REPUBLIC OF TÜRKİYE
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

GRADUATE SCHOOL
ELECTRICAL AND ELECTRONIC ENGINEERING DEPARTMENT

ROBOT DOCTORS: RECOGNITIONS, ASSESSMENTS, AND DIALOG-BASED INTERVENTIONS

FATMA GÖNGÖR

Ph.D.

ADANA 2024

REPUBLIC OF TÜRKİYE
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

GRADUATE SCHOOL
ELECTRICAL AND ELECTRONIC ENGINEERING DEPARTMENT

ROBOT DOCTORS: RECOGNITIONS, ASSESSMENTS, AND DIALOG-BASED INTERVENTIONS

FATMA GÖNGÖR

Ph.D.

THESIS ADVISOR
Prof. Dr. ÖNDER TUTSOY

ADANA, 2024

DECLARATION OF CONFORMITIY

In this thesis study, which was prepared following the thesis writing rules of Adana Alparslan Türkeş Science and Technology University Institute of Graduate School, I declare that I provide all the information, documents, evaluations and results in accordance with scientific ethics and moral codes without resorting to any means or assistance that would be contrary to scientific ethics and traditions. I also declare that I refer to all of the articles I used in this study with appropriate references and accept all moral and legal consequences if a situation is found contrary to my statement regarding my work.

27/09/2024

[Signature]

Fatma GÖNGÖR

ÖZET

ROBOT DOKTORLAR: TANIMLAMALAR, DEĞERLENDİRMELER VE DİYALOG TABANLI MÜDAHALELER

Fatma GÖNGÖR
Doktora, Elektrik-Elektronik Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Önder TUTSOY
Eylül 2024, 210 sayfa
Sağlık, insan yaşamının temel bir bileşeni olarak yalnızca hastalığın yokluğunu değil, aynı zamanda optimal fiziksel ve zihinsel iyilik halinin sağlanmasını da içerir. Son yıllarda yaşanan küresel zorluklar, sağlık hizmetleri alanında köklü değişimlere ivme kazandırmış ve yenilikçi sağlık hizmeti sunum yöntemlerinin geliştirilmesi ve uygulanmasının kritik bir gereklilik olduğunu ortaya koymuştur. Bu değişen ortamda, insansı robotlar (IR'ler), sağlık izleme uygulamalarında devrim niteliğinde bir rol üstlenerek, vazgeçilmez araçlar haline gelmiştir. Bu tez, IR'lerin kritik sağlık parametrelerini otonom bir şekilde tanımlamasını, kapsamlı değerlendirmeler yapmasını ve diyalog tabanlı müdahaleler gerçekleştirmesini sağlayacak bir dizi sofistike algoritma geliştirmekte ve böylece bu robotları bireylerin günlük yaşamlarında kişisel sağlık hizmeti sunucuları olarak konumlandırmaktadır. Tezin ilk bölümü, yardımcı robotik alanındaki son gelişmeleri ve bunların sağlık hizmetlerine etkilerini ayrıntılı bir literatür incelemesi ile ele almaktadır. Devam eden bölümlerde, kalp atış hızı izleme algoritmasının tasarımı ve uygulanması detaylı bir şekilde incelenmekte, ardından kişiselleştirilmiş sıcaklık tahmini algoritmasının derinlemesine bir analizi sunulmaktadır. COVID-19 pandemisinin yol açtığı sosyal izolasyon ve karantinanın psikolojik etkilerine yanıt olarak, bir sonraki bölümde karantina altındaki hastalar için özel olarak geliştirilmiş bir konuşma bazlı duygu analizi algoritması sunulmaktadır. Bunu takiben, pandeminin hafifletilmesine yönelik olarak geliştirilen yüz maskesi takma kontrolü algoritması ele alınmaktadır. Tez, IR'lerin geriatrik hastalar için fiziksel egzersizlerin kolaylaştırılmasındaki rolünü inceleyen bir vaka çalışması ile sonlanmaktadır. Geliştirilen algoritmalar, bu tez kapsamında özellikle NAO, Alpha Mini ve Alpha 1Pro insansı robotlarda uygulanmış olmakla birlikte, tüm diğer IR'lere de uyarlanabilir niteliktedir. Uygulanan her bir algoritmanın simülasyon ve deney sonuçları, yardımcı IR'lerin sağlık hizmetlerindeki kritik rolünü vurgulamakta ve sağlık müdahalelerinin ilerletilmesinde sağladıkları olumlu etki ve etkinlikleri gözler önüne sermektedir.

Anahtar Kelimeler: Diyalog Tabanlı Müdahaleler, Nabız Takibi, Robot Doktorlar, Sağlık Hizmeti, Vücut Sıcaklık Tahmini.

ABSTRACT

ROBOT DOCTORS: RECOGNITIONS, ASSESSMENTS, AND DIALOG-BASED INTERVENTIONS

Fatma GÖNGÖR

Ph.D., Department of Electrical and Electronic Engineering

Supervisor: Prof. Dr. Önder TUTSOY

September 2024, 210 pages

Health represents a fundamental aspect of human existence, encompassing not only the absence of disease but also the achievement of optimal physical and mental well-being. Recent global challenges have expedited transformative changes within the health-care domain, emphasizing the critical need for the development and implementation of innovative health-care delivery methodologies. In this evolving landscape, humanoid robots (HRs) have emerged as crucial entities in revolutionizing health monitoring practices. This thesis advances a series of sophisticated algorithms designed to empower HRs to autonomously recognize critical health parameters, conduct comprehensive assessments, and implement dialog-based interventions, thereby positioning them as personal health-care providers in individuals' daily routines. The initial section offers an extensive literature review, exploring the latest advancements in assistive robotics and their health-care implications. The subsequent sections provide a detailed exploration of the design and implementation of the heart rate monitoring algorithm, followed by an in-depth analysis of the personalized temperature estimation algorithm. In response to the psychological impacts of social isolation and quarantine induced by the COVID-19 pandemic, the following section presents a speech-based emotion analysis algorithm specifically designed for patients in quarantine. This is followed by the development of a facial mask-wearing control and regulation algorithm aimed at mitigating the effects of the pandemic. Finally, the thesis culminates in a case study that examines the application of HRs in facilitating physical exercise for geriatric patients. While the developed algorithms are expressly implemented on NAO, Alpha Mini and Alpha 1Pro HRs within the scope of this thesis, but its applicability extends to encompass all other HRs. The simulation and experimental results of each implemented algorithm highlight the significant role of assistive HRs in health-care, emphasizing their positive impact and effectiveness in advancing health interventions.

Keywords: Dialogue-Based Interventions, Health-Care, Heart Rate Monitoring, Robot Doctors, Temperature Estimation.

Dedication

to my beloved family…

ACKNOWLEDGEMENTS

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Professor Önder TUTSOY, for his unwavering support throughout my research. His enthusiasm, dedication, patience, and understanding have deeply inspired and encouraged me. His excellent guidance, strong assistance, and insightful critiques have been invaluable in shaping this thesis. I am truly fortunate to have had him as my supervisor. Dear Professor TUTSOY, your guidance has been profoundly instrumental in my growth both academically and personally throughout my MSc and Ph.D. studies. Thank you for everything! I could not have imagined a better supervisor and mentor during my Ph.D. journey.

Secondly, I would like to convey my profound gratitude to my current and former Thesis Monitoring Committee Members, Assoc. Prof. Dr. Ali İNAN and Assist. Prof. Dr. Kemal BALIKÇI. Their exceptional kindness and patience have been deeply appreciated throughout our engagements. Their profound expertise and unwavering support have been pivotal to the success of my research.

I would also like to thank my other Advisory Committee Members, Assoc. Prof. Dr. Fatih ABUT, Assoc. Prof. Dr. Gökay DİŞKEN, and Assist. Prof. Dr. Çağatay CEBECİ for their valuable contributions. I am also grateful for their role in making my defence a memorable experience, and for their insightful comments and constructive suggestions.

Finally, I must convey my deepest gratitude to my family for their unwavering support and continuous encouragement throughout my years of study and the entire process of researching and writing this thesis. This achievement would not have been possible without their steadfast presence and encouragement.

This thesis was supported by the Turkish Council of Higher Education (YÖK 100/2000 Doctorate Scholarship) and The Scientific and Research Council of Turkiye (TÜBİTAK 2211-C Doctorate Scholarship).

PUBLICATIONS

The key contributions of this Ph.D. thesis have been published or submitted to various leading robotics journals and conferences.

Journals:

1. Gongor, F., & Tutsoy, O. (2024). Doctor Robots: Design and Implementation of a Personalized Temperature Estimation Algorithm Utilizing a Swarm Intelligence-Based Hierarchical Vision Transformer and Multi-Modal Fusion Architectures. *International Journal of Social Robotics*, submitted.(Q1).
2. Gongor, F., & Tutsoy, O. (2024). On the Remarkable Advancement of Assistive Robotics in Human-Robot Interaction-Based Health-Care Applications: An Exploratory Overview of the Literature. *International Journal of Human–Computer Interaction*, 1-41 (Q1).
3. Gongor, F., & Tutsoy, O. (2022). Doctor Robots: Design and Implementation of a Heart Rate Estimation Algorithm. *International Journal of Social Robotics*, 14(6), 1435-1461 (Q1).

Conferences:

1. Gongor, F., & Tutsoy, O. (2024). A Review of Recent Advancements in Deep Machine Learning, Artificial Intelligence, Object Detection, and Human-Robot Interactions Approaches for Assistive Robotics. In 3. International Paris Congress on Applied Sciences (pp. 111-138).
2. Gongor, F., & Tutsoy, O. (2024). Physical Exercise of Geriatric Patients with Assistive Health-Care Humanoid Robots Utilizing Social Dialogue and Machine Learning Algorithms. In 3. International Paris Congress on Applied Sciences (pp. 378-403).
3. Gongor, F., & Tutsoy, O. (2022). Social companion robots: A speech-based emotion analysis algorithm for quarantined patients. In 4th International Congress on Engineering Sciences and Multidisciplinary Approaches (pp. 275-302).
4. Gongor, F., & Tutsoy, O. (2021). Social distance measurement algorithm and its application to the humanoid robot to halt the pandemic diseases spread. In *Proceedings of the 8th International Congress on Engineering, Architecture and Design* (pp. 254-274).
5. Gongor, F., & Tutsoy, O. (2021). Facial mask-wearing control and regulation by the humanoid robots during the COVID-19 pandemic outbreak. In 8th International Congress on Engineering, Architecture and Design (pp. 303-326).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

LIST OF ABBREVIATIONS

AI	: Artificial Intelligence
BA	: Backpropagation Algorithm
BDT-FSM	: Bayesian Decision-Tree guided Finite State Machine
DBI	: Dialogue-Based Intervention
FD	: Face Detection
FFT	: Fast Fourier Transform
GUI	: Graphical User Interface
GA	: Genetic Algorithm
HVT	: Hierarchical Vision Transformer
HRE	: Heart Rate Estimation
HHI	: Human-Human Interaction
HRI	: Human-Robot Interaction
HR	: Humanoid Robot
HMRT	: Hierarchical Multi Resolution Based Tracking
ICA	: Independent Component Analysis
KLT	: Kanade-Lucas-Tomasi
LSTM	: Long Short-Term Memory
MFCC	: Mel-Frequency-Cepstral Coefficient
M-GWO	: Modified Grey Wolf Optimizer
NN	: Neural Network
RNN	: Recurrent Neural Network
SI	: Swarm Intelligence
TE	: Temperature Estimation
TSMA	: Tripled Self-and-Mutual Attention
VJA	: Viola-Jones Algorithm
WHO	: World Health Organization

1. INTRODUCTION

As personalized health-care advances, the integration of humanoid robots (HRs) promises to substantially enhance early detection and intervention capabilities, marking a significant advancement in the pursuit of comprehensive and proactive health management. Although the HRs are mechanical systems, it is possible to convert them into social-assistive and even into the doctor robots by equipping them with vital health-check-up capabilities such as such as instant heart rate and temperature estimation from human faces. In addition, enhancing the HRs with various personalized Human-Human Interaction (HHI) properties such as being able to extract human emotions and analyse the body movements during the interaction are necessary. These functionalities are particularly critical in health-care environments such as hospitals, nursing homes, and home-care settings, where accurate and timely monitoring is crucial for the early detection of illnesses, infection control, and ongoing health assessment.

According to the World Health Organization (WHO), the early detection and management of critical health parameters are crucial for preventing the spread of infectious diseases, lowering global mortality rates, and enhancing patient recovery outcomes. In this context, the integration of advanced artificial intelligence (AI) algorithms into HRs facilitates precise, non-invasive monitoring, reducing the need for physical contact and providing a significant advantage in infectious disease management. These AI-driven algorithms enable HRs to deliver highly personalized and accurate assessments of vital health indicators, thereby enhancing their effectiveness and reliability in addressing individual healthcare needs. For example, if an HR detects an elevated heart rate in someone who is sensitive, frightened, or anxious, it can calmly reassure the individual and encourage them to contact appropriate healthcare services.

Likewise, if a patient displays an abnormal facial temperature indicative of postoperative complications, an HR can promptly initiate the required interventions such as calming strategies, administering antibiotics, adjusting treatment plans, and alerting healthcare professionals or emergency services for further evaluation and care. This proactive approach not only supports early diagnosis but also ensures timely medical intervention, potentially reducing the severity of health conditions, minimizing the risk of complications, and lowering healthcare costs associated with extended treatments. Recognizing the efficacy of HRs, this thesis introduces a set of advanced algorithms designed to empower HRs to autonomously detect critical health parameters, perform thorough assessments, and execute dialog-based interventions, thereby positioning them as personal health-care providers within individuals' daily lives.

1.1. Thesis Objectives

- The distinctive objective of this thesis is to initiate the development of a social interactive personal HR which can act as a personal doctor by continuously measuring and monitoring the critical physiological indicators such as heart rate and temperatures of the humans, performing necessary assessments, and executing dialog-based interventions in their daily lives.
- The ultimate goal is to transform the HRs in a form of personal/social entity where everyone can benefit from them both in domestic and professional settings.

1.2. Thesis Contributions

1) The first contribution of this thesis is to provide a thorough literature review that examines recent advancements in assistive robotics and their implications for health-care. This includes:

- **Comprehensive coverage:** This paper offers a number of studies pertaining to each concept, method, and application of HRI in the realm of health-care.
- **In-depth analysis:** The paper emphasizes the theoretical underpinnings, strengths, and significant challenges and related recommendations of these fields in a highly exhaustive manner.
- **Guiding resource:** It provides an opportunity for scholars and researchers actively engaged in the pertinent field to obtain comprehensive additional insights, serving as a guiding resource for their academic endeavours.

2) The second contribution is to provide a detailed exploration of the design and implementation of the heart rate monitoring algorithm. This includes:

- **Advanced motion tracking:** The Hierarchical Multi-Resolution Based Tracking technique enhances HRs' ability to track small and large head movements, ensuring precise heart rate monitoring in dynamic environments.
- **Improved signal accuracy:** The adverse effects of tracked head movements are mitigated through the application of the adaptive Independent Component Analysis

(ICA) technique, thereby enhancing the accuracy and reliability of heart rate measurements.

- **False detection reduction:** False detections occurring due to sudden fluctuations in the heart rate is eliminated with the constructed rule-based algorithm.
- **Novel social robotics application:** Even though the laptop/phone camera and video-based heart rate estimations have been extensively studied in the literature, the heart rate estimation algorithm, for the best of author's knowledge, has not been considered yet in the social robotics literature and there are no studies applied to HRs.
- **Real-world applicability:** The algorithm's effectiveness makes it suitable for real-time heart rate monitoring in unpredictable user environments, increasing HRs' practicality in health-care.
- **Broader health-care applications:** The developed algorithm allows HRs to perform continuous, non-invasive health monitoring, making them valuable in health-care settings for geriatric or chronically ill patients.

3) The third contribution is to present an in-depth analysis of the personalized temperature estimation algorithm. This encompasses:

- **Enhanced accuracy and adaptability with personalization:** A new personalized temperature estimation algorithm utilizing a swarm intelligence-based hierarchical vision transformer architecture with shifted-window mechanism and multi-modal fusion architecture is developed.
- **Improved user engagement:** A new custom graphical user interface and a Bayesian decision-tree guided finite state machines (BDT-FSM)-based dialogue model is developed for providing user-friendly engagements.
- **Pioneering social robotics integration:** Even though the laptop/phone camera and video-based temperature estimations from infrared thermal and depth cameras have been extensively studied in the literature, the temperature estimation algorithm from a general RGB cameras, for the best of authors' knowledge, has not been considered yet in the social-interactive health-care robotics literature and there are no studies applied to HRs.

- **Scalability and cost-efficiency:** The use of RGB cameras over more expensive alternatives makes the system scalable and cost-effective, facilitating broader adoption in health-care and robotic assistance.

4) The fourth contribution is to introduce a speech-based emotion analysis algorithm for patients in quarantine. This outlines:

- **Advanced emotion recognition:** The algorithm significantly enhances emotion detection through robust feature extraction and sophisticated temporal sequence modelling, thereby providing greater accuracy and depth in understanding emotional states.
- **Insightful performance evaluation:** The comprehensive analysis of this methodology yields valuable insights into its efficacy, highlighting its potential to advance interactive systems and emotional engagement across multiple applications.
- **Enhanced patient interaction:** By providing real-time, nuanced emotional feedback, the algorithm can improve patient interaction, offering more personalized and empathetic communication, which is crucial in managing the emotional well-being of quarantined individuals.
- **Integration with health monitoring systems:** The algorithm's ability to integrate with existing health monitoring systems enhances the overall capability of quarantine management by adding an emotional dimension to patient assessments, facilitating a more holistic approach to care.

5) The fifth contribution involves the development of a facial mask-wearing control algorithm aimed at mitigating the impact of pandemics. This includes:

- **Real-time monitoring:** The algorithm enables continuous and real-time supervision, allowing HRs to immediately detect improper mask usage and prompt corrective actions.
- **Increased compliance:** Regular reminders can improve adherence to safety protocols, reducing the risk of health hazards in public or clinical settings.
- **Reduced human effort:** Automating this process with HRs minimizes the need for constant human oversight, freeing up health-care or supervisory staff for more critical tasks.

- **Consistent enforcement:** HRs provide consistent and unbiased reminders, ensuring that mask-wearing guidelines are followed uniformly without human error or fatigue.
- **Adaptability:** The system can be easily adapted to different environments, such as hospitals, public spaces, or workplaces, ensuring compliance across various settings.
- **Enhanced safety:** By promoting mask adherence, HRs can help limit the spread of airborne diseases, contributing to overall public health and safety during pandemics or flu seasons.

6) Finally, the last contribution is to provide a case study that examines the application of HRs in facilitating physical exercise for geriatric patients. This outlines:

- **Enhanced personalization in physical therapy:** By integrating social dialogue and machine learning-based solutions, it enhances the capabilities of HRs to function as personal physical therapy doctor for geriatric individuals by facilitating periodic physical exercises.
- **Consistent monitoring and feedback:** HRs provide real-time monitoring and feedback on patients' exercise performance, enabling timely adjustments and continuous improvement in therapy.
- **Transformation into personalized and social entities:** The HRs evolved into personalized and social entities capable of positively impacting the lives of geriatric individuals at every stage of life.
- **Reduction in caregiver burden:** The automation of exercise routines helps alleviate the workload on caregivers, allowing them to focus on other critical aspects of patient care.
- **Cost-Effectiveness:** By reducing the need for frequent in-person therapy sessions and improving patient outcomes, HRs present a cost-effective solution for physical therapy in geriatric-care.

1.3. Thesis Outline

This thesis is organized as follows:

In chapter 2, a three-stage exploratory overview of the literature on the remarkable advancement of assistive robotics in HRI-based health-care applications are given. The first stage initiates an assessment of assistive robotics spanning historical epochs from ancient to modern times. Following this, the second stage comprehensively explores assistive robotics investigations in the realm of HRI-based health-care with its four sub-fields including rehabilitation, geriatric-care, paediatric-care, and nursing. Finally, the third stage entails a thorough analysis of the common challenges encountered in these pertinent investigations and provides a set of recommendations. This comprehensive chapter not only provides an abundance of studies for each concept, method, and application in HRI, but it also presents their theoretical foundations, strengths, gaps, critical challenges, and recommendations.

In chapter 3, a detailed exploration of the design and implementation of the heart rate monitoring algorithm and its application to the NAO HR is presented. The 9-stages cover the recognition of the face with the Viola-Jones algorithm, determination of the facial regions with the geometric-based facial distance measurement technique, extraction of the forehead and cheek regions, tracking of these facial regions with the Hierarchical Multi Resolution algorithm, decomposition of the facial regions in the Red-Green-Blue (RGB) colour channels, averaging and normalization of the RGB colour data, elimination of the artefacts with the Adaptive ICA technique, calculating the power spectrum of the data with the Fast Fourier Transform (FFT) technique, and finally determining the peaks inside the threshold reflecting the human heart rate boundaries.

In chapter 4, an in-depth analysis of a new four-stage personalized temperature estimation algorithm and its application to HRs is exhibited. The first stage utilizes the Viola-Jones algorithm for accurate face detection across RGB and its complementary LAB, and mapped thermal image modalities. The second stage applies a modified grey wolf optimizer-based hierarchical vision transformer to extract relevant thermal features from these modalities. These features are then fused through a multi-modal architecture to estimate facial temperature precisely. Finally, the algorithm is integrated into Alpha Mini HR with a custom graphical user interface and a Bayesian decision-tree guided dialogue model.

In chapter 5, a 3-stage speech-based emotion analysis algorithm and its application to NAO HR is introduced. At the initial stage of the algorithm, the most informative speech features are extracted with Mel-Frequency Cepstral Coefficients (MFCC) technique. Then, in the next stage, each extracted speech features are classified to obtain related emotional statements by using a Long Short-Term Memory - based Recurrent Neural Network (LSTM-RNN) machine learning algorithm. Finally, in the last stage, engagement is performed with the help of the created decision tree structure.

In chapter 6, a 4-stage facial mask-wearing algorithm that detects the absence of a facial mask or improper facial mask-wearing situations and its implementation to NAO HR is presented. At the initial stage of the algorithm, HR detects face and facial mask with the Viola-Jones algorithm. Then, in the next stage, crucial face and facial mask distance measurements are determined with the geometric based distance measurement technique. Then, these measured face and facial mask models are matched each other to label and evaluate the facial mask-wearing situations. Finally, in the last stage, the labelled and evaluated results are classified with a 3-layer NN algorithm.

In chapter 7, a four-stage physical exercise algorithm and its implementation to Alpha Mini and Alpha 1Pro HRs is presented. The initial stage of the proposed algorithm involves the Viola-Jones algorithm to accurately detect and continuously track the human body within the robot perceptual framework. Following this, geometric-based distance measurements are employed to characterize human actions, thereby enabling the algorithm to comprehend physical movements relevant to targeted exercises for geriatric individuals. Then, in the third stage, the characterized human actions are learned through an LSTM-RNN combined with a genetic algorithm. Finally, in the last stage, a structured and user-friendly engagement is executed by employing a Finite State Machine (FSM) dialogue model.

Finally, **in chapter 8,** the conclusions drawn from the research presented in this thesis are summarized, and potential future directions for further investigation are discussed.

2. LITERATURE REVIEW OF THE RECENT AND RELATED APPROACHES

With the incorporation of advanced AI algorithms, assistive robotic systems have become essential tools in a range of Human-Robot Interaction (HRI)-based healthcare applications, including nursing, geriatric, paediatric care, and beyond. In this framework, these robots are specifically designed to enhance interaction and collaboration with humans, offering support in daily physical activities, participating in mental therapy sessions, monitoring vital signs, and providing emotional assistance. This is achieved through their autonomous perception, advanced reasoning, and precise execution capabilities.

