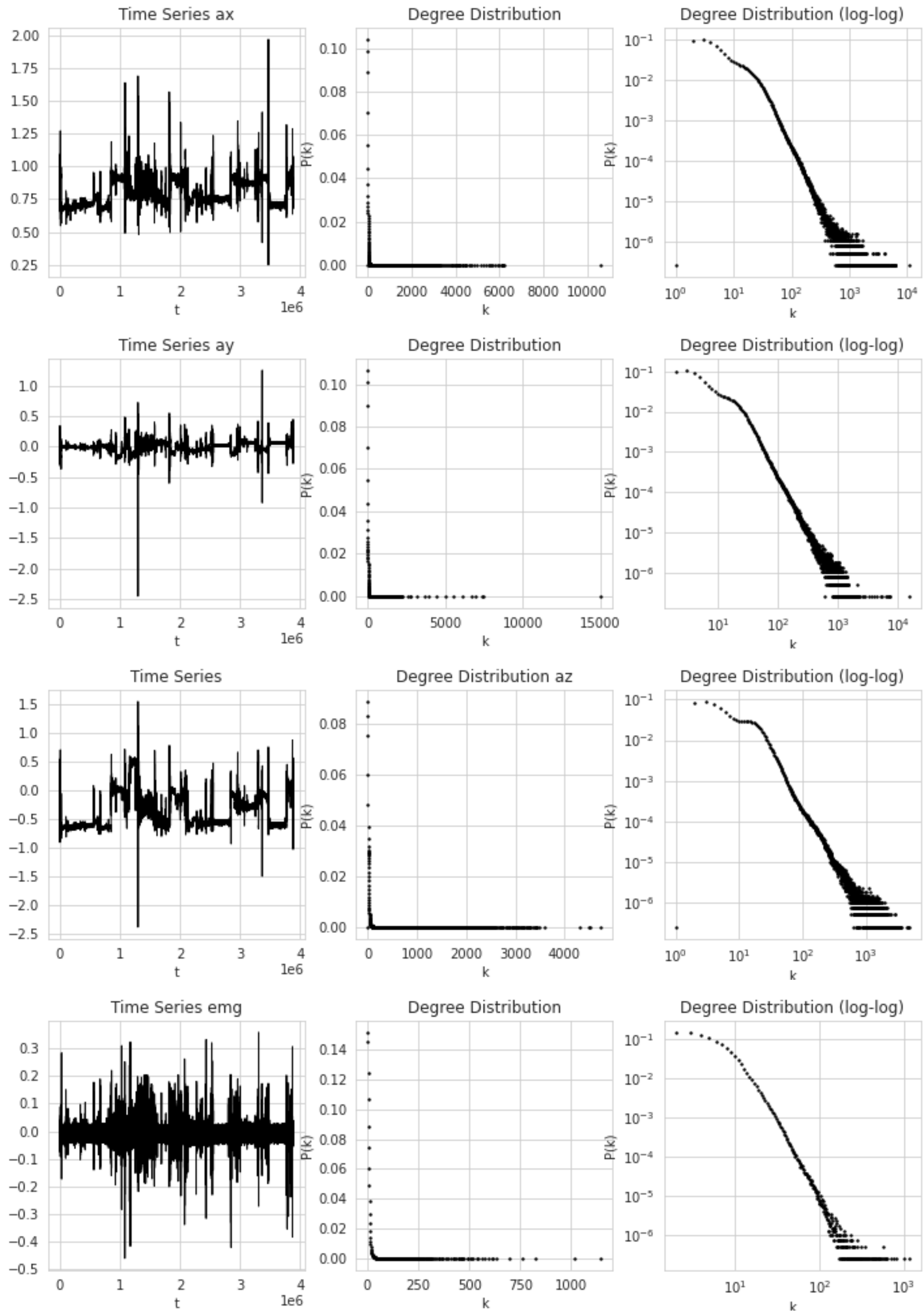


Figure B.21: Degree distribution of networks which present physiological time series data of the participant with identifier number 14 from the WESAD dataset. The data are 3-axis acceleromete data (x, y, z) and EMG. From left to right, the signal, the degree distribution, and the log-log graph

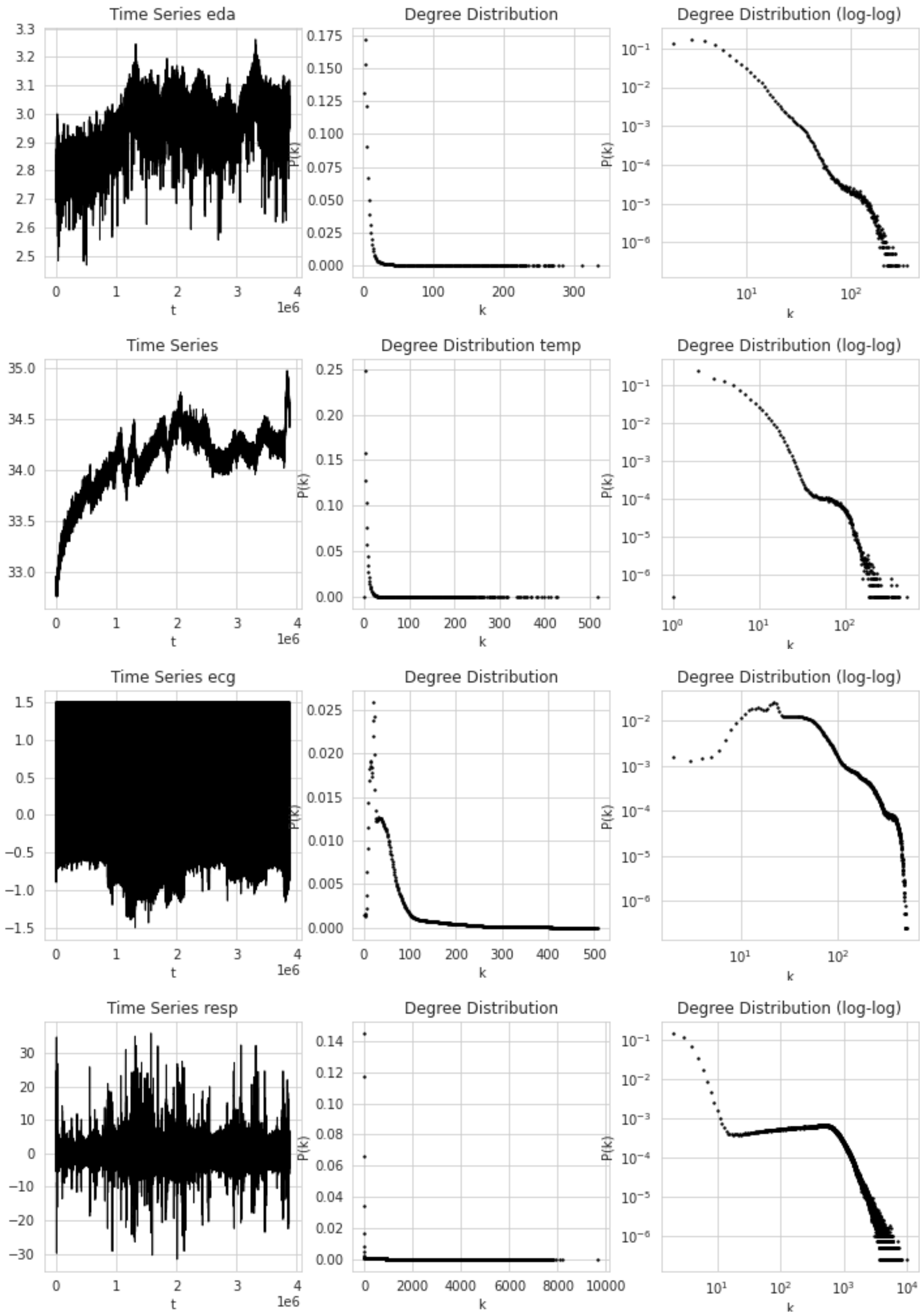


Figure B.22: Degree distribution of networks which present physiological time series data of the participant with identifier number 14 from the WESAD dataset. The data is EDA, temperature data, ECG, and respiration data. From left to right, the signal, the degree distribution, and the log-log graph

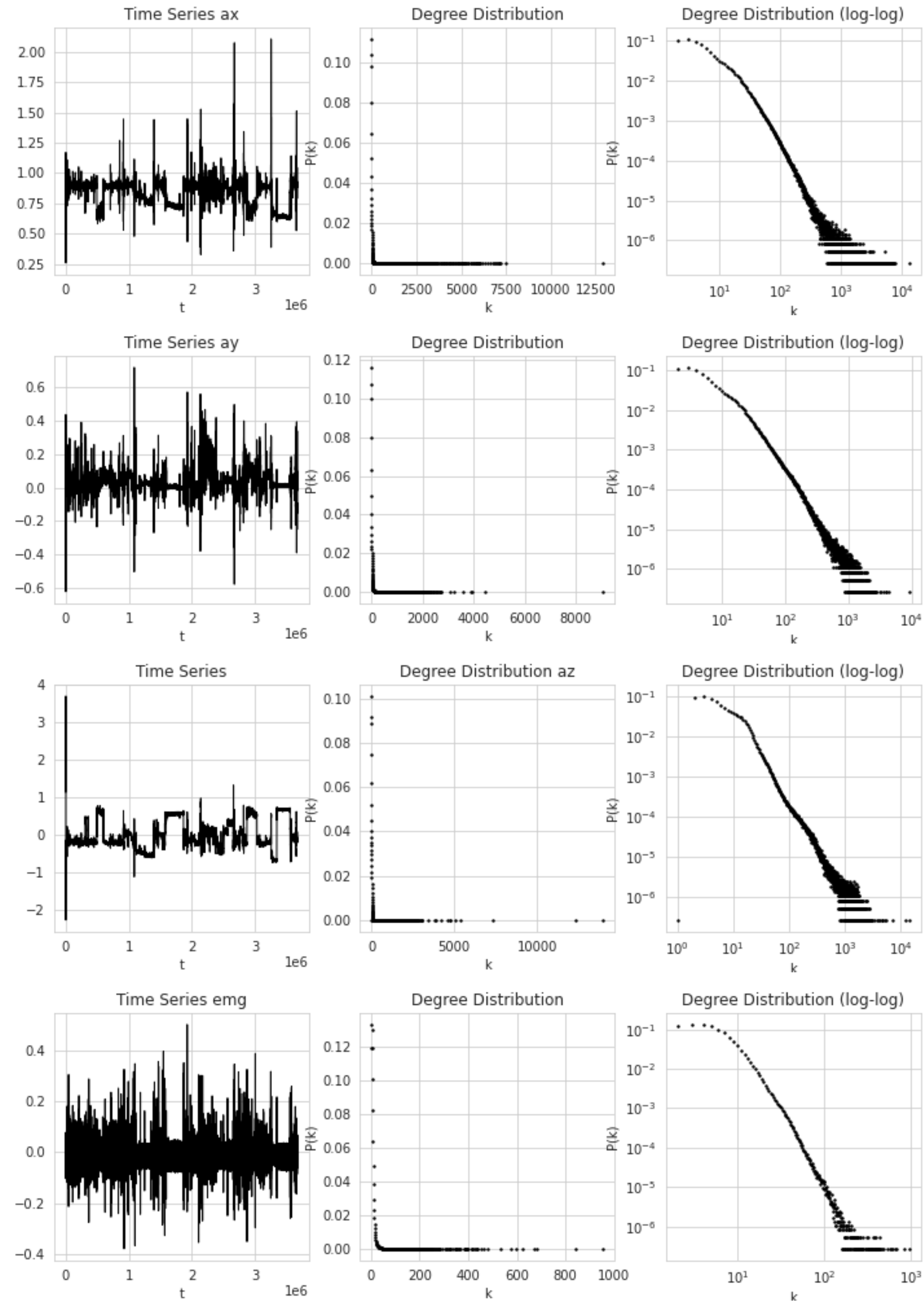


Figure B.23: Degree distribution of networks which present physiological time series data of the participant with identifier number 15 from the WESAD dataset. The data are 3-axis acceleromete data (x, y, z) and EMG. From left to right, the signal, the degree distribution, and the log-log graph

