Doctoral thesis

Doctoral theses at NTNU, 2023:285

Muhammad Ali Fauzi

The Impact of Stress on Healthcare Staff’s Cybersecurity Practices

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology and Electrical
Engineering
Dept. of Information Security and
Communication Technology

NTNU
Norwegian University of
Science and Technology

Muhammad Ali Fauzi

The Impact of Stress on Healthcare Staff's Cybersecurity Practices

Thesis for the Degree of Philosophiae Doctor

Gjøvik, September 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

ISBN 978-82-326-7270-7 (printed ver.)
ISBN 978-82-326-7269-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2023:285

Printed by NTNU Grafisk senter

"The best of people are those that bring most benefit to the rest of mankind."

(Muhammad PBUH.)

Declaration of Authorship

I, Muhammad Ali Fauzi, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signed:

(Muhammad Ali Fauzi)

Date:

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Information Security and Communication Technology at the Norwegian University of Science and Technology, Norway.

The study was carried out during the period from June 2019 to May 2023. The thesis is written on the basis of 6 published research papers, 2 in press papers, and 2 research papers under review. The articles are reformatted to fit the thesis's structure and the contents of the original articles, including the table formats are maintained.

Summary

Digitalization has revolutionized the healthcare industry, offering numerous advantages. However, it has also introduced the risk of data breaches through cyber-attacks. Healthcare information systems, containing valuable data that can be sold at high prices, are often targeted by adversaries. Surprisingly, a recent report revealed that 82% of data leaks involved a human element. As a result, the study of human behavior in cybersecurity has gained significant attention. However, the relationship between stress levels and cybersecurity practices, particularly in the healthcare setting, has only been the subject of a small number of peer-reviewed studies. This study aims to fill this gap by examining the relationship between stress levels and risky cybersecurity practices among hospital workers. Additionally, it investigates how stress impacts email judgment performance. Furthermore, the study compares different strategies to develop effective multimodal stress detection systems. To achieve these objectives, the research methodology employs correlation analysis, causal analysis utilizing a randomized controlled trial (RCT), and comparative analysis of various machine learning models.

The correlation analysis reveals a positive correlation between stress levels and risky cybersecurity practices in Ghana, Indonesia, and the combined dataset from three countries (Ghana, Norway, and Indonesia), indicating that individuals experiencing higher stress are more likely to engage in behaviors compromising cybersecurity. The causal analysis shows that while stress does not directly compromise participants' ability to detect phishing emails, higher stress levels are significantly correlated with lower accuracy in the Indonesian context. Furthermore, completion time is identified as a potential mediator of the impact of stress on email judgment performance, with longer time associated with better performance. While the result from Norway showed no significant difference, the result from Indonesia suggested that participants in the non-stress group took a significantly longer time to complete judging emails than participants in the stress group.

The comparative analysis of multimodal stress detection systems demonstrates the superiority of multiple sensor fusion models over individual sensors with the weighted score-level fusion approach getting the best performance. Furthermore, preprocessing such as feature normalization and feature selection was proven to improve system performance. In term of classifier, using Logistic Regression as the classifier yield the best results.

This study contributes to our understanding of the impact of stress on cybersecurity practices and email judgment performance. These findings have important implications for hospital management, emphasizing the need for targeted training programs and support systems to enhance cybersecurity practices among staff. The comparative analysis provides insights into effective multimodal stress detection systems, promoting privacy and accuracy. In conclusion, It offers practical recommendations for healthcare organizations to enhance cybersecurity and provides insights for the development of effective multimodal stress detection systems.

Acknowledgments

> ”When eating fruit, remember the one who planted the tree”
>
> VIETNAMESE PROVERB

I would like to express my deepest gratitude to my principal supervisor, Associate Professor Bian Yang, for his unwavering commitment, invaluable guidance, continuous support, and insightful advice throughout the entire journey of this research and thesis writing. His encouragement has allowed me to explore additional research areas of personal interest and engage in fruitful collaborations, which have contributed to establishing my network within the academic community. My supervisor has been an exceptional guiding force with his humility, fair assessment methods, and dedicated mentorship. I would also like to extend my appreciation to Professor Christoph Busch, my co-supervisor, for his valuable advice during this journey.

I am incredibly grateful to my wife, Syarifatun Nadhiroh Qomariyah, whose unwavering encouragement during the challenging moments of this PhD journey provided me with peace of mind. My deepest gratitude extends to our two children, Muhammad Hanif Alfaruq and Kaja Najmatun Fauzi, my entire beloved family, and my friends for their unwavering support throughout this journey.

Furthermore, I am profoundly fortunate to have received tremendous support during my PhD work. I would like to take this opportunity to express my heartfelt appreciation to the participants of our data collection and the people who facilitate the data collection including the hospital staff in Norway, Ghana, and Indonesia (RS Aisyiyah Bojonegoro), and academic staff and students from NTNU Gjøvik, Brawijaya University, and the University of British Columbia. Their cooperation has been instrumental to the success of my research. I also extend my gratitude to my wonderful office mates, as well as the academic and administrative staff of the Department of Information Security and Communication Technology. Lastly, I would like to express my gratitude to the entire Norwegian and Indonesian society for their support and contributions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> ”Amateurs hack systems,
> professionals hack people”
>
> BRUCE SCHNEIER

1.1 Motivation, research aim, and scope

In recent years, the healthcare industry has increasingly relied on technology for managing patient data and improving patient care. However, the heightened use of technology has also led to new cybersecurity risks that can compromise sensitive healthcare data. Data breaches, financial losses, reputational harm, and, most crucially, jeopardized patient care are all possible outcomes of these attacks [114]. Recent investigations reveal that healthcare institutions are one of the top sectors that are at the highest risk of data breaches [147]. This is particularly concerning since healthcare organizations are responsible for safeguarding extremely private and sensitive information, such as medical records, personal data, and financial information.

Despite implementing advanced cybersecurity measures and technologies, the human element is often regarded as the weakest link in the security chain [149, 156]. Risky human behavior can create vulnerabilities and weaknesses that can be exploited by cybercriminals, thereby leaving digital systems and data vulnerable to attack [77, 132]. Human errors, such as using weak passwords, falling for phishing emails, or failing to follow security protocols, can compromise the security of healthcare systems. Hence, while technical measures such as firewalls and encryption can help protect against cyber threats, the human factor is also critical in ensuring the security of healthcare systems. The Verizon report [147] also highlights the significant role of human error in data breaches, with 82% of breaches involving a human element. The report also reveals that social engineering is a common pattern in data breaches. Thus, it has prompted numerous studies aimed at understanding human behavior in relation to the use of computers and the internet, as well as identifying the factors that affect cybersecurity practices.

Stress is a key human factor that has received little attention in the context of healthcare cybersecurity. Healthcare staff faces a range of stressors in their work, including long hours, heavy workloads, the emotional cost of caring, lack of reward, etc. [35, 100, 61]. Psychological research has revealed that heightened stress levels can greatly hinder decision-making abilities and task performance, resulting in less advantageous choices and diminished patient safety. Starcke and Brand [136] conducted a study on acute stress and decision-making performance, concluding that individuals experiencing high stress tend to take more risks, particularly in ambiguous situations, which could lead to unfavorable outcomes. Additionally, stress has been shown to impede learning and knowledge acquisition, making it more difficult to comprehend and assimilate new information. A study by Schoofs et al. [127] investigated the impact of chronic stress on learning and found that individuals under high stress have slower learning rates and poorer recall than those with lower stress levels. Moreover, Michailidis and Banks [101] discovered that employees experiencing burnout are more likely to make impulsive or illogical decisions than those who feel more fulfilled at work. Wemm and colleagues [150] also found that stressed individuals tend to learn new information at a slower pace and make less profitable decisions. In terms of healthcare, a systematic review conducted by Tawfik et al. [140] explored the impact of stress on patient safety in healthcare settings, revealing that high levels of stress among healthcare providers are linked to increased medical errors, adverse events, and lower patient satisfaction.

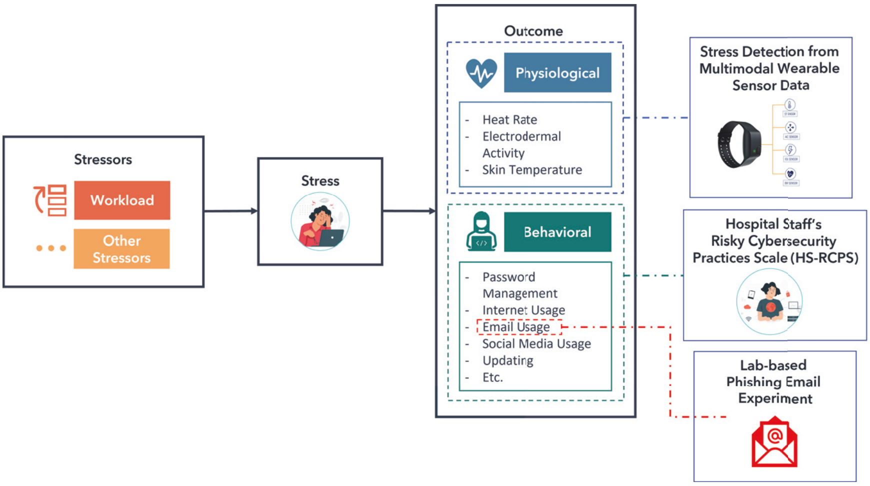


Figure 1.1: Study Framework

Given the significant impact of stress on human behavior, it is crucial to understand how stress affects cybersecurity practices. However, research on this topic, especially in the healthcare sector, is limited. Moreover, most studies were conducted in developed countries such as Australia and the

United Kingdom [97, 53, 146]. This study investigated the impact of stress levels on hospital staff's cybersecurity practices in three countries: Ghana, Norway, and Indonesia. In this research context, healthcare staff refers to the individuals who work in various roles within the healthcare industry. In this study, we focused on the hospital setting because it can be a valuable representation of healthcare settings. Hospitals are comprehensive healthcare institutions that offer a wide range of medical services and typically serve a large and diverse patient population. Participants targeted in this study include all of the staff in the hospital, involving a diverse range of professionals and staff members from clinical healthcare professionals (e.g. doctors, nurses) to support staff (e.g. IT staff). The framework of this study is depicted in Figure 1.1.

The research activities conducted in this thesis are classified into four (4) major parts. In the first part, as depicted in Figure 1.2, a correlation analysis was conducted to examine the relationship between stress and cybersecurity practices among hospital staff in Ghana, Norway, and Indonesia. Besides, we also compared cybersecurity practices across these countries and explored the association between demographic variables and cybersecurity practices. An online survey was utilized to collect data on the healthcare staff's demographic details, stress levels, and cybersecurity practices. The participants' stress levels and risky cybersecurity practices were assessed using the Perceived Stress Scale (PSS) and Hospital Staff's Risky Cybersecurity Practices Scale (HS-RCPS), respectively.

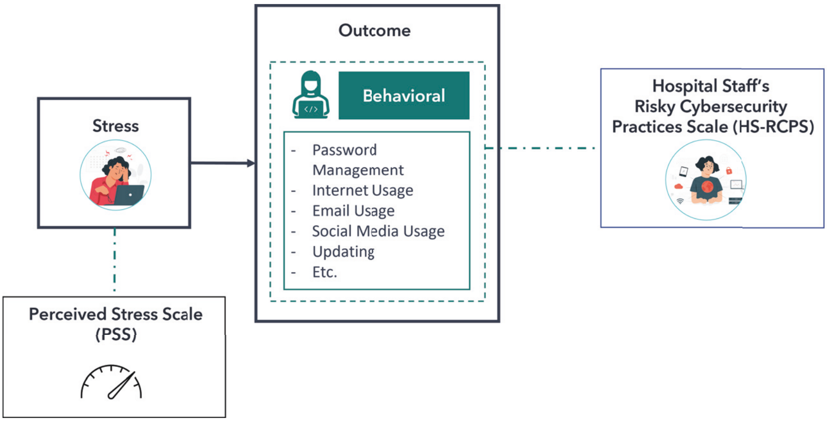


Figure 1.2: Study Part 1

In the second part, as shown in Figure 1.3, a causal analysis using a randomized controlled trial was performed to assess the impact of stress on phishing emails in Norway and Indonesia. Participants were randomly allocated to either the control or intervention group. The participants in the experimental group were administered the Purple Multitasking Framework (MTF) to induce stress. Stress (stress versus no stress) served as the independent variable, while completion time and email judgment performance, including accuracy, sensitivity, and specificity, were the dependent variables.

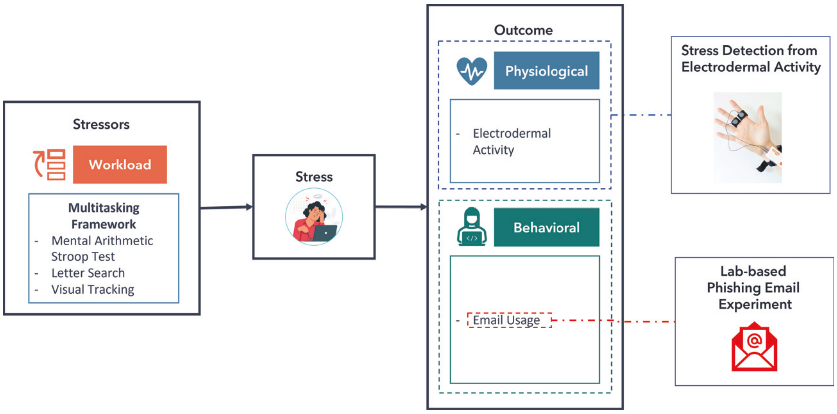


Figure 1.3: Study Part 2

Furthermore, monitoring hospital workers' stress levels has many advantages as one of the steps in stress management. Knowing their stress level can help them stay aware and better manage their response to situations while also identifying when they need to take action to address stress [87]. While questionnaires like the Perceived Stress Scale [25] and Perceived Stress Questionnaire [84] are commonly used to assess stress levels, they can be time-consuming and not practical for continuous monitoring. Another method for assessing stress levels is by measuring physiological responses related to stress, such as heart rate, blood pressure, and skin conductance, using sensors like the electrocardiogram (ECG) and galvanic skin response (GSR). Advances in wearable technology have made it possible to collect data on multiple physiological responses continuously and passively. However, some wearable devices can be inconvenient to wear during work, such as chest-worn devices and finger-placed GSR sensors [134]. Smartwatches have emerged as a promising platform for stress monitoring due to their built-in sensors, including Blood Volume Pulse, Electrodermal Activity, temperature, and accelerometer, which make multimodal-based stress detection possible. Additionally, smartwatches have a high degree of social acceptance and are widely used, making them a convenient and practical option for continuous stress monitoring [81].

Machine learning (ML) technologies have been remarkable in empowering practical artificial intelligence (AI) applications, including in medical fields. By using multiple sensor data, richer information as the result of a combination of many environmental views can be used to train the machine learning algorithm so that the trained model can be more robust. However, most prior research used feature-level fusion to combine numerous sensor data (e.g. [125, 55, 67, 130]). Therefore, the third part was about building an effective multimodal stress detection system, as depicted in Figure 1.4. In this part, we examined several methods for combining multimodal data from sensors, combining several classifiers, and using preprocessing methods in order to improve the effectiveness of the stress detection system. including accuracy, precision, recall, and F-1 measure.

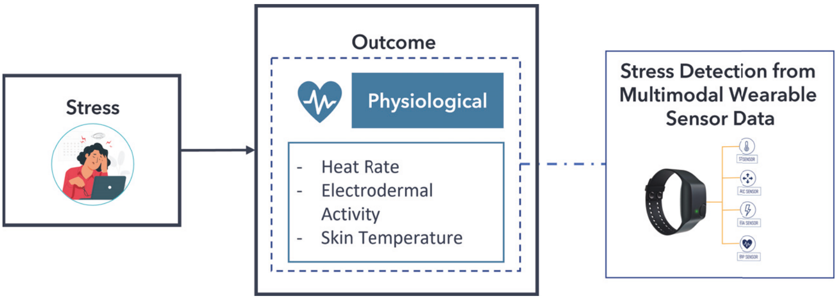


Figure 1.4: Study Part 3

Furthermore, the data for stress detection contain sensitive information that can jeopardize the user's privacy. Thus, a growing number of studies put attention on safeguarding private data in analysis processes. In the last part of this study, we implemented a privacy-preserving technology for stress detection and conducted a comparative analysis with traditional methods. Three machine learning strategies were compared: individual learning, centralized learning, and federated learning. Federated learning (FL) can solve privacy challenges in stress detection by allowing each data register to train models on separate, isolated datasets while only sharing the trained models, which do not contain any personal information. as depicted in Figure 1.5. The comparison includes several factors, especially the effectiveness and privacy of the stress detection system. The performance metrics for effectiveness used were accuracy, precision, recall, and F-1 measure. In addition, usability, and the need for hardware and computational power were also discussed.

The articles and their related mapping to the various parts of the study are shown in Figure 1.6.

1.2 Research questions

Based on the research aim, objective, and motivation, four research questions were formulated to guide this thesis study. The relation between arti-

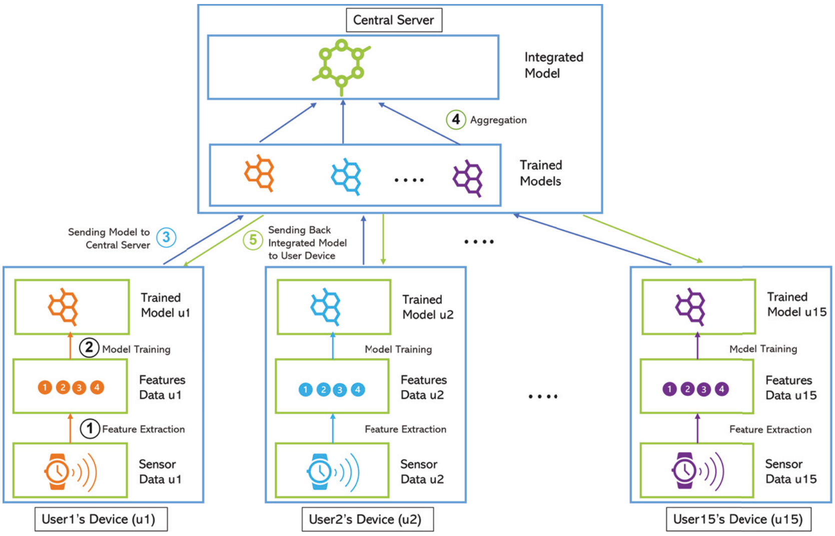


Figure 1.5: Study Part 4

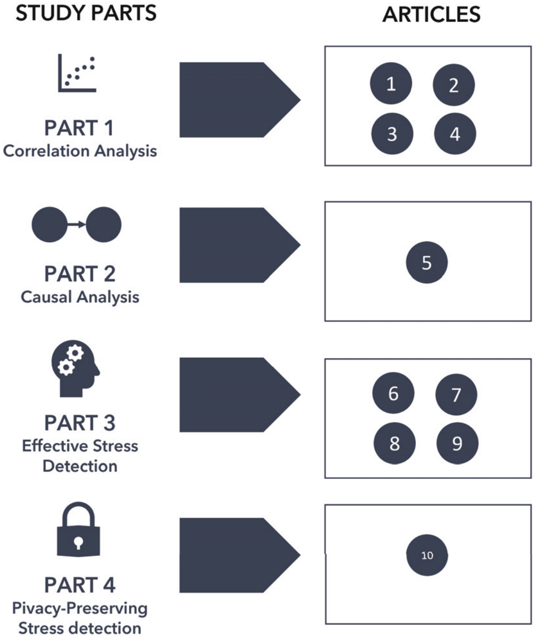


Figure 1.6: Study Parts

cles and the research questions is depicted in Figure 1.7. The research questions are outlined in the following paragraphs.

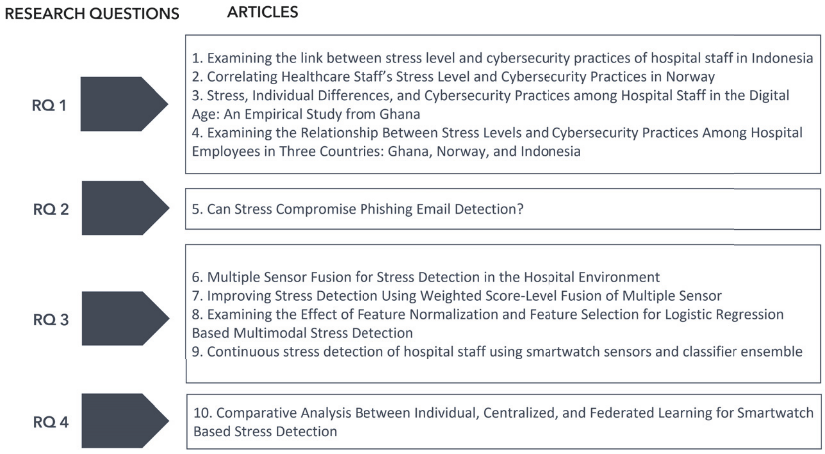

RESEARCH QUESTIONS
ARTICLES
RQ 1
1. Examining the link between stress level and cybersecurity practices of hospital staff in Indonesia
2. Correlating Healthcare Staff's Stress Level and Cybersecurity Practices in Norway
3. Stress, Individual Differences, and Cybersecurity Practices among Hospital Staff in the Digital Age: An Empirical Study from Ghana
4. Examining the Relationship Between Stress Levels and Cybersecurity Practices Among Hospital Employees in Three Countries: Ghana, Norway, and Indonesia
RQ 2
5. Can Stress Compromise Phishing Email Detection?
RQ 3
6. Multiple Sensor Fusion for Stress Detection in the Hospital Environment
7. Improving Stress Detection Using Weighted Score-Level Fusion of Multiple Sensor
8. Examining the Effect of Feature Normalization and Feature Selection for Logistic Regression Based Multimodal Stress Detection
9. Continuous stress detection of hospital staff using smartwatch sensors and classifier ensemble
RQ 4
10. Comparative Analysis Between Individual, Centralized, and Federated Learning for Smartwatch Based Stress Detection

Figure 1.7: Research questions and their mapping to the articles.

Research question 1 (RQ1): What is the relationship between stress level and cybersecurity practices among hospital staff? The human element is frequently viewed as the weakest link in the security chain. Despite the best cybersecurity measures and technologies, human behavior can still introduce vulnerabilities and weaknesses that cybercriminals can exploit. One key human factor that has received only a little attention in the context of healthcare cybersecurity is stress. Healthcare staff faces a range of stressors in their work, including long hours, heavy workloads, the emotional cost of caring, lack of reward, etc. Research in psychology has found that high levels of stress can significantly impair decision-making abilities and task performance, leading to less profitable decisions and lower levels of patient safety. This research question aims to investigate the relationship between stress levels and cybersecurity practices among hospital workers in Ghana, Norway, and Indonesia. This study hypothesized that employees with higher stress levels are more likely to engage in risky security practices, including password management, internet usage, email usage, updating, backup, etc. By understanding the relationship between stress on cybersecurity practice, healthcare organizations can develop targeted interventions and training programs to mitigate the risks associated with human factors in cybersecurity. It can be one of the basis for informing the development of more effective cybersecurity policies, procedures, and training programs. Ultimately, this research has the potential to improve the security of healthcare systems and protect sensitive patient data.

Research question 2 (RQ2): Can stress compromise phishing email

detection? Phishing attacks are a prevalent form of cybercrime. Phishing attacks aim to trick people into divulging sensitive information or providing access to their computer systems through fraudulent emails. The consequences of falling victim to a phishing email can be severe, ranging from identity theft to financial loss. In some cases, it can even lead to data breaches that compromise the security of entire organizations. As a result, detecting phishing emails has become a critical cybersecurity task. One factor that has received considerable attention in recent years is stress. Stress can affect an individual's cognitive functioning, including attention, memory, impulsivity, and decision-making abilities, which are essential for detecting phishing emails. This research question aims to investigate the impact of stress on participants' ability to detect phishing emails using a Randomized controlled trial (RCT). Stress (stress versus no stress) served as the independent variable, while completion time and email judgment performance, including accuracy, sensitivity, and specificity, were the dependent variables. This study aimed to contribute to a better understanding of the factors that influence people's ability to detect phishing emails and provide insights into how to improve cybersecurity education and training programs.

Research question 3 (RQ3): How to build an effective stress detection system from multimodal wearable sensor data? Monitoring hospital workers' stress level has many advantages. Knowing their own stress level can help them stay aware and feel more in control of their response to situations and know when it is time to relax or take some actions to treat it properly. Smartwatch has recently emerged as a new platform that provides many successful applications. These devices have several built-in sensors that are useful for stress monitoring including Blood Volume Pulse (BVP), Electrodermal Activity (EDA), temperature, accelerometer, etc. Besides, the use of watches is well known and has a high degree of social acceptance of their ubiquity in everyday life, making it suitable to be used in a working environment such as a hospital. Therefore, it has a high potential to be applied for multi-modal-based stress detection. This research question aims to examine several methods for combining multimodal data from sensors, combining several classifiers, and using preprocessing methods in order to improve the stress detection system effectiveness, including accuracy, precision, recall, and F-1 measure.

Research question 4 (RQ4): What is the difference between individual, centralized, and federated learning for stress detection in terms of effectiveness, privacy, usability, and hardware and computational power needed? Machine learning techniques generally need a sufficient amount of data for training to perform well. Therefore, to create a robust method, we need to collect sensor data from several users and collect them at a central server for processing. However, the uploaded medical data may contain individual privacy-related and sensitive information. Privacy breaches can happen if the central server is compromised. As a result, a growing number of studies put attention on safeguarding private data in analysis processes. One of the most popular methods is federated learning. However, usually, there is a trade-off in the classification performance. This research question aims to analyze the difference between individual, centralized, and federated Learning for stress detection in terms of effectiveness, privacy, usability, and hardware and computational power needed. This study aimed to contribute to a better understanding of machine learning strategies as the basis of selecting the best one depending on the needs and priority.

1.3 Background

This section presents relevant background and an overview of the thesis to facilitate a better understanding of the remaining aspect of this thesis.

1.3.1 Stress in healthcare environment

Stress is a complex psychological and physiological response to challenging or threatening situations, events, or stimuli. It involves a range of physical, emotional, cognitive, and behavioral changes that prepare the body and mind to cope with potential dangers or stressors [26, 98]. The transactional model of stress, proposed by Lazarus and Folkman [82], emphasizes the dynamic relationship between individuals and their environment in the experience of stress. It suggests that stress is not solely determined by external events but rather by how individuals appraise and interpret those events. The transactional model consists of two evaluative components: primary appraisal and secondary appraisal. According to the model, the stress process begins with stimuli from the internal or external environment. Stimuli that lead to stress also are referred to as stressors. These stimuli are constantly evaluated through the primary appraisal processes, which determine whether they are irrelevant, benign-positive, or stressful to the individual's well-being. Once a situation is appraised as relevant and potentially stressful, the individual engages in secondary appraisal. Secondary appraisal focuses on the individual's perceived ability to cope with the demands of the situation and the potential outcomes of their coping efforts. Psychological stress is perceived when coping-option resources are perceived to be insufficient to overcome a stressful-appraised situation [82]. The outcome of stress is referred to as strain, which can manifest in physiological, psychological, or behavioral forms [27]. A physiological strain involves the physiological arousal of the body. This involves the release of stress hormones (such as cortisol and adrenaline), increased heart rate, elevated blood pressure, heightened muscle tension, and other physiological changes [121, 64, 12].

Psychological strain relates to emotional reactions to stressors. Individuals may experience emotional distress, anxiety, irritability, mood swings, cognitive disruptions, difficulty concentrating, negative thoughts, job dissatisfaction, or depression [118, 128, 99, 35]. Behavioral strain is characterized by reduced productivity, disruptive behavior, or poor task performance [139, 151, 63].

Healthcare staff are particularly susceptible to high levels of stress due to the nature of their work, which often involves many stressors such as long hours, high workloads, emotionally challenging situations, and critical decision-making [128, 99]. Healthcare staff can experience stress from a variety of sources, including patient care, work environment, organizational culture, interpersonal relationships, and personal factors such as financial or family problems [11, 35, 128]. Stress in healthcare staff can have significant consequences for individual health and wellbeing, as well as the quality of patient care. High levels of stress can lead to physical and emotional exhaustion, burnout, reduced job satisfaction, decreased productivity, and increased absenteeism [128, 99, 35]. In addition, stress can affect cognitive functioning, decision-making, and interpersonal relationships, which can compromise the quality of patient care [151, 63, 128, 35].

1.3.2 Healthcare staff cybersecurity Practice

Healthcare staff cybersecurity practice refers to the actions and practices that healthcare employees take to protect sensitive patient information and healthcare systems from cyber threats [59]. This includes following best practices such as using strong passwords, regularly updating software, and being vigilant about phishing attacks, as well as implementing technical measures such as firewalls, antivirus software, encryption, etc. [34, 57]. Hospitals, being complex healthcare institutions, involve a diverse range of professionals and staff members who collectively contribute to providing comprehensive patient care, administrative support, and specialized services. The diversity of roles within a hospital leads to a range of cybersecurity practices, each informed by the unique demands and responsibilities of the respective positions.

Healthcare staff play a crucial role in upholding information security principles by adhering to policies and procedures that safeguard patient records, whether they are in electronic or physical format. Information security practices of hospital staff involve a set of guidelines and protocols designed to protect sensitive patient data and ensure the confidentiality, integrity, and availability of information. As most healthcare data is now digitalized, cybersecurity practices become critical to protect electronic health records (EHRs), medical devices, and other digital assets from cyber threats. Poor cybersecurity practices among healthcare staff can lead to a range of cybersecurity incidents, including data breaches, malware infections, and phishing attacks, which can compromise patient information and put individuals at risk [9]. The consequences of these incidents extend beyond data privacy concerns, as patient safety can also be jeopardized. For instance, a cyberattack on medical devices or healthcare systems could disrupt critical patient care and treatment processes, leading to potential harm or delays in providing essential medical services [114]. A data breach in a healthcare setting can expose patients to identity theft, fraud, and other forms of cybercrime [56, 142]. Given the critical nature of healthcare data and the potential consequences of cyber threats, healthcare staff must adopt a balanced approach that places equal emphasis on both cybersecurity and patient care. By recognizing the significance of cybersecurity practices, healthcare workers can contribute to maintaining the trust and well-being of patients while safeguarding their data from cyber incidents.

1.3.3 Stress impact on healthcare staff cybersecurity practice

One of the forms of strain caused by stress is behavioral [27]. The behavioral strain theory, also known as the General Strain Theory (GST) [3], can provide insights into the impact of stress on healthcare staff cybersecurity practices. This theory suggests that when individuals experience stress or strain, they are more likely to engage in deviant or maladaptive behaviors as a means of coping with or escaping from the stressors. Several studies suggested that stress can lead to deviant workplace behavior [117, 107]. In the context of healthcare staff cybersecurity practices, such behaviors may include several practices. For example, under stress, healthcare staff may be more inclined to deviate from established cybersecurity policies and procedures. Stress can drive individuals to seek quick solutions or workarounds to cope with time pressures or perceived inefficiencies. They may take shortcuts or neglect security measures, such as sharing passwords, using unsecured devices, or failing to update security software. Besides, stress can also affect an individual's cognitive functioning, including attention, memory, impulsivity, and decision-making abilities ([131, 101, 150]), leading to errors or carelessness in conducting some crucial cybersecurity task such as phishing email handling.

1.3.4 Stress detection system

The most common way to assess stress levels is by using questionnaires (e.g. Perceived Stress Scale [25], Perceived Stress Questionnaire [84], etc.). However, this method takes time so that it is not convenient to be performed every day for continuous monitoring. The other stress level assessment method is by measuring the physiological responses related to stress such as heart rate, blood pressure, skin conductance, respiration activity, etc. Some sensors can be used to conduct the measurement task. For ex-

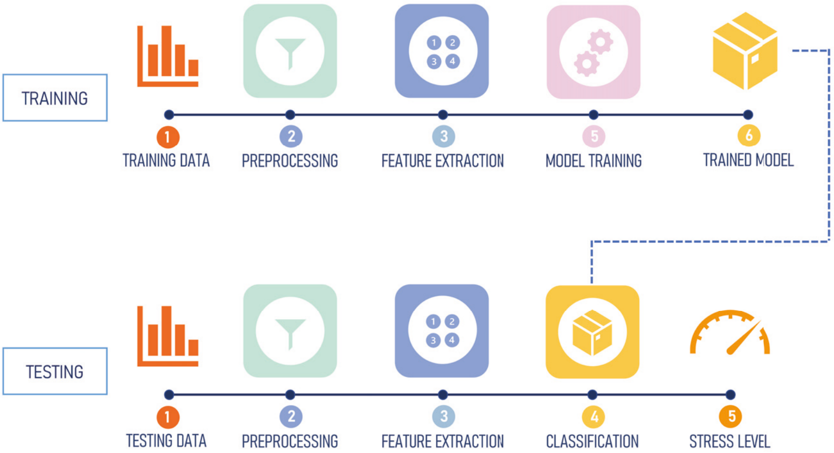


Figure 1.8: Stress Detection Pipeline.

ample, an electrocardiogram (ECG) can be used to measure the heart rate, galvanic skin response (GSR) for skin conductance, etc. The recent advance in wearable devices with sophisticated built-in sensors makes it feasible to passively collect multimodal data from people's daily lives for automatic continuous stress detection purposes. Many previous works have been successfully leveraging multimodal sensor data and machine learning methods to build automatic stress detection. The popular machine learning methods used are Random Forest, Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression [125, 55, 67, 130].

Generally, the stress detection pipeline is presented in Figure 1.8 [14, 49]. The pipeline is divided into two main parts: training and testing. The final result of the training phase is to create a trained ML model while the final product of the testing phase is the classification result. The first step in the pipeline for both training and testing is preprocessing. Raw data often requires preprocessing to transform it into a suitable format for training the ML model. This step involves activities like cleaning the data by removing inconsistencies or missing values, handling outliers, and converting data types. The next step is extracting features from the raw data. In the case of stress detection using multimodal sensor data, statistical or frequency domain features can be extracted from the raw sensor data. Furthermore, in the training phase, the features are used to train an ML algorithm to produce a trained ML model. This ML model is then saved and used in the testing phase. In the testing phase, the features are classified by using the trained ML model from the training phase. Finally, the stress level of the testing data was determined. To evaluate the effectiveness of the stress classification system, several evaluation metrics can be used such as accuracy, precision, recall, and F_1-measure (F_1) [130, 67, 125].

1.4 Related work and identified gap

Identification of studies via databases was conducted in order to explore the current state of the art on the impact of stress on cybersecurity practices following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement [102]. The eligibility criteria were original studies that empirically examine the relationship between stress and cybersecurity practices. The language of publication was restricted to English. Papers not meeting eligibility criteria were excluded from the review. The articles were identified by conducting a literature search through 3 bibliographic databases: Scopus, Web Science, and PubMed. A manual search based on the reference lists of retrieved publications was also conducted. The literature search was conducted using the following keywords: (“stress” AND ((“information security” OR “cyber?security” OR “data security” OR “password” OR “phishing”) AND (practice* OR behavio* OR perception OR awareness OR decision OR choice OR judgment OR error))) without published time limitation. A PRISMA flow diagram of the literature search process is shown in Fig. 1.9. The studies included are displayed in Table 1.1.

Table 1.1: Studies included in the review

Year	Ref	Methodology	Participants	Country	Cybersecurity practices
2014	[30]	Lab based experiment	University students	US	Choosing safe apps
2014	[28]	Online Questionnaire	Company workers	US	ISP compliance
2017	[70]	Field experiment	University faculty, staff, and students	US	Phishing email susceptibility
2018	[65]	Paper-based questionnaire	Company workers	South Korea	ISP Compliance
2018	[53]	Lab based experiment	university staff and students	UK	Password Choice

2018	[97]	Online Questionnaire	Company workers	Australia	Information Security Awareness
2019	[33]	Experience Sampling Method (ESM) with online questionnaire	Company workers	US	ISP compliance
2021	[1]	Online web-based experiment	General participants from Mturk	Not specified	Email judgement
2021	[48]	Online Questionnaire	Hospital workers	Indonesia	Several cybersecurity practices including password management, email, social media, information sharing, updating, backup, reporting, patient privacy, and visiting public website
2021	[143]	Online simulation	70.3 % of them are students	88.2 percent from Germany	ISP compliance
2021	[66]	Online and paper-based questionnaire	Company workers	South Korea	ISP compliance

2021	[144]	Questionnaire	University students	Croatia	Potentially risky behavior, information security awareness, user's beliefs about information security, and the quality and security of passwords.
2021	[79]	Questionnaire	General	US	Protective responses such as changing passwords and notifying others that a breach has occurred
2021	[104]	Online Questionnaire	Company workers	US	ISP compliance
2021	[71]	Online Questionnaire	Company workers	South Korea	ISP compliance
2021	[146]	Online Questionnaire	University students	France	Information Security Awareness
2022	[158]	Online Questionnaire	Company workers	US	ISP compliance
2022	[159]	Online Questionnaire	Company workers	US	ISP compliance
2022	[72]	Online Questionnaire	computer-using employees with financial responsibilities	US	Computer Fraud

2022	[5]	Online Questionnaire	Oil and Gas companies employees	Malaysia	ISP compliance
2022	[22]	Questionnaire	Company workers	China	ISP compliance
2022	[21]	Online Questionnaire	Company workers	Not specified	ISP compliance
2023	[73]	Questionnaire	Company workers	Not specified	Voluntary ISP violating behavior

1.4.1 Stress and Cybersecurity Practices

Most studies analyzing the impact of stress on cybersecurity practices used questionnaires. In the study conducted by McCormack et al. [97], the relationship between employee job stress and information security awareness was investigated using a quantitative method. A survey questionnaire was administered to working Australians, where the Human Aspects of Information Security Awareness - Questionnaire (HAIS-Q) by Parsons et al. [110] was utilized to assess information security awareness, and the Job Stress Scale developed by Lamber et al.[80] was used to measure job stress. The study revealed that lower levels of job stress were associated with better knowledge, attitude, and behavior toward mitigating cyber hazards. Fauzi et al. [48] examined the correlation between stress levels and risky cybersecurity practices among hospital staff in Indonesia. The employees' stress levels were measured using the Perceived Stress Scale (PSS) by Cohen et al. [23] while a new scale based on the security behavior intentions scale (SeBIS) by Egelman et al. [34] and HAIS-Q by Parsons et al. [110] was employed to assess the staff's risky cybersecurity practices. The findings demonstrated a significant link between higher stress levels and riskier cybersecurity practices, such as clicking on links in emails from unknown senders and failing to create strong passwords. The study by Venard [146] focused on the impact of stress related to COVID-19 on the cybersecurity behaviors of students in a higher education institution in the west of France. A new stress measurement scale was developed based on the work of Cohen et al. [24] to assess stress related to COVID-19 while cybersecurity behaviors were measured using HAIS-Q. The study revealed that the stress related to COVID-19 did not directly affect cybersecurity behavior.

Velki and Milić [144] investigated the mediating role of stress in the associations between risky online behaviors and risk factors (real-life risky behaviors and information security awareness) and a protective factor (life satisfaction). The study involved the distribution of questionnaires among stu-

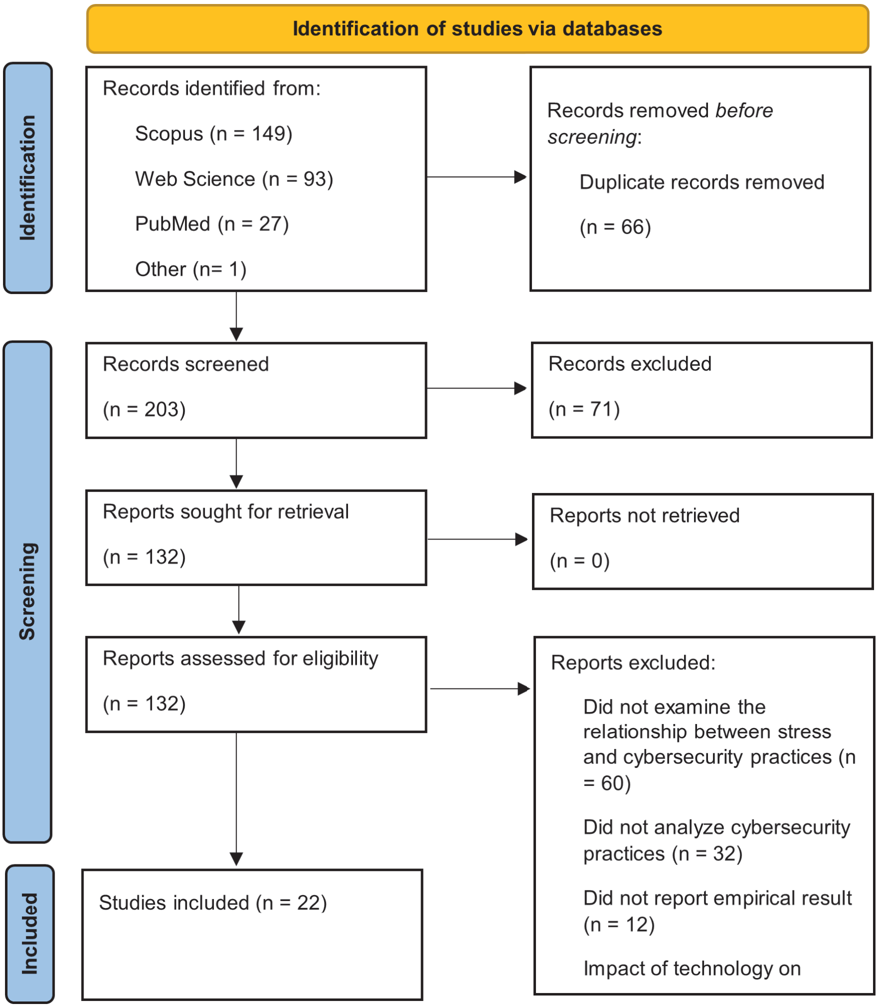

Identification of studies via databases
Identification
Screening
Included
Records identified from:
Scopus (n = 149)
Web Science (n = 93)
PubMed (n = 27)
Other (n= 1)
Records removed *before screening*:
Duplicate records removed
(n = 66)
Records screened
(n = 203)
Records excluded
(n = 71)
Reports sought for retrieval
(n = 132)
Reports not retrieved
(n = 0)
Reports assessed for eligibility
(n = 132)
Reports excluded:
Did not examine the relationship between stress and cybersecurity practices (n = 60)
Did not analyze cybersecurity practices (n = 32)
Did not report empirical result (n = 12)
Impact of technology on
Studies included (n = 22)

Figure 1.9: Flow diagram of literature search following PRISMA guidelines.

dents from four Croatian universities. Stress was measured using the PSS, while risky online behaviors were assessed using the Users' Information Security Awareness Questionnaire by Velki et al. [145]. The findings showed that stress significantly correlated with risky online behaviors. The results also suggested that the association between real-life risky behaviors and risky online behaviors became stronger under stress. Stress also fully mediates between life satisfaction and risky online behaviors. However, stress failed to mediate the association between information security awareness and risky online behaviors. Furthermore, Labrecque et al. [79] analyzed the relationship between stress and consumer protective responses following data breaches using questionnaires. The findings showed that stress following data breaches had a positive significant effect on protective behaviors including changing passwords and notifying others that a breach has occurred. The findings from this study are interesting because stress can lead to good cybersecurity behaviors.

In their two studies, Jiang and Zhang [72, 73] analyzed the relationship between work pressure and cybersecurity practices. Both studies were conducted in the US among company workers. The first study [72] found a statistically significant relationship between work pressure and computer fraud intention. Meanwhile, the second study [73] indicated that work pressure was positively related to nonmalicious ISP violation intentions. Furthermore, Jeon et al. [71] employed questionnaires to collect data from company workers in South Korea to examine the association between frustration and ISP compliance. They reported that frustration was negatively related to ISSP compliance intentions. Another stress that can be caused by work is role stress. Hwang et al. [65] used paper-based questionnaires to collect data from company workers in South Korea. They examined the relationship between security-related role stress (RS) and security-related technostress creators (TC) and ISP compliance. The findings showed that security-related technostress creators negatively affected information security compliance through organizational commitment. Meanwhile, the increased level of security-related role stress due to security-related technostress creators further served as another antecedent to decrease organizational commitment that could lead to ISP violation behavior. Shadbad and Biros [104] also assessed the relationship between role stress and ISP compliance intention using online questionnaires. The results suggested that role stress has a significant negative effect on ISP compliance intention.

Several prior works also used questionnaires to focus on the relationship between security-related stress (SRS) and information security policy (ISP) compliance. D'Arcy et al. [28] assessed the relationship between security-related stress (SRS) and information security policy (ISP) compliance including password sharing, password write-down, copying sensitive data to an insecure USB device, and failure to log off the workstation among company workers in US. The findings showed that security requirements perceived as overload, complex, and uncertain can induce employee rationalizations of ISP violations, which in turn increase susceptibility to this behavior. In another study, D'Arcy and Teh [33] employed an Experience Sampling Method (ESM) with online questionnaires to collect data about the association between security-related stress (SRS) and ISP compliance among 138 company workers in the US. The findings revealed that SRS had a positive association with frustration and fatigue, and these negative emotions were associated with the neutralization of ISP violations. Additionally, frustration and fatigue make employees more likely to follow through on their rationalizations of ISP violations that lead to ISP violation behaviors. Yazdanmehr et al. [159] also studied the association between SRS and ISP compliance. The results indicate that SRS triggers all three coping responses. The first coping response, problem-focused coping, then decreases ISP violation intention, whereas inward and outward emotion-focused coping increases it. Meanwhile, Ali and Dominic [5] also examined the relationship between SRS and ISP compliance. The participants were oil and gas company workers in Malaysia. The results indicated that employees perceive security requirements as stressful. The stress causes avoidance coping, which later leads to non-compliance behavior. Chen et al. [21] also found that employees with high levels of SRS tend to experience information security fatigue, and this negative emotion decreases their ISP compliance intention. In a similar focus, Hwang et al. [66] analyzed the association between security technostress creators and ISP compliance among company employees in South Korea. The results found that the more employees encounter security technostressors, the more negative the adherence to information security.

Some other studies focused on the challenge and hindrance aspects of ISP. Yazdanmehr et al. [158] used questionnaires to analyze employee reactions to information security policies. The results show that the challenge aspect of ISP demands elicits a positive psychological response from employees, which in turn triggers their planful problem-solving to deal with these demands. In contrast, the hindrance aspect of ISP demands provokes a negative psychological response that triggers employees' wishful thinking about ISP demands. Subsequently, planful problem-solving reduces employees' intention to violate the ISP, while wishful thinking increases their intention. Chen et al. [22] examined the relationship between the challenge aspect of information security and ISP compliance using questionnaires. The result indicated that challenge information security stress has a significantly positive influence on ISP compliance. In addition, challenge information security stress has a significantly positive influence on positive emotions and a significantly negative influence on negative emotions. Furthermore, positive emotions have a significantly positive influence on ISP compliance but negative emotions did not significantly influence ISP compliance.