The utilization of such assistive robots has attracted scholarly attention, leading to numerous literature reviews on the subject. For instance, in the field of geriatric-care, Shibata and Wada (2011) explored the impact of robotic therapy on dementia patients and the broader geriatric population. Bemelmans et al. (2012) assessed the effectiveness of assistive robots across seventeen studies, highlighting positive outcomes but emphasizing the need for further validation. Mordoch et al. (2013) reviewed studies on robots for dementia patients, highlighting a lack of randomized clinical trials and limited exploration of robots' social aspects. In a further study, Kachouie et al. (2014) investigated socially assistive robots in geriatric care, highlighting the need to understand caregiver expectations, multi-modal interactions, and personalized care. Kulpa et al. (2021) reviewed research on socially assistive robots in geriatric psychiatry, identifying gaps and needs in outcome measures. Ghafurian et al. (2021) reviewed fifty-three articles from sixteen countries on dementia-assistive robots, noting a focus on therapy and engagement while highlighting underrepresented areas such as health guidance. Naseer et al. (2023a) examined telepresence robots in healthcare during the pandemic, evaluating technical and ethical challenges, and advocating for advanced, adaptive systems to enhance healthcare delivery and address limitations identified during COVID-19.

In the realm of paediatric-care, Littler et al. (2021) systematically assessed the impact of assistive robots on children's anxiety and distress during hospital visits, finding that interactions with socially assistive robots generally reduced anxiety and distress. However, the review noted that the studies had low-quality evidence, highlighting the need for improved research designs to better understand the effectiveness of assistive robots in paediatric health interventions. Additionally, Van Bindsbergen et al. (2022) investigated healthcare providers' perspectives on assistive robots in paediatric oncology, analysing opinions through an online questionnaire across a global sample and examining variations based on providers' background characteristics. In a subsequent systematic review, Triantafyllidis et al. (2023) evaluated assistive robotics in paediatric care, considering factors such as intervention targets, robot features, participant age, and study duration. The review found positive outcomes in technology acceptance, feasibility, enjoyment, engagement, therapeutic goal achievement, pain reduction, and mental health improvements. However, it noted that many studies involved only single-session interactions and limited participant numbers.

In the field of nursing, Christoforou et al. (2020) provided a comprehensive overview of the integration of assistive robotics in nursing, highlighting both the benefits and challenges of implementing these technologies in clinical practice, while also considering end-user perspectives. In another work, Nieto Agraz et al. (2022) examined and classified 133 robotic systems relevant to nursing, evaluating their technical features and applications, and identifying gaps for future research. In a recent study, Ohneberg et al. (2023) reviewed forty-seven publications, categorizing them into studies, technical articles, and opinion pieces, and identified thirty-one distinct assistive robotic systems.

Lastly, Akalın and Loutfi (2021) performed a comprehensive review of reinforcement learning applications in assistive robotics, focusing on its role in health-care environments. They analysed real-world interactions involving social physical robots and categorized reinforcement learning approaches by method and reward mechanism design. The review identified three major themes: interactive reinforcement learning, intrinsically motivated methods, and task performance-driven methods. The review highlighted the benefits and challenges of reinforcement learning in social robotics, evaluated current assessment methods, and proposed solutions to practical issues, including areas that have been underexplored in the literature.

Despite significant advancements in AI-driven assistive robotics for healthcare, notable gaps persist. Although numerous studies have highlighted the potential benefits of these robots in specific care settings, there remains a consistent lack of high-quality, large-scale randomized clinical trials to validate their long-term efficacy and safety. Reviews frequently call for more rigorous trials to evaluate the clinical impact of robotic therapy and address limitations in current research designs. Additionally, while some studies report positive outcomes in paediatric and geriatric care, they also highlight issues related to study quality, sample size, and short-term interactions, indicating a need for more extensive research. There is also a significant gap in addressing the integration of assistive robots into existing health-care systems and their alignment with caregiver expectations. Moreover, practical applications and real-world effectiveness of advanced AI methodologies are insufficiently explored.

To address these limitations, this section presents a three-stage exploratory review. The first stage evaluates the historical development of assistive robotics from ancient to modern times. The second stage explores current research across three sub-fields: geriatric-care, paediatric-care, and nursing. The third stage analyses common challenges and offers recommendations. The primary goal of this section is to detail advancements in assistive robotics, provide a comprehensive overview of current and future developments, and guide the creation of new algorithms to address identified research gaps.

2.1. Historical Evaluation of Assistive Robotics

From the earliest cranes in Mesopotamia to the rockets of today, a number of the machines produced have become indispensable prostheses and have been integrated into human environments to expand and enhance their limited capabilities. Notably, around 450 BCE, Greek engineers pioneered early automated devices using compressed air, steam, and hydraulics. Similarly, between 10 and 70 AD, the mathematician and engineer İsmâil bin er-Rezzâz al-Cezerî gained acclaim for his innovative creations, including automated gates and service waiters (Romdhane & Zeghloul, 2009). His ground-breaking work inspired subsequent inventors, notably Leonardo da Vinci, who conceptualized an autonomous knight equipped with tambours and gears around 1495 (Pasek, 2014).

Along with the industrial revolution, the development of machines was accelerated with the emergence of electricity and steam power. The burgeoning consumerism era compelled manufacturers to innovate and devise advanced, automated tools to expedite production processes and replace human labour in perilous environments. In pursuit of this goal, George Moore initiated a pioneering project in 1893 to create an early prototype of a humanoid machine, termed the "Steam Man," which was powered by a steam engine capable of walking at speeds of 9 miles per hour and transporting light loads (Nonami et al., 2014). Concurrently, Henry Ford, an American industrialist, transformed the manufacturing landscape by introducing the first movable assembly line in 1913, a breakthrough that revolutionized automobile mass production. This innovation dramatically decreased the time required to construct an automobile from 12 hours to just 1 hour and 33 minutes by substituting manual labour (Graves, 2013).

The swift evolution of machines necessitated the development of a more nuanced term to encapsulate their nature. Consequently, in 1920, the Czechoslovakian writer Karel Čapek (1971) coined the term "robot" to describe these advanced machines. The term is derived from the Czech word "robota," meaning "forced labourer" or "servant," and was popularized in his science fiction play, Rossum's Universal Robots (RUR). The play depicts a factory where thousands of efficient and tireless synthetic humanoid machines are utilized for labour, thereby not only establishing the modern nomenclature for robots but also intensifying societal fears about the potential displacement of human workers by such machines. Following this, in 1928, Alan Reffell and William Richard exhibited an early version of a voice-controlled humanoid robot named "ERIC" at the annual Society of Model Engineers exhibition in London (Newton, 2018). Subsequently, in 1938, the Westinghouse Company commenced the development of a second humanoid robot, "ELECTRO," which was presented and demonstrated at the prestigious World's Fair (Sharkey & Sharkey, 2008).

Long after the term "robot" had been established, the concept of "robotics" was introduced by science fiction author Isaac Asimov in his 1941 short story "Runaround." Asimov coined this term to describe the field encompassing robots and their interactions with humans. Unlike Karel Čapek, Asimov adopted a more optimistic view of the role of robots in society, depicting them as valuable assistants and formulating the "Three Laws of Robotics" as a framework for safe and effective HRIs. The First Law asserts that a robot must not harm a human or allow a human to come to harm through inaction. The Second Law mandates that a robot must obey human commands, provided these do not conflict with the First Law. The Third Law requires a robot to protect its own existence as long as this does not violate the First or Second Laws (Clarke, 2020). These principles not only contested the notion of robots as mere emotionless machines programmed for singular tasks with binary moral tendencies but also laid the groundwork for ethical considerations and heralded a new era in robotics research.

Following Isaac Asimov's introduction of the "Three Laws of Robotics," the concept of HRI gained significant traction in scholarly literature, prompting a multitude of researchers to explore its complexities and underscore its significance. For instance, Goodrich and Schultz (2007) define HRI as "a field of study dedicated to understanding, modelling, and evaluating advanced robotic systems involving human-robot interactions through direct, safe, and effective communication." This definition emphasizes the goal of creating a synergistic relationship between robotic capabilities and human ingenuity. Humans contribute their experience, task management knowledge, perceptual abilities, understanding of control strategies, adaptability, and learning, while robots offer precision, speed, and power. Unlike traditional interaction models, HRI is characterized by sophisticated, dynamic, and intelligent frameworks that facilitate autonomous operation in unpredictable real-world environments.

In the following decades, researchers have focused on understanding and analysing human behaviour by investigating brain structure and integrating these insights into robots to enhance HRI. Notably, in 1943, Warren McCulloch and Walter Pitts introduced the concept of neural networks, employing specific modelling frameworks to explore brain function and simulate intelligent behaviour through interconnected circuits (Piccinini, 2004). Subsequently, in 1948, William Gray Walter produced the first electric-powered, tortoise-shaped robots, named Elmer and Elsie (Bhaumik, 2018). These robots utilized the modelled neural networks to navigate back to a charging station, manoeuvre around obstacles, and move toward a light source, demonstrating rudimentary autonomous behaviours.

The remarkable architecture of the human brain spurred significant advancements in the field of HRI in the following years. A pivotal moment in this evolution occurred in 1950 when Alan Turing, a British computer scientist and mathematician, conducted an experiment to assess whether a robot could exhibit independent thought and autonomous cognition (Muggleton, 2014). This evaluation of a machine's intellectual capabilities and its potential resemblance to human intelligence marked a seminal achievement in the field of Artificial Intelligence (AI), profoundly shaping research perspectives in this area. A notable example of this influence is the development of the "Stochastic Neural Analog Reinforcement Computer (SNARC)" by Marvin Minsky and Dean Edmonds in 1951. The SNARC, the first neurocomputer, was designed to solve maze problems, representing a significant milestone in the evolution of computational models for intelligent behaviour (Hillis et al., 2007).

Following this innovation, John McCarthy coined the term "Artificial Intelligence" (AI) in 1955 to describe the field dedicated to advancing robotic intelligence through scientific and engineering methodologies (Rajaraman, 2014). Subsequently, in 1963, the development of "Rancho," the first AI-based and computer-controlled robotic arm, marked a significant advancement. Building on this conceptual foundation, the first AI-based computer-controlled robotic arm, named "Rancho," was developed in 1963 to assist disabled patients at Rancho Los Amigos Hospital in Downey, California (Moran, 2007). Subsequently, Rancho was acquired by Stanford University, where it contributed to "robotics and prosthetics research" and inspired the development of a new category of human-centred robots known as "Cobots. Just three years later, the Stanford Research Institute introduced "Shakey,"the first AI-based mobile robot capable of reasoning about its environment and executing tasks through its own planning, eliminating the need for step-by-step instructions.

During the 1970s, significant strides were witnessed in humanoid robot development. Notably, Ichiro Kato's creation of the "WAP-1" marked a milestone as the first fully functional bipedal humanoid robot, possessing an intelligence level comparable to that of an 18-month-old human infant (Bezerra & Zampieri, 2004). This achievement represented a major advancement in humanoid robotics. Building on this progress, Honda unveiled the "ASIMO" in 2000, the world's first advanced multifunctional humanoid robot, designed to assist the bedridden and disabled while featuring a friendly and approachable appearance (Shigemi, 2018).

In the subsequent decades, the development of assistive robots advanced to include capabilities for understanding human emotions and engaging in meaningful interactions. An exemplary case is "Kismet," an emotionally intelligent and interactive robot created by Breazeal in 1998 as part of an affective computing experiment. This innovation aimed to develop robots capable of participating in significant social exchanges with humans. Similarly, in 1999, Sony introduced the AIBO robot dog, marking the debut of the first interactive robotic pet in the consumer market (Fujita, 2001). A landmark achievement occurred in 2017 when "Sophia," a social robot developed by Hanson Robotics, was granted Saudi Arabian citizenship, becoming the first robot to receive legal personhood (Retto, 2017). This milestone highlights the capability of robots to adapt to various environments and exhibit human-like behaviour to meet their objectives.

Throughout the subsequent years, the development of assistive robots has continued to aid doctors in performing health-care operations. A prime example of this is the inception of the "HelpMate" service robot created by Engelberger (1988). Taking a step further, Hanson Robotics (2021) has manufactured a nurse robot named 'Grace' that provides emotional support to isolated elderly patients. Through the use of patient data, "Grace" is able to engage in therapeutic interactions and cognitive stimulation.

Figure 2.1. Historical timeline of the notable assistive robotic inventions from antiquity to modern times.

As illustrated in Figure 2.1, recent advancements have transformed assistive robots into highly sophisticated machines with advanced engineering and mechanical systems. Unlike traditional welding robots in the automotive industry, which repetitively perform the same task regardless of their environment, these modern robots possess the capability to perceive and interpret their surroundings dynamically. They can navigate freely within human society, interact meaningfully with their environment, and respond to direct instructions from users rather than relying solely on pre-programmed behaviours. A critical factor contributing to their enhanced functionality is the implementation of advanced AI algorithms, which enable these robots to understand, interpret, and address user needs with greater accuracy. Consequently, these robots have achieved a level of autonomy, allowing them to learn independently and adapt to diverse situations.

2.2. Existing HRI-Based Health-Care Applications

Following the historical overview of assistive robots, this sub-section explores the diverse aspects, significant advancements, and key findings related to HRI in health-care applications. It addresses three specific sub-fields: geriatric care, paediatric care, and nursing, each focusing on how robotic technologies are applied to improve care and outcomes in these areas.

2.2.1. Assistive Robots for Geriatric-Care

With the ongoing increase in the senior population, there is a growing demand for robotic solutions aimed at improving quality of life, enhancing relaxation and motivation, and providing both physical and psychological support. In response to these demands, the literature has highlighted the development of both animal-like and human-like robots designed to replicate the companionship typically offered through animal and human-assisted therapies in geriatric care (Helm et al., 2022).

For instance, Pineau et al. (2003) developed a mobile robotic assistant named Pearl to aid geriatric patients with mild cognitive and physical impairments and to assist nurses in their tasks. The system comprises three modules: an automated reminder system for scheduling notifications, a human-tracking system using particle filters, and a high-level controller utilizing probabilistic reasoning (POMDP) for decision-making under uncertainty. Pearl was tested in a nursing home and demonstrated effective communication with patients, including appointment reminders and daily information. However, the study revealed the need for manual adjustments to accommodate individual variations in walking speed, volume sensitivity, and hearing. This underscores the necessity of incorporating advanced techniques to address the diverse needs of geriatric users.

In a distinct study, Tamura et al. (2004) investigated the therapeutic effects of the AIBO robotic companion on 13 patients with severe dementia in a geriatric-care facility, comparing it with a traditional toy dog. Both interventions led to increased patient activity, with individuals engaging in conversation and commenting on their experiences. AIBO was particularly effective in eliciting immediate emotional responses, while the toy dog stimulated the recall of past memories. However, the study did not include dementia scores for the patients and lacked a thorough categorization of the robot and toy’s features.

In another work, Tapus et al. (2008) developed a socially assistive robot aimed at engaging post-stroke geriatric patients during rehabilitation exercises without physical contact. The study explored the impact of the robot's personality on therapy outcomes and its correlation with user extroversion or introversion, using the Psychoticism, Extraversion, and Neuroticism (PEN) model. Initial experiments mapped user personality traits to various robot therapy styles, while subsequent experiments employed a reinforcement learning algorithm to dynamically adjust the robot's style. Despite these efforts, achieving optimal interaction between the robot and user proved challenging due to factors such as limited adaptation steps relative to the parameter space and distracting personality variations. The study found that varying one parameter at a time would improve the measurement of its impact on user performance. Future research should focus on integrating empathy into robot behaviour control and utilizing physiological data to better understand and respond to user needs with a larger participant pool.

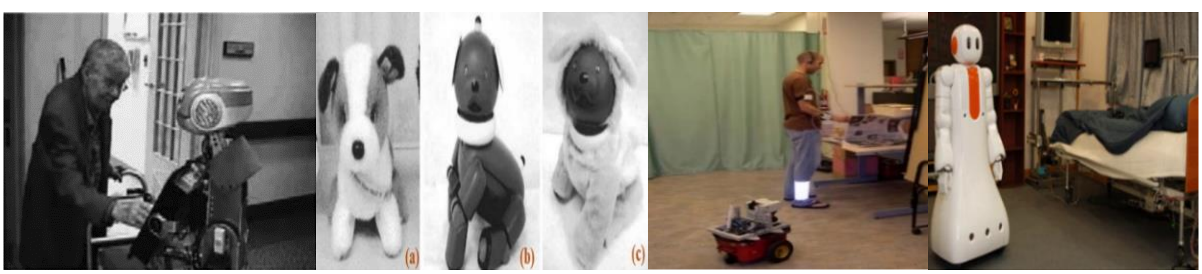


Figure 2.2. Left to right: (1) Nursebot Pearl (Pineau et al., 2003); (2) AIBO robot (Tamura et al., 2004); (3) Experimental setup of (Tapus et al., 2008); (4) Steward HR (Park et al., 2008).

In a further work, Gadde et al. (2011) developed an interactive robot coach, RoboPhilo, to assist older adults with physician-prescribed exercise routines. The robot guided patients through exercises by mirroring their movements and providing verbal instructions, but required the patient to stand at a specific distance from the camera and in well-lit conditions. Despite its effective performance, these limitations affected the robot's usability.

Similarly, McColl and Nejat (2013) investigated the use of the robot Brian 2.1 as a social companion for older adults during mealtimes. Brian 2.1 featured communication abilities, such as nonverbal gestures, facial expressions, and natural speech, to engage users by greeting them, offering jokes, and encouraging social interaction. The robot operated using a finite-state acceptor to determine appropriate behaviours based on mealtime events. However, challenges arose with the robot’s load cells, which failed to detect small amounts of food, impairing its ability to prompt eating behaviours. Additionally, the haar classifiers used for detecting eye and head orientation were insufficiently accurate. Future research should address these issues by enhancing sensor sensitivity, combining visual and weight detection methods, and conducting extended trials to assess long-term user engagement and safety in care settings.

During a further investigation, Wu et al. (2014) examined the acceptance of assistive robots among older adults, focusing on experiences over a one-month period. The study involved six older adults with mild cognitive impairment (MCI) and five healthy older adults, who interacted with a robot weekly in a simulated living lab. Methods included surveys, semi-structured interviews, usability assessments, and focus groups. Results indicated that while both groups learned to use the robot, those with MCI took longer to complete tasks after a break. Despite similar acceptance ratings from both groups, there was a general lack of motivation and negative perceptions towards the robot. Challenges to robot acceptance included resistance to technology, feelings of isolation, and ethical concerns. The study's limitations included a small, predominantly female sample from France and limited exposure to the robot. Future research should involve larger, more diverse samples and real-world settings to better understand robot acceptance.

In another work, Bemelmans et al. (2015) explored the use of the Paro seal robot in psychogeriatric care, involving 71 geriatric patients and their caregivers. The study aimed to evaluate the short-term effects of Paro on patients' mental and psychosocial well-being and on caregivers' daily tasks. The primary outcome was assessed using the Individually Prioritized Problems Assessment (IPPA) score, while a mood scale served as a secondary measure to corroborate caregivers' reports of patients' emotional states. The sample size was determined using the Wilcoxon signed-rank test. Results showed no significant differences between genders or dementia stages, suggesting preliminary effects that need further validation. The study's small subgroup sizes warrant additional research to confirm these findings and better understand the implications of robotic care interventions across different patient groups.

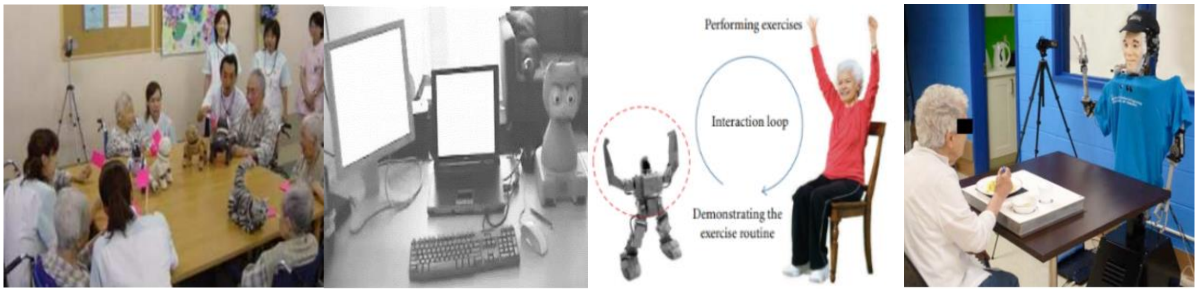


Figure 2.3. Left to right: (1) Robots interacting with geriatric patients (Naganuma et al., 2008); (2) Experimental setup of (Looije et al., 2010); (3) Personal trainer robot with geriatric patient (Gadde et al., 2011); (4) Brian 2.1 robot (McColl and Nejat, 2013).

In another study, Fischinger et al. (2016) developed the Hobbit PT1 robot to assist elderly individuals with six specific tasks. The robot was tested with 49 geriatric participants in a laboratory setting, using a multimodal interface that included speech recognition, text-to-speech, gesture recognition, and a touch-based graphical interface. Although slightly over half of the participants believed the robot could be a long-term addition to their homes, nearly half doubted its practicality. The robot's remote operation by researchers highlighted its lack of autonomy. Future evaluations should be conducted in participants' actual homes to test autonomous navigation, object recognition, and other functionalities in real-world settings, along with the multimodal communication interface.

Likewise, Görer et al. (2017) introduced a self-directed exercise instructor for seniors using the NAO robot to guide users through a two-stage exercise program. Initially, a therapist trains the robot to replicate exercises, providing verbal or physical demonstrations. In the subsequent stage, the NAO robot independently instructs the exercises, monitoring and evaluating user performance with pre-programmed routines. The robot starts each session with a greeting, introduces exercises verbally, replicates movements, and evaluates performance upon session completion. Despite its innovative approach, the system has limitations: feedback accuracy is constrained by a fixed score range, the program is tailored for healthy older adults and may not suit those with specific needs, and monitoring seated exercises is hampered by inaccurate skeletal data.

In a distinct study, Gerłowska et al. (2018) evaluated the initial version of the RAMCIP robotic assistant in a controlled setting to gauge its societal impact and user desirability. The study involved structuring tasks around scenarios such as social interaction, hazard avoidance, and medication adherence. Although the small sample size could affect the results, the trials showed no significant differences between groups in terms of the assistant's appropriateness, functionality, and societal impact.

Figure 2.4. Left to right: (1) Kompaï robot (Wu et al., 2014); (2) Paro robot (Bemelmans et al., 2015); (3) Hobbit PT1 robot (Fischinger et al., 2016); (4) Nao HR (Görer et al., 2017).

In another work, Martinez-Martin et al. (2019) assessed the Pharos 2.0 system for monitoring and assisting elderly individuals with physical exercise in their homes. The system employed the Pepper robot to capture initial poses through image processing, utilizing the OpenPose library to extract and reshape human skeletons. These were then analysed using a Convolutional Neural Network (CNN) with a 50-layer residual network to identify user poses. Exercises were ranked using the Glicko2 system, and adjustments were made based on comparisons among exercises. The system achieved high accuracy with training, testing, and validation scores of 93.98%, 91.37%, and 90.70%, respectively. However, challenges included inconsistencies in recognizing exercises performed on different sides, a lack of feedback based on error severity, and limited participant interest in the proposed exercises, which were already part of their routine. Future research should explore alternative algorithms for skeleton estimation to improve accuracy and address these limitations.