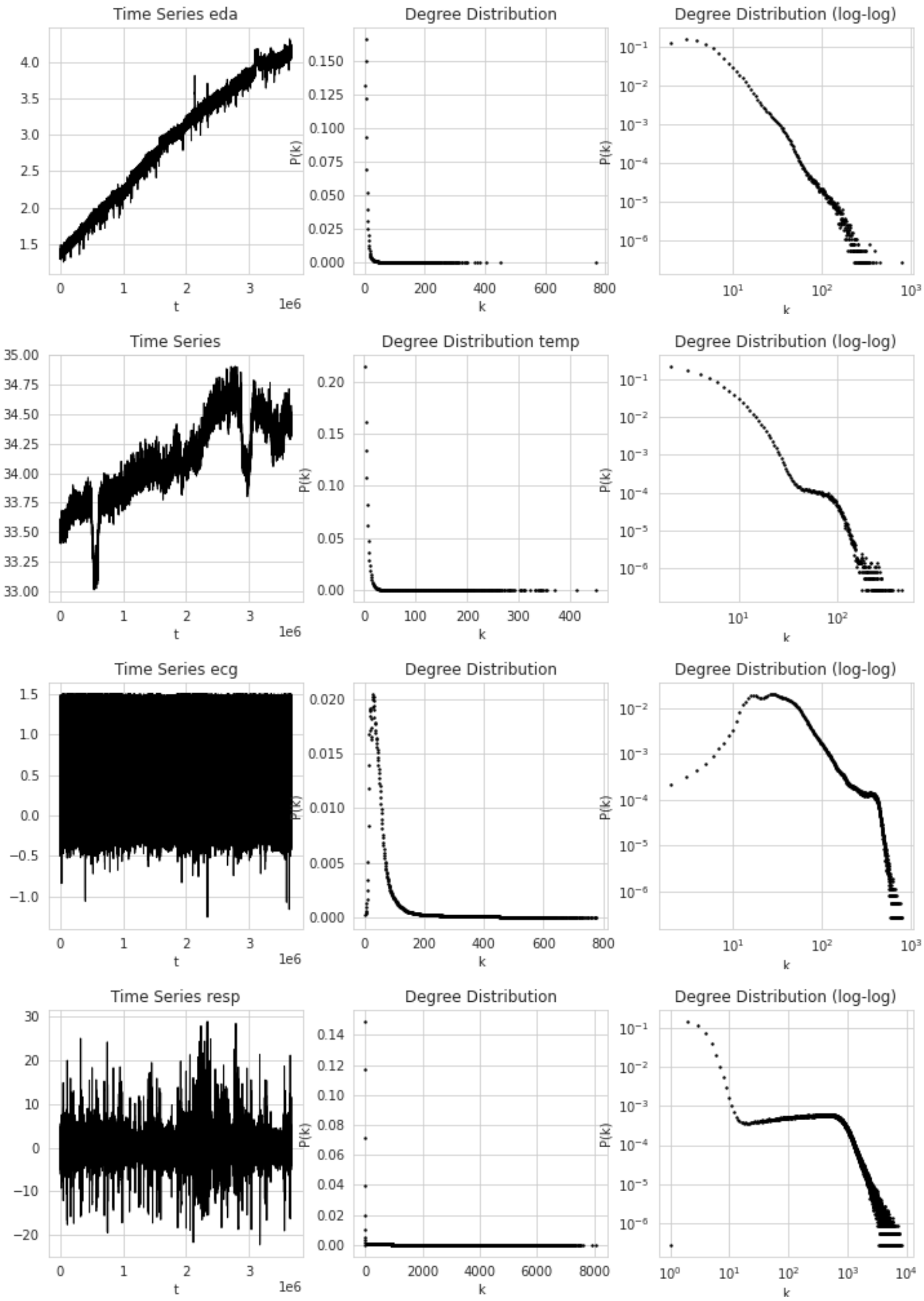


Figure B.24: Degree distribution of networks which present physiological time series data of the participant with identifier number 15 from the WESAD dataset. The data is EDA, temperature data, ECG, and respiration data. From left to right, the signal, the degree distribution, and the log-log graph

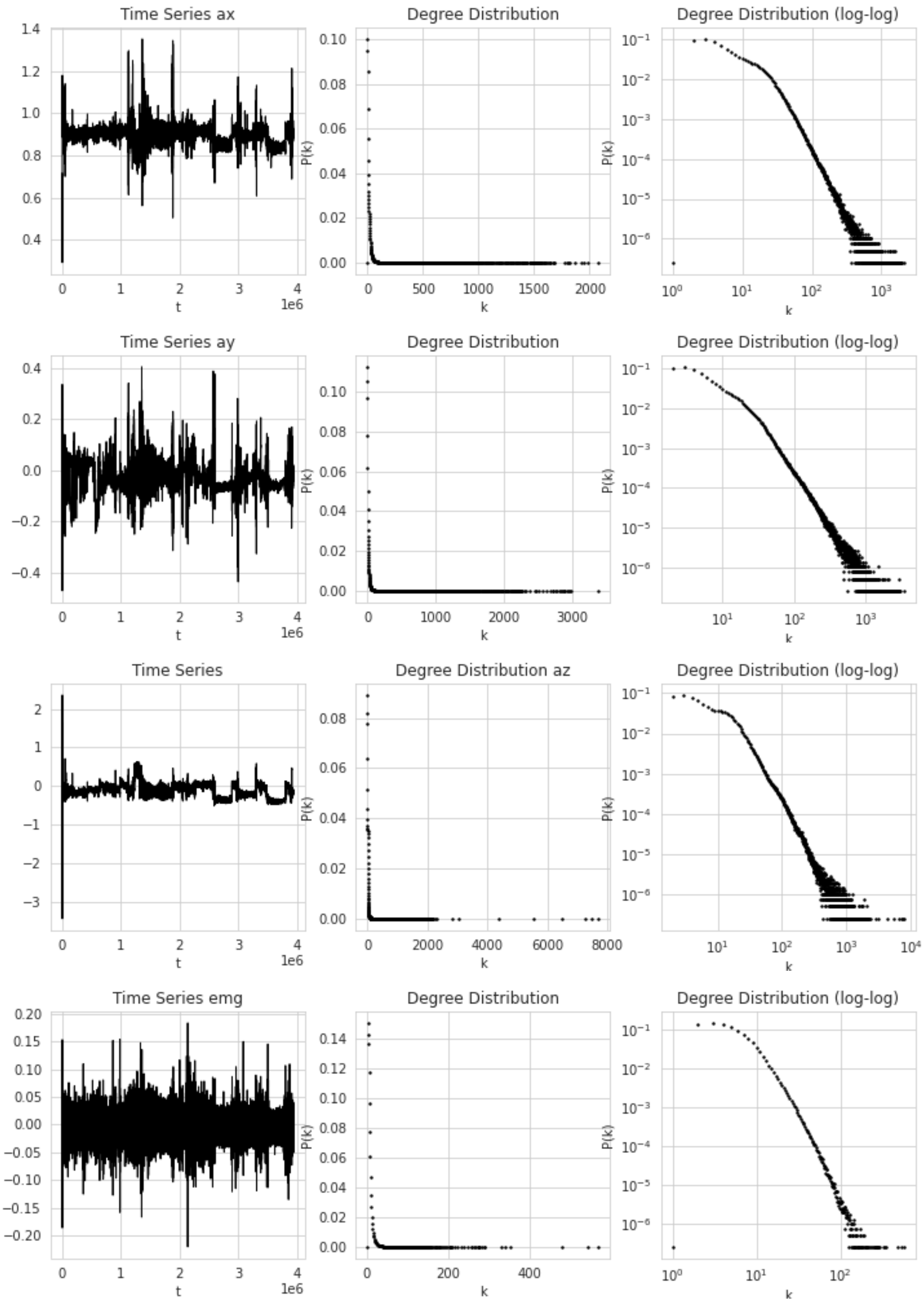


Figure B.25: Degree distribution of networks which present physiological time series data of the participant with identifier number 16 from the WESAD dataset. The data are 3-axis acceleromete data (x, y, z) and EMG. From left to right, the signal, the degree distribution, and the log-log graph

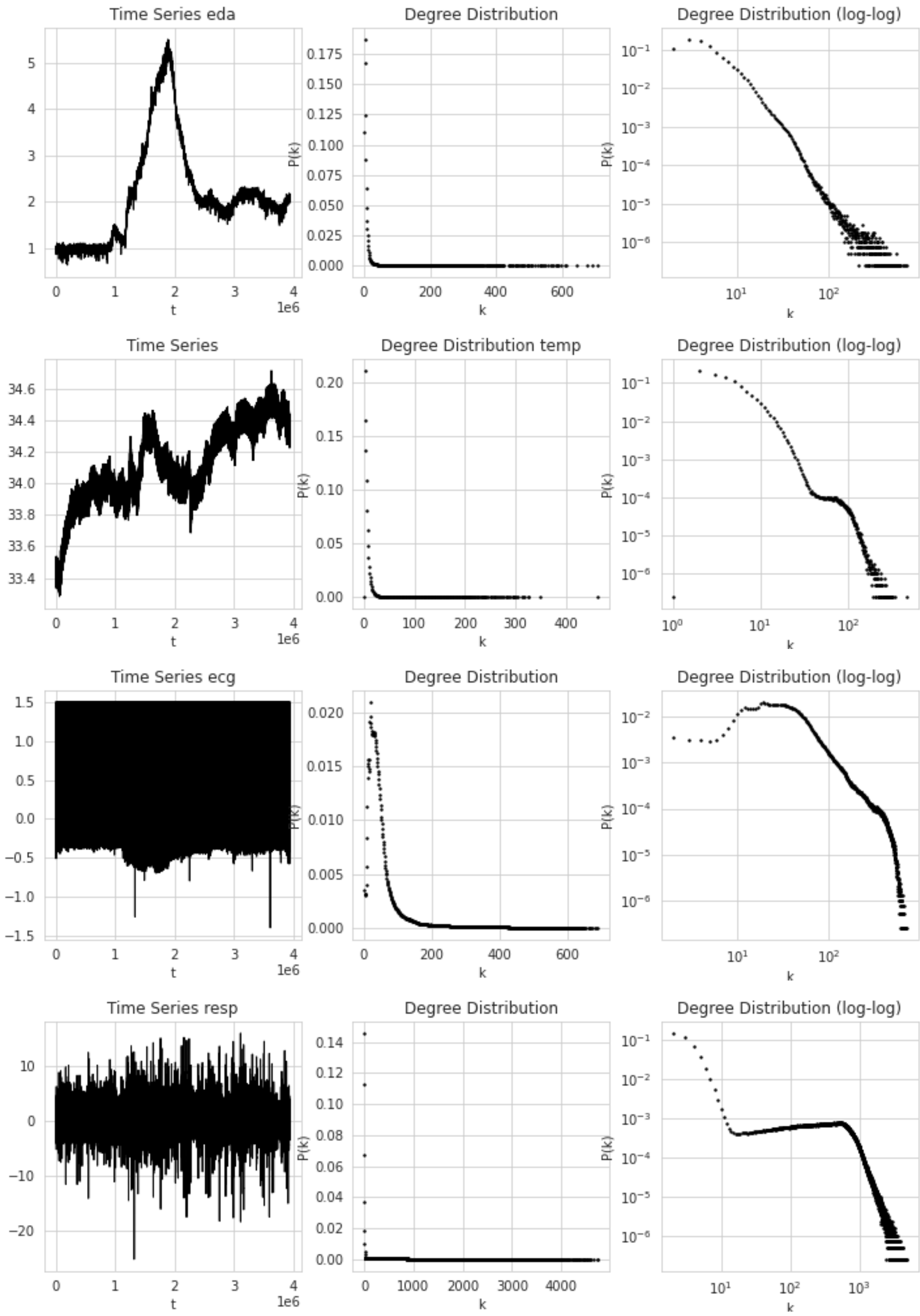


Figure B.26: Degree distribution of networks which present physiological time series data of the participant with identifier number 16 from the WESAD dataset. The data is EDA, temperature data, ECG, and respiration data. From left to right, the signal, the degree distribution, and the log-log graph