Some studies employed a different approach by using lab-based experiments. Fordyce et al. [53] conducted a lab-based experiment to examine the link between stress and password choice. They recruited participants from university staff and students in the United Kingdom. The study was conducted over a period of two days, with the first day involving a manipulation designed to induce stress and requiring participants to ”choose” a password. On the second day, participants were informed that they needed to set a new password for their personal data under the pretext of a security incident, without receiving any manipulation. The manipulation used in the experiment involved two tasks designed to induce stress, namely the Serial Subtraction Task ([86]) and the isometric handgrip task ([96]). During the task, the participants were informed that the principal investigator of the study would review their results to make them more stressful, taking inspiration from the Trier Social Stress Test ([76]). The stress level of participants was measured using two instruments, namely the Short State Stress Questionnaire ([62]) and the State-Trait Anxiety Inventory ([135]). The strength of passwords was measured using the Password Guessability Service zxcvbn by Wheeler [154]. The results indicated no statistically significant difference in the mean zxcvbn password strengths. Davis et al. [30] also conducted a lab-based experiment to examine the effect of stress on secure application selection. Several manipulations were applied to the participants including no stress (control), loud crowd noise played through computer speakers, multi-tasking stress where participants had to switch every 30 seconds to another app and answer a question, and a time stress condition where the safe app choice must be made in a limited amount of time. The participants were asked to choose the safest app from three available apps in a fake app store. The results indicated that stress did not significantly affect accuracy in choosing a safe application. Another study by Trang and Nastjuk [143] used an in-basket experiment to measure the effect of time pressure on stress and information security compliance behavior. Participants were instructed to assume the role of an Acme company employee and respond to an email backlog. They were randomly assigned to one of two groups: a group with strict time constraints and a group with no time limit to complete the email task. The perceived time stress was assessed using four items, which were adapted from [32] and [95]. Information security compliance behavior was evaluated based on participants’ replies to the emails. Each of the four emails contained a request from a colleague to violate the Acme company’s information security policies. The findings of the study suggest that time constraints can induce stress and increase the likelihood of non-compliance with information security policies in the workplace.

Most research on the understanding of the relationship between stress and cybersecurity practices has primarily focused on developed countries such as Australia, the United Kingdom, South Korea, Croatia, China, and the United States. Only two studies were conducted on developing countries with one of them being our study in Indonesia [48] and the other was a study on oil and gas company workers from Malaysia [5]. Studying this topic in developing countries could be interesting to get a new perspective. Besides, there are no studies that address this topic in the hospital setting. Additionally, most of the studies on the human factor impact on cybersecurity practices use cross-sectional study methods so it is difficult to infer a causal effect. Only three studies used stress manipulation. Fordyce et al. [53] used stress manipulation in a lab-based experiment to examine the effect of stress on password selection while Davis et al. [30] conducted a lab-based experiment to examine the effect of stress on secure application selection. Trang and Nastjuk [143] used an online experiment to measure the effect of time pressure on stress and ISP compliance about sharing sensitive data. Studies using randomized experiments on other cybersecurity practices such as phishing email detection can be the future direction to infer the causal effect of stress on several particular cybersecurity practices.

1.4.2 The impact of stress and other human factors on phishing emails detection performance

In relation to phishing emails, to the best of our knowledge and through searches in peer-reviewed databases, there is no prior randomized experiment study specifically examining the role of stress on phishing email detection performance. Studies related to phishing emails include the role of time pressure and email load that may elicit stress. Both [18] and [74] found that when users were given less time, they made more mistakes in classifying the legitimacy of emails. Meanwhile, [124] suggested that a higher email load also decreases the email classification accuracy. Furthermore, Jensen et al. [70] examined the effect of mindfulness training on the phishing email response among university faculty, staff, and students. The result showed that the mindfulness approach to training significantly reduced the likelihood that participants responded to the phishing attack. However, these studies did not measure the stress level of the participants. Meanwhile, Abroshan et al. [1] used questionnaires and online experiments to examine the relationship between stress and phishing email detection ability. The results suggested a statistically significant relationship between stress and phishing email detection ability. However, this study did not have a stress manipulation so that it is difficult to infer a causal relationship. Given the limited studies on this topic, a causal analysis to examine the impact of stress on phishing email detection performance is needed to fill the gap.

Meanwhile, regarding human factors in general, numerous studies have investigated the role of human factors in detecting phishing emails. Several studies examined the link between phishing email susceptibility and demographic information such as age, gender, position, work experience, and training. Regarding age, the result was not always consistent. [109], [69], and [124] suggested that younger targets were more vulnerable to phishing email attacks. However, [138] found that older participants were more vulnerable. Regarding gender, [69] and [2] found that women were more likely to click on phishing links. However, [124] found no gender differences in email classification while [138] reported that female participants were less likely to report but found no effect of gender on click-through rates. In terms of work position, [138] found that managers were less likely to click on phishing emails than non-managers. With regard to work experience, [111] found that individuals with work experience were better at identifying phishing emails. Interestingly, they also reported that participants with formal information system training performed more poorly. In addition, [16] and [109] found that trust in technical measures, such as spam filters, was negatively related to the ability to detect phishing emails.

The decision-making style's effect on phishing email detection ability has also been frequently studied. [111] found that priming individuals about phishing risks led to a more diligent screening approach and better detection. [17] found that users who were less impulsive in decision-making were less likely to judge a link as safe in fraudulent emails. [2] found that high levels of general risk-taking increased the possibility of clicking on a phishing link. Additionally, [148] reported that systematic processing attenuated phishing susceptibility slightly, while heuristic processing and strong email habits made people tend not to read their email carefully and increased victimization significantly.

1.5 Research methodology

The research methodology in this study is depicted in Figure 1.10. The research problem about the impact of stress on cybersecurity practices was first defined. After literature reviews and gap identifications, research questions were formulated. Based on these research questions, the research hypothesis, study framework, and models were constructed. Then, three different methods were used to answer the research questions. Finally, recommendations and future works were extracted based on the results and findings.

1.5.1 Correlation Analysis

Correlation is a statistical technique utilized to evaluate the potential linear relationship between two continuous variables [103]. This study employed correlation analysis to investigate the relationship between stress

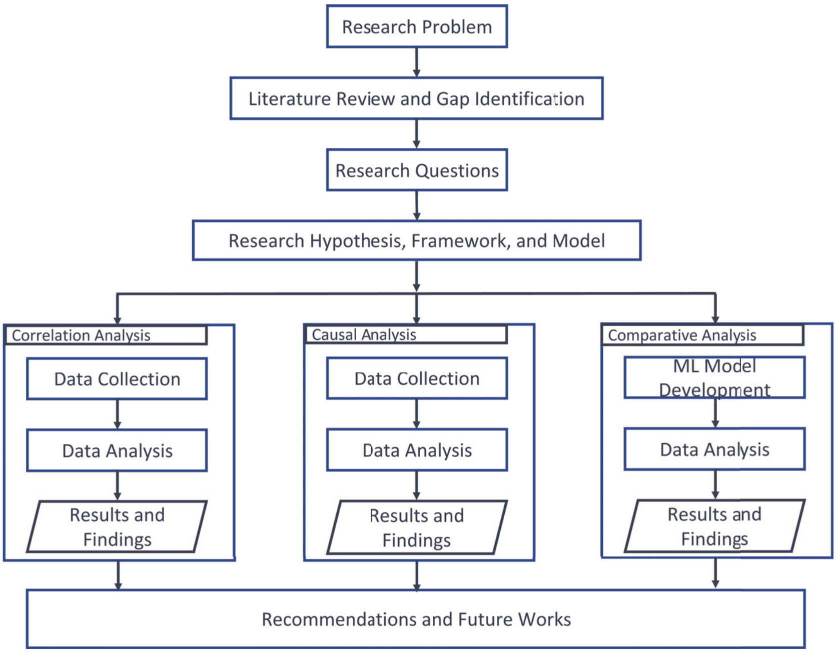


Figure 1.10: Research methodology

levels and risky cybersecurity practices among hospital workers as shown in Figure 1.11. Data were collected through a web-based questionnaire, including demographic information, stress levels, and cybersecurity practices over the past month. The Perceived Stress Scale (PSS) was utilized to assess the stress levels of the respondents, while Hospital Staff's Risky Cybersecurity Practices Scale (HS-RCPS) was used to evaluate the risky cybersecurity practices of hospital staff. PSS is a widely used and well-validated self-report questionnaire designed to measure an individual's perception of stress in their life [25]. PSS has demonstrated good reliability and validity in measuring perceived stress across various populations and settings [83, 6, 20, 91, 120, 10, 108]. The PSS items examine the degree to which individuals feel about their life, specifically how unpredictable, uncontrollable, and overloaded they find their lives during the last month. Rather than concentrating on specific experiences or events, the items in the questionnaire are general in nature [25]. Pearson's correlation coefficient was then used to evaluate the relationship between PSS and HS-RCPS. Additionally, other statistical analyses (t-test, ANOVA, etc.) were conducted to examine the relationship between individual differences (age, gender, work experience, etc.) and cybersecurity practices in this study.

This method was selected because it is inexpensive and allows us to

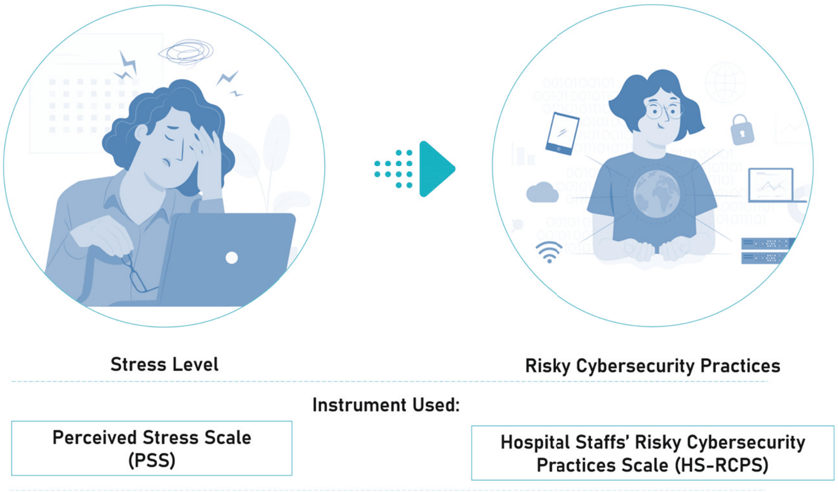

Stress Level
Risky Cybersecurity Practices
Instrument Used:
Perceived Stress Scale (PSS)
Hospital Staffs' Risky Cybersecurity Practices Scale (HS-RCPS)

Figure 1.11: Correlation analysis to investigate the relationship between stress levels and risky cybersecurity practices among hospital workers

reach participants from three countries in a short amount of time, despite being separated by great geographic distances [155, 141, 92]. This method also allows us to cover several aspects of cybersecurity practices such as password management, email usage, updating, etc. However, it is important to note that correlation coefficients indicate associations, not causal relationships [103].

The participants of this study were hospital staff from a hospital in Norway, three hospitals in Ghana, and a hospital in Indonesia. All the hospitals that were part of this study have adopted electronic health record (EHR) systems. Notably, the Norwegian hospital has been utilizing this system since the 2000s. while the four remaining hospitals, three Ghanaian hospitals and one Indonesian hospital, started using the system in the 2010s. Regarding security measures, the Norwegian hospital has established specific security policies for its staff. On the contrary, the Ghanaian and Indonesian hospitals currently don't have formal security policies for their staff.

1.5.2 Causal Analysis

Causal analysis entails the systematic examination of evidence to establish a cause-and-effect relationship between a specific treatment or intervention and the observed outcome [54]. Unlike the inference of association, causal inference goes beyond assessing the likelihood of events under static conditions and instead focuses on understanding the dynamics of events under changing conditions, such as changes induced by treatments or external interventions [113]. Randomized controlled trials (RCTs) are widely considered the gold standard for causal inference in medicine and social science [19].

This study employed a randomized controlled trial (RCT) to investigate the impact of stress on performance and completion time in email judgment. Participants were randomly assigned to either the control or intervention group, and their participation was voluntary with informed consent obtained from those who agreed to participate. The independent variable was stress (stress versus no stress), while the dependent variables included completion time and email judgment performance, measured by accuracy, sensitivity, and specificity.

Participants in the intervention group were exposed to the Purple Multitasking Framework (MTF), a computerized stressor developed by Purple Research Solutions in the UK [153] aimed to simulate the cognitive overload experienced during multitasking by presenting participants with four performance-based tasks simultaneously on a computer screen. Participants were instructed to achieve the highest possible score by completing the tasks quickly and accurately. Previous research has demonstrated the effectiveness of the Purple MTF in inducing cognitive demand, stress, and negative mood in participants [126, 75, 152]. All participants were asked to complete an email judgment test consisting of simulated phishing and legitimate emails. The emails were presented one at a time, with the order randomized at the beginning of the study and maintained consistently for each participant. Email judgment performance was assessed using three evaluation metrics: accuracy, sensitivity, and specificity. Additionally, the time taken by participants to complete the email judgment task was recorded. To evaluate the stress levels of participants, two self-report questionnaires were utilized: the State-Trait Anxiety Inventory for Adults (STAI-6) [93] and the Visual Analogue Scale (VAS). They were used because of their simplicity and ease of use. Both STAI-6 and VAS are relatively simple and easy to administer [119]. They involve straightforward instructions and can be quickly completed by participants without much training or assistance. This makes them convenient for researchers to implement in a lab setting. Besides, both measures have been widely used in psychological research, have undergone extensive validation to assess stress, and have been validated in numerous studies [90, 8, 122, 123, 29]. Additionally, a Shimmer3 galvanic skin response (GSR) sensor [15] was employed to measure participants' skin conductance, as it can indicate increased stress levels. Previous research by Jacobs et al. [68] has shown that healthy individuals exhibit increased skin conductance in response to stress.

The data analysis involved t-tests to compare stress level scores (VAS and STAI-6) between the non-stress and stress groups. Furthermore, t-tests were also performed to compare email judgment performance between the two groups, including measures of accuracy, sensitivity, specificity, and completion time. Pearson correlation analyses were conducted to explore the relationship between completion time and the three email judgment performance metrics. Moreover, paired t-tests were used to assess the statistical difference in skin conductance between baseline and after the stress intervention. Several statistical analyses (t-tests, ANOVA, Mann-Whitney U tests, etc.) were also carried out to examine the impact of individual differences (e.g., age, gender, phishing detection self-efficacy) on the email judgment task.

1.5.3 Comparative Analysis

In this part of the study, several machine learning models for stress detection based on multimodal sensor data were built using several strategies and then compared. The strategies compared include the strategy on how to combine the data from several sensors, how to combine several classifiers to build an ensemble method, and which preprocessing step need to be incorporated in order to improve the stress detection system. Furthermore, strategies on how to train the data considering the privacy of the users were also compared. The comparison especially focuses on user privacy and detection performance, such as accuracy, precision, recall, and F-1 measure. The experiment was conducted on the publicly available stress detection dataset called WESAD (Wearable Stress and Affect Detection) [125] due to its popularity for multimodal stress detection studies. The source code for the stress detection in this study can be found at `https://github.com/cahkanor/WESAD-Multiple-Sensor-Fusion-Stress-Detection` and `https://github.com/cahkanor/WESAD-Stress-Detection-Logistic-Regression`.

1.5.4 Ethical considerations

Prior to conducting this research, ethical approval was obtained from the Norwegian Centre for Research Data (NSD) and the Regional Committees for Medical and Health Research Ethics (REK) of Norway. Furthermore, since the research extended to Ghana and Indonesia, ethical clearance was obtained from each institution in both countries. Additionally, explicit permissions and informed consent were obtained from the healthcare facilities and individuals who participated in the study.

1.6 Summary of contribution

Various contributions have been made in an effort to answer the specified research questions as outlined in section 1.2. In the following section, the contribution of each part will be described.

1.6.1 List of included research publications

In total, ten (10) articles were included in this study from the four parts of the study to answer the five research questions. Four (4) papers were published for Part 1 (Correlation Analysis), one paper was published for Part 2 (Causal Analysis), four (4) articles were published for Part 3 (Effective Stress Detection), and one article was published for Part 4 (Privacy-Preserving Stress Detection). The following sections outlined the list of publications in the various parts of the study.

1.6.1.1 Part 1 (Correlation Analysis)

1. [48] Fauzi, M. A., Yeng, P., Yang, B., & Rachmayani, D. (2021, August). Examining the link between stress level and cybersecurity practices of hospital staff in Indonesia. In Proceedings of the 16th International Conference on Availability, Reliability and Security (pp. 1-8).

2. [50] Fauzi, M. A., Yeng, P., & Yang, B. (2023). Correlating Healthcare Staff's Stress Level and Cybersecurity Practices in Norway (In Proceedings of IEEE Conference on the Intelligent Methods, Systems, and Applications (IMSA)).

3. [51] Fauzi, M. A., Yeng, P. K., Yang, B., Nimbe, P., & Rachmayani, D. (2023). Stress and Cybersecurity Practices among Hospital Staff in the Digital Age: An Empirical Study from Ghana (Under review: IEEE Access).

4. [52] Fauzi, M. A., Yeng, P. K., Yang, B., Rachmayani, D., & Nimbe, P. (2023). Examining the Relationship Between Stress Levels and Cybersecurity Practices Among Hospital Employees in Three Countries: Ghana, Norway, and Indonesia. In Proceedings of the IEEE Annual Computers, Software, and Applications Conference (COMPSAC).

1.6.1.2 Part 2 (Causal Analysis)

5. [37] Fauzi, M. A., Yang, B., Katrien De Moor, K. D., Yeng, P. K., Rachmayani, D., Busch, C., & Wetherell, M. (2023). Can Stress Compromise Phishing Email Detection?. (Under review: Decision Support Systems).

1.6.1.3 Part 3 (Effective Stress Detection)

6. [40] Fauzi, M. A., & Yang, B. (2022). Multiple Sensor Fusion for Stress Detection in the Hospital Environment. In Proceedings of the 6th EAI International Conference on Computer Science and Engineering (COMPSE).

7. [46] Fauzi, M. A., Yang, B., & Yeng, P. (2022, November). Improving Stress Detection Using Weighted Score-Level Fusion of Multiple Sensor. In Proceedings of the 7th International Conference on Sustainable Information Engineering and Technology (pp. 65-71).

8. [47] Fauzi, M. A., Yang, B., & Yeng, P. K. (2022, September). Examining the Effect of Feature Normalization and Feature Selection for Logistic Regression Based Multimodal Stress Detection. In 2022 International Conference on Theoretical and Applied Computer Science and Engineering (ICTASCE) (pp. 90-94). IEEE.

9. [39] Fauzi, M. A., & Yang, B. (2021). Continuous stress detection of hospital staff using smartwatch sensors and classifier ensemble. In pHealth 2021 (pp. 245-250). IOS Press.

1.6.1.4 Part 4 (Privacy-Preserving Stress Detection)

10. [41] Fauzi, M. A., Yang, B., & Blobel, B. (2022). Comparative Analysis Between Individual, Centralized, and Federated Learning for Smartwatch Based Stress Detection. Journal of Personalized Medicine, 12(10), 1584.

1.6.1.5 List of additional research publications not included

1. [160] Yeng, P., Yang, B., Fauzi, M. A., Nimbe, P., Priharsari, D., & Priharsari, D. (2022, November). A Framework for Assessing Motivational Methods Towards Incentivizing Cybersecurity Practice in Healthcare. In Proceedings of the 7th International Conference on Sustainable Information Engineering and Technology (pp. 325-330).

2. [164] Yeng, P. K., Fauzi, M. A., Yang, B., & Nimbe, P. (2022). Investigation into Phishing Risk Behaviour among Healthcare Staff. Information, 13(8), 392.

3. [161] Yeng, P. K., Fauzi, M. A., Sun, L., & Yang, B. (2022). Assessing the Legal Aspects of Information Security Requirements for Health Care in 3 Countries: Scoping Review and Framework Development. JMIR Human Factors, 9(2), e30050.

4. [165] Yeng, P. K., Fauzi, M. A., Yang, B., & Yayilgan, S. Y. (2022, May). Analysing digital evidence towards enhancing healthcare security practice: The KID model. In 2022 1st International Conference on AI in Cybersecurity (ICAIC) (pp. 1-9). IEEE.

5. [166] Yeng, P. K., Nweke, L. O., Yang, B., Ali Fauzi, M., & Snekkenes, E. A. (2021). Artificial intelligence–based framework for analyzing health

care staff security practice: Mapping review and simulation study. JMIR medical informatics, 9(12), e19250.

6. [38] Fauzi, M. A., & Yang, B. (2021). Audiouth: Multi-factor authentication based on audio signal. In Proceedings of the Future Technologies Conference (FTC) 2020, Volume 3 (pp. 935-946). Springer International Publishing.

7. [45] Fauzi, M. A., Yang, B., & Martiri, E. (2021). PassGAN for Honeywords: Evaluating the Defender and the Attacker Strategies. In Advances on Smart and Soft Computing: Proceedings of ICACIn 2020 (pp. 391-401). Springer Singapore.

8. [163] Yeng, P. K., Fauzi, M. A., & Yang, B. (2020, December). Workflow-based anomaly detection using machine learning on electronic health records' logs: A Comparative Study. In 2020 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 753-760). IEEE.

9. [94] Martiri, E., Yang, B., & Fauzi, M. A. (2020, December). Indistinguishability of biometric honey templates: comparing human testers and SVM classifiers. In 2020 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 76-82). IEEE.

10. [162] Yeng, P. K., Fauzi, M. A., & Yang, B. (2020, December). Comparative analysis of machine learning methods for analyzing security practice in electronic health records' logs. In 2020 IEEE International Conference on Big Data (Big Data) (pp. 3856-3866). IEEE.

11. [44] Fauzi, M. A., Yang, B., & Martiri, E. (2020, July). Password guessing-based legacy-UI honeywords generation strategies for achieving flatness. In 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC) (pp. 1610-1615). IEEE.

12. [43] Fauzi, M. A., Yang, B., & Martiri, E. (2020, May). PassGAN based honeywords system for machine-generated passwords database. In 2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS) (pp. 214-220). IEEE.

13. [36] Fauzi, M. A., & Bours, P. (2020, April). Ensemble method for sexual predators identification in online chats. In 2020 8th international workshop on biometrics and forensics (IWBF) (pp. 1-6). IEEE.

14. [42] Fauzi, M. A., Yang, B., & Martiri, E. (2019, December). PassGAN-based honeywords system. In 2019 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 179-184). IEEE.

1.6.2 List of major contributions

In this section, the contribution made in each of the included papers is highlighted and presented in their respective parts of the study and in relation to how they answered the specified research question. The mapping of the research questions to the respective papers along with the key findings from each paper is listed in Table 1.2.

Table 1.2: Research questions (RQ), papers, and key findings

RQ	Paper	Key Findings
RQ1	1	The study finding revealed that in Indonesia, hospital workers' high stress levels correlate significantly with risky cybersecurity practices
	2	The findings indicated that in Norway, there was no significant correlation between stress levels and cybersecurity practices. The findings showed that individual differences including gender, age, position, and position level did not have a significant impact on healthcare professionals' risky cybersecurity activities. However, years of working experience were found to be a crucial factor in predicting cybersecurity practices among hospital employees with staff who had more than 25 years of work experience having the riskiest cybersecurity practices
	3	The findings from this study demonstrated a significant correlation between higher levels of stress among hospital employees and riskier cybersecurity practices. This research found a sig nificant difference in risky practices across different staff groups based on years of working experience. However, a significant difference in the practices was not observed across different staff groups based on other demographic factors such as gender, age, position, and position level

	4	The results base on the data combined from Indonesia, Norway, and Ghana indicated a statistically significant correlation between the stress levels of hospital staff and their engagement in risky cybersecurity practices. Specifically, the study finds that staff members' inclination to click on links from unknown sources is the cybersecurity practice most strongly influenced by stress levels
RQ2	5	The study found that Purple Multitasking Framework effectively increased participants' stress levels in both Indonesian and Norwegian participants. However, no significant difference in email judgment performance, including accuracy, sensitivity, and specificity, between participants in the non-stress and stress groups. Furthermore, the findings indicated that completion time may be a valuable measure of email judgment performance, as participants who took longer to analyze the emails significantly associated with better email judgment performance. While the result from Norway showed no significant difference, the result from Indonesia suggested that participants in the non-stress group took a significantly longer time to complete judging emails than participants in the stress group.
RQ3	6	The experiment results showed that all of the models of the multiple sensor fusion methods achieved better performance compared to all of the individual sensor models. The best result was achieved by score-level fusion with 0.844 of F$_{1$-measure and 0.921 of accuracy.
	7	The experiment results showed that multiple sensor fusion models obtained better performance than models from the individual sensor strategy. The weighted score-level fusion strategy achieved a better performance than the feature-level strategy with accuracy, precision, recall, and F$_{1$-measure of 0.931, 0.824, 0.939, and 0.868, respectively.

	8	The experiment results showed that the stress classification system with feature normalization performs better than without feature normalization. The stress detection system with Min-Max normalization got the best performance in terms of all evaluation metrics with accuracy, precision, recall, and F$_{1$-measure of 0.891, 0.814, 0.855, and 0.812, respectively. Meanwhile, the results of the feature selection experiment showed that the use of the fewest features gives the worst performance. The performance of the stress classification system increased as the number of features increases but the performance slightly declines at a particular point. The best performance was obtained when we used 90\% of the total features (378 features) with accuracy, precision, recall, and F$_{1$-measure of 0.894, 0.819, 0.859, and 0.817, respectively.
	9	The results showed that the ensemble method obtained higher performance compared to all of the individual classifiers. Furthermore, ES (soft voting) ensemble strategy had higher accuracy than EH (hard voting) strategy in this study but EH had a better F1-measure than ES.

RQ4	10	In terms of accuracy, the individual learning strategy beats both centralized learning and federated learning. In terms of privacy, centralized learning requires all of the data to be shared with a centralized server. There is a risk of privacy breach when the central server got compromised. In contrast, the individual learning strategy offers a very high level of privacy, since it does not require any user data or model to leave the user's device. Federated learning also offers a high level of privacy, since only the learned model, and no raw user data are processed in the central server. In terms of usability, individual learning has low usability for a new user. For centralized and federated learning, the new user can use the integrated model to infer her/his stress level right after registration. In contrast, for individual learning, the user must collect training data first to build the personalized model. In terms of the need for hardware and computational power, both centralized and federated learning need a server while individual learning does not. Furthermore, individual learning demands a user's device have enough computing power for feature extraction, model training, and stress detection tasks. Meanwhile, centralized learning requires less computing power for a user's device, because all of the processes can be done on the central server. However, the device has to be always online since the device has to send the data to the central server. Federated learning needs a user's device that has enough computing power to do the local training as well as a communication channel to exchange data between the device and the centralized server.

1.6.2.1 Correlation Analysis (Part I)

1. **Examining the link between stress level and cybersecurity practices of hospital staff in Indonesia:** This paper presented an empirical study to examine the link between stress levels and cybersecurity practices among hospital employees in Indonesia. The contributions in this paper include:

    - The study on the human factor impact on cybersecurity practices

   was limited in Indonesia. This is one of the earlier studies that assess the topic in Indonesia. Our study is among the first to consider stress as an important antecedent of cybersecurity practices. Existing research on stress and cybersecurity practices has primarily focused on settings other than hospitals. To the best of our knowledge and through searches in peer-reviewed databases, it is the first study assessing the relationship between stress and cybersecurity practices to Indonesia, especially in Indonesian hospitals. The study finding revealed that hospital workers' high stress levels correlate significantly with risky cybersecurity practices.
   - The study extends the limited use and validation of the Indonesian version of PSS to measure stress levels. Based on the survey results in this study, the Cronbach α of the PSS scale was 0.733, suggesting that the items in the scale have relatively good internal consistency.

2. **Correlating Healthcare Staff's Stress Level and Cybersecurity Practices in Norway:** This paper examined the relationship between stress levels and cybersecurity practices among hospital employees in Norway using questionnaires. In addition, the relationship between individual differences (e.g. gender, age, position, position level, and work experience) and cybersecurity practices was also analyzed. The contributions in this paper include:

   - The study extends the limited research on the understanding of the relationship between stress and cybersecurity practices. This paper is the first study assessing the relationship between stress and cybersecurity practices in Norway, especially in a Norwegian hospital. The findings indicated that there was no significant correlation between stress levels and cybersecurity practices.
   - This paper is also one of the earlier studies that used and validated the Norwegian version of PSS to measure stress levels. The survey result showed that the Norwegian PSS version had good reliability with a Cronbach's α of 0.844.
   - This paper enriches the study about correlating individual differences and cybersecurity behavior. The findings showed that individual differences including gender, age, position, and position level did not have a significant impact on healthcare professionals' risky cybersecurity activities. However, years of working experience were found to be a crucial factor in predicting cybersecurity practices among hospital employees with staff who had more than 25 years of work experience having the riskiest cybersecurity practices.

3. **Stress and Cybersecurity Practices among Hospital Staff in the Digital Age: An Empirical Study from Ghana:** This paper presented an empirical study from Ghana to examine the relationship between two factors, stress and individual differences, and cybersecurity practices among hospital staff. The contributions in this paper include:

    - This study extends the limited research on the understanding of the relationship between stress and cybersecurity practices. To the best of our knowledge and through searches in peer-reviewed databases, it is the first study in Africa about this topic. Existing research on this topic has primarily focused on developed countries such as Australia, the United Kingdom, and the United States. The findings from this study demonstrated a significant correlation between higher levels of stress among hospital employees and riskier cybersecurity practices.
    - This paper also examined the relationship between individual differences and cybersecurity practices. Since this topic is also limited in developing countries, this paper is one of the earlier studies to empirically assess this topic. This research found a significant difference in risky practices across different staff groups based on years of working experience. However, a significant difference in the practices was not observed across different staff groups based on other demographic factors such as gender, age, position, and position level.

4. **Examining the Relationship Between Stress Levels and Cybersecurity Practices Among Hospital Employees in Three Countries: Ghana, Norway, and Indonesia:** This paper presented an empirical study based on combined data from three countries to examine the relationship between two factors, stress and individual differences, and cybersecurity practices among hospital staff. In addition, this paper also compared cybersecurity practices between the three countries. The contributions in this paper include:

    - This study extends the limited research on the understanding of the relationship between stress and cybersecurity practices. This study assesses this topic by covering three countries on three continents. The results indicated a statistically significant positive correlation between the stress levels of hospital staff and their engagement in risky cybersecurity practices. Specifically, the study finds that staff members' inclination to click on links from unknown sources is the cybersecurity practice most strongly influenced by stress levels.

- The study did not observe any significant differences in cybersecurity practices based on gender, age, job, position level, or work experience.
- This study provides a comparison of cybersecurity practices between three countries. This study highlighted notable differences in cybersecurity practices across countries, with Norwegian hospital staff exhibiting better cybersecurity practices than their counterparts from Ghana and Indonesia.

1.6.2.2 Causal Analysis (Part 2)

5. **Can Stress Compromise Phishing Email Detection?:** This paper presented a randomized controlled trial to examine the influence of stress on performance and completion time in email judgment. The study involved the recruitment of participants from Norway and Indonesia, which resulted in a total of 150 participants. Participants were randomly allocated to either the control or intervention group. Stress served as the independent variable, while completion time and email judgment performance were the dependent variables. The contributions in this paper include:

   - Most of the studies on the human factor impact on cybersecurity practices use cross-sectional study methods so it was difficult to infer causal effects. To the best of our knowledge and through searches in peer-reviewed databases, it is the first study to use a Randomized Controlled Trial (RCT) to assess the impact of stress on phishing email detection ability. It is also one of the earlier studies that use RCT to study the impact of stress on cybersecurity practices. The study found that Purple Multitasking Framework effectively increased participants' stress levels in both Indonesian and Norwegian participants. However, no significant difference in email judgment performance, including accuracy, sensitivity, and specificity, between participants in the non-stress and stress groups.
   - This paper enriches the study about assessing the relationship between individual differences and phishing email detection ability. The findings indicated that almost there was no significant effect of individual differences on email judgment performance. Based on the data from Indonesia, only gender had a significant difference. Specifically, male participants in the non-stress group had significantly higher accuracy and sensitivity scores than female participants. In Norway, a significant correlation was found between age and both accuracy and sensitivity. The result showed

that the older the participants in Norway, the better the phishing detection ability. Data from Norway also suggested that education was significantly correlated with sensitivity with participants who completed doctoral degrees getting the highest score.

- This paper also assesses the impact of stress on phishing judgment task completion time and the relationship between completion time and phishing email detection performance. The findings indicated that completion time may be a valuable measure of email judgment performance, as participants who took longer to analyze the emails significantly associated with better email judgment performance. While the result from Norway showed no significant difference, the result from Indonesia suggested that participants in the non-stress group took a significantly longer time to complete judging emails than participants in the stress group.

1.6.2.3 Effective Stress Detection (Part 3)

6. **Multiple Sensor Fusion for Stress Detection in the Hospital Environment:** The most popular strategy to combine data from several sensors for machine learning-based stress detection is to combine them at the feature level. In this paper, we implement and propose some multiple sensor fusion strategies for stress detection using machine learning, including the combination of feature level, decision level, and score level. A comparative analysis of the stress detection performance of the strategies is provided. The experiment results showed that accelerometer sensor models had the best performance compared to other sensor models with 0.758 of F_1-measure and 0.866 of accuracy. The results also reported that all of the models of the multiple sensor fusion methods achieved better performance compared to all of the individual sensor models. The best result was achieved by score-level fusion with 0.844 of F_1-measure and 0.921 of accuracy.

7. **Improving Stress Detection Using Weighted Score-Level Fusion of Multiple Sensor:** This paper proposes a new method called weighted score-level fusion strategy to combine the data from several sensors in order to improve the machine learning-based stress detection performance. A comparison of the stress detection performance between the proposed method and established methods is provided. The experiment results showed that multiple sensor fusion models obtained better performance than models from the individual sensor strategy. The weighted score-level fusion strategy achieved a better performance than the feature-level strategy with accuracy, precision, recall, and F_1-measure of 0.931, 0.824, 0.939, and 0.868, respectively.

8. **Examining the Effect of Feature Normalization and Feature Selection for Logistic Regression Based Multimodal Stress Detection:** This study builds a multimodal-based stress detection system using a machine learning method and investigates the effects of feature normalization and feature selection on performance. A comparative analysis of the stress detection performance between several normalizations and feature selection methods is provided. The experiment results showed that the stress classification system with feature normalization performs better than without feature normalization. The stress detection system with Min-Max normalization got the best performance in terms of all evaluation metrics with accuracy, precision, recall, and F_1-measure of 0.891, 0.814, 0.855, and 0.812, respectively. Meanwhile, the results of the feature selection experiment showed that the use of the fewest features gives the worst performance. The performance of the stress classification system increased as the number of features increases but the performance slightly declines at a particular point. The best performance was obtained when we used 90% of the total features (378 features) with accuracy, precision, recall, and F_1-measure of 0.894, 0.819, 0.859, and 0.817, respectively.

9. **Continuous stress detection of hospital staff using smartwatch sensors and classifier ensemble:** This paper compares several classifiers for stress detection systems. It also gives a comparative analysis of the use of classifier ensembles to improve machine learning-based stress detection systems. The experiment results showed that all of the classifiers work quite well to detect stress with an accuracy of more than 70%. RF obtained the best accuracy with 86.61% while KNN had the lowest accuracy with 73%. In terms of F1-measure, LR achieved the best F1-measure with 76.25%. Similar to the accuracy result, KNN also had the lowest F1-measure (52.43%). The results also showed that the ensemble method obtained higher performance compared to all of the individual classifiers. Furthermore, ES (soft voting) ensemble strategy had higher accuracy than EH (hard voting) strategy in this study but EH had a better F1-measure than ES.

1.6.2.4 Privacy-Preserving Stress detection (Part 4)

10. **Comparative Analysis Between Individual, Centralized, and Federated Learning for Smartwatch Based Stress Detection:** This paper implemented several machine learning strategies for machine learning-based stress detection systems including individual, centralized, and federated learning. A comparative analysis of several aspects including accuracy, privacy, usability, and the need for computational power is provided.

- In terms of accuracy, the individual learning strategy beats both centralized learning and federated learning. This is quite reasonable because different participants may react differently to stressors, so a personalized model is needed. The integrated model aims to build a single model for all so that it cannot adjust for each user.
- In terms of privacy, centralized learning requires all of the data to be shared with a centralized server. There is a risk of privacy breach when the central server got compromised. In contrast, the individual learning strategy offers a very high level of privacy, since it does not require any user data or model to leave the user's device. Federated learning also offers a high level of privacy, since only the learned model, and no raw user data are processed in the central server.
- In terms of usability, individual learning has low usability for a new user. For centralized and federated learning, the new user can use the integrated model to infer her/his stress level right after registration. In contrast, for individual learning, the user must collect training data first to build the personalized model.
- In terms of the need for hardware and computational power, both centralized and federated learning need a server while individual learning does not. Furthermore, individual learning demands a user's device have enough computing power for feature extraction, model training, and stress detection tasks. Meanwhile, centralized learning requires less computing power for a user's device, because all of the processes can be done on the central server. However, the device has to be always online since the device has to send the data to the central server. Federated learning needs a user's device that has enough computing power to do the local training as well as a communication channel to exchange data between the device and the centralized server.

1.7 Discussion

In the first part of the study, we analyzed the relationship between stress levels and risky cybersecurity behavior of hospital staff in three countries: Indonesia, Norway, and Ghana. A statistically significant correlation was found between staff stress levels and their engagement in riskier cybersecurity practices based on the data from Ghana, Indonesia, and the combination of the three countries. This implies that individuals experiencing higher levels of stress are more likely to exhibit behaviors that compromise cybersecurity. However, no significant correlation was observed between these two factors based on the data from Norway. This can be because stress does not affect the cybersecurity practices of hospital workers in Norway. However, the number of respondents from Norway was too low. Hence, further studies in Norway need to be conducted. The overall result that suggests a significant correlation between hospital staff's stress levels and their engagement in riskier cybersecurity practices is consistent with the broader literature on the negative effects of stress on decision-making [101, 150]. This result was also supported by McCormac et al. [97] who found that workers with greater levels of stress had worse information security awareness (ISA). Furthermore, the finding that clicking on links from unknown sources was the riskiest cybersecurity behavior most influenced by stress levels is also reasonable since stress can harm an individual's cognitive functioning, impairing their ability to make rational decisions and increasing the likelihood of impulsive behavior [131]. In addition, stress can lead to feelings of anxiety or overwhelm, causing individuals to rush through tasks or pay less attention to details, making them more likely to overlook the signs of a phishing email [89].

From a practical perspective, these findings highlight the importance of addressing stress and well-being in the context of cybersecurity training and awareness programs. To be noted, addressing stress-related risky cybersecurity practices is an important effort that requires collaboration between employees, organizations, and other stakeholders in healthcare. Every stakeholder plays a pivotal role to achieve the goals. Employees should be encouraged to practice stress management techniques, take breaks, and report incidents promptly. Organizations should consider incorporating stress management techniques and well-being training into their cybersecurity training programs to help employees manage stress and reduce their engagement in risky cybersecurity practices. Having an understanding of how stress can influence an individual's cybersecurity practices, one can take measures to regulate their stress levels and maintain a heightened awareness of their cybersecurity practices. These measures could comprise tactics such as taking breaks to alleviate stress, exercising increased mindfulness with regard to cybersecurity practices while experiencing stress, and seeking assistance as necessary. Additionally, policies would also be needed to enforce the employee to adhere to good cybersecurity practices. However, enforcing individual rules solely through individual responsibility, particularly in a high-stress environment like a hospital, can indeed present challenges and potential drawbacks. While encouraging personal accountability is essential, relying solely on individuals to uphold cybersecurity measures can lead to various issues such as high workload. To address these challenges and avoid overwhelming hospital employees with undue responsibility, the organization should also show support and help such as creating some automated tools and alerts. For example, the organization can implement automated tools that scan and filter incoming emails for potential threats. These tools can flag suspicious emails, reducing the burden on individuals to identify threats manually. Furthermore, regular and tailored cybersecurity training sessions can equip hospital employees with the knowledge and skills needed to identify and respond to threats effectively. These sessions should be designed to fit into busy schedules and emphasize practical application.

Regarding demographic factors, they almost did not have any significant effect on healthcare staff cybersecurity practices. However, the research results revealed a significant difference in cybersecurity practices between healthcare professionals in Norway, a developed country, and those in Ghana and Indonesia, two developing countries. Developing nations have historically slowly adopted and utilized computer and internet technologies. As identified by Ben-David et al.'s research [13], developing nations' security landscape is affected by five fundamental factors: inadequate "security hygiene," unique resource constraints (such as one computer for multiple users), novice internet users, use of pirated software, and limited comprehension of cybersecurity adversaries. These factors could explain why people in developing countries generally exhibit poorer cybersecurity practices than their counterparts in developed nations. Insufficient IT education and a lack of computer and internet manuals in local languages have also contributed to unsafe cybersecurity practices [58]. Moreover, Norway's healthcare systems and infrastructure are comparatively advanced and better equipped to implement and enforce cybersecurity protocols than Indonesia and Ghana. Future research can investigate cultural factors and explore how they may be leveraged to improve cybersecurity practices in different regions.

In the second part of the study, the results from the lab-based experiment showed that stress did not significantly affect the participants' performance on the email judgment task. In both Indonesia and Norway, there were no significant differences between the stress and non-stress groups regarding accuracy, sensitivity, and specificity. A prior study on the effect of stress on a cybersecurity task also showed a similar result. [53] found no statistically significant difference in created password strengths between stressed and non-stressed participants. This finding is unexpected, as previous research has shown that stress can impair cognitive performance, including decision-making abilities ([101, 150]). However, we noted that individual differences in resilience and coping strategies may also have moderated the effect of stress on performance. There were variations in stress responses among participants in both the stress and non-stress groups, and this might have contributed to the lack of significant differences in email judgment performance. To further explore the relationship between stress and email judgment performance, we conducted a correlation analysis. The results indicated that although there were no significant differences in email judgment performance between the stress and non-stress groups, there was a significant negative correlation between stress level (indicated by STAI-6 score) and accuracy. This finding suggests that higher levels of stress may lead to lower accuracy in detecting phishing emails. A similar study by [116] on the effect of stress on medical students' clinical reasoning also showed that differences were found in performance in non-stress and stress conditions but correlational analyses revealed a negative correlation between multiple-stress measures and the one aspect of students' clinical ability due to the individual difference in stress responses to the stressor. Hence, based on this result, stress could not be fully ignored in terms of phishing email detection.

Regarding completion time, in the case of the Indonesian study, participants in the non-stress group took a significantly longer time to analyze the email before they judged them. A study by [7] stated that elevated stress was associated with higher levels of impulsivity in many cases could lead to faster decision-making. [4] also found that more stressed people finished the task of replying to emails faster. Furthermore, the completion time was significantly correlated with sensitivity across participants from Indonesia and correlated with both accuracy and sensitivity and sensitivity across participants from Norway. These results suggest that completion time may be a useful measure of email judgment performance. Participants who took longer tended to perform better in detecting phishing emails. One possible explanation for the finding is that participants who took more time to analyze an email may be engaging in a more careful and thorough evaluation of the email's contents. As a result, participants who take more time may be more likely to detect subtle clues that indicate an email is fraudulent, leading to higher accuracy and sensitivity scores. This result is consistent with a study by [18] that found that participants who were given a shorter time got lower scores in detecting phishing emails. Another study by [115] also suggested that dentists' diagnostic performance decreased when given less time.

Regarding individual differences, almost there was no significant effect of individual differences on email judgment performance. Based on the data from Indonesia, only gender had a significant difference. Specifically, male participants in the non-stress group had significantly higher accuracy and sensitivity scores than female participants. [129] also reported a similar result where females were more susceptible to phishing emails than males. However, the difference between gender was not observed when the participants were stressed. In Norway, a significant correlation was found between age and both accuracy and sensitivity. In contrast to data from Indonesia, where the age of the participants was uniform since almost all of them are undergraduate students, the ages of the participants in Norway tended to be more diverse. The result showed that the older the participants in Norway, the better the phishing detection ability. This was also reported by [112], who suggested that older adults in Australia were more risk-averse, with regard to their information security behaviors than younger adults. These results could be different if the study is conducted in developing countries because older people in developing countries tend to lack skills in using computers and the internet. Unfortunately, almost all of the participants from Indonesia were young adults so that we could not analyze this factor. Data from Norway also suggested that education was significantly correlated with sensitivity with participants who completed doctoral degrees getting the highest score. In line with these results, [57] reported that faculty/staff had better cybersecurity behaviors than students.

Furthermore, the result from both Indonesia and Norway showed that self-efficacy in phishing email detection had a significant positive correlation with email judgment. People with higher self-efficacy are usually equipped with the required skills to have the confidence to practice the appropriate security behavior. Some factors could influence the relationship between self-efficacy and performance, such as prior experience, knowledge, or training. [105] also found that self-efficacy was positively associated with good behavior in cybersecurity.

In the third part of this study, the comparative analysis of several strategies for effective multimodal stress detection systems showed that the multiple sensor fusion models exhibited superior performance compared to the individual sensor strategy. By combining some sensor data, richer information can be used to decide the stress levels. However, in terms of computational cost, the individual sensor strategy is better because less information means less processing time in terms of feature extraction and classification which can make the process faster. Another anticipated result is that the weighted score-level fusion strategy can outperform the feature-level strategy. This result is reasonable because stress is personal and each subject can have a different reaction to the stress. Hence, some sensors may be a good indicator of stress for one subject but maybe it is not the case for other others. In the feature-level fusion strategy, for each subject, we cannot give different weights for each sensor. In contrast, the weighted score-level fusion strategy enables us to be adaptive by giving more weight to the sensor model that can predict better on the subject data. Therefore, a weighted score-level fusion strategy can obtain a better result. Furthermore, the experiment results also demonstrated that the classifier ensemble method outperformed the individual classifiers. Generally, most individual classifiers have their own inherent defects [85] and their performance is also domain-dependent [78]. By combining some classifiers, the advantage of one classifier is expected to cover the shortcomings of other classifiers so that the performance can be improved.

Regarding feature normalization, the results show that the stress classification system with feature normalization performs better than without feature normalization. These results align with many other works that prove that feature normalization can improve the performance of classification tasks because this normalization process can reduce the variation of feature value range so that each feature contributes proportionally to the classification result. For the feature selection experiment, the results show that the use of too few features can lead to a bad performance because we remove too many important features. As the number of features increases, the performance also increases because more important features are employed. However, at some points, increasing the number of features will reduce performance because there are some features involved that have a bad contribution to the stress classification result. Therefore, finding the suitable number of features used and which features should be used is important to improve the classification performance. Besides, by using feature selection, the computational complexity of the stress classification task can be reduced because fewer features need to be processed.

In the last part of the study, a comparison of individual learning, centralized learning, and federated learning dataset was discussed. Generally, more data will make the machine learning model better and more accurate, because the more information we give to the model, the more it will learn and the more cases it will be able to correctly infer [137]. Therefore, integrated models such as centralized and federated learning are expected to be more accurate than individual learning. Surprisingly, the individual model surpasses in this study both centralized and federated learning. The WESAD dataset labels the data based on the stimulus given to the participants. Different participants may react differently to each stimulus. In this case, the personalized approach such as the individual learning model can adjust the model to the user's behavior. The integrated model aims at building a single model for all so that it cannot adjust for each user. This study outcome is in line with another study about stress detection that also reported that a personalized model outperformed an integrated model [88].