Similarly, Sato et al. (2020) investigated the use of the Pepper robot for facilitating rehabilitation activities in geriatric patients with schizophrenia and dementia at geriatric health facilities. The study involved programming Pepper to lead 40-minute Care Prevention Gymnastic Exercises (CPGE) sessions, combining body-brain gymnastics and allowing patients to exercise at their own pace. Using a case-oriented approach, including observation notes, field texts, and video recordings, the researchers observed both proactive and reactive interactions. While Pepper effectively encouraged physical activity, limitations emerged, such as patients struggling to understand or hear instructions and Pepper's inability to engage in deep conversations or provide nonverbal communication. The study highlighted the need for advancements in robotic technology to enhance communication, physical activity, and motivation for elderly patients with mental disorders. It suggests incorporating activities like singing to improve interaction and emphasizes the importance of integrating human care in robotic-assisted rehabilitation.

In a further work, Luperto et al. (2022) presented the MoveCare system, which integrates socially assistive robots into ambient assisted living environments to support frail older individuals at home. Their extensive 300-week pilot study assessed the system's acceptability and feasibility, demonstrating its effectiveness in promoting engagement with the robots. However, the robots slightly reduced overall acceptability. The study identified technical, organizational, and logistical challenges that affect system performance. Future research should focus on improving interactions, addressing challenges related to long-term robot deployment, and correcting inconsistencies in integrated systems. Expanding sample sizes and conducting longer experiments are crucial for understanding the evolution of HRIs and evaluating the benefits of adaptive robot behaviour over extended periods.

Figure 2.5. Left to right: (1) RAMCIP robot (Gerłowska et al., 2018); (2) and (3) Pepper HR (Martinez-Martin et al., 2019; Sato et al., 2020); (4) Sanbot Elf robot (Di Napoli et al., 2023).

More recently, Di Napoli et al. (2023) reported on the UPA4SAR project, which developed an assistive system tailored for home care of elderly patients with mild cognitive impairments. The system employs personalized and adaptable tasks based on user characteristics and environmental conditions. Field tests with 7 patients over 118 days showed the system's reliability, positive patient responses, and high acceptability. Despite these successes, challenges persist in updating user profiles, incorporating entertainment preferences, addressing technology familiarity, and integrating natural language processing. Future research should explore cost-effective natural language solutions and incorporate caregiver feedback to enhance the effectiveness of assistive care.

Lastly, Shaik et al. (2023) introduced an innovative multi-agent deep reinforcement learning framework for patient monitoring in assistive robotics. This framework employs specialized learning agents to monitor physiological features such as heart rate, respiration, and temperature from related measurement devices. Operating within a standard health-care environment, these agents learn patient behaviour patterns and autonomously alert medical staff when emergencies are estimated. The framework was tested with real-world data from the PPG-DaLiA and WESAD datasets, demonstrating superior performance compared to baseline models in monitoring vital signs. The study also included hyper-parameter optimization to enhance each agent's learning process. Despite its advancements in handling complex environments and making real-time decisions, challenges related to data scale and predicting future vital signs suggest directions for future research.

2.2.1.1. Discussions

Despite significant advancements in assistive robots for geriatric care, a number of critical gaps remain. One major issue is the need for enhanced adaptability of these robots to the diverse and complex health conditions of older adults. Improvements in customizing robotic assistance to individual needs and preferences are essential. To fully assess the sustainability of these robots' effects, research should focus on their long-term impact and determine the optimal duration for their use. Future studies should also replicate findings across various settings and user groups, including individuals with different levels of dementia, to improve result generalizability. Furthermore, relying solely on self-report questionnaires for measuring robot acceptance may yield subjective responses; therefore, future research should combine self-reported data with objective measures for a more accurate assessment. Addressing these gaps is vital for maximizing the benefits of assistive robots and fostering a patient-centred approach to aging-related healthcare challenges.

2.2.2. Assistive Robots for Paediatric-Care

The emergence of assistive robots in paediatric care has opened up new possibilities, offering significant benefits to both children and their caregivers. This sub-section explores two main themes: first, how these robots can provide emotional support to children dealing with illnesses and enhance their well-being during hospital stays, and second, how they can offer distraction during medical procedures.

2.2.2.1. Assistive Robots for Emotional and Moral Encouragement

A number of studies in the literature have extensively examined how emotional support and moral encouragement can be provided to paediatric patients undergoing hospitalization while they cope with illnesses. In one such study, Lu et al. (2011) developed the interactive robotic companion, MediRobbi, to reduce children's anxiety during hospitalization by guiding them through medical procedures and teaching about allergies using a Stop-and-Go game. The robot featured five interaction modes: activation via a touch sensor, navigation along a track with light sensors, detection of red spots on a map, reactivation at red spots, and providing visual and auditory feedback. The study utilized paired interviews with children aged 3 to 7 and their parents, along with field observations in a simulated hospital environment, to evaluate the robot's effectiveness. MediRobbi successfully provided a personalized playmate, enhancing comfort and trust for paediatric patients. However, challenges such as the limitations of Lego Mindstorms and the robot's size and weight impacting usability were noted. Future designs should prioritize optimizing the robot's dimensions and weight to enhance user interaction.

In another study, Csala et al. (2012) explored the use of the NAO robot with three children undergoing leukemia treatment, each with different ages and cognitive levels. In collaboration with medical professionals and a psychologist, they designed four brief activities to make daily routines—such as eating, taking medication, waking up, and bathing—more engaging. While NAO was recognized as a potentially valuable tool in medical practice, challenges included coordinating its use with the hospital's schedule and addressing children's confusion about the robot's purpose. To address these issues, NAO's role within the hospital needs to be clearly defined, and personalization of interactions is crucial to maintain children's interest. Instead of relying on passive activities like television, incorporating a variety of puzzles and math problems using infrared-remote control technology can enhance engagement and provide a more stimulating experience for the children.

Similarly, Okita (2013) investigated the use of robotic animals as companions to alleviate emotional distress and pain in paediatric patients and their parents. The study, conducted with eighteen children aged six to sixteen and their parents at a children's hospital in Palo Alto, California, compared two conditions: interaction with Paro alone and interaction with Paro alongside their parents. Results showed that patients who interacted with both their parents and Paro experienced a greater reduction in negative emotions and pain compared to those who only interacted with Paro. However, the study's limited data did not allow for a detailed analysis of the factors influencing these outcomes. The study also lacked male paediatric patients and their caregivers. Future research should broaden the scope to include various hospital wards and diverse patient demographics, including male paediatric patients and adults. Additionally, exploring the effects of different types of robots, such as humanoid or other animal robots, could provide further insights into their impact.

Likewise, Jeong et al. (2015) conducted a comprehensive study to evaluate the effectiveness of the Huggable robot compared to a virtual character and a plush teddy bear in reducing stress and anxiety among children aged 3 to 10, hospitalized for severe chronic pain. The study involved observing patient and parent behaviour during interventions, with participants wearing sensors to measure Electro Dermal Activity (EDA) and being video-recorded to assess facial expressions. The results indicated that children interacting with the robot were more engaged and perceived the robot as a social peer. However, further research with a larger sample and more quantitative measures is needed to confirm these findings. Future work should include developing algorithms for automatic emotional state detection in hospitalized children using video-coded facial expressions and EDA.

Figure 2.6. Left to right: (1) MediRobbi robot (Lu et al., 2011); (2) Nao HR (Csala et al., 2012); (3) Experimental setup of (Okita, 2013); (4) Plush, virtual, and huggable bear robot (Jeong et al., 2015).

In another study, Looije et al. (2016) created a framework to facilitate educational and enjoyable interactions between children and robots during hospital visits. The program included three one-hour sessions where 17 children, aged six to ten, participated in activities such as a quiz, a sorting game, and video watching. The diabetic children particularly responded positively to the robot's dependence on them, enhancing their willingness to socialize. Nonetheless, the significant emotional bond created between the children and the robot raised ethical dilemmas, particularly during discussions of sensitive topics. Technical difficulties also led to inconsistent sessions, complicating data analysis. Future work should focus on creating an independent prototype for home use, addressing challenges in speech recognition, dialog management, personalization, and self-management. Additionally, the effectiveness and cost of implementing such a robot in care facilities need to be rigorously evaluated.

Likewise, Beraldo et al. (2019) investigated the use of Pepper and Sanbot Elf humanoid robots as non-pharmacological interventions to manage negative emotions and promote positive moods in hospitalized children undergoing sedation and analgesia. Both robots were controlled via Wizard of Oz mode, utilizing text-to-speech synthesis and gesture-based animations. Emotional assessments conducted before, during, and after the intervention revealed that the Sanbot Elf was more effective than the Pepper robot in reducing agitation and increasing serenity, likely due to its calming, whimsical appearance. The study highlighted challenges related to robot autonomy, particularly the limitations in current sensors and software for detecting children's low voices. Future studies should involve larger sample sizes and incorporate advanced data collection tools like cameras and EDA sensors to gather quantitative data and deepen the analysis of these initial findings.

In another work, Gamborino et al. (2019) developed an autonomous robotic system designed to provide emotional support to children through extended interactions. They employed an emotion estimation algorithm based on visual cues and used deep interactive reinforcement learning to help the robot understand the user's social tendencies. Testing with 15 elementary school-aged children showed that the robot could adapt to user preferences and generate positive responses. However, questions remain about the algorithm's accuracy in assessing social tendencies and the need for a larger sample size for statistically significant results. Future research should focus on expanding the participant pool for more robust quantitative analysis to validate the approach.

In a further one, Trost et al. (2020) investigated the effectiveness of an empathetic socially assistive robot in alleviating pain and fear during peripheral IV insertion in children. The randomized study compared interactions with a 3D printed humanoid robot featuring empathy or distraction functions to no robot intervention. Empathy was assessed via questionnaires completed by children and parents, while pain and fear were measured using self-report scales (FEAR and FACES) and video analysis (CHEOPS and FLACC). Results showed that the empathy group reported and exhibited the lowest levels of pain and fear, while the distraction group had the highest levels. Limitations include the study’s single-centre design, English-speaking sample, and small size, which may introduce selection bias. Future research should explore the mechanisms behind the empathy robot's effectiveness by incorporating biochemical, electroencephalographic, and neuroimaging data, and extend investigations to other medical procedures and settings to evaluate the broader impact of assistive robots.

Similarly, Smakman et al. (2021) explored the use of a social robot as a distraction to reduce stress and anxiety in children during blood draws. The robot, designed with input from health-care professionals, was tested in a randomized controlled trial in a children's hospital. Results showed that children interacting with the robot experienced significantly lower anxiety levels, particularly those in middle primary school classes. Parents also viewed the robot positively. However, the study's findings are limited by the specific cultural and hospital context, as well as privacy constraints that restricted the robot's camera and microphone use, limiting personalized responses. Neerincx et al. (2023) further investigated the role of emotional gesturing in social robots during group vaccinations, finding that such gestures improved children's engagement, reduced anxiety, and increased trust in the robot.

Figure 2.7. Left to right: Patient-robot interaction setups for child individuals (Looije et al., 2016; Beraldo et al., 2019; Gamborino et al., 2019; Trost et al., 2020).

Finally, this sub-section concludes by summarizing the discussion and transitions to a detailed literature review on assistive robots used for distraction during medical procedures. The subsequent section offers a thorough examination of their advancements, contributions, and key features in the context of research and development.

2.2.2.2. Assistive Robots for Distraction during Medical Procedures

Numerous in-depth studies have investigated the use of assistive robots to provide distraction during medical procedures. For instance, Beran et al. (2013) explored the effectiveness of a Nao robot as a diversion during vaccinations. The study involved 57 children, divided into two groups: one receiving standard care and the other receiving care with the robot, which used cognitive-behavioural strategies for interaction. The results, analysed through multivariate variance analysis, showed a significant reduction in pain and distress among children interacting with the robot.

Despite these findings, the study applied a uniform treatment plan, regardless of individual differences, which could be improved by customizing the robot's speech and actions to suit the child's age and preferences. The use of modern distractions like music, stories, and games could further enhance the robot's effectiveness. Additionally, the study's findings, conducted at a single location with informed nurses, may not be broadly generalizable. Future research should involve diverse locations and personnel to test the broader applicability of the results. Moreover, further investigation is needed to determine whether the robot's presence, actions, or specific distractions contribute to reduced distress and to evaluate the impact of the robot's appearance and behaviours.

Alemi et al. (2015) explored the use of the NAO humanoid robot as a psychotherapist to alleviate distress in young oncology patients. The study aimed to foster a friendship between the robot and children aged 7 to 12 diagnosed with cancer to reduce their pain and anxiety. Eleven children were randomly assigned to two groups: one received robot therapy over eight sessions, while the other received standard psychotherapy. The NAO robot played various roles and provided information and empathy to the children. Analysis using descriptive statistics and MANOVA demonstrated that robot therapy significantly reduced stress, depression, and anger. Challenges included ensuring participants could attend all sessions, which was difficult given their treatment schedules, and assembling a mid-sized group, as both parents and patients were hesitant to visit medical centers during their free time.

In a similar work, Henkemans et al. (2017) evaluated the effectiveness of a personalized robot for diabetes self-management education in children aged 7 to 12 with type 1 diabetes mellitus (T1DM). The study assessed participant enjoyment, engagement, and motivation through a diabetes quiz, with the robot's behaviour based on Self-Determination Theory (SDT) to foster autonomy, competence, and relatedness. After three sessions, children interacting with the personalized robot showed higher SDT scores, greater enjoyment, better quiz performance, and increased motivation for subsequent sessions compared to those with a neutral robot. Despite these positive outcomes, the study identified the need to adjust the timing of the robot's verbal responses and suggested incorporating varied activities to cater to different learning styles. Additionally, there was no definitive evidence regarding how long children remain interested in the robot, indicating the need for further investigation into this aspect.

In a further study, Rossi et al. (2020) examined the effectiveness of the NAO social robot in providing distraction during medical procedures for 139 children over two months. The robot, equipped with interactive behaviours and emotional cues, was compared for its impact on emotional versus non-emotional distraction, as well as no distraction. Grounded in Cognitive Behavioural Therapy (CBT) principles, the study aimed to mitigate mild depression and anxiety by addressing negative thoughts and attitudes. The findings revealed that emotional distraction significantly reduced fear and anxiety, enhancing overall happiness. The robot's emotional behaviours effectively captured and maintained children's attention. However, the study faced challenges such as collecting data on critical anxiety levels, dealing with false positives in anxiety measurement, and ensuring proper synchronization among team members. These issues highlighted the need for real-time adjustments in distraction techniques to better address varying levels of distress.

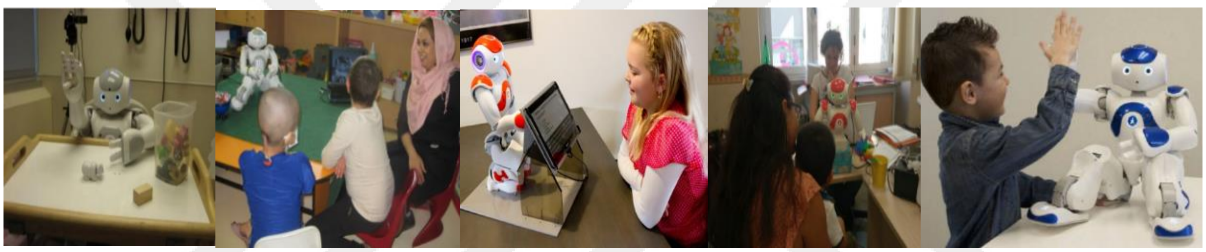


Figure 2.8. Left to right: Nao HR engagement setups for child individuals (Beran et al., 2013; Alemi et al., 2015; Henkemans et al., 2017; Rossi et al., 2020; Nishat et al., 2023).

In another work, Shenoy et al. (2021) developed adaptive humanoid robots to alleviate pain in children using non-pharmacological methods. They employed Attentional Convolutional Networks and Deep Time-Delay Markov models to analyse facial expressions and voice quality, which were then used to train an Extreme Gradient Boosting Classifier for predicting emotion labels. These labels guided a reinforcement learning model to select optimal actions for distraction and pain relief. Despite positive results, the study faced challenges such as detection errors due to environmental factors like lighting and background noise, which affected facial expression recognition and the robot's ability to hear the child. To address these issues, creating a diverse interaction database for reinforcement learning and selecting modular actions suitable for various age groups are recommended. Emphasizing reactive actions is also suggested to reduce delays between emotion detection and robot responses.

Likewise, Tanaka et al. (2022) conducted a non-randomized controlled trial to evaluate the effect of the Aibo robot on children's distress during vaccination. The study involved 57 children aged 3 to 12 receiving the Japanese encephalitis vaccine, who were assigned to either the Aibo intervention group or a control group using a stuffed dog. Pain levels were self-reported by the children using the Faces Pain Rating Scale, and behaviour was assessed by an observer. Statistical analysis showed that the Aibo group experienced significantly lower pain and distress and calmed more quickly post-vaccination compared to the control group. Despite these positive results, the study faced limitations including potential observer bias, proxy responses from caregivers, and unaddressed confounding factors such as developmental disabilities and vaccination history. Further research is needed to optimize the timing of Aibo interventions and to explore its broader application in preventing medical trauma across different procedures and settings.

More recently, Nishat et al. (2023) performed a qualitative needs assessment to understand the views of children and caregivers on an AI-enhanced socially assistive robot for procedural support during intravenous insertion (IVI). Data from semi-structured interviews and focus groups with 11 caregivers and 19 children across two Canadian paediatric emergency departments revealed three main themes: clinical experience, acceptance and concerns about the robot, and features that promote child engagement. While most participants had a favourable view of the robot technology, they expressed concerns about privacy, technical reliability, and potential intrusiveness. Recommendations included improving robot movement and language customization to better engage with children and meet developmental needs. The study acknowledges that user feedback might not fully reflect real-world implementation challenges.

2.2.2.3. Discussions

While initial findings suggest that assistive robots may help alleviate anxiety and stress in children, further investigation is required to fully understand their impact on mental health outcomes. Existing research primarily focuses on free interaction, structured interaction, and robotic distraction, but there is limited evidence on how different types of robotic interaction affect mental health. Studying children poses unique challenges, such as their susceptibility to suggestion and the lack of standardized outcome measurements. Despite some studies employing qualitative assessments and validated measures, inconsistencies in control conditions and the absence of standardized outcomes hinder comparability and generalizability. Additionally, there is a lack of longitudinal studies assessing the long-term effects of social robot interventions. Future research should prioritize the development of comprehensive and longitudinal studies to establish robust evidence on the effectiveness of assistive robots in improving children's health and well-being. Finally, the current sub-section has been thoroughly concluded alongside the discussion. The upcoming sub-section will provide an in-depth literature review of nursing and doctor robots, detailing their specific features and contributions in a comprehensive manner.

2.2.3. Nursing and Doctor Robots

The increasing adoption of nursing and doctor robots in healthcare facilities is attributed to their numerous benefits for patients and healthcare professionals. These advanced robots are capable of performing various tasks, including medication administration, patient monitoring, and social interaction. Research into their effectiveness has generally been positive. Research generally supports their effectiveness.

For instance, Kang et al. (2005) addressed nurse shortages by developing "Clara," a socially adept robot designed for hospital environments. In a simulated hospital room, Clara demonstrated four key behaviours: patient location identification, engagement, breathing monitoring, and room departure. Evaluations showed Clara's control systems were effective, with an average satisfaction rating of 84.6%. Nevertheless, improvements are needed based on user feedback. Enhancements to Clara's voice recognition system are required to boost accuracy and minimize patient discomfort. Additionally, a more advanced collision avoidance system should be integrated to ensure safety in dynamic hospital settings. Expanding Clara's social behaviours and potentially modifying its physical appearance could further enhance patient engagement and motivation.

In other work, Onishi et al. (2007) demonstrated the RI-MAN robot's ability to lift a simulated human dummy using immersion-type dynamic simulation technology to emulate human movements and gather visual and tactile feedback. Despite its technological sophistication, RI-MAN had practical limitations: its mechanical structure constrained joint movement, motion accuracy, and payload capacity, restricting it to lifting dummies up to 18.5 kg and preventing safe lowering. Additionally, the robot required a controlled environment and lacked essential safety features for human interaction. To improve, future work should focus on optimizing the tactile sensor controller, increasing measurement resolution, and enhancing sensor noise tolerance.

To address the limitations of RI-MAN, Mukai et al. (2010) developed the RIBA robot, featuring human-like arms designed for tasks requiring a delicate touch. RIBA demonstrated its effectiveness by transferring individuals between a bed and a wheelchair, utilizing a tactile guidance technique that allowed caregivers to adjust the robot's movements via tactile sensors, thus minimizing patient stress. The robot also incorporates speech recognition for following verbal commands. The robot's strength was tested on its elbow-bending, shoulder-rotating, and waist-bending joints, demonstrating its capability to lift individuals up to 63 kg with a payload-to-weight ratio of 0.35. Future development should enhance payload capacity, ensure user comfort, and improve safety measures.

In a subsequent study, Iwata and Sugano (2009) developed the multifunctional nursing robot "TWENDY-ONE" to assist with kitchen tasks while ensuring safety during physical interactions. They based its ergonomic design on data from the Research Institute of Human Engineering for Quality Life (HQL) in Japan, identifying 44 tasks such as carrying, cooking, cleaning, and garment handling. Using graphical computing techniques, they simulated the robot's movements to optimize joint specifications under various conditions. Despite its advanced features, TWENDY-ONE faced challenges in reliability, maintenance, and battery life. Experiments highlighted the need for human-centred cooperative control methods that adapt to individual motion properties and leverage the robot's passivity for improved performance.

Figure 2.9. Left to right: (1) Clara assisting with spirometry exercise (Kang et al., 2005); (2) Human care task with RI-MAN (Onishi et al., 2007); (3) RIBA lifting a subject (Mukai et al., 2010); (4) TWENDY-ONE aiding in breakfast preparation (Iwata and Sugano, 2009); (5) Cody positioned by the patient bed (King et al., 2010).

Likewise, King et al. (2010) developed and implemented a behaviour for the robot Cody to autonomously clean specific body areas using the Equilibrium Point Control (EPC) method. This approach enabled precise arm guidance by adjusting the Cartesian-space Equilibrium Point (CEP) over time. An interface allowed the operator to select the cleaning area, but Cody's effectiveness was limited to surfaces aligned with gravity. Further studies could include robotic bed baths, automated body part identification, and comparisons of user preference and performance between autonomous and tele-operational modes. Expanding research to include the cleaning of different body parts and other hygiene tasks is also suggested.

In another work, Ding et al. (2012) introduced the RIBA nursing-care assistant robot designed to lift and transfer individuals between a bed and a wheelchair. The robot's model incorporates factors such as posture and lifting status for both the robot and the human, estimating the forces and torques applied to human joints to assess comfort. The model's accuracy was validated by comparing comfort level calculations with EMG data and questionnaire results. The analysis confirmed the model's effectiveness in determining comfort during lifting. Future developments could involve integrating optimized weight parameters with the model to enhance the robot's ability to automatically identify the most efficient lifting motions.

Similarly, Ding et al. (2014) developed the RoNA robotic nursing assistant for lifting tasks in healthcare settings. RoNA features intelligent sensors for intuitive guidance by nurses and offers three control modes: Cartesian, posture-adjusting, and trajectory-following, allowing for precise navigation along predetermined paths. Its control system includes three layers—application, real-time, and hardware—enabling user inputs via joystick and force sensors, high-level control computations, and 3D virtual environment generation for trajectory planning. The gravity compensation mode supports effortless lifting from beds or wheelchairs by integrating velocity and force controllers. Future enhancements may include advanced algorithms for improved interaction and performance evaluation in hospital settings with user-centred design modifications.

In their investigation, Gombolay et al. (2018) investigated trust and reliance on robotic decision support systems in labour and delivery environments. They assessed how nurses and doctors responded to recommendations of varying quality from both robotic and computer-based systems. Recommendations from the robotic system, which used action-driven learning from expert demonstrations, achieved a high acceptance rate of 90%. Future research should focus on improving machine learning algorithms for semi-supervised data integration to reduce manual labelling while maintaining recommendation quality. Additionally, further studies are needed to explore the use of robotic decision support systems as training tools, leveraging expert knowledge to assist in training less experienced healthcare professionals.