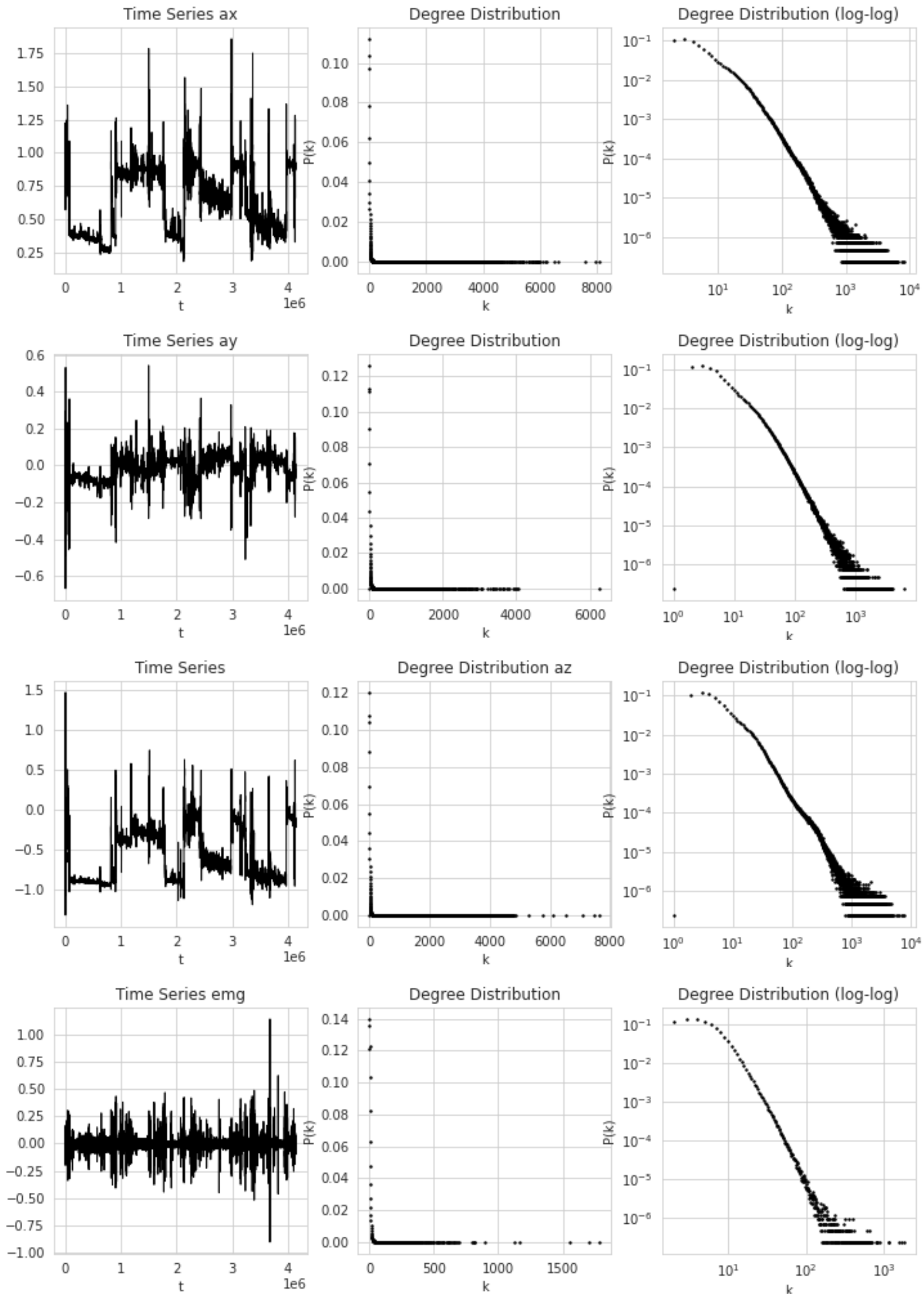


Figure B.27: Degree distribution of networks which present physiological time series data of the participant with identifier number 17 from the WESAD dataset. The data are 3-axis acceleromete data (x, y, z) and EMG. From left to right, the signal, the degree distribution, and the log-log graph

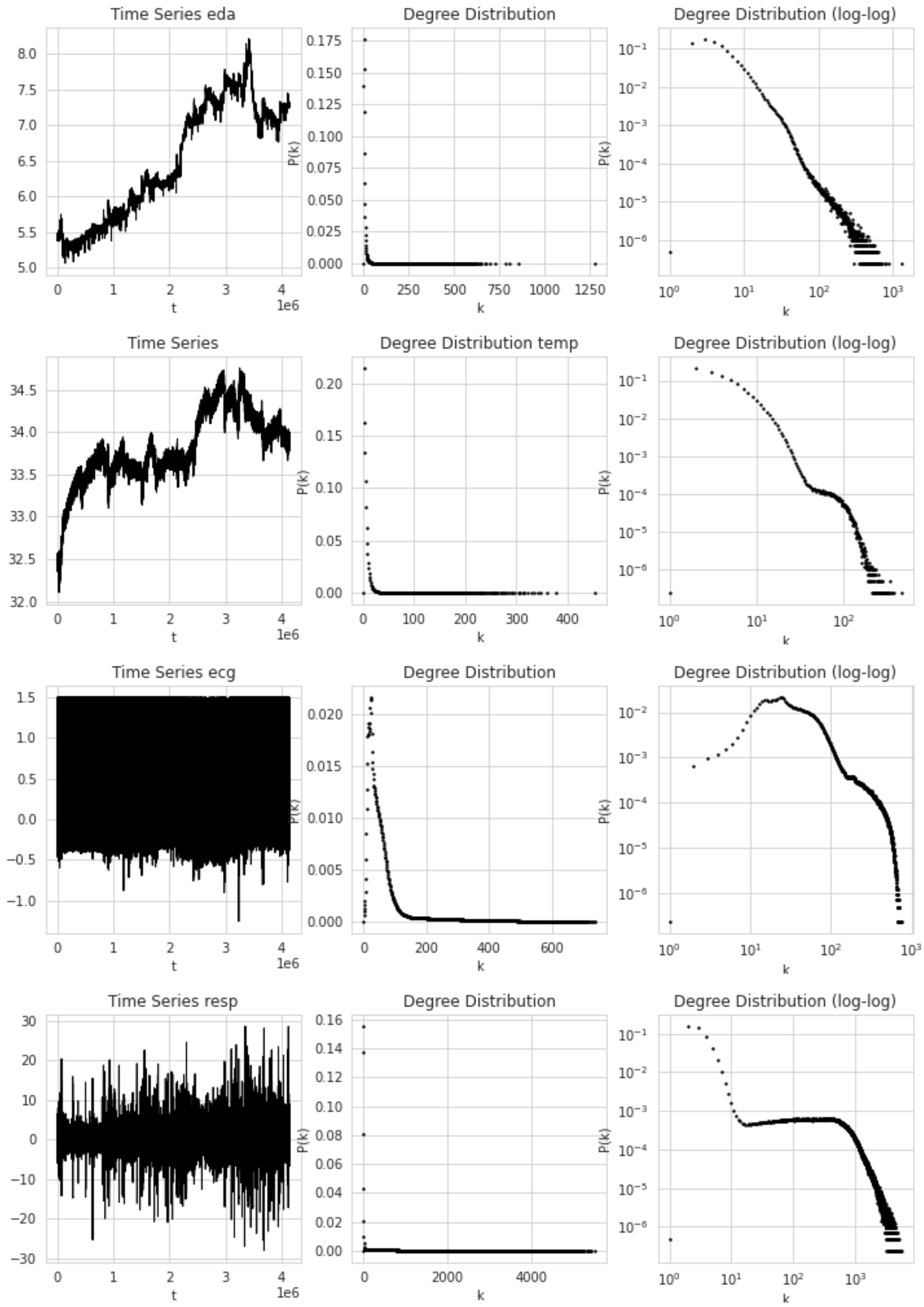


Figure B.28: Degree distribution of networks which present physiological time series data of the participant with identifier number 17 from the WESAD dataset. The data is EDA, temperature data, ECG, and respiration data. From left to right, the signal, the degree distribution, and the log-log graph

Bibliography

Abdulsadig, R. S. & Rodriguez-Villegas, E., 2024, ‘A comparative study in class imbalance mitigation when working with physiological signals’, *Frontiers in Digital Health*, vol. 6, p. 1377165.

Acikmese, Y. & Alptekin, S. E., 2019, ‘Prediction of stress levels with LSTM and passive mobile sensors’, *Procedia Computer Science*, vol. 159, pp. 658–667.

Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F. & Ziegelmeier, L., 2017, ‘Persistence images: A stable vector representation of persistent homology’, *Journal of Machine Learning Research*, vol. 18, no. 8, pp. 1–35.

Airij, A. G., Sudirman, R. & Sheikh, U. U., 2018, ‘Gsm and gps based real-time remote physiological signals monitoring and stress levels classification’, *2018 2nd International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*, IEEE, p. 130–135.

Aktas, M. E., Akbas, E. & Fatmaoui, A. E., 2019, ‘Persistence homology of networks: methods and applications’, *Applied Network Science*, vol. 4, no. 1, pp. 1–28.

Al-Fahoum, A. S. & Al-Fraihat, A. A., 2014, ‘Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains’, *International Scholarly Research Notices*, vol. 2014, no. 1, p. 730218.

Alberdi, A., Aztiria, A. & Basarab, A., 2016, ‘Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review’, *Journal of Biomedical Informatics*, vol. 59, pp. 49–75.

Albert, R. & Barabási, A.-L., 2002, 'Statistical mechanics of complex networks', *Reviews of modern physics*, vol. 74, no. 1, p. 47.

Aldwin, C. M., 2007, *Stress, coping, and development: An integrative perspective*, Guilford Press.

Aloysius, N. & Geetha, M., 2017, 'A review on deep convolutional neural networks', *2017 international conference on communication and signal processing (ICCSP)*, IEEE, p. 588.

Alshamrani, M., 2021, 'An advanced stress detection approach based on processing data from wearable wrist devices', *Int.J.Adv.Comput.Sci.Appl*, vol. 12, pp. 399–405.

Amaral, I., 2022, *Complex networks*, Encyclopedia of Big Data, Springer, pp. 198–201.

Anisman, H. & Merali, Z., 1999, 'Understanding stress: Characteristics and caveats', *Alcohol Research Health*, vol. 23, no. 4, p. 241, pmid:10890820.

Asif, A., Majid, M. & Anwar, S. M., 2019, 'Human stress classification using eeg signals in response to music tracks', *Computers in biology and medicine*, vol. 107, p. 182–196.

Assaad, R. & El-adaway, I. H., 2021, 'Impact of dynamic workforce and workplace variables on the productivity of the construction industry: New gross construction productivity indicator', *Journal of Management in Engineering*, vol. 37, no. 1, pp. 1–15.

Asselbergs, J., Ruwaard, J., Ejdys, M., Schrader, N., Sijbrandij, M. & Riper, H., 2016, 'Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study', *Journal of Medical Internet research*, vol. 18, no. 3, p. e72.

Babu, M., Lautman, Z., Lin, X., Sobota, M. H. & Snyder, M. P., 2024, 'Wearable devices: Implications for precision medicine and the future of health care', *Annual Review of Medicine*, vol. 75, pp. 401–415.

Badr, Y., Tariq, U., Al-Shargie, F., Babiloni, F., Mughairbi, F. A. & Al-Nashash, H., 2024, 'A review on evaluating mental stress by deep learning using EEG signals', *Neural Computing and Applications*, p. 1–26.

Baldoumas, G., Peschos, D., Tatsis, G., Votis, C. I., Chronopoulos, S. K., Christofilakis, V., Kostarakis, P., Sarlis, N. V., Skordas, E. S. & Naka, K. K., 2018, 'Comparison of the RR intervals in ECG and oximeter signals to be used in complexity measures of natural time analysis', *2018 7th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, IEEE, pp. 1–4.

Barabási, A.-L., 2013, 'Network science', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1987, p. 20120375.

Barabási, A.-L. & Albert, R., 1999, 'Emergence of scaling in random networks', *Science*, vol. 286, no. 5439, pp. 509–512.

Barabási, A.-L. & Bonabeau, E., 2003, 'Scale-free networks', *Scientific American*, vol. 288, no. 5, pp. 60–69.

Barrat, A., Barthelemy, M. & Vespignani, A., 2008, *Dynamical processes on complex networks*, Cambridge University Press.

BenjamTorr, 2021, 'TDA DeepL TimeSeries', https://github.com/BenjamTorr/TDA_DeepL_TimeSeries.

Bernard, C., 1957, *An introduction to the study of experimental medicine*, vol. 400, Courier Corporation.