Generally, federated learning is expected to perform worse than centralized learning. It is because centralized learning has direct access to all data while federated learning trains the model locally and only communicates an updated model to a central server [106]. Surprisingly, the performance difference between the two strategies is very big. A more complex model such as Deep Neural Network (DNN) is needed to build a better federated learning model. Some previous work shows that federated learning with DNN can obtain performance levels comparable to those models trained using a centralized learning scheme [106, 88]. Another study also suggested that less complex models perform worse than more complex models in federated learning [133]. However, a more complex model requires the user's device to have a higher computational power to train the model. Additionally, a more complex model will also lead to higher communication costs between the user's device and the central server. Thus, there will be a challenge to use a complex model for communication-sensitive applications [133].

Another factor that can also be considered is the usability of the three learning schemes for a new user. For centralized and federated learning, the new users can use the integrated model to predict their stress level right after the registration. For individual learning, however, the user must collect training data first. The users should record their data using the smartwatch during stress and non-stress condition. The users must also give the correct label to the data because the quality of the model heavily depends on the training data quality. This training data is used to train the personalized model for the users before they can infer their stress level automatically.

In addition, the computational cost is also different between these three schemes. Individual learning demands that a user's device has enough computing power for feature extraction, model training, and stress detection tasks. Meanwhile, centralized learning requires less computing power for a user's device, because all of the processes can be done on the central server. However, the device has to be always online since the device has to send the data to the central server. Federated learning needs a user's device that has enough computing power to do the local training as well as a communication channel to exchange data between the device and the centralized server.

Finally, stress data are considered sensitive as they can be used to disclose the user's health status. Based on a study on health data privacy, most of the interview subjects are worried about their data privacy on an individual level [31]. Therefore, the processing of this kind of data needs to pay more attention to privacy concerns. In centralized learning, all the data are collected on a centralized server. When these data are shared with the central server, privacy leaks can occur if the central server is compromised. Therefore, centralized learning can jeopardize users' privacy. On the contrary, individual and federated learning strategies offer a high level of privacy. In federated learning, only the learning model, and no raw user data, is processed centrally. Meanwhile, individual learning provides a higher level of privacy as it does not require any user data or model to leave the user's device.

Federated learning can protect raw sensitive data. However, the trained model needs to be sent to the server in this learning strategy. Even though it is not a straightforward step, it is possible to reconstruct or approximate the training data used to train the ML model. Numerous techniques have been designed to address the extraction of sensitive information from trained models. One approach is model inversion which uses the outputs or predictions of the trained model to infer or reconstruct inputs that would likely produce those outputs [157]. Another approach is GradInversion which is able to reconstruct individual images in a batch, given averaged gradients

[167]. Another study proposed a reconstruction scheme based on the implicit bias in training neural networks with gradient-based methods [60]. To be noted, studies only suggested that the reconstruction is only possible if the model is based on a neural network and still there are several limitations. The trained model using traditional ML methods (e.g. logistic regression) is still safe from the data reconstruction techniques.

1.8 Recommendations for individuals, organizations, and related stakeholders based on the study findings

The proposed recommendations encompass a diverse range of suggestions aimed at employees, organizations, and other stakeholders in healthcare. The recommendations stem from the understanding that every stakeholder plays a pivotal role to achieve the goals. Therefore, collaboration and collective efforts among all stakeholders are important.

1.8.1 Correlation analysis between stress and cybersecurity practices

The following are the recommended actions for hospital employees based on the research findings on stress levels and risky cybersecurity behavior in order to better manage stress, enhance their cybersecurity awareness and practices, and contribute to creating a more secure work environment:

1. Recognize the impact of stress on cybersecurity: Based on the study findings, the high-stress level is significantly correlated with risky cybersecurity practices. Hospital employees should be aware of the potential influence of stress on their cybersecurity practices. Understand that high-stress levels can impair decision-making and increase the likelihood of engaging in risky online behaviors.

2. Stay vigilant during stressful times: When experiencing high levels of stress, hospital staff should be extra cautious and maintain a heightened awareness of their cybersecurity practices. The research findings showed that staff with high-stress levels tend to click a link from an unknown sender. The findings also suggested that clicking on links from unknown sources was the riskiest cybersecurity behavior most influenced by stress levels. Therefore, they should take their time to review emails, messages, and online content carefully before clicking on links or providing sensitive information, especially during stressful times.

3. Prioritize stress management: Hospital employees should take proactive steps to manage stress effectively. Practice stress reduction tech-

niques such as deep breathing exercises, mindfulness meditation, physical activity, and taking regular breaks to recharge and maintain mental well-being.

4. Engage in cybersecurity training and education and practice good cybersecurity hygiene: Hospital staff should participate actively in cybersecurity training and awareness programs provided by the hospital. They also should adhere to established cybersecurity protocols and guidelines within the hospital in order to have better cybersecurity practices.

5. Seek support and report concerns: Hospital staff should seek support from their supervisor, IT department, or designated cybersecurity personnel if they feel overwhelmed by stress or encounter suspicious emails, messages, or cybersecurity incidents. They are also recommended to report any potential security threats promptly and follow the established incident response protocols.

6. Engage in ongoing self-assessment: The staff should reflect on their own cybersecurity practices and identify areas for improvement. They are recommended to continuously assess their own stress levels and consider implementing personal stress management strategies to minimize the impact on their cybersecurity practices.

The following are the recommended actions for hospital management based on the research findings on stress levels and risky cybersecurity practices:

1. Assess and address stress levels: Hospital management should conduct regular assessments of stress levels among hospital staff and identify factors contributing to stress, such as workload, organizational culture, and job demands. Implement strategies to reduce stress, such as workload management, promoting work-life balance, and fostering a supportive work environment.

2. Integrate stress management into cybersecurity training: Hospital management should incorporate stress management techniques and coping strategies into cybersecurity training programs. They are also recommended to provide staff with resources and education on stress reduction, resilience-building, and self-care practices to help them manage stress effectively while maintaining secure cybersecurity practices.

3. Increase awareness of the impact of stress on cybersecurity: The management should educate hospital staff about the relationship between stress levels and risky cybersecurity behavior. They are recommended

to make a guideline that highlights the potential consequences of stress-related lapses in cybersecurity and emphasizes the importance of maintaining vigilance even during stressful periods.

4. Provide ongoing training and reinforcement: The hospital management should offer regular training sessions and refreshers on cybersecurity best practices, emphasizing the role of stress management in maintaining secure behaviors. They could provide resources such as tip sheets, posters, and newsletters to reinforce key messages and keep cybersecurity practices at the forefront of employees' minds.

5. Regularly evaluate cybersecurity practices: The management should continuously monitor and evaluate the effectiveness of cybersecurity practices within the organization to identify areas for improvement and track progress in reducing risky behaviors influenced by stress.

6. Establish comprehensive cybersecurity policies and guidelines: To enhance the implementation of cybersecurity practices at the individual level, it is strongly recommended that organizations establish comprehensive policies and guidelines. These policies should outline the specific cybersecurity practices that individuals are expected to follow in their daily operations. By creating such policies, organizations can provide a clear framework for employees and stakeholders to understand their responsibilities and the necessary steps to mitigate potential security risks.

7. Engage developers or system providers to enhance security: A holistic approach to mitigating the risks associated with phishing emails necessitates the active involvement of developers and user interface designers. They should recognize the demanding nature of hospital environments and acknowledge that employees often operate under time pressure by designing applications with streamlined workflows that do not exacerbate stress. The goal is to enhance cybersecurity without imposing additional cognitive load. Interfaces should be intuitive, minimizing the need for extensive training or adaptation.

Furthermore, the study found a significant disparity in cybersecurity practices between healthcare professionals in Norway, a developed country, and those in Ghana and Indonesia, two developing countries. Developing nations have traditionally been slower in adopting and utilizing computer and internet technologies. Ben-David et al.'s research [13] identifies five key factors that affect the security landscape in developing nations: inadequate "security hygiene," resource limitations (such as multiple users sharing a single computer), novice internet users, use of pirated software, and limited understanding of cybersecurity threats. These factors may contribute to the lower cybersecurity practices observed in developing countries compared to their counterparts in developed nations. Insufficient IT education and the absence of computer and internet manuals in local languages further contribute to unsafe cybersecurity practices [58]. Additionally, Norway's advanced healthcare systems and infrastructure provide better capabilities for implementing and enforcing cybersecurity protocols compared to Indonesia and Ghana. Making significant strides for developing countries in improving their cybersecurity practices in the healthcare sector requires a multi-faceted approach that combines education, capacity building, regulatory frameworks, infrastructure development, and international collaboration. The following are some recommendations for developing countries, especially Ghana and Indonesia:

1. Enhance cybersecurity education and awareness: Develop and implement comprehensive cybersecurity education programs targeted at healthcare professionals and the general population. Increase awareness about cybersecurity risks, best practices, and the potential consequences of poor cybersecurity practices. Translate educational materials into local languages to ensure accessibility.

2. Strengthen IT education and training: Improve IT education and training programs to enhance the digital literacy and technical skills of healthcare professionals and other relevant stakeholders. Provide training on basic computer and internet usage, safe browsing habits, and secure handling of sensitive data.

3. Establish local cybersecurity frameworks and regulations: Develop and enforce cybersecurity frameworks and regulations tailored to the specific needs and challenges of the country. Collaborate with government agencies, industry stakeholders, and international partners to establish standards, guidelines, and compliance requirements for cybersecurity in the healthcare sector.

4. Invest in secure infrastructure and technologies: Allocate resources to invest in secure IT infrastructure, hardware, and software solutions. Prioritize the adoption of licensed and up-to-date software to minimize vulnerabilities associated with the use of pirated software.

5. Build local cybersecurity expertise: Develop local cybersecurity talent by investing in training programs, certifications, and professional development opportunities. Nurture a pool of skilled cybersecurity professionals who can contribute to safeguarding healthcare systems and networks.

6. Foster public-private partnerships: Encourage collaborations between the public and private sectors to address cybersecurity challenges ef-

fectively. Foster partnerships that leverage the expertise and resources of both sectors to develop innovative solutions, share knowledge, and promote cybersecurity awareness.

7. International collaboration and support: Seek international collaboration and support from organizations, governments, and agencies that specialize in cybersecurity capacity building. Engage in partnerships that provide technical assistance, training, and knowledge sharing to enhance cybersecurity practices.

1.8.2 Causal analysis between stress and phishing email detection performance

The following are the recommendations for individuals based on the research findings on the impact of stress on phishing email detection performance:

1. Enhance phishing email knowledge: The study findings indicate that participants tend to have difficulty detecting phishing emails. They should take the initiative to educate themselves about cybersecurity best practices, especially related to phishing emails.

2. Develop self-efficacy in cybersecurity: They should build confidence in their ability to detect and respond to cybersecurity threats by improving their skills through practice and learning. The are recommended to seek out resources, online courses, or workshops that can help enhance their cybersecurity knowledge and capabilities.

3. Manage stress levels: They should recognize the potential impact of stress on cybersecurity practices. Develop strategies to effectively manage stress, such as practicing mindfulness, engaging in regular physical exercise, and taking breaks when feeling overwhelmed. Prioritize self-care to maintain cognitive functioning and make rational decisions.

4. Prioritize accuracy over speed, be thorough, and avoid impulsive decisions: The study findings suggested that the time taken to analyze the email has a significant positive correlation with email judgment performance. When analyzing emails, especially those that seem suspicious or require action, it is strongly recommended to avoid rushing through the process. They are recommended to allocate sufficient time to carefully evaluate the email's content, sender information, and any attachments or links. Slowing down can help them identify subtle indicators of phishing attempts and make more accurate judgments.

The following are the recommendations for organizations or workplace management based on the research findings on the impact of stress on phishing email detection performance:

1. Improve employees' phishing email detection skills: Provide training and educational programs to enhance employees' ability to detect phishing emails.

2. Address the impact of stress: They should recognize the impact of stress on employees' phishing email susceptibility. Integrate stress management techniques into training programs to help employees effectively manage stress levels. The workplace should provide resources and support systems to assist employees in coping with stress and maintaining their cybersecurity vigilance.

3. Allow Sufficient Time: Employers should encourage employees to allocate an appropriate amount of time to analyze and evaluate emails, particularly those that may be potentially fraudulent or phishing attempts. Rushed decision-making due to time constraints can lead to lower accuracy and sensitivity in detecting phishing emails.

4. Provide prior experience, knowledge, and training: Offer opportunities for employees to gain prior experience, knowledge, and training in cybersecurity. Foster self-efficacy by equipping participants with the necessary skills and confidence to practice appropriate security behaviors.

1.8.3 Comparative analysis on machine learning based stress detection system

Based on the results of the study on stress classification using smartwatch sensor data, the following recommendations can be made:

1. Utilize Multiple Sensor Fusion: Employing a fusion strategy that combines data from multiple sensors can improve the accuracy and performance of stress classification systems. The study found that fusion strategies outperformed the individual sensor strategy. Therefore, researchers developing stress classification systems should consider incorporating data from multiple sensors to enhance the system's effectiveness.

2. Assess Weighted Score-Level Fusion Strategy: The study reported that the weighted score-level fusion strategy achieved better performance than the feature-level strategy. Researchers developing stress classification systems should consider exploring and implementing weighted

score-level fusion techniques to improve the effectiveness of their systems.

3. Utilize Feature Normalization: The study found that incorporating feature normalization in the stress classification system leads to better performance compared to without feature normalization. Feature normalization reduces the variation in feature values, ensuring that each feature contributes proportionally to the classification result. Therefore, it is recommended to include feature normalization techniques to achieve optimal results.

4. Select an Optimal Number of Features: The results of the feature selection experiment indicated that using the fewest features resulted in the worst performance. The performance of the stress classification system generally improved as the number of features increased, reaching a peak at a certain point before slightly declining. Based on the study, the best performance was achieved when utilizing 90% of the total features. However, the number of optimal features can be different for different cases. Therefore, it is recommended to carefully select and include an appropriate number of features in the stress classification system to achieve optimal performance.

5. Consider Individual Learning for Performance and Privacy Aspect: The study found that the individual learning model for stress detection surpassed both centralized and federated learning in stress detection. This suggests that a personalized approach, such as the individual learning model, may be more effective in adjusting to users' behavior and improving performance. Therefore, organizations should consider implementing individual learning models for stress detection systems. Besides, individual learning is also able to preserve user data privacy by not sharing data with others. However, individual learning cannot fully accommodate usability because it requires data collection from each user and training before the stress detection model can be built. In contrast, centralized and federated learning allow new users to use the integrated model immediately after registration. Individual learning also needs a personal device with enough computational power to train the model.

6. Explore Complex Models for Federated Learning: While federated learning is generally expected to perform worse than centralized learning, the study revealed a significant performance difference between the two strategies. To enhance the performance of federated learning models, the use of more complex models like Deep Neural Networks (DNNs) is recommended. However, it is important to consider the

computational power and communication costs associated with complex models in federated learning.

1.9 Limitations

There are several limitations of this study that need to be acknowledged. For the first part of the study, firstly, the sample size from Norway was relatively small, which may limit the generalizability of our findings. Second, the p-value to conclude the significance was quite high in some of the analyses. Furthermore, despite the fact that anonymity was ensured, it is still possible to have a desirability bias since we utilized a self-reported questionnaire to gather the data. Participants may provide answers that they believe are socially desirable or that they think the researcher wants to hear, rather than providing honest answers. Besides, memory bias could also occur when participants have trouble remembering details correctly, particularly if the details relate to previous events or behaviors. In addition, by using the rating scales, this study may also suffer from central-tendency bias where people frequently hesitate to provide excessive reactions and often lean toward the middle. To minimize the user doing random clicking and ensure the participants really read the questions. we utilized attention-checking questions.

For the second part of the study, the lab-based experiment also had some limitations, among which the foremost is the predominantly university student sample utilized for recruitment. Consequently, the outcomes of the study may not be generalizable to the broader population, thereby limiting the generalizability of the findings. Furthermore, while lab-based phishing email tests can provide valuable insights into people's susceptibility to phishing attacks, the study setting may not reflect the real-world context in which people encounter phishing emails. Participants in a lab setting may be aware that they are participating in a study and maybe more vigilant or cautious than they would be in their everyday lives. This can influence their behavior and responses to the phishing email test, potentially leading to a less accurate representation of their true behavior and response patterns in real-world settings.

In this study, we explored two distinct approaches for stress level detection: utilizing self-report questionnaires (PSS, STAI-6, and VAS) and leveraging machine learning models on smartwatch sensor data. Questionnaires have been extensively used to assess subjective stress levels due to their ease of administration. The PSS, STAI-6, and VAS are well-established tools that provide direct self-report data on an individual's perceived stress levels. Many studies have reported the reliability and validity of the three scales. One of the key advantages of questionnaires lies in their ability to capture the cognitive and emotional aspects of stress experienced by the user. However, questionnaire-based stress detection comes with limitations. Self-report measures are subjective and may be influenced by various biases, such as social desirability or recall bias. The reliance on user awareness and willingness to report stress levels accurately might lead to potential inaccuracies. Furthermore, questionnaires are not suitable for real-time monitoring of stress levels and also require people to get out of their daily routine activities.

Leveraging machine learning models on smartwatch sensor data offers a promising objective approach to stress level detection. The advantage of smartwatches lies in their ability to capture continuous and real-time physiological data, such as heart rate, skin conductance, and physical activity, which are indicative of stress-related responses. Machine learning models can analyze and interpret this data to make stress level predictions, potentially providing personalized stress assessment for individuals. However, using smartwatch sensor data for stress detection presents its own set of challenges. First, data privacy and ethical concerns arise due to the collection of sensitive physiological information from users. Additionally, the cost and availability of wearable sensors like smartwatches may limit their use in certain populations. Furthermore, one limitation of machine learning-based stress detection using smartwatch sensor data is its potential inability to capture the context surrounding the user's physiological responses. For instance, physical activity can significantly influence skin conductance and heart rate, leading to variations in the captured physiological data that might be misinterpreted as changes in stress levels. When a user engages in physical activities such as exercise or strenuous movements, their heart rate and skin conductance levels may increase, irrespective of their stress levels. This could result in false positive stress predictions by the machine learning model. Besides, ML models used for stress detection based on physiological sensor data are their inability to directly identify the psychological strain. These models primarily rely on the analysis of physiological responses such as heart rate and skin conductance to infer the presence of stress. However, stress is a multifaceted construct that encompasses both physiological and psychological components, and the machine learning models are limited to capturing the physiological aspect. While physiological responses like increased heart rate or skin conductance can indicate an activation of the body's stress response system, they do not inherently capture the cognitive and emotional aspects of stress. Psychological strain can manifest as feelings of overwhelm, anxiety, or unease that might not necessarily manifest in easily measurable physiological changes.

To improve stress level detection, researchers may consider integrating data from both questionnaires and ML models based on wearable sensor data. The complementary nature of these approaches can lead to a more comprehensive understanding of an individual's stress experience. While questionnaires provide subjective insights, wearable sensor data offer objective physiological information, combining to create a holistic stress profile.

1.10 Recommendation for future work

1.10.1 Stress and cybersecurity practices

In future studies, there are several potential research directions that can be pursued to further explore the relationship between stress levels and risky cybersecurity practices in healthcare organizations. These studies can provide valuable insights and contribute to the development of effective strategies for stress management and improving cybersecurity behaviors.

Firstly, conducting a longitudinal study would be beneficial in establishing a causal relationship between stress levels and risky cybersecurity practices over time. By measuring stress levels and cybersecurity behavior at multiple time points, researchers can examine changes over time and identify any potential cause-effect relationships. This longitudinal approach would provide more robust evidence on the impact of stress on cybersecurity practices and help in identifying specific stressors or periods of heightened vulnerability.

Another important area for future research is to explore coping mechanisms used by hospital staff in response to stress and how these strategies relate to their cybersecurity practices. Understanding the coping mechanisms employed by individuals in high-stress environments can provide insights into effective stress management techniques that can be incorporated into training programs. This can include investigating the role of mindfulness, relaxation techniques, social support, and other stress reduction strategies in improving cybersecurity practices.

Assessing the effectiveness of stress management interventions within healthcare organizations is another promising avenue for future research. Implementing stress management interventions and evaluating their impact on reducing stress levels and improving cybersecurity practices can provide practical insights and evidence-based recommendations. These interventions can include mindfulness training programs, relaxation techniques, employee support programs, and other stress reduction initiatives. Such studies can help inform organizational policies and practices aimed at promoting employee well-being and cybersecurity awareness.

To broaden the understanding of the relationship between stress levels and risky cybersecurity practices, future research can extend beyond the healthcare sector and conduct comparative studies across industries. By comparing different sectors, researchers can determine if the relationship between stress and cybersecurity practices is specific to healthcare or if it applies more broadly. This comparative approach can provide valuable insights into the commonalities and differences in stress-cybersecurity dynamics across various organizational contexts.

Additionally, exploring technological solutions to mitigate the impact of stress on risky cybersecurity practices can be a fruitful area of investigation. Researchers can explore the potential of intelligent email filtering systems, behavior monitoring tools, or other technological interventions to help identify and prevent employees from falling victim to phishing attempts and other cybersecurity threats. Assessing the effectiveness of such technologies in reducing the negative impact of stress on cybersecurity practices can inform the development and implementation of security measures in healthcare organizations.

Furthermore, future research can delve into the role of personal resilience in mitigating the impact of stress on cybersecurity practices. Investigating how individual differences in resilience affect employees' ability to manage stress and make secure decisions in the face of cybersecurity threats can provide valuable insights for intervention and training programs. Understanding the protective role of personal resilience can help in developing targeted strategies to enhance employees' ability to cope with stress and make secure choices in the digital environment.

Finally, a causal analysis of how stress impacts individuals' ability to detect and respond to other types of cyber threats, such as other social engineering attacks or malware infections, could also be interesting future work.

1.10.2 Stress detection system

In future work on stress detection systems, several areas can be explored to enhance their performance and applicability. First, the optimization of federated learning should be a focus, aiming to narrow the performance gap between federated learning and centralized learning observed in the study. Advanced techniques such as improved aggregation methods, model compression, and secure computation protocols can be investigated to enhance the accuracy and efficiency of federated learning models.

Second, there is a need to further explore personalized approaches to stress detection. Research should investigate methods to adapt the model to individual users' behavior, preferences, and physiological responses, leading to improved accuracy and user satisfaction. Developing techniques for dynamic model updating and personalization over time can enhance the long-term performance of stress detection systems.

Third, collecting stress data across multiple sessions can provide valuable insights into the system's performance in different contexts and scenarios. Understanding how the stress detection system performs over time and adapting the model accordingly can improve its robustness and generalizability.

Privacy preservation is a crucial aspect to consider in stress detection systems. Future research should focus on developing robust privacy-preserving techniques that enable accurate stress detection while safeguarding users' privacy. Exploring methods such as secure multiparty computation, homomorphic encryption, and differential privacy can contribute to the development of privacy-enhancing stress detection systems.

Lastly, conducting field studies and real-world deployments of stress detection systems is essential for evaluating their practical effectiveness, user acceptance, and integration into everyday life. Evaluating the performance and user experience in real-world settings will help identify potential challenges and opportunities for improvement, ensuring that stress detection systems meet the needs and expectations of users in practical applications.

1.11 Conclusion

This research study examined the relationship between stress levels and risky cybersecurity practices among hospital workers and investigates the impact of stress on email judgment performance. The study also compares different strategies for effective multimodal stress detection systems. The research methodology includes correlation analysis, causal analysis using a randomized controlled trial (RCT), and comparative analysis of machine learning models.

First, a correlation analysis was conducted to investigate the relationship between stress levels and risky cybersecurity practices among hospital workers in three countries: Ghana, Norway, and Indonesia. In total, 353 qualified participants were finally included in the study, with 212 participants from Ghana, 42 from Norway, and 99 from Indonesia. A statistically significant positive correlation was found between staff stress levels and their engagement in riskier cybersecurity practices based on the data from Ghana, Indonesia, and the combination of the three countries. However, no significant correlation was observed between these two factors based on the data from Norway. This implies that individuals experiencing higher levels of stress are more likely to exhibit behaviors that compromise cybersecurity. Specifically, the staff's tendency to click on links from unknown sources was found to be the risky cybersecurity practice most heavily associated with higher stress levels. The findings of this study have important implications for hospital management. Understanding the relationship between stress levels and cybersecurity practices can serve as a foundation for improving the effectiveness and efficiency of cybersecurity measures within healthcare organizations. By identifying the factors that influence cybersecurity practices among hospital employees, management can design targeted training programs, awareness campaigns, and support systems to enhance cybersecurity awareness and practices. These interventions can help mitigate the risk of cyberattacks, protect patient privacy and data, and promote a culture of cybersecurity within healthcare settings. Additionally, variations in risky cybersecurity practices were observed across staff groups based on the country of origin. The study findings revealed a significant difference in cybersecurity practices between healthcare professionals in Norway and those in Ghana and Indonesia. This suggests that cultural and contextual factors may play a role in shaping cybersecurity practices among hospital staff. Developing nations have historically slowly adopted and utilized computer and internet technologies and do not place cybersecurity as their main priority.

Second, a causal analysis was conducted using a randomized controlled trial (RCT) to investigate the impact of stress on performance and completion time in email judgment in two countries, Indonesia and Norway. In total, 150 participants participated in our study, with 100 participants from Indonesia and 50 participants from Norway. The results revealed that participants in both the non-stress and stress groups exhibited difficulty in detecting phishing emails, with an average accuracy rate of approximately 60%. However, the randomized controlled trial (RCT) analysis showed no significant difference in email judgment performance between the two groups, suggesting that stress did not directly compromise participants' ability to detect phishing emails. Nevertheless, correlation analysis conducted specifically with Indonesian participants revealed a noteworthy finding. It demonstrated that higher levels of stress were significantly correlated with lower accuracy in detecting phishing emails. This suggests that stress may have an impact on email judgment performance in the Indonesian context. Additionally, the study explored the completion time as a potential measure of email judgment performance. Interestingly, participants who took longer to complete the task tended to perform better. Based on the data from Indonesia, participants in the non-stress group took significantly more time to judge emails compared to those in the stress group. However, in Norway, no significant difference in completion time was observed between the two groups. Furthermore, the study found that individual differences had minimal impact on email judgment performance, except for gender in the Indonesian sample. Male participants exhibited significantly higher accuracy and sensitivity scores compared to their female counterparts. This study suggests that completion time could serve as a valuable measure of email judgment performance. The findings derived from both correlation and causal analyses unequivocally underscore the formidable challenge of ascertaining the exact impacts of stress on behavior within the intricate web of confounding factors. These outcomes are further corroborated by the evidence obtained from the comprehensive literature review. Furthermore, it is crucial to acknowledge the expensive data collection efforts. The prevailing circumstances during the COVID-19 pandemic in the hospital setting presented formidable obstacles in enlisting a sufficiently large and diverse participant pool.

Third, a comparative analysis of several strategies for effective multimodal stress detection systems was conducted on the publicly available stress detection dataset called WESAD (Wearable Stress and Affect Detection). The experimental results demonstrated that the multiple sensor fusion models exhibited superior performance compared to the individual sensor strategy. Among the fusion strategies, the weighted score-level fusion approach outperformed the feature-level strategy, yielding accuracy, precision, recall, and F1-measure of 0.931, 0.824, 0.939, and 0.868, respectively. Additionally, the experimental findings demonstrated the effectiveness of feature normalization in the stress classification system. The stress detection system with Min-Max normalization achieved the best performance across all evaluation metrics, with accuracy, precision, recall, and F1-measure of 0.891, 0.814, 0.855, and 0.812, respectively. Furthermore, the feature selection experiment indicated that using a larger number of features improved the performance of the stress classification system, reaching its peak at 90% of the total features (378 features) with accuracy, precision, recall, and F1-measure of 0.894, 0.819, 0.859, and 0.817, respectively. Regarding the classifiers, in terms of F1-measure, Logistic Regression achieved the best result with 0.76 compared to several classifiers including Naive Bayes, Support Vector Machine, Neural Network, K-Nearest Neighbours, Random Forest, and Decision Tree. The classifier ensemble method outperformed the individual classifiers, benefiting from the strengths of each classifier to improve accuracy. Additionally, the results indicated that the soft voting strategy achieved higher accuracy (0.87), while the hard voting strategy performed better in terms of F1-measure (0.77).

Regarding the comparison between individual, centralized, and federated learning approaches for stress detection, the findings indicated that the individual learning strategy outperformed centralized learning and federated learning in terms of accuracy. This can be attributed to the fact that individuals may respond differently to stressors so that it is better to make personalized models. In terms of privacy, centralized learning poses a risk of privacy breaches as all data must be shared with a centralized server, which could be compromised. On the other hand, the individual learning strategy offers a high level of privacy since no user data or model leaves the user's device. Similarly, federated learning also maintains a high level of privacy as only the learned model, and not raw user data, is processed in the central server. However, one drawback of the individual learning strategy is its limited usability for new users. In centralized and federated learning, new users can immediately utilize the integrated model to infer their stress levels upon registration. In contrast, for individual learning, users must collect training data first to build a personalized model, resulting in more work for the user.

1.12 Bibliography

[1] ABROSHAN, H., DEVOS, J., POELS, G., AND LAERMANS, E. Covid-19 and phishing: Effects of human emotions, behavior, and demographics on the success of phishing attempts during the pandemic. *Ieee Access 9* (2021), 121916–121929. 14, 21

[2] ABROSHAN, H., DEVOS, J., POELS, G., AND LAERMANS, E. Effects of email users' behaviour and demographics on respond to each step of a phishing attack. In *15th International Conference on Communications and Information Technology Security (ICCITS 2021)* (2021). 22, 163

[3] AGNEW, R., AND BREZINA, T. General strain theory. *Handbook on crime and deviance* (2019), 145–160. 11, 119

[4] AKBAR, F., BAYRAKTAROGLU, A. E., BUDDHARAJU, P., DA CUNHA SILVA, D. R., GAO, G., GROVER, T., GUTIERREZ-OSUNA, R., JONES, N. C., MARK, G., PAVLIDIS, I., ET AL. Email makes you sweat: Examining email interruptions and stress using thermal imaging. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), pp. 1–14. 42, 188

[5] ALI, R. F., AND DOMINIC, P. Investigation of information security policy violations among oil and gas employees: A security-related stress and avoidance coping perspective. *Journal of Information Science* (2022), 01655515221087680. 16, 19, 21

[6] ANDREOU, E., ALEXOPOULOS, E. C., LIONIS, C., VARVOGLI, L., GNARDELLIS, C., CHROUSOS, G. P., AND DARVIRI, C. Perceived stress scale: reliability and validity study in greece. *International journal of environmental research and public health 8*, 8 (2011), 3287–3298. 23, 83, 123

[7] ANSELL, E. B., GU, P., TUIT, K., AND SINHA, R. Effects of cumulative stress and impulsivity on smoking status. *Human Psychopharmacology: Clinical and Experimental 27*, 2 (2012), 200–208. 42, 188

[8] ANTON, N. E., HOWLEY, L. D., PIMENTEL, M., DAVIS, C. K., BROWN, C., AND STEFANIDIS, D. Effectiveness of a mental skills curriculum to reduce novices' stress. *Journal of surgical research 206*, 1 (2016), 199–205. 25

[9] ARGAW, S. T., TRONCOSO-PASTORIZA, J. R., LACEY, D., FLORIN, M.-V., CALCAVECCHIA, F., ANDERSON, D., BURLESON, W., VOGEL, J.-M., O'LEARY, C., ESHAYA-CHAUVIN, B., ET AL. Cybersecurity of

hospitals: discussing the challenges and working towards mitigating the risks. *BMC medical informatics and decision making 20* (2020), 1–10. 11

[10] BASTIANON, C. D., KLEIN, E. M., TIBUBOS, A. N., BRÄHLER, E., BEUTEL, M. E., AND PETROWSKI, K. Perceived stress scale (pss-10) psychometric properties in migrants and native germans. *BMC psychiatry 20*, 1 (2020), 1–9. 23, 83, 123

[11] BECK, C. T. Secondary traumatic stress in nurses: A systematic review. *Archives of psychiatric nursing 25*, 1 (2011), 1–10. 10, 118

[12] BEMENT, M. H., WEYER, A., KELLER, M., HARKINS, A. L., AND HUNTER, S. K. Anxiety and stress can predict pain perception following a cognitive stress. *Physiology & behavior 101*, 1 (2010), 87–92. 9, 118, 168

[13] BEN-DAVID, Y., HASAN, S., PAL, J., VALLENTIN, M., PANJWANI, S., GUTHEIM, P., CHEN, J., AND BREWER, E. A. Computing security in the developing world: A case for multidisciplinary research. In *Proceedings of the 5th ACM workshop on Networked systems for developing regions* (2011), pp. 39–44. 41, 48, 155

[14] BOBADE, P., AND VANI, M. Stress detection with machine learning and deep learning using multimodal physiological data. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)* (2020), IEEE, pp. 51–57. 12

[15] BURNS, A., DOHENY, E. P., GREENE, B. R., FORAN, T., LEAHY, D., O'DONOVAN, K., AND MCGRATH, M. J. Shimmer™: an extensible platform for physiological signal capture. In *2010 annual international conference of the IEEE engineering in medicine and biology* (2010), IEEE, pp. 3759–3762. 25, 168

[16] BUTAVICIUS, M., PARSONS, K., LILLIE, M., MCCORMAC, A., PATTINSON, M., AND CALIC, D. When believing in technology leads to poor cyber security: Development of a trust in technical controls scale. *Computers & Security 98* (2020), 102020. 22, 163

[17] BUTAVICIUS, M., PARSONS, K., PATTINSON, M., AND MCCORMAC, A. Breaching the human firewall: Social engineering in phishing and spear-phishing emails. *arXiv preprint arXiv:1606.00887* (2016). 22, 163

[18] BUTAVICIUS, M., TAIB, R., AND HAN, S. J. Why people keep falling for phishing scams: The effects of time pressure and deception cues on the detection of phishing emails. *Computers & Security 123* (2022), 102937. 21, 42, 165, 188

[19] CARTWRIGHT, N. What are randomised controlled trials good for? *Philosophical studies 147*, 1 (2010), 59. 25

[20] CHAAYA, M., OSMAN, H., NAASSAN, G., AND MAHFOUD, Z. Validation of the arabic version of the cohen perceived stress scale (pss-10) among pregnant and postpartum women. *BMC psychiatry 10*, 1 (2010), 1–7. 23, 83, 123

[21] CHEN, H., LIU, M., AND LYU, T. Understanding employees' information security–related stress and policy compliance intention: the roles of information security fatigue and psychological capital. *Information & Computer Security*, ahead-of-print (2022). 16, 19

[22] CHEN, L., XIE, Z., ZHEN, J., AND DONG, K. The impact of challenge information security stress on information security policy compliance: The mediating roles of emotions. *Psychology Research and Behavior Management* (2022), 1177–1191. 16, 19

[23] COHEN, S. Perceived stress in a probability sample of the united states. 16, 82, 120, 164

[24] COHEN, S., KAMARCK, T., AND MERMELSTEIN, R. A global measure of perceived stress. *Journal of health and social behavior* (1983), 385–396. 16, 120, 164

[25] COHEN, S., KAMARCK, T., AND MERMELSTEIN, R. Perceived stress scale (pss). *J Health Soc Beh 24* (1983), 285. 4, 11, 23, 81, 82, 102, 123, 145, 246, 258

[26] COHEN, S., KESSLER, R. C., GORDON, L. U., ET AL. Strategies for measuring stress in studies of psychiatric and physical disorders. *Measuring stress: A guide for health and social scientists 28* (1995), 3–26. 9, 118

[27] COOPER, C. L., DEWE, P. J., DEWE, P. J., O'DRISCOLL, M. P., AND O'DRISCOLL, M. P. Organizational stress: A review and critique of theory, research, and applications. 9, 11, 118, 119

[28] D'ARCY, J., HERATH, T., AND SHOSS, M. K. Understanding employee responses to stressful information security requirements: A coping perspective. *Journal of management information systems 31*, 2 (2014), 285–318. 13, 18

[29] DAVEY, H. M., BARRATT, A. L., BUTOW, P. N., AND DEEKS, J. J. A one-item question with a likert or visual analog scale adequately measured current anxiety. *Journal of clinical epidemiology 60*, 4 (2007), 356–360. 25

[30] Davis, A., Shashidharan, A., Liu, Q., Enck, W., McLaughlin, A., and Watson, B. Insecure behaviors on mobile devices under stress. In *Proceedings of the 2014 Symposium and Bootcamp on the Science of Security* (2014), pp. 1–2. 13, 20, 21

[31] De Maeyer, C., and Markopoulos, P. Are digital twins becoming our personal (predictive) advisors?:'our digital mirror of who we were, who we are and who we will become'. In *22st International Conference on Human-Computer Interaction'20: HCI International 2020* (2020), Springer, pp. 250–268. 45, 272

[32] Dhar, R., and Nowlis, S. M. The effect of time pressure on consumer choice deferral. *Journal of Consumer research 25*, 4 (1999), 369–384. 20, 121, 165

[33] D'Arcy, J., and Teh, P.-L. Predicting employee information security policy compliance on a daily basis: The interplay of security-related stress, emotions, and neutralization. *Information & Management 56*, 7 (2019), 103151. 14, 19

[34] Egelman, S., and Peer, E. Scaling the security wall: Developing a security behavior intentions scale (sebis). In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (2015), pp. 2873–2882. 10, 16, 83, 119, 120, 123, 164

[35] Embriaco, N., Azoulay, E., Barrau, K., Kentish, N., Pochard, F., Loundou, A., and Papazian, L. High level of burnout in intensivists: prevalence and associated factors. *American journal of respiratory and critical care medicine 175*, 7 (2007), 686–692. 2, 10, 100, 117, 118, 119

[36] Fauzi, M. A., and Bours, P. Ensemble method for sexual predators identification in online chats. In *2020 8th International Workshop on Biometrics and Forensics (IWBF)* (2020), IEEE, pp. 1–6. 29, 207, 248

[37] Fauzi, M. A., Yang, B., , Moor, K. D., Yeng, P., Rachmayani, D., Busch, C., and Wetherell, M. Can stress compromise phishing email detection? *Under review: Decision Support Systems* (2023). 27

[38] Fauzi, M. A., and Yang, B. Audiouth: Multi-factor authentication based on audio signal. In *Proceedings of the Future Technologies Conference (FTC) 2020, Volume 3* (2021), Springer, pp. 935–946. 29

[39] Fauzi, M. A., and Yang, B. Continuous stress detection of hospital staff using smartwatch sensors and classifier ensemble. In *pHealth 2021*. IOS Press, 2021, pp. 245–250. 28, 202, 211, 219, 228, 232, 238, 262

[40] FAUZI, M. A., AND YANG, B. Multiple sensor fusion for stress detection in the hospital environment. In *2022 6th EAI International Conference on Computer Science and Engineering (COMPSE)* (2022), Springer. 27

[41] FAUZI, M. A., YANG, B., AND BLOBEL, B. Comparative analysis between individual, centralized, and federated learning for smartwatch based stress detection. *Journal of Personalized Medicine 12*, 10 (2022), 1584. 28

[42] FAUZI, M. A., YANG, B., AND MARTIRI, E. Passgan-based honeywords system. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)* (2019), IEEE, pp. 179–184. 30

[43] FAUZI, M. A., YANG, B., AND MARTIRI, E. Passgan based honeywords system for machine-generated passwords database. In *2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)* (2020), IEEE, pp. 214–220. 29

[44] FAUZI, M. A., YANG, B., AND MARTIRI, E. Password guessing-based legacy-ui honeywords generation strategies for achieving flatness. In *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)* (2020), IEEE, pp. 1610–1615. 29

[45] FAUZI, M. A., YANG, B., AND MARTIRI, E. Passgan for honeywords: Evaluating the defender and the attacker strategies. In *Advances on Smart and Soft Computing: Proceedings of ICACIn 2020* (2021), Springer, pp. 391–401. 29

[46] FAUZI, M. A., YANG, B., AND YENG, P. Improving stress detection using weighted score-level fusion of multiple sensor. In *Proceedings of the 7th International Conference on Sustainable Information Engineering and Technology* (2022), pp. 65–71. 28

[47] FAUZI, M. A., YANG, B., AND YENG, P. K. Examining the effect of feature normalization and feature selection for logistic regression based multimodal stress detection. In *2022 International Conference on Theoretical and Applied Computer Science and Engineering (ICTASCE)* (2022), IEEE, pp. 90–94. 28

[48] FAUZI, M. A., YENG, P., YANG, B., AND RACHMAYANI, D. Examining the link between stress level and cybersecurity practices of hospital staff in indonesia. In *Proceedings of the 16th International Conference on Availability, Reliability and Security* (2021), pp. 1–8. 14, 16, 21, 27, 100, 108, 117, 120, 132, 153, 164, 202, 218, 232, 246, 258

[49] Fauzi, M. A., and Yuniarti, A. Ensemble method for indonesian twitter hate speech detection. *Indonesian Journal of Electrical Engineering and Computer Science 11*, 1 (2018), 294–299. 12, 247, 248

[50] Fauzi, Muhammad Ali, Y. P., and Yang, B. Correlating healthcare staff's stress level and cybersecurity practices in norway. In *Under review: IEEE Conference on the Intelligent Methods, Systems, and Applications (IMSA)* (2023), IEEE. 27

[51] Fauzi, Muhammad Ali, Y. P., Yang, B., Nimbe, P., and Rachmayani, D. Stress and cybersecurity practices among hospital staff in the digital age: An empirical study from ghana. *Under review: IEEE Access* (2023). 27

[52] Fauzi, Muhammad Ali, Y. P., Yang, B., Rachmayani, D., and Nimbe, P. Examining the relationship between stress levels and cybersecurity practices among hospital employees in three countries: Ghana, norway, and indonesia. In *IEEE Annual Computers, Software, and Applications Conference (COMPSAC)* (2023), IEEE. 27

[53] Fordyce, T., Green, S., and Gross, T. Investigation of the effect of fear and stress on password choice. In *Proceedings of the 7th Workshop on Socio-Technical Aspects in Security and Trust* (2018), pp. 3–15. 3, 13, 20, 21, 41, 120, 144, 164, 187

[54] Frey, B. B. *The SAGE encyclopedia of educational research, measurement, and evaluation*. Sage Publications, 2018. 24

[55] Garg, P., Santhosh, J., Dengel, A., and Ishimaru, S. Stress detection by machine learning and wearable sensors. In *26th International Conference on Intelligent User Interfaces* (2021), pp. 43–45. 5, 12, 202, 218, 247, 258

[56] Goutam, R. K. Importance of cyber security. *International Journal of Computer Applications 111*, 7 (2015). 11

[57] Gratian, M., Bandi, S., Cukier, M., Dykstra, J., and Ginther, A. Correlating human traits and cyber security behavior intentions. *computers & security 73* (2018), 345–358. 10, 43, 80, 100, 117, 119, 144, 189

[58] Grobler, M., and van Vuuren, J. J. Broadband broadens scope for cyber crime in africa. In *2010 Information Security for South Africa* (2010), IEEE, pp. 1–8. 41, 49, 155

[59] Hadlington, L. Employees attitudes towards cyber security and risky online behaviours: an empirical assessment in the united kingdom. 10, 81, 119

[60] Haim, N., Vardi, G., Yehudai, G., Shamir, O., and Irani, M. Reconstructing training data from trained neural networks. *Advances in Neural Information Processing Systems 35* (2022), 22911–22924. 46

[61] Happell, B., Dwyer, T., Reid-Searl, K., Burke, K. J., Caperchione, C. M., and Gaskin, C. J. Nurses and stress: recognizing causes and seeking solutions. *Journal of nursing management 21*, 4 (2013), 638–647. 2, 100, 117

[62] Helton, W. S. Validation of a short stress state questionnaire. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (2004), vol. 48, Sage Publications Sage CA: Los Angeles, CA, pp. 1238–1242. 20, 121, 165

[63] Hoboubi, N., Choobineh, A., Ghanavati, F. K., Keshavarzi, S., and Hosseini, A. A. The impact of job stress and job satisfaction on workforce productivity in an iranian petrochemical industry. *Safety and health at work 8*, 1 (2017), 67–71. 10, 118, 119

[64] Hulsman, R. L., Pranger, S., Koot, S., Fabriek, M., Karemaker, J. M., and Smets, E. M. How stressful is doctor–patient communication? physiological and psychological stress of medical students in simulated history taking and bad-news consultations. *International Journal of Psychophysiology 77*, 1 (2010), 26–34. 9, 118, 168

[65] Hwang, I., and Cha, O. Examining technostress creators and role stress as potential threats to employees' information security compliance. *Computers in Human Behavior 81* (2018), 282–293. 13, 18

[66] Hwang, I., Kim, S., and Rebman, C. Impact of regulatory focus on security technostress and organizational outcomes: The moderating effect of security technostress inhibitors. *Information Technology & People 35*, 7 (2022), 2043–2074. 14, 19

[67] Indikawati, F. I., and Winiarti, S. Stress detection from multimodal wearable sensor data. In *IOP Conference Series: Materials Science and Engineering* (2020), vol. 771, IOP Publishing, p. 012028. 5, 12, 13, 202, 218, 247, 258

[68] Jacobs, S. C., Friedman, R., Parker, J. D., Tofler, G. H., Jimenez, A. H., Muller, J. E., Benson, H., and Stone, P. H. Use of skin conductance changes during mental stress testing as an index of autonomic arousal in cardiovascular research. *American heart journal 128*, 6 (1994), 1170–1177. 25, 168, 175, 181

[69] Jagatic, T. N., Johnson, N. A., Jakobsson, M., and Menczer, F. Social phishing. *Communications of the ACM 50*, 10 (2007), 94–100. 22, 163

[70] Jensen, M. L., Dinger, M., Wright, R. T., and Thatcher, J. B. Training to mitigate phishing attacks using mindfulness techniques. *Journal of Management Information Systems 34*, 2 (2017), 597–626. 13, 21

[71] Jeon, S., Son, I., and Han, J. Understanding employee's emotional reactions to issp compliance: focus on frustration from security requirements. *Behaviour & Information Technology* (2022), 1–18. 15, 18

[72] Jiang, R. Exploring employees' computer fraud behaviors using the fraud triangle theory. *Pacific Asia Journal of the Association for Information Systems 14*, 4 (2022), 4. 15, 18

[73] Jiang, R., and Zhang, J. The impact of work pressure and work completion justification on intentional nonmalicious information security policy violation intention. *Computers & Security 130* (2023), 103253. 16, 18

[74] Jones, H. S., Towse, J. N., Race, N., and Harrison, T. Email fraud: The search for psychological predictors of susceptibility. *PloS one 14*, 1 (2019), e0209684. 21, 165

[75] Kennedy, D. O., Little, W., and Scholey, A. B. Attenuation of laboratory-induced stress in humans after acute administration of melissa officinalis (lemon balm). *Psychosomatic medicine 66*, 4 (2004), 607–613. 25, 166, 187

[76] Kirschbaum, C., Pirke, K.-M., and Hellhammer, D. H. The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology 28*, 1-2 (1993), 76–81. 20, 121, 165, 203, 219, 233, 260

[77] Krombholz, K., Hobel, H., Huber, M., and Weippl, E. " advanced social engineering attacks"; journal of information security and applications, 22 (2015), s. 113-122. 1, 100, 116

[78] Kumar, B. S., and Ravi, V. Text document classification with pca and one-class svm. In *Proceedings of the 5th international conference on frontiers in intelligent computing: theory and applications* (2017), Springer, pp. 107–115. 43, 250

[79] Labrecque, L. I., Markos, E., Swani, K., and Peña, P. When data security goes wrong: Examining the impact of stress, social contract violation, and data type on consumer coping responses following a data breach. *Journal of Business Research 135* (2021), 559–571. 15, 18