In a different study, Abubakar et al. (2020) developed the ARNA service robot to assist nurses with physical tasks such as helping patients walk and sit in hospital settings. The robot features autonomous navigation, object manipulation, environmental sensing, and communication capabilities. Its usability is enhanced by the Directed Observer-Led Assistant (DOLA) software framework, which includes directive, observer, and lead assistant components, and utilizes neuro-adaptive and PID controllers for movement. The neuro-adaptive controller, based on a two-loop neural network algorithm with Lyapunov stability proofs, showed that ARNA's features are effective and user-friendly. The robot's adaptable design fits well with the DOLA framework, providing flexibility for various autonomy levels. Future research should focus on advanced modelling and longitudinal studies in real healthcare environments to further evaluate ARNA's practical utility.

Recently, Narayanan et al. (2022) developed an autonomous nursing robot for smart hospital environments. This robot can identify patient rooms, plan optimal medication delivery routes, and measure patient parameters using a Bluetooth Low-Energy (BLE) beacon system. It navigates by following wireless beacon signals and employs a fuzzy controller with ultrasonic sensors and a camera for obstacle detection and collision avoidance. Patient data is gathered via sensors and transmitted to the hospital's database through a BLE tag module. Testing in a private hospital revealed high accuracy in obstacle detection and avoidance. Future research should focus on AI-based obstacle avoidance, enhanced surface analysis for better navigation, and advanced multimedia interaction. Integrating AI with edge computing could further improve the system's practical effectiveness.

Likewise, Miyake et al. (2022) developed a humanoid robot to assist with physical exercise by utilizing Mediapipe-based skeleton recognition to track human poses. The robot uses its 7-DOF arm to align with the coordinates of the human right shoulder. While the robot successfully generated motion under controlled conditions, issues arose when participants' backs were partially visible or when the robot was positioned behind them. Challenges included the lack of automatic positional adjustments for the robot, noise in skeleton data due to participant movement, and difficulties in adjusting motion trajectories during arm reaching. Future improvements could include automatic position and head angle adjustments, interactive waypoint modifications based on human feedback, and enhanced motion planning to better accommodate participant movements.

In a further study, Zhang and Demiris (2022) developed a dressing pipeline using the Baxter robot to assist patients with putting on hospital gowns. This system involves grasping, unfolding, navigating around a bed, and dressing the patient. They tackled challenges in learning manipulation policies for garment handling and translating simulated object manipulation to real-world applications. The pipeline combines prehensile and non-prehensile actions for effective garment handling and approximates garment physics to bridge the simulation-reality gap. Although the method showed high success rates with medical manikins, pre-training policies were more effective with human demonstration data. Future work could aim to reduce demonstration data through self-supervised learning, develop models for various garment types, and test the system with real patients, especially those with physical impairments. Enhanced robot payload capacity and safety features will be necessary. Further research could explore real-time reasoning and scenarios such as dressing patients with bent elbows.

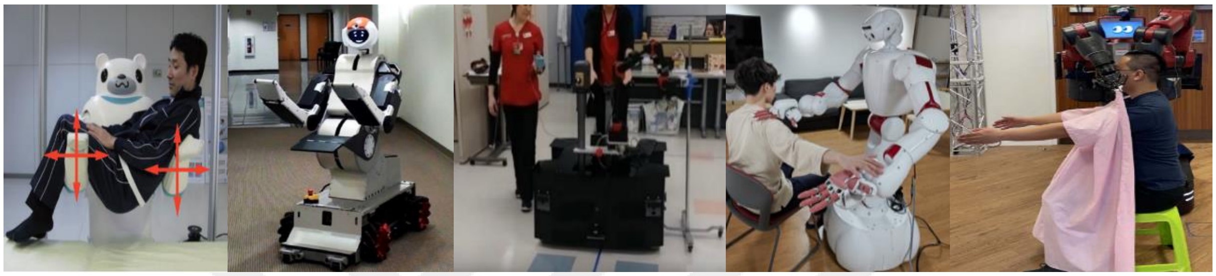


Figure 2.10. left to right: (1) Lifting experiment with RIBA (Ding et al., 2012); (2) RoNA prototype for patient lifting (Ding et al., 2014); (3) DOLA assisting subjects (Abubakar et al., 2020); (4) Physical exercises with Dry-AIREC (Miyake et al., 2022); (5) Robot dressing patient (Zhang and Demiris, 2023).

Furthermore, Jeffcock et al. (2023) proposed a robotic implementation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) to facilitate early delirium detection. The system employs Human-Robot Interaction (HRI) for interactive assessment features and a vision transformer-based deep learning model to predict Richmond Agitation Sedation Scale (RASS) levels from image sequences, utilizing thermal imaging for patient anonymity. A user study with 18 participants validated the system, achieving accuracies of 1.0 and 0.833 for the CAM-ICU's inattention and disorganized thinking features, respectively. The action recognition model attained a mean accuracy of 0.852 in classifying RASS levels, showcasing the potential for real-world hospital deployment.

In another work, Cen et al. (2023) introduced a medical service robot for cuffless blood pressure (BP) estimation using the OCViTnet model. This model features an omni-scale convolution subnet for single-period waveform characteristics, a vision Transformer subnet for multi-period waveform features, and a regressor that fuses phenotypic information with pulse wave features for BP regression. Their method excels in non-contact BP estimation from forehead imaging photoplethysmography and utilizes the radial artery pulse wave for BP estimation. Adhering to the British Hypertension Society standards, their approach achieved grade C/B performance for systolic/diastolic BP estimation.

More recently, Guo et al. (2024) introduced a transformer-based multi-modal fusion prototype for health monitoring, which integrates visual, audio, and sensor data to assess cognitive workload, biomechanical strain, and muscle fatigue in HRIs. Their method leverages the transformer's ability to capture complex sequential relationships, enabling precise health monitoring through effective data fusion and representation learning. Experimental results highlight the approach's efficacy in detecting intricate patterns and predicting health conditions with future potential for real-time medical-care.

2.2.3.1. Discussions

While assistive nursing and doctor robots show great potential for enhancing patient care, several critical gaps and limitations need addressing. A key issue is the lack of standardized metrics to evaluate these robots' effectiveness, which often relies on subjective measures, hindering definitive conclusions. Additionally, comparative studies are needed to identify the most beneficial robot features. Most current robots are designed with non-anthropomorphic appearances, which impacts user behaviour and empathic responses. More anthropomorphic designs may enhance user trust and attitudes. Research has largely focused on short-term outcomes, highlighting the need for longitudinal studies to assess long-term effects. Future research should improve language and visual algorithms to better adapt to patient needs. Ultimately, developing sophisticated assistive robotic systems that can adapt to individual personalities, recognize physical cues, respect cultural diversity, and adhere to ethical standards is essential for effective implementation in healthcare applications.

3. DESIGN AND IMPLEMENTATION OF A HEART RATE ESTIMATION ALGORITHM

Considering the prevalent challenges faced by assistive robotics in HRI-based health-care, this section introduces a 9-stage heart rate estimation algorithm, marking a first in social robotics literature. The algorithm begins with capturing images from the NAO HR's camera. Next, it employs the Viola–Jones algorithm for face detection and applies a geometric-based technique to measure facial distances. The forehead and cheek regions are then extracted and tracked using a Hierarchical Multi-Resolution based tracking algorithm. These tracked regions are decomposed into RGB colour channels, which are subsequently averaged and normalized. Motion artefacts are removed using an adaptive ICA technique. The FFT is applied to derive power spectrums, and heart rate is estimated by counting peaks within the power spectrum threshold.

3.1. Face Detection with Viola-Jones Algorithm

This sub-section provides a brief overview of the Viola-Jones face detection algorithm, which comprises four main steps: Integral image, haar-like features, Ada-Boost, and Cascade classifier, as illustrated in Figure 3.1.

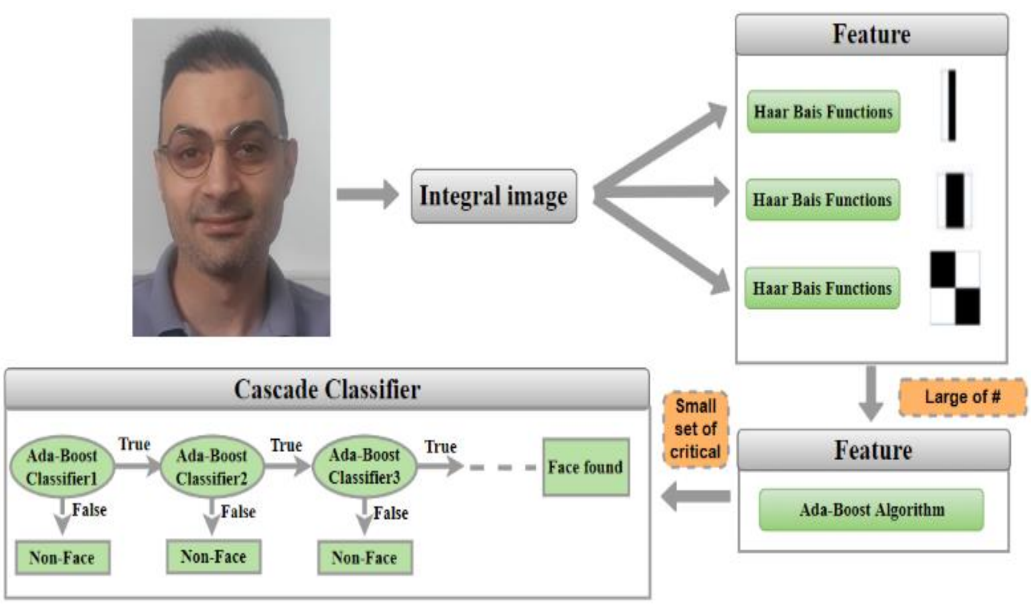


Figure 3.1. The Viola-Jones algorithm architecture.

The process begins by dividing the image into rectangles to accelerate feature extraction for face detection. Typically, two to four rectangles are used. After extracting features from these rectangles, their sums are calculated and normalized to determine a temporary labelling error. This error helps classify features as weak or strong candidates based on a threshold value. The Ada-Boost algorithm uses these summed features to label candidates accordingly. These labelled candidates are then compared with testing samples for each facial component. Candidates that pass each stage are forwarded to the next stage until the face is successfully detected. Upon successful detection, the candidates proceed to the facial distance measurement phase, as outlined in the following section.

3.2. Facial Distance Measurements

After detecting the face, it is mapped onto a coordinate frame to measure facial distances, with the centre of the face and other highlighted key regions as shown in the first two images on the left in Figure 3.2. Once the centres of each facial region are identified, 28 facial points are positioned within the corresponding areas, as illustrated in the last two images on the right in Figure 3.2.

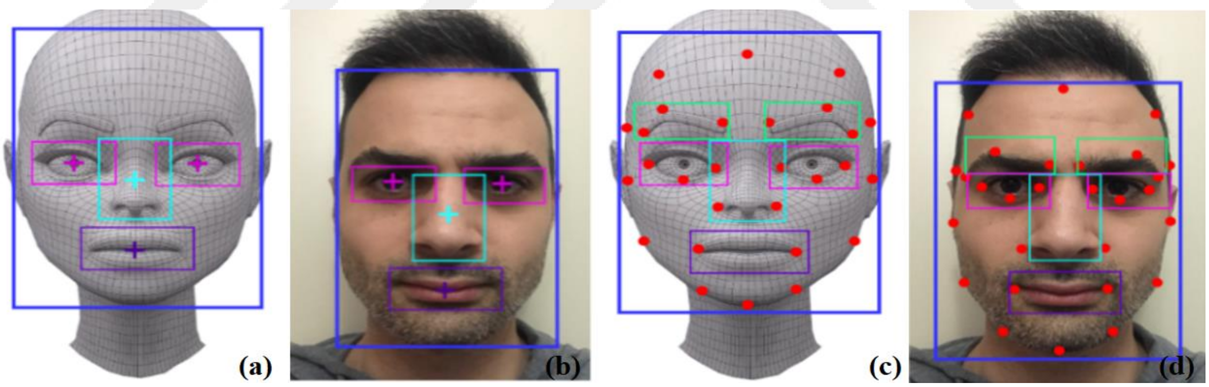


Figure 3.2. Rectangular representation of the face with the positioning of facial points within geometric frames.

Each facial distance measurement within the geometric frame is calculated using the Pythagorean Theorem, with each point assigned to its respective coordinates. For instance, to measure the distance between the eyes, a separate box is designated for each eye, as illustrated in Figure 3.3. As can be seen, the exterior eye shapes are represented using two parabolas respectively, one for the upper eye lid and one for the bottom eye lid. Herein, each eye is assigned by 6 points in x,y coordinates, starting at the left corner of the eye and then draw a clockwise shape around the remainder of the region. For the left eye; the terms p_{Eleft} and p_{Eright} are left and right eye corner points in the x and y coordinates. The measurements p_{Ue1} and p_{Ue2} represent the upper eye lid points, p_{Le1} and p_{Le2} represent the lower eye lid points.

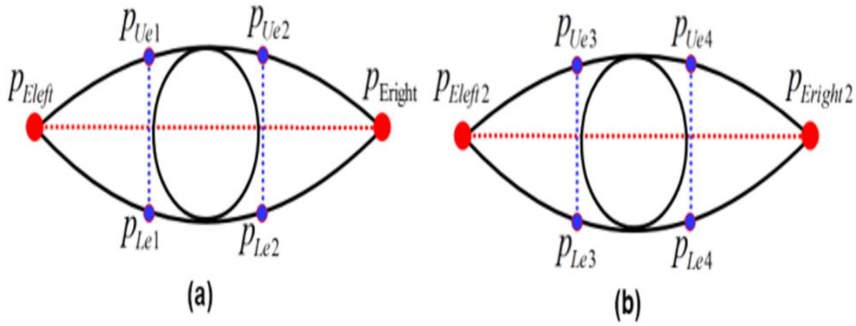


Figure 3.3. Eye measurement configurations: (a) Measurements for the left eye and (b) right eye**.**

For the right eye; p_{Eleft2} and $p_{Eright2}$ are left and right eye corner points in the x and y coordinates., p_{Ue3} and p_{Ue4} correspond to the upper eye lid points, p_{Le3} and p_{Le4} represent the lower eye lid points. These coordinates generate the roundness for each eyes formulated as

$$LE_{Roundness} = \frac{\|p_{Ue1} - p_{Le1}\| + \|p_{Ue2} - p_{Le2}\|}{2\|p_{Eleft} - p_{Eright}\|} \tag{3.1}$$

$$RE_{Roundness} = \frac{\|p_{Ue3} - p_{Le3}\| + \|p_{Ue4} - p_{Le4}\|}{2\|p_{Eleft2} - p_{Eright2}\|} \tag{3.2}$$

In Equations (3.1) and (3.2), the numerator represents the distance between the lateral eye points of each eye, while the denominator calculates the distance between the sagittal eye points, respectively.

Figure 3.4. Eyebrow measurement configurations: (a) Measurements for the left eyebrow and (b) right eyebrow.

In addition to eye distance calculations, measurements for the eyebrows, nose, mouth, and face shape must be determined for the forehead and cheek regions. After measuring the eye shape, eyebrow measurements are calculated, as shown in Figure 3.4. In this figure, p_{EBleft}, p_{EB1}, $p_{EBright}$,

$p_{EBleft2}$, p_{EB2}, and $p_{EBright2}$ measurements represent: the left corner of the outer left eyebrow, the midpoint of the left eyebrow, the right corner of the inner left eyebrow, the left corner of the outer right eyebrow, the midpoint of the right eyebrow, and the right corner of the inner right eyebrow, respectively.

To extract the lengths of the left and right eyebrows, the following distance measurements are determined respectively:

$$LEB_{Length} = \| p_{EBleft} - p_{EBright} \| \tag{3.3}$$

$$REB_{Length} = \| p_{EBleft2} - p_{EBright2} \| \tag{3.4}$$

To extract nose lengths, five points are placed in the x, y coordinates, as illustrated on the left side of Figure 3.5.

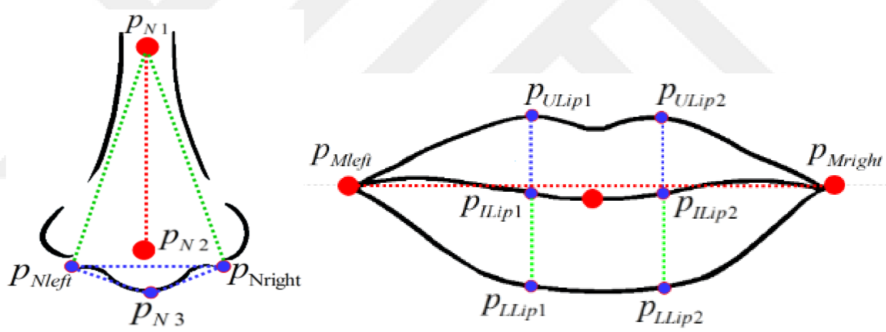


Figure 3.5. Measurement configurations for the nose and mouth.

In this context, P_{Nleft} and P_{Nright} represent the left and right corners of the nose along the x direction, respectively. Meanwhile, p_{N1} and p_{N2} represent the nose's distance measurements from the upper to the lower corner in the y direction. To determine the length, width, and shape of the nose, the following measurements are calculated:

$$N_{Vlength} = \| p_{N1} - p_{N2} \| \tag{3.5}$$

$$N_{Width} = \frac{\| p_{Nleft} - p_{N3} \| + \| p_{N3} - p_{Nright} \|}{\| p_{Nleft} - p_{Nright} \|} \tag{3.6}$$

$$N_{Shape} = \frac{\| p_{N1} - p_{Nleft} \| + \| p_{N1} - p_{Nright} \|}{N_{Width}} \tag{3.7}$$

To determine the coordinates for the mouth length and shape, the same procedure used for the nose is applied, as shown on the right side of Figure 3.5. The following measurements are then calculated to determine the length of the mouth and the ratios of the upper and lower lips:

$$M_{HLength} = \left\| p_{Mleft} - p_{Mright} \right\| \tag{3.8}$$

$$ULip_{Ratio} = \frac{\left\| p_{ULip1} - p_{ILip1} \right\| + \left\| p_{ULip2} - p_{ILip2} \right\|}{M_{HLength}} \tag{3.9}$$

$$LLip_{Ratio} = \frac{\left\| p_{ILip1} - p_{LLip1} \right\| + \left\| p_{ILip2} - p_{LLip2} \right\|}{M_{HLength}} \tag{3.10}$$

$$M_{Shape} = ULip_{Ratio} + LLip_{Ratio} \tag{3.11}$$

In these measurements, $M_{HLength}$ represents the horizontal length of the mouth, $ULip_{Ratio}$ and $LLip_{Ratio}$ denote the upper lip ratio and lower lip ratio, respectively. P_{Mleft} and indicate the left and right corners of the mouth in the x direction, respectively. p_{ULip1} and p_{ULip2} correspond to the upper lip points. p_{ILip1} and p_{ILip2} describe the inner lip points, and p_{LLip1} and p_{LLip2} represent the lower lip points. Finally, these assigned facial feature points are used to extract the face shape.

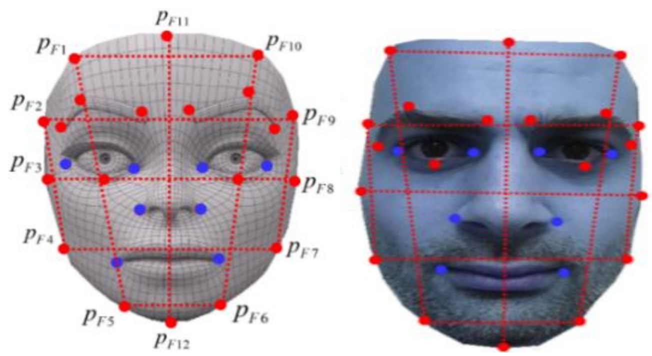


Figure 3.6. Measurements of facial and jaw contours.

and calculated as:

$$F_S = \frac{\left(\left\| p_{F1} - p_{F10} \right\| + \left\| p_{F2} - p_{F9} \right\| + \left\| p_{F3} - p_{F8} \right\| + \left\| p_{F4} - p_{F7} \right\| + \left\| p_{F5} - p_{F6} \right\| \right)}{\left(\left\| p_{F1} - p_{F5} \right\| + \left\| p_{F11} - p_{F12} \right\| + \left\| p_{F10} - p_{F6} \right\| \right)} \tag{3.12}$$

where F_S represents the extracted face shape using the facial feature points p_{F1} through p_{F12}.

Once the facial distance measurements are successfully completed, the calculated distances are forwarded to the facial region extraction stage, as described in the subsequent section.

3.3. Facial Region Extraction

Following the acquisition of related facial measurements, the forehead and cheek regions are extracted accordingly. The choice of these regions is due to their provision of rich data for heart rate estimation, as they expose superficial blood vessels that allow light to easily transmit and reflect back to the HR's camera. Additionally, these regions are less influenced by muscle movements compared to other facial areas. Initially, the corresponding forehead region is extracted using the specified facial feature points in the coordinates and is calculated as follows:

$$F_{FH} = \frac{\left(\left\|p_{F1} - p_{F11}\right\| + \left\|p_{F11} - p_{F10}\right\| + \left\|p_{F10} - p_{rb2}\right\| + \left\|p_{rb2} - p_{rb1}\right\| + \left\|p_{rb1} - p_{lb1}\right\| + \left\|p_{lb1} - p_{lb2}\right\| + \left\|p_{lb2} - p_{F1}\right\| \right)}{F_S} \quad (3.13)$$

where F_{FH} represents the extracted forehead region. Specifically, p_{F1}, p_{F11} and p_{F10} denote the top border of the extracted face shape, while p_{lb1} and indicate the left inner and middle eyebrow points, respectively. Similarly, p_{rb1} and p_{rb2} represent the right inner and middle eyebrow points, respectively.

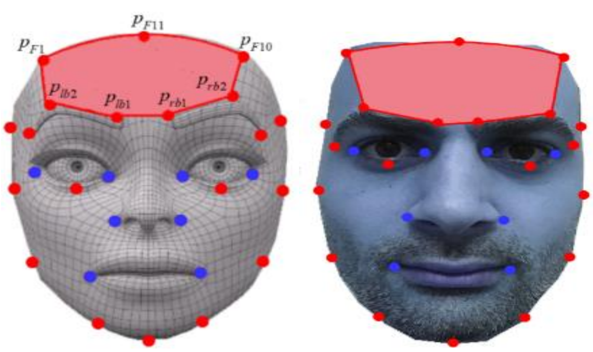


Figure 3.7. Measurement of the forehead area.

Following the extraction of the forehead region, the left and right cheek regions are extracted based on the specified facial points in the x, y coordinates. The corresponding calculations are conducted as follows:

$$F_{LC} = \frac{\left(\left\|p_{F3} - p_{l\_le}\right\| + \left\|p_{l\_le} - p_{Nleft}\right\| + \left\|p_{Nleftt} - p_{Mleftt}\right\| + \left\|p_{Mleft} - p_{F4}\right\| + \left\|p_{F4} - p_{F3}\right\| \right)}{F_S} \quad (3.14)$$

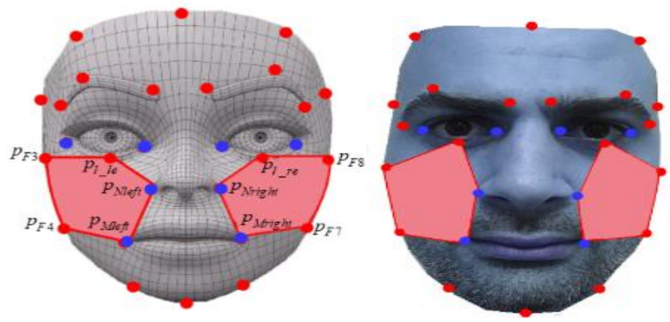


Figure 3.8. Measurement of the left and right cheek areas.

$$F_{RC}=\frac{\left(\ \left\|p_{F8}-p_{l\_re}\right\|+\left\|p_{r\_le}-p_{Nright}\right\|+\left\|p_{Nright}-p_{Mright}\right\|+\left\|p_{Mrightrt}-p_{F7}\right\|+\left\|p_{F7}-p_{F8}\right\|\right)}{F_S} \tag{3.15}$$

where F_{LC} and F_{RC} represent the extracted left and right cheek regions, respectively. Specifically, p_{F3}, p_{F4}, p_{F5} and p_{F7}, p_{F8} denote the feature points at the left and right borders of the extracted face shape. The labels of the left and right lower eye's middle feature points are represented by p_{l_le} and p_{l_re}, respectively. The left and right corners of the nose and mouth are represented by p_{Nleft}, p_{Nright}, p_{Mleft}, p_{Mright}, respectively.