Betti, S., Lova, R. M., Rovini, E., Acerbi, G., Santarelli, L., Cabiati, M., Ry, S. D. & Cavallo, F., 2017, 'Evaluation of an integrated system of wearable physiological sensors for stress monitoring in working environments by using biological markers', *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 8, p. 1748–1758.

Bhavsar, H. & Ganatra, A., 2012, 'Variations of support vector machine classification technique: a survey', *International Journal of Advanced Computer Research*, vol. 2, no. 6, p. 230–236.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U., 2006, 'Complex networks: Structure and dynamics', *Physics reports*, vol. 424, no. 4-5, pp. 175–308.

Boissonnat, J.-D. & Pritam, S., 2020, 'Edge collapse and persistence of flag complexes', *36th International Symposium on Computational Geometry (SoCG 2020)*, Schloss-Dagstuhl-Leibniz Zentrum für Informatik.

Bota, P. J., Wang, C., Fred, A. L. & Silva, H. P. D., 2019, 'A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals', *IEEE Access*, vol. 7, pp. 140990–141020.

Boudin, F., 2013, 'A comparison of centrality measures for graph-based keyphrase extraction', *Sixth Int.Jt.Conf. on Natural Language Processing*, pp. 834–838.

Broido, A. D. & Clauset, A., 2019, 'Scale-free networks are rare', *Nature communications*, vol. 10, no. 1, p. 1017.

Brouwer, A.-M. & Hogervorst, M. A., 2014, 'A new paradigm to induce mental stress: the Sing-a-Song Stress Test (ssst)', *Frontiers in neuroscience*, vol. 8, p. 224.

Bubenik, P., 2015, 'Statistical topological data analysis using persistence landscapes.', *J.Mach.Learn.Res.*, vol. 16, no. 1, pp. 77–102.

Can, Y. S., Arnrich, B. & Ersoy, C., 2019a, 'Stress detection in daily life scenarios using smart phones and wearable sensors: A survey', *Journal of Biomedical Informatics*, vol. 92, p. 103139.

Can, Y. S., Chalabianloo, N., Ekiz, D. & Ersoy, C., 2019b, 'Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study', *Sensors*, vol. 19, no. 8, p. 1849.

Cannon, W. B., 1926, 'Physiological regulation of normal states: some tentative postulates concerning biological homeostatics', *Ses Amis, ses Colleges, ses Eleves*.

Carlsson, G., Zomorodian, A., Collins, A. & Guibas, L. J., 2005, 'Persistence barcodes for shapes', *International Journal of Shape Modeling*, vol. 11, no. 02, pp. 149–187.

Carneiro, M. G. & Zhao, L., 2013, ‘High level classification totally based on complex networks’, *2013 BRICS Congress on Computational Intelligence and 11th Brazilian Congress on Computational Intelligence*, IEEE, pp. 507–514.

Carrière, M., Chazal, F., Ike, Y., Lacombe, T., Royer, M. & Umeda, Y., 2020, ‘Perslay: A neural network layer for persistence diagrams and new graph topological signatures’, *International Conference on Artificial Intelligence and Statistics*, PMLR, pp. 2786–2796.

Castaldo, R., Montesinos, L., Melillo, P., Massaro, S. & Pecchia, L., 2017, *To What Extent Can We Shorten HRV Analysis in Wearable Sensing? A Case Study on Mental Stress Detection.*, EMBEC & NBC 2017, Springer, pp. 643–646.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L. & Lopez, A., 2020, ‘A comprehensive survey on support vector machine classification: Applications, challenges and trends’, *Neurocomputing*, vol. 408, pp. 189–215.

Chattopadhyay, S., Das, A. K. & Ghosh, K., 2020, ‘Finding patterns in the degree distribution of real-world complex networks: going beyond power law’, *Pattern Analysis and Applications*, vol. 23, no. 2, pp. 913–932.

Chazal, F. & Michel, B., 2021, ‘An introduction to topological data analysis: fundamental and practical aspects for data scientists’, *Frontiers in Artificial Intelligence*, vol. 4, pp. 1–28.

Chazal, F., Steiner, D. C. & Mérigot, Q., 2011, ‘Geometric inference for probability measures’, *Foundations of Computational Mathematics*, vol. 11, no. 6, pp. 733–751.

Cheema, A. & Singh, M., 2019, ‘An application of phonocardiography signals for psychological stress detection using non-linear entropy based features in empirical mode decomposition domain’, *Applied Soft Computing*, vol. 77, p. 24–33.

Chen, J., Wang, H. & Hua, C., 2018, ‘Assessment of driver drowsiness using EEG signals based on multiple functional brain networks’, *Int.J.Psychophysiol.*, vol. 133, pp. 120–130.

Chen, L. L., Zhao, Y., Ye, P. F., Zhang, J. & Zou, J. Z., 2017, 'Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers', *Expert Systems with Applications*, vol. 85, pp. 279–291.

Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A. & Batista, G., 2015, 'The UCR time series classification archive', www.cs.ucr.edu/~eamonn/time_series_data/.

Chung, Y.-M., Cruse, W. & Lawson, A., 2020, 'A persistent homology approach to time series classification', *arXiv e-prints*, p. arXiv: 2003.06462.

Cohen, E., Delling, D., Pajor, T. & Werneck, R. F., 2014, 'Computing classic closeness centrality, at scale', *Proceedings of the second ACM conference on Online social networks*, pp. 37–50.

Cohen, M. X., 2004, 'Advantages and limitations of time-and time frequency-domain analyses', *Analyzing neural time series data: Theory and practice*.

Cohen, S., Janicki-Deverts, D. & Miller, G. E., 2007, 'Psychological stress and disease', *The Journal of the American Medical Association*, vol. 298, no. 14, pp. 1685–1687.

Colliri, T., Ji, D., Pan, H. & Zhao, L., 2018, 'A network-based high level data classification technique', *Int.Jt.Conf. on Neural Networks (IJCNN)*, IEEE, pp. 1–8.

Cosoli, G., Antognoli, L., Panni, L. & Scalise, L., 2024, 'Indirect estimation of breathing rate using wearable devices', *IEEE Transactions on Instrumentation and Measurement*.

Cosoli, G., Antognoli, L. & Scalise, L., 2023, 'Methods for the metrological characterization of wearable devices for the measurement of physiological signals: state of the art and future challenges', *MethodsX*, vol. 10, p. 102038.

Craik, A., He, Y. & Contreras-Vidal, J. L., 2019, 'Deep learning for electroencephalogram (EEG) classification tasks: a review', *Journal of Neural Engineering*, vol. 16, no. 3, p. 031001.

Cunningham, P. & Delany, S. J., 2021, 'K-nearest neighbour classifiers-a tutorial', *ACM computing surveys (CSUR)*, vol. 54, no. 6, pp. 1–25.

Curiskis, S. A., Osborn, T. R. & Kennedy, P. J., 2015, 'Link prediction and topological feature importance in social networks', *Australian Data Mining Conference*, Australian Computer Society, pp. 39–50.

Cyranka, J., Georges, A. & Meyer, D., 2019, 'Mapper based classifier', *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, IEEE, pp. 1099–1106.

da F. Costa, L., Rodrigues, F. A., Travieso, G. & Boas, P. R. V., 2007, 'Characterization of complex networks: A survey of measurements', *Advances in Physics*, vol. 56, no. 1, pp. 167–242.

da Fontoura Costa, L., Jr, O. N. O., Travieso, G., Rodrigues, F. A., Boas, P. R. V., Antiqueira, L., Viana, M. P. & Rocha, L. E. C., 2011, 'Analyzing and modeling real-world phenomena with complex networks: a survey of applications', *Advances in Physics*, vol. 60, no. 3, pp. 329–412.

Davis, M. T., Holmes, S. E., Pietrzak, R. H. & Esterlis, I., 2017, 'Neurobiology of chronic stress-related psychiatric disorders: evidence from molecular imaging studies', *Chronic Stress*, vol. 1, p. 2470547017710916.

Dawson, M. E., Schell, A. M. & Filion, D. L., 2017, *The electrodermal system.*, Handbook of psychophysiology, Cambridge University Press, p. 217–243.

Dedovic, K., Renwick, R., Mahani, N. K., Engert, V., Lupien, S. J. & Pruessner, J. C., 2005, 'The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain', *Journal of psychiatry & neuroscience : JPN*, vol. 30, no. 5, pp. 319–325.

Deldari, S., Smith, D. V., Sadri, A. & Salim, F., 2020, 'Espresso: Entropy and shape aware time-series segmentation for processing heterogeneous sensor data', *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, pp. 1–24.

Dhanabal, S. & Chandramathi, S., 2011, 'A review of various k-nearest neighbor query processing techniques', *International Journal of Computer Applications*, vol. 31, no. 7, p. 14–22.

Ding, J., Liu, C., Zheng, Y., Zhang, Y., Yu, Z., Li, R., Chen, H., Piao, J., Wang, H. & Liu, J., 2024, 'Artificial intelligence for complex network: Potential, methodology and application', *arXiv preprint arXiv:2402.16887*.

Dirafzoon, A., Lokare, N. & Lobaton, E., 2016, 'Action classification from motion capture data using topological data analysis', *2016 IEEE global conference on signal and information processing (globalSIP)*, IEEE, pp. 1260–1264.

Diykh, M., Li, Y. & Wen, P., 2017, 'Classify epileptic EEG signals using weighted complex networks based community structure detection', *Expert Systems with Applications*, vol. 90, pp. 87–100.

Eckmann, J.-P., Kamphorst, S. O. & Ruelle, D., 1995, 'Recurrence plots of dynamical systems', *World Scientific Series on Nonlinear Science Series A*, vol. 16, pp. 441–446.

Edelsbrunner, H., Letscher, D. & Zomorodian, A., 2000, 'Topological persistence and simplification', *Proceedings 41st Annual Symposium on Foundations of Computer Science*, IEEE, pp. 454–463.

Emrani, S., Gentimis, T. & Krim, H., 2014, 'Persistent homology of delay embeddings and its application to wheeze detection', *IEEE Signal Processing Letters*, vol. 21, no. 4, pp. 459–463.

Erdos, P. & Rényi, A., 1959, 'On random graphs I', *Publicationes Mathematicae Debrecen*, vol. 6, no. 290-297, p. 18.

Eroglu, D., Marwan, N., Stebich, M. & Kurths, J., 2018, 'Multiplex recurrence networks', *Physical Review E*, vol. 97, no. 1, pp. 1–9.

Fawaz, H. I., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D. F., Weber, J., Webb, G. I., Idoumghar, L., Muller, P.-A. & Petitjean, F., 2020, 'Inceptiontime: Finding AlexNet for time series classification', *Data Mining and Knowledge Discovery*, vol. 34, no. 6, pp. 1936–1962.

Feldhoff, J. H., Donner, R. V., Donges, J. F., Marwan, N. & Kurths, J., 2012, 'Geometric detection of coupling directions by means of inter-system recurrence networks', *Physics Letters A*, vol. 376, no. 46, pp. 3504–3513.

Feldhoff, J. H., Donner, R. V., Donges, J. F., Marwan, N. & Kurths, J., 2013, 'Geometric signature of complex synchronisation scenarios', *EPL (Europhysics Letters)*, vol. 102, no. 3, pp. 1–8.

Figner, B. & Murphy, R. O., 2011, 'Using skin conductance in judgment and decision making research', *A handbook of process tracing methods for decision research*, pp. 163–184.

Fink, G., 2016, *Stress, definitions, mechanisms, and effects outlined: lessons from anxiety*, Stress: Concepts, cognition, emotion, and behavior, Elsevier, pp. 3–11.

Frosini, P., 1990, 'A distance for similarity classes of submanifolds of a euclidean space', *Bulletin of the Australian Mathematical Society*, vol. 42, no. 3, pp. 407–415.

Fugacci, U., Scaramuccia, S., Iuricich, F. & Floriani, L. D., 2016, 'Persistent homology: a step-by-step introduction for newcomers.', *Smart Tools and Apps in Computer Graphics*, pp. 1–10.

Gabrielsson, R. B., Nelson, B. J., Dwaraknath, A. & Skraba, P., 2020, 'A topology layer for machine learning', *International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, pp. 1553–1563.

Gao, Z., Dang, W., Wang, X., Hong, X., Hou, L., Ma, K. & Perc, M., 2021, 'Complex networks and deep learning for EEG signal analysis', *Cognitive Neurodynamics*, vol. 15, no. 3, pp. 369–388.