[80] LAMBERT, E. G., HOGAN, N. L., CAMP, S. D., AND VENTURA, L. A. The impact of work–family conflict on correctional staff: A preliminary study. *Criminology & Criminal Justice 6*, 4 (2006), 371–387. 16, 120, 164

[81] LAZARO, M. J. S., LIM, J., KIM, S. H., AND YUN, M. H. Wearable technologies: acceptance model for smartwatch adoption among older adults. In *International Conference on Human-Computer Interaction* (2020), Springer, pp. 303–315. 4, 202, 232, 233, 246, 258

[82] LAZARUS, R. S., AND FOLKMAN, S. *Stress, appraisal, and coping*. Springer publishing company, 1984. 9, 118, 245, 257

[83] LEE, E.-H. Review of the psychometric evidence of the perceived stress scale. *Asian nursing research 6*, 4 (2012), 121–127. 23, 82, 123

[84] LEVENSTEIN, S., PRANTERA, C., VARVO, V., SCRIBANO, M. L., BERTO, E., LUZI, C., AND ANDREOLI, A. Development of the perceived stress questionnaire: a new tool for psychosomatic research. *Journal of psychosomatic research 37*, 1 (1993), 19–32. 4, 11, 246, 258

[85] LI, L., ZHANG, Y., ZOU, L., LI, C., YU, B., ZHENG, X., AND ZHOU, Y. An ensemble classifier for eukaryotic protein subcellular location prediction using gene ontology categories and amino acid hydrophobicity. *PLoS One 7*, 1 (2012), e31057. 43, 250

[86] LI-MEI LIAO, R., AND CAREY, M. G. Laboratory-induced mental stress, cardiovascular response, and psychological characteristics. *Rev. Cardiovasc. Med 16* (2015), 28–35. 20, 121, 164

[87] LIAO, W., ZHANG, W., ZHU, Z., AND JI, Q. A real-time human stress monitoring system using dynamic bayesian network. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)-workshops* (2005), IEEE, pp. 70–70. 4, 202, 232, 246, 258

[88] LIU, J. C., GOETZ, J., SEN, S., AND TEWARI, A. Learning from others without sacrificing privacy: Simulation comparing centralized and federated machine learning on mobile health data. *JMIR mHealth and uHealth 9*, 3 (2021), e23728. 44, 270

[89] LIU, Q., LIU, Y., LENG, X., HAN, J., XIA, F., AND CHEN, H. Impact of chronic stress on attention control: Evidence from behavioral and event-related potential analyses. *Neuroscience bulletin 36* (2020), 1395–1410. 40, 155

[90] Lukasik, K. M., Waris, O., Soveri, A., Lehtonen, M., and Laine, M. The relationship of anxiety and stress with working memory performance in a large non-depressed sample. *Frontiers in psychology 10* (2019), 4. 25

[91] Maroufizadeh, S., Zareiyan, A., and Sigari, N. Reliability and validity of persian version of perceived stress scale (pss-10) in adults with asthma. *Archives of Iranian medicine 17*, 5 (2014), 0–0. 23, 83, 123

[92] Marshall, G. The purpose, design and administration of a questionnaire for data collection. *Radiography 11*, 2 (2005), 131–136. 24

[93] Marteau, T. M., and Bekker, H. The development of a six-item short-form of the state scale of the spielberger state—trait anxiety inventory (stai). *British journal of clinical Psychology 31*, 3 (1992), 301–306. 25, 167, 168

[94] Martiri, E., Yang, B., and Fauzi, M. A. Indistinguishability of biometric honey templates: comparing human testers and svm classifiers. In *2020 International Conference on Computational Science and Computational Intelligence (CSCI)* (2020), IEEE, pp. 76–82. 29

[95] Matteson, M. T., and Ivancevich, J. M. *Controlling work stress: Effective human resource and management strategies.* Jossey-Bass, 1987. 20, 121, 165

[96] Matthews, K. A., and Stoney, C. M. Influences of sex and age on cardiovascular responses during stress. *Psychosomatic medicine 50*, 1 (1988), 46–56. 20, 121, 164

[97] McCormac, A., Calic, D., Parsons, K., Butavicius, M., Pattinson, M., and Lillie, M. The effect of resilience and job stress on information security awareness. *Information & Computer Security* (2018). 3, 14, 16, 40, 100, 108, 117, 120, 132, 144, 153, 163, 165

[98] McEwen, B. S. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiological reviews 87*, 3 (2007), 873–904. 9, 118

[99] McHugh, M. D., Kutney-Lee, A., Cimiotti, J. P., Sloane, D. M., and Aiken, L. H. Nurses' widespread job dissatisfaction, burnout, and frustration with health benefits signal problems for patient care. *Health affairs 30*, 2 (2011), 202–210. 10, 118

[100] McVicar, A. Workplace stress in nursing: a literature review. *Journal of advanced nursing 44*, 6 (2003), 633–642. 2, 100, 117

[101] MICHAILIDIS, E., AND BANKS, A. P. The relationship between burnout and risk-taking in workplace decision-making and decision-making style. *Work & Stress 30*, 3 (2016), 278–292. 2, 11, 40, 41, 80, 119, 144, 153, 162, 187

[102] MOHER, D., LIBERATI, A., TETZLAFF, J., ALTMAN, D. G., AND PRISMA GROUP*, T. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Annals of internal medicine 151*, 4 (2009), 264–269. 13

[103] MUKAKA, M. M. A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal 24*, 3 (2012), 69–71. 22, 24

[104] NASIRPOURI SHADBAD, F., AND BIROS, D. Understanding employee information security policy compliance from role theory perspective. *Journal of Computer Information Systems 61*, 6 (2021), 571–580. 15, 18

[105] NG, B.-Y., KANKANHALLI, A., AND XU, Y. C. Studying users' computer security behavior: A health belief perspective. *Decision Support Systems 46*, 4 (2009), 815–825. 43, 189

[106] NILSSON, A., SMITH, S., ULM, G., GUSTAVSSON, E., AND JIRSTRAND, M. A performance evaluation of federated learning algorithms. In *Proceedings of the second workshop on distributed infrastructures for deep learning* (2018), pp. 1–8. 44, 270

[107] OMAR, F., HALIM, F. W., ZAINAH, A., FARHADI, H., NASIR, R., AND KHAIRUDIN, R. Stress and job satisfaction as antecedents of workplace deviant behavior. *World Applied Sciences Journal 12*, 16 (2011), 45–51. 11, 119

[108] PARK, J. O., AND SEO, Y. S. Validation of the perceived stress scale (pss) on samples of korean university students. *Korean Journal of Psychology: General 29*, 3 (2010), 611–629. 23, 83, 123

[109] PARSONS, K., BUTAVICIUS, M. A., LILLIE, M., CALIC, D., MCCORMAC, A., AND PATTINSON, M. R. Which individual, cultural, organisational and interventional factors explain phishing resilience? In *HAISA* (2018), pp. 1–11. 22, 163

[110] PARSONS, K., MCCORMAC, A., BUTAVICIUS, M., PATTINSON, M., AND JERRAM, C. Determining employee awareness using the human aspects of information security questionnaire (hais-q). *Computers & security 42* (2014), 165–176. 16, 120, 164

[111] PARSONS, K., MCCORMAC, A., PATTINSON, M., BUTAVICIUS, M., AND JERRAM, C. Phishing for the truth: A scenario-based experiment

of users' behavioural response to emails. In *Security and Privacy Protection in Information Processing Systems: 28th IFIP TC 11 International Conference, SEC 2013, Auckland, New Zealand, July 8-10, 2013. Proceedings 28* (2013), Springer, pp. 366–378. 22, 163

[112] Pattinson, M., Butavicius, M., Parsons, K., McCormac, A., and Calic, D. Factors that influence information security behavior: An australian web-based study. In *Human Aspects of Information Security, Privacy, and Trust: Third International Conference, HAS 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015. Proceedings 3* (2015), Springer, pp. 231–241. 43, 189

[113] Pearl, J. Causal inference. *Causality: objectives and assessment* (2010), 39–58. 25

[114] Perakslis, E. D. Cybersecurity in health care. *N Engl J Med 371*, 5 (2014), 395–397. 1, 11, 80, 144

[115] Plessas, A., Nasser, M., Hanoch, Y., O'Brien, T., Delgado, M. B., and Moles, D. Impact of time pressure on dentists' diagnostic performance. *Journal of dentistry 82* (2019), 38–44. 42, 188

[116] Pottier, P., Dejoie, T., Hardouin, J., Le Loupp, A., Planchon, B., Bonnaud, A., and Leblanc, V. Effect of stress on clinical reasoning during simulated ambulatory consultations. *Medical teacher 35*, 6 (2013), 472–480. 42, 188

[117] Radzali, F. M., Ahmad, A., and Omar, Z. Workload, job stress, family-to-work conflict and deviant workplace behavior. *International Journal of Academic Research in Business and Social Sciences 3*, 12 (2013), 109. 11, 119

[118] Ragu-Nathan, T., Tarafdar, M., Ragu-Nathan, B. S., and Tu, Q. The consequences of technostress for end users in organizations: Conceptual development and empirical validation. *Information systems research 19*, 4 (2008), 417–433. 10, 118

[119] Rajendran, V., Jayalalitha, S., Adalarasu, K., and Usha, G. A review on mental stress detection using pss method and eeg signal method. *ECS Transactions 107*, 1 (2022), 1845. 25

[120] Remor, E. Psychometric properties of a european spanish version of the perceived stress scale (pss). *The Spanish journal of psychology 9*, 1 (2006), 86. 23, 83, 123

[121] Riedl, R. On the biology of technostress: literature review and research agenda. *ACM SIGMIS database: the DATABASE for advances in information systems 44*, 1 (2012), 18–55. 9, 118

[122] Rodrigues, S., Paiva, J. S., Dias, D., Aleixo, M., Filipe, R. M., and Cunha, J. P. S. Cognitive impact and psychophysiological effects of stress using a biomonitoring platform. *International journal of environmental research and public health 15*, 6 (2018), 1080. 25, 168

[123] Rodrigues, S., Paiva, J. S., Dias, D., Pimentel, G., Kaiseler, M., and Cunha, J. P. S. Wearable biomonitoring platform for the assessment of stress and its impact on cognitive performance of firefighters: an experimental study. *Clinical practice and epidemiology in mental health: CP & EMH 14* (2018), 250. 25

[124] Sarno, D. M., and Neider, M. B. So many phish, so little time: Exploring email task factors and phishing susceptibility. *Human Factors 64*, 8 (2022), 1379–1403. 21, 22, 163, 165

[125] Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., and Van Laerhoven, K. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 20th ACM international conference on multimodal interaction* (2018), pp. 400–408. 5, 12, 13, 26, 202, 203, 211, 218, 219, 228, 233, 247, 258, 259

[126] Scholey, A., Haskell, C., Robertson, B., Kennedy, D., Milne, A., and Wetherell, M. Chewing gum alleviates negative mood and reduces cortisol during acute laboratory psychological stress. *Physiology & behavior 97*, 3-4 (2009), 304–312. 25, 166, 187

[127] Schoofs, D., Wolf, O. T., and Smeets, T. Cold pressor stress impairs performance on working memory tasks requiring executive functions in healthy young men. *Behavioral neuroscience 123*, 5 (2009), 1066. 2, 100, 117

[128] Shanafelt, T. D., Boone, S., Tan, L., Dyrbye, L. N., Sotile, W., Satele, D., West, C. P., Sloan, J., and Oreskovich, M. R. Burnout and satisfaction with work-life balance among us physicians relative to the general us population. *Archives of internal medicine 172*, 18 (2012), 1377–1385. 10, 118, 119

[129] Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., and Downs, J. Who falls for phish? a demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2010), pp. 373–382. 42, 188

[130] Siirtola, P. Continuous stress detection using the sensors of commercial smartwatch. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Pro-*

ceedings of the 2019 ACM International Symposium on Wearable Computers (2019), pp. 1198–1201. 5, 12, 13, 202, 211, 218, 228, 247, 258

[131] Simon, L., Jiryis, T., and Admon, R. Now or later? stress-induced increase and decrease in choice impulsivity are both associated with elevated affective and endocrine responses. *Brain Sciences 11*, 9 (2021), 1148. 11, 40, 119, 144, 154, 162

[132] Siponen, M., and Vance, A. Neutralization: New insights into the problem of employee information systems security policy violations. *MIS quarterly* (2010), 487–502. 1, 100, 116

[133] Sozinov, K., Vlassov, V., and Girdzijauskas, S. Human activity recognition using federated learning. In *2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom)* (2018), IEEE, pp. 1103–1111. 44, 45, 271

[134] Spagnolli, A., Guardigli, E., Orso, V., Varotto, A., and Gamberini, L. Measuring user acceptance of wearable symbiotic devices: validation study across application scenarios. In *International Workshop on Symbiotic Interaction* (2015), Springer, pp. 87–98. 4, 202, 232, 233, 246, 258

[135] Spielberger, C. D., Gorsuch, R. L., Lushene, R. E., Vagg, P., and Jacobs, G. A. Manual for the state-trait anxiety inventory. *Consulting Psychologist* (1983). 20, 121, 165, 167

[136] Starcke, K., and Brand, M. Decision making under stress: a selective review. *Neuroscience & Biobehavioral Reviews 36*, 4 (2012), 1228–1248. 2, 90, 91, 117, 132

[137] Surden, H. Machine learning and law. *Wash. L. Rev. 89* (2014), 87. 44, 270

[138] Taib, R., Yu, K., Berkovsky, S., Wiggins, M., and Bayl-Smith, P. Social engineering and organisational dependencies in phishing attacks. In *Human-Computer Interaction–INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part I* (2019), Springer, pp. 564–584. 22, 163

[139] Tarafdar, M., Tu, Q., and Ragu-Nathan, T. Impact of technostress on end-user satisfaction and performance. *Journal of management information systems 27*, 3 (2010), 303–334. 10, 118

[140] TAWFIK, D. S., SCHEID, A., PROFIT, J., SHANAFELT, T., TROCKEL, M., ADAIR, K. C., SEXTON, J. B., AND IOANNIDIS, J. P. Evidence relating health care provider burnout and quality of care: a systematic review and meta-analysis. *Annals of internal medicine 171*, 8 (2019), 555–567. 2, 100, 117

[141] TAYLOR, H. Does internet research work? *International journal of market research 42*, 1 (2000), 1–11. 24

[142] THOMAS, J. Individual cyber security: Empowering employees to resist spear phishing to prevent identity theft and ransomware attacks. *Thomas, JE (2018). Individual cyber security: Empowering employees to resist spear phishing to prevent identity theft and ransomware attacks. International Journal of Business Management 12*, 3 (2018), 1–23. 11

[143] TRANG, S., AND NASTJUK, I. Examining the role of stress and information security policy design in information security compliance behaviour: An experimental study of in-task behaviour. *Computers & Security 104* (2021), 102222. 14, 20, 21, 121, 165

[144] VELKI, T., AND MILIĆ, M. Stress as a mediator between risk and protective factors and online risky behaviors in adolescents. *Primenjena psihologija 14*, 2 (2021), 149–171. 15, 16, 120, 164

[145] VELKI, T., SOLIC, K., AND OCEVCIC, H. Development of users' information security awareness questionnaire (uisaq)—ongoing work. In *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (2014), IEEE, pp. 1417–1421. 18, 120, 164

[146] VENARD, B. Cybersecurity behavior under covid-19 influence. In *2021 international conference on cyber situational awareness, data analytics and assessment (CyberSA)* (2021), IEEE, pp. 1–9. 3, 15, 16, 120, 164

[147] VERIZON. 2022 data breach investigations report. *Available online: https://enterprise.verizon.com/resources/reports/dbir (accessed on 18 March 2023)* (2022). 1, 80, 100, 116, 144, 162

[148] VISHWANATH, A. Examining the distinct antecedents of e-mail habits and its influence on the outcomes of a phishing attack. *Journal of Computer-Mediated Communication 20*, 5 (2015), 570–584. 22, 163

[149] WARKENTIN, M., AND WILLISON, R. Behavioral and policy issues in information systems security: the insider threat. *European Journal of Information Systems 18*, 2 (2009), 101–105. 1, 80, 116, 144

[150] WEMM, S. E., AND WULFERT, E. Effects of acute stress on decision making. *Applied psychophysiology and biofeedback 42*, 1 (2017), 1–12. 2, 11, 40, 41, 80, 119, 144, 153, 162, 187, 202, 246, 258

[151] WEST, C. P., SHANAFELT, T. D., AND KOLARS, J. C. Quality of life, burnout, educational debt, and medical knowledge among internal medicine residents. *Jama 306*, 9 (2011), 952–960. 10, 118, 119

[152] WETHERELL, M. A., AND CARTER, K. The multitasking framework: The effects of increasing workload on acute psychobiological stress reactivity. *Stress and Health 30*, 2 (2014), 103–109. 25, 166, 187

[153] WETHERELL, M. A., AND SIDGREAVES, M. C. Secretory immunoglobulin-a reactivity following increases in workload intensity using the defined intensity stressor simulation (diss). *Stress and Health: Journal of the International Society for the Investigation of Stress 21*, 2 (2005), 99–106. 25, 166

[154] WHEELER, D. L. zxcvbn: Low-budget password strength estimation. In *USENIX security symposium* (2016), pp. 157–173. 20, 121, 165

[155] WRIGHT, K. B. Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of computer-mediated communication 10*, 3 (2005), JCMC1034. 24

[156] YAN, Z., ROBERTSON, T., YAN, R., PARK, S. Y., BORDOFF, S., CHEN, Q., AND SPRISSLER, E. Finding the weakest links in the weakest link: How well do undergraduate students make cybersecurity judgment? *Computers in Human Behavior 84* (2018), 375–382. 1, 80, 116, 144

[157] YANG, Z., ZHANG, J., CHANG, E.-C., AND LIANG, Z. Neural network inversion in adversarial setting via background knowledge alignment. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (2019), pp. 225–240. 45

[158] YAZDANMEHR, A., LI, Y., AND WANG, J. Does stress reduce violation intention? insights from eustress and distress processes on employee reaction to information security policies. *European Journal of Information Systems* (2022), 1–19. 15, 19

[159] YAZDANMEHR, A., LI, Y., AND WANG, J. Employee responses to information security related stress: Coping and violation intention. *Information Systems Journal* (2023). 15, 19

[160] YENG, P., YANG, B., FAUZI, M. A., NIMBE, P., PRIHARSARI, D., AND PRIHARSARI, D. A framework for assessing motivational methods

towards incentivizing cybersecurity practice in healthcare. In *Proceedings of the 7th International Conference on Sustainable Information Engineering and Technology* (2022), pp. 325–330. 28

[161] YENG, P. K., FAUZI, M. A., SUN, L., AND YANG, B. Assessing the legal aspects of information security requirements for health care in 3 countries: Scoping review and framework development. *JMIR Human Factors 9*, 2 (2022), e30050. 28

[162] YENG, P. K., FAUZI, M. A., AND YANG, B. Comparative analysis of machine learning methods for analyzing security practice in electronic health records' logs. In *2020 IEEE International Conference on Big Data (Big Data)* (2020), IEEE, pp. 3856–3866. 29

[163] YENG, P. K., FAUZI, M. A., AND YANG, B. Workflow-based anomaly detection using machine learning on electronic health records' logs: A comparative study. In *2020 International Conference on Computational Science and Computational Intelligence (CSCI)* (2020), IEEE, pp. 753–760. 29

[164] YENG, P. K., FAUZI, M. A., YANG, B., AND NIMBE, P. Investigation into phishing risk behaviour among healthcare staff. *Information 13*, 8 (2022), 392. 28

[165] YENG, P. K., FAUZI, M. A., YANG, B., AND YAYILGAN, S. Y. Analysing digital evidence towards enhancing healthcare security practice: The kid model. In *2022 1st International Conference on AI in Cybersecurity (ICAIC)* (2022), IEEE, pp. 1–9. 28

[166] YENG, P. K., NWEKE, L. O., YANG, B., ALI FAUZI, M., AND SNEKKENES, E. A. Artificial intelligence–based framework for analyzing health care staff security practice: Mapping review and simulation study. *JMIR medical informatics 9*, 12 (2021), e19250. 28

[167] YIN, H., MALLYA, A., VAHDAT, A., ALVAREZ, J. M., KAUTZ, J., AND MOLCHANOV, P. See through gradients: Image batch recovery via gradinversion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 16337–16346. 46

Part I

Correlation Analysis

Chapter 2

Examining the Link Between Stress Level and Hospital Staffs' Cybersecurity Practices in Indonesia

Muhammad Ali Fauzi; Prosper Yeng; Bian Yang; Dita Rachmayani

Abstract

Since healthcare information systems have many important data that can attract many adversaries, it is important to take the right steps to prevent data breaches. Recent studies suggested that 85% of breaches involved a human element and the frequent patterns used are social engineerings. Therefore, many studies focus on making a better understanding of human behavior in cybersecurity and the factors that affect cybersecurity practices. However, there are only a few peer-reviewed studies that focus on the link between stress level and cybersecurity practices. In this study, we examined the link between stress level and cybersecurity practices among hospital employees in Indonesia by surveying 99 hospital workers. Perceived Stress Scale (PSS) was used to measure the employees' stress level and a new scale to measure hospital staff's risky cybersecurity practices was proposed. This study showed that both PSS and proposed cybersecurity practices scales are reliable with Cronbach's α value of more than 0.7. The survey results also revealed that hospital worker's higher stress levels correlate significantly with riskier cybersecurity practices (rs = 0.305, $p < 0.01$). Besides, a higher stress level is also significantly linked to certain cybersecurity practices, such as clicking on a link in an email from an unknown sender, not preventing colleagues from viewing patients' information for a non-therapeutic purpose, posting patient information on social media, ignoring colleagues who engage in negative information security practices, and failing to create strong passwords.

2.1 Introduction

The rise of digitalization in healthcare has a huge potential to improve patient care performance. However, it also carries a hazardous side-effect: healthcare system vulnerabilities. Cybercrime in healthcare can lead to not only data and financial loss but also medical devices and infrastructure damage [26]. Based on a recent investigations report [37], healthcare has become one of the top sectors that are in the biggest exposure of a data leak. The report also suggested that 85% of breaches involved a human element and around 35% patterns in breaches are social engineering, the highest compared to any other patterns.

Humans are frequently referred to as the weakest link in cybersecurity [38, 42]. No matter how sophisticated the technology is developed to improve cybersecurity systems, the system is still prone to be hacked due to human error. For example, users that use a weak password, share their password with others, or forget to log out after using a public computer could lead to data breaches. Therefore, many studies focus on making a better understanding of human behavior in using computers and the internet and the factors that influence their cybersecurity practices.

Whitty et al [41] studied the link between several factors including impulsivity, self-monitoring, and internal-external control and password sharing behavior. Halevi et al [15] evaluated the connection between cultural, personality, and demographic variables and cybersecurity practices. Gratian et al [12] investigated human characteristics such as personality traits, risk-taking preferences, and decision-making styles with cybersecurity behavior intentions. Yeng et al proposed a framework to analyze security practices of hospital employees that combine demographic and psycho-socio-cultural factors [44, 43]. Kennison and Chan-Tin [17] also studied how personality traits, risk-taking preferences, and secure password knowledge can be correlated to risky cybersecurity practices.

One of the factors that also have a significant effect on human behavior is their mental state such as their stress level. Research in psychology has shown that high stress level has a deleterious effect on decision making. Michailidis and Banks [22] found that employees who experienced burnout were more likely to make spontaneous or irrational decisions than those who felt more satisfied with their works. Another study [40] reported that stressed people will be slow in learning something new and may choosing less profitable decisions. In healthcare, some researchers also examined the link between stress level and staff's performance especially related to patient safety [24, 39, 35]. However, these studies did not include cybersecurity practices.

In this study, we specifically examine the link between stress levels and cybersecurity practices among hospital workers in Indonesia. This work follows the hypothesis that hospital employees with higher stress levels have riskier security practices. This study aims for future use as basic data to include stress factors for promoting good cybersecurity behavior.

2.2 Methodology

2.2.1 Research approach

As pictured in Figure 2.1, the goal of this study was to determine the influence of stress levels on hospital staffs' cybersecurity practices. An online questionnaire was used to collect data about healthcare staffs' demographic information, stress level, and cybersecurity practices in the last month. One month period is selected since the predictive validity of the Perceived Stress Scale (PSS), the scale used to measure stress level in this study, is expected to drop rapidly after four to eight weeks [7]. Since not all cybersecurity practices are carried out every day (e.g. updating, backup data, etc.), one month period is also considered ideal. Some prior researches also use one month period to collect user's cybersecurity practices [14, 28].

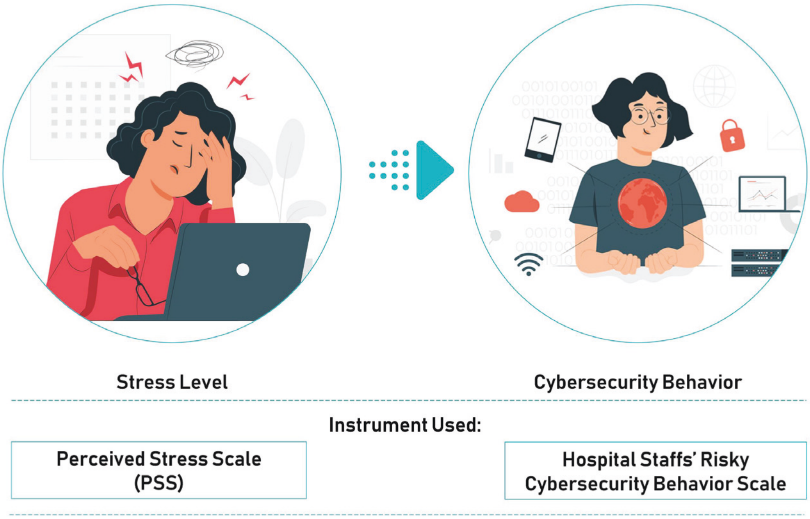

Stress Level
Cybersecurity Behavior
Instrument Used:
Perceived Stress Scale (PSS)
Hospital Staffs' Risky Cybersecurity Behavior Scale

Figure 2.1: Proposed approach.

2.2.2 Participants and Procedures

This study was conducted in a hospital in East Java, Indonesia and granted approval from the hospital's ethics committees. The participants of this study are recruited from the hospital's staff. Written consent was obtained electronically from all participants and questionnaires were completed and analyzed anonymously. In total, 112 participants completed the survey but 13 of them failed to answer an attention-checking question correctly resulting in 99 qualified participants.

2.2.3 Survey Instrument

We developed a web-based questionnaire on Nettskjema, a tool for creating and administering online surveys managed by the University of Oslo [36]. Nettskjema ensures a high degree of security and privacy, which is very crucial for data collection. The questionnaire is in the Indonesian language and consists of three parts: demographic data, the stress level in the last month, and cybersecurity practices in the last month. Perceived stress scale (PSS) was used to measure respondents' stress levels in the last month while a new scale was proposed to measure hospital staff's risky cybersecurity practices. In addition, one attention-checking question was inserted in the middle of cybersecurity practices questions. The question used a 5-point Likert scale (0 = disagree, 1 = slightly disagree, 2 = neutral, 3 = slightly agree, and 4 = agree). The question text in English was as follows: This is an attention-checking question, select '2 (Neutral)' to show you really read this question.

2.2.3.1 Perceived Stress Scale (PSS)

The instrument used to measure stress level in this study is the Perceived Stress Scale (PSS), the most common psychological instrument employed for measuring the perception of stress created by Cohen et al. [7]. It is a short and easy-to-use self-reported questionnaire established with acceptable psychometric properties to measure the extent to which situations in an individual's life are assessed as stressful. The PSS items examine the degree to which individuals feel about their life, specifically how unpredictable, uncontrollable, and overloaded they find their lives during the last month. Rather than concentrating on specific experiences or events, the items in the questionnaire are general in nature.

The PSS is available in three different versions. The original instrument is a 14-item scale (PSS-14) with 7 positive and 7 negative items assessed on a 5-point Likert scale that was developed in English [7]. The PSS-14 was trimmed to ten questions (PSS-10) utilizing factor analysis based on data from 2,387 U.S. residents five years after its launch [6]. A four-item PSS (PSS-4) was also designed for scenarios requiring a very limited time or telephone interviews [6]. PSS-10 is the most widely used version. Based on the systematic review of the PSS psychometric evidence conducted by Lee [19], the 10-item PSS was found to have better psychometric qualities than the

14-item PSS, while the 4-item scale performed the worst. PSS-10 also have been translated into many languages other than English such as Greek[1], Arabic [5], Persian[21], Spanish[29], German[2], Korean[23], etc. Therefore, PSS-10 was used in this study.

Since the study was administered in Indonesia, an Indonesian version of PSS-10 is used. This version has been tested and has a Cronbach Alpha coefficient value of 0.96 [27]. The possible scores range from 0–40. Higher scores indicate that the respondent had a higher stress level in the last month.

2.2.3.2 Hospital Staff's Risky Cybersecurity Practices Scale

The hospital staffs' cybersecurity practices scale is developed partially based on the Human Aspects of Information Security Questionnaire (HAIS-Q) [25], Security Behavior Intentions Scale (SeBIS) [10]. Since these scales are designed for general computer users, we also conducted some interviews with 36 people including hospital staff and cybersecurity experts from several hospitals and universities in Indonesia, Ghana, and Norway to collect inputs and feedback about the scales. Then, based on their inputs and feedback, we modified the scale items to cover cybersecurity practices specifically for hospital staff. We were planning to use this scale to measure cybersecurity practices in hospitals in Indonesia, Ghana, and Norway. However, we only collect data from Indonesia for this study.

The scale asked participants to rate, on a scale of 0–4 (0 = disagree, 1 = slightly disagree, 2 = neutral, 3 = slightly agree, and 4 = agree), how often they engaged in the specific practices in the last month. As displayed in Table 2.1, the final scale included 14 items with 11 items represent risky practices while only 3 items (item number 8, 13, and 14) depict good practices. The possible scores ranged from 0–56 where the scores from the three good practices items were reversed. Higher scores indicate that the respondent had riskier cybersecurity practices in the last month.

2.2.4 Reliability Testing and Correlation Analysis

Since the proposed hospital staff's risky cybersecurity practices scale is new, a reliability analysis was performed by measuring Cronbach's α [8], one of the most popular measures of reliability in the social sciences [4]. Cronbach's α measures how closely related a set of items in the scale are as a group, on the survey result. Based on sources, Cronbach's α value of more than 0.7 is considered acceptable [3, 34, 33]. In addition, we also tested the reliability of the Indonesian version of PSS-10 using the same measure.

Meanwhile, in order to examine the link between the two scales (PSS and Hospital Staff's Risky Cybersecurity Practices Scale), Spearman's rank correlation coefficient (r_s) [30] was used. It is one of the most widely used correlation measures in the psychology field [9].

Table 2.1: Items for the Hospital Staffs' Cybersecurity Practices Scale

	Item
1	In the last month, I usually write my user name and passwords on a piece of paper and stick the paper onto my computer for easy access
2	In the last month, I sometimes visit at least one of the following websites using the hospital's computer: social media; Dropbox and other public file storage systems; online music or videos sites; online newspapers and magazines; personal e-mail accounts; games; instant messaging services, etc
3	In the last month, I did not often read the alert messages/emails concerning security
4	In the last month, I sometimes click on a link in an email from an unknown sender
5	In the last month, I usually postpone software updating activities (restarting, clicking to run an update, accepting to update or follow update schedule) of my computers at my workplace
6	In the last month, I usually postpone backup activities when I am prompted
7	In the last month, I usually do not prevent my colleagues from seeing patients' records for a non-therapeutic purpose when I am working on a patients information on my laptop
8	In the last month, I did not post patient information on social media
9	In the last month, I sometimes share my passwords with my colleagues in hospital
10	In the last month, I usually do not take any action when I notice my colleague ignoring information security rules
11	In the last month, I usually talk about the patient condition in a shared patient ward in a hospital
12	In the last month, I usually disclose sensitive personal health information (patients diagnosis and personal data) in the hospital
13	In the last month, I used a combination of letters, numbers, and symbols in my work passwords
14	In the last month, I have changed my passwords

2.3 Result

2.3.1 General characteristics of participants

In total, 112 staff participated in the survey but only 99 of them were considered as qualified as they answered an attention-checking question correctly. The characteristic of participants is displayed in Table 2.2.

In total, more females participated in this survey (67.68%) than males (32.32%). The age of participants varies from 21 to over 50 with about half of them aged between 31-40 (49.49%). The percentage of participants in the age range 21-30 becomes the second most with 31.31% while the proportion of participants aged between 41-50 is 17.17%. In the last place, participants aged over 50 contribute 2.02% of the total participants. Furthermore, by position, the majority of the participants are Nurses (60.61%) and Pharmacy staff (14.14%). From the position level, one of the hospital's executives participated in this survey. As expected, almost all of the participants are operational staff (91,92%). Meanwhile, 7.07% of participants had manager and supervisor position levels. Concerning work experience, there are no participants with less than one year of work experience. The participant with 1-5 years, 6-10 years, and 11-15 years of work experience share a similar proportion with 22.22%, 26.26%, and 28.28% respectively. In addition, 10.10% of participants had 16-20 years of work experience, 11.11% of them had been working for 21-25 years and 2.02% of them had experience of more than 25 years working in the hospital.

2.3.2 PSS Score

The distribution of the PSS scores is presented in Figure 2.2. The figure depicts a left-skewed distribution. It means that more people are in a lower level of stress. The range of the PSS score is between 0-40 where higher scores indicate that the respondent had a higher stress level in the last month. The survey results show that the average of the participants' PSS score is 13.89 with a standard deviation (SD) of 4.41. The lowest score obtained is 3 (1 participant) while the highest score is 21 (4 participants).

In addition, we also conduct reliability testing of the PSS scale. Based on the survey results, the Cronbach α in this study for the PSS scale was 0.733, suggesting that the items in the scale have relatively good internal consistency.

2.3.3 Hospital Staffs' Cybersecurity Practices Score

The statistic of the cybersecurity practices Score scores is presented in Figure 2.3 and Table 2.3. According to Figure 2.3, the score frequency distribution is left-skewed so that it is good news for the hospital as fewer employees

Table 2.2: Participant Characteristics

Variable	Category	n	%
Gender	F	67	67.68 %
	M	32	32.32 %
	Prefer not to say	0	0.00 %
Age	21-30	31	31.31 %
	31-40	49	49.49 %
	41-50	17	17.17 %
	Over 50	2	2.02 %
Position	Top Level Management	1	1.01 %
	Doctor	3	3.03 %
	Nurse	60	60.61 %
	Lab staff	1	1.01 %
	Pharmachy staff	14	14.14 %
	Nutritionist	3	3.03 %
	Medical record staff	5	5.05 %
	IT staff	1	1.01 %
	Other	11	11.11 %
Position level	Executive	1	1.01 %
	Managers and supervisors	7	7.07 %
	Operational staff	91	91.92 %
Work experience	<1 Year	0	0.00 %
	1-5 Years	22	22.22 %
	6-10 Years	26	26.26 %
	11-15 Years	28	28.28 %
	16-20 Years	10	10.10 %
	21-25 Years	11	11.11 %
	>25 Years	2	2.02 %

have riskier cybersecurity practices. With regards to the possible score of 0-56, the minimum score obtained is 0 (1 participant) while the highest score achieved is 31 (1 participant). Based on Table 2.3, the average score obtained is 16.37 with an SD of 7.66. On this scale, higher scores indicate that the respondent had riskier cybersecurity practices in the last month.

The score range of each item in the hospital employees' cybersecurity practices is between 0-4. Table 2.3 shows that all of the items have an average risky cybersecurity practices score of less than 2. Moreover, some of them have a score averaging less than 1. Item 2 becomes the one with the highest risk score average of 1.91. It means that more people visit external websites using the hospital's computer. Meanwhile, item 8 obtained the lowest risk score average of 0.38. It means that almost all of the employees never post patient information on social media that can lead to personal information

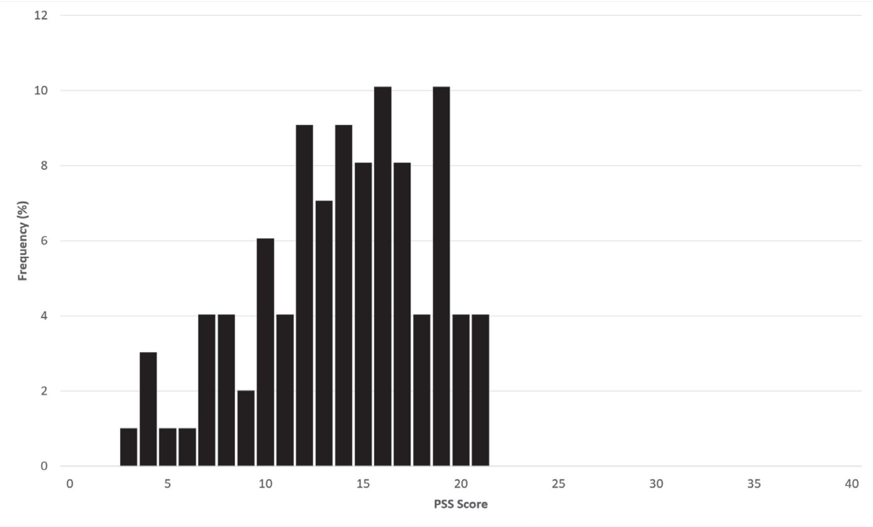


Figure 2.2: Frequency distribution of the hospital staffs' Perceived Stress Scale (PSS) scores.

leakage. It is very reasonable because this practice is against the law.

Besides, we also tested the reliability of the hospital employees' cybersecurity practices scale using the survey results. The Cronbach's α for the scale in this study was 0.732, indicating that the items in the scale had relatively good internal consistency. As displayed in Table 2.4, most of the items are significantly correlated with one another.

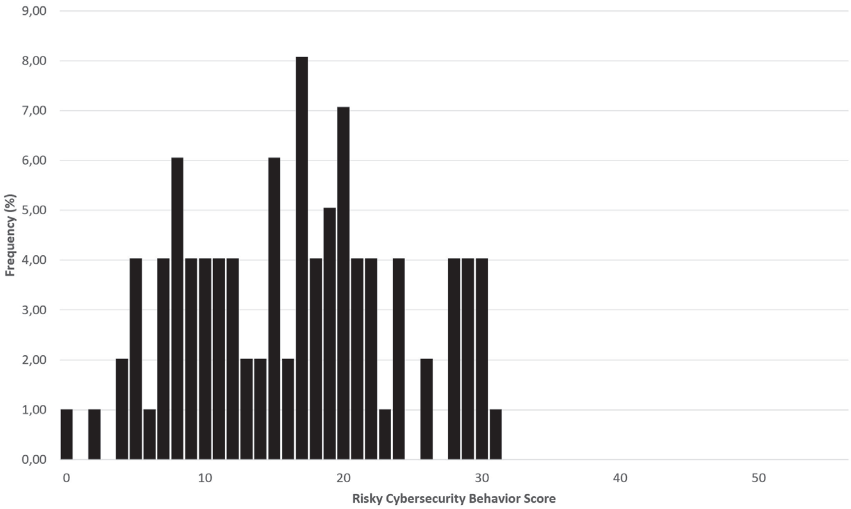


Figure 2.3: Frequency distribution of the hospital staffs' cybersecurity practices scores.

Table 2.3: Means, standard deviations, and ranges for the cybersecurity practices scale items.

Item	Mean	SD	Range
1	1.30	1.46	0-4
2	1.91	1.37	0-4
3	1.68	1.27	0-4
4	0.89	1.08	0-4
5	2.00	1.03	0-4
6	1.73	1.05	0-4
7	0.63	1.13	0-4
8	0.38	1.15	0-4
9	0.87	1.38	0-4
10	1.09	1.34	0-4
11	0.60	1.36	0-4
12	1.49	1.79	0-4
13	1.01	1.67	0-4
14	1.68	1.89	0-4
Total Score	16.37	7.66	0-56

Table 2.4: Correlation of 14 items in the hospital staff's cybersecurity practices scale

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	Item11	Item12	Item13	Item14	Total
Item1	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Item2	0.03	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-
Item3	0.14	0.02	1.00	-	-	-	-	-	-	-	-	-	-	-	-
Item4	0.33**	0.10	0.17	1.00	-	-	-	-	-	-	-	-	-	-	-
Item5	-0.07	0.12	0.11	0.15	1.00	-	-	-	-	-	-	-	-	-	-
Item6	0.14	0.08	0.16	0.28**	0.35**	1.00	-	-	-	-	-	-	-	-	-
Item7	0.25*	0.14	0.13	0.38**	0.05	0.18	1.00	-	-	-	-	-	-	-	-
Item8	0.39**	0.11	0.07	0.36**	0.02	0.17	0.38**	1.00	-	-	-	-	-	-	-
Item9	0.43**	0.13	0.12	0.39**	-0.13	0.11	0.35**	0.29**	1.00	-	-	-	-	-	-
Item10	0.13	-0.01	0.05	0.40**	0.17	0.17	0.39**	0.26*	0.49**	1.00	-	-	-	-	-
Item11	0.30**	0.18	0.18	0.38**	0.11	0.25*	0.58**	0.40**	0.54**	0.44**	1.00	-	-	-	-
Item12	0.23*	0.09	0.11	0.28**	0.05	0.29**	0.32**	0.10	0.29**	0.25*	0.30**	1.00	-	-	-
Item13	0.17	0.06	0.16	0.35**	0.26*	0.22*	0.31**	0.36**	0.19	0.40**	0.26*	0.19	1.00	-	-
Item14	0.09	-0.03	0.48**	0.10	0.27**	0.17	0.11	0.14	0.00	0.20*	0.06	0.05	0.58**	1.00	-
Total	0.51**	0.33**	0.43**	0.60**	0.35**	0.50**	0.56**	0.48**	0.53**	0.55**	0.62**	0.51**	0.59**	0.49**	1.00

*$p < 0.05$

**$p < 0.01$

2.3.4 Correlation Between Stress Level and Cybersecurity Practices

The correlation between hospital staff's stress level and their cybersecurity practices is shown in Table 2.5. In general, stress level had a significant correlation with staff's cybersecurity practices (r_s = 0.305, p $<$ 0.01). It means that employees with higher PSS scores tend to have higher risky cybersecurity practices scores. In other words, hospital staff with higher stress levels tend to practice riskier practices in terms of cybersecurity.

Furthermore, stress level has a significant correlation with several specific cybersecurity practices, namely item 4, 7, 8, 10, and 13 with correlation coefficient of 0.237 (p $<$ 0.05), 0.291 (p $<$ 0.01), 0.257 (p $<$ 0.05), 0.308 (p $<$ 0.01), and 0.228 (p $<$ 0.05) respectively. The results suggest that there is a significant relationship between stress level and several cybersecurity practices in the items. Specifically, stress has a significant relationship with several risky practices including clicking on a link in an email from an unknown sender, not preventing colleagues from seeing patients' information, posting patient information on social media, ignoring colleagues practicing bad information security practices, and not creating strong passwords. For other risky practices items in the survey, stress level also has a positive correlation, even though it is not significant, except for item 3. It means that higher stress levels also have a positive link with other risky practices even though it is not significant. Regarding item 3, unexpectedly, higher stress levels have a relationship with good practices in terms of reading the alert messages/emails concerning security. However, the relationship is very weak and not significant (r_s = 0.006).

2.3.5 Principal Finding and Practical Application

In this study, we looked at the links between stress levels and cybersecurity activities among Indonesian hospital personnel. Overall, we found evidence that stress is significantly associated with risky cybersecurity practices. This result agreed with many previous studies that reported a significant correlation between the stressful condition and decreased performance outcomes (e.g. [24, 18] etc.). In particular, stress correlates significantly with several specific cybersecurity practices, namely clicking on a link in an email from an unknown sender, not preventing colleagues from seeing patients' information, posting patient information on social media, ignoring colleagues practicing bad information security practices, and not creating strong passwords.

Acute stress affects our brains to consider reward and punishment in a way that can make us focus on pleasure and neglect the possible negative outcomes of our decisions [31, 32]. We are more likely to do things that feel good at the moment but are terrible for us in the long run when we are in

Table 2.5: PSS score correlation to cybersecurity practices score.

Item	Correlation coefficient (r_s)
Item 1	0.095
Item 2	0.070
Item 3	-0.006
Item 4	0.237*
Item 5	0.167
Item 6	0.109
Item 7	0.291**
Item 8	0.257*
Item 9	0.180
Item 10	0.308**
Item 11	0.152
Item 12	0.073
Item 13	0.228*
Item 14	0.155
Total cybersecurity practices score	0.305**

*p < 0.05
**p < 0.01

a stressful condition. For example, when we are stressed and need to create a password for our new account, we are more likely to choose an easy-to-remember password or using the same password as for the other website. It makes us feel good because we will not be stressed to remember a new or complicated password even though we know that it may be compromised easily by some attackers in the future. In the same vein, a stressed person is more likely to access their social media during their work to search for joy. However, a stressed person tends to make errors and choose irrational decisions that possibly lead to posting sensitive data from work on social media. Furthermore, stress can make individuals feel exhausted, limited their attention and cognitive resources, and reduce their executive functioning [31, 16] so that they are more vulnerable to phishing attacks. Besides, these factors are also more likely to make the workers ignore their colleague's risky practices.

Regarding the practical application, this study contributes input for hospital management to concern about stress prevention among hospital workers as one of the factors to reduce risky cybersecurity practices. Our findings highlight the need of preventing stress in order to ensure appropriate cybersecurity practices and avoid negative consequences for the hospital such as data breaches. In addition, reducing hospital worker's stress levels can also contribute positively to their physical and mental health and work performance [20, 11].

2.3.6 Limitation

There are several limitations to our work. First, since a self-reported questionnaire was used to collect the data, even though anonymity is guaranteed, it is possible to have a desirability bias [13] where the participant may choose to answer in a way that they got low risky cybersecurity practices score or low stress level. Second, it is also possible that the user gave inaccurate answers because of fatigue, failures to understand questions, or even random clicking. In this study, we tried to minimize it by using an attention-checking question. Next, the result of this study may have been affected by selection bias. Employees with a high workload, which are more likely to have high stress levels, may have declined to participate due to limited time. As we can see from Figure 2.2, the stress level in our sample was relatively low. Finally, this study was cross-sectional, hence, no conclusions about causal relationships can be formed.

2.4 Conclusion

Many hackers consider healthcare information systems as their target because of the abundance of important data contained in them. Therefore, it's crucial to take the proper precautions to avoid data breaches. In cybersecurity, humans are frequently referred to as the weakest link. According to recent reports, a large number of breaches featured a human element, with social engineering being the most common method exploited. As a result, many studies are devoted to gaining a better knowledge of human behavior in cybersecurity and the elements that influence cybersecurity practices. However, only a few research have looked into the relationship between stress and cybersecurity practices.

In this work, we surveyed 99 hospital workers in Indonesia to understand if there was a relationship between their stress levels and their cybersecurity practices. This study is based on the hypothesis that hospital staff who are under a lot of stress exhibit riskier security practices. The employees' stress levels were measured using the Perceived Stress Scale (PSS). To assess hospital staff's risky cybersecurity practices, a custom scale was developed based on the HAIS-Q, SeBIS, and interviews with 36 hospital staff and cybersecurity experts from several hospitals and universities in Indonesia, Ghana, and Norway.