Following the completion of the extraction of specific facial regions, the tracking of these extracted regions is conducted using a hierarchical multi-resolution tracking algorithm, as detailed in the subsequent sub-section.

3.4. Facial Region Tracking with the Hierarchical Multi-Resolution Based Tracking Algorithm

To track the motion of specific facial regions, two consecutive frames, $I_{t-1}(u)$ and $I_t(u)$ are analysed. The goal is to determine the updated facial point location vector $v=u+d$ on the second frame I_t. Here, the displacement vector $d=\left[d_x, d_y\right]^T$ represents the shift of the extracted facial region at point $u=[x \quad y]^T$. The displacement d is subsequently computed by minimizing the specified cost function as follows:

$$\varepsilon(d)=\sum_{x=u_x^L-w_x}^{u_x^L+w_x}\sum_{y=u_y^L-w_y}^{u_y^L+w_y}\left[I_{t-1}(u)-I_t(u+d)\right]^2 \tag{3.16}$$

Therein, w_x and w_y represent the size of integration window value in the x and y coordinates respectively. The size of the integration window presents a trade-off between prediction sensitivity and accuracy. A smaller window size improves detection of minute displacements around point u but may overlook larger displacements, while a larger window size detects broader displacements with reduced resolution. To address these issues, a hierarchical multi-resolution approach is employed. This method reconstructs the integration window at various levels, each with a smaller size, enabling analysis at different resolutions and more effective detection of displacements exceeding the integration window's size. Consequently, this hierarchical adjustment of image frame resolution enhances both sensitivity and accuracy in detecting larger movements.

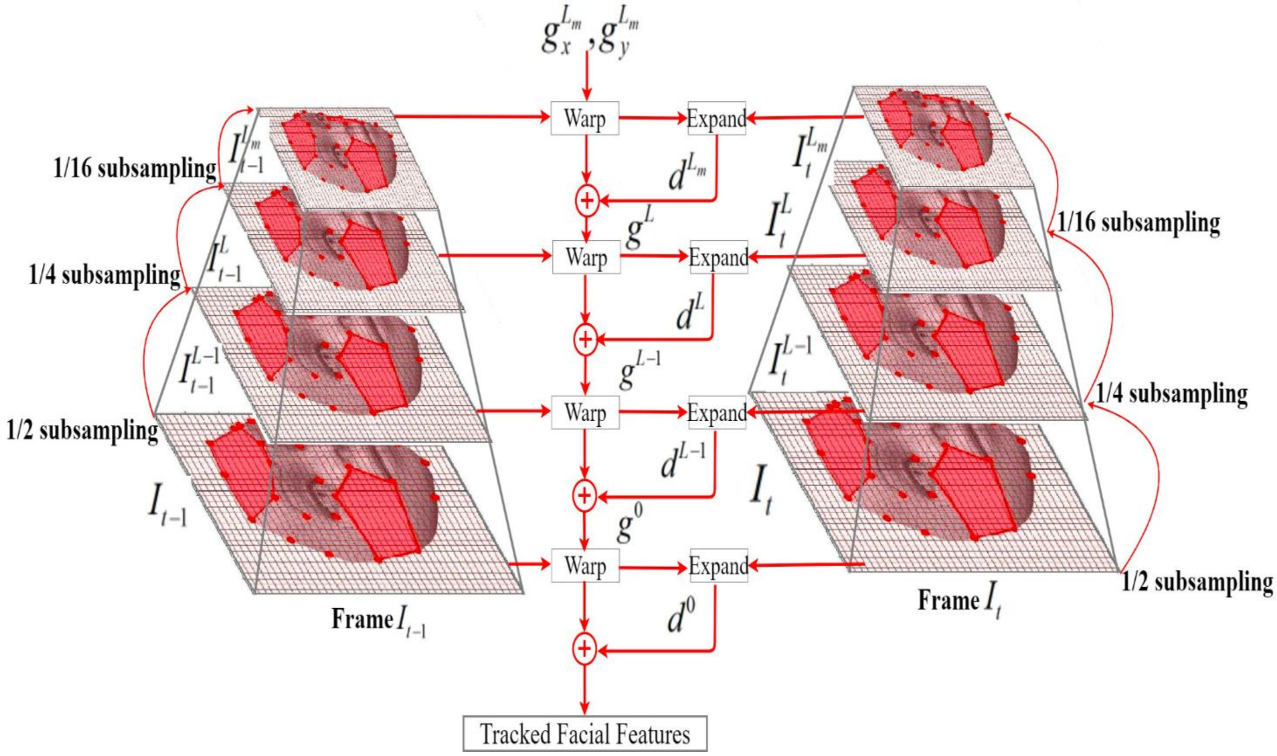


Figure 3.9. Hierarchical multi-resolution based tracking algorithm structure.

The recursive hierarchical multi-resolution structure for the consecutive image frames I_{t-1} and I_t is established for a specific case with the following height/level configuration:

$$I^L(x,y)=\frac{I^{L-1}(2x,2y)}{4}+\frac{\left(I^{L-1}(2x+1,2y)+I^{L-1}(2x-1,2y)+I^{L-1}(2x,2y+1)+I^{L-1}(2x,2y-1)\right)}{8}+\frac{\left(I^{L-1}(2x+1,2y+1)+I^{L-1}(2x-1,2y-1)+I^{L-1}(2x+1,2y-1)+I^{L-1}(2x-1,2y-1)\right)}{16} \tag{3.17}$$

As detailed in Equation (3.17), each image frame is down sampled by a factor of 1/2 at each hierarchical level, creating a structure with four distinct resolutions, as depicted in Figure 3.9. The algorithm begins at the top coarsest resolution level L_m, where it estimates the optimal motion by calculating the displacement vector to minimize the error function ε or that level. The resulting displacement estimation is then passed to the lower level L as the initial estimate for facial feature point displacement. This process continues through the hierarchy, with refined displacement patterns calculated at each subsequent level L, and L^0, up to the finest resolution. Each level's estimation serves as the initial value for the next, ultimately predicting the final displacement. The window size remains consistent across all levels, allowing for the detection of large head movements at coarser resolutions. These stages are provided in a sequence by the pseudo-code as:

Algorithm 3.1. Hierarchical multi-resolution based facial region tracking algorithm.

Input: u feature point of the extracted facial region on the first image frame I_{t-1}.
Output: Updated location v of u on the next image frame I_t.
Initialization:

1. Build the hierarchical multi resolution representations of I_{t-1} and I_t as $\left\{I_{t-1}^L\right\}_{L=0,\ldots,L_m}$ and $\left\{I_t^L\right\}_{L=0,\ldots,L_m}$ with each corresponding resolution level L.

2. Initialize the displacement estimation value of the facial feature points as

$$g^{L_m} = \left[g_x^{L_m} \quad g_y^{L_m}\right]^T = \left[0 \quad 0\right]^T \tag{3.18}$$

for $L = L_m$ to coarsest/lowest resolution level $L = L_m$

3. Define location of u on image frame I_{t-1}^L as

$$u^L = \left[p_x \quad p_y\right]^T = \frac{u}{2^L} \tag{3.19}$$

with p_x and p_y pixel values of u in the x and y coordinates.

4. Take derivative of I_{t-1}^L with respect to x coordinate as

$$\left(I_{t-1}\right)_x (x,y) = \frac{I_{t-1}^L(x+1,y) - I_{t-1}^L(x-1,y)}{2} \tag{3.20}$$

5. Take derivative of I_{t-1}^L with respect to y coordinate as

$$\left(I_{t-1}\right)_y (x,y) = \frac{I_{t-1}^L(x,y+1) - I_{t-1}^L(x,y-1)}{2} \tag{3.21}$$

6. Obtain the spatial gradient matrix as

$$G = \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} \begin{bmatrix} \left(I_{t-1}\right)_x^2 (x,y) & \left(I_{t-1}\right)_x (x,y) I_y (x,y) \\ \left(I_{t-1}\right)_x (x,y) \left(I_{t-1}\right)_y (x,y) & \left(I_{t-1}\right)_y^2 (x,y) \end{bmatrix} \tag{3.22}$$

7. Initialize iterative tracking algorithm as $v^0 = \left[0 \quad 0\right]^T$ (3.23) *for* $k=1$ to K iteration

8. Determine difference value of the first image frame and second wrapped image frame as

$$\delta(I_{t-1})(x,y)=I_{t-1}^{L}(x,y)-I_{t}^{L}\left(x+g_x^L+v_x^{k-1},y+g_y^L+v_y^{k-1}\right) \quad (3.24)$$

9. Obtain image frame mismatch vector as

$$b_k=\sum_{x=p_x-w_x}^{p_x+w_x}\sum_{y=p_y-w_y}^{p_y+w_y}\begin{bmatrix}\delta(I_{t-1})_k(x,y)(I_{t-1})_x(x,y)\\ \delta(I_{t-1})_k(x,y)(I_{t-1})_y(x,y)\end{bmatrix} \quad (3.25)$$

10. Attain the pattern of apparent displacement as $\eta^k=G^{-1}b_k$ (3.26)

11. Take initial displacement estimation value of facial feature point for the next iteration as $v^k=v^{k-1}+\eta^k$ (3.27)

end for

12. Determine the final pattern of apparent displacement at level L as $d^L=v^k$.

13. Take displacement estimation value of the facial feature point for the next level $L-1$ as $g^{L-1}=\begin{bmatrix}g_x^{L-1} & g_y^{L-1}\end{bmatrix}^T=2\left(g^L+d^L\right)$ (3.28)

end for

14. Take the final pattern of apparent displacement vector as

$$d=g^0+d^0 \quad (3.29)$$

15. Determine the location of translated and updated facial feature point on the second image frame I_t as

$$v=u+d \quad (3.30)$$

Finally, the facial regions are tracked in real-time using the completed tracking algorithm. The proposed algorithm mitigates the impact of both small and large head motion variations on the identified facial regions, providing more robust and stable coordinates for the face.

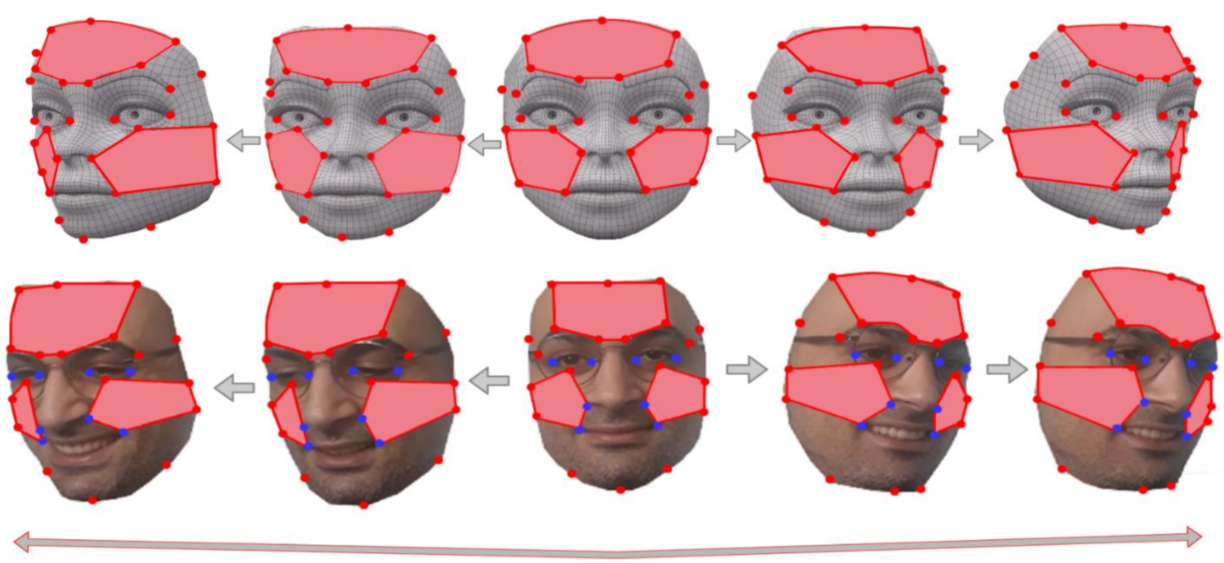


Figure 3.10. Illustration of the instantaneous facial region tracking.

Consequently, it enables the real-time extraction of heart rate information from the facial regions, even in the presence of head movements. After the tracking of the extracted regions using a hierarchical multi-resolution tracking algorithm, the process advances to the heart rate estimation stage.

3.5. Heart Rate Estimation

In this sub-section, the key steps for the heart rate are explained in detail.

3.5.1. Normalization of the Tracked Facial Regions

Once the facial regions are tracked, the next step is to average the features within these regions across each colour channel to obtain $R_{avg}(t)$, $G_{avg}(t)$, and $B_{avg}(t)$, which correspond to the average red, green, and blue facial features at the frame t as :

$$R_{avg}(t) = \frac{1}{n_{roi}} \sum_{i}^{roi} \sum_{j}^{roi} R_{i,j}(t) \tag{3.31}$$

$$G_{avg}(t) = \frac{1}{n_{roi}} \sum_{i}^{roi} \sum_{j}^{roi} G_{i,j}(t) \tag{3.32}$$

$$B_{avg}(t) = \frac{1}{n_{roi}} \sum_{i}^{roi} \sum_{j}^{roi} B_{i,j}(t) \tag{3.33}$$

where the roi (region of interest) represents the tracked facial regions.

After calculating the average values, they should be further normalized as follows:

$$R_{norm}(t) = \frac{R_{i,j}(t) - R_{avg}(t)}{\sigma_R} \tag{3.34}$$

$$G_{norm}(t) = \frac{G_{i,j}(t) - G_{avg}(t)}{\sigma_G} \tag{3.35}$$

$$B_{norm}(t) = \frac{B_{i,j}(t) - B_{avg}(t)}{\sigma_B} \tag{3.36}$$

where σ_R, σ_G, and σ_B represent the standard deviations of the red, green, and blue channels, respectively.

Once the normalized RGB values are obtained, they are stored in vector form for subsequent processing. While the human heart rate falls within a narrow frequency range, the normalized colour channels often exhibit additional movements at frequencies outside this range due to various nonlinear head motion artefacts. To mitigate these artefacts and accurately measure the heart rate over a broad frequency range, a Butterworth filter is applied. This filter, which passes frequencies between 0.5 Hz (30 BPM) and 4 Hz (240 BPM) while rejecting those outside this range, helps eliminate unwanted frequencies. Additionally, the Butterworth filter includes a pre-whitening process for further refinement. Consequently, the filtered channels are obtained by applying this band-pass filter to the normalized values as follows:

$$R_{BP}(t) = BP(t)R_{norm}(t) \tag{3.37}$$

$$G_{BP}(t) = BP(t)G_{norm}(t) \tag{3.38}$$

$$B_{BP}(t) = BP(t)B_{norm}(t) \tag{3.39}$$

The next sub-section reviews the Adaptive ICA technique to eliminate artefacts in the filtered data.

3.5.2. Motion Artefacts Elimination with the Adaptive ICA Technique

After normalization, the data are further processed by decomposing the normalized and filtered signals into three independent source signals using the Adaptive ICA technique. This statistical method represents multivariate data as a linear combination of independent components. Unlike correlation-based methods, Adaptive ICA not only de-correlates the resulting values but also reduces high statistical dependence. It is particularly effective compared to classical methods due to its capacity to handle non-Gaussian distributed data. Let the initial observed data be denoted as $X = [X_1, X_2, X_3]^T$ whose m elements are mixtures of the independent components of the random data $S = [S_1, S_2, \ldots, S_m]^T$ given by:

$$X = AS \tag{3.40}$$

where dimension of X is 3×1, dimension of A is $3 \times m$ and dimension of S is $m \times 1$ respectively.

The mixing matrix is denoted by A, and the sample values of X_j are assigned by x_j and $j = 1,2,\ldots,m$. The ultimate goal is to find a separating or un-mixing matrix W, which is the inverse of the original mixing matrix A, that provides the best possible approximation of S as follows:

$$S \cong WX \tag{3.41}$$

Equation (3.41) calculates each independent component of S from the mixture of RGB signals and can be represented in matrix form as:

$$\begin{bmatrix} S_1(t) \\ S_2(t) \\ S_3(t) \end{bmatrix} = \begin{bmatrix} R_{BP}(t) \\ G_{BP}(t) \\ B_{BP}(t) \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \begin{bmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{bmatrix} \tag{3.42}$$

In this context, it is essential to emphasize that the sample W_p is represented by, and the goal is to maximize the non-Gaussianity of each source component w_p. To accomplish this, the unmixing matrix corresponding to the components of the source data is calculated iteratively using the following procedure:

$$\overline{w}_p = w_p + \eta \left(\underbrace{\mu\left[x\sigma\left(w_p^T x\right)\right]}_{(a)} - \underbrace{\mu\left[\sigma'\left(w_p^T x\right)\right] w_p}_{(b)} \right) \tag{3.43}$$

where refers to the non-quadratic function, while σ' denotes its derivative. The term μ represents the empirical mean, and η represents learning rate with being in $0 < \eta < \frac{2}{x^T x}$.

In equation (3.43), the term (a) represents the μ of the product of the data x and the non-linear transformed projection $\sigma\left(w_p^T x\right)$. In this case, it involves summing across all observed samples of x and subsequentially applying σ, thereby modifying w_p in a matter that maximizes the non-Gaussianity of the projected data.

Conversely, the term (b) considers the μ value of the $\left[\sigma'\left(w_p^T x\right)\right]$ which gives a scalar that adjusts the direction of w_p. The absence of this term could result in an unstable solution during the gradient descent process. Similar to the previous term, this adjustment is also approximated by averaging over all observed data points.

In each iteration, the computed $\overline{w}_p$ is adjusted by subtracting the projection of the previously estimated vectors, using the Gram-Schmidt process, as follows:

$$\overline{\overline{w}}_p = \overline{w}_p - \sum_{j=1}^{p-1} (\overline{w}_p^T w_j) w_j \tag{3.44}$$

Herein, it is essential to recognize that the recursion for specific independent components of $\overline{w}_p$ may not converge to the same maximum values. To tackle this issue, the components are sequentially extracted from the mixture through the use of orthogonal projections, as articulated in Equation (3.45).

Following each update, the $\overline{\overline{w}}_p$ values are normalized as follows:

$$\vec{w}_p = \overline{\overline{w}}_p / \|\overline{\overline{w}}_p\| \tag{3.45}$$

Once convergence is achieved with the normalized and updated $\vec{w}_p$, the subsequent component ($\vec{w}_{p+1}$) must be made orthogonal to it and all previously extracted values using Equations (3.44) and (3.45) to ensure its distinctiveness. If converge is attained, the process increments p and proceeds to the next component. The procedure continues until the condition $p > n$ is satisfied, where n denotes the total number of independent components. However, if convergence of the estimated values is not achieved, the update process must be repeated from Equation (3.43) and continued until convergence is confirmed by verifying the constraint on $\left\|\overline{w}_p - \overline{\overline{w}}_p\right\| < \varepsilon$.

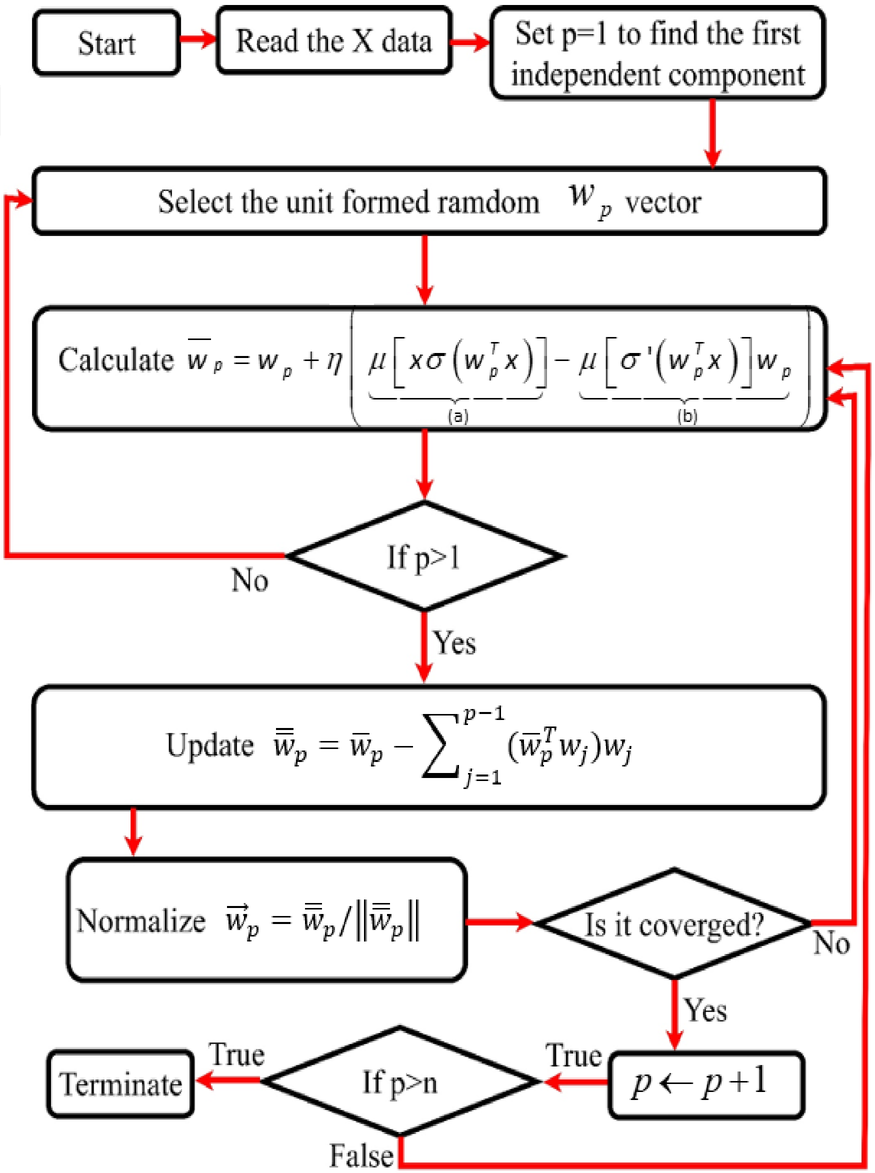


Figure 3.11. Flow diagram of the implemented adaptive ICA algorithm.

Following this, it proceeds to the subsequent stage.

3.5.3. Frequency Analysis using FFT Technique

In the subsequent step, the independent data are transformed into the frequency domain using the FFT method to extract dominant frequencies. This transformation is achieved through the inverse discrete FT, defined as follows:

$$X[n]=\sum_{n=0}^{N-1}\log\left(\left|\sum_{n=0}^{N-1}x[n]\exp\left(-j\frac{2\pi}{N}kn\right)\right|\right)\exp\left(-j\frac{2\pi}{N}kn\right) \tag{3.46}$$

wherein, the term j represents one of the complex coefficients, which are described as the Fourier descriptors of the boundary. n and k denote the time and frequency indices, respectively, while N represents the number of decomposed samples. The windowing process involves multiplying a segment of the signal by a window function with a gradually tapering amplitude, which reduces leakage effects by ensuring a smooth transition towards zero. Upon the conclusion of this process, the subject is subsequently advanced to the final stage of estimation.