Gao, Z. K., Guo, W., Cai, Q., Ma, C., Zhang, Y. B. & Kurths, J., 2019a, 'Characterization of SSMVEP-based EEG signals using multiplex limited penetrable horizontal visibility graph', *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 29, no. 7, p. 073119.

Gao, Z. K., Li, Y.-L., Yang, Y.-X. & Ma, C., 2019b, 'A recurrence network-based convolutional neural network for fatigue driving detection from EEG', *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 29, no. 11, p. 113126.

García, V., Mollineda, R. A. & Sánchez, J. S., 2008, 'On the k-NN performance in a challenging scenario of imbalance and overlapping', *Pattern Analysis and Applications*, vol. 11, no. 3, pp. 269–280.

Gaskin, J., Jenkins, J., Meservy, T., Steffen, J. & Payne, K., 2017, 'Using wearable devices for non-invasive, inexpensive physiological data collection', *50th Hawaii International Conference on System Sciences*, pp. 597–605.

Gedam, S. & Paul, S., 2021, 'A review on mental stress detection using wearable sensors and machine learning techniques', *IEEE Access*, vol. 9, pp. 84045–84066.

Gharghabi, S., Ding, Y., Yeh, C.-C. M., Kamgar, K., Ulanova, L. & Keogh, E., 2017, 'Matrix profile VIII: domain agnostic online semantic segmentation at superhuman performance levels', *2017 IEEE International Conference on Data Mining (ICDM)*, IEEE, pp. 117–126.

Gidea, M., 2017, 'Topological data analysis of critical transitions in financial networks', *International conference and school on network science*, Springer, pp. 47–59.

Gidea, M. & Katz, Y., 2018, 'Topological data analysis of financial time series: Landscapes of crashes', *Physica A: Statistical Mechanics and its Applications*, vol. 491, pp. 820–834.

Gjoreski, M., Gjoreski, H., Luštrek, M. & Gams, M., 2016, 'Continuous stress detection using a wrist device: in laboratory and real life', *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pp. 1185–1193.

Gopali, S., Siami-Namini, S., Abri, F. & Namin, A. S., 2024, 'A comparative multivariate analysis of VAR and deep learning-based models for forecasting volatile time series data', *IEEE Access*.

Goshvarpour, A., Abbasi, A. & Goshvarpour, A., 2017, 'An accurate emotion recognition system using ECG and GSR signals and matching pursuit method', *Biomedical journal*, vol. 40, no. 6, p. 355–368.

Granados, L. S., Organero, M. M., Gonzalez, G. R., Abdulhay, E. & Arunkumar, N., 2018, 'Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS)', *IEEE Access*, vol. 7, pp. 57–67.

Greene, S., Thapliyal, H. & Caban-Holt, A., 2016, 'A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health', *IEEE Consumer Electronics Magazine*, vol. 5, no. 4, pp. 44–56.

Grossman, P., 1983, 'Respiration, stress, and cardiovascular function', *Psychophysiology*, vol. 20, no. 3, pp. 284–300.

Guo, C., Lu, M. & Chen, J., 2020, 'An evaluation of time series summary statistics as features for clinical prediction tasks', *BMC medical informatics and decision making*, vol. 20, no. 1, p. 48.

Gupta, D., Bhatia, M. & Kumar, A., 2021, 'Resolving data overload and latency issues in multivariate time-series iomt data for mental health monitoring', *IEEE Sensors Journal*, vol. 21, no. 22, pp. 25421–25428.

Hag, A., Handayani, D., Pillai, T., Mantoro, T., Kit, M. H. & Al-Shargie, F., 2021, 'EEG mental stress assessment using hybrid multi-domain feature sets of functional connectivity network and time-frequency features', *Sensors*, vol. 21, no. 18, p. 6300.

Haidri, S., 2023, 'Analytic: Understanding decision boundaries and dimensionality reduction in machine learning', *arXiv preprint arXiv:2401.05418*.

Han, Y., Zhang, H., Chen, Y., Liu, C., Zhang, Y., Zhang, Z., Jing, R., Guo, L., Li, D. & Chu, W., 2025, 'Small-world network and neuroscience', *Brain-X*, vol. 3, no. 1, p. e70025.

Haouij, N. E., 2018, *Biosignals for driver's stress level assessment: functional variable selection and fractal characterization*, Ph.D. thesis, Université Paris-Saclay (ComUE), Paris, France.

Hassan, M. M., Alam, M. G. R., Uddin, M. Z., Huda, S., Almogren, A. & Fortino, G., 2019, 'Human emotion recognition using deep belief network architecture', *Information Fusion*, vol. 51, pp. 10–18.

He, Q. & Yu, F., 2023, 'Trend recurrence analysis and time series classification via trend fuzzy granular recurrence plot method', *Chaos, Solitons & Fractals*, vol. 176, p. 114158.

Healey, J. A., 2000, 'Wearable and automotive systems for affect recognition from physiology', *MIT Libraries*.

Healey, J. A. & Picard, R. W., 2005, 'Detecting stress during real-world driving tasks using physiological sensors', *IEEE Transactions on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166.

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. & Gräler, B., 2018, 'Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables', *PeerJ*, vol. 6, p. e5518.

Hensel, F., Moor, M. & Rieck, B., 2021, 'A survey of topological machine learning methods', *Frontiers in Artificial Intelligence*, vol. 4, p. 681108.

Hernandez, J., Morris, R. R. & Picard, R. W., 2011, 'Call center stress recognition with person-specific models', *International Conference on Affective Computing and Intelligent Interaction*, Springer, pp. 125–134.

Hickey, B. A., Chalmers, T., Newton, P., Lin, C.-T., Sibbritt, D., McLachlan, C. S., Clifton-Bligh, R., Morley, J. & Lal, S., 2021, 'Smart devices and wearable technologies to detect and monitor mental health conditions and stress: A systematic review', *Sensors*, vol. 21, no. 10, p. 3461.

Hofer, C., Kwitt, R., Niethammer, M. & Uhl, A., 2017, 'Deep learning with topological signatures', *Advances in neural information processing systems*, vol. 30, pp. 1–11.

Hofer, C. D., Kwitt, R. & Niethammer, M., 2019, 'Learning representations of persistence barcodes', *Journal of Machine Learning Research*, vol. 20, no. 126, pp. 1–45.

Hong, J.-H., Ramos, J. & Dey, A. K., 2012, 'Understanding physiological responses to stressors during physical activity', *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pp. 270–279.

Hosseini, S. H. & Kesler, S. R., 2013, 'Comparing connectivity pattern and small-world organization between structural correlation and resting-state networks in healthy adults', *NeuroImage*, vol. 78, pp. 402–414.

Hovsepian, K., Al'Absi, M., Ertin, E., Kamarck, T., Nakajima, M. & Kumar, S., 2015, 'cStress: Towards a gold standard for continuous stress assessment in the mobile environment', *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pp. 493–504.

Huynh, L., Nguyen, T., Nguyen, T., Pirttikangas, S. & Siirtola, P., 2021, 'Stress-nas: Affect state and stress detection using neural architecture search', *ACM Int.Jt.Conf. on Pervasive and Ubiquitous Computing*, pp. 121–125.

Hyndman, R. J. & Athanasopoulos, G., 2018, *Forecasting: principles and practice*, OTexts, Australia.

Iacobello, G., Ridolfi, L. & Scarsoglio, S., 2021, 'A review on turbulent and vortical flow analyses via complex networks', *Physica A: Statistical Mechanics and its Applications*, vol. 563, p. 125476.

Inibhunu, C., 2020, *A method to detect and represent temporal patterns from time series data and its application for analysis of physiological data streams*, Ph.D. thesis, University of Ontario Institute of Technology, Ontario, Canada.

Inibhunu, C., Yeung, J., Gates, A., Chicoine, B. & McGregor, C., 2021, 'A decoupled data pipeline and its reliability assessment: Case study in extreme climatic humans studies', *2021 IEEE International Conference on Big Data (Big Data)*, IEEE, pp. 3075–3084.

Iwanski, J. S. & Bradley, E., 1998, 'Recurrence plots of experimental data: To embed or not to embed?', *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 8, no. 4, pp. 861–871.

James, G., Witten, D., Hastie, T. & Tibshirani, R., 2013, *An Introduction to Statistical Learning*, vol. 112, Springer.

Jiang, M., Gao, X., An, H., Li, H. & Sun, B., 2017, 'Reconstructing complex network for characterizing the time-varying causality evolution behavior of multivariate time series', *Scientific reports*, vol. 7, no. 1, pp. 1–12.

Jiménez-Mijangos, L. P., Rodríguez-Arce, J., Martínez-Méndez, R. & Reyes-Lagos, J. J., 2023, 'Advances and challenges in the detection of academic stress and anxiety in the classroom: A literature review and recommendations', *Education and Information Technologies*, vol. 28, no. 4, pp. 3637–3666.

Kanjo, E., Younis, E. M. & Ang, C. S., 2019, 'Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection', *Information Fusion*, vol. 49, pp. 46–56.

Karan, A. & Kaygun, A., 2021, 'Time series classification via topological data analysis', *Expert Systems with Applications*, pp. 1–14.

Karim, F., Majumdar, S., Darabi, H. & Harford, S., 2019, 'Multivariate LSTM-FCNs for time series classification', *Neural Networks*, vol. 116, pp. 237–245.

Kazanskiy, N. L., Khonina, S. N. & Butt, M. A., 2024, 'A review on flexible wearables-recent developments in non-invasive continuous health monitoring', *Sensors and Actuators A: Physical*, p. 114993.

Kemme, A. J. & Agyingi, C. A., 2025, 'Persistent homology: A pedagogical introduction with biological applications', *arXiv preprint arXiv:2505.06583*.

Kemper, K. J., Hamilton, C. & Atkinson, M., 2007, 'Heart rate variability: impact of differences in outlier identification and management strategies on common measures in three clinical populations', *Pediatric research*, vol. 62, no. 3, pp. 337–342.

Keshan, N., Parimi, P. V. & Bichindaritz, I., 2015, 'Machine learning for stress detection from ECG signals in automobile drivers', *2015 IEEE International conference on big data (Big Data)*, IEEE, p. 2661–2669.

Kestel, D., Lewis, S., Freeman, M., Chisholm, D., Siegl, O. G. & van Ommeren, M., 2022, 'A world report on the transformation needed in mental health care', *Bulletin of the World Health Organization*, vol. 100, no. 10, p. 583, pmid:36188024.

Khusainov, R., Azzi, D., Achumba, I. E. & Bersch, S. D., 2013, 'Real-time human ambulation, activity, and physiological monitoring: Taxonomy of issues, techniques, applications, challenges and limitations', *Sensors*, vol. 13, no. 10, pp. 12852–12902.

Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H. & Koo, B. H., 2018, 'Stress and heart rate variability: A meta-analysis and review of the literature', *Psychiatry investigation*, vol. 15, no. 3, pp. 235–245.

Kim, K., Kim, J., Zaheer, M., Kim, J., Chazal, F. & Wasserman, L., 2020, 'PLLay: Efficient topological layer based on persistence landscapes', *NeurIPS 2020-34th Conference on Neural Information Processing Systems*, .

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M. & Inman, D. J., 2021, '1d convolutional neural networks and applications: A survey', *Mechanical systems and signal processing*, vol. 151, p. 107398.

Kirschbaum, C., Pirke, K. M. & Hellhammer, D. H., 1993, 'The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting', *Neuropsychobiology*, vol. 28, no. 1-2, pp. 76–81.

Kleiger, R. E., Stein, P. K. & Jr, J. T. B., 2005, 'Heart rate variability: measurement and clinical utility', *Annals of Noninvasive Electrocardiology*, vol. 10, no. 1, pp. 88–101.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A. & Patras, I., 2011, 'DEAP: A database for emotion analysis; using physiological signals', *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31.

Koldijk, S., Neerincx, M. A. & Kraaij, W., 2016, 'Detecting work stress in offices by combining unobtrusive sensors', *IEEE Transactions on affective computing*, vol. 9, no. 2, pp. 227–239.

Koldijk, S., Sappelli, M., Verberne, S., Neerincx, M. A. & Kraaij, W., 2014, 'The SWELL knowledge work dataset for stress and user modeling research', *Proceedings of the 16th international conference on multimodal interaction*, pp. 291–298.