Based on the reliability analysis, both PSS and Hospital Staffs' Cybersecurity Practices scales have acceptable Cronbach's alpha values, which are 0.733 and 0.732 respectively. The survey also found that higher stress levels among hospital employees are significantly associated with riskier cybersecurity practices (r_s= 0.305, p 0.01). Furthermore, specifically, higher stress level also has a significant relationship with some specific cybersecurity practices, namely clicking on a link in an email from an unknown sender, not preventing colleagues from seeing patients' information, posting patient information on social media, ignoring colleague practicing bad information security practices, and not creating strong passwords.

The result of this study presents input for hospital management to give concern on stress prevention among their workers in order to reduce risky cybersecurity practices. Our findings highlight the need of preventing stress to ensure appropriate cybersecurity practices and avoid negative outcomes for the hospital such as data breaches. However, it was a cross-sectional study so that no conclusions about causal relationships can be drawn. Therefore, analyzing the causal relationships between stress level and cybersecurity practices could be an important future work.

2.5 Bibliography

[1] Andreou, E., Alexopoulos, E. C., Lionis, C., Varvogli, L., Gnardellis, C., Chrousos, G. P., and Darviri, C. Perceived stress scale: reliability and validity study in greece. *International journal of environmental research and public health 8*, 8 (2011), 3287–3298. 23, 83, 123

[2] Bastianon, C. D., Klein, E. M., Tibubos, A. N., Brähler, E., Beutel, M. E., and Petrowski, K. Perceived stress scale (pss-10) psychometric properties in migrants and native germans. *BMC psychiatry 20*, 1 (2020), 1–9. 23, 83, 123

[3] Bland, J. M., and Altman, D. G. Statistics notes: Cronbach's alpha. *Bmj 314*, 7080 (1997), 572. 83

[4] Bonett, D. G., and Wright, T. A. Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of organizational behavior 36*, 1 (2015), 3–15. 83

[5] Chaaya, M., Osman, H., Naassan, G., and Mahfoud, Z. Validation of the arabic version of the cohen perceived stress scale (pss-10) among pregnant and postpartum women. *BMC psychiatry 10*, 1 (2010), 1–7. 23, 83, 123

[6] Cohen, S. Perceived stress in a probability sample of the united states. 16, 82, 120, 164

[7] Cohen, S., Kamarck, T., and Mermelstein, R. Perceived stress scale (pss). *J Health Soc Beh 24* (1983), 285. 4, 11, 23, 81, 82, 102, 123, 145, 246, 258

[8] Cronbach, L. J. Coefficient alpha and the internal structure of tests. *psychometrika 16*, 3 (1951), 297–334. 83

[9] de Winter, J. C., Gosling, S. D., and Potter, J. Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological methods 21*, 3 (2016), 273. 83

[10] Egelman, S., and Peer, E. Scaling the security wall: Developing a security behavior intentions scale (sebis). In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (2015), pp. 2873–2882. 10, 16, 83, 119, 120, 123, 164

[11] Giluk, T. L. Mindfulness-based stress reduction: Facilitating work outcomes through experienced affect and high-quality relationships. 91

[12] Gratian, M., Bandi, S., Cukier, M., Dykstra, J., and Ginther, A. Correlating human traits and cyber security behavior intentions. *computers & security 73* (2018), 345–358. 10, 43, 80, 100, 117, 119, 144, 189

[13] Grimm, P. Social desirability bias. *Wiley international encyclopedia of marketing* (2010). 92

[14] Hadlington, L. Employees attitudes towards cyber security and risky online behaviours: an empirical assessment in the united kingdom. 10, 81, 119

[15] Halevi, T., Memon, N., Lewis, J., Kumaraguru, P., Arora, S., Dagar, N., Aloul, F., and Chen, J. Cultural and psychological factors in cyber-security. In *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services* (2016), pp. 318–324. 80, 100, 117, 144

[16] Kassam, K. S., Koslov, K., and Mendes, W. B. Decisions under distress: Stress profiles influence anchoring and adjustment. *Psychological science 20*, 11 (2009), 1394–1399. 91, 132

[17] Kennison, S. M., and Chan-Tin, E. Taking risks with cybersecurity: Using knowledge and personal characteristics to predict self-reported cybersecurity behaviors. *Frontiers in Psychology 11* (2020), 3030. 80, 100, 117, 144

[18] LeBlanc, V. R. The effects of acute stress on performance: implications for health professions education. *Academic Medicine 84*, 10 (2009), S25–S33. 90, 132

[19] Lee, E.-H. Review of the psychometric evidence of the perceived stress scale. *Asian nursing research 6*, 4 (2012), 121–127. 23, 82, 123

[20] Mackenzie, C. S., Poulin, P. A., and Seidman-Carlson, R. A brief mindfulness-based stress reduction intervention for nurses and nurse aides. *Applied nursing research 19*, 2 (2006), 105–109. 91

[21] Maroufizadeh, S., Zareiyan, A., and Sigari, N. Reliability and validity of persian version of perceived stress scale (pss-10) in adults with asthma. *Archives of Iranian medicine 17*, 5 (2014), 0–0. 23, 83, 123

[22] Michailidis, E., and Banks, A. P. The relationship between burnout and risk-taking in workplace decision-making and decision-making style. *Work & Stress 30*, 3 (2016), 278–292. 2, 11, 40, 41, 80, 119, 144, 153, 162, 187

[23] Park, J. O., and Seo, Y. S. Validation of the perceived stress scale (pss) on samples of korean university students. *Korean Journal of Psychology: General 29*, 3 (2010), 611–629. 23, 83, 123

[24] Park, Y.-M., and Kim, S. Y. Impacts of job stress and cognitive failure on patient safety incidents among hospital nurses. *Safety and health at work 4*, 4 (2013), 210–215. 80, 90, 132

[25] Parsons, K., Calic, D., Pattinson, M., Butavicius, M., McCormac, A., and Zwaans, T. The human aspects of information security questionnaire (hais-q): two further validation studies. *Computers & Security 66* (2017), 40–51. 83, 123

[26] Perakslis, E. D. Cybersecurity in health care. *N Engl J Med 371*, 5 (2014), 395–397. 1, 11, 80, 144

[27] Pin, T. L. Hubungan kebiasaan berolahraga dengan tingkat stres pada mahasiswa fakultas kedokteran universitas sumatera utara tahun masuk 2008. *Skripsi. Medan: Fakultas Kedokteran Universitas Sumatera Utara* (2011). 83, 146

[28] Priestman, W., Anstis, T., Sebire, I. G., Sridharan, S., and Sebire, N. J. Phishing in healthcare organisations: Threats, mitigation and approaches. *BMJ health & care informatics 26*, 1 (2019). 81

[29] Remor, E. Psychometric properties of a european spanish version of the perceived stress scale (pss). *The Spanish journal of psychology 9*, 1 (2006), 86. 23, 83, 123

[30] Spearman, C. The proof and measurement of association between two things. 83

[31] Starcke, K., and Brand, M. Decision making under stress: a selective review. *Neuroscience & Biobehavioral Reviews 36*, 4 (2012), 1228–1248. 2, 90, 91, 117, 132

[32] Szalavitz, M. Decision making under stress: The brain remembers rewards, forgets punishments. time, 2012. 90

[33] Taber, K. S. The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education 48*, 6 (2018), 1273–1296. 83, 105, 106, 126, 127, 149, 150

[34] Tavakol, M., and Dennick, R. Making sense of cronbach's alpha. *International journal of medical education 2* (2011), 53. 83

[35] Tsiga, E., Panagopoulou, E., and Montgomery, A. Examining the link between burnout and medical error: A checklist approach. *Burnout Research 6* (2017), 1–8. 80, 202, 218, 232, 246, 258

[36] UiO. Hva er nettskjema, June 2010. Available from: `http://www.uio.no/tjenester/it/applikasjoner/nettskjema/mer-om/`. 82

[37] Verizon. 2021 data breach investigations report, Mar. 2021. Available from: `https://enterprise.verizon.com/resources/reports/dbir`. 1, 80, 100, 116, 144, 162

[38] Warkentin, M., and Willison, R. Behavioral and policy issues in information systems security: the insider threat. *European Journal of Information Systems 18*, 2 (2009), 101–105. 1, 80, 116, 144

[39] Welp, A., Meier, L. L., and Manser, T. Emotional exhaustion and workload predict clinician-rated and objective patient safety. *Frontiers in psychology 5* (2015), 1573. 80, 202, 218, 232, 246, 258

[40] Wemm, S. E., and Wulfert, E. Effects of acute stress on decision making. *Applied psychophysiology and biofeedback 42*, 1 (2017), 1–12. 2, 11, 40, 41, 80, 119, 144, 153, 162, 187, 202, 246, 258

[41] Whitty, M., Doodson, J., Creese, S., and Hodges, D. Individual differences in cyber security behaviors: an examination of who is sharing passwords. *Cyberpsychology, Behavior, and Social Networking 18*, 1 (2015), 3–7. 80, 100, 117, 144

[42] Yan, Z., Robertson, T., Yan, R., Park, S. Y., Bordoff, S., Chen, Q., and Sprissler, E. Finding the weakest links in the weakest link: How well do undergraduate students make cybersecurity judgment? *Computers in Human Behavior 84* (2018), 375–382. 1, 80, 116, 144

[43] Yeng, P., Yang, B., and Snekkenes, E. Observational measures for effective profiling of healthcare staffs' security practices. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)* (2019), vol. 2, IEEE, pp. 397–404. 80

[44] Yeng, P. K., Yang, B., and Snekkenes, E. A. Framework for healthcare security practice analysis, modeling and incentivization. In *2019 IEEE International Conference on Big Data (Big Data)* (2019), IEEE, pp. 3242–3251. 80, 100, 117, 144

Chapter 3

Correlating Healthcare Staff's Stress Level and Cybersecurity Practices in Norway

Muhammad Ali Fauzi; Prosper Yeng; Bian Yang;

Abstract

This paper examines the relationship between stress levels and cybersecurity practices among hospital employees in Norway. As the healthcare sector increasingly uses technology to improve patient care, hospital data became more vulnerable to cyber attacks and the human factor remains a significant vulnerability that can be exploited by hackers. Stress is one such factor that can affect cybersecurity practices. This study hypothesized that employees with higher stress levels are more likely to engage in risky security practices. An online survey was conducted to collect data on demographic details, stress levels, and cybersecurity practices of healthcare staff. Respondents' stress levels were assessed using the Perceived Stress Scale (PSS), and hospital staff's hazardous cybersecurity practices were evaluated using the Hospital Staff's Risky Cybersecurity Practices Scale (HS-RCPS). Regarding cybersecurity practices, the results show that hospital workers in Norway tend to have good cybersecurity practices. Our findings also indicate that there is no significant correlation between stress levels and cybersecurity practices (r=0.101). In addition, demographic variables including gender, age, position, and position level did not have a significant impact on healthcare professionals' dangerous cybersecurity activities. However, years of working experience were found to be a crucial factor in determining cybersecurity practices among hospital employees (F(3, 38) = 3.146, p = 0.036) with staff who had more than 25 years of work experience getting the highest average HS-RCPS score.

3.1 Introduction

The healthcare sector has increasingly utilized technology to manage patient data and improve patient care. However, despite the implementation of robust cybersecurity measures and advanced technologies, the human factor remains a significant vulnerability that can be exploited by hackers [10, 15]. Human error, resulting from actions such as selecting weak passwords, falling for phishing scams, or disregarding security protocols, can compromise digital systems and data security, exposing them to cyber attacks. Notably, a recent Verizon study [19] highlights the critical role of human error in data breaches, with approximately 82% of incidents attributed to this factor. Consequently, the human element in cybersecurity has been extensively investigated in prior research [20, 7, 6, 21, 9].

Stress is one of the human factors that needs attention in the context of cybersecurity in the healthcare setting. Healthcare professionals experience various job-related stressors, including long working hours, heavy workloads, low compensation, organizational challenges, etc. [4, 12, 8]. Psychological studies indicate that high levels of stress can negatively affect decision-making skills and task performance, leading to poor patient safety and fewer profitable choices [14, 17]. However, The relationship between stress and cybersecurity practices has only been briefly studied in peer-reviewed papers. Fauzi et al. [5] have evaluated the association between stress levels and the cybersecurity practices of hospital staff in Indonesia while McCormac et al [11] examined the relationship between job stress and information security awareness (ISA) among company workers in Australia. There have been no studies on this topic for hospital professionals in Norway.

This paper aims to investigate the relationship between stress levels and cybersecurity practices among hospital workers in Norway, hypothesizing that employees with higher stress levels are more likely to engage in risky security practices. In addition, we will also evaluate the relationship between the workers' demographic variables and their cybersecurity practices. By understanding the impact of stress on cybersecurity practice, healthcare organizations can develop targeted interventions and training programs to mitigate the risks associated with human factors in cybersecurity. The findings of this study have important implications for the healthcare industry and cybersecurity professionals. It can inform the development of more effective cybersecurity policies, procedures, and training programs. Finally, this research has the potential to improve the security of healthcare systems and protect sensitive patient data.

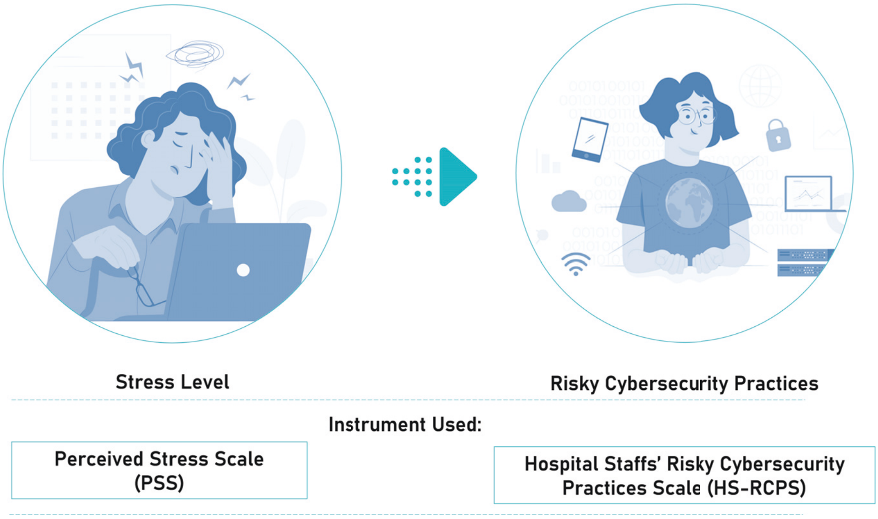

Stress Level
Risky Cybersecurity Practices
Instrument Used:
Perceived Stress Scale (PSS)
Hospital Staffs' Risky Cybersecurity Practices Scale (HS-RCPS)

Figure 3.1: Proposed Approach.

3.2 Materials and Methods

3.2.1 Research approach

In this study, the primary objective was to examine the relationship between stress levels and cybersecurity practices among hospital staff. The research approach is outlined in Figure 3.1. To achieve this objective, an online survey developed using Nettskjema was utilized to collect data on the demographic details, stress levels, and cybersecurity practices of healthcare staff within the past month. Nettskjema is an online survey platform that places a high priority on data privacy and security run by the University of Oslo [18]. The survey was composed in Norwegian. Respondents' stress levels were assessed using the Perceived Stress Scale (PSS), and hospital staff's hazardous cybersecurity practices were evaluated using the Hospital Staff's Risky Cybersecurity Practices Scale (HS-RCPS). Additionally, to guarantee the quality of the response, the questionnaire also contained an attention-checking question. Hospital employees from a hospital in Norway were invited to participate in the study. All respondents provided their written consent electronically, and the surveys' completion and analysis were completely anonymous.

3.2.2 Perceived Stress Scale (PSS)

The Perceived Stress Scale (PSS) is a self-report survey used to measure individuals' perceptions of stress in their lives. It uses a 5-point Likert scale to assess the frequency of thoughts and feelings connected to stress during the previous month. PSS assesses the subjective experience of stress, rather than specific stressors. PSS has three variations: PSS-14, PSS-10, and PSS-4, with PSS-10 having superior psychometric properties than the others [2]. PSS has been translated and validated in many languages, indicating its cross-cultural applicability. This study will use the Norwegian language version of PSS. This version was translated by CheckWare AS, Norway [3]. We will also evaluate the reliability of this Norwegian version in this study.

3.2.3 Hospital Staff's Risky Cybersecurity Practices Scale (HS-RCPS)

The Hospital Staff's Risky Cybersecurity Practices Scale (HS-RCPS) was developed to evaluate hospital staff's cybersecurity practices. This scale is partially based on the Human Aspects of Information Security Questionnaire (HAIS-Q) and the Security Behavior Intentions Scale (SeBIS). The scale was tailored specifically to measure the cybersecurity practices of healthcare workers based on feedback from interviews with 36 healthcare employees and cybersecurity professionals from various universities and hospitals in Ghana, Indonesia, and Norway. As seen in Table 3.1 The scale consists of 12 items with a possible total score ranging from 0 to 48. Higher scores indicate riskier cybersecurity practices over the past month. Respondents were asked to rate their engagement in specific cybersecurity practices using a scale of 0 to 4 (”disagree” to ”agree”).

3.2.4 Data Analysis

SPSS was employed to analyze the data. The reliability of both PSS and HS-RCPS was evaluated using Cronbach's α. Furthermore, the correlation between the two scales, PSS and HS-RCPS, was assessed using Pearson's correlation coefficient. In addition, a t-test was used to analyze the mean difference of HS-RCPS scores between different gender and position level groups while ANOVA was utilized to test the mean difference between different age, position, and work experience groups.

3.3 Results

3.3.1 General characteristics of respondents

In total, 44 hospital staff took part in the survey but only 42 of them were judged to be eligible as they correctly answered the attention-checking question. The characteristic of the eligible respondents is displayed in Table 3.2.

Regarding the gender distribution, the majority of the respondents (85.70%) identified as female, while a smaller percentage (14.30%) identified as male.

Table 3.1: Items for the HS-RCPS

	Item
1	In the last month, I usually write my username and password on a piece of paper and stick the paper onto my computer for easy access
2	In the last month, I sometimes visit at least one of the following websites using the hospital's computer: social media; Dropbox and other public file storage systems; online music or videos sites; online newspapers and magazines; personal e-mail accounts; games; instant messaging services, etc
3	In the last month, I did not often read the alert messages/emails concerning security
4	In the last month, I sometimes click on a link in an email from an unknown sender
5	In the last month, I usually postpone software updating activities (restarting, clicking to run an update, accepting to update, or following the update schedule) of my computers at my workplace
6	In the last month, I usually postpone backup activities when I am prompted
7	In the last month, I usually do not prevent my colleagues from seeing patients' records for a non-therapeutic purpose when I am working on patient information on my laptop
8	In the last month, I did not post patient information on social media
9	In the last month, I sometimes share my passwords with my colleagues in the hospital
10	In the last month, I usually do not take any action when I notice my colleague ignoring information security rules
11	In the last month, I used a combination of letters, numbers, and symbols in my work passwords
12	In the last month, I have changed my passwords

Table 3.2: Respondent Characteristics

Variable	Category	n	%
Gender	Female	36	85.70 %
	Male	6	14.30 %
Age	21-31	11	26.20 %
	31-40	14	33.30 %
	41-50	10	23.80 %
	Over 50	7	16.70 %
Position	Top Level Management	2	4.80 %
	Doctor	14	33.30 %
	Nurse	21	50.00 %
	Other	5	11.90 %
Position level	Executive	0	0.09 %
	Managers and supervisors	6	14.30 %
	Operational staff	36	85.70 %
Work experience	<6 Year	7	16.70 %
	6-15 Years	20	47.60 %
	16-25 Years	10	23.80 %
	>25 Years	5	11.90 %

The survey respondents were fairly evenly distributed across different age ranges, with the largest group falling in the 31-40 age range (33.30%) and the smallest group was respondents over the age of 50 (16.70%). Meanwhile, the percentages of respondents in the 21-31 and 41-50 age ranges were 26.30% and 23.80%, respectively. Based on position, a majority of the respondents identified as nurses (50.00%) and followed by doctors (33.30%). Only a small percentage (4.80%) occupied a management top-level management position while 11.90% of them worked in other positions. The vast majority of respondents (85.70%) identified as operational staff, while a smaller percentage (14.30%) identified as managers or supervisors. No respondents identified as executive level. Furthermore, regarding years of work experience, the majority of respondents (47.60%) reported having 6-15 years of work experience, followed by 16-25 years (23.80%). Only a small percentage (11.90%) reported having over 25 years of work experience.

3.3.2 PSS Score

The PSS scores' distribution is shown in Figure 3.2. Higher PSS scores indicate that the respondent has experienced more stress during the last month. The PSS scores range from 0 to 40. According to the result, the PSS score average was 14.05 with a standard deviation of 6.4. One staff had a PSS score of 1, which is the lowest score recorded among all respondents. In contrast, the highest PSS score ever recorded was 29, which was recorded by one staff. Furthermore, we also assessed the reliability of the PSS. According to the survey results, the scale had a Cronbach's α of 0.844 which indicates that its items had a good level of internal consistency [16, 22].

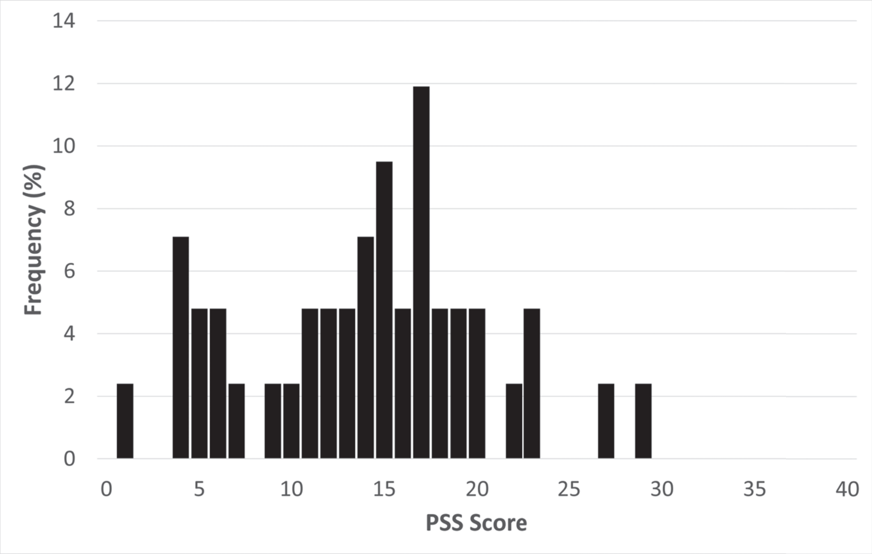


Figure 3.2: Frequency distribution of the PSS scores.

3.3.3 HS-RCPS Score

The distribution and statistics of the HS-RCPS scores are shown in Figure 3.3 and Table 3.3. The riskiness of the cybersecurity practices was measured on a scale from 0 to 48, with 0 being the lowest and 48 being the greatest. The lowest risky cybersecurity practice score among the respondents over the last month was 2, which was achieved by two staff. One staff, on the other hand, had a score of 26, indicating that they had the riskiest practices within the same time period. The standard deviation of the score was 4.90 with the average being 10.88.

The range of each item was 0-4. The mean response values for the items range from 0.00 to 1.90, with item 6 having the highest mean value (1.90). It means that generally, the staff had good cybersecurity practices. Moreover, two items (items 1 and 8) had a mean value of 0, the safest cybersecurity practice. It suggests that none of the staff write their credential in public places for easy access (item 1) or post patient information on social media (item 8). Finally, we assessed the reliability of the HS-RCPS. According to

Table 3.3: Descriptive statistic of HS-RCPS items.

Item	Min	Max	Mean	SD	Range
I-1	0	0	0.00	0.00	0-4
I-2	0	4	0.95	1.43	0-4
I-3	0	4	1.21	1.32	0-4
I-4	0	4	0.14	0.65	0-4
I-5	0	4	1.88	1.15	0-4
I-6	0	3	1.90	0.48	0-4
I-7	0	4	1.64	1.38	0-4
I-8	0	0	0.00	0.00	0-4
I-9	0	4	0.12	0.63	0-4
I-10	0	4	1.40	1.13	0-4
I-11	0	4	0.74	1.36	0-4
I-12	0	4	0.88	1.42	0-4
HS-RCPS	2	26	10.88	4.90	0-48

the survey results, the scale had a Cronbach's α of 0.502 which indicates that its items had an acceptable level of internal consistency [1, 16, 22].

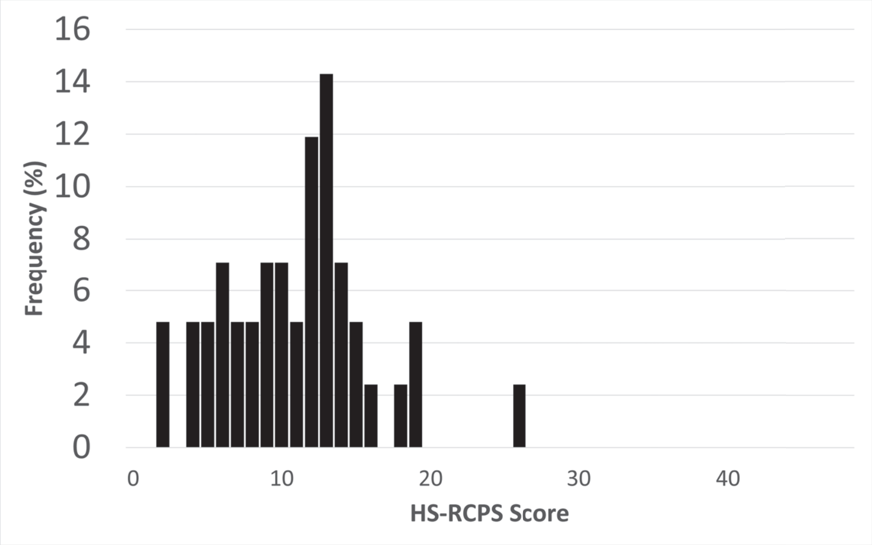


Figure 3.3: Frequency distribution of the HS-RCPS scores.

Table 3.4: Descriptive statistic of risky cybersecurity practices score based on gender.

Gender	Number of respondents	Mean	SD
Female	36	10.4722	5.0793
Male	6	13.3333	2.8752

Table 3.5: Descriptive statistic of risky cybersecurity practices score based on age.

Age (years)	Number of respondents	Mean	SD
Less than 31	11	12.0000	3.6878
31-40	14	11.0714	4.7307
41-50	10	8.4000	3.8355
Over 50	7	12.2857	7.4992

3.3.4 Demographic and Risky Cybersecurity Practices

The descriptive statistic of risky cybersecurity practices score based on gender, age, position, position level, and work experience are displayed in Table 3.4, 3.5, 3.6, 3.7, and 3.8 respectively. The t-test results indicate that there was no significant difference in the levels of risky cybersecurity practices between male (M = 13.33, SD = 2.88) and female (M = 10.47, SD = 5.08) respondents, t(40) = 2.039, p = 0.161. Similarly, the difference was also not significant between staff in operational (M = 11.22, SD = 4.92) and manager/supervisor (M = 8.83, SD = 4.71) positions, t(40) = 0.003, p = 0.959.

Additionally, the ANOVA results indicate that there was no significant difference in mean scores of risky cybersecurity practices among age groups (F(3, 38) = 1.266, p = 0.300) and position groups (F(3, 38) = 0.326, p = 0.828). However, there was a significant difference in mean scores of risky cybersecurity practices between staff groups based on years of work experience, F(3, 38) = 3.146, p = 0.036. More specifically, staff with over 25 years of work experience had significantly higher mean HS-RCPS scores compared to those with 16-25 years of work experience (p = 0.038). Staff with more than 25 years of work experience got the highest average HS-RCPS score compared to other groups with 14.40.

3.3.5 Correlation Between Stress Level and Cybersecurity Practices

Table 3.9 depicts the correlation between the stress levels of hospital staff and their risky cybersecurity practices. The results indicate that stress levels did not significantly correlate with cybersecurity practices. The correlation value between PSS and HS-RCPS-14 was r = 0.101. Furthermore, there are

Table 3.6: Descriptive statistic of risky cybersecurity practices score based on position.

Position	Number of respondents	Mean	SD
Management	2	10.0000	5.6569
Doctor	14	11.7143	4.2141
Nurse	21	10.7619	5.6649
Other	5	9.4000	3.8471

Table 3.7: Descriptive statistic of risky cybersecurity practices score based on position level.

Position Level	Number of respondents	Mean	SD
Managers and supervisors	6	8.8333	4.7082
Operational staff	36	11.2222	4.9171

Table 3.8: Descriptive statistic of risky cybersecurity practices score based on years of work experience.

Work Experience	Number of respondents	Mean	SD
<6 Year	7	12.0000	2.8868
6-15 Years	20	11.3500	4.2584
16-25 Years	10	7.4000	4.5510
>25 Years	5	14.4000	7.2319

no single specific risky cybersecurity practices that had a significant relationship with the stress level. The table also shows the correlation between PSS and either item 1 or item 8 cannot be computed because at least one of the variables is constant. Based on the survey data, both item 1 and item 8 scores are always constant, which is 0.

3.4 Discussion

In this paper, we conducted a correlation analysis between healthcare staff's stress level and cybersecurity practices in Norway. Despite previous work such as by Fauzi et. al. [5] and McCormac et al. [11] reported that higher stress level was significantly associated with riskier cybersecurity practices, this study could not provide significant evidence about that correlation. This can be because stress does not affect the cybersecurity practices of hospital workers in Norway. However, the number of respondents is too low. Hence, further studies in Norway need to be conducted.

Regarding cybersecurity practices, hospital workers in Norway tend to have a low HS-RCPS score which indicates they have good cybersecurity practices. Moreover, all of them have zero scores on two items, namely,

Table 3.9: PSS score correlation to cybersecurity practices score.

Item	Correlation coefficient
I-1	a.
I-2	-0.013
I-3	-0.125
I-4	0.109
I-5	0.165
I-6	0.259
I-7	0.137
I-8	a.
I-9	0.100
I-10	-0.130
I-11	0.001
I-12	0.131
HS-RCPS	0.101

*p < 0.05.
**p < 0.01.
a. Cannot be computed because at least one of the variables is constant.

writing their credentials in public places for easy access and posting patient information on social media. It means that they never do that practice regardless of their stress level.

In examining the relationship between demographic variables and risky cybersecurity practices of healthcare professionals, it was found that factors such as gender, age, position, and position level had minimal to no significant impact on cybersecurity practices. However, when the years of working experience were taken into account, significant differences in unsafe practices were observed between various groups of healthcare professionals. More specifically, staff with over 25 years of work experience had significantly riskier cybersecurity practices compared to those with 16-25 years of work experience. Staff with more than 25 years of work experience got the highest average HS-RCPS score compared to other groups. Experienced healthcare professionals may become complacent or overconfident in their abilities to maintain cybersecurity, leading to risky behaviors. Additionally, they may have developed bad habits over time that make them more susceptible to security breaches. In addition, generally, people with more than 25 years of work experience are older people. Mubarak and Nycyk [13] reported suggested older people still face challenges in learning internet skills.

Regarding the PSS in the Norwegian version, we also have evaluated its reliability. The result shows that this version has a good reliability with a Cronbach's α of 0.844. This contributes to the limited studies that validate the Norwegian version of PSS.

For practical application, this research can be one of the basis for hospital management to make their cybersecurity policies. Furthermore, the result from this paper can also be used as guidance to make an effective and efficient training program to improve the cybersecuirty practices of hospital staff.

3.5 Limitation

Several limitations to our study need to be acknowledged. First, the sample size was relatively small, which may limit the generalizability of our findings. Secondly, we used a self-reported questionnaire to collect data, which could be subject to desirability bias despite anonymity being guaranteed. Respondents may provide socially desirable responses or responses that they believe the researcher wants to hear rather than truthful ones. In addition, memory bias could also affect the accuracy of the responses, particularly for questions related to past events or behaviors. Moreover, this study was cross-sectional, which means that we collected data at a single point in time. Therefore, we cannot determine the causal relationship between stress and cybersecurity practices.

3.6 Conclusions

Our study examined the relationship between stress levels and cybersecurity practices among hospital employees in Norway. Our findings revealed that there is no significant correlation between these two factors. Moreover, demographic variables such as gender, age, position, and degree of position did not have a significant impact on healthcare professionals' dangerous cybersecurity activities. However, years of working experience were found to be a crucial factor in determining cybersecurity practices among hospital employees.

Our study has several implications for hospital management. The results can serve as a foundation for improving the effectiveness and efficiency of cybersecurity measures in healthcare settings. By identifying the factors that influence cybersecurity practices among hospital employees, management can design targeted training programs to enhance cybersecurity awareness and practices. Such measures can lower the risk of cyberattacks and improve patient safety and privacy.

For future studies, more respondents from hospital workers in Norway are needed. In addition, future research to explore the causal relationship between stress levels and cybersecurity practices is also important.

3.7 Bibliography

[1] BERGER, R., AND HÄNZE, M. Impact of expert teaching quality on novice academic performance in the jigsaw cooperative learning method. *International Journal of Science Education 37*, 2 (2015), 294–320. 106, 127, 150

[2] COHEN, S., KAMARCK, T., AND MERMELSTEIN, R. Perceived stress scale (pss). *J Health Soc Beh 24* (1983), 285. 4, 11, 23, 81, 82, 102, 123, 145, 246, 258

[3] DEPARTMENT OF PSYCHOLOGY, C. M. U. Scales - laboratory for the study of stress, immunity, and disease, Feb. 2015. Available from: `https://www.cmu.edu/dietrich/psychology/stress-immunity-disease-lab/scales/index.html`. 102, 146

[4] EMBRIACO, N., AZOULAY, E., BARRAU, K., KENTISH, N., POCHARD, F., LOUNDOU, A., AND PAPAZIAN, L. High level of burnout in intensivists: prevalence and associated factors. *American journal of respiratory and critical care medicine 175*, 7 (2007), 686–692. 2, 10, 100, 117, 118, 119

[5] FAUZI, M. A., YENG, P., YANG, B., AND RACHMAYANI, D. Examining the link between stress level and cybersecurity practices of hospital staff in indonesia. In *Proceedings of the 16th International Conference on Availability, Reliability and Security* (2021), pp. 1–8. 14, 16, 21, 27, 100, 108, 117, 120, 132, 153, 164, 202, 218, 232, 246, 258

[6] GRATIAN, M., BANDI, S., CUKIER, M., DYKSTRA, J., AND GINTHER, A. Correlating human traits and cyber security behavior intentions. *computers & security 73* (2018), 345–358. 10, 43, 80, 100, 117, 119, 144, 189

[7] HALEVI, T., MEMON, N., LEWIS, J., KUMARAGURU, P., ARORA, S., DAGAR, N., ALOUL, F., AND CHEN, J. Cultural and psychological factors in cyber-security. In *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services* (2016), pp. 318–324. 80, 100, 117, 144

[8] HAPPELL, B., DWYER, T., REID-SEARL, K., BURKE, K. J., CAPERCHIONE, C. M., AND GASKIN, C. J. Nurses and stress: recognizing causes and seeking solutions. *Journal of nursing management 21*, 4 (2013), 638–647. 2, 100, 117

[9] KENNISON, S. M., AND CHAN-TIN, E. Taking risks with cybersecurity: Using knowledge and personal characteristics to predict self-reported cybersecurity behaviors. *Frontiers in Psychology 11* (2020), 3030. 80, 100, 117, 144

[10] Krombholz, K., Hobel, H., Huber, M., and Weippl, E. " advanced social engineering attacks"; journal of information security and applications, 22 (2015), s. 113-122. 1, 100, 116

[11] McCormac, A., Calic, D., Parsons, K., Butavicius, M., Pattinson, M., and Lillie, M. The effect of resilience and job stress on information security awareness. *Information & Computer Security* (2018). 3, 14, 16, 40, 100, 108, 117, 120, 132, 144, 153, 163, 165

[12] McVicar, A. Workplace stress in nursing: a literature review. *Journal of advanced nursing 44*, 6 (2003), 633–642. 2, 100, 117

[13] Mubarak, F., and Nycyk, M. Teaching older people internet skills to minimize grey digital divides: Developed and developing countries in focus. *Journal of Information, Communication and Ethics in Society* (2017). 109, 133

[14] Schoofs, D., Wolf, O. T., and Smeets, T. Cold pressor stress impairs performance on working memory tasks requiring executive functions in healthy young men. *Behavioral neuroscience 123*, 5 (2009), 1066. 2, 100, 117

[15] Siponen, M., and Vance, A. Neutralization: New insights into the problem of employee information systems security policy violations. *MIS quarterly* (2010), 487–502. 1, 100, 116

[16] Taber, K. S. The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education 48*, 6 (2018), 1273–1296. 83, 105, 106, 126, 127, 149, 150

[17] Tawfik, D. S., Scheid, A., Profit, J., Shanafelt, T., Trockel, M., Adair, K. C., Sexton, J. B., and Ioannidis, J. P. Evidence relating health care provider burnout and quality of care: a systematic review and meta-analysis. *Annals of internal medicine 171*, 8 (2019), 555–567. 2, 100, 117

[18] UiO. Hva er nettskjema, June 2010. Available from: `http://www.uio.no/tjenester/it/applikasjoner/nettskjema/mer-om/`. 101, 122

[19] Verizon. 2022 data breach investigations report. *Available online: https://enterprise.verizon.com/resources/reports/dbir (accessed on 18 March 2023)* (2022). 1, 80, 100, 116, 144, 162

[20] Whitty, M., Doodson, J., Creese, S., and Hodges, D. Individual differences in cyber security behaviors: an examination of who is sharing passwords. *Cyberpsychology, Behavior, and Social Networking 18*, 1 (2015), 3–7. 80, 100, 117, 144

[21] Yeng, P. K., Yang, B., and Snekkenes, E. A. Framework for healthcare security practice analysis, modeling and incentivization. In *2019 IEEE International Conference on Big Data (Big Data)* (2019), IEEE, pp. 3242–3251. 80, 100, 117, 144

[22] Yusoff, M. S. B. Stability of dreem in a sample of medical students: a prospective study. *Education Research International 2012* (2012). 105, 106, 126, 127, 149, 150

Chapter 4

Stress and Cybersecurity Practices among Hospital Staff in the Digital Age: An Empirical Study from Ghana

Muhammad Ali Fauzi; Prosper Yeng; Bian Yang; Peter Nimbe; Dita Rachmayani

This paper is awaiting publication and is not included

Chapter 5

Examining the Relationship Between Stress Levels and Cybersecurity Practices Among Hospital Employees in Three Countries: Ghana, Norway, and Indonesia

Muhammad Ali Fauzi; Prosper Yeng; Bian Yang; Dita Rachmayani; Peter Nimbe

Abstract

This study aims to investigate the relationship between stress levels among hospital staff and their risky cybersecurity practices. A web-based survey was conducted with a sample of 353 hospital staff from Ghana, Norway, and Indonesia. The results indicate a statistically significant positive correlation between the stress levels of hospital staff and their engagement in unsafe cybersecurity practices (r = 0.201, p $<$ 0.01). Specifically, the study finds that staff members' inclination to click on links from unknown sources is the cybersecurity practice most strongly influenced by stress levels. The study did not observe any significant differences in cybersecurity practices based on gender, age, job, position level, or work experience. However, it does highlight notable differences in cybersecurity practices across countries, with Norwegian hospital staff exhibiting better cybersecurity practices than their counterparts from Ghana and Indonesia.

5.1 Introduction

Electronic health records (EHRs), telemedicine, and remote patient monitoring systems have all been adopted in recent years, undergoing a considerable digital transition in the healthcare sector. Even though these technological developments have improved patient outcomes and the quality

~~of treatment, they have also presented new cybersecurity threats~~ to hospitals and their workers. Due to the enormous volumes of sensitive patient data that are electronically kept and exchanged, the healthcare sector is especially susceptible to cyber-attacks [18]. Data breaches, financial losses, reputational harm, and, most crucially, jeopardized patient care are all possible outcomes of these attacks [14].

It is generally known that human factors are one of the major causes of cybersecurity breaches. Human error may compromise even the most sophisticated technological security measures [19, 22]. For instance, according to a recent Verizon research, humans were involved in 82% of all data leaks [18]. Therefore, many prior studies focused on understanding how the human factor can affect cybersecurity practices and identify the factors that affect cybersecurity practices [21, 9, 7, 23, 10].

Stress is one of the human factors that can affect cybersecurity practices. Stress can lead to lapses in judgment, increased impulsivity, and a reduced ability to make rational decisions [16, 13, 20]. In the context of cybersecurity, stress may lead to unsafe cybersecurity practices, such as clicking on suspicious links or responding to phishing emails. Hence, it is essential to understand the impact of stress on cybersecurity practices. However, only a few studies focused on this topic, especially in the healthcare setting. Moreover, most of the studies were conducted in developed countries. McCormac et al. [12] analyzed the effect of job stress on information security awareness among company workers in Australia while Fordyce et al. [6] investigated the effect of stress on password choice among students in United Kindom. There is no study on this topic conducted in developing countries.

This study aims to investigate the relationship between hospital staff stress levels and cybersecurity practices in Norway and two developing countries, Ghana and Indonesia. This study follows the hypothesis that hospital workers with higher stress levels engage in riskier security practices. Additionally, we will also compare the cybersecurity practices between these three countries and examine the relationship between demographic variables and cybersecurity practices. By examining this relationship, this study can contribute to the existing literature on cybersecurity in the healthcare industry and provide insights for hospitals to improve their cybersecurity practices.

5.2 Materials and Methods

5.2.1 Research approach

This study's primary objective was to examine the relationship between stress levels and cybersecurity practices among hospital staff. The research approach is outlined in Figure 5.1. To achieve this objective, an online sur-

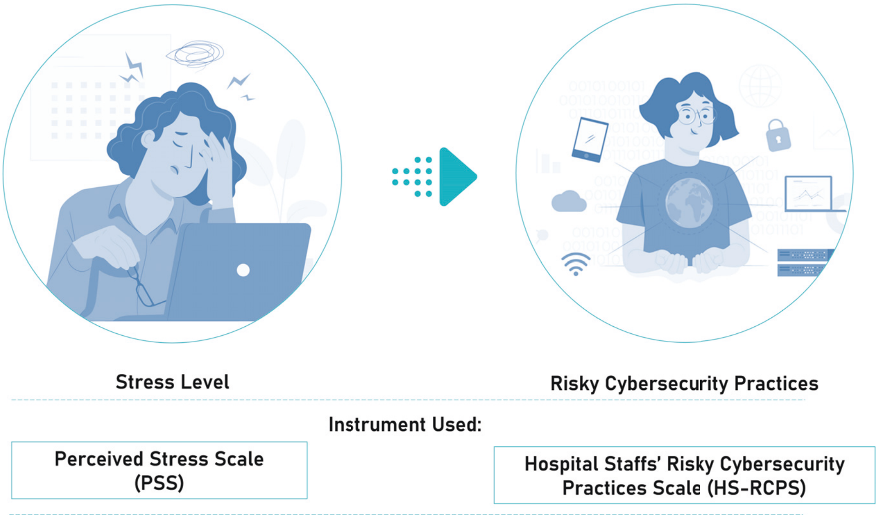

Stress Level
Risky Cybersecurity Practices
Instrument Used:
Perceived Stress Scale (PSS)
Hospital Staffs' Risky Cybersecurity Practices Scale (HS-RCPS)

Figure 5.1: Proposed Approach.

vey developed using Nettskjema was utilized to collect data on healthcare staff's demographic details, stress levels, and cybersecurity practices within the past month. Nettskjema is an online survey platform that places a high priority on data privacy and security run by the University of Oslo. The survey was composed in English for participants from Ghana, Norwegian for participants from Norway, and Indonesian for participants from Indonesia. The participants' stress levels and risky cybersecurity practices were assessed using the Perceived Stress Scale (PSS) and Hospital Staff's Risky Cybersecurity Practices Scale (HS-RCPS), respectively. Additionally, the questionnaire included an attention-checking question to guarantee the response's quality. Hospital employees from three hospitals in Ghana, a hospital in Norway, and a hospital from Indonesia were invited to participate in the study. All participants provided their written consent electronically, and the surveys' completion and analysis were completely anonymous.

5.2.2 Perceived Stress Scale (PSS)

The Perceived Stress Scale (PSS) is a self-report survey used to measure individuals' perceptions of stress. It uses a 5-point Likert scale to assess the frequency of thoughts and feelings connected to stress during the previous month. PSS assesses the subjective experience of stress rather than specific stressors. PSS has three variations: PSS-14, PSS-10, and PSS-4, with PSS-10 having superior psychometric properties than the others [3]. PSS has been translated and validated in many languages, indicating its cross-cultural ap-

~~plicability. This study will use the Norwegian language version~~ of PSS. In this study, we used the original English version of PSS-10 for participants from Ghana. Meanwhile, the Indonesian version by Pin [15] was used for Indonesian participants and the Norwegian version translated by CheckWare AS [4] was employed for Norwegian participants.

5.2.3 Hospital Staff's Risky Cybersecurity Practices Scale (HS-RCPS)

The Hospital Staff's Risky Cybersecurity Practices Scale (HS-RCPS) was developed to evaluate hospital staff's cybersecurity practices based on the Human Aspects of Information Security Questionnaire (HAIS-Q) and the Security Behavior Intentions Scale (SeBIS). The scale was tailored specifically to measure the cybersecurity practices of healthcare workers based on feedback from interviews with 36 healthcare employees and cybersecurity professionals from various universities and hospitals in Ghana, Indonesia, and Norway. The scale consists of 12 items with a possible total score ranging from 0 to 48 as depicted in Table 5.1. Higher scores indicate riskier cybersecurity practices over the past month. Participants were asked to rate their engagement in specific cybersecurity practices using a scale of 0 to 4 (”disagree” to ”agree”). This scale is available in English, Norwegian, and Indonesian versions.

5.2.4 Data Analysis

The present study employed SPSS software to analyze the collected data. The reliability of the PSS and HS-RCPS was measured using Cronbach's alpha. Furthermore, Pearson's correlation coefficient was utilized to evaluate the relationship between the PSS and HS-RCPS scales. The mean differences among various demographic groups, such as age, position, position level, and work experience, were assessed using ANOVA. A t-test was utilized to evaluate the mean difference in HS-RCPS scores between males and females. Additionally, a Kruskal-Wallis test with Bonferroni-Dunn posthoc analysis was conducted to examine the variance in HS-RCPS scores among staff groups based on their country.