3.5.4. Determination of Dominant Frequency Peaks for Preliminary Heart Rate Assessment

After computing the FFT, the highest peak value among the dominant frequencies is identified as the preliminary heart rate. This is achieved by calculating the maxima and their respective indices within each local window for each time interval. The largest peaks are then selected from these local maxima through a three-step process. In the first step, each local maximum is compared with its preceding and succeeding values to identify the peaks, as illustrated with number 1. To meet this step, the current local maxima must be greater than both the preceding and succeeding values. In the second step, the local maxima are assessed against a predefined threshold value, as depicted with number 2; each maximum must exceed this threshold. In the final step, the largest peak candidates are evaluated based on the time intervals between them, compared against a predefined time threshold, as shown with number 3. This involves checking whether there are other maxima close to the candidate peaks within the specified time threshold. If multiple maxima are present, the peak with the highest value is selected as the heart rate peak f_p (highlighted with red circles), and the corresponding heart rate value is computed as:

$$Hr = \frac{60}{t_0} f_p \tag{3.47}$$

wherein, t_0 represents the current time, ranging from 1 to n seconds.

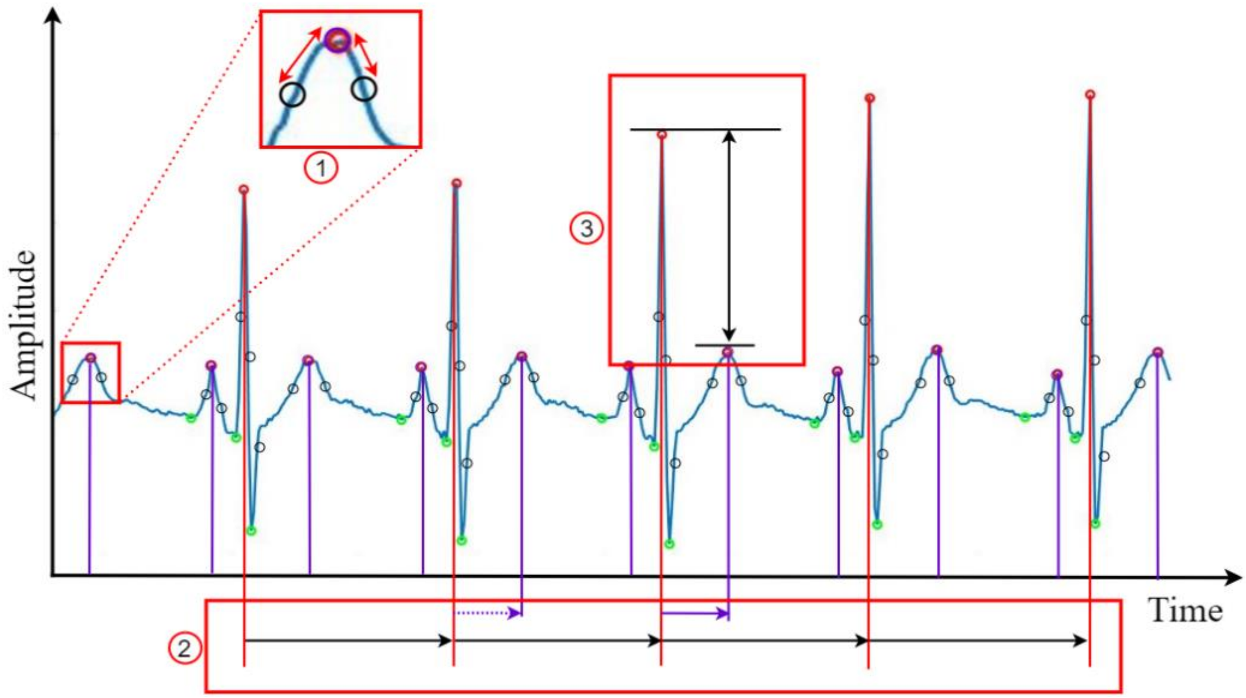


Figure 3.12. Illustration of peak search process for heart rate estimation.

As indicated by Equation (3.47), heart rate estimation is performed over the initial 60 seconds. After determining the final heart rate value, the absolute difference between the current heart rate value $Hr(t_o)$ and the previous heart rate value $Hr(t_o - \Delta t)$ s examined to eliminate false detections caused by abrupt fluctuations in the heart rate, as expressed by:

$$Hr_{diff} = \left| Hr(t_o - \Delta t) - Hr(t_o) \right| \tag{3.48}$$

If the Hr_{diff} exceeds 12 BPM, a sliding average is calculated using the most recent 12 values, as follows:

$$Hr_{final} = \begin{cases} Hr(t_o) & \text{if } Hr_{diff} < 12 \text{ BPM} \\ \frac{1}{12} \sum_{t=t_o-11}^{t_o} Hr(t) & \text{otherwise} \end{cases} \tag{3.49}$$

Once the Hr_{diff} exceeds 12 BPM, a timer is initiated for 4 seconds. During this period, $Hr(t_o - \Delta t)$ s used as the output. After 4 seconds, the new heart rate value is compared again. If Hr_{diff} is still above 12 BPM, the average of the last 12 values is then used as the final output. Upon concluding the explanation of the final stage of the proposed algorithm, the subsequent section provides a detailed examination of the simulation and experimental results of the proposed algorithm.

3.6. Simulation and Experimental Results

This sub-section evaluates the proposed algorithm through both simulation and experimental environments. It begins by describing the experimental setup and subsequently assesses the results.

3.6.1. Simulation and Experimental Settings

To assess the reliability and effectiveness of the proposed algorithm, a self-collected dataset was established. This dataset was gathered in a lab environment with sunlight as the sole light source. It comprises five sessions, each lasting 30 seconds, involving 10 subjects. In the first session, subjects maintain a neutral expression and still head position in front of the NAO HR camera. In the second session, they move their heads horizontally (right and left). The third session involves head movements at 15, 30, and 40 degrees in various directions (yaw, pitch, roll). During the fourth session, subjects display a range of basic emotional expressions, including laughter, surprise, sadness, anxiety, disgust, and fear. In the final session, subjects sit casually with normal head movements, simulating an interaction scenario where the HR detects heart rate while the subjects are at ease. All sessions are performed at distances of 0.5m and 1m from the HR camera to evaluate the impact of distance.

In addition to the described sessions, an electrocardiograph is attached to the subjects to record the ground truth heart rate, providing a basis for validating the accuracy of the proposed algorithm. Statistical metrics, including Standard Deviation (SD), Mean Square Error (MSE), Correlation Coefficient (CC), and Bland-Altman analysis, are employed to assess the agreement and investigate the discrepancies between the estimated and reference heart rate values, thus quantitatively evaluating the algorithm's accuracy. Following these analyses, the proposed algorithm is integrated into the NAO HR interface using the Choregraphe program to demonstrate its applicability and reliability.

3.6.2. An Overview of NAO HR

The NAO HR is an autonomous, programmable bipedal robot developed by Aldebaran Robotics. Standing 58 cm tall, it is capable of movement, human recognition, auditory perception, and verbal communication. It serves as a platform in the Two-Legged Standard League.

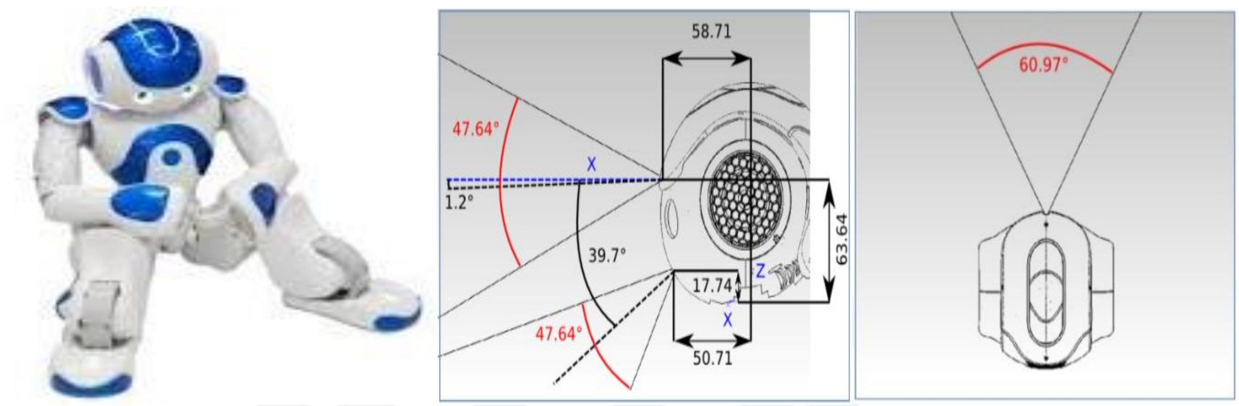


Figure 3.13. NAO HR with its vertical and horizontal camera ranges.

NAO uses two identical video cameras, positioned on its forehead and inside its mouth, to interpret emotions through facial expressions. Both cameras offer a resolution of 1280×960 pixels at 30 frames per second. The vertical field of view for each camera is 47.64 degrees, while the horizontal field of view is 60.97 degrees.

3.6.3. Results

The first experiment evaluates the tracking accuracy of the proposed algorithm for a single subject across the described five sessions. Figure 3.14 illustrates the tracking performance using forehead data for each session, compared to the actual data measured at a distance of 0.5 meters over 50 seconds. The swinging and rotational motions fall within the algorithm's sample frequency, allowing it to detect the corresponding measurement points and continue heart rate estimation without interruption. Furthermore, as the measurements are taken from the forehead, the results are minimally affected by emotions and conversations. Figure 3.14 shows a strong correlation between the estimated and actual values across all sessions. Similarly, Figure 3.15 plots the estimated values against the actual values for each session.

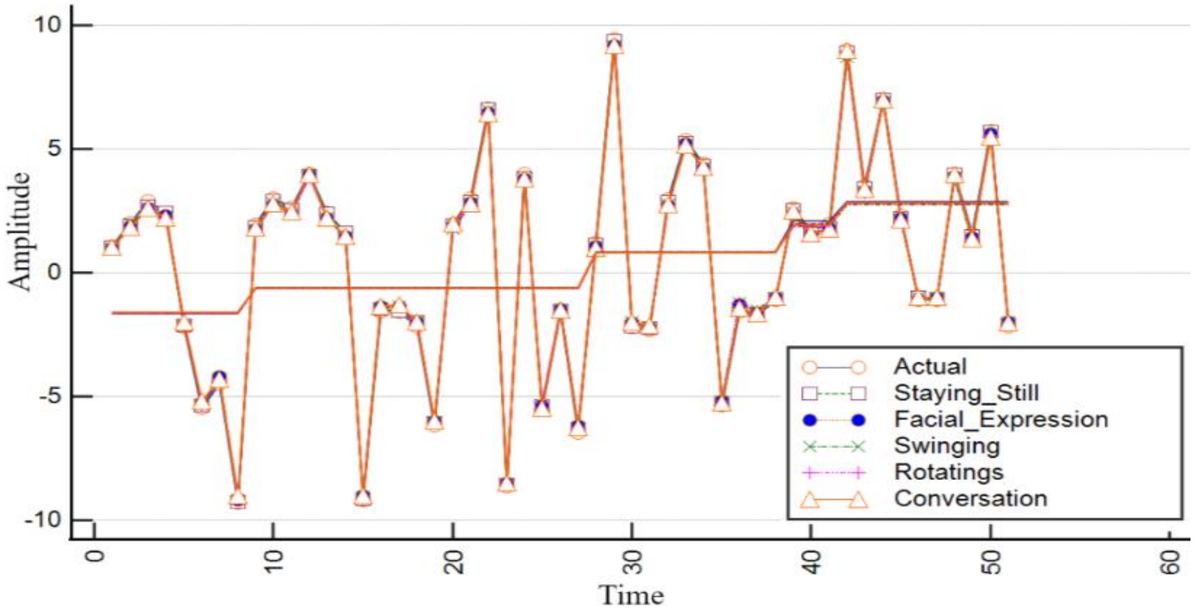


Figure 3.14. Isotonic curve analysis of tracking performance for one subject across five conditions.

Table 3.1 reveals that the correlation coefficient exceeds 0.99 in every session, indicating a nearly perfect positive linear relationship with the actual values.

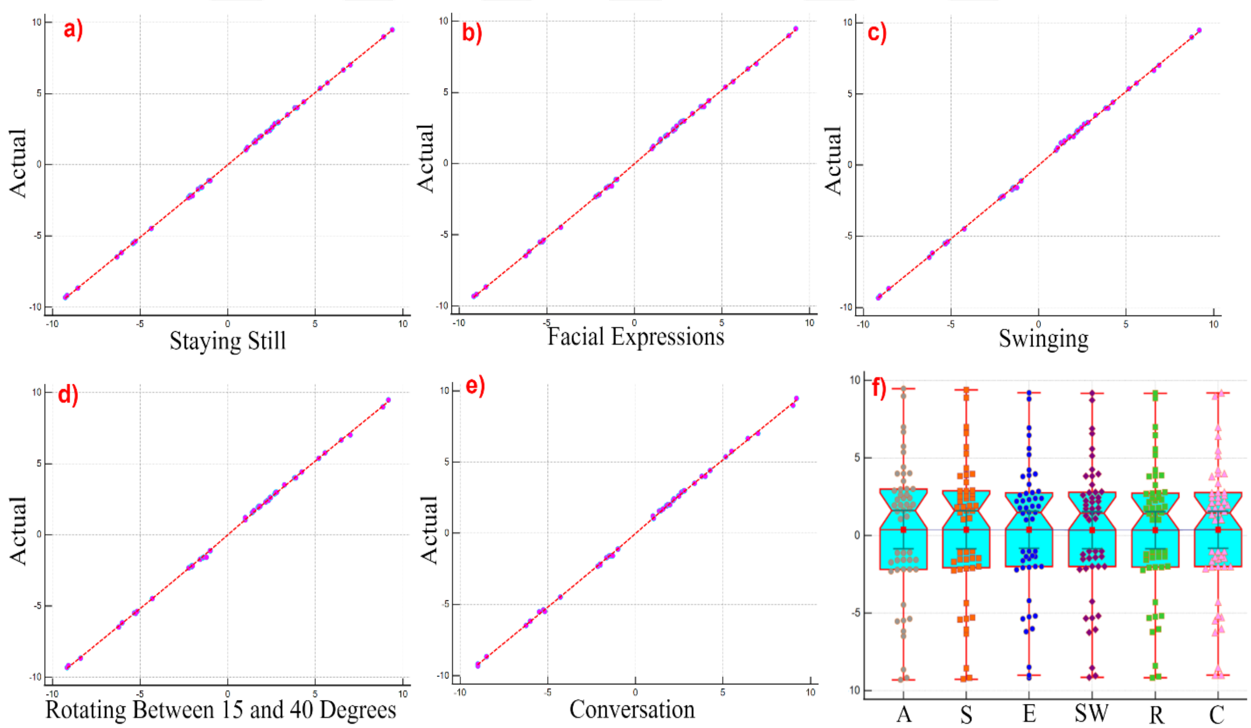


Figure 3.15. Correlation analysis of the proposed algorithm's tracking performance for one subject: (a) Session 1, (b) Session 2, (c) Session 3, (d) Session 4, (e) Session 5, and (f) Error values for each condition (A: Actual, S: Staying Still, E: Expressions, SW: Swinging, R: Rotating, C: Conversation).

Herein, the average error values are close to zero, and the standard deviations align closely with the RMSE values for each session. The Bland-Altman analysis, presented in Figure 3.16, graphically illustrates the accuracy of the experiment, confirming over 95% reliability in the experimental data.

Table 3.1. Statistical analysis of the proposed algorithm's tracking performance for one subject over 50 seconds.

Sessions	Conditions	Correlation Coefficient (CC)	Mean Square Error (MSE)
1	Staying Still	0,9909	0,00422
2	Facial Expressions	0,9907	0,01126
3	Swinging	0,9906	0,01143
4	Rotating	0,9906	0,01436
5	Conversation	0,9906	0,01391

Thus, the variability and precision of the compared data are quantitatively verified. The average values for each session are plotted against the actual values on the horizontal axis, while the differences between them are shown on the vertical axis.

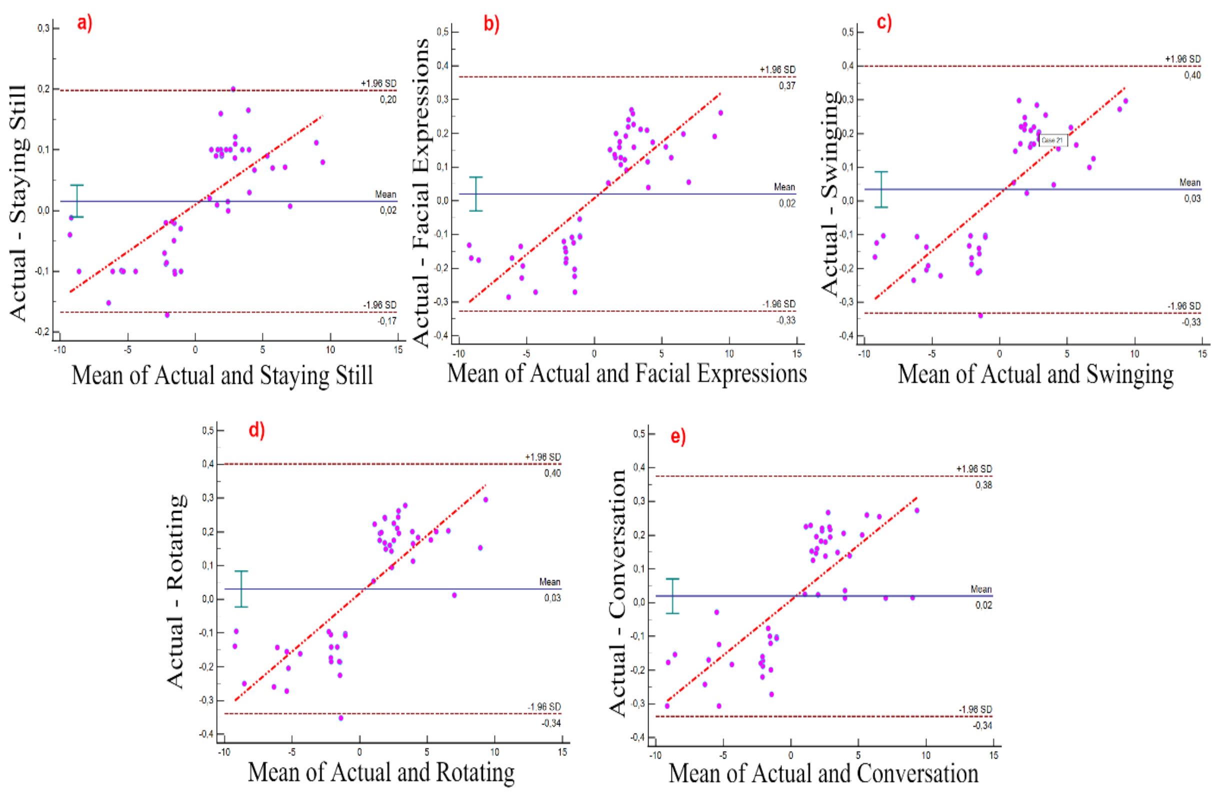


Figure 3.16. Bland–Altman analysis results with 95% limits of agreement: (a) Session 1, (b) Session 2, (c) Session 3, (d) Session 4, (e) Session 5.

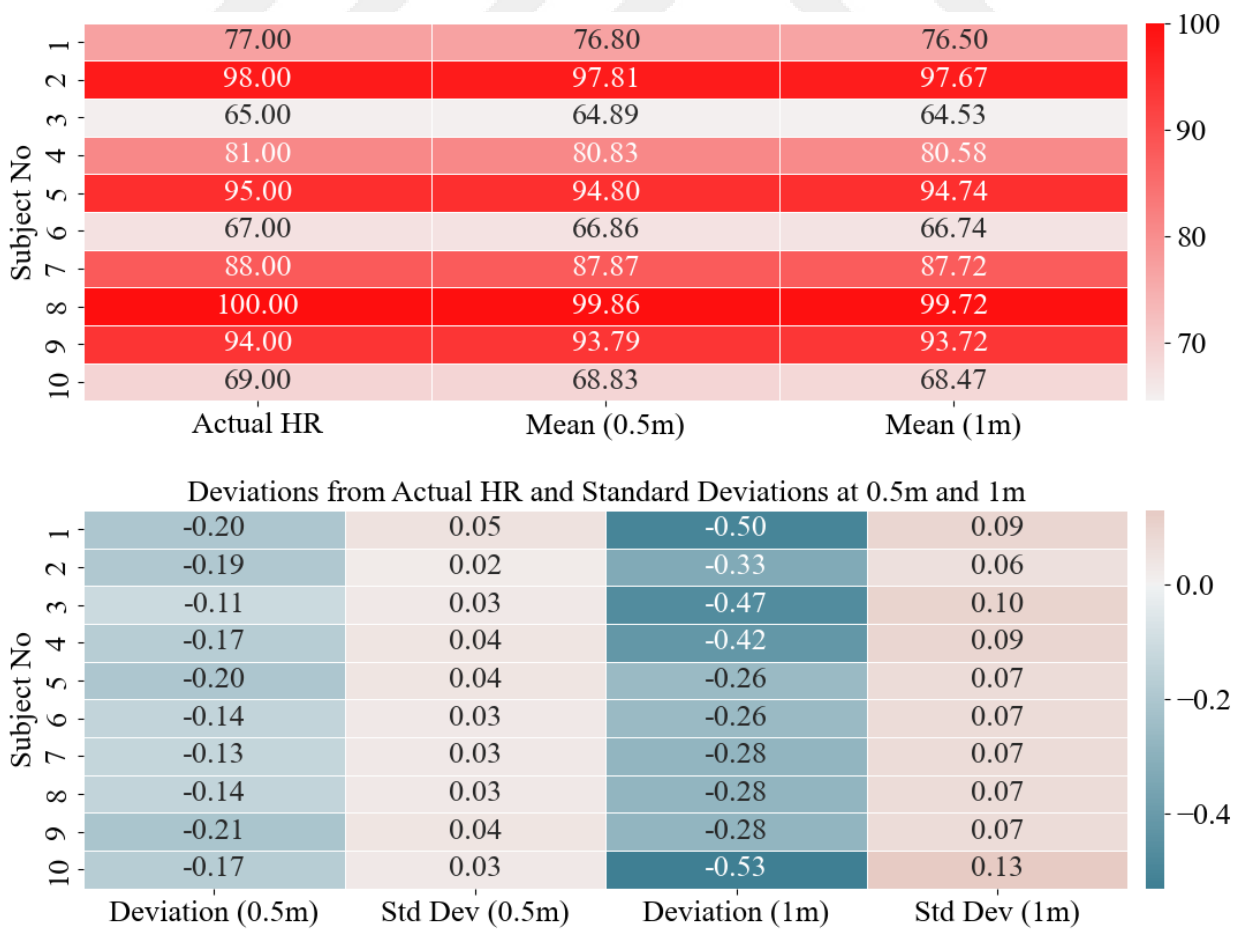
Additionally, Table 3.2 presents a reasonable limit of agreement (LOA), showing that 95% of the data fall within 1.96 standard deviations.

Table 3.2. Limit of agreement (LOA) values for each session.