Kong, Y., Wang, Z., Nie, Y., Zhou, T., Zohren, S., Liang, Y., Sun, P. & Wen, Q., 2025, 'Unlocking the power of lstm for long term time series forecasting', *Proceedings of the AAAI Conference on Artificial Intelligence*, , vol. 39pp. 11968–11976.

Konstantinou, P., Trigeorgi, A., Georgiou, C., Gloster, A. T., Panayiotou, G. & Karekla, M., 2020, 'Comparing apples and oranges or different types of citrus fruits? using wearable versus stationary devices to analyze psychophysiological data', *Psychophysiology*, vol. 57, no. 5, p. e13551.

Kwon, O., Jeong, J., Kim, H. B., Kwon, I. H., Park, S. Y., Kim, J. E. & Choi, Y., 2018, 'Electrocardiogram sampling frequency range acceptable for heart rate variability analysis', *Healthcare informatics research*, vol. 24, no. 3, pp. 198–206.

Lacasa, L., Luque, B., Ballesteros, F., Luque, J. & Nuno, J. C., 2008, 'From time series to complex networks: The visibility graph', *Proceedings of the National Academy of Sciences*, vol. 105, no. 13, pp. 4972–4975.

Lacasa, L., Nicosia, V. & Latora, V., 2015, 'Network structure of multivariate time series', *Scientific reports*, vol. 5, no. 1, pp. 1–9.

Lan, X., Mo, H., Chen, S., Liu, Q. & Deng, Y., 2015, 'Fast transformation from time series to visibility graphs', *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 8.

Lanius, M., Zheng, J. & Abebe, A., 2024, 'Exploring undergraduate students' psychological stress in the mathematics classroom with fitness trackers', *Investigations in Mathematics Learning*, vol. 16, no. 1, p. 18–35.

Latora, V. & Marchiori, M., 2002, 'Is the boston subway a small-world network?', *Physica a: statistical mechanics and its applications*, vol. 314, no. 1-4, pp. 109–113.

LeCun, Y., Bengio, Y. & Hinton, G., 2015, 'Deep learning', *Nature*, vol. 521, no. 7553, p. 436–444.

Ledoyt, T., 2019, 'Breathing and the nervous system', <https://www.onthegophysicaltherapy.com/blog/2019/10/12/breathing-part-2>.

Lee, D. S., Chong, T. W. & Lee, B. G., 2016, 'Stress events detection of driver by wearable glove system', *IEEE Sensors Journal*, vol. 17, no. 1, pp. 194–204.

Li, H., Jia, R. & Wan, X., 2022, 'Time series classification based on complex network', *Expert Systems with Applications*, vol. 194, p. 116502.

Li, R. & Liu, Z., 2020, 'Stress detection using deep neural networks', *BMC Medical Informatics and Decision Making*, vol. 20, no. 11, pp. 1–10.

Li, Y., Liang, X. & Xie, J., 2019, 'Constructing an effective model for mental stress detection with small–scale analysis', *2019 IEEE 3rd Information Technology,* *Networking, Electronic and Automation Control Conference (ITNEC)*, IEEE, p. 310–314.

Lisowska, A., Wilk, S. & Peleg, M., 2021, ‘Catching patient’s attention at the right time to help them undergo behavioural change: Stress classification experiment from blood volume pulse’, *19th Int.Conf. on Artificial Intelligence in Medicine, AIME 2021, June 15–18*, Springer, pp. 72–82.

Liu, R. & Gillies, D. F., 2016, ‘Overfitting in linear feature extraction for classification of high-dimensional image data’, *Pattern Recognition*, vol. 53, pp. 73–86.

Loscalzo, J., 2012, ‘Personalized cardiovascular medicine and drug development: time for a new paradigm’, *Circulation*, vol. 125, no. 4, pp. 638–645.

Lowery, D., Fillingim, R. B. & Wright, R. A., 2003, ‘Sex differences and incentive effects on perceptual and cardiovascular responses to cold pressor pain’, *Psychosomatic medicine*, vol. 65, no. 2, pp. 284–291.

Lum, P. Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J. & Carlsson, G., 2013, ‘Extracting insights from the shape of complex data using topology’, *Nature Scientific reports*, vol. 3, no. 1, pp. 1–8.

Längkvist, M., Karlsson, L. & Loutfi, A., 2014, ‘A review of unsupervised feature learning and deep learning for time-series modeling’, *Pattern Recognition Letters*, vol. 42, pp. 11–24.

Ma, Q., Liu, Z., Zheng, Z., Huang, Z., Zhu, S., Yu, Z. & Kwok, J. T., 2023, ‘A survey on time-series pre-trained models’, *arXiv preprint arXiv:2305.10716*.

Maaoui, C., Pruski, A. & Abdat, F., 2008, ‘Emotion recognition for human-machine communication’, *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 1210–1215.

MacQuarrie, A., Sidhu, J., Deetlefs, C., Whitfield, S. & Stainer, M., 2024, ‘Establishing the validity and reliability of the astroskin® biometric shirt’, *Topics in Exercise Science and Kinesiology*, vol. 5, no. 1, p. 4.

Majumdar, S. & Laha, A. K., 2020, ‘Clustering and classification of time series using topological data analysis with applications to finance’, *Expert Systems with Applications*, vol. 162, p. 113868.

Makridakis, S., Spiliotis, E. & Assimakopoulos, V., 2018a, ‘The M4 competition: Results, findings, conclusion and way forward’, *International Journal of Forecasting*, vol. 34, no. 4, pp. 802–808.

Makridakis, S., Spiliotis, E. & Assimakopoulos, V., 2018b, ‘Statistical and machine learning forecasting methods: Concerns and ways forward’, *PloS one*, vol. 13, no. 3, p. e0194889.

Mandal, S., Ghosh, B. & Naskar, R., 2023, ‘A photoplethysmography (PPG) sensor based stress level monitoring system for undergraduate students of technical education’, *2023 IEEE 20th India Council International Conference (INDICON)*, IEEE, p. 197–202.

Marchese, A. & Maroulas, V., 2018, ‘Signal classification with a point process distance on the space of persistence diagrams’, *Advances in Data Analysis and Classification*, vol. 12, pp. 657–682.

Marwan, N., Donges, J. F., Zou, Y., Donner, R. V. & Kurths, J., 2009, ‘Complex network approach for recurrence analysis of time series’, *Physics Letters A*, vol. 373, no. 46, pp. 4246–4254.

Mason, J. W., 1975, ‘A historical view of the stress field’, *Journal of human stress*, vol. 1, no. 2, pp. 22–36.

McCraty, R. & Shaffer, F., 2015, ‘Heart rate variability: new perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk’, *Global advances in health and medicine*, vol. 4, no. 1, pp. 46–61.

McCullough, M., Small, M., Stemler, T. & Iu, H. H.-C., 2015, ‘Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems’, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 5, pp. 1–12.

McGregor, C., 2021, ‘A platform for real-time space health analytics as a service utilizing space data relays’, *2021 IEEE Aerospace Conference (50100)*, IEEE, pp. 1–14.

Mentis, A.-F. A., Lee, D. & Roussos, P., 2023, 'Applications of artificial intelligence machine learning for detection of stress: a critical overview', *Molecular psychiatry*, pp. 1–13.

Mishra, V., Sen, S., Chen, G., Hao, T., Rogers, J., Chen, C.-H. & Kotz, D., 2020, 'Evaluating the reproducibility of physiological stress detection models', *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 4, pp. 1–29.

Miyakoshi, T. & Ito, Y. M., 2024, 'Assessing the current utilization status of wearable devices in clinical research', *Clinical Trials*, p. 17407745241230287.

Mohammed, O. D. & Rantatalo, M., 2020, 'Gear fault models and dynamics-based modelling for gear fault detection–a review', *Engineering Failure Analysis*, vol. 117, p. 104798.

Moor, M., Horn, M., Rieck, B. & Borgwardt, K., 2020, 'Topological autoencoders', *International conference on machine learning*, PMLR, pp. 7045–7054.

Moskowitz, D. S. & Young, S. N., 2006, 'Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology', *Journal of psychiatry & neuroscience : JPN*, vol. 31, no. 1, pp. 13–20.

Muaremi, A., Bexheti, A., Gravenhorst, F., Arnrich, B. & Tröster, G., 2014, 'Monitoring the impact of stress on the sleep patterns of pilgrims using wearable sensors', *IEEE-EMBS international conference on biomedical and health informatics (BHI)*, IEEE, pp. 185–188.

Munch, E., 2017, 'A user's guide to topological data analysis', *Journal of Learning Analytics*, vol. 4, no. 2, p. 47–61.

Musa, S. M. S. S., Noorani, M. S. M., Razak, F. A., Ismail, M., Alias, M. A. & Hussain, S. I., 2021, 'Using persistent homology as preprocessing of early warning signals for critical transition in flood', *Nature Scientific Reports*, vol. 11, no. 1, pp. 1–14.

Myers, A., Munch, E. & Khasawneh, F. A., 2019, 'Persistent homology of complex networks for dynamic state detection', *Physical Review E*, vol. 100, no. 2, pp. 1–15.

Namini, S. S., Tavakoli, N. & Namin, A. S., 2018, 'A comparison of ARIMA and LSTM in forecasting time series', *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, pp. 1394–1401.

Neligan, P. J. & Baranov, D., 2013, 'Trauma and aggressive homeostasis management', *Anesthesiology clinics*, vol. 31, no. 1, pp. 21–39, pmid:23351532.

Oates, T., Firoiu, L. & Cohen, P. R., 1999, 'Clustering time series with hidden markov models and dynamic time warping', *Proceedings of the IJCAI-99 workshop on neural, symbolic and reinforcement learning methods for sequence learning*, , vol. 17,Citeseer, p. 21.

Ollander, S., Godin, C., Campagne, A. & Charbonnier, S., 2016, 'A comparison of wearable and stationary sensors for stress detection', *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, pp. 4362–4366.

Oreshkin, B. N., Carpov, D., Chapados, N. & Bengio, Y., 2019, 'N-BEATS: Neural basis expansion analysis for interpretable time series forecasting', *arXiv preprint arXiv:1905.10437*.

Panicker, S. S. & Gayathri, P., 2019, 'A survey of machine learning techniques in physiology based mental stress detection systems', *Biocybernetics and Biomedical Engineering*, vol. 39, no. 2, pp. 444–469.

Park, J. & Barabási, A.-L., 2007, 'Distribution of node characteristics in complex networks', *Proceedings of the National Academy of Sciences*, vol. 104, no. 46, pp. 17916–17920.

Pavlopoulos, G. A., Secrier, M., Moschopoulos, C. N., Soldatos, T. G., Kossida, S., Aerts, J., Schneider, R. & Bagos, P. G., 2011, 'Using graph theory to analyze biological networks', *BioData mining*, vol. 4, pp. 1–27.

Peper, E., Harvey, R., Lin, I.-M., Tylova, H. & Moss, D., 2007, 'Is there more to blood volume pulse than heart rate variability, respiratory sinus arrhythmia, and cardiorespiratory synchrony?', *Biofeedback*, vol. 35, no. 2.

Pereira, C. M. & de Mello, R. F., 2015, 'Persistent homology for time series and spatial data clustering', *Expert Systems with Applications*, vol. 42, no. 15-16, pp. 6026–6038.

Pereira, T. M., Conceição, R. C., Sencadas, V. & Sebastião, R., 2023, 'Biometric recognition: A systematic review on electrocardiogram data acquisition methods', *Sensors*, vol. 23, no. 3, p. 1507.

Perera, S., Bell, M. G. & Bliemer, M. C., 2017, 'Network science approach to modelling the topology and robustness of supply chain networks: a review and perspective', *Applied network science*, vol. 2, pp. 1–25.

Peternel, K., Pogačnik, M., Tavčar, R. & Kos, A., 2012, 'A presence-based context-aware chronic stress recognition system', *Sensors*, vol. 12, no. 11, pp. 15888–15906.

Phutela, N., Relan, D., Gabrani, G., Kumaraguru, P. & Samuel, M., 2022, 'Stress classification using brain signals based on lstm network', *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, p. 7607592.

Plarre, K., Raij, A., Hossain, S. M., Ali, A. A., Nakajima, M., Al'Absi, M., Ertin, E., Kamarck, T., Kumar, S. & Scott, M., 2011, 'Continuous inference of psychological stress from sensory measurements collected in the natural environment', *Proceedings of the 10th ACM/IEEE international conference on information processing in sensor networks*, IEEE, pp. 97–108.