5.3 Results

5.3.1 General characteristics of participants

In total, 389 hospital employees participated but 36 of them failed to answer the attention-checking question correctly. As seen in Table 5.2, 353 qualified participants were finally included in the study, with 212 participants from Ghana, 42 from Norway, and 99 from Indonesia. Based on gender, 209

Table 5.1: Items for the HS-RCPS

	Item
1	In the last month, I usually write my username and password on a piece of paper and stick the paper onto my computer for easy access
2	In the last month, I sometimes visit at least one of the following websites using the hospital's computer: social media; Dropbox and other public file storage systems; online music or videos sites; online newspapers and magazines; personal e-mail accounts; games; instant messaging services, etc
3	In the last month, I did not often read the alert messages/emails concerning security
4	In the last month, I sometimes click on a link in an email from an unknown sender
5	In the last month, I usually postpone software updating activities (restarting, clicking to run an update, accepting to update, or following the update schedule) of my computers at my workplace
6	In the last month, I usually postpone backup activities when I am prompted
7	In the last month, I usually do not prevent my colleagues from seeing patients' records for a non-therapeutic purpose when I am working on patient information on my laptop
8	In the last month, I did not post patient information on social media
9	In the last month, I sometimes share my passwords with my colleagues in the hospital
10	In the last month, I usually do not take any action when I notice my colleague ignoring information security rules
11	In the last month, I used a combination of letters, numbers, and symbols in my work passwords
12	In the last month, I have changed my passwords

Table 5.2: Participant Characteristics

Variable	Category	n	%
Country	Ghana	212	60.10 %
	Norway	42	11.90 %
	Indonesia	99	28.00 %
Gender	Female	209	59.20 %
	Male	143	40.50 %
	Prefer not to say	1	0.30 %
Age	21-31	117	33.10 %
	31-40	168	47.60 %
	41-50	47	13.30 %
	Over 50	21	5.90 %
Position	Top Level Management	15	4.20 %
	Doctor	34	9.60 %
	Nurse	180	51.00 %
	Lab staff	22	6.20 %
	Pharmacy staff	28	7.90 %
	IT staff	14	4.00 %
	Researcher	3	0.80 %
	Nutritionist	3	0.80 %
	Other	54	15.30 %
Position level	Executive	5	1.40 %
	Managers and supervisors	59	16.70 %
	Operational staff	289	81.90 %
Work experience	<6 Year	131	37.10 %
	6-15 Years	166	47.00 %
	16-25 Years	46	13.00 %
	>25 Years	10	2.80 %

(59.20%) of them are females and 143 of them are males (40.50%). One participant (0.30%) preferred not to disclose their gender. The age range of the participants varied, with 117 (33.10%) falling in the 21-31 years category, 168 (47.60%) falling in the 31-40 years category, 47 (13.30%) falling in the 41-50 years category, and 21 (5.90%) falling in the over 50 years category. Regarding participants' positions, 15 (4.20%) were in top-level management, 34 (9.60%) were doctors, 180 (51.00%) were nurses, 22 (6.20%) were lab staff, 28 (7.90%) were pharmacy staff, 14 (4.00%) were IT staff, 3 (0.80%) were researchers, 3 (0.80%) were nutritionists, and 54 (15.30%) reported other positions. The participants' position level was categorized as executives (1.40%), managers and supervisors (16.70%), and operational staff (81.90%). Regarding work experience, 131 (37.10%) participants had less than six years of experience, 166 (47.00%) had 6-15 years of experience, 46 (13.00%) had 16-25

Table 5.3: Descriptive statistic of PSS score in Ghana, Norway, and Indonesia.

Country	Min	Max	Mean	SD
Ghana	1	27	16.12	5.23
Norway	1	29	14.05	6.40
Indonesia	3	21	13.89	4.41
All	1	29	15.25	5.28

years of experience, and 10 (2.80%) had more than 25 years of experience.

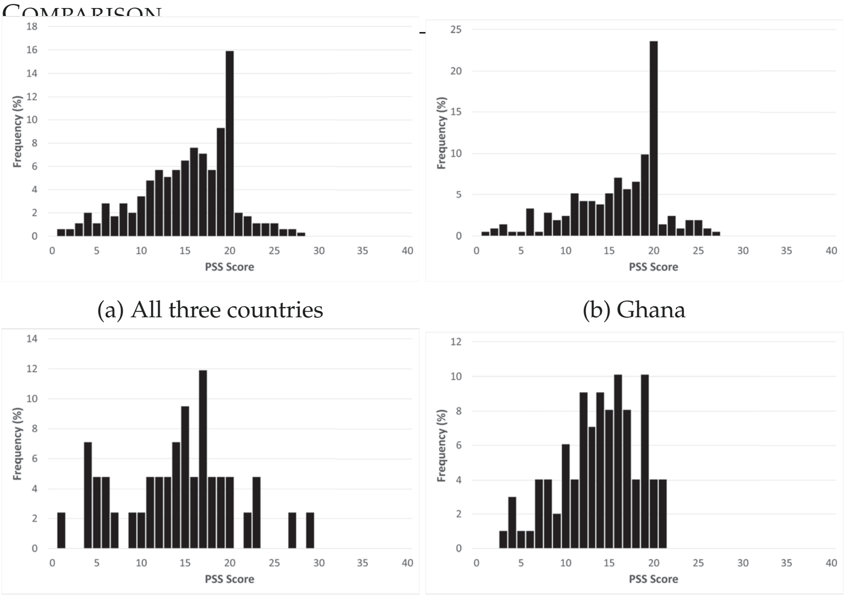
5.3.2 PSS Score

Figure 5.2 displays the distribution of Perceived Stress Scale (PSS) scores among the study participants. The PSS is a self-reported scale that measures the degree to which individuals perceive their lives as stressful. The scores range from 0 to 40, with higher scores indicating higher levels of perceived stress during the past month. The statistic of the PSS score from the three countries is depicted in table 5.3. Ghana had the highest PSS score average, followed by Norway, and Indonesia became the last. From Ghana, the PSS scores reported ranged from 1 to 27, with an average score of 16.12 and a standard deviation of 5.23. The lowest PSS score was obtained by one participant with 1, while the highest score was also reported by one participant with 27. From Norway, the PSS scores reported ranged from 1 to 29, with an average score of 14.05 and a standard deviation of 6.4. The lowest PSS score was obtained by one staff member, while the highest score was also reported by one staff member. From Indonesia, the PSS scores reported ranged from 3 to 21, with an average score of 13.89 and a standard deviation of 4.41. The lowest PSS score was obtained by one participant, while the highest score was reported by four participants. Combining all of the results from these three countries, the mean PSS score was 15.25 with a standard deviation of 5.28.

Finally, we assessed the reliability of the PSS. According to the survey results, PSS in English, Norwegian, and Indonesian versions had Cronbach's α of 0.750, 0.844, and 0.733, respectively. It indicates that the items in all three PSS versions had a good level of internal consistency [17, 24].

5.3.3 HS-RCPS Score

The distribution and statistics of the HS-RCPS scores are shown in Figure 5.3 and Table 5.4. HS-RCPS is a scale of 0 to 48, with 0 denoting the lowest risky cybersecurity practice and 48 denoting the highest. Overall, the results showed that the mean HS-RCPS score for all three countries was 14.94, with



(a) All three countries

(b) Ghana

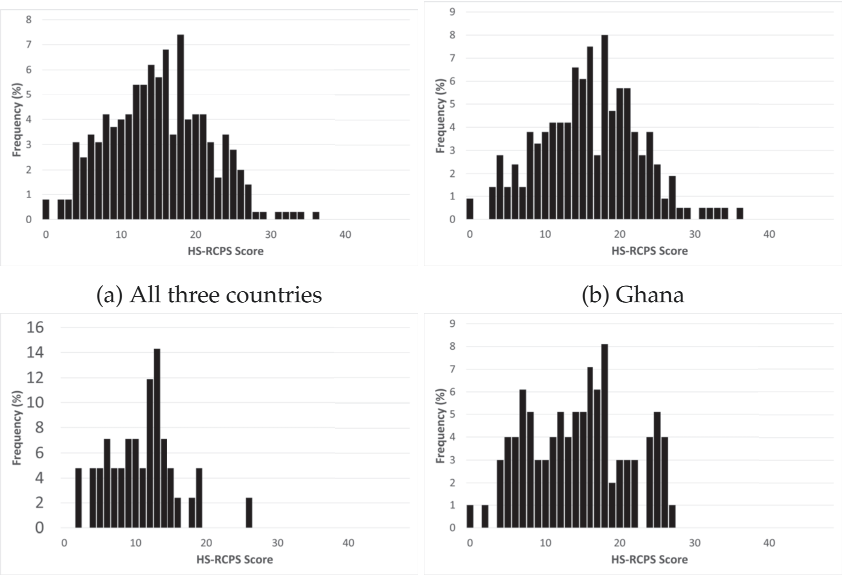
(c) Norway

(d) Indonesia

Figure 5.2: Frequency distribution of the PSS score

a standard deviation of 6.64. From Ghana, the results indicate that the minimum HS-RCPS score among the participants was 0, while the maximum was 36. The mean score was 15.95 with a standard deviation of 6.64. Meanwhile, the minimum and maximum scores in Norway were 2 and 26, respectively, with a mean of 10.88 and a standard deviation of 4.90. Finally, in Indonesia, the minimum and maximum scores were 0 and 27, respectively, with a mean of 14.49 and a standard deviation of 6.64. The findings suggest that risky cybersecurity practice is relatively low among individuals in the three countries. Comparatively, Ghana had the highest mean score while Norway had the lowest.

Furthermore, the reliability of the HS-RCPS was assessed through survey results obtained from Ghana, Norway, and Indonesia. The scale's internal consistency was evaluated using Cronbach's α coefficient. The survey results from Ghana, Norway, and Indonesia indicated that the scale had a Cronbach's α of 0.595, 0.502, and 0.697, respectively. Overall, the HS-RCPS demonstrated acceptable internal consistency across the surveyed populations [2, 17, 24].



(a) All three countries

(b) Ghana

(c) Norway

(d) Indonesia

Figure 5.3: Frequency distribution of the HS-RCPS scores

Table 5.4: Descriptive statistic of HS-RCPS score in Ghana, Norway, and Indonesia.

Country	Min	Max	Mean	SD
Ghana	0	36	15.95	6.64
Norway	2	26	10.88	4.90
Indonesia	0	27	14.49	6.64
All	0	36	14.94	6.64

5.3.4 Demographic and Risky Cybersecurity Practices

The descriptive statistic of risky cybersecurity practices score based on gender, age, position, position level, and work experience are displayed in Table 5.5, 5.6, 5.7, 5.8, and 5.9 respectively. The statistical analysis revealed no significant differences in the levels of risky cybersecurity practices between male and female participants. Technically, the t-test results indicated that the mean scores for females (M = 14.65, SD = 6.47) and males (M = 15.32, SD = 6.90) were not significantly different, t(350) = 0.980, p = 0.323. In addition, the ANOVA results indicated that there were no significant differences in mean scores of risky cybersecurity practices among various groups, including age (F(3, 349) = 0.347, p = 0.791), position (F(8, 344) = 1.774, p = 0.081),

Table 5.5: Descriptive statistic of risky cybersecurity practices score based on gender.

Gender	Number of Participants	Mean	SD
Female	209	14.6507	6.4657
Male	143	15.3217	6.9023

Table 5.6: Descriptive statistic of risky cybersecurity practices score based on age.

Age (years)	Number of Participants	Mean	SD
Less than 31	117	15.3932	6.7924
31-40	168	11.0714	4.7307
41-50	47	8.4000	3.8355
Over 50	21	12.2857	7.4992

Table 5.7: Descriptive statistic of risky cybersecurity practices score based on position.

Position	Number of Participants	Mean	SD
Top Level Management	15	14.2667	9.0984
Doctor	34	14.7353	5.8946
Nurse	180	15.0500	6.2929
Lab staff	22	16.9091	6.6541
Pharmacy staff	28	17.2143	5.4525
IT staff	14	13.5000	6.4896
Researcher	3	18.3333	7.7675 %
Nutritionist	3	19.6667	4.5093
Other	54	12.8148	7.6160

position level (F(2, 350) = 0.144, p = 0.866), and work experience (F(3, 349) = 1.369, p = 0.252).

On the other hand, the Kruskal-Wallis test showed that the scores for risky cybersecurity practices varied significantly across various staff groups based on country ($\chi^2(2)$ = 23.124, $p < 0.001$). Specifically, the scores for hospital staff from Norway were significantly lower than those from Ghana and Indonesia (p = 0.000 and p=0.04, respectively), suggesting that hospital staff from Norway have better cybersecurity practices.

5.3.5 Correlation Between Stress Level and Cybersecurity Practices

Table 5.10 presents the correlation between the perceived stress levels of hospital staff and their risky cybersecurity practices. The results reveal that

Table 5.8: Descriptive statistic of risky cybersecurity practices score based on position level.

Position Level	Number of Participants	Mean	SD
Executive	5	15.6000	8.5615
Managers and supervisors	59	14.5424	7.1615
Operational staff	289	15.0069	6.52026

Table 5.9: Descriptive statistic of risky cybersecurity practices score based on years of work experience.

Work Experience	Number of Participants	Mean	SD
<6 Year	131	14.7481	7.0769
6-15 Years	166	15.4639	6.2935
16-25 Years	46	14.3043	6.5553
>25 Years	10	11.6000	6.3631

there was a statistically significant positive correlation between staff's stress levels and their cybersecurity practices, as indicated by a Pearson's correlation coefficient of $r = 0.201$ ($p < 0.01$). This finding suggests that employees who reported higher levels of stress, as measured by the Perceived Stress Scale (PSS), were also more likely to engage in riskier cybersecurity practices, as assessed by the Hospital Staff Risky Cybersecurity Practices Scale (HS-RCPS). Specifically, item 4 of the HS-RCPS, which measures staff's tendency to click on links from unknown sources, had the highest positive correlation with stress levels, indicating that this is the riskiest cybersecurity behavior that is most influenced by stress levels among hospital staff. In addition, this significant correlation also appears when we analyze only the data from Ghana or only the data from Indonesia with $r = 0.138$ ($p < 0.05$) and $r = 0.311$ ($p < 0.01$), respectively. However, a significant correlation between stress and risky cybersecurity practices was not found in Norway ($r = 0.101$).

5.4 Discussion

The results of this study have important implications for organizations concerned with cybersecurity and employee well-being. The positive correlation between stress levels and risky cybersecurity practices supports the notion that stress can impair cognitive functioning and increase the likelihood of individuals engaging in risky behavior, including online behavior. This is consistent with the broader literature on the negative effects of stress on decision-making [13, 20]. This result was also supported by Fauzi et al. [5] and McCormac et al. [12] who found that workers with greater levels of

Table 5.10: PSS score correlation to cybersecurity practices score.

Item	Correlation coefficient
I-1	-0.064
I-2	0.058
I-3	0.111*
I-4	0.259**
I-5	0.046
I-6	0.121*
I-7	0.12*
I-8	0.162**
I-9	0.070
I-10	0.037
I-11	0.091
I-12	0.066
HS-RCPS	0.201**
HS-RCPS in Ghana	0.138*
HS-RCPS in Norway	0.101
HS-RCPS in Indonesia	0.311**

*$p < 0.05$.
**$p < 0.01$.

stress engaged in riskier cybersecurity practices or had worse information security awareness (ISA).

From a practical perspective, these findings highlight the importance of addressing stress and well-being in the context of cybersecurity training and awareness programs. Specifically, organizations should consider incorporating stress management techniques and well-being training into their cybersecurity training programs to help employees manage stress and reduce their engagement in risky cybersecurity practices. Having an understanding of how stress can influence an individual's cybersecurity practices, one can take measures to regulate their stress levels and maintain a heightened awareness of their cybersecurity practices. These measures could comprise tactics such as taking breaks to alleviate stress, exercising increased mindfulness with regard to cybersecurity practices while experiencing stress, and seeking assistance as necessary.

Furthermore, the finding that clicking on links from unknown sources was the riskiest cybersecurity behavior most influenced by stress levels is also reasonable since stress can harm an individual's cognitive functioning, impairing their ability to make rational decisions and increasing the likelihood of impulsive behavior [16]. In addition, stress can lead to feelings of anxiety or overwhelm, causing individuals to rush through tasks or pay less attention to details, making them more likely to overlook the signs of a phishing email [11].

In addition, the research results also revealed a significant difference in cybersecurity practices between healthcare professionals in Norway, a developed country, and those in Ghana and Indonesia, two developing countries. Developing nations have historically slowly adopted and utilized computer and internet technologies. As identified by Ben-David et al.'s research [1], developing nations' security landscape is affected by five fundamental factors: inadequate "security hygiene," unique resource constraints (such as one computer for multiple users), novice internet users, use of pirated software, and limited comprehension of cybersecurity adversaries. These factors could explain why people in developing countries generally exhibit poorer cybersecurity practices than their counterparts in developed nations. Insufficient IT education and a lack of computer and internet manuals in local languages have also contributed to unsafe cybersecurity practices [8]. Moreover, Norway's healthcare systems and infrastructure are comparatively advanced and better equipped to implement and enforce cybersecurity protocols than Indonesia and Ghana. Future research can investigate cultural factors and explore how they may be leveraged to improve cybersecurity practices in different regions.

5.5 Limitation

There are several limitations of this study that need to be acknowledged. First, the study used a self-report survey to collect data, which may result in social desirability bias, meaning that participants may have needed to be more honest in their responses. Second, memory bias could also occur when participants have trouble remembering details correctly, particularly if the details relate to previous events or behaviors. Finally, the study's cross-sectional design precludes the establishment of causality. Using this study design, it is difficult to determine if high-stress levels cause risky cybersecurity practices or if it is the other way around.

5.6 Conclusions

In conclusion, this study explored the relationship between stress levels and risky cybersecurity practices among hospital staff in three countries. The results showed a statistically significant positive correlation between staff stress levels and their engagement in riskier cybersecurity practices. Specifically, the staff's tendency to click on links from unknown sources was found to be the risky cybersecurity practice most heavily associated with higher stress levels. Interestingly, no significant differences were found in the levels of risky cybersecurity practices between male and female participants or among different age groups, positions, position levels, and work experi-

~~ence. However, a significant difference was observed in risky cy~~bersecurity practices scores across staff groups based on the country of origin, with hospital staff from Norway showing significantly lower scores than those from Ghana and Indonesia, suggesting Norwegian healthcare staff had safer cybersecurity practices.

There are several directions that future studies can take based on the findings of this study. Firstly, further research can explore the causal relationship between stress levels and cybersecurity practices. Second, future studies can explore other factors influencing risky cybersecurity practices among hospital staff, such as personality traits, motivation, or job satisfaction. By gaining a more comprehensive understanding of the various factors that influence cybersecurity practices, interventions can be developed that target these factors to promote safer cybersecurity behaviors among employees. Finally, future studies can examine the effectiveness of various interventions aimed at promoting safer cybersecurity practices among hospital staff. Such interventions may include training programs, awareness campaigns, or technological solutions such as secure communication platforms.

5.7 Bibliography

[1] Ben-David, Y., Hasan, S., Pal, J., Vallentin, M., Panjwani, S., Gutheim, P., Chen, J., and Brewer, E. A. Computing security in the developing world: A case for multidisciplinary research. In *Proceedings of the 5th ACM workshop on Networked systems for developing regions* (2011), pp. 39–44. 41, 48, 155

[2] Berger, R., and Hänze, M. Impact of expert teaching quality on novice academic performance in the jigsaw cooperative learning method. *International Journal of Science Education 37*, 2 (2015), 294–320. 106, 127, 150

[3] Cohen, S., Kamarck, T., and Mermelstein, R. Perceived stress scale (pss). *J Health Soc Beh 24* (1983), 285. 4, 11, 23, 81, 82, 102, 123, 145, 246, 258

[4] Department of Psychology, C. M. U. Scales - laboratory for the study of stress, immunity, and disease, Feb. 2015. Available from: `https://www.cmu.edu/dietrich/psychology/stress-immunity-disease-lab/scales/index.html`. 102, 146

[5] Fauzi, M. A., Yeng, P., Yang, B., and Rachmayani, D. Examining the link between stress level and cybersecurity practices of hospital staff in indonesia. In *Proceedings of the 16th International Conference on*

Availability, Reliability and Security (2021), pp. 1–8. 14, 16, 21, 27, 100, 108, 117, 120, 132, 153, 164, 202, 218, 232, 246, 258

[6] Fordyce, T., Green, S., and Gross, T. Investigation of the effect of fear and stress on password choice. In *Proceedings of the 7th Workshop on Socio-Technical Aspects in Security and Trust* (2018), pp. 3–15. 3, 13, 20, 21, 41, 120, 144, 164, 187

[7] Gratian, M., Bandi, S., Cukier, M., Dykstra, J., and Ginther, A. Correlating human traits and cyber security behavior intentions. *computers & security 73* (2018), 345–358. 10, 43, 80, 100, 117, 119, 144, 189

[8] Grobler, M., and van Vuuren, J. J. Broadband broadens scope for cyber crime in africa. In *2010 Information Security for South Africa* (2010), IEEE, pp. 1–8. 41, 49, 155

[9] Halevi, T., Memon, N., Lewis, J., Kumaraguru, P., Arora, S., Dagar, N., Aloul, F., and Chen, J. Cultural and psychological factors in cyber-security. In *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services* (2016), pp. 318–324. 80, 100, 117, 144

[10] Kennison, S. M., and Chan-Tin, E. Taking risks with cybersecurity: Using knowledge and personal characteristics to predict self-reported cybersecurity behaviors. *Frontiers in Psychology 11* (2020), 3030. 80, 100, 117, 144

[11] Liu, Q., Liu, Y., Leng, X., Han, J., Xia, F., and Chen, H. Impact of chronic stress on attention control: Evidence from behavioral and event-related potential analyses. *Neuroscience bulletin 36* (2020), 1395–1410. 40, 155

[12] McCormac, A., Calic, D., Parsons, K., Butavicius, M., Pattinson, M., and Lillie, M. The effect of resilience and job stress on information security awareness. *Information & Computer Security* (2018). 3, 14, 16, 40, 100, 108, 117, 120, 132, 144, 153, 163, 165

[13] Michailidis, E., and Banks, A. P. The relationship between burnout and risk-taking in workplace decision-making and decision-making style. *Work & Stress 30*, 3 (2016), 278–292. 2, 11, 40, 41, 80, 119, 144, 153, 162, 187

[14] Perakslis, E. D. Cybersecurity in health care. *N Engl J Med 371*, 5 (2014), 395–397. 1, 11, 80, 144

[15] Pin, T. L. Hubungan kebiasaan berolahraga dengan tingkat stres pada mahasiswa fakultas kedokteran universitas sumatera utara tahun masuk 2008. *Skripsi. Medan: Fakultas Kedokteran Universitas Sumatera Utara* (2011). 83, 146

[16] Simon, L., Jiryis, T., and Admon, R. Now or later? stress-induced increase and decrease in choice impulsivity are both associated with elevated affective and endocrine responses. *Brain Sciences 11*, 9 (2021), 1148. 11, 40, 119, 144, 154, 162

[17] Taber, K. S. The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education 48*, 6 (2018), 1273–1296. 83, 105, 106, 126, 127, 149, 150

[18] Verizon. 2022 data breach investigations report. *Available online: https://enterprise.verizon.com/resources/reports/dbir (accessed on 18 March 2023)* (2022). 1, 80, 100, 116, 144, 162

[19] Warkentin, M., and Willison, R. Behavioral and policy issues in information systems security: the insider threat. *European Journal of Information Systems 18*, 2 (2009), 101–105. 1, 80, 116, 144

[20] Wemm, S. E., and Wulfert, E. Effects of acute stress on decision making. *Applied psychophysiology and biofeedback 42*, 1 (2017), 1–12. 2, 11, 40, 41, 80, 119, 144, 153, 162, 187, 202, 246, 258

[21] Whitty, M., Doodson, J., Creese, S., and Hodges, D. Individual differences in cyber security behaviors: an examination of who is sharing passwords. *Cyberpsychology, Behavior, and Social Networking 18*, 1 (2015), 3–7. 80, 100, 117, 144

[22] Yan, Z., Robertson, T., Yan, R., Park, S. Y., Bordoff, S., Chen, Q., and Sprissler, E. Finding the weakest links in the weakest link: How well do undergraduate students make cybersecurity judgment? *Computers in Human Behavior 84* (2018), 375–382. 1, 80, 116, 144

[23] Yeng, P. K., Yang, B., and Snekkenes, E. A. Framework for healthcare security practice analysis, modeling and incentivization. In *2019 IEEE International Conference on Big Data (Big Data)* (2019), IEEE, pp. 3242–3251. 80, 100, 117, 144

[24] Yusoff, M. S. B. Stability of dreem in a sample of medical students: a prospective study. *Education Research International 2012* (2012). 105, 106, 126, 127, 149, 150

Part II

Causal Analysis

Chapter 6

Can Stress Compromise Phishing Email Detection?

Muhammad Ali Fauzi; Bian Yang; Katrien De Moor; Prosper Yeng; Dita Rachmayani; Christoph Busch; Mark Wetherell

This paper is awaiting publication and is not included

Part III

Effective Stress Detection

Chapter 7

Multiple Sensor Fusion for Stress Detection in the Hospital Environment

Muhammad Ali Fauzi; Bian Yang

This paper is awaiting publication and is not included

Chapter 8

Improving Stress Detection Using Weighted Score-Level Fusion of Multiple Sensor

Muhammad Ali Fauzi; Bian Yang; Prosper Yeng

Abstract

Work-related stress is now a widespread issue in our modern life. Consequently, stress management is crucial. Automated stress detection is one of the innovations in helping people manage their well-being better by providing information about their stress levels. The advancement of sensor technology and artificial intelligence has made this task easier. The assessment of stress levels by using a variety of sensors from a smartwatch and machine learning algorithms has been very popular in recent years. The use of multiple sensor data enables richer information to train the machine learning algorithm so that the trained model can be more robust. However, if we use a feature-level fusion, since it is the most popular fusion strategy, generally, each sensor data will have the same significance. In fact, stress is personal and each subject can have a different reaction to the stress so that a particular sensor may be an effective stress indicator for some subjects but it might not be for others. Therefore, we propose a personalized stress detection system based on a weighted score-level multiple sensor fusion strategy. For each individual, this strategy gives different weights to each sensor based on the performance of the sensor on the individual's data. The experiment results show that both feature-level and weighted score-level fusion models obtained better performance than models from the individual sensor strategy. The weighted score-level fusion strategy achieved a better performance than the feature-level strategy with accuracy, precision, recall, and F_1-measure of 0.931, 0.824, 0.939, and 0.868, respectively.

8.1 Introduction

Work-related stress has become a common problem in modern society. The introduction of COVID-19 made it worse. Based on Gallup's State of the Global Workplace 2022 Report [5], 44% of the workforce reported high levels of everyday stress in 2021. This number is an all-time high, increasing from 43% in 2020 (the previous all-time high). Stress may harm one's physical and mental health [18, 13]. In terms of work performance, higher stress level correlates with higher absenteeism, lower employer responsibility, poor customer care, and riskier cybersecurity behavior [9, 20, 19, 4]. Therefore, stress management is important. Automated stress detection is one of the crucial parts to help people manage their mental being better by giving information about their stress levels.

One of the most popular techniques for determining stress levels in recent years is the assessment of physiological reactions to stress using a variety of sensors and machine learning (ML) algorithms. One of the main concerns of stress detection in a working environment is usability. A lot of sensors are not very convenient to be used at work and can limit the user's activity (e.g. chest-worn sensors, finger-placed EDA sensors).

Smartwatch is one suitable device for this task due to its usability and its high degree of social acceptance in society. This kind of device also has several built-in sensors that can be utilized for stress detection using multiple sensor data, which has more potential than using only one single sensor data. By using multiple sensor data, richer information as the result of a combination of many environmental views can be used to train the machine learning algorithm so that the trained model can be more robust.

Regarding how to combine numerous sensors data, there are several different ways with feature-level fusion has become the most common approach (e.g. [14, 6, 8, 16]). However, stress is personal and each subject can have a different reaction to the stress. Various conditions tend to induce different patterns of stress responses and also there are individual differences in stress responses to the same situation [15]. As a result, whereas a particular sensor may be an effective stress indicator for some subjects, it might not be for others. For example, a study by Sommerfeldt et al. [17] shows that people had different heart rates even though they reported the same stress level. In the feature-level fusion approach, we cannot assign distinct weights to each sensor for each subject. Therefore, in this paper, we propose a personalized stress detection system based on a weighted score-level multiple sensor fusion strategy. Different from feature-level fusion where the combination is conducted before the data is fed to the ML algorithm, the proposed strategy does the combination after the ML model predicts the result so that this strategy can be categorized as decision-level fusion. By using this strategy, for each subject, we will assign a greater weight to the sensor model that can make better predictions on the subject data. Different from the decision-level strategy that usually combines the category result (e.g. using majority voting), the proposed weighted score-level fusion strategy employed the probability score of the testing data belonging to the stress category. This weighted probability score combination is expected to improve classification performance. Furthermore, the ML algorithm that will be used in this work is Logistic Regression because it did well in previous work for stress categorization [3, 11, 2].

8.2 Materials and Methods

8.2.1 Dataset

The dataset used in this study is a publicly available stress detection dataset called WESAD (Wearable Stress and Affect Detection) [14]. This dataset includes data from motion and physiological sensors recorded using a chest-worn RespiBAN device and smartwatch Empatica E4 device from 15 subjects. Three sessions were conducted in the data collection including baseline, amusement, and stress situations. Trier Social Stress Test (TSST) protocol [10] is employed for producing stress while amusement is induced using funny video clips. Meanwhile, in the baseline situation, the subject was asked to read given neutral reading materials.

Even though chest-worn data were also available in this dataset, this study exclusively uses Empatica E4 data because of the usability of the smartwatch. Chest-worn sensors do not have high usability to be worn every day in the working environment. As displayed in Figure 8.1, Empatica E4 provides several sensors including electrodermal activity (EDA), accelerometers (ACC), skin temperature (ST), and blood volume pulse (BVP) sensors.

8.2.2 Features

Several steps were conducted to extract the features. First, the raw signal data were segmented using a 60-second sliding window with a 0.25-second sliding step. In the next step, we produced new signal data from accelerometer data called magnitude using Equation (8.1):

$$ACC_{norm} = \sqrt{ACC_x^2 + ACC_y^2 + ACC_z^2} \tag{8.1}$$

where ACC_{norm} is the magnitude value of the accelerometer data and ACC_x, ACC_y, and ACC_z are the data of accelerometer from axis x, y, and z, respectively.

In the third step, for each signal data, first and second derivatives, as well as three transformed signal data using a Discrete Wavelet Transform (DWT) with the Haar wavelet at three distinct frequencies (1 Hz, 2 Hz, and 4 Hz), were generated for every original sensor's signal data. Hence, we

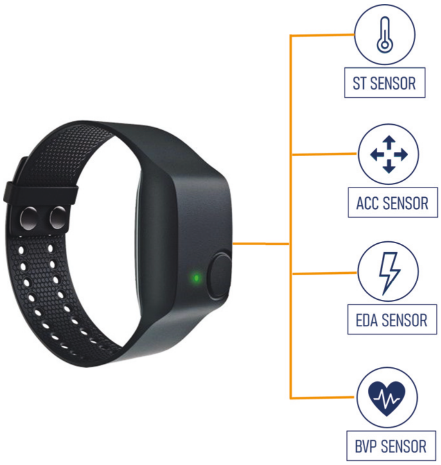


Figure 8.1: Empatica E4 that contains several sensors including electrodermal activity (EDA), accelerometers (ACC), skin temperature (ST), and blood volume pulse (BVP) sensors.

Table 8.1: Statistical Features Extracted From Each Signal

No.	Statistical Features
1	Mean of the Signal
2	Minimum value of the signal
4	Maximum value of the signal
4	Median of the signal
5	Maximum signal amplitude
6	Signal variance
7	Standard signal deviation
8	Absolute signal deviation
9	Signal kurtosis
10	Signal skewness

got 42 distinct signal data in total. In the final step, using the BioSPPy and Numpy libraries [7], 10 statistical features were calculated from each signal as depicted in Table 8.1. In addition, we normalize the features using Min-Max method because of its ability to improve classification performance.

8.2.3 Classification Strategies

In this study, we will compare the performance of using individual sensors and using multiple sensors. The machine learning (ML) model is fed data from just one sensor in the individual sensor strategy while in contrast, the

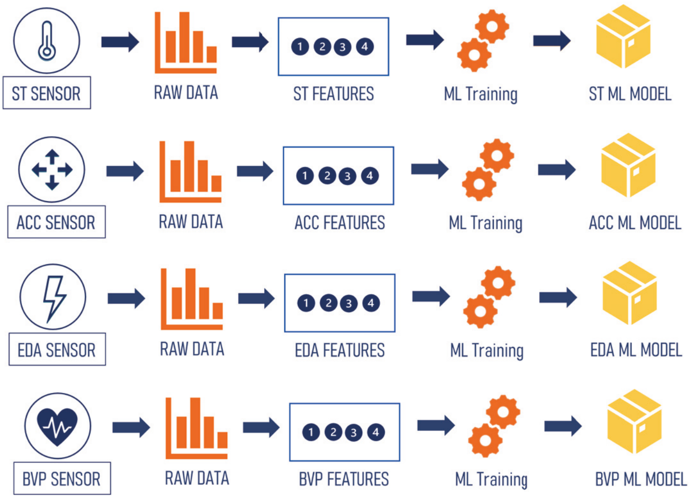


Figure 8.2: Training Process Using Individual Sensor.

multiple sensors strategy uses data from numerous sensors to gather richer and more varied information to improve the model robustness. The fusion of multiple sensors can be conducted before or after the data were fed to the ML model. In this study, for the former, the fusion was implemented on the feature level while the latter was implemented using the weighted score-level fusion strategy. Logistic Regression is chosen in this study as the ML model and was implemented using the Scikit-learn library [12].

8.2.4 Individual Sensor Strategy

Generally, classification tasks contain two processes: training and testing. As shown in Figure 8.2, using this strategy, the features extracted from the sensor data are used to train the ML method to create a stress classification model. This training process produces four separate models because we have four sensors. In the testing phase, as pictured in Figure 8.3, these models are then used to classify test data from each sensor. Each model may yield a different result.

8.2.5 Multiple Sensor Fusion Strategy

Generally, multiple sensor fusion strategy can be divided into two types: data or feature level fusion that is conducted before the data is fed to the ML model and decision-level fusion after the model determine the outcome based on the data fed to the model. In this study, we implemented a feature-level fusion as the first type. Meanwhile, for the second type, we propose a

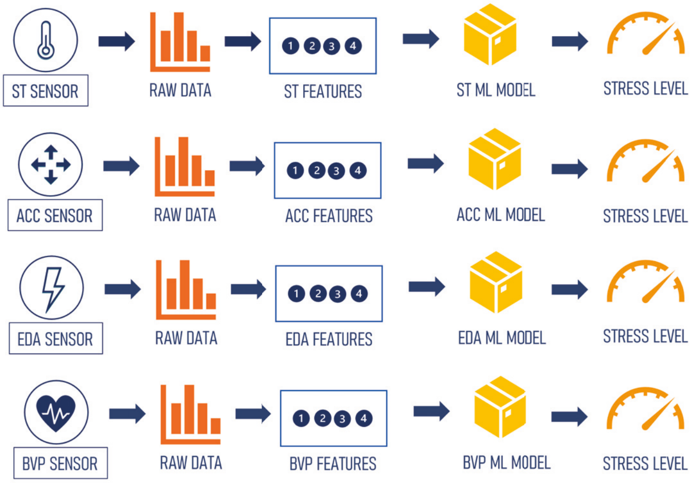


Figure 8.3: Individual Sensor ML Model Testing.

weighted score-level fusion. The details of feature-level and weighted score-level strategies are explained as follows:

- `Feature-Level Fusion`: Using the feature-level fusion, all of the extracted features from each sensor are combined into one vector. This feature vector is then used in the training phase of the ML algorithm to create an ML model as pictured in Figure 8.4. In the testing phase, the extracted features from each sensor are also combined before the trained ML model predicts which class this data belongs to as shown in Figure 8.5.

- `Weighted Score-Level Fusion`: This multiple sensor fusion strategy utilizes the probability score of each class computed using the ML model. This strategy's training procedure is identical to that of the individual sensor strategy's training procedure as displayed in Figure 8.2 where we have four separate models in the end. These models are then employed in the testing stage to predict stress levels as displayed in Figure 8.6. First, assuming we have n sensors, we compute the stress score of the test data t from subject p by using the following formula:

$$StressScore_{pt} = \frac{\sum_{i=0}^{n} w_{ip} s_{it}}{n} \tag{8.2}$$

  where w_{ip} is the weight of the ML model of sensor i on subject p, s_{it} is the probability score of the testing data t belonging to the stress class computed by the ML model of sensor i, and n is the number of sensor models used. The weight of each sensor model for each subject is

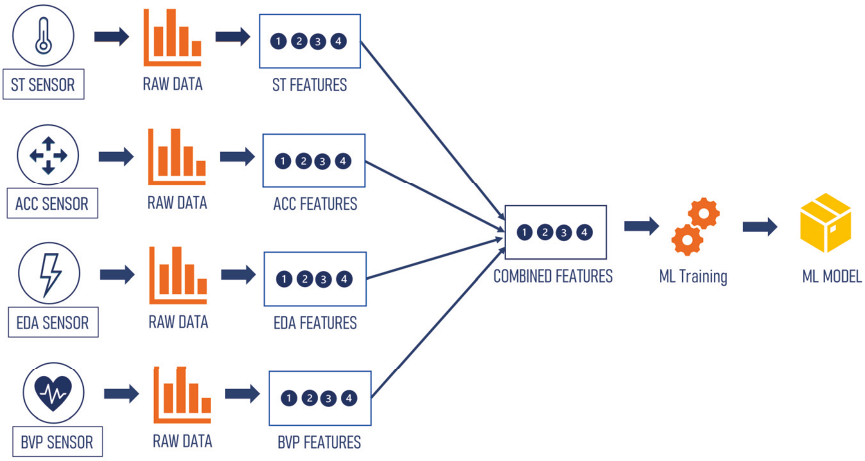


Figure 8.4: Training Process Using Multiple Sensor Fusion on Feature Level.

computed based on the performance (ie. F_1-measure) of each sensor model on some subject data that become the testing data. The formula to compute the weight of the ML model of sensor i on subject p is displayed in the following:

$$w_{ip} = \frac{F1_{ip}}{\sum_{j=0}^{n} F1_{jp}} \tag{8.3}$$

where $F1_{ip}$ is the F_1-measure value of the ML model of sensor i tested on some subject p's data and n is the number of sensor models used. These weights are computed for each sensor's model for each subject so that the weight of the same sensor is highly likely to be different for each subject. In other words, the weight is personalized based on the performance of the sensor model on the subject's data. Finally, the class of the test data t from subject p is determined by the following formula:

$$Class(pt) = \begin{cases} Non-stress & StressScore_{pt} \leq 0.5 \\ Stress & StressScore_{pt} > 0.5 \end{cases} \tag{8.4}$$

8.2.6 Evaluation

The leave-one-subject-out cross-validation (LOSOCV) procedure was used to measure the performance of the stress classification system. LOSOCV is a particular form of k-fold CV, where each fold represents each subject. Hence, the value of k, the number of folds, is equal to the number of subjects in the

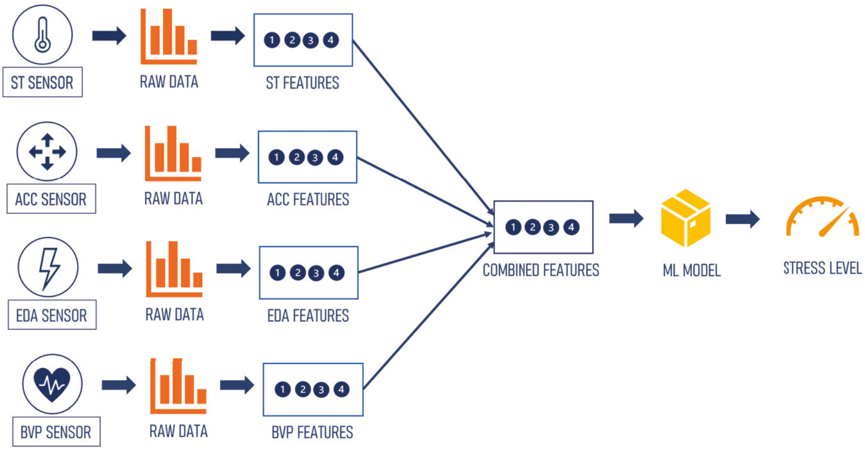


Figure 8.5: Feature Level-Multiple Sensor Fusion ML Model Testing.

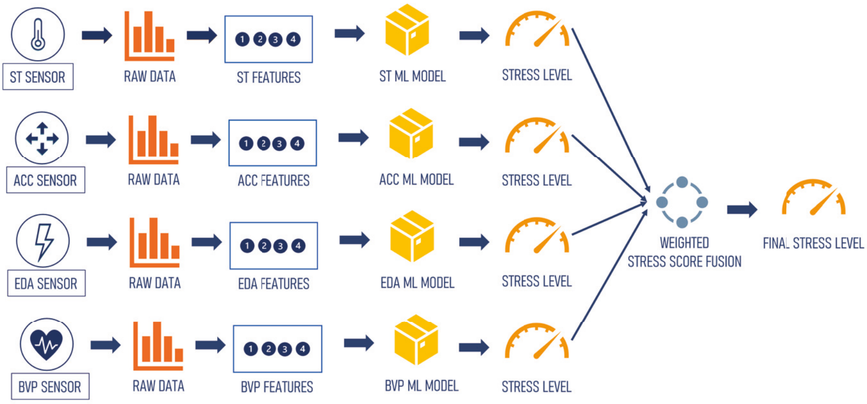


Figure 8.6: Score Level-Multiple Sensor Fusion ML Model Testing.

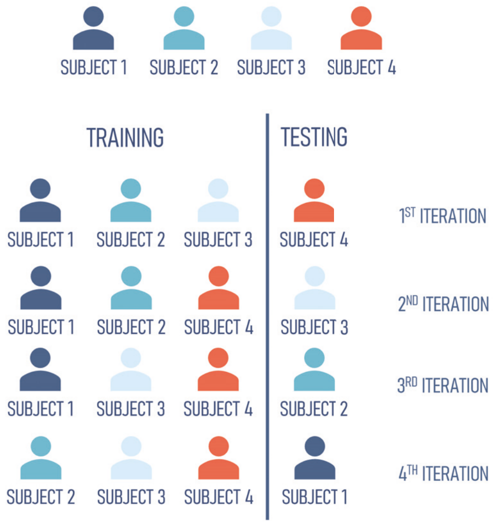


Figure 8.7: Leave-One-Subject-Out Cross-Validation Procedure.

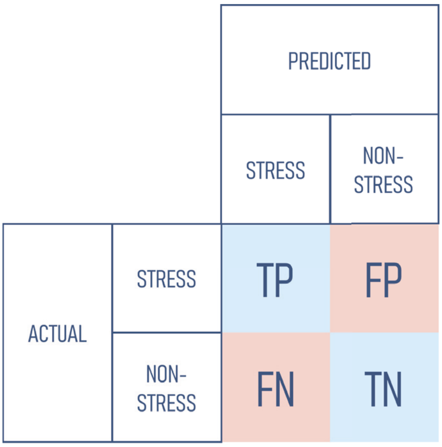
dataset. As demonstrated in Figure 8.7, for $k = 4$, we iterate four times where in each iteration we train the model on $k - 1$ subjects' data as the training set and then test the model on the one "left out" subject's data. Hence, we will have four different evaluation metrics, one from each fold. In the last step, we have to compute the mean of these evaluation metrics in order to assess the performance of the overall model. In this study, since the number of subjects is 15, the value of k is 15 and 15 iterations have to be completed to assess the stress classification system. In this study, since we used a weighted scoring strategy, we divided the testing data into two parts. The first part is 10% of data that were used to determine the weight of each sensor's model for each subject. Meanwhile, the rest 90% of the testing data were used for evaluating the model performance.

Four evaluation metrics are used to measure the system performance based on the confusion matrix displayed in Figure 8.8:

- `Accuracy (Acc)`: Accuracy measures how many times the model was correct overall, both in predicting stress or non-stress class. The formula is displayed in Equation (8.5).

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{8.5}$$

- `Precision (P)`: Precision measures how many of the stress class predictions made are correct. The formula is displayed in Equation (8.6).



		PREDICTED	
		STRESS	NON-STRESS
ACTUAL	STRESS	TP	FP
	NON-STRESS	FN	TN

Figure 8.8: Confusion Matrix.

$$P = \frac{TP}{TP + FP} \tag{8.6}$$

- `Recall (R)`: Recall measures how many of the stress data are correctly retrieved by the model, over all the stress data in the data. The formula is displayed in Equation (8.7).

$$R = \frac{TP}{TP + FN} \tag{8.7}$$

- F$_1$-measure (F$_1$): F$_1$-measure is the harmonic mean of precision and recall. The formula is displayed in Equation (8.8).

$$F_1 = 2\frac{P \cdot R}{P + R} \tag{8.8}$$

8.3 Result and Discussion

The stress classification performance using individual sensor strategy can be seen in Table 8.2 while the performance using multiple sensor fusion is shown in Table 8.3. As shown in Table 8.2, ACC sensor model got the best performance with accuracy, precision, recall, and F$_1$-measure of 0.867, 0.753, 0.813, and 0.758, respectively. BVP sensor model came second with 0.853 accuracy and 0.726 F$_1$-measure while the TEMP sensor model followed behind with 0.823 accuracy and 0.683 F$_1$-measure. EDA sensor model achieved the

Table 8.2: Stress Classification Performance Using Individual Sensor

Sensor	Acc	P	R	F1
EDA	0.828	0.556	0.812	0.610
ACC	**0.867**	**0.753**	**0.813**	**0.758**
TEMP	0.823	0.660	0.743	0.683
BVP	0.853	0.702	0.793	0.726

Table 8.3: Stress Classification Performance Using Multiple Sensor Fusion Strategies

Sensor	Acc	P	R	F1
Feature-Level Fusion	0.891	0.813	0.855	0.812
Weighted Score-Level Fusion	**0.931**	**0.824**	**0.939**	**0.868**

worst performance among all individual sensors with accuracy of 0.828 and F_1-measure of 0.610.

Based on the experiment results displayed in Table 8.3, the best performance was achieved by the weighted score-level fusion with accuracy, precision, recall, and F_1-measure of 0.931, 0.824, 0.939, and 0.868, respectively. Meanwhile, the feature-level fusion strategy got accuracy, precision, recall, and F_1-measure of 0.891, 0.813, 0.855, and 0.812, respectively.

The experiment results show that both multiple sensor fusion models obtained better performance than models from the individual sensor strategy. This result is expected because more sensors means more information that can make the model more robust as the result of learning from more perspectives. However, in terms of computational cost, the individual sensor strategy is better because less information means less processing time in terms of feature extraction and classification which can make the process faster.

Another anticipated result is that the weighted score-level fusion strategy can outperform the feature-level strategy. This result is reasonable because stress is personal and each subject can have a different reaction to the stress. Hence, some sensors may be a good indicator of stress for one subject but maybe it is not the case for other others. In the feature-level fusion strategy, for each subject, we cannot give different weights for each sensor. In contrast, the weighted score-level fusion strategy enables us to be adaptive by giving more weight to the sensor model that can predict better on the subject data. Therefore, a weighted score-level fusion strategy can obtain a better result.

The comparison of our work with prior work is shown in Table 8.4. The prior work selected are studies that used smartwatch data from the WESAD

Table 8.4: Comparison With Other Studies

Method	Accuracy
Schmidt et al. [14] (2018)	0.883
Siirtola (2019) [16]	0.874
Alshamrani (2021) [1]	0.850
Fauzi and Yang (2021) [3]	0.871
Zhu et al. (2022) [21]	0.864
Our best method	**0.931**

dataset for stress detection tasks with two classes (stress and non-stress). We solely compare accuracy because that is the only evaluation metric used by nearly all of the prior publications. Table 8.4 demonstrates that our best method outperformed the other methods from prior studies in terms of accuracy.