Sessions	Conditions	Limit Of Agreement (LOA)
1	Staying Still	[-0,17 0,20]
2	Facial Expressions	[-0,33 0,37]
3	Swinging	[-0,33 0,40]
4	Rotating	[-0,34 0,40]
5	Conversation	[-0,34 0,38]

As observed in Table 3.3, the proposed algorithm demonstrates high precision compared to the actual values, with most data points falling within the agreed limits.

Table 3.3. Performance analysis of the proposed heart rate estimation algorithm for ten subjects at distances of 0.5 and 1 meters.



This indicates that the difference between the actual and estimated values remains within a satisfactory range.

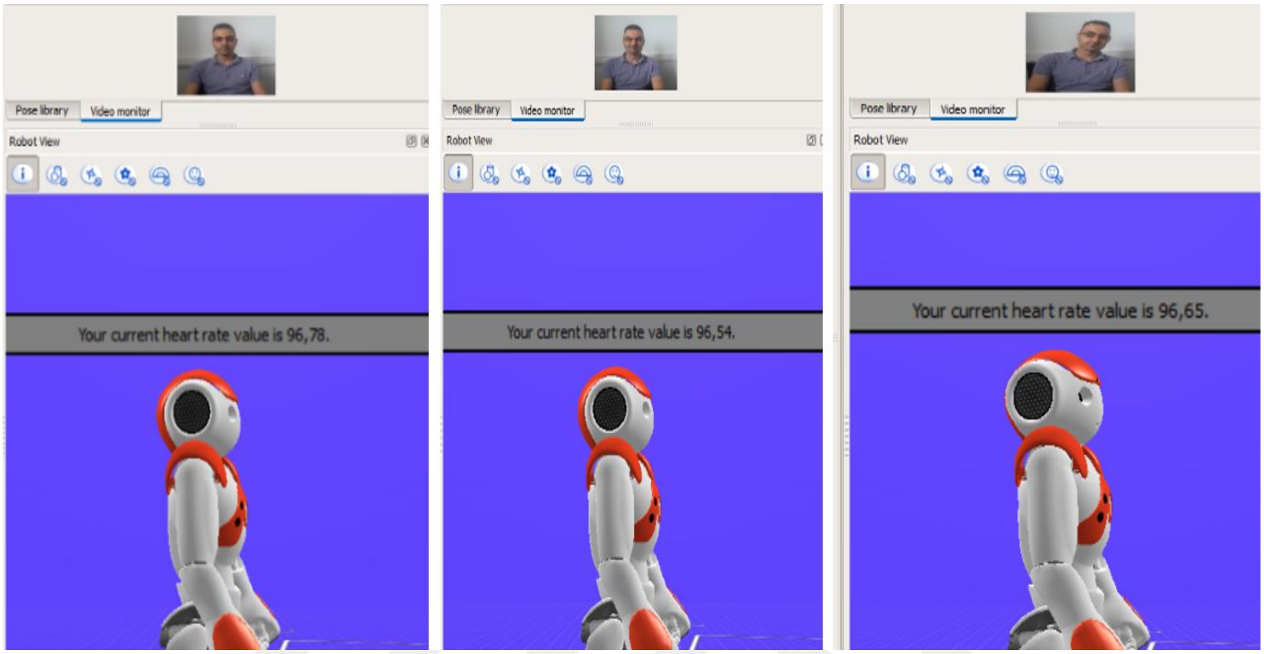


Figure 3.17. From left to right: Heart rate estimation application using Choreographe in sessions 1, 2, and 3.

Subsequently, the heart rate values for 10 subjects were estimated from distances of 0.5 and 1 meter under 33 conditions.

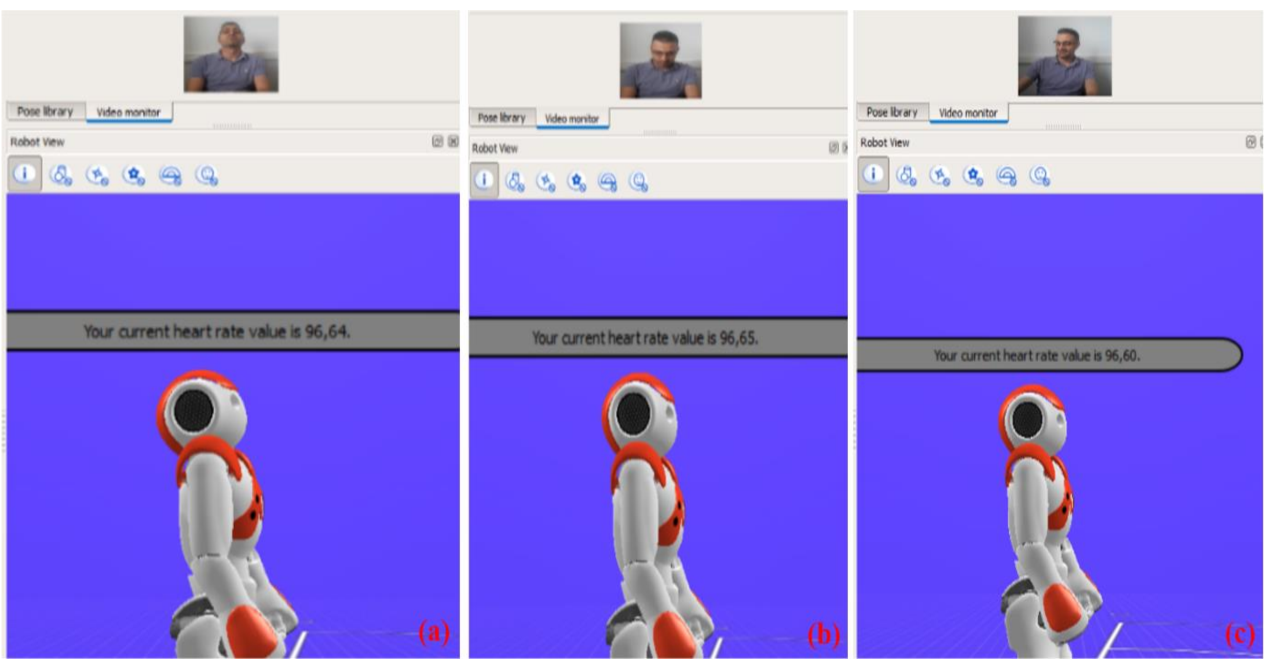


Figure 3.18. Heart rate estimation application using Choreographe in sessions 4 and 5.

Herein, the tracking performances of the forehead, right cheek, and left cheek regions are analysed separately using two statistical features (mean and standard deviation) for 33 conditions at a distance of 0.5 meters. The analysis yields mean and standard deviation values of 99% and 0.13, respectively. These results demonstrate that the proposed algorithm maintains high tracking performance and accuracy, even with independent evaluations of each facial region.

To assess the impact of distance, the same experiments were conducted at a distance of 1 meter, yielding mean and standard deviation values of 98.91% and 0.17, respectively. Although there is a slight decrease in these values, the tracking performance for the relevant facial regions remains effective. In addition to the conducted experiments, the social dialogue engagement capability of the NAO HR was also observed during the heart rate estimation processes in a natural environment, as illustrated Figure 3.19.

After accurately obtaining the heart rate information for 98.61% of the subjects, as shown in Figure 3.19 (a), the NAO HR engages in communication with the subject by analysing their emotional statements, as depicted in Figure 3.20. As illustrated in Figure 3.20, the NAO HR first analyses the subject's heart rate information and then initiates the dialogue by greeting the subject. The communication continues by incorporating both the heart rate data and the detected emotional statements. In this scenario, the NAO HR detects a serious facial expression and requests permission to tell a funny joke to entertain the subject. Upon receiving consent, the NAO HR begins the joke-telling process.

Unlike the previous social interaction experiment shown in Figure 3.19 (a), this interaction involves verifying the consistency between the heart rate data and the detected emotional statement. The "sad" emotion is specifically chosen for analysis due to its distinctive pattern. During sadness, the heart rate typically increases while blood flow to the heart decreases, facilitating a clear comparison between the emotional statement and the heart rate information. During this process, the NAO HR monitors any changes in the subject's emotional statements. If a positive change is detected, the NAO HR expresses its happiness and continues the interaction, as illustrated in Figure 3.20. In a separate experiment shown in Figure 3.19 (b), the NAO HR accurately obtains the subject's heart rate information with a 98.55% accuracy rate and engages in communication by analysing the subject's emotional statements, as depicted in Figure 3.21.

Table 3.4. Numerical results of the conducted experiments in Figure 3.17 and Figure 3.18.

Conditions	**Measured Values**
Actual Heart Rate Value	97
Staying Still	96,78
Facial Expressions	96,54
Swinging (Fast)	96,65
Rotating (up, pitch direction in 40 degrees, Fast)	96,64
Rotating (down, pitch direction in 40 degrees, Fast)	96,65
Conversation	96,60

Herein, the NAO HR initially retrieves and communicates the subject's heart rate information. It then evaluates the subject's emotional statement and compares it with the obtained heart rate data.

Figure 3.19. Social dialogue engagement experiment 1 based on the detected heart rate value.

Upon detecting a mismatch between the heart rate information and the emotional statement, the NAO HR informs the subject of this discrepancy and requests the true emotional statement. If the subject confirms being genuinely sad, the NAO HR inquiries about the reason and re-initiates the analysis.

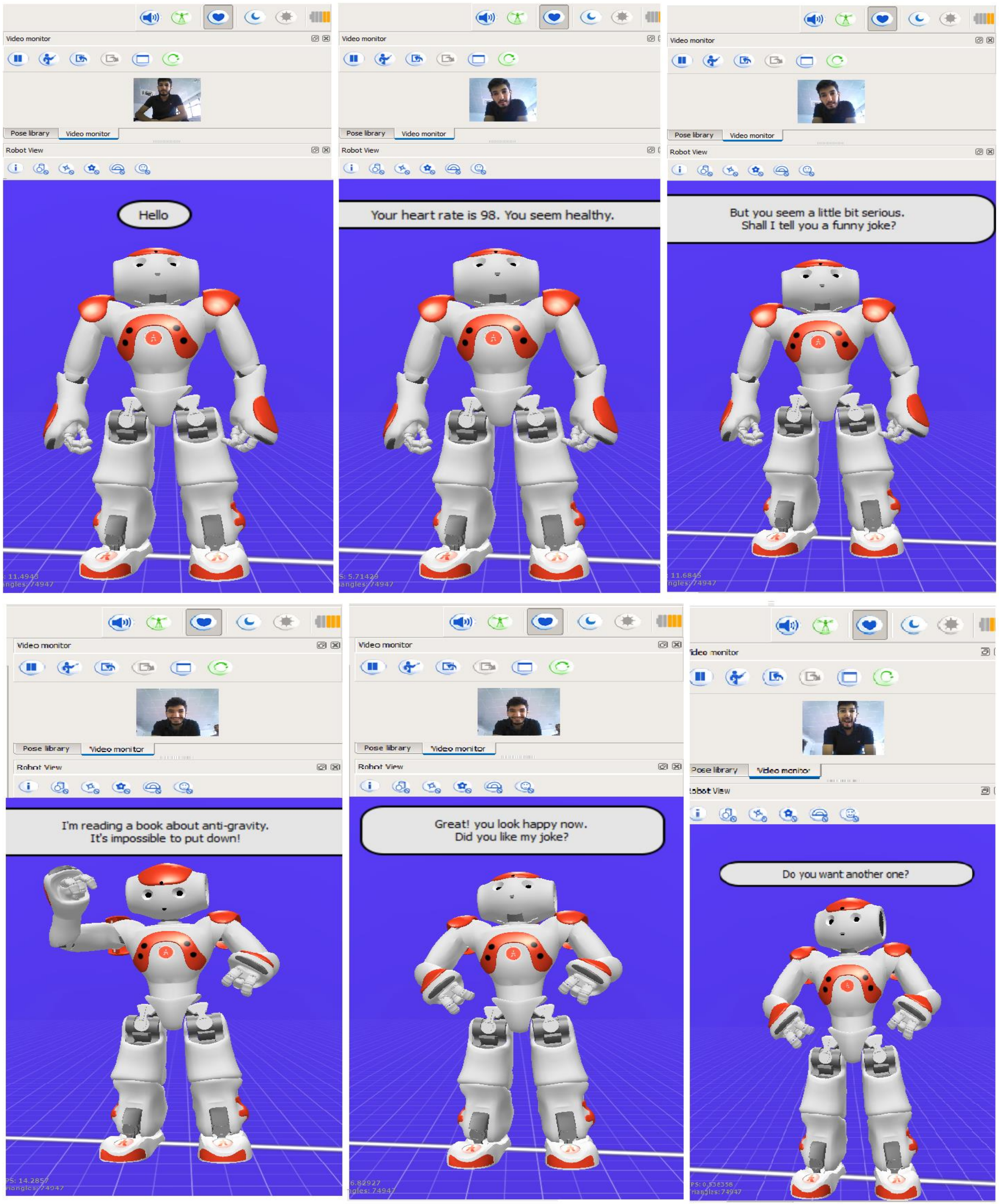


Figure 3.20. Social dialogue engagement experiment 2 based on the detected heart rate value.

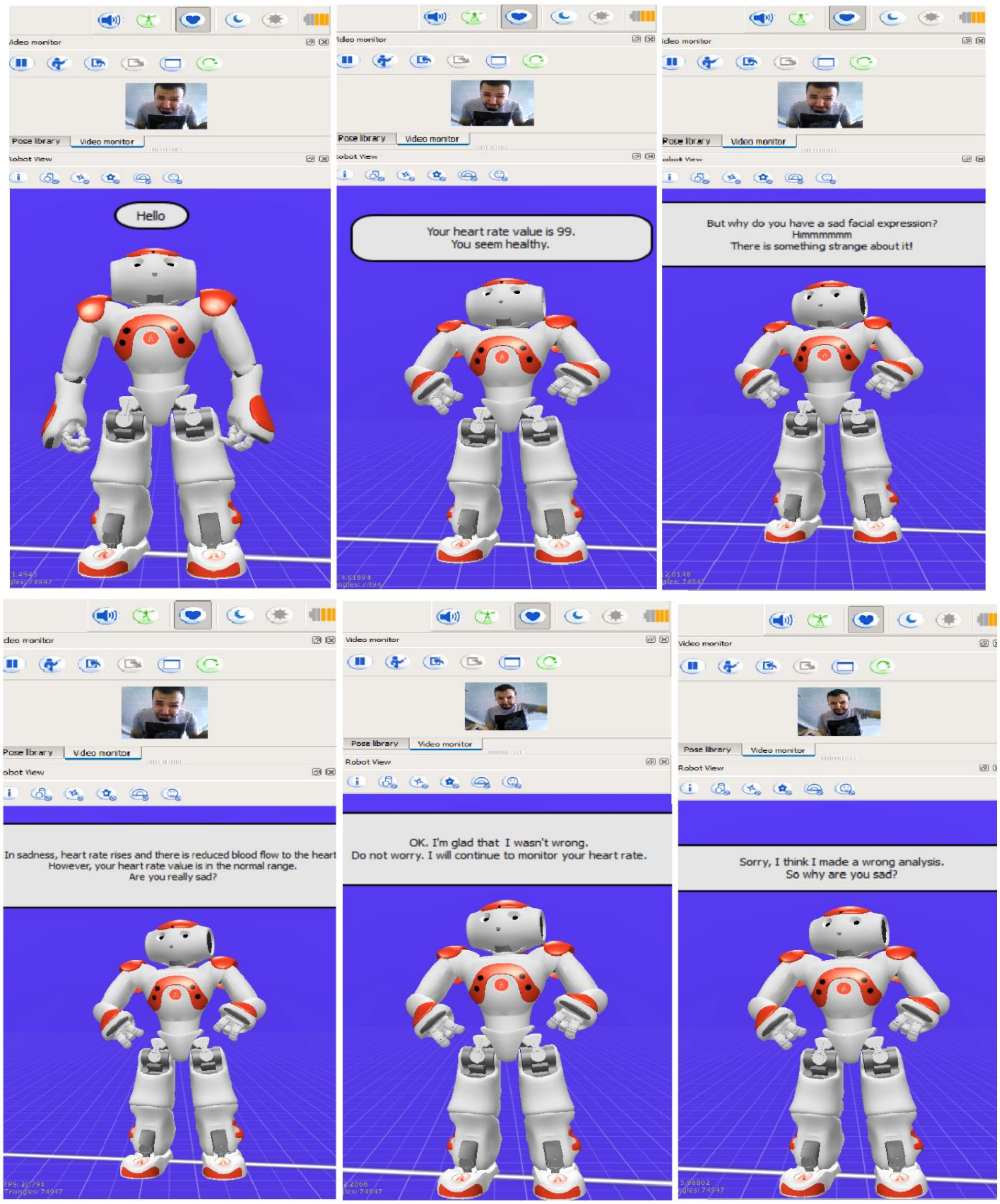


Figure 3.21. Social dialogue engagement experiment 3 based on the detected heart rate value.

Conversely, if the subject indicates that the facial expression was deceptive, the NAO HR acknowledges the successful analysis and expresses satisfaction with the results.

3.7. Discussions

In this section, a heart rate estimation algorithm is presented and implemented to a HR in real time. The algorithm begins with face detection using the Viola-Jones method and measures facial distances geometrically. It then extracts the forehead and cheek regions and tracks these using a Hierarchical Multi-Resolution Tracking approach. The algorithm processes colour information by decomposing, averaging, and normalizing RGB channels. Motion artefacts are addressed with ICA, and FFT is applied to analyse power spectra. Heart rate is determined by identifying prominent peaks in the spectrum. This section provides a Hierarchical Multi Resolution Based Tracking technique that enables HRs to track both small and large head motion variations. It provides very low estimation error and a smooth heart rate tracking performance. Additionally, detrimental impacts of the tracked head movements are eliminated using the ICA with combination of the hierarchical multi-resolution based tracking technique. Although extensive research exists on heart rate estimation using cameras and video, the application of such algorithms to social interactions with HRs is novel and has not been explored in the existing social robotics literature.

4. DESIGN AND IMPLEMENTATION OF A PERSONALIZED TEMPERATURE ESTIMATION ALGORITHM

This section introduces a new four-stage algorithm for personalized temperature estimation and its application to HRs. The first stage utilizes the Viola-Jones algorithm for accurate face detection across RGB and its complementary LAB, and mapped thermal image modalities. The second stage applies a modified grey wolf optimizer-based hierarchical vision transformer to extract relevant thermal features from these modalities. These features are then fused through a multi-modal architecture to estimate facial temperature precisely. Finally, the algorithm is integrated into Alpha Mini HR with a custom graphical user interface and a BDT-FSMs dialogue model.

4.1. Face Detection with Viola-Jones Algorithm

In the initial phase of the proposed algorithm, RGB facial images are detected in conjunction with their complementary LAB and mapped thermal image modalities using the Viola-Jones algorithm. This approach of integrating LAB and mapped thermal image modalities with RGB input face images for temperature estimation offers a robust and comprehensive method that notably improves the accuracy and reliability of the estimation process. By aligning each RGB image with its corresponding LAB and thermal data, the method ensures precise synchronization of spatial and colorimetric information.

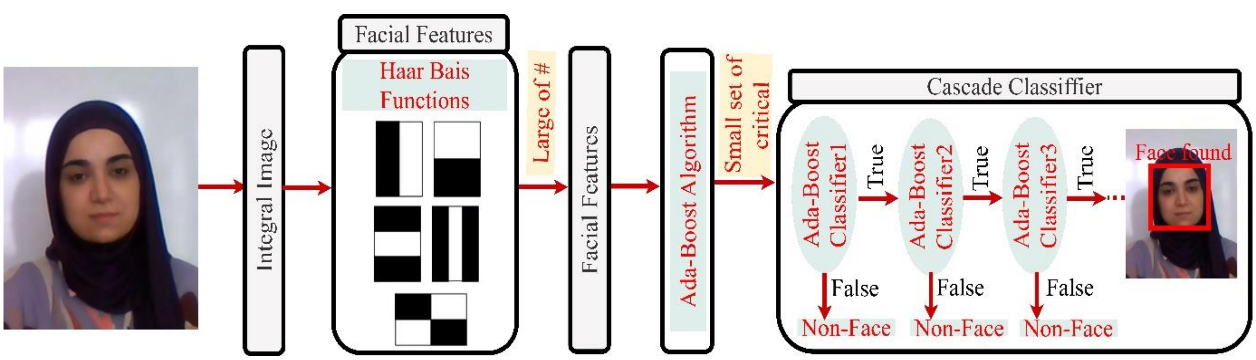


Figure 4.1. The basic architecture of the Viola-Jones algorithm.

After detailing the rationale for selecting the complementary image modalities used in this study, the Viola-Jones algorithm operates through four fundamental steps: Haar-like features, the integral image, AdaBoost, and the cascade classifier as described in the previous section and shown in the Figure 4.1.

4.2. Hierarchical Vision Transformer Architecture with Shifted-Window Mechanism

After completing the face detection stage, in the next phase, a hierarchical vision transformer mechanism shifted-window mechanism is designed with four stages as shown in the upper part of Figure 4.3. In stage 1, the detected RGB face image together with its complementary LAB and mapped thermal image modalities are concatenated to the next stage as inputs as:

$$X_{rgb} = \begin{bmatrix} x_{b_1} & x_{b_2} & \dots & x_{b_N} \end{bmatrix} \tag{4.1}$$

$$X_{lab} = \begin{bmatrix} x_{l_1} & x_{l_2} & \dots & x_{l_N} \end{bmatrix} \tag{4.2}$$

$$X_{thermal} = \begin{bmatrix} x_{t_1} & x_{t_2} & \dots & x_{t_N} \end{bmatrix} \tag{4.3}$$

where b represents the base RGB face image, l represents each LAB image modality and t represents each mapped thermal image modality, N represents the number of elements in each $X_{rgb} \in \Re^{bN \times D}$, $X_{lab} \in \Re^{lN \times D}$, and $X_{thermal} \in \Re^{tN \times D}$ sequences with ground truth temperature output labels $y \in \{35.5, 35.6, 35.7, \ldots, 45.0\}$ at 0.01 intervals.

As depicted in the second column of Figure 4.2, the proposed architecture divides each image modalities a (1), b (1), and c (1) into non-overlapping patches, resulting in a (2), b (2), and c (2). In this regard, let X_p^{rgb} , X_p^{lab}, and $X_p^{thermal}$ represent the patches for each modality as:

$$X_p^{rgb} = \begin{bmatrix} x_{p_1}^{b_1} & x_{p_2}^{b_2} & \dots & x_{p_N}^{b_N} \end{bmatrix} \tag{4.4}$$

$$X_p^{lab} = \begin{bmatrix} x_{p_1}^{l_1} & x_{p_2}^{l_2} & \dots & x_{p_N}^{l_N} \end{bmatrix} \tag{4.5}$$

$$X_p^{thermal} = \begin{bmatrix} x_{p_1}^{t_1} & x_{p_2}^{t_2} & \dots & x_{p_N}^{t_N} \end{bmatrix} \tag{4.6}$$

where each $x_{p_i}^{b_i}$, $x_{p_i}^{l_i}$, and, $x_{p_i}^{t_i}$ corresponds to a patch of the RGB, and its complementary LAB and mapped thermal image modalities, respectively. Each modality indexed from p_1 to p_N with the bias term is omitted for the sake of simplicity. Herein, the separate and simultaneous inputs to the subsequent stage can be represented as a combination of these sequences of patches as follows:

$$X_p^m = \left\{ X_p^{rgb}, X_p^{lab}, X_p^{thermal} \right\} \tag{4.7}$$

As illustrated in Figure 4.2, these patches in a (3), b(3), and c(3) are subsequently projected into 1D token vectors as:

$$X_p^m \in \Re^{N \times D} \rightarrow \left(X_p^m \right)^P \in \Re^{N \times C} \tag{4.8}$$

where C represents the projection dimension, and power P encompasses all grouped modalities' patches, as illustrated below:

$$\left(X_p^m\right)^P = \left\{\left(X_p^{rgb}\right)^P, \left(X_p^{lab}\right)^P, \left(X_p^{thermal}\right)^P\right\} \tag{4.9}$$

with

$$\left(X_p^{rgb}\right)^P = W_{rgb} X_p^{rgb} \tag{4.10}$$

$$\left(X_p^{lab}\right)^P = W_{lab} X_p^{lab} \tag{4.11}$$

$$\left(X_p^{thermal}\right)^P = W_{thermal} X_p^{thermal} \tag{4.12}$$

Upon the completion of this 1D projection procedure, the primary transformer block is activated in the next stage as depicted in Figure 4.4 (a).