Postolache, O., Girão, P. S., Pinheiro, E. & Postolache, G., 2010, *Unobtrusive and non-invasive sensing solutions for on-line physiological parameters monitoring*, Wearable and Autonomous Biomedical Devices and Systems for Smart Environment, Springer, pp. 277–314.

Pun, C. S., Lee, S. X. & Xia, K., 2022, 'Persistent-homology-based machine learning: a survey and a comparative study', *Artificial Intelligence Review*, vol. 55, no. 7, p. 5169–5213.

Pyakillya, B., Kazachenko, N. & Mikhailovsky, N., 2017, 'Deep learning for ECG classification', *Journal of physics: conference series*, , vol. 913IOP Publishing, Moscow, Russian Federation, pp. 1–6.

Qi, C. R., Su, H., Mo, K. & Guibas, L. J., 2017, 'Pointnet: Deep learning on point sets for 3d classification and segmentation', *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660.

Quadrini, M., Daberdaku, S., Blanda, A., Capuccio, A., Bellanova, L. & Gerard, G., 2022, 'Stress detection from wearable sensor data using Gramian Angular Fields and CNN', *International Conference on Discovery Science*, Springer, pp. 173–183.

Quazi, M., Mukhopadhyay, S., Suryadevara, N. & Huang, Y.-M., 2012, 'Towards the smart sensors based human emotion recognition', *2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, IEEE, pp. 2365–2370.

Ramos, J., Hong, J.-H. & Dey, A. K., 2014, 'Stress recognition: A step outside the lab.', *PhyCS*, pp. 107–118.

Ramírez, M. L. G., Vázquez, J. P. G., Rodríguez, M. D., Padilla-López, L. A., Galindo-Aldana, G. M. & Cuevas-González, D., 2023, 'Wearables for stress management: A scoping review', *Healthcare*, , vol. 11MDPI, p. 2369.

Rasheed, N., 2016, 'Prolonged stress leads to serious health problems: Preventive approaches', *International journal of health sciences*, vol. 10, no. 1, p. V, pmid:27004066.

Rastgoo, M. N., Nakisa, B., Maire, F., Rakotonirainy, A. & Chandran, V., 2019, 'Automatic driver stress level classification using multimodal deep learning', *Expert Systems with Applications*, vol. 138, p. 112793.

Ray, A., Goyal, V. & Bandyopadhyay, S. K., 2023, *Student Stress Detection in Online Learning During Outbreak*, Computational Methods in Psychiatry, Springer, p. 259–281.

Ray, S., 2019, 'A quick review of machine learning algorithms', *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*, IEEE, pp. 35–39.

Reinauer, R., Caorsi, M. & Berkouk, N., 2021, 'Persformer: A transformer architecture for topological machine learning', *arXiv preprint arXiv:2112.15210*.

Rieck, B. A., Togninalli, M., Bock, C., Moor, M., Horn, M., Gumbsch, T. & Borgwardt, K., 2019, 'Neural persistence: A complexity measure for deep neural networks using algebraic topology', *International Conference on Learning Representations (ICLR 2019)*, OpenReview.

Riondato, M. & Kornaropoulos, E. M., 2014, 'Fast approximation of betweenness centrality through sampling', *Proceedings of the 7th ACM international conference on Web search and data mining*, pp. 413–422.

Rizwan, M. F., Farhad, R., Mashuk, F., Islam, F. & Imam, M. H., 2019, 'Design of a biosignal based stress detection system using machine learning techniques', *2019 international conference on robotics, electrical and signal processing techniques (ICREST)*, IEEE, p. 364–368.

Rochette, L. & Vergely, C., 2017, 'Hans Selye and the stress response: 80 years after his "letter" to the editor of nature', *Annales de Cardiologie et d'Angeiologie*, vol. 66, no. 4, pp. 181–183.

Rosales, F., Garcia-Dopico, A., Bajo, R. & Nevado, A., 2015, 'An efficient implementation of the synchronization likelihood algorithm for functional connectivity', *Neuroinformatics*, vol. 13, pp. 245–258.

Rosenfield, D., Zhou, E., Wilhelm, F. H., Conrad, A., Roth, W. T. & Meuret, A. E., 2010, 'Change point analysis for longitudinal physiological data: Detection of cardio-respiratory changes preceding panic attacks', *Biological psychology*, vol. 84, no. 1, pp. 112–120.

Ruan, Y., Donner, R. V., Guan, S. & Zou, Y., 2019, 'Ordinal partition transition network based complexity measures for inferring coupling direction and delay from time series', *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 29, no. 4, pp. 1–14.

Rubinov, M. & Sporns, O., 2010, 'Complex network measures of brain connectivity: uses and interpretations', *NeuroImage*, vol. 52, no. 3, pp. 1059–1069.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. & Sejdinovic, D., 2019, 'Detecting and quantifying causal associations in large nonlinear time series datasets', *Science advances*, vol. 5, no. 11, pp. 1–16.

Runge, J., Petoukhov, V., Donges, J. F., Hlinka, J., Jajcay, N., Vejmelka, M., Hartman, D., Marwan, N., Paluš, M. & Kurths, J., 2015, 'Identifying causal gateways and mediators in complex spatio-temporal systems', *Nature communications*, vol. 6, no. 1, pp. 1–10.

Sanderson, N., Shugerman, E., Molnar, S., Meiss, J. D. & Bradley, E., 2017, 'Computational topology techniques for characterizing time-series data', *Advances in Intelligent Data Analysis XVI: 16th International Symposium, IDA 2017, London, UK, October 26–28, 2017, Proceedings 16*, Springer, pp. 284–296.

Sandulescu, V., Andrews, S., Ellis, D., Bellotto, N. & Mozos, O. M., 2015, 'Stress detection using wearable physiological sensors', *International work-conference on the interplay between natural and artificial computation*, Springer, pp. 526–532.

Sannino, S., Stramaglia, S., Lacasa, L. & Marinazzo, D., 2017, 'Visibility graphs for fMRI data: Multiplex temporal graphs and their modulations across resting-state networks', *Network Neuroscience*, vol. 1, no. 3, pp. 208–221.

Sano, A. & Picard, R. W., 2013, 'Stress recognition using wearable sensors and mobile phones', *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, IEEE, pp. 671–676.

Sano, A., Taylor, S., McHill, A. W., Phillips, A. J., Barger, L. K., Klerman, E. & Picard, R., 2018, 'Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study', *Journal of Medical Internet research*, vol. 20, no. 6, p. e210.

Saramäki, J., Kivelä, M., Onnela, J.-P., Kaski, K. & Kertesz, J., 2007, 'Generalizations of the clustering coefficient to weighted complex networks', *Physical Review E*, vol. 75, no. 2, p. 027105.

Savle, K., Zadrozny, W. & Lee, M., 2019, 'Topological data analysis for discourse semantics?', *Proceedings of the 13th International Conference on Computational Semantics-Student Papers*, pp. 34–43.

Scarciglia, A., Catrambone, V., Bonanno, C. & Valenza, G., 2023, 'Physiological noise: Definition, estimation, and characterization in complex biomedical signals', *IEEE Transactions on Biomedical Engineering*.

Scarpina, F. & Tagini, S., 2017, 'The Stroop Color and Word Test', *Frontiers in psychology*, vol. 8, p. 557.

Schmidt, P., Reiss, A., Duerichen, R. & Laerhoven, K. V., 2018a, 'Wearable affect and stress recognition: A review', *arXiv preprint arXiv:1811.08854*.

Schmidt, P., Reiss, A., Duerichen, R., Marberger, C. & Laerhoven, K. V., 2018b, 'Introducing WESAD, a multimodal dataset for wearable stress and affect detection', *20th ACM Int.Conf. on Multimodal Interaction*, pp. 400–408.

Schober, P. & Vetter, T. R., 2021, 'Logistic regression in medical research', *Anesthesia & Analgesia*, vol. 132, no. 2, p. 365–366, pmid:33449558.

Schuler, R. S., 1980, 'Definition and conceptualization of stress in organizations', *Organizational behavior and human performance*, vol. 25, no. 2, pp. 184–215.

Schäfer, P., Ermshaus, A. & Leser, U., 2021, 'Clasp-time series segmentation', *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, pp. 1578–1587.

Selye, H., 1936, 'A syndrome produced by diverse nocuous agents', *Nature*, vol. 138, no. 3479, pp. 32–32.

Selye, H., 1950, 'Stress and the general adaptation syndrome', *British medical journal*, vol. 1, no. 4667, pp. 1383–1392.

Selye, H., 1978, *The stress of life, Rev*, McGraw Hill.

Seo, S.-H., Lee, J.-T. & Crisan, M., 2010, 'Stress and EEG', *Convergence and hybrid information technologies*, vol. 27, pp. 413–424.

Seversky, L. M., Davis, S. & Berger, M., 2016, 'On time-series topological data analysis: New data and opportunities', *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 59–67.

Shaffer, F. & Ginsberg, J. P., 2017, 'An overview of heart rate variability metrics and norms', *Frontiers in public health*, vol. 5, p. 258.

Sierra, A. D. S., Ávila, C. S., Casanova, J. G. & del Pozo, G. B., 2011, 'A stress-detection system based on physiological signals and fuzzy logic', *IEEE Transactions on Industrial Electronics*, vol. 58, no. 10, pp. 4857–4865.

Sierra, A. D. S., Ávila, C. S., Casanova, J. G., del Pozo, G. B. & Vera, V. J., 2010, 'Two stress detection schemes based on physiological signals for real-time applications', *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, IEEE, pp. 364–367.

Silva, T. C. & Zhao, L., 2012, 'Network-based high level data classification', *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 6, pp. 954–970.

Silva, V. F., Silva, M. E., Ribeiro, P. & Silva, F., 2021, 'Time series analysis via network science: Concepts and algorithms', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 11, no. 3, pp. 1–44.

Silva, V. F., Silva, M. E., Ribeiro, P. & Silva, F., 2022, 'Novel features for time series analysis: a complex networks approach', *Data Mining and Knowledge Discovery*, vol. 36, no. 3, pp. 1062–1101.

Skaf, Y. & Laubenbacher, R., 2022, 'Topological data analysis in biomedicine: A review', *Journal of Biomedical Informatics*, vol. 130, p. 104082.

Small, M., 2013, 'Complex networks from time series: Capturing dynamics', *IEEE Int.Symp. on Circuits and Systems (ISCAS)*, IEEE, pp. 2509–2512.

Song, T., Zheng, W., Song, P. & Cui, Z., 2018, 'EEG emotion recognition using dynamical graph convolutional neural networks', *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532–541.

Sousa, J. & Barata, J., 2021, *Mining Sociotechnical Patterns of Enterprise Systems with Complex Networks: A Guiding Framework*, Handbook of Research on Autopoiesis and Self-Sustaining Processes for Organizational Success, IGI Global, pp. 38–57.

Sporns, O., Chialvo, D. R., Kaiser, M. & Hilgetag, C. C., 2004, 'Organization, development and function of complex brain networks', *Trends in cognitive sciences*, vol. 8, no. 9, pp. 418–425, pmid:15350243.

Stephen, M., Gu, C. & Yang, H., 2015, 'Visibility graph based time series analysis', *PloS one*, vol. 10, no. 11, p. e0143015.

Stoltzfus, J. C., 2011, 'Logistic regression: a brief primer', *Academic Emergency Medicine*, vol. 18, no. 10, p. 1099–1104.

Strogatz, S. H., 2001, 'Exploring complex networks', *Nature*, vol. 410, no. 6825, pp. 268–276.

Stroop, J. R., 1935, 'Studies of interference in serial verbal reactions.', *Journal of experimental psychology*, vol. 18, no. 6, p. 643.

Subramaniyam, N. P. & Hyttinen, J., 2015, 'Dynamics of intracranial electroencephalographic recordings from epilepsy patients using univariate and bivariate recurrence networks', *Physical Review E*, vol. 91, no. 2, pp. 1–18.

Suhara, Y., Xu, Y. & Pentland, A., 2017, 'Deepmood: Forecasting depressed mood based on self-reported histories via recurrent neural networks', *Proceedings of the 26th International Conference on World Wide Web*, pp. 715–724.

Supriya, S., Siuly, S., Wang, H., Cao, J. & Zhang, Y., 2016, 'Weighted visibility graph with complex network features in the detection of epilepsy', *IEEE Access*, vol. 4, pp. 6554–6566.