8.4 Conclusion

In this study, we built a stress classification system based on smartwatch sensor data and logistic regression as the classifiers. We analyzed two multiple sensor strategies (feature-level and weighted score-level fusion strategy) and compare their results with the results from the individual sensor strategy.

The experiment results show that for individual sensor strategy, the ACC sensor model got the best performance among all other sensor models with accuracy, precision, recall, and F_1-measure of 0.867, 0.753, 0.813, and 0.758, respectively. Furthermore, the result also reports that both multiple sensor fusion models obtained better performance than models from the individual sensor strategy. The weighted score-level fusion strategy achieved a better performance than the feature-level strategy with accuracy, precision, recall, and F_1-measure of 0.931, 0.824, 0.939, and 0.868, respectively.

In this study, only statistical features are used. For some sensor data, other types of features can be better to improve the stress classification result. Therefore, for future work, it is important to test other types of features to make a more robust stress detection system. Besides, other ML algorithms and fusion strategies are also interesting to experiment with.

8.5 Bibliography

[1] Alshamrani, M. An advanced stress detection approach based on processing data from wearable wrist devices. *Int. J. Adv. Comput. Sci. Appl 12* (2021), 399–405. 211, 228

[2] Can, Y. S., Chalabianloo, N., Ekiz, D., and Ersoy, C. Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study. *Sensors 19*, 8 (2019), 1849. 202, 219, 232

[3] Fauzi, M. A., and Yang, B. Continuous stress detection of hospital staff using smartwatch sensors and classifier ensemble. In *pHealth 2021*. IOS Press, 2021, pp. 245–250. 28, 202, 211, 219, 228, 232, 238, 262

[4] Fauzi, M. A., Yeng, P., Yang, B., and Rachmayani, D. Examining the link between stress level and cybersecurity practices of hospital staff in indonesia. In *The 16th International Conference on Availability, Reliability and Security* (2021), pp. 1–8. 14, 16, 21, 27, 100, 108, 117, 120, 132, 153, 164, 202, 218, 232, 246, 258

[5] Gallup. State of the global workplace: 2022 report, June 2022. Available from: `https://www.gallup.com/workplace/349484/state-of-the-global-workplace.aspx`. 218

[6] Garg, P., Santhosh, J., Dengel, A., and Ishimaru, S. Stress detection by machine learning and wearable sensors. In *26th International Conference on Intelligent User Interfaces* (2021), pp. 43–45. 5, 12, 202, 218, 247, 258

[7] Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., et al. Array programming with numpy. *Nature 585*, 7825 (2020), 357–362. 204, 220, 235, 260

[8] Indikawati, F. I., and Winiarti, S. Stress detection from multimodal wearable sensor data. In *IOP Conference Series: Materials Science and Engineering* (2020), vol. 771, IOP Publishing, p. 012028. 5, 12, 13, 202, 218, 247, 258

[9] Jacobson, B. H., Aldana, S. G., Goetzel, R. Z., Vardell, K., Adams, T. B., and Pietras, R. J. The relationship between perceived stress and self-reported illness-related absenteeism. *American Journal of Health Promotion 11*, 1 (1996), 54–61. 218

[10] Kirschbaum, C., Pirke, K.-M., and Hellhammer, D. H. The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology 28*, 1-2 (1993), 76–81. 20, 121, 165, 203, 219, 233, 260

[11] Kurniawan, H., Maslov, A. V., and Pechenizkiy, M. Stress detection from speech and galvanic skin response signals. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems* (2013), IEEE, pp. 209–214. 202, 219, 232, 262

[12] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M., AND DUCHESNAY, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12* (2011), 2825–2830. 204, 221, 235, 262

[13] PICKERING, T. G. Mental stress as a causal factor in the development of hypertension and cardiovascular disease. *Current hypertension reports 3*, 3 (2001), 249–254. 202, 218, 232, 246, 258

[14] SCHMIDT, P., REISS, A., DUERICHEN, R., MARBERGER, C., AND VAN LAERHOVEN, K. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 20th ACM international conference on multimodal interaction* (2018), pp. 400–408. 5, 12, 13, 26, 202, 203, 211, 218, 219, 228, 233, 247, 258, 259

[15] SCHNEIDERMAN, N., IRONSON, G., AND SIEGEL, S. D. Stress and health: psychological, behavioral, and biological determinants. *Annual review of clinical psychology 1* (2005), 607. 218

[16] SIIRTOLA, P. Continuous stress detection using the sensors of commercial smartwatch. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (2019), pp. 1198–1201. 5, 12, 13, 202, 211, 218, 228, 247, 258

[17] SOMMERFELDT, S. L., SCHAEFER, S. M., BRAUER, M., RYFF, C. D., AND DAVIDSON, R. J. Individual differences in the association between subjective stress and heart rate are related to psychological and physical well-being. *Psychological science 30*, 7 (2019), 1016–1029. 218

[18] TENNANT, C. Work-related stress and depressive disorders. *Journal of psychosomatic research 51*, 5 (2001), 697–704. 202, 218, 232

[19] TSIGA, E., PANAGOPOULOU, E., AND MONTGOMERY, A. Examining the link between burnout and medical error: A checklist approach. *Burnout Research 6* (2017), 1–8. 80, 202, 218, 232, 246, 258

[20] WELP, A., MEIER, L. L., AND MANSER, T. Emotional exhaustion and workload predict clinician-rated and objective patient safety. *Frontiers in psychology 5* (2015), 1573. 80, 202, 218, 232, 246, 258

[21] ZHU, L., SPACHOS, P., AND GREGORI, S. Multimodal physiological signals and machine learning for stress detection by wearable devices. In *2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA)* (2022), IEEE, pp. 1–6. 211, 228

Chapter 9

Examining the Effect of Feature Normalization and Feature Selection for Logistic Regression Based Multimodal Stress Detection

Muhammad Ali Fauzi; Bian Yang; Prosper Yeng

Abstract

Automated multimodal stress detection using smartwatches and machine learning (ML) has been very popular nowadays. One of the processes in ML-based classification is preprocessing, which includes feature normalization and feature selection because it can enhance classification performance. In this study, we construct a multimodal-based stress detection system using Logistic Regression and investigate the effects of feature normalization and feature selection on performance. The experiment results show that the stress classification system with feature normalization performs better than without feature normalization. The results also show that the use of the fewest features gives the worst performance. The performance of the stress classification system increases as the number of features increases but the performance slightly declines at a particular point. The best performance was obtained when Min-Max normalization and ANOVA-based feature selection were employed with accuracy, precision, recall, and F_1-measure of 0.894, 0.819, 0.859, and 0.817, respectively. This best result was achieved when 90% of the total features (378 features) were used.

9.1 Introduction

Workplace stress is a problem that affects many people in our modern world and is becoming more well recognized. This problem got worse in 2019 with the introduction of COVID-19. According to Gallup's State of the Global Workplace 2022 Report, 44% of the workforce reported experiencing high levels of daily stress in 2021. This figure surpasses the previous record which was set in 2020 (43%).

In the long run, the health and mental stability of an individual are at risk due to daily stress [18]. The cardiovascular system, immunological system, neuroendocrine system, and metabolic system all suffer damage as a result of the allostatic load brought on by prolonged exposure to stress [2]. Repeated exposure to stress can cause the neurons in the hippocampal area to die, which results in poorer memory performance [16]. In extreme circumstances, stress also contributes to numerous mental illnesses such as anxiety and depression [25, 14]. Conditions including hypertension and coronary artery disease, diabetes, and asthma, among others, can be brought on by the long-term effects of stress [20]. In addition, employees' work-related stress can also have negative impacts on the company's productivity, finances, and safety [26, 24, 5, 23]. Therefore, stress detection is important as one crucial part of stress management. People may be able to feel more in control of how they respond to situations if they are aware of their own level of stress [13]. It can make them stay alert and know when to relax or take action to handle stress.

Smartwatch has been a popular device nowadays for a lot of purposes including affective computing. The advantage of this kind of device is its usability and high level of social acceptance [22, 12]. Besides, smartwatches have several integrated sensors, such as a temperature sensor, an accelerometer, an electrodermal activity sensor (EDA), and a Blood Volume Pulse (BVP) sensor, that may be utilized for multimodal stress detection.

Automated stress detection has been successfully built in several earlier research using data from multimodal sensors and machine learning (ML) techniques. One of the popular machine learning methods for this task is Logistic Regression as it has proven to give good performances in this task [4, 11, 1]. One of the processes in ML-based classification is preprocessing including feature normalization and feature selection as they can improve the classification performance. In this work, we implement and examine the impact of feature normalization and feature selection on the performance of a multimodal-based stress detection method using Logistic Regression. We will compare three scenarios for feature normalization including without normalization, with Min-Max normalization, and with Z-Score normalization. Meanwhile, for the feature selection experiment, we will analyze the impact of the number of features selected and the best number of features used for the stress classification task.

9.2 Materials and Methods

9.2.1 Dataset

The dataset used in this study is a publicly available stress detection dataset called WESAD (Wearable Stress and Affect Detection) [21]. WESAD dataset contains multimodal data from 15 participants, consisting of 13 male and 2 female subjects. Three different situations were created. The first 20 minutes were spent collecting data in a neutral situation in which the participants were instructed to read a magazine while sitting/standing at a table. In the next session called the amusement scenario, the participants viewed 11 humorous videos for a total of 392 seconds. Finally, for the stress scenario, the participants completed a Trier social stress test (TSST) for a total of 10 minutes. TSST is a well-known laboratory-based procedure used to reliably induce stress in human research participants [10]. In this test, the participants were asked to perform public speaking and a mental arithmetic task to impose a high mental load on them. In this study, the neutral and amusement scenarios were merged into one non-stress class for the stress detection task resulting in a binary classification problem (stress and non-stress).

The data on the WESAD dataset was recorded using two different types of sensors, chest-worn (RespiBAN) and wrist-worn devices (Empatica E4 smartwatch). In this study, only the data from the smartwatch were used since the use of watches is well known and has a high degree of social acceptance by their ubiquity in everyday life compared to the chest-worn devices that have very low usability and are not convenient to wear in the working environment [12, 22]. Empatica E4 provides several sensors including electrodermal activity (EDA), accelerometers (ACC), skin temperature (ST), and blood volume pulse (BVP) sensors as depicted in Figure 9.1. The details of the sensors are explained in Table 9.1.

9.2.2 Stress Detection Pipeline

The stress detection pipeline in this study is presented in Figure 9.2. The pipeline is divided into two main parts: training and testing. The final result of the training phase is to create a trained ML model while the final product of the testing phase is the classification result.

In the training phase, 420 features were extracted from the raw training data. Then the features were normalized using MinMax or Z-score normalization method. In the third step, n best features were selected while others were removed based on their importance. ANOVA was the method used in this study to determine the importance of each feature. Furthermore, the selected features were used to train an ML algorithm to produce a trained ML model. This ML model was saved and then used in the testing phase. The ML algorithm used in this study was Logistic Regression.

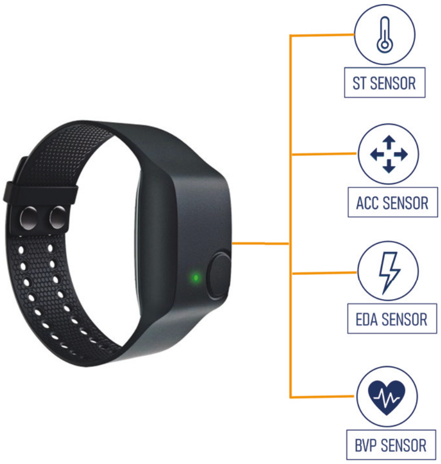


Figure 9.1: Empatica E4.

Table 9.1: Empatica E4 data sensors in the WESAD dataset

Label	Sensor	Detail	Sampling Rate (Hz)
ST	Skin Temperature	Reads peripheral skin temperature	4
ACC	Accelerometer	Captures 3-axis motion-based activity	32
EDA	Electrodermal Activity	Measures the constantly fluctuating changes in certain electrical properties of the skin	4
BVP	Blood Volume Pulse	Measures Blood Volume Pulse from which heart rate variability can be derived	64

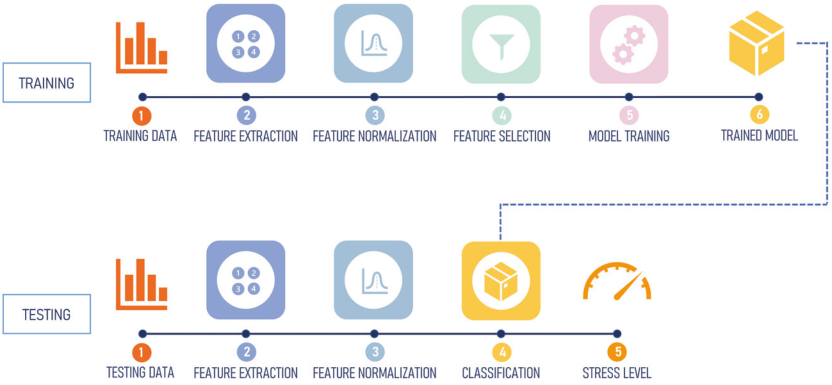


Figure 9.2: Stress Detection Pipeline.

In the testing phase, some features were also extracted from the raw testing data. The difference between the feature extraction process in the training and testing phase is the number of features extracted. In the training phase, from the raw training data, we extracted all of the features that we initially defined (420 features). Meanwhile, in the testing phase, we only extracted n features from the testing data. The n features used in the testing data were based on the n best features selected from the feature selection process in the training phase. Therefore, we did not need a feature selection process in the testing phase. After the features were extracted, they were normalized. In the next step, the normalized features were classified by using the trained ML model. Finally, the stress level of the testing data was determined. To be noted, all of the processes in the stress detection pipeline were implemented in Python using the Scikit-learn library [19].

9.2.3 Feature Extraction

The feature extraction process contains several steps. In the first step, the raw signal data were segmented using a 60-second sliding window with a 0.25-second sliding step. For the accelerometer data, in addition to the 3-axis signal data, we also calculated the magnitude by using Equation (9.1). In the next step, we generated 6 different signals for each signal data: the original signal; its first and second derivatives; and three modified signal data using a Discrete Wavelet Transform (DWT) with the Haar wavelet at three different frequencies (1 Hz, 2 Hz, and 4 Hz). As the result, we had 42 distinct signal data in total. Finally, 10 statistical features described in Table 9.2 were calculated from each signal using the BioSPPy and Numpy libraries [6]. In total, 420 features were used in this study.

Table 9.2: Statistical Features Extracted From Each Signal

No.	Statistical Features
1	Mean of the Signal
2	Minimum value of the signal
4	Maximum value of the signal
4	Median of the signal
5	Maximum signal amplitude
6	Signal variance
7	Standard signal deviation
8	Absolute signal deviation
9	Signal kurtosis
10	Signal skewness

$$ACC_{norm} = \sqrt{ACC_x^2 + ACC_y^2 + ACC_z^2} \tag{9.1}$$

9.2.4 Feature Normalization

The feature value range in this study varies greatly. Some machine learning algorithms' objective functions won't operate effectively without normalization because features with a high value range may dominate the classification result. To ensure that each feature contributes proportionally to the classification result, the range of all feature values should be normalized. The Min-Max and Z-score normalization are the two most often used feature normalization techniques for a classification task. In this study, we will compare the stress classification performance in three scenarios: without normalization, with Min-Max normalization, and with Z-score normalization. The details of the two normalization techniques are presented in the following subsections.

9.2.4.1 Min-Max Normalization

Min-max normalization is the simplest normalizing method. This method is also well-known as feature scaling. The situation when the boundaries (maximum and minimum values) of the feature values are known is best suited for min-max normalization. Using this method, the minimum and maximum scores are changed to 0 and 1, respectively, and then all the feature values are transformed into a common range of 0 and 1 [8]. Even if the feature value boundaries are unknown, we can still use this method by estimating the minimum and maximum values of the features or using the minimum and maximum values of the features in the training set. Given a set of feature values $\{f_k\}, k = 1, 2, ..., n$, the normalized feature values are calculated by using the following formula:

$$f'_k = \frac{f_k - min(f_k)}{max(f_k) - min(f_k)} \tag{9.2}$$

where f'_k is the normalized feature values, $min(f_k)$ is the minimum feature values and $max(f_k)$ is the maximum feature values.

9.2.4.2 Z-Score Normalization

Z-score normalization is a technique for normalizing data based on the mean and standard deviation of the data [7]. When the features' lowest and maximum values are unknown, this approach is highly helpful. However, this approach is not robust when the features contain outliers because the mean and standard deviation are both susceptible to outliers. Given a set of feature values $\{f_k\}, k = 1, 2, ..., n$, the normalized feature values are calculated by using the following formula:

$$f'_k = \frac{f_k - \mu}{\sigma} \tag{9.3}$$

where f'_k is the normalized feature values, μ is the arithmetic mean of the feature values in the training data and σ is the standard deviation of the feature values in the training data.

9.2.5 Feature Selection

Feature selection is a process of selecting the best features to improve the effectiveness and efficiency of classification. By using a feature selection, the computational complexity can be reduced because fewer data will be processed by the next process. Besides, feature selection also often increases the accuracy of the classification task because it can remove bad features that can hinder the classification performance [3].

The Analysis of Variance (ANOVA) f-test statistic is one of the most often used feature selection techniques for a task in which the input is numerical data and the output is categorical such as stress detection using signal data from sensors. ANOVA is a well-known statistical method for comparing many independent means [9, 17]. It is a quick and effective technique for determining if two groups' means differ from one another.

The null hypothesis for feature selection in this technique is that each feature (variable) has no mean difference between distinct classes (e.g., stress and non-stress), and if this null hypothesis could not be rejected by an F-test, the importance of the feature is low [15]. In this study, a one-way ANOVA F-test statistic was employed by computing the ratio of variances within and across groups. The ANOVA F-value was used to rank each feature's importance. The following formula was used to get each feature's F-value:

$$F = \frac{variation between classes}{variation within classes} \tag{9.4}$$

9.2.6 Machine Learning Model

To train an ML model, for the first step, we need to select the ML algorithm that we want to use. The ML method used in this study is Logistic regression. This method is selected due to its good performance in stress detection tasks [4]. This ML method was trained using the selected features from the previous process. In the model training process, we fed the features to the Logistic Regression model. The final product of this training was a trained ML model that we can use to classify the incoming testing data. Furthermore, in the classification process, we classified the features from the testing data to determine the stress level.

9.2.7 Evaluation

The effectiveness of the stress classification system was evaluated using the leave-one-subject-out cross-validation (LOSOCV) method. The example of this evaluation procedure is displayed in Figure 9.3. For example, if the number of participants $k = 4$, we have to do four iterations. In every iteration, we train the ML model using the data from $k - 1$ participants and test the model on the one left-out participant. As a result, each fold will have one of our four assessment measures. To evaluate the effectiveness of the entire model, we must compute the mean of these assessment metrics in the last phase. In this study, since the number of participants is 15, the value of k is 15 and 15 iterations have to be conducted to evaluate the stress classification system.

Four evaluation metrics are used to measure the system performance based on the confusion matrix displayed in Figure 9.4 including Accuracy (Acc), Precision (P), Recall (R), and F$_1$-measure (F$_1$). The following are the formulas to compute the four evaluation metrics:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{9.5}$$

$$P = \frac{TP}{TP + FP} \tag{9.6}$$

$$R = \frac{TP}{TP + FN} \tag{9.7}$$

$$F_1 = 2\frac{P \cdot R}{P + R} \tag{9.8}$$

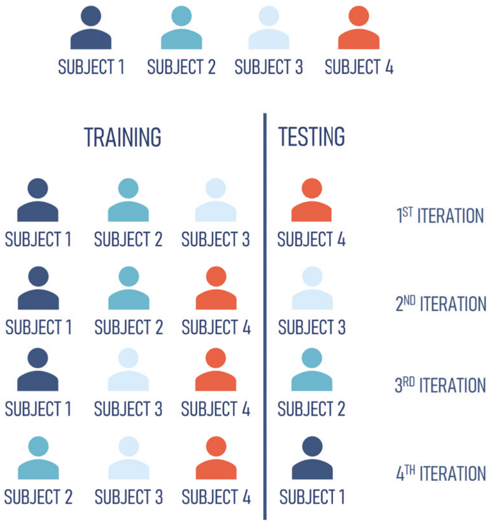
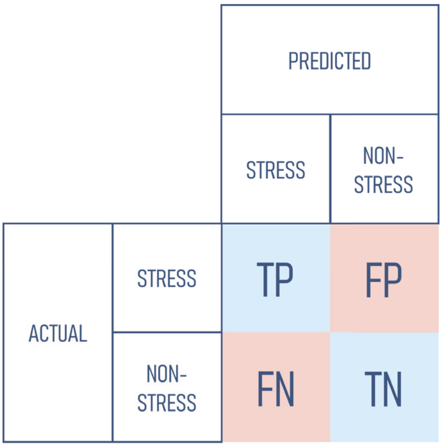


Figure 9.3: Leave-One-Subject-Out Cross-Validation Procedure.



		PREDICTED	
		STRESS	NON-STRESS
ACTUAL	STRESS	TP	FP
	NON-STRESS	FN	TN

Figure 9.4: Confusion Matrix.

Table 9.3: Feature Normalization Experiment Result

Method	Acc	P	R	F_1
Without Normalization	0.874	0.728	0.842	0.742
With Min-Max Normalization	**0.891**	**0.814**	**0.855**	**0.812**
With Z-Score Normalization	0.889	0.822	0.846	0.806

9.3 Result

In this study, we have two experiments. The first experiment is to analyze the use of feature normalization while the second one focuses on the use of feature selection. Therefore, for the first experiment, we did not use the feature selection method. The stress classification results from the first experiment using logistic regression and three normalization scenarios are displayed in Table 9.3. The results show that the stress classification system with feature normalization performs better than without feature normalization. The accuracy and F_1-measure of the system without feature normalization are 0.874 and 0.742, respectively. Meanwhile, the stress detection system with Min-Max normalization got the best performance in terms of all evaluation metrics with accuracy, precision, recall, and F_1-measure of 0.891, 0.814, 0.855, and 0.812, respectively. These results align with many other works that prove that feature normalization can improve the performance of classification tasks because this normalization process can reduce the variation of feature value range so that each feature contributes proportionally to the classification result.

Since Min-Max normalization obtained the best result in the first experiment, this normalization was used for the second experiment. The stress classification results from the second experiment using logistic regression, Min-Max normalization, and ANOVA feature selection are displayed in Table 9.4. In this experiment, we tested 10 different scenarios of the number of features used for the stress classification task from 10% of total features (42 features) to 100% of total features (all 420 features). The results show that the use of the fewest features gives the worst performance. The use of 10% of total features only obtained the accuracy of 0.869 and F_1-measure of 0.747. The performance of the stress classification system increases as the number of features increases. The best performance was obtained when we used 90% of the total features (378 features) with accuracy, precision, recall, and F_1-measure of 0.894, 0.819, 0.859, and 0.817, respectively. When the number of features was increased to 100% (without feature selection), the performance slightly declines.

These results show that the use of too few features can lead to a bad performance because we remove too many important features. As the number of features increases, the performance also increases because more impor-

Table 9.4: Feature Selection Experiment Result

Percentage of Features Used (%)	Acc	P	R	F_1
10	0.869	0.728	0.860	0.747
20	0.879	0.756	0.866	0.774
30	0.870	0.751	0.837	0.761
40	0.867	0.756	0.830	0.759
50	0.866	0.779	0.809	0.767
60	0.884	0.788	0.842	0.789
70	0.880	0.802	0.827	0.791
80	0.882	0.814	0.831	0.797
90	**0.894**	**0.819**	**0.859**	**0.817**
100	0.891	0.814	0.855	0.812

tant features are employed. However, at some points, increasing the number of features will reduce performance because there are some features involved that have a bad contribution to the stress classification result. Therefore, finding the suitable number of features used and which features should be used is important to improve the classification performance. Besides, by using feature selection, the computational complexity of the stress classification task can be reduced because fewer features need to be processed.

9.4 Conclusion

Machine learning-based automated multimodal stress detection has gained a lot of attention today. One device that is appropriate for this job is the smartwatch since it can be used conveniently in a working environment and has several sensors built in that can allow multimodal stress detection. Preprocessing, which includes feature normalization and feature selection, is one of the procedures used in ML-based classification since it can enhance classification performance. In this study, we construct a multimodal-based stress detection system using Logistic Regression and investigate the effects of feature normalization and feature selection on performance.

The experiment results show that the stress classification system with feature normalization performs better than without feature normalization because this normalization process can reduce the variation of feature value range so that each feature contributes proportionally to the classification result. The stress detection system with Min-Max normalization got the best performance in terms of all evaluation metrics with accuracy, precision, recall, and F_1-measure of 0.891, 0.814, 0.855, and 0.812, respectively. Meanwhile, the results of the feature selection experiment show that the use of the fewest features gives the worst performance. The performance of the stress classification system increases as the number of features increases but the performance slightly declines at a particular point. The best performance was obtained when we used 90% of the total features (378 features) with accuracy, precision, recall, and F_1-measure of 0.894, 0.819, 0.859, and 0.817, respectively.

In this study, we conducted a binary classfication because the WESAD dataset only employs two classes (stress and non-stress). In future work, a dataset with various stress levels may be used to examine the impact of feature normalization and feature selection (e.g. low stress, moderate stress, and high stress). Besides, other methods for feature normalization and feature selection can also be explored for future experiment.

9.5 Bibliography

[1] Can, Y. S., Chalabianloo, N., Ekiz, D., and Ersoy, C. Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study. *Sensors 19*, 8 (2019), 1849. 202, 219, 232

[2] Edes, A. N., and Crews, D. E. Allostatic load and biological anthropology. *American Journal of Physical Anthropology 162* (2017), 44–70. 232

[3] Fauzi, M. A., Arifin, A. Z., Gosaria, S. C., and Prabowo, I. S. Indonesian news classification using naïve bayes and two-phase feature selection model. *Indonesian Journal of Electrical Engineering and Computer Science 2*, 3 (2016), 401–408. 237

[4] Fauzi, M. A., and Yang, B. Continuous stress detection of hospital staff using smartwatch sensors and classifier ensemble. In *pHealth 2021*. IOS Press, 2021, pp. 245–250. 28, 202, 211, 219, 228, 232, 238, 262

[5] Fauzi, M. A., Yeng, P., Yang, B., and Rachmayani, D. Examining the link between stress level and cybersecurity practices of hospital staff in indonesia. In *The 16th International Conference on Availability, Reliability and Security* (2021), pp. 1–8. 14, 16, 21, 27, 100, 108, 117, 120, 132, 153, 164, 202, 218, 232, 246, 258

[6] Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., et al. Array programming with numpy. *Nature 585*, 7825 (2020), 357–362. 204, 220, 235, 260

[7] Henderi, H., Wahyuningsih, T., and Rahwanto, E. Comparison of min-max normalization and z-score normalization in the k-nearest

neighbor (knn) algorithm to test the accuracy of types of breast cancer. *International Journal of Informatics and Information Systems 4*, 1 (2021), 13–20. 237

[8] Jain, A., Nandakumar, K., and Ross, A. Score normalization in multimodal biometric systems. *Pattern recognition 38*, 12 (2005), 2270–2285. 236

[9] Johnson, K. J., and Synovec, R. E. Pattern recognition of jet fuels: comprehensive gc× gc with anova-based feature selection and principal component analysis. *Chemometrics and Intelligent Laboratory Systems 60*, 1-2 (2002), 225–237. 237

[10] Kirschbaum, C., Pirke, K.-M., and Hellhammer, D. H. The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology 28*, 1-2 (1993), 76–81. 20, 121, 165, 203, 219, 233, 260

[11] Kurniawan, H., Maslov, A. V., and Pechenizkiy, M. Stress detection from speech and galvanic skin response signals. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems* (2013), IEEE, pp. 209–214. 202, 219, 232, 262

[12] Lazaro, M. J. S., Lim, J., Kim, S. H., and Yun, M. H. Wearable technologies: acceptance model for smartwatch adoption among older adults. In *International Conference on Human-Computer Interaction* (2020), Springer, pp. 303–315. 4, 202, 232, 233, 246, 258

[13] Liao, W., Zhang, W., Zhu, Z., and Ji, Q. A real-time human stress monitoring system using dynamic bayesian network. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)-workshops* (2005), IEEE, pp. 70–70. 4, 202, 232, 246, 258

[14] Liu, X., Kakade, M., Fuller, C. J., Fan, B., Fang, Y., Kong, J., Guan, Z., and Wu, P. Depression after exposure to stressful events: lessons learned from the severe acute respiratory syndrome epidemic. *Comprehensive psychiatry 53*, 1 (2012), 15–23. 232

[15] Lu, S., Shen, S., Huang, J., Dong, M., Lu, J., and Li, W. Feature selection of laser-induced breakdown spectroscopy data for steel aging estimation. *Spectrochimica Acta Part B: Atomic Spectroscopy 150* (2018), 49–58. 237

[16] Mcewen, B. S., Albeck, D., Cameron, H., Chao, H. M., Gould, E., Hastings, N., Kuroda, Y., Luine, V., Magarinos, A. M., Mckittrick, C. R., et al. Stress and the brain: a paradoxical role for adrenal steroids. *Vitamins & hormones 51* (1995), 371–402. 232

[17] Nasiri, H., and Alavi, S. A. A novel framework based on deep learning and anova feature selection method for diagnosis of covid-19 cases from chest x-ray images. *Computational intelligence and neuroscience 2022* (2022). 237

[18] Nath, R. K., and Thapliyal, H. Smart wristband-based stress detection framework for older adults with cortisol as stress biomarker. *IEEE Transactions on Consumer Electronics 67*, 1 (2021), 30–39. 232

[19] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12* (2011), 2825–2830. 204, 221, 235, 262

[20] Pickering, T. G. Mental stress as a causal factor in the development of hypertension and cardiovascular disease. *Current hypertension reports 3*, 3 (2001), 249–254. 202, 218, 232, 246, 258

[21] Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., and Van Laerhoven, K. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 20th ACM international conference on multimodal interaction* (2018), pp. 400–408. 5, 12, 13, 26, 202, 203, 211, 218, 219, 228, 233, 247, 258, 259

[22] Spagnolli, A., Guardigli, E., Orso, V., Varotto, A., and Gamberini, L. Measuring user acceptance of wearable symbiotic devices: validation study across application scenarios. In *International Workshop on Symbiotic Interaction* (2015), Springer, pp. 87–98. 4, 202, 232, 233, 246, 258

[23] Tennant, C. Work-related stress and depressive disorders. *Journal of psychosomatic research 51*, 5 (2001), 697–704. 202, 218, 232

[24] Tsiga, E., Panagopoulou, E., and Montgomery, A. Examining the link between burnout and medical error: A checklist approach. *Burnout Research 6* (2017), 1–8. 80, 202, 218, 232, 246, 258

[25] Wang, Y., Chen, R., and Zhang, L. Reliability and validity of generalized anxiety scale-7 in inpatients in chinese general hospital. *J Clin Psychiatr 28* (2018), 168–71. 202, 232, 246, 258

[26] Welp, A., Meier, L. L., and Manser, T. Emotional exhaustion and workload predict clinician-rated and objective patient safety. *Frontiers in psychology 5* (2015), 1573. 80, 202, 218, 232, 246, 258

Chapter 10

Continuous Stress Detection of Hospital Staff Using Smartwatch Sensors and Classifier Ensemble

Muhammad Ali Fauzi; Bian Yang

Abstract

High stress levels among hospital workers could be harmful to both workers and the institution. Enabling the workers to monitor their own stress level has many advantages. Knowing their own stress level can help them to stay aware and feel more in control of their response to situations and know when it is time to relax or take some actions to treat it properly. This monitoring task can be enabled by using wearable devices to measure physiological responses related to stress. In this work, we propose a smartwatch sensors based continuous stress detection method using some individual classifiers and classifier ensembles. The experiment results show that all of the classifiers work quite well to detect stress with an accuracy of more than 70%. The results also show that the ensemble method obtained higher accuracy and F1-measure compared to all of the individual classifiers. The best accuracy was obtained by the ensemble with soft voting strategy (ES) with 87.10% while the hard voting strategy (EH) achieved the best F1-measure with 77.45%.

10.1 Introduction

Over recent years, stress has become an interesting topic in today's hectic world. There has been increasing awareness in many countries about the rise of work-related stress. Stress can be defined as a unique affective state that occurs when an individual considers that his or her perceived resources or ability cannot cope with the perceived demand of a stimulus [11]. The latest survey by Acas in 2019 [22] about stress and anxiety at work suggested that about 66% of working people have experienced work-related stress in the last 12 months. Hospital is possibly one of the most important workplaces to be alarmed about this issue. Many studies reported that many hospital workers suffer from work-related stress [15, 24, 1]. This is frequently due to high demands placed on healthcare personnel, as well as a lack of time, skills, and social support at work [15].

Although stress at some level is normal, chronic stress can harm our physical, mental, and emotional wellbeing. Many studies reported that stress has a significant contribution to the development of hypertension and coronary artery disease, diabetes, asthma, etc [16]. Moreover, excessive stress also has a negative impact on the employee's productivity, increases absenteeism, and plays a crucial role in mental illness development, such as generalized anxiety disorder and depression [23].

Specifically for hospital, many studies suggested that higher stress level has a relationship with low patient safety [25, 21]. Another study also reported that higher stress level is significantly correlated with riskier cybersecurity practices [4]. These studies are in line with a prior study [26] reporting that stressed people will be slow in learning something new and may choosing less profitable decisions.

Monitoring hospital workers' stress level has many advantages. Knowing their own stress level can help them stay aware and feel more in control of their response to situations and know when it is time to relax or take some actions to treat it properly [14]. Besides, this monitoring can help for early diagnosis of mental illness and disorders. The most common way to assess stress level is by using questionnaires (e.g. Perceived Stress Scale [2], Perceived Stress Questionnaire [12], etc.). However, this method takes time so that it is not convenient to be performed every day for continuous monitoring.

The other stress level assessment method is by measuring the physiological responses related to stress such as heart rate, blood pressure, skin conductance, respiration activity, etc. Some sensors can be used to conduct the measurement task. For example, electrocardiogram (ECG) can be used to measure the heart rate, galvanic skin response (GSR) for skin conductance, etc. The recent advance in wearable devices with sophisticated built-in sensors makes it feasible to passively collect multimodal data from people's daily lives for automatic continuous stress detection purposes. However, some wearable devices have a very low usability and not convenient to wear during work (e.g. chest-worn devices, finger placed GSR sensors, etc.) [20].

Smartwatch has recently emerged as a new platform that provides many successful applications. These devices have several built-in sensors that are useful for stress monitoring including Blood Volume Pulse (BVP), Electrodermal Activity (EDA), temperature, accelerometer, etc. Besides, the use of watches is well known and has a high degree of social acceptance by their ubiquity in everyday life [10]. Therefore, it has a high potential to be applied for multi-modal-based continuous stress detection.

Many previous works have been successfully leveraging multi-modal sensors data and machine learning methods to build automatic stress detection. The popular machine learning methods used are Random Forest, Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression [18, 6, 7, 19]. In this work, we propose a multi-modal based continuous stress detection method using classifier ensemble and give comparative analysis between individual classifiers. Classifiers ensemble is a set of base classifiers whose individual classification outputs are combined in some way in order to enhance classification accuracy [5]. The individual classifiers used for this works include Naive Bayes (NB), Support Vector Machine (SVM), Neural Network (NN), K-Nearest Neighbours (KNN), Logistic Regression (LR), Random Forest (RF), and Decision Tree (DT).

10.2 Proposed Work

10.2.1 Dataset

This research is based on the WESAD [18] dataset, which is available to the public. It includes data from 15 people who were measured with the Empatica E4 wrist-worn device and chest-worn RespiBAN device. However, because the focus of this work is on smartwatch sensors, only E4 data is used in this analysis. The E4 gadget incorporates skin temperature (ST), accelerometers (ACC), electrodermal activity (EDA), and blood volume pulse sensors (BVP) sensors. Data from three separate affective states (stress, amusement, and relaxation) were obtained during the data collection process. The stress situation lasted about 10 minutes, the amused situation 6.5 minutes, and the relaxed situation 20 minutes. For the stress detection task in this study, the amusement and relaxation classes were merged into one class: non-stress. As a result, the problem under investigation was binary (stress and non-stress).

10.2.2 Features

In this study, we used the data from all of the sensors available in the smartwatch including ACC, EDA, ST, and BVP. To extract the features, a sliding window with a window shift of 0.25 seconds was used to segment the data. Furthermore, the ACC features were computed with a five-second window size, as this is a common window length for acceleration-based context detection [17]. Meanwhile, all other physiological features were calculated with a window size of 60 seconds following the suggestion by Kreibig et al. [8]. The AC, EDA, and ST features were extracted based on prior work by [27]. The features extracted including some statistical features (mean, standard deviation, maximum, and minimum). Besides, some derivatives and

Discrete Wavelet Transform (DWT) were also applied to the data to extract other statistical features. Meanwhile, for BVP, statistical features (mean, standard deviation) were also computed. Moreover, some features based on energy in different frequency bands were also calculated.

10.2.3 Classifier

Seven machine learning methods were used as classifiers for stress detection tasks including Naive Bayes (NB), Support Vector Machine (SVM), Neural Network (NN), K-Nearest Neighbours (KNN), Logistic Regression (LR), Random Forest (RF), and Decision Tree (DT). In addition, we also used two ensemble methods. In order to do stress detection, the ensemble technique trains numerous classification methods and then combines them using particular approach [3]. It is important to take note that the performance of the ensemble methods cannot be guaranteed to be higher than the best individual method in the ensemble. However, it would significantly minimize the chances of picking a poor-performing classifier [5].

In this study, we employed three classification methods to build the ensemble learning method. Three individual classifiers with the highest accuracy were selected for the ensemble. Two ensemble strategies were used in this work as follows:

1. Hard voting (hard): As depicted in Figure 10.1, each classifier had one vote, and the class of the data was determined by the majority vote.

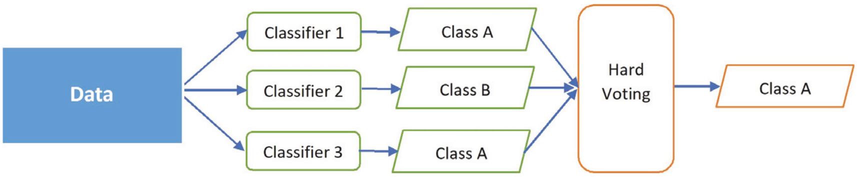


Figure 10.1: The hard voting strategy

2. Soft voting (soft): As depicted in Figure 10.2, each classifier calculated the probability of each class in the first step. Then, the probabilities of each class from all classifiers were averaged, and the final class of the data was the one with the greatest average probability value.

10.2.4 Performance Evaluation

All of the classifiers were tested using the leave-one-subject-out (LOSO) cross-validation (CV) approach, which shows how a model will general-

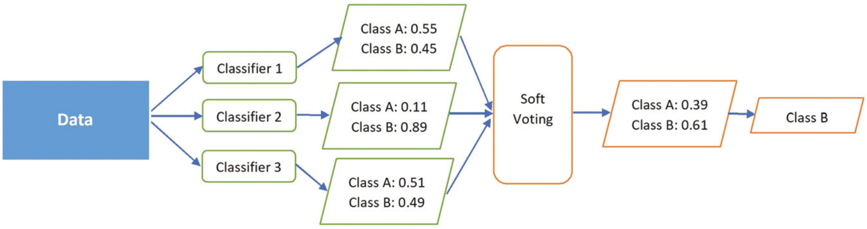
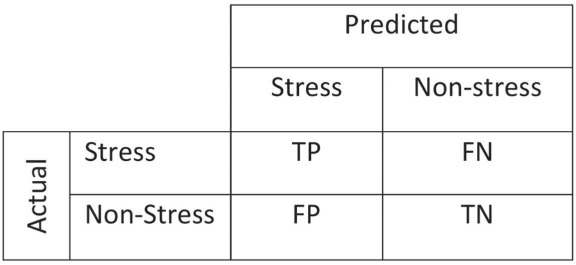


Figure 10.2: The soft voting strategy



		Predicted	
		Stress	Non-stress
Actual	Stress	TP	FN
	Non-Stress	FP	TN

Figure 10.3: Confusion Matrix

ize and perform on previously unseen data. Several measurements including Accuracy (Acc), Precision (P), Recall (R), and F$_1$-measure (F$_1$) were employed for classifier performance evaluation. The formulas for all of the measurements are based on the confusion matrix depicted in Figure 10.3 and displayed in the Eq. (10.1), Eq. (10.2), Eq. (10.3), and Eq. (10.4) respectively.

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{10.1}$$

$$P = \frac{TP}{TP + FP} \tag{10.2}$$

$$R = \frac{TP}{TP + FN} \tag{10.3}$$

$$F_1 = 2\frac{P \cdot R}{P + R} \tag{10.4}$$

Table 10.1: Stress Detection Result Using Individual Classifiers (%)

Method	Accuracy	Precision	Recall	F1-measure
NB	79.26	57.58	73.31	60.67
SVM	84.60	76.29	80.51	75.01
NN	**84.76**	**76.53**	80.12	74.97
KNN	73.71	52.31	63.74	52.43
LR	**85.46**	77.53	82.16	**76.25**
RF	**86.61**	69.17	**89.87**	73.05
DT	79.25	66.90	73.59	66.81

Table 10.2: Stress Detection Result Using Classifiers Ensemble (%)

Method	Accuracy	Precision	Recall	F1-measure
EH	86.99	76.00	88.02	**77.45**
ES	**87.10**	76.11	86.75	75.91

10.3 Result

The stress detection result using individual classifiers is shown in Table 10.1 while the result using classifiers ensemble is displayed in Table 10.2. Table 10.1 depicts that all of the classifiers work quite well to conduct a stress detection task. All of the classifiers show adequate performance with an accuracy of more than 70%. RF obtained the best accuracy with 86.61% following by LR with a slight difference (85.46%). At third place was NN that has a slight margin to the first and second place (84.76%). These three top-classifiers were then used for the ensemble methods. Meanwhile, The lowest accuracy was achieved by KNN with a value of only 73%.

In terms of precision, NN has the best precision among other individual classifiers with a value of 76.53%. Furthermore, in terms of recall, RF has the highest value with 89.87%. However, the precision of RF is quite low (69.17%) so that it could not obtain the highest F1-measure. It means that RF tends to successfully detect almost all of the stress data available but many non-stress data are incorrectly labeled as stress. Meanwhile, LR has a more balance precision and recall so that it could achieve the best F1-measure with 76.25%. Similar to the accuracy result, KNN also has the lowest F1-measure (52.43%).

The ensemble methods were build using the three best individual classifiers from the previous results (RF, LR, and NN). Table 10.2 shows that both of the ensemble methods obtained higher performance compared to all of the individual classifiers. Generally, most individual classifiers have their own inherent defects [13] and their performance is also domain-dependent [9]. By combining some classifiers, the advantage of one classifier is expected to cover the shortcomings of other classifiers so that the performance can be improved. Soft and hard voting have different strategies to combine the result from the individual classifiers so that they can lead to different decisions.

The result displayed in Table 10.2 shows that ES (soft voting) has a higher performance than EH (hard voting) in this study in terms of accuracy. In contrast, EH has a better performance in terms of F1-measure. Generally, the soft voting strategy tends to get better performance than the hard voting strategy as it takes into account more information. Soft voting is smoother as it uses probability information to get the final decision. However, the additional information could also lead to a worse decision. In this study, the best accuracy is obtained by ES with 87.10%. Meanwhile, the best F1-measure was achieved by EH with 77.45%.

10.4 Conclusion

Several studies suggested that many hospital workers suffer from work-related stress. This condition can be harmful to both workers and the institution. Staff with acute stress levels can develop some diseases and mental problems. Furthermore, a high stress level also has a negative impact on the employee's work performance including low patient care performance and riskier cybersecurity practices. Enabling the workers to monitor their own stress level has many advantages. Knowing their own stress level can help them stay aware and feel more in control of their response to situations and know when it is time to relax or take some actions to treat it properly.

This monitoring task can be enabled by using wearable devices to measure physiological responses related to stress. Smartwatch is one of the devices that can be used for this task due to its usability for the working environment and its built-in sensors. In this work, we propose a multi-modal based continuous stress detection method using some individual classifiers and classifier ensembles.

The experiment results show that all of the classifiers work quite well to detect stress with an accuracy of more than 70%. RF obtained the best accuracy with 86.61% while KNN has the lowest accuracy with 73%. In terms of F1-measure, LR could achieve the best F1-measure with 76.25%. Similar to the accuracy result, KNN also has the lowest F1-measure (52.43%). The results also show that the ensemble method obtained higher performance compared to all of the individual classifiers. In this study, the advantage of one classifier can cover the shortcomings of other classifiers so that the accuracy can be improved. Furthermore, the results also show that ES (soft voting) has higher accuracy than EH (hard voting) in this study but EH has a better F1-measure than ES. In this study, the best accuracy is obtained by ES with 87.10%. Meanwhile, the best F1-measure was achieved by EH with 77.45%.

Our experimental study for the effect classifier ensemble is limited by the WESAD dataset that uses only two classes: stress and non-stress. In future work, the effect of the use of the ensemble method can be tested on a dataset that provides different stress levels (e.g. low stress, moderate stress, and high stress). Besides, a new dataset with more subjects could be created in the future in order to test the reliability of the proposed methods. The future dataset could also include not only label based on the intervention like in the WESAD dataset, but also the label from user-filled questionnaires (e.g. PSS).