Takens, F., 1981, 'Detecting strange attractors in turbulence', *Dynamical Systems and Turbulence, Warwick 1980: proceedings of a symposium held at the University of Warwick 1979/80*, Springer, pp. 366–381.

Tang, J., Alelyani, S. & Liu, H., 2014, 'Feature selection for classification: A review', *CRC Press Data classification: Algorithms and applications*, pp. 37–64.

Tang, R., Zeng, F., Chen, Z., Wang, J.-S., Huang, C.-M. & Wu, Z., 2020, 'The comparison of predicting storm-time ionospheric TEC by three methods: ARIMA, LSTM, and Seq2Seq', *Atmosphere*, vol. 11, no. 4, p. 316.

Tauzin, G., Lupo, U., Tunstall, L., Pérez, J. B., Caorsi, M., Medina-Mardones, A. M., Dassatti, A. & Hess, K., 2021, 'giotto-tda:: A topological data analysis toolkit for machine learning and data exploration', *Journal of Machine Learning Research*, vol. 22, no. 39, p. 1–6.

Taylor, S., Jaques, N., Nosakhare, E., Sano, A. & Picard, R., 2017, 'Personalized multitask learning for predicting tomorrow's mood, stress, and health', *IEEE Transactions on Affective Computing*, vol. 11, no. 2, pp. 200–213.

Temur, A. S., Akgün, M. & Temur, G., 2019, 'Predicting housing sales in Turkey using ARIMA, LSTM and hybrid models', *Journal of Business Economics and Management*, vol. 20, no. 5, pp. 920–938.

Tian, Y., 2022, 'Identification and modeling of college students' psychological stress indicators for deep learning', *Scientific Programming*, vol. 2022, no. 1, p. 6048088.

Tiwari, S., Chandra, R. & Agarwal, S., 2023, 'An optimized hybrid solution for IoT based lifestyle disease classification using stress data', *29th Int.Conf. on Neural Information Processing, ICONIP, November 22–26*, Springer, pp. 433–445.

Tran, V. H., Cheong, S. A. & Bui, N. D., 2019, 'Complex network analysis of the robustness of the hanoi, vietnam bus network', *Journal of Systems Science and Complexity*, vol. 32, pp. 1251–1263.

Triguero, I., Maillo, J., Luengo, J., García, S. & Herrera, F., 2016, 'From big data to smart data with the k-nearest neighbours algorithm', *2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, IEEE, p. 859–864.

Trirat, P., Shin, Y., Kang, J., Nam, Y., Na, J., Bae, M., Kim, J., Kim, B. & Lee, J.-G., 2024, 'Universal time-series representation learning: A survey', *arXiv preprint arXiv:2401.03717*.

Tu, J. V., 1996, 'Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes', *Journal of clinical epidemiology*, vol. 49, no. 11, pp. 1225–1231.

Tymochko, S., Singhal, K. & Heo, G., 2021, 'Classifying sleep states using persistent homology and Markov chains: A pilot study', *Advances in Data Science*, pp. 253–289.

Umematsu, T., Sano, A. & Picard, R. W., 2019a, 'Daytime data and LSTM can forecast tomorrow's stress, health, and happiness', *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, pp. 2186–2190.

Umematsu, T., Sano, A., Taylor, S. & Picard, R. W., 2019b, 'Improving students' daily life stress forecasting using LSTM neural networks', *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, IEEE, pp. 1–4.

Umematsu, T., Sano, A., Taylor, S., Tsujikawa, M. & Picard, R. W., 2020, 'Forecasting stress, mood, and health from daytime physiology in office workers and students', *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, pp. 5953–5957.

van den Burg, G. J. & Williams, C. K., 2020, 'An evaluation of change point detection algorithms', *arXiv preprint arXiv:2003.06222*.

Vila, G., Godin, C., Charbonnier, S., Labyt, E., Sakri, O. & Campagne, A., 2018, 'Pressure-specific feature selection for acute stress detection from physiological recordings', *2018 IEEE international conference on systems, man, and cybernetics (SMC)*, IEEE, p. 2341–2346.

Vilca, E. & Zhao, L., 2020, 'A network-based high-level data classification algorithm using betweenness centrality', *arXiv preprint arXiv:2009.07971*.

Vildjiounaite, E., Kallio, J., Kyllönen, V., Nieminen, M., Määttänen, I., Lindholm, M., Mäntyjärvi, J. & Gimel'farb, G., 2018, 'Unobtrusive stress detection on the basis of smartphone usage data', *Personal and Ubiquitous Computing*, vol. 22, no. 4, pp. 671–688.

Wagstaff, E., Fuchs, F., Engelcke, M., Posner, I. & Osborne, M. A., 2019, 'On the limitations of representing functions on sets', *International Conference on Machine Learning*, PMLR, pp. 6487–6494.

Wang, H., Chang, W. & Zhang, C., 2016, 'Functional brain network and multichannel analysis for the p300-based brain computer interface system of lying detection', *Expert Systems with Applications*, vol. 53, pp. 117–128.

Wang, J.-S., Lin, C.-W. & Yang, Y.-T. C., 2013, 'A k-nearest-neighbor classifier with heart rate variability feature-based transformation algorithm for driving stress recognition', *Neurocomputing*, vol. 116, pp. 136–143.

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D. & Campbell, A. T., 2014, 'StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones', *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, pp. 3–14.

Wang, Z., Yan, W. & Oates, T., 2017, 'Time series classification from scratch with deep neural networks: A strong baseline', *Int.Jt.Conf. on Neural Networks (IJCNN)*, IEEE, pp. 1578–1585.

Waring, J., Lindvall, C. & Umeton, R., 2020, 'Automated machine learning: Review of the state-of-the-art and opportunities for healthcare', *Artificial Intelligence in Medicine*, vol. 104, p. 101822.

Watts, D. J. & Strogatz, S. H., 1998, 'Collective dynamics of 'small-world' networks', *Nature*, vol. 393, no. 6684, pp. 440–442.

Weber, E. J., Molenaar, P. C. & Molen, M. W. V. D., 1992, 'A nonstationarity test for the spectral analysis of physiological time series with an application to respiratory sinus arrhythmia', *Psychophysiology*, vol. 29, no. 1, pp. 55–62.

Weytjens, H., Lohmann, E. & Kleinsteuber, M., 2021, 'Cash flow prediction: MLP and LSTM compared to ARIMA and Prophet', *Electronic Commerce Research*, vol. 21, no. 2, pp. 371–391.

Wijsman, J., Grundlehner, B., Penders, J. & Hermens, H., 2013, 'Trapezius muscle EMG as predictor of mental stress', *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 12, no. 4, pp. 1–20.

Wolpert, D. H., 1996, 'The lack of a priori distinctions between learning algorithms', *Neural computation*, vol. 8, no. 7, pp. 1341–1390.

Woodward, K., Kanjo, E., Brown, D., McGinnity, T. M., Inkster, B., MacIntyre, D. & Tsanas, T., 2020, 'Beyond mobile apps: a survey of technologies for mental well-being', *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1216–1235.

World Health Organization, 2021, 'Investing in mental health', <https://www.who.int/publications/i/item/9241562579>.

Wu, C. & Hargreaves, C. A., 2021, 'Topological machine learning for multivariate time series', *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 34, no. 2, pp. 311–326.

Wyatt, J. C. & Liu, J. L., 2002, 'Basic concepts in medical informatics', *Journal of epidemiology and community health*, vol. 56, no. 11, pp. 808–812.

Xia, L., Malik, A. S. & Subhani, A. R., 2019, *A physiological signal-based method for early mental-stress detection*, Cyber-Enabled Intelligence, Taylor & Francis, p. 259–289.

Xing, X., Li, Z., Xu, T., Shu, L., Hu, B. & Xu, X., 2019, 'SAE LSTM: A New framework for emotion recognition from multi-channel EEG', *Frontiers in neurorobotics*, vol. 13, p. 37.

Yamak, P. T., Yujian, L. & Gadosey, P. K., 2019, 'A comparison between ARIMA, LSTM, and GRU for time series forecasting', *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*, ACM, New York United States, pp. 49–55.

Yen, P. T.-W. & Cheong, S. A., 2021, 'Using topological data analysis (TDA) and persistent homology to analyze the stock markets in Singapore and Taiwan', *Frontiers in Physics*, vol. 9, pp. 1–19.

Yen, P. T.-W., Xia, K. & Cheong, S. A., 2021, 'Understanding changes in the topology and geometry of financial market correlations during a market crash', *Entropy*, vol. 23, no. 9, pp. 1–48.

Yoo, G., Seo, S., Hong, S. & Kim, H., 2018, 'Emotion extraction based on multi bio-signal using back-propagation neural network', *Multimedia Tools and Applications*, vol. 77, p. 4925–4937.

Yu, H., Klerman, E. B., Picard, R. W. & Sano, A., 2019, 'Personalized wellbeing prediction using behavioral, physiological and weather data', *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, IEEE, pp. 1–4.

Yuan, Y. & Mitra, G., 2016, 'Market regime identification using hidden Markov models', *SSRN Electronic Journal*.

Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R. R. & Smola, A. J., 2017, 'Deep sets', *Advances in neural information processing systems*, vol. 30, pp. 1–11.

Zhai, J. & Barreto, A., 2006, 'Stress detection in computer users based on digital signal processing of noninvasive physiological variables', *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, pp. 1355–1358.

Zhang, H., Li, K. & Fu, X., 2010, 'On pinning control of some typical discrete-time dynamical networks', *Communications in Nonlinear Science and Numerical Simulation*, vol. 15, no. 2, pp. 182–188.

Zhang, H., Zhu, Y., Maniyeri, J. & Guan, C., 2014, 'Detection of variations in cognitive workload using multi-modality physiological sensors and a large margin unbiased regression machine', *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, pp. 2985–2988.

Zhang, J. & Small, M., 2006, 'Complex network from pseudoperiodic time series: Topology versus dynamics', *Physical Review Letters*, vol. 96, no. 23, pp. 1–4.

Zhang, J., Yin, Z., Chen, P. & Nichele, S., 2020a, 'Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review', *Information Fusion*, vol. 59, pp. 103–126.

Zhang, J., Zhou, J., Tang, M., Guo, H., Small, M. & Zou, Y., 2017, 'Constructing ordinal partition transition networks from multivariate time series', *Scientific reports*, vol. 7, no. 1, pp. 1–13.

Zhang, M., Huang, T., Guo, Z. & He, Z., 2022, 'Complex-network-based traffic network analysis and dynamics: A comprehensive review', *Physica A: Statistical Mechanics and its Applications*, vol. 607, p. 128063.

Zhang, X., Gao, Y., Lin, J. & Lu, C.-T., 2020b, 'TapNet: Multivariate time series classification with attentional prototypical network', *Proceedings of the AAAI Conference on Artificial Intelligence*, , vol. 34pp. 6845–6852.

Zhao, B., Lu, H., Chen, S., Liu, J. & Wu, D., 2017, 'Convolutional neural networks for time series classification', *Journal of Systems Engineering and Electronics*, vol. 28, no. 1, pp. 162–169.

Zheng, M., Domanskyi, S., Piermarocchi, C. & Mias, G. I., 2021, 'Visibility graph based temporal community detection with applications in biological time series', *Scientific reports*, vol. 11, no. 1, p. 5623.

Zhou, C., Dong, Z. & Lin, H., 2022, 'Learning persistent homology of 3d point clouds', *Computers & Graphics*, vol. 102, pp. 269–279.

Zia, A., Khamis, A., Nichols, J., Tayab, U. B., Hayder, Z., Rolland, V., Stone, E. & Petersson, L., 2024, 'Topological deep learning: a review of an emerging paradigm', *Artificial Intelligence Review*, vol. 57, no. 4, p. 77.

Zubair, M., Yoon, C., Kim, H., Kim, J. & Kim, J., 2015, 'Smart wearable band for stress detection', *2015 5th International Conference on IT Convergence and Security (ICITCS)*, IEEE, pp. 1–4.

Ávila, G. R., Gapelyuk, A., Marwan, N., Stepan, H., Kurths, J., Walther, T. & Wessel, N., 2013, 'Classifying healthy women and preeclamptic patients from cardiovascular data using recurrence and complex network methods', *Autonomic Neuroscience*, vol. 178, no. 1-2, pp. 103–110.