10.5 Bibliography

[1] Aasland, O. G., Olff, M., Falkum, E., Schweder, T., and Ursin, H. Health complaints and job stress in norwegian physicians: the use of an overlapping questionnaire design. *Social science & medicine 45*, 11 (1997), 1615–1629. 246, 258

[2] Cohen, S., Kamarck, T., and Mermelstein, R. Perceived stress scale (pss). *J Health Soc Beh 24* (1983), 285. 4, 11, 23, 81, 82, 102, 123, 145, 246, 258

[3] Fauzi, M. A., and Bours, P. Ensemble method for sexual predators identification in online chats. In *2020 8th International Workshop on Biometrics and Forensics (IWBF)* (2020), IEEE, pp. 1–6. 29, 207, 248

[4] Fauzi, M. A., Yeng, P., Yang, B., and Rachmayani, D. Examining the link between stress level and cybersecurity practices of hospital staff in indonesia. In *The 16th International Conference on Availability, Reliability and Security* (2021), pp. 1–8. 14, 16, 21, 27, 100, 108, 117, 120, 132, 153, 164, 202, 218, 232, 246, 258

[5] Fauzi, M. A., and Yuniarti, A. Ensemble method for indonesian twitter hate speech detection. *Indonesian Journal of Electrical Engineering and Computer Science 11*, 1 (2018), 294–299. 12, 247, 248

[6] Garg, P., Santhosh, J., Dengel, A., and Ishimaru, S. Stress detection by machine learning and wearable sensors. In *26th International Conference on Intelligent User Interfaces* (2021), pp. 43–45. 5, 12, 202, 218, 247, 258

[7] Indikawati, F. I., and Winiarti, S. Stress detection from multimodal wearable sensor data. In *IOP Conference Series: Materials Science and Engineering* (2020), vol. 771, IOP Publishing, p. 012028. 5, 12, 13, 202, 218, 247, 258

[8] KREIBIG, S. D. Autonomic nervous system activity in emotion: A review. *Biological psychology 84*, 3 (2010), 394–421. 247, 260

[9] KUMAR, B. S., AND RAVI, V. Text document classification with pca and one-class svm. In *Proceedings of the 5th international conference on frontiers in intelligent computing: theory and applications* (2017), Springer, pp. 107–115. 43, 250

[10] LAZARO, M. J. S., LIM, J., KIM, S. H., AND YUN, M. H. Wearable technologies: acceptance model for smartwatch adoption among older adults. In *International Conference on Human-Computer Interaction* (2020), Springer, pp. 303–315. 4, 202, 232, 233, 246, 258

[11] LAZARUS, R. S., AND FOLKMAN, S. *Stress, appraisal, and coping*. Springer publishing company, 1984. 9, 118, 245, 257

[12] LEVENSTEIN, S., PRANTERA, C., VARVO, V., SCRIBANO, M. L., BERTO, E., LUZI, C., AND ANDREOLI, A. Development of the perceived stress questionnaire: a new tool for psychosomatic research. *Journal of psychosomatic research 37*, 1 (1993), 19–32. 4, 11, 246, 258

[13] LI, L., ZHANG, Y., ZOU, L., LI, C., YU, B., ZHENG, X., AND ZHOU, Y. An ensemble classifier for eukaryotic protein subcellular location prediction using gene ontology categories and amino acid hydrophobicity. *PLoS One 7*, 1 (2012), e31057. 43, 250

[14] LIAO, W., ZHANG, W., ZHU, Z., AND JI, Q. A real-time human stress monitoring system using dynamic bayesian network. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)-workshops* (2005), IEEE, pp. 70–70. 4, 202, 232, 246, 258

[15] MARINE, A., RUOTSALAINEN, J. H., SERRA, C., AND VERBEEK, J. H. Preventing occupational stress in healthcare workers. *Cochrane Database of Systematic Reviews*, 4 (2006). 246, 258

[16] PICKERING, T. G. Mental stress as a causal factor in the development of hypertension and cardiovascular disease. *Current hypertension reports 3*, 3 (2001), 249–254. 202, 218, 232, 246, 258

[17] REISS, A., AND STRICKER, D. Introducing a new benchmarked dataset for activity monitoring. In *2012 16th international symposium on wearable computers* (2012), IEEE, pp. 108–109. 247

[18] SCHMIDT, P., REISS, A., DUERICHEN, R., MARBERGER, C., AND VAN LAERHOVEN, K. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 20th ACM international conference on multimodal interaction* (2018), pp. 400–408. 5, 12, 13, 26, 202, 203, 211, 218, 219, 228, 233, 247, 258, 259

[19] Siirtola, P. Continuous stress detection using the sensors of commercial smartwatch. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (2019), pp. 1198–1201. 5, 12, 13, 202, 211, 218, 228, 247, 258

[20] Spagnolli, A., Guardigli, E., Orso, V., Varotto, A., and Gamberini, L. Measuring user acceptance of wearable symbiotic devices: validation study across application scenarios. In *International Workshop on Symbiotic Interaction* (2015), Springer, pp. 87–98. 4, 202, 232, 233, 246, 258

[21] Tsiga, E., Panagopoulou, E., and Montgomery, A. Examining the link between burnout and medical error: A checklist approach. *Burnout Research 6* (2017), 1–8. 80, 202, 218, 232, 246, 258

[22] Wakeling, A. *Stress and anxiety at work: personal or cultural?* ACAS, 2019. 245, 257

[23] Wang, Y., Chen, R., and Zhang, L. Reliability and validity of generalized anxiety scale-7 in inpatients in chinese general hospital. *J Clin Psychiatr 28* (2018), 168–71. 202, 232, 246, 258

[24] Weinberg, A., and Creed, F. Stress and psychiatric disorder in healthcare professionals and hospital staff. *the Lancet 355*, 9203 (2000), 533–537. 246, 258

[25] Welp, A., Meier, L. L., and Manser, T. Emotional exhaustion and workload predict clinician-rated and objective patient safety. *Frontiers in psychology 5* (2015), 1573. 80, 202, 218, 232, 246, 258

[26] Wemm, S. E., and Wulfert, E. Effects of acute stress on decision making. *Applied psychophysiology and biofeedback 42*, 1 (2017), 1–12. 2, 11, 40, 41, 80, 119, 144, 153, 162, 187, 202, 246, 258

[27] Zhang, Y., Haghdan, M., and Xu, K. S. Unsupervised motion artifact detection in wrist-measured electrodermal activity data. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers* (2017), pp. 54–57. 247, 260

Part IV

Privacy-Preserving Stress Detection

Chapter 11

Comparative Analysis Between Individual, Centralized, and Federated Learning for Smartwatch Based Stress Detection

Muhammad Ali Fauzi; Bian Yang; Bernd Blobel

Abstract

Machine learning has been proven to provide good performances on stress detection tasks using multi-modal sensor data from a smartwatch. Generally, machine learning techniques need a sufficient amount of data to train a robust model. Thus, we need to collect data from several users and send them to a central server to feed the algorithm. However, the uploaded data may contain sensitive information that can jeopardize the user's privacy. Federated learning can tackle this challenge by enabling the model to be trained using data from all users without the user's data leaving the user's device. In this study, we implement federated learning-based stress detection and provide a comparative analysis between individual, centralized, and federated learning. The experiment was conducted on WESAD dataset by using Logistic Regression as the classifier. The experiment results show that in terms of accuracy, federated learning cannot reach the performance level of both individual and centralized learning. The individual learning strategy performs best with an average accuracy of 0.9998 and an average F_1-measure of 0.9996.

11.1 Introduction

In today's busy world, stress has become an interesting issue in recent years, gaining awareness in many countries. Stress can be defined as a unique affective state that occurs when an individual considers that their perceived resources or ability cannot cope with the perceived demand of a stimulus [17]. The latest survey by Acas in 2019 [34] about stress and anxiety at work reported that about 66% of working people have experienced work-related stress in the last 12 months. Hospital employees, who in fact are very familiar with this issue, are also exposed to high levels of work-related stress [22, 36, 1].

Stress at a low level is acceptable or maybe even positive, also called eustress. However, prolonged stress can have a negative impact on our physical, mental, and emotional health. Many studies reported that stress has a significant impact on the development of hypertension and coronary artery disease, diabetes, asthma, etc. [26]. Moreover, excessive stress also harms the employee's productivity, increases absenteeism, and plays a crucial role in mental illness development, such as generalized anxiety disorder and depression [35]. According to studies, in the hospital setting for example, a higher stress level is significantly correlated with low patient safety [37, 33]. Another study also suggested that a higher stress level of hospital staff results in riskier cybersecurity practices [8]. These studies are in line with a prior study [38], reporting that stressed people will be slow in learning something new and may choose less profitable decisions.

Monitoring an individual's stress level has many advantages. Knowing their own stress level can help them in staying aware and feeling more in control of their response to situations and knowing when it is time to relax or take some actions to treat it properly [20]. Furthermore, this monitoring can help to early diagnose mental illness and disorders. The most common way to assess a stress level is the use of questionnaires (e.g., Perceived Stress Scale [4], Perceived Stress Questionnaire [18], etc.). However, this method takes time, so it is not convenient to use every day for continuous monitoring. Another approach for determining stress levels is to measure stress-related physiological reactions using sensors. The smartwatch is one of the most suitable devices to perform this stress monitoring task, especially in the working environment. A smartwatch offers a number of built-in sensors that can be used for multimodal-based stress detection including blood volume pulse, electrodermal activity, skin temperature, accelerometer, etc. Unlike many wearable devices that have very low usability and are not convenient to wear during work (e.g., chest-worn devices, finger-placed galvanic skin response (GSR) sensors, etc.), the smartwatch is well known and has a high degree of social acceptance due to their ubiquity in everyday life [31, 16].

There has been a remarkable success of machine learning (ML) technologies in empowering practical artificial intelligence (AI) applications, including in medical fields. Many prior studies have used multi-modal sensor data and machine learning methods to develop stress detection systems such as Decision Tree, K-Nearest Neighbors (KNN), Random Forest, and Logistic Regression [28, 9, 11, 29]. Machine learning techniques generally need a sufficient amount of data for training to perform well. Therefore, to create a

Table 11.1: Participants' demographic characteristics in the WESAD dataset (N = 15).

Characteristic	Value, Mean (SD)
Age (years)	27.5 (2.4)
Height (cm)	177.6 (6.7)
Weight (kg)	73.1 (10.3)

robust method, we need to collect sensor data from several users and collect them at a central server for processing. However, the uploaded medical data may contain individual privacy-related and sensitive information. Privacy breaches can happen if the central server is compromised. Furthermore, the leakage can also happen even when well-intentioned individuals, who have access to the server, share the data for legitimate purposes. As a result, a growing number of studies place attention on safeguarding private data in analysis processes. Federated learning (FL) can be the solution to this privacy challenge. FL works by allowing each data register to train models on separate, isolated datasets while only sharing the trained models, which do not contain any personal information. The registers then send their models to a central server for aggregating them to a single, integrated model. This process is repeated for a number of iterations until a high-quality model is produced. In this work, we implement FL-based stress detection and provide a comparative analysis between individual, centralized, and federated learning.

The remainder of this paper is organized as follows. The introduction part is given in Section 11.1. Dataset, features, learning strategies, and evaluation methods for the stress detection task are explained in Section 11.2. The results and discussion of this paper are described in Sections 11.3 and 11.4, while conclusions are provided in Section 11.5.

11.2 Materials and Methods

11.2.1 Dataset

A public dataset called WESAD (Wearable Stress and Affect Detection) [28] was used in this study. The dataset was created in the lab by the Ubiquitous Computing research group at the University of Siegen, Germany, and was made public in 2018. The data came from 15 participants consisting of 12 males and 3 females. The demographic information of the participants in this dataset is displayed in Table 11.1.

The data in the WESAD study were acquired using an Empatica E4 smartwatch and a RespiBAN chest band at the same time during specified tasks designed to capture three different affective states: neutral, stress, and amusement. Only Empatica E4 data are used in this study because the focus of this work is on smartwatch sensors. The built-in sensors on the smartwatch are skin temperature (ST), accelerometers (ACC), electrodermal activity (EDA), and blood volume pulse sensors (BVP). Each individual had a data collection session of at least 36.5 min, which included the neutral position for approximately 20 min, the stress situation for 10 min, and the amusement situation for around 6.5 minutes. During the neutral position, the participants were sitting/standing and neutrally reading provided magazines. During the stress situation, the participants faced the Trier Social Stress Test (TSST) [13] to induce their stress, whereas during the amusement situation, the participants watched a set of funny video clips. The neutral and relaxation sessions were combined into one non-stress class for the stress detection task in this study so that the classification problem was binary (stress and non-stress).

11.2.2 Features

In this study, we employed all the sensors' data on the smartwatch including ST, ACC, EDA, and BVP. To extract the features, the signal data were segmented by using a 60-second sliding window with a sliding step of 0.25 s following the recommendation by Kreibig et al. [14]. Furthermore, we constructed 6 different signals for each sensor's data: the original signal; its first and second derivatives; and the transformed signal data using a Discrete Wavelet Transform (DWT) with the Haar wavelet at 3 different frequencies (1 Hz, 2 Hz, and 4 Hz). Wavelet transforms can catch both frequency and time information, while immediate changes in signals can be captured by the Haar wavelet [39]. For the ACC data, in addition to the 3-dimensional signal data (x, y, and z-axis that are represented by ACC_x, ACC_y, and ACC_z, respectively), we also calculated their magnitude (ACC_{norm}) using Equation (11.1). In total, we have used signals consisting of 6 ST signals, 24 ACC signals, 6 EDA signals, and 6 BVP signals as displayed in Table 11.2. In the last step, we extracted 10 statistical features using BioSPPy and Numpy libraries [10] in Python as displayed in Table 11.3. In total, 420 features were analyzed for this study.

$$ACC_{norm} = \sqrt{ACC_x^2 + ACC_y^2 + ACC_z^2} \tag{11.1}$$

11.2.3 Learning Strategies

In this study, three learning strategies are compared: individual learning; centralized learning; and federated learning. All those learning strategies used Logistic Regression (LR) as the machine learning model. LR is selected

Table 11.2: Signal data used in this study.

Sensor	**Signal**
Skin temperature (ST)	ST original signal
	ST first derivative signal
	ST second derivative signal
	ST signal with DWT with the Haar wavelet at 4 Hz
	ST signal with DWT with the Haar wavelet at 2 Hz
	ST signal with DWT with the Haar wavelet at 1 Hz
Accelerometers (ACC)	ACC_x original signal
	ACC_x first derivative signal
	ACC_x second derivative signal
	ACC_x signal with DWT with the Haar wavelet at 4 Hz
	ACC_x signal with DWT with the Haar wavelet at 2 Hz
	ACC_x signal with DWT with the Haar wavelet at 1 Hz
	ACC_y original signal
	ACC_y first derivative signal
	ACC_y second derivative signal
	ACC_y signal with DWT with the Haar wavelet at 4 Hz
	ACC_y signal with DWT with the Haar wavelet at 2 Hz
	ACC_y signal with DWT with the Haar wavelet at 1 Hz
	ACC_z original signal
	ACC_z first derivative signal
	ACC_z second derivative signal
	ACC_z signal with DWT with the Haar wavelet at 4 Hz
	ACC_z signal with DWT with the Haar wavelet at 2 Hz
	ACC_z signal with DWT with the Haar wavelet at 1 Hz
	ACC_{norm} original signal
	ACC_{norm} first derivative signal
	ACC_{norm} second derivative signal
	ACC_{norm} signal with DWT with the Haar wavelet at 4 Hz
	ACC_{norm} signal with DWT with the Haar wavelet at 2 Hz
	ACC_{norm} signal with DWT with the Haar wavelet at 1 Hz
Electrodermal activity (EDA)	EDA original signal
	EDA first derivative signal
	EDA second derivative signal
	EDA signal with DWT with the Haar wavelet at 4 Hz
	EDA signal with DWT with the Haar wavelet at 2 Hz
	EDA signal with DWT with the Haar wavelet at 1 Hz
Blood volume pulse sensors (BVP)	BVP original signal
	BVP first derivative signal
	BVP second derivative signal
	BVP signal with DWT and the Haar wavelet at 4 Hz
	BVP signal with DWT and the Haar wavelet at 2 Hz
	BVP signal with DWT and the Haar wavelet at 1 Hz

Table 11.3: Statistical Features.

No.	Features
1	Mean of the Signal
2	Minimum value of the signal
4	Maximum value of the signal
4	Median of the signal
5	Maximum signal amplitude
6	Signal variance
7	Standard signal deviation
8	Absolute signal deviation
9	Signal kurtosis
10	Signal skewness

due to its good performance in stress detection tasks [7, 15, 40]. LR also provides relatively low computational complexity, compared to Deep Neural Networks (DNN), for example. Thus, it does not need a device with high computational power. LR in this study is implemented using the Scikit-learn library [25].

11.2.3.1 Individual Learning

In this scheme, each user had their own model. As displayed in Figure 11.1, the user's data never left their device. Using this scheme, the user's device captured the sensor data, extracted the features, and then trained their individual machine learning model using their own data. In the end, each user attained a model personalized for them. Since there are 15 participants, there have been 15 separate models for each participant in this study. Like the raw sensor data, this model never left the user's device and has never been shared with other users. The model will be used later on to detect the user's stress. To be noted, this learning strategy needs a device that has enough computational power to perform the feature extraction and model training tasks.

This scheme offers a very high level of privacy because no data or model left the user's device. Unlike the two other schemes, individual learning does not need a central server to combine the data or model, so it can minimize the cost. However, it prevents information sharing across users that generally can improve the performance of a machine learning model. In addition, if there is a new user, they cannot use the stress detection system right after the registration. The new user must collect their own stress data to train their individual model.

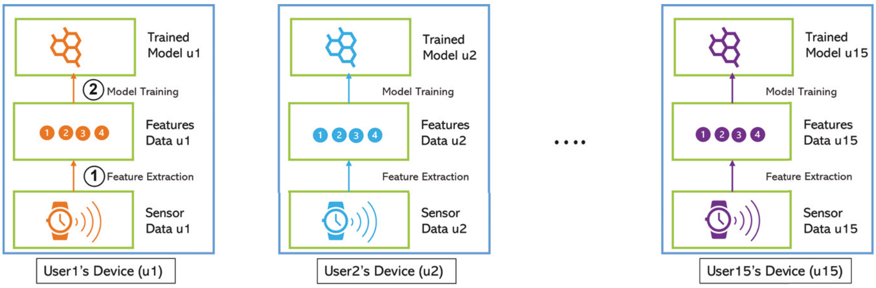


Figure 11.1: Individual Learning Scheme.

11.2.3.2 Centralized Learning

In this scheme, we only have a single integrated model. Unlike individual learning, this learning strategy needs a central server to combine the data and train the integrated model. As shown in Figure 11.2, each user's device captures the sensor data and then sends the raw data to the central server. Thereafter, the central server combines all the data from all users, extracts the features, and then trains a machine learning model using the combined data. As result, a single integrated model is created. This model is then sent to each user's device and is used later to detect the user's stress. Since the feature extraction and model training tasks are conducted on the central server side, this learning strategy does not need a device with high computational power. The user device only needs to do the stress detection/inference task using the model. Depending on the size of the dataset, training often takes several hours or more to complete. This stage of the process demands the greatest CPU or GPU power. The inference task on the other hand usually needs far less computing power than the training task. To minimize the computing power needed on the user's device, the integrated model in this scheme can be stored on the server. When the user needs to perform the inference task on new data, the device can send the data to the server, and the server will detect the stress level of the data using the model and send the result back. However, this strategy requires the user's device to be always online. If the integrated model is saved on each user's device, the user's device does not need to be online to predict the stress level.

This scheme offers a very low level of privacy because the user data leaves her/his device. This is sensitive data that can be used to disclose users' personal information and their health status. However, it enables information sharing across users that generally can increase the robustness of a machine learning model. The other advantage of using this scheme is that a new user can use the stress detection system right after the registration by deploying the integrated model. The new user does not need to collect their

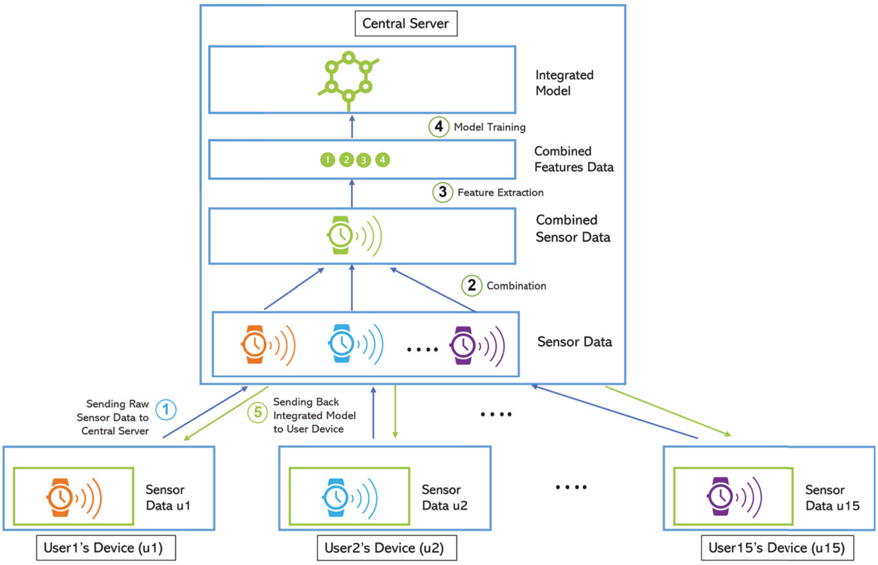


Figure 11.2: Centralized Learning Scheme.

own stress data and do the data labeling.

11.2.3.3 Federated Learning (FL)

As displayed in Figure 11.3, the federated learning scheme is similar to centralized learning in terms of needing a central server and having just a single integrated model. The main difference between centralized and federated learning is that the user's data will never leave the user's device in federated learning, that way maintaining the user's privacy. Federated learning in this study is implemented using Flower [3] with FederatedAveraging (FedAvg) aggregation strategy [23].

Stress data from sensors contain sensitive information that can be used to disclose users' personal information and their health status. Therefore, the stress detection system needs to give more attention to privacy concerns. In Europe, the General Data Protection Regulation (GDPR) protects the users' privacy by limiting the exchange of sensitive data [6]. On the other hand, the use of sensor data has many potential benefits. Therefore, a new family of privacy-preserving technologies is emerging to solve this problem. The goal of privacy-preserving technologies is to make the most of the data without jeopardizing users' privacy. This technology employs strategies to reduce the amount of personal data held while maintaining the analysis operation. Several privacy-preserving methods have been proposed, and one of the techniques with high potential is Federated Learning.

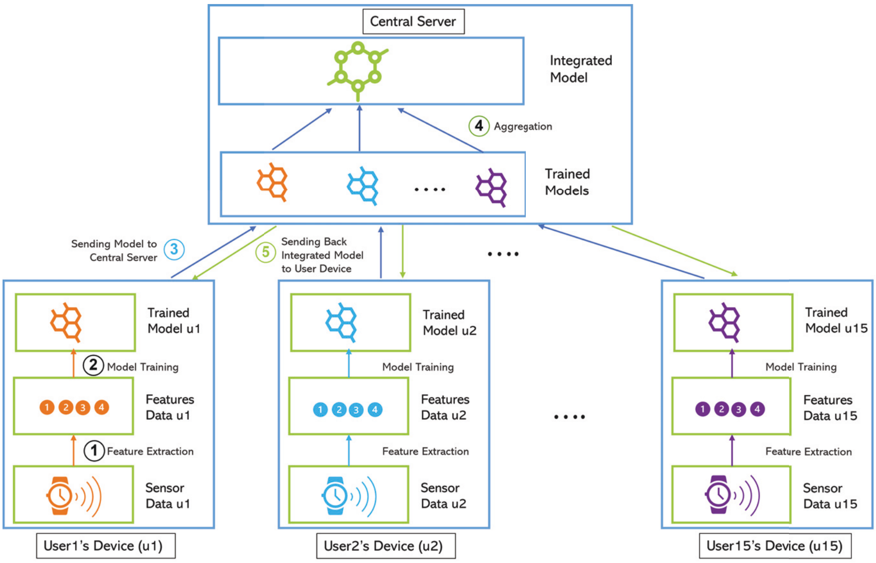


Figure 11.3: Federated Learning Scheme.

Federated learning is a learning paradigm that aims to solve the problem of data privacy by collectively training algorithms without transferring data [23]. It has recently acquired popularity in healthcare applications [27, 2]. FL allows for collaboratively using datasets without transferring the raw patient data outside of the institutions' databases. As shown in Figure 11.3, each user's device captures the sensor data and extracts the features. Furthermore, the machine learning model is trained locally on each user's device. Next, the trained model is uploaded to the central server so that the central server can combine all the models and share the integrated model with each user's device. This model will be used later to infer the user's stress level. Some works show that models trained by FL can obtain performance levels comparable to those trained on centrally hosted data sets and exceeds models that only see isolated single-device data [19]. Successful implementation of FL could have a huge impact on enabling large-scale precision medicine, resulting in unbiased models while also respecting privacy issues [27]. To be noted, this learning strategy needs a device that has enough computational power to do the feature extraction and local model training tasks.

The federated learning scheme offers a very high level of privacy, because no data is leaving the user's device. This scheme also enables information sharing across users that generally can improve the robustness of a machine learning model. In addition, if there is a new user, she/he can use

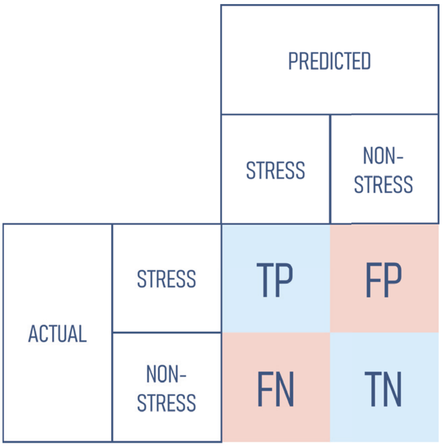


Figure 11.4: Confusion Matrix. Blue square means the data are correctly predicted while red square means the data are incorrectly predicted.

the stress detection system right after the registration by using the integrated model without doing data collection first.

11.2.4 Evaluation

In this study, each data set is divided into two parts: training and testing data with a split ratio of 80:20. All the strategies use the training data for model training and testing data to evaluate the model performance. Several measurements including Accuracy (Acc), Precision (P), Recall (R), and F$_1$-measure (F_1) were deployed for classifier performance evaluation. All measurements were calculated based on the confusion matrix displayed in Figure 11.4. True Positive (TP) and True Negative (TN) are the numbers of data that were correctly predicted. TP represents the number of stress data that were correctly predicted as stress, while TN represents the number of non-stress data that were correctly predicted as non-stress. Meanwhile, False Positive (FP), often called Type I Error, is the number of non-stress data that were incorrectly predicted as stress data, and False Negative (FN) or Type II Error represents the number of stress data that were incorrectly predicted as non-stress data.

The formulas for all measurements are displayed in Equation (11.2), Equation (11.3), Equation (11.4), and Equation (11.4) respectively.

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{11.2}$$

$$P = \frac{TP}{TP + FP} \tag{11.3}$$

Table 11.4: Individual Learning Result.

Participant	**Acc**	**P**	**R**	**F_1**
1	1.0000	1.0000	1.0000	1.0000
2	1.0000	1.0000	1.0000	1.0000
3	1.0000	1.0000	1.0000	1.0000
4	1.0000	1.0000	1.0000	1.0000
5	1.0000	1.0000	1.0000	1.0000
6	1.0000	1.0000	1.0000	1.0000
7	1.0000	1.0000	1.0000	1.0000
8	1.0000	1.0000	1.0000	1.0000
9	1.0000	1.0000	1.0000	1.0000
10	1.0000	1.0000	1.0000	1.0000
11	1.0000	1.0000	1.0000	1.0000
12	0.9994	0.9980	1.0000	0.9990
13	0.9970	0.9960	0.9941	0.9951
14	1.0000	1.0000	1.0000	1.0000
15	1.0000	1.0000	1.0000	1.0000
Average	**0.9998**	**0.9996**	**0.9996**	**0.9996**

$$R = \frac{TP}{TP + FN} \tag{11.4}$$

$$F_1 = 2\frac{P \cdot R}{P + R} \tag{11.5}$$

11.3 Results

The results of stress detection using individual learning, centralized learning, and federated learning are presented in Tables 11.4–11.6. The experimental results show that individual learning is the most appropriate strategy for this task by obtaining an almost perfect performance with an average accuracy of 0.9998, an average precision of 0.9996, an average recall of 0.9996, and an average F_1-measure of 0.9996. All individual models of the participants achieved 100% accuracy and F_1-measure. Even the poorest individual model provided an accuracy of 0.9970 and F_1-measure of 0.9951, which can still be considered almost perfect.

Meanwhile, centralized learning had also a good performance with an average accuracy of 0.9355, an average precision of 0.9125, an average recall of 0.8698, and an average F_1-measure of 0.8783. The single integrated model from the centralized learning is excellent for inferring the stress level of most of the participants. The model achieved an accuracy below 0.9 just for three participants' data (participant 5, 8, and 13). In terms of F_1-measure,

Table 11.5: Centralized Learning Result.

Participant	Acc	P	R	F_1
1	0.9414	0.8250	1.0000	0.9041
2	0.9317	0.9809	0.7809	0.8696
3	0.9660	0.8916	1.0000	0.9427
4	0.9571	0.8716	1.0000	0.9314
5	0.8833	0.9658	0.5853	0.7288
6	0.9511	0.8726	0.9720	0.9196
7	0.9772	0.9827	0.9401	0.9609
8	0.8545	0.9674	0.4771	0.6390
9	0.9244	1.0000	0.7495	0.8568
10	0.9957	0.9880	0.9980	0.9930
11	0.9475	0.8540	0.9851	0.9149
12	0.9353	0.8812	0.9127	0.8967
13	0.8837	0.8575	0.7475	0.7987
14	0.9437	0.8400	1.0000	0.9130
15	0.9404	0.9098	0.8994	0.9046
Average	**0.9355**	**0.9125**	**0.8698**	**0.8783**

the model achieved a value below 0.9 for six participants' data. The model best performed on the data of participant 10 with an accuracy of 0.9957, precision of 0.9880, recall of 0.9980, and F_1-measure of 0.9930. In contrast, the worst result was gathered when detecting the stress level of participant 8 with an accuracy of 0.8545, precision of 0.9674, recall of 0.4771, and F_1-measure of 0.6390.

Based on Table 11.6, federated learning had a relatively mediocre performance for the stress detection tasks in this study. It obtained an average accuracy of 0.8575, an average precision of 0.9892, an average recall of 0.5208, and an average of F_1-measure of 0.6339. The integrated model from federated learning performed quite well on most of the participants' data but performed very poorly on the data of some participants. This model achieved an F_1-measure below 0.5 for 5 participants (participant 2, 4, 8, 9, and 13). The integrated model achieved the best result on the data of participant 3 with an accuracy of 0.9969, precision of 1.0000, recall of 0.9887, and F_1-measure of 0.9943. On the contrary, the model performs the worst inferring the stress level of participant 4, with an accuracy of 0.7259, precision of 1.0000, recall of 0.0589, and F_1-measure of 0.1113.

The study results suggest that the individual model achieved the best stress detection performance. This scheme outperformed both centralized learning and federated learning because it offers personalization by training the model separately for each user, using the user's own data. The WESAD

Table 11.6: Federated Learning Result.

Participant	**Acc**	**P**	**R**	**F_1**
1	0.9131	0.8675	0.8089	0.8372
2	0.7565	0.9872	0.1670	0.2857
3	0.9969	1.0000	0.9887	0.9943
4	0.7259	1.0000	0.0589	0.1113
5	0.8511	1.0000	0.4447	0.6156
6	0.8700	1.0000	0.5484	0.7083
7	0.8578	1.0000	0.5227	0.6866
8	0.7796	1.0000	0.1835	0.3101
9	0.7820	1.0000	0.2781	0.4352
10	0.9390	0.9950	0.8016	0.8879
11	0.9524	1.0000	0.8337	0.9093
12	0.9097	0.9917	0.7123	0.8291
13	0.7620	1.0000	0.2288	0.3724
14	0.8880	0.9967	0.6232	0.7669
15	0.8778	1.0000	0.6110	0.7585
Average	**0.8575**	**0.9892**	**0.5208**	**0.6339**

dataset labels the data based on the stimulus given to the participants. All the data recorded during the neutral and amusement condition, where the participants were reading magazines and watching funny videos, were labeled as non-stress, whereas all of the data recorded during the TSST session were labeled as stress. Different individuals will react to the stressors with varying intensity or duration [12]. Therefore, the personalized approach like the individual learning model surpasses the integrated model provided by centralized learning and federated learning. The integrated model aims at building a single model for all, so that it cannot adjust for each user.

These results also demonstrate that some models achieved quite good accuracy on some participants, but had a very poor F_1-measure. To be noted, the stress dataset used in this study is imbalanced. It has more non-stress data than stress data. Therefore, accuracy is not good enough to be used as the evaluation measure. We need to perform the evaluation using precision, recall, and F_1-measure. High accuracy means that the model can well predict the class. However, it is important to mention that accuracy is based on True Positive (TP) and True Negative (TN). In an imbalanced dataset where the number of non-stress data is higher than stress data, high accuracy may be achieved because the value of TN is very high even though the value of TP is very low. As an extreme example, if we have 100 testing data containing 90 non-stress data and 10 stress data and the model predicts all of the testing data as non-stress, the model will still get very good accuracy with

0.9. In this example, the model gets 90 TN and 0 TP. This model is actually not good because it cannot predict any stress data even though the accuracy is very high. In contrast with accuracy, the F_1-measure of this model will be very low. Picking an example from the experimental result, the integrated model from federated learning applied to participant 4's data achieved an accuracy of 0.7259, precision of 1.0000, recall of 0.0589, and F_1-measure of 0.1113. The low recall with high precision means that the data predicted as stress by the model are very few, but most of the predicted labels are correct. In other words, this model mostly predicts the data as non-stress so that the TN value is very high, resulting in a high-value accuracy even though the TP value is very low because only a small amount of data were predicted as stress. In contrast with the accuracy, the F_1-measure of this model is very low. Therefore, in an imbalanced dataset, F_1-measure is a better measurement than accuracy.

11.4 Discussion

This paper discusses the comparison of individual learning, centralized learning, and federated learning on the WESAD stress detection dataset. Generally, more data will make the machine learning model better and more accurate, because the more information we give to the model, the more it will learn and the more cases it will be able to correctly infer [32]. Therefore, integrated models such as centralized and federated learning are expected to be more accurate than individual learning. Surprisingly, the individual model surpasses in this study both the centralized and the federated learning as depicted in Figure 11.5. The WESAD dataset labels the data based on the stimulus given to the participants. Different participants may react differently to each stimulus. In this case, the personalized approach such as the individual learning model can adjust the model to the user's behavior. The integrated model aims at building a single model for all so that it cannot adjust for each user. This study outcome is in line with another study about stress detection that also reported that a personalized model outperformed an integrated model [21].

Generally, federated learning is expected to perform worse than centralized learning. It is because centralized learning has direct access to all data while federated learning train the model locally and only communicates an updated model to a central server [24]. Surprisingly, the performance difference between the two strategies is very big. A more complex model such as Deep Neural Network (DNN) is needed to build a better federated learning model. Some previous work shows that federated learning with DNN can obtain performance levels comparable to those models trained using a centralized learning scheme [24, 21]. Another study also suggested that less complex models perform worse than more complex models in federated

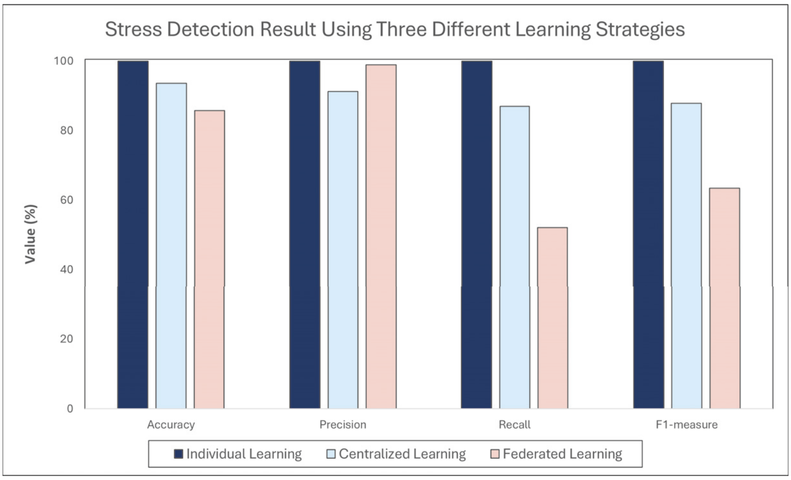


Figure 11.5: Stress Detection Results Using Three Different Learning Strategies.

learning [30]. However, a more complex model requires the user's device to have a higher computational power to train the model. Additionally, a more complex model will also lead to higher communication costs between the user's device and the central server. Thus, there will be a challenge to use a complex model for communication-sensitive applications [30].

Furthermore, since the WESAD dataset in this study is labeled based on the stimulus, there may be the possibility that the labels do not represent the participants' actual stress levels. For example, during the TSST situation, there is the possibility that the participant was not feeling stressed (e.g., because they are good at public speaking) but all their gathered data during that session will be labeled as stress. Another issue could be that a participant was feeling stressed while watching the funny videos, because it reminded them of some traumatic events, for example, but all of their data during that session will be labeled as non-stress. Therefore, it will be of interest to see the comparison between the personalized and the integrated model on the stress dataset that is labeled based on the user's subjective stress level measurement. In addition, the WESAD data collection was conducted in one session, which will make the data very similar. Thus, it is also of interest to see the comparison on the stress dataset, that is collected on multiple sessions to see how the model can perform across sessions.

Another factor that can also be considered is the usability of the three learning schemes for a new user. For centralized and federated learning, the new users can use the integrated model to predict their stress level right after the registration. For individual learning, however, the user must collect training data first. The users should record their data using the smartwatch during stress and non-stress condition. The users must also give the correct label to the data because the quality of the model heavily depends on the training data quality. This training data is used to train the personalized model for the users before they can infer their stress level automatically.

In addition, the computational cost is also different between these three schemes. Individual learning demands that a user's device has enough computing power for feature extraction, model training, and stress detection tasks. Meanwhile, centralized learning requires less computing power for a user's device, because all of the processes can be done on the central server. However, the device has to be always online since the device has to send the data to the central server. Federated learning needs a user's device that has enough computing power to do the local training as well as a communication channel to exchange data between the device and the centralized server.

Finally, stress data are considered sensitive as they can be used to disclose the user's health status. Based on a study on health data privacy, most of the interview subjects are worried about their data privacy on an individual level [5]. Therefore, the processing of this kind of data needs to pay more attention to privacy concerns. In centralized learning, all the data are collected on a centralized server. When these data are shared with the central server, privacy leaks can occur if the central server is compromised. Therefore, centralized learning can jeopardize users' privacy. On the contrary, individual and federated learning strategies offer a high level of privacy. In federated learning, only the learning model, and no raw user data, is processed centrally. Meanwhile, individual learning provides a higher level of privacy as it does not require any user data or model to leave the user's device.

11.5 Conclusions

In this study, the comparison between individual, centralized, and federated learning for smartwatch-based stress detection is discussed. In terms of accuracy, the individual learning strategy beats both centralized learning and federated learning. This is quite reasonable because different participants may react differently to stressors, so a personalized model is needed. The integrated model aims to build a single model for all so that it cannot adjust for each user. In terms of privacy, centralized learning requires all of the data to be shared with a centralized server. There is a risk of privacy breach, when the central server got compromised. In contrast, the individual learning strategy offers a very high level of privacy, since it does not require any user data or model to leave the user's device. Federated learning also offers a high level of privacy, since only the learned model, and no raw user data, is processed in the central server. The only disadvantage of individual learning is the low usability for a new user. For centralized and federated learning, the new users can use the integrated model to infer their stress level right after the registration. In contrast, for individual learning, the users must collect training data first to build the personalized model.

In future work, a more complex model such as DNN can be used to improve the federated learning scheme performance. In addition, it will be interesting to see the comparison between individual learning, centralized, and federated learning on the stress dataset that is labeled based on the user's subjective stress level measurement and collected on multi sessions, instead of only a single session.

11.6 Bibliography

[1] Aasland, O. G., Olff, M., Falkum, E., Schweder, T., and Ursin, H. Health complaints and job stress in norwegian physicians: the use of an overlapping questionnaire design. *Social science & medicine 45*, 11 (1997), 1615–1629. 246, 258

[2] Ahmed, S. T., Kumar, V. V., Singh, K. K., Singh, A., Muthukumaran, V., and Gupta, D. 6g enabled federated learning for secure iomt resource recommendation and propagation analysis. *Computers and Electrical Engineering 102* (2022), 108210. 265

[3] Beutel, D. J., Topal, T., Mathur, A., Qiu, X., Parcollet, T., de Gusmão, P. P., and Lane, N. D. Flower: A friendly federated learning research framework. *arXiv preprint arXiv:2007.14390* (2020). 264

[4] Cohen, S., Kamarck, T., and Mermelstein, R. Perceived stress scale (pss). *J Health Soc Beh 24* (1983), 285. 4, 11, 23, 81, 82, 102, 123, 145, 246, 258

[5] De Maeyer, C., and Markopoulos, P. Are digital twins becoming our personal (predictive) advisors?:'our digital mirror of who we were, who we are and who we will become'. In *22st International Conference on Human-Computer Interaction'20: HCI International 2020* (2020), Springer, pp. 250–268. 45, 272

[6] Domingo-Ferrer, J., and Blanco-Justicia, A. Privacy-preserving technologies. In *The Ethics of Cybersecurity*. Springer, Cham, 2020, pp. 279–297. 264

[7] Fauzi, M. A., and Yang, B. Continuous stress detection of hospital staff using smartwatch sensors and classifier ensemble. In *pHealth 2021*. IOS Press, 2021, pp. 245–250. 28, 202, 211, 219, 228, 232, 238, 262

[8] Fauzi, M. A., Yeng, P., Yang, B., and Rachmayani, D. Examining the link between stress level and cybersecurity practices of hospital staff in indonesia. In *The 16th International Conference on Availability, Reliability and Security* (2021), pp. 1–8. 14, 16, 21, 27, 100, 108, 117, 120, 132, 153, 164, 202, 218, 232, 246, 258

[9] Garg, P., Santhosh, J., Dengel, A., and Ishimaru, S. Stress detection by machine learning and wearable sensors. In *26th International Conference on Intelligent User Interfaces* (2021), pp. 43–45. 5, 12, 202, 218, 247, 258

[10] Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., et al. Array programming with numpy. *Nature 585*, 7825 (2020), 357–362. 204, 220, 235, 260

[11] Indikawati, F. I., and Winiarti, S. Stress detection from multimodal wearable sensor data. In *IOP Conference Series: Materials Science and Engineering* (2020), vol. 771, IOP Publishing, p. 012028. 5, 12, 13, 202, 218, 247, 258

[12] Jacoby, R., Greenfeld Barsky, K., Porat, T., Harel, S., Hanalis Miller, T., and Goldzweig, G. Individual stress response patterns: Preliminary findings and possible implications. *Plos one 16*, 8 (2021), e0255889. 269

[13] Kirschbaum, C., Pirke, K.-M., and Hellhammer, D. H. The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology 28*, 1-2 (1993), 76–81. 20, 121, 165, 203, 219, 233, 260

[14] Kreibig, S. D. Autonomic nervous system activity in emotion: A review. *Biological psychology 84*, 3 (2010), 394–421. 247, 260

[15] Kurniawan, H., Maslov, A. V., and Pechenizkiy, M. Stress detection from speech and galvanic skin response signals. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems* (2013), IEEE, pp. 209–214. 202, 219, 232, 262

[16] Lazaro, M. J. S., Lim, J., Kim, S. H., and Yun, M. H. Wearable technologies: acceptance model for smartwatch adoption among older adults. In *International Conference on Human-Computer Interaction* (2020), Springer, pp. 303–315. 4, 202, 232, 233, 246, 258

[17] Lazarus, R. S., and Folkman, S. *Stress, appraisal, and coping.* Springer publishing company, 1984. 9, 118, 245, 257

[18] LEVENSTEIN, S., PRANTERA, C., VARVO, V., SCRIBANO, M. L., BERTO, E., LUZI, C., AND ANDREOLI, A. Development of the perceived stress questionnaire: a new tool for psychosomatic research. *Journal of psychosomatic research 37*, 1 (1993), 19–32. 4, 11, 246, 258

[19] LI, W., MILLETARÌ, F., XU, D., RIEKE, N., HANCOX, J., ZHU, W., BAUST, M., CHENG, Y., OURSELIN, S., CARDOSO, M. J., ET AL. Privacy-preserving federated brain tumour segmentation. In *International workshop on machine learning in medical imaging* (2019), Springer, pp. 133–141. 265

[20] LIAO, W., ZHANG, W., ZHU, Z., AND JI, Q. A real-time human stress monitoring system using dynamic bayesian network. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)-workshops* (2005), IEEE, pp. 70–70. 4, 202, 232, 246, 258

[21] LIU, J. C., GOETZ, J., SEN, S., AND TEWARI, A. Learning from others without sacrificing privacy: Simulation comparing centralized and federated machine learning on mobile health data. *JMIR mHealth and uHealth 9*, 3 (2021), e23728. 44, 270

[22] MARINE, A., RUOTSALAINEN, J. H., SERRA, C., AND VERBEEK, J. H. Preventing occupational stress in healthcare workers. *Cochrane Database of Systematic Reviews*, 4 (2006). 246, 258

[23] MCMAHAN, B., MOORE, E., RAMAGE, D., HAMPSON, S., AND Y ARCAS, B. A. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* (2017), PMLR, pp. 1273–1282. 264, 265

[24] NILSSON, A., SMITH, S., ULM, G., GUSTAVSSON, E., AND JIRSTRAND, M. A performance evaluation of federated learning algorithms. In *Proceedings of the second workshop on distributed infrastructures for deep learning* (2018), pp. 1–8. 44, 270

[25] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M., AND DUCHESNAY, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12* (2011), 2825–2830. 204, 221, 235, 262

[26] PICKERING, T. G. Mental stress as a causal factor in the development of hypertension and cardiovascular disease. *Current hypertension reports 3*, 3 (2001), 249–254. 202, 218, 232, 246, 258

[27] Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., et al. The future of digital health with federated learning. *NPJ digital medicine 3*, 1 (2020), 1–7. 265

[28] Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., and Van Laerhoven, K. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 20th ACM international conference on multimodal interaction* (2018), pp. 400–408. 5, 12, 13, 26, 202, 203, 211, 218, 219, 228, 233, 247, 258, 259

[29] Siirtola, P. Continuous stress detection using the sensors of commercial smartwatch. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (2019), pp. 1198–1201. 5, 12, 13, 202, 211, 218, 228, 247, 258

[30] Sozinov, K., Vlassov, V., and Girdzijauskas, S. Human activity recognition using federated learning. In *2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom)* (2018), IEEE, pp. 1103–1111. 44, 45, 271

[31] Spagnolli, A., Guardigli, E., Orso, V., Varotto, A., and Gamberini, L. Measuring user acceptance of wearable symbiotic devices: validation study across application scenarios. In *International Workshop on Symbiotic Interaction* (2015), Springer, pp. 87–98. 4, 202, 232, 233, 246, 258

[32] Surden, H. Machine learning and law. *Wash. L. Rev. 89* (2014), 87. 44, 270

[33] Tsiga, E., Panagopoulou, E., and Montgomery, A. Examining the link between burnout and medical error: A checklist approach. *Burnout Research 6* (2017), 1–8. 80, 202, 218, 232, 246, 258

[34] Wakeling, A. *Stress and anxiety at work: personal or cultural?* ACAS, 2019. 245, 257

[35] Wang, Y., Chen, R., and Zhang, L. Reliability and validity of generalized anxiety scale-7 in inpatients in chinese general hospital. *J Clin Psychiatr 28* (2018), 168–71. 202, 232, 246, 258

[36] Weinberg, A., and Creed, F. Stress and psychiatric disorder in healthcare professionals and hospital staff. *the Lancet 355*, 9203 (2000), 533–537. 246, 258

[37] Welp, A., Meier, L. L., and Manser, T. Emotional exhaustion and workload predict clinician-rated and objective patient safety. *Frontiers in psychology 5* (2015), 1573. 80, 202, 218, 232, 246, 258

[38] Wemm, S. E., and Wulfert, E. Effects of acute stress on decision making. *Applied psychophysiology and biofeedback 42*, 1 (2017), 1–12. 2, 11, 40, 41, 80, 119, 144, 153, 162, 187, 202, 246, 258

[39] Zhang, Y., Haghdan, M., and Xu, K. S. Unsupervised motion artifact detection in wrist-measured electrodermal activity data. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers* (2017), pp. 54–57. 247, 260

[40] Zubair, M., Yoon, C., Kim, H., Kim, J., and Kim, J. Smart wearable band for stress detection. In *2015 5th International Conference on IT Convergence and Security (ICITCS)* (2015), IEEE, pp. 1–4. 262

ISBN 978-82-326-7270-7 (printed ver.)
ISBN 978-82-326-7269-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU
Norwegian University of
Science and Technology