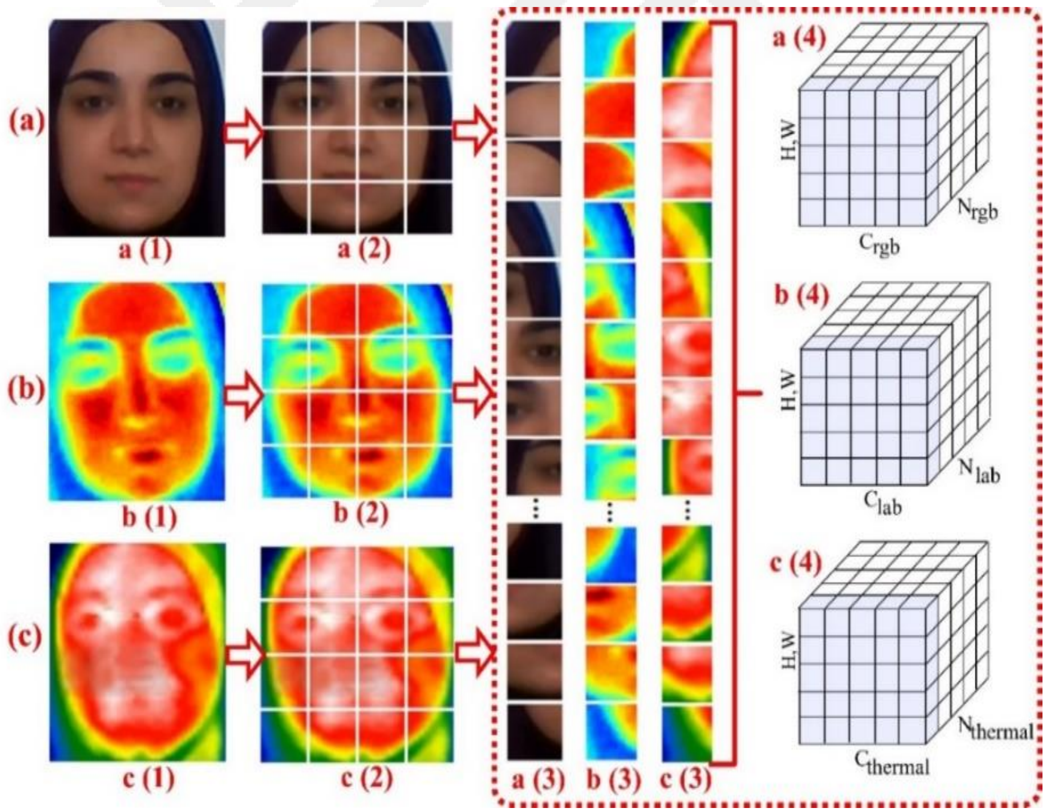


Figure 4.2. Demonstration of the proposed concatenation, patch partitioning, and linear projection methodologies applied to a) detected RGB facial image and its complementary b) LAB, and c) mapped thermal image modalities respectively.

Initially, a modified layer normalization is applied to each patch $\left(X_p^m\right)^P$ within each image modality. This step identifies regions with elevated energy levels within the projected patches, which may indicate significant temperature variations. Integrating these energy metrics into the layer normalization process enhances the architecture's ability to detect and differentiate thermal patterns characterized by intensity variations across distinct facial regions, each reflecting specific temperature gradients. This increased sensitivity to subtle variations is expected to improve the accuracy of temperature estimation from facial features.

Following this, the next step involves quantifying the energy of each projected patch modality. This quantification entails the summation of squared values contained within the corresponding projected patches as:

$$\zeta_p = \sum_{i=1}^{C}\left(\left(X_{p_i}^m\right)^P\right)^2 \tag{4.13}$$

Subsequently, the mean and variance values of each projected patch modality are determined as:

$$\mu_p^m = \frac{1}{C}\sum_{i=1}^{C}\left(\left(X_{p_i}^m\right)^P\right) \tag{4.14}$$

and

$$\left(\eta_p^m\right)^2 = \frac{1}{C}\sum_{i=1}^{C}\left(\left(\left(X_{p_i}^m\right)^P\right) - \mu_p^m\right)^2 \tag{4.15}$$

from here, they combined with ζ_p as:

$$\left(X_{p_i}^m\right)^P = \frac{\left(\left(X_{p_i}^m\right)^P + \zeta_p\right) - \mu_p}{\sqrt{\left(\eta_p^m\right)^2 + \varepsilon}} \tag{4.16}$$

wherein, ε is a small constant introduced to prevent division by zero, and

$$\left(X_{p_i}^m\right)^P = \left\{\left(X_{p_i}^{rgb}\right)^P, \left(X_{p_i}^{lab}\right)^P, \left(X_{p_i}^{thermal}\right)^P\right\} \tag{4.17}$$

are obtained respectively. The final modified normalized layer is then derived as follows:

$$LN\left(\left(X_{p_i}^m\right)^P\right) = \gamma\left(X_{p_i}^m\right)^P + \beta \tag{4.18}$$

where LN denotes layer normalization, γ and β are learnable parameters.

To enable the thorough integration of all N entities, then the normalized input patch sequence $LN\left(\left(X_p^m\right)^P\right)$ is projected onto three learnable unknown parameter matrices. This process transforms each input tensor into three distinct tensors: query, key, and value, each with a concatenation size of 4x4 respectively.

This transformation is initially applied to the RGB image modality as follows:

$$Q_p^{rgb} = W_{Q_p}^{rgb} LN\left(\left(X_p^{rgb}\right)^P\right) \tag{4.19}$$

$$K_p^{rgb} = W_{K_p}^{rgb} LN\left(\left(X_p^{rgb}\right)^P\right) \tag{4.20}$$

$$V_p^{rgb} = W_{V_p}^{rgb} LN\left(\left(X_p^{rgb}\right)^P\right) \tag{4.21}$$

with

$$W_{Q_p}^{rgb} = \begin{bmatrix} w_{Q_{p_1}}^{b_1} & w_{Q_{p2}}^{b_2} & \dots & w_{Q_{p_N}}^{b_N} \end{bmatrix} \tag{4.22}$$

$$W_{K_p}^{rgb} = \begin{bmatrix} w_{K_{p_1}}^{b_1} & w_{K_{p2}}^{b_2} & \dots & w_{K_{p_N}}^{b_N} \end{bmatrix} \tag{4.23}$$

$$W_{V_p}^{rgb} = \begin{bmatrix} w_{V_{p_1}}^{b_1} & w_{V_{p2}}^{b_2} & \dots & w_{V_{p_N}}^{b_N} \end{bmatrix} \tag{4.24}$$

Subsequently, analogous computations are performed for the LAB image modality as follows:

$$Q_p^{lab} = W_{Q_p}^{lab} LN\left(\left(X_p^{lab}\right)^P\right) \tag{4.25}$$

$$K_p^{lab} = W_{K_p}^{lab} LN\left(\left(X_p^{lab}\right)^P\right) \tag{4.26}$$

$$V_p^{lab} = W_{V_p}^{lab} LN\left(\left(X_p^{lab}\right)^P\right) \tag{4.27}$$

with

$$W_{Q_p}^{lab} = \begin{bmatrix} w_{Q_{p_1}}^{l_1} & w_{Q_{p2}}^{l_2} & \dots & w_{Q_{p_N}}^{l_N} \end{bmatrix} \tag{4.28}$$

$$W_{K_p}^{lab} = \begin{bmatrix} w_{K_{p_1}}^{l_1} & w_{K_{p2}}^{l_2} & \dots & w_{K_{p_N}}^{l_N} \end{bmatrix} \tag{4.29}$$

$$W_{V_p}^{lab} = \begin{bmatrix} w_{V_{p_1}}^{l_1} & w_{V_{p2}}^{l_2} & \dots & w_{V_{p_N}}^{l_N} \end{bmatrix} \tag{4.30}$$

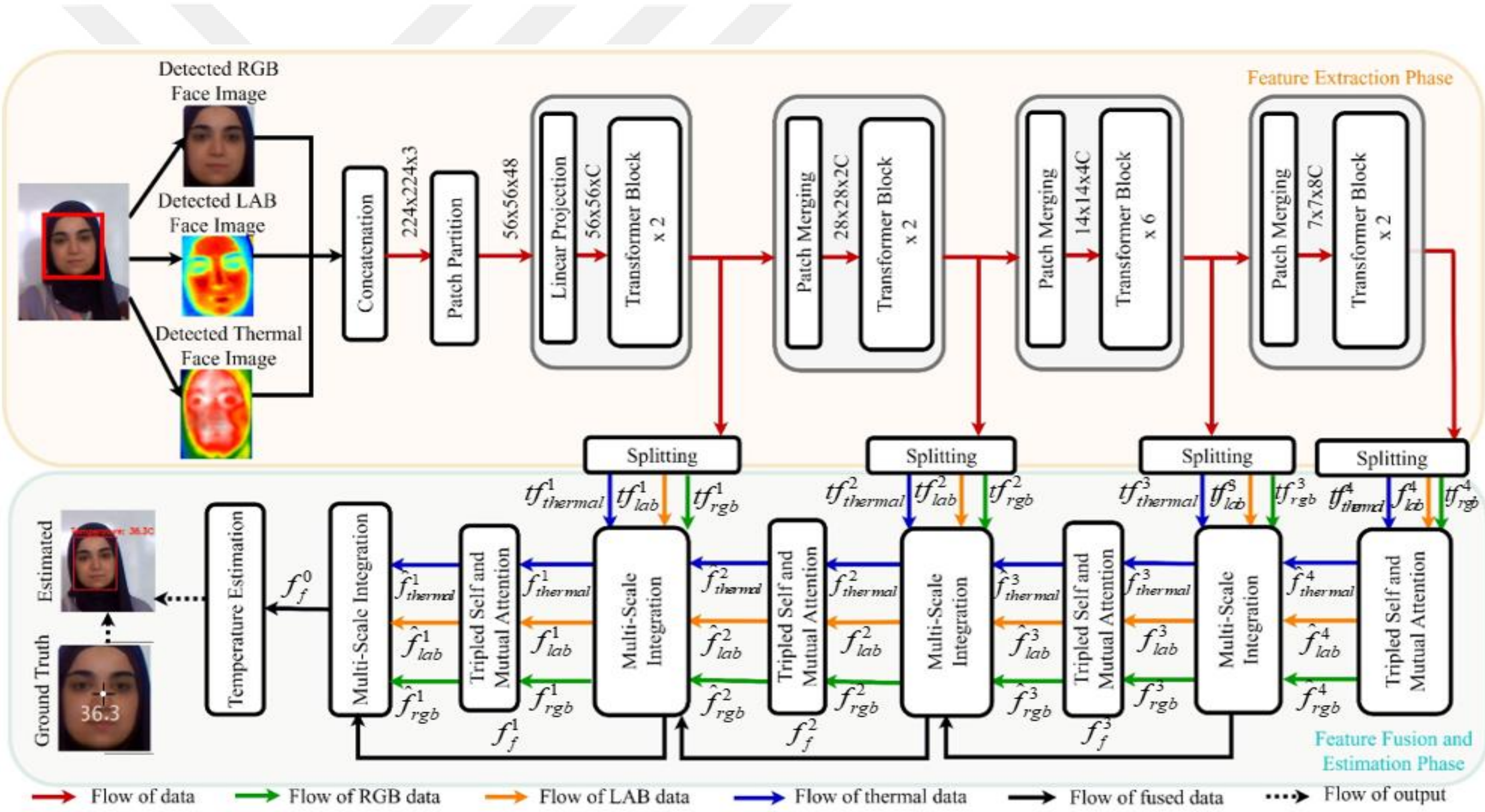


Figure 4.3. Proposed personalized temperature estimation algorithm utilizing a swarm intelligence- based hierarchical vision transformer architecture with shifted-window mechanism and multi-modal fusion architecture.

and, finally the equivalent procedures are applied to the mapped thermal image modality as:

$$Q_p^{thermal} = W_{Q_p}^{thermal} LN\left(\left(X_p^{thermal}\right)^P\right) \tag{4.31}$$

$$K_p^{thermal} = W_{K_p}^{thermal} LN\left(\left(X_p^{thermal}\right)^P\right) \tag{4.32}$$

$$V_p^{thermal} = W_{V_p}^{thermal} LN\left(\left(X_p^{thermal}\right)^P\right) \tag{4.33}$$

with

$$W_{Q_p}^{thermal} = \begin{bmatrix} w_{Q_{p1}}^{t_1} & w_{Q_{p2}}^{t_2} & \dots & w_{Q_{pN}}^{t_N} \end{bmatrix} \tag{4.34}$$

$$W_{K_p}^{thermal} = \begin{bmatrix} w_{K_{p1}}^{t_1} & w_{K_{p2}}^{t_2} & \dots & w_{K_{pN}}^{t_N} \end{bmatrix} \tag{4.35}$$

$$W_{V_p}^{thermal} = \begin{bmatrix} w_{V_{p1}}^{t_1} & w_{V_{p2}}^{t_2} & \dots & w_{V_{pN}}^{t_N} \end{bmatrix} \tag{4.36}$$

In these processes, the relevance score derived from the query tensor (Q) is utilized to assess the significance of each value vector relative to the query. The key tensor (K) serves as a reference for measuring the similarity between itself and the query tensor. Each K is compared to the query tensor using a scaled dot-product approach, which generates a score indicating the importance of the corresponding value to the query. The value tensor (V) contains the actual information associated with each input element. By integrating these components, a normalized one-headed tensor/self-attention mechanism is generated for each modality as:

$$H_{rgb}^{SA} = \sigma\left(\frac{Q_p^{rgb}\left(K_p^{rgb}\right)^T}{\sqrt{d}} + B\right)V_p^{rgb} \tag{4.37}$$

$$H_{lab}^{SA} = \sigma\left(\frac{Q_p^{lab}\left(K_p^{lab}\right)^T}{\sqrt{d}} + B\right)V_p^{lab} \tag{4.38}$$

$$H_{thermal}^{SA} = \sigma\left(\frac{Q_p^{thermal}\left(K_p^{thermal}\right)^T}{\sqrt{d}} + B\right)V_p^{thermal} \tag{4.39}$$

where B denotes the relative position bias and d denotes a scaling factor.

Subsequently, each computed self-attention mechanism is extended into multi-head self-attention mechanisms. This extension allows the attention mechanism to account for multiple attention distributions, thereby enabling the architecture to capture information from diverse perspectives.

In this process, the attention function is executed h times simultaneously and the resulting outputs are concatenated to form the multi-head self-attention mechanism as:

$$MH_{rgb}^{SA} = \left\lfloor H_{rgb}^{SA_1} \quad H_{rgb}^{SA_2} \quad \dots \quad H_{rgb}^{SA_N} \right\rfloor W_{rgb}^{o} \tag{4.40}$$

and,

$$MH_{lab}^{SA} = \left\lfloor H_{lab}^{SA_1} \quad H_{lab}^{SA_2} \quad \dots \quad H_{lab}^{SA_N} \right\rfloor W_{lab}^{o} \tag{4.41}$$

finally,

$$MH_{thermal}^{SA} = \left\lfloor H_{thermal}^{SA_1} \quad H_{thermal}^{SA_2} \quad \dots \quad H_{thermal}^{SA_N} \right\rfloor W_{thermal}^{o} \tag{4.42}$$

where $\lfloor \bullet \rfloor$ denotes the concatenation of the outputs of each H_{rgb}^{SA}, H_{lab}^{SA}, and $H_{thermal}^{SA}$ self-attention heads and finally, W_{rgb}^{o}, W_{lab}^{o}, and $W_{thermal}^{o}$ are learnable unknown parameters, which allow the architecture to effectively combine the outputs from individual attention heads.

For the sake of clarity, these formulations can be expressed in a generalized form as presented in the following equations:

$$Q_p^m = W_Q^m LN\left(X_p^m\right)^P \tag{4.43}$$

$$K_p^m = W_K^m LN\left(X_p^m\right)^P \tag{4.44}$$

$$V_p^m = W_V^m LN\left(X_p^m\right)^P \tag{4.45}$$

with

$$W_Q^m = \left\{W_{Q_p}^{rgb}, W_{Q_p}^{lab}, W_{Q_p}^{thermal}\right\}, \tag{4.46}$$

$$W_K^m = \left\{W_{K_p}^{rgb}, W_{K_p}^{lab}, W_{K_p}^{thermal}\right\}, \tag{4.47}$$

and

$$W_V^m = \left\{W_{V_p}^{rgb}, W_{V_p}^{lab}, W_{V_p}^{thermal}\right\} \tag{4.48}$$

and, finally

$$H_m^{SA} = \sigma\left(\frac{Q_p^m \left(K_p^m\right)^T}{\sqrt{d}} + B\right) V_p^m \tag{4.49}$$

is obtained for one head with

$$H_m^{SA} = \left\{H_{rgb}^{SA}, H_{lab}^{SA}, H_{thermal}^{SA}\right\} \tag{4.50}$$

with its multi-head representation as:

$$MH_m^{SA} = \left\{ MH_{rgb}^{SA}, MH_{lab}^{SA}, MH_{thermal}^{SA} \right\} \tag{4.51}$$

As demonstrated in Equation (49), the final step involves multiplying the normalized tensor by its corresponding value tensor using a scaled-dot product operation. This process finalizes the self-attention mechanism for a single head (as shown in Figure 4.4 (d)). The multi-head self-attention mechanism is subsequently applied, facilitating nuanced feature extraction across multiple modalities (Figure 4.4 (c)) through the implementation of the window-based multi-head self-attention ($w\_MH_m^{SA}$) and shifted window-based multi-head self-attention ($sw\_MH_m^{SA}$) mechanisms, as illustrated in Figure 4.4 (a).

Herein, the $w\_MH_m^{SA}$ mechanism is pivotal in capturing contextual dependencies across various regions of an image. This mechanism divides the feature map into M × M local windows and computes the self-attention output within each window. By restricting the self-attention computation to patches within individual windows, $w\_MH_m^{SA}$ effectively reduces computational complexity while maintaining the ability to identify correlations among neighbouring pixels within each image modality. This approach enhances both computational efficiency and estimation accuracy by focusing attention on relevant local regions of the image.

Following the computational process of $w\_MH_m^{SA}$, as depicted in the initial layer of Figure 4(a), $\overline{f}^{l}$ is subsequently computed in the next stage as:

$$\overline{f}^{l} = w\_\left(MH_m^{SA}\right)^{l-1} + \left(LN\left(\left(X_{p_i}^{m}\right)^{P}\right)\right)^{l-1} \tag{4.52}$$

To complete the computations related to the initial layer of the proposed transformer block, $\overline{\overline{f}}^{l}$ is forwarded to the 2-layer multilayer perceptron (MLP) head module, as depicted in Figure 4.4 (b), and is computed as follows:

$$\overline{\overline{f}}^{l} = \delta\left(W_1^{h} LN\left(\overline{f}^{l}\right)\right) W_2^{h} + \overline{f}^{l} \tag{4.53}$$

where δ represents Gaussian Error Linear Units (GELU) activation function, W_1^{h} and W_2^{h} denote the learnable parameters associated with the first layer of the MLP.

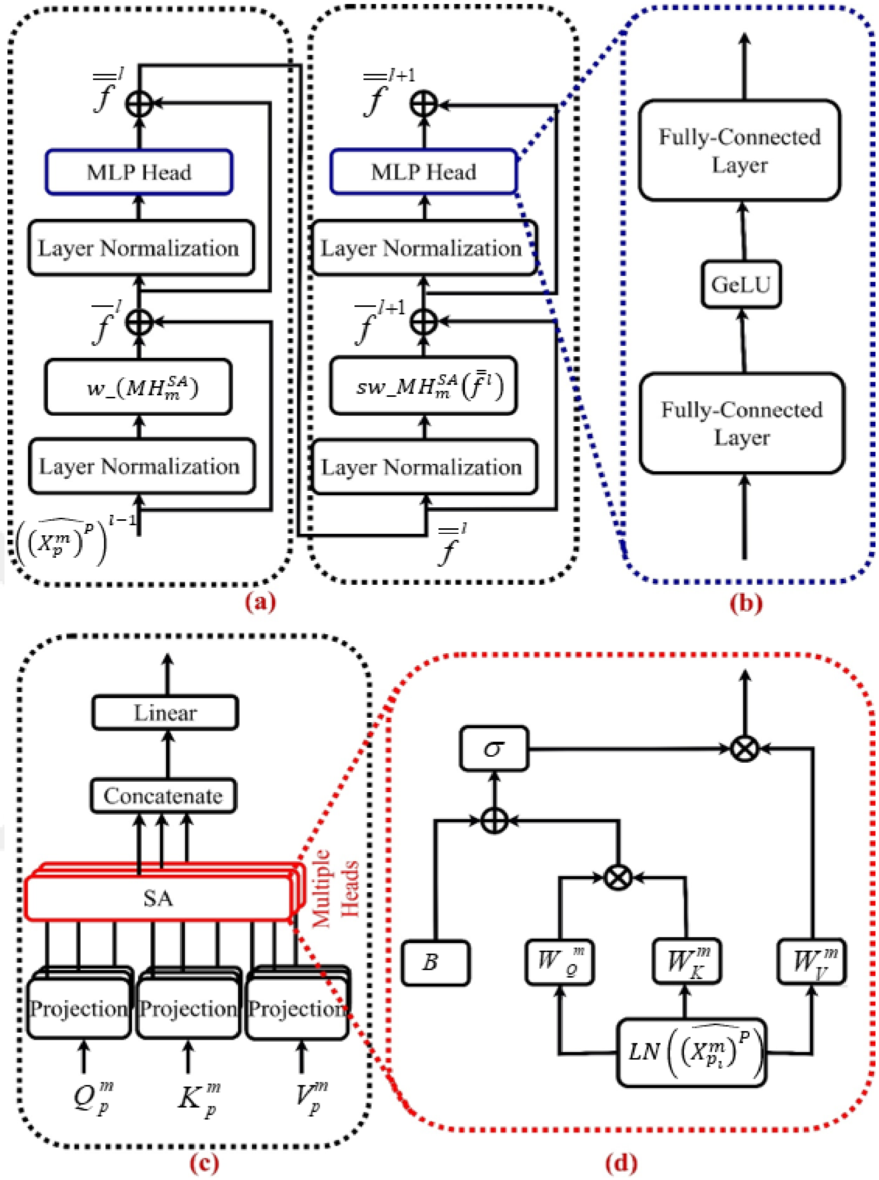


Figure 4.4. Proposed hierarchical vision transformer with shifted-window: Depicting a) Successive blocks, b) MLP head, c) MHSA, and d) SA Modules.

Despite the effectiveness of the computation outlined in equation (4.53), the $w\_MH_m^{SA}$ framework exhibits limitations in facilitating information exchange between non-overlapping windows. This shortcoming can lead to inaccuracies in the extraction of temperature features spread across distinct windows. To address this issue, the $sw\_MH_m^{SA}$ introduces a cyclic shifting operation coupled with mask padding. Specifically, the feature map is circularly shifted along both the height and width axes by a distance equal to half the patch size. This process ensures that neighbouring patches, previously confined to separate windows, are now integrated within the same window, enabling more comprehensive feature extraction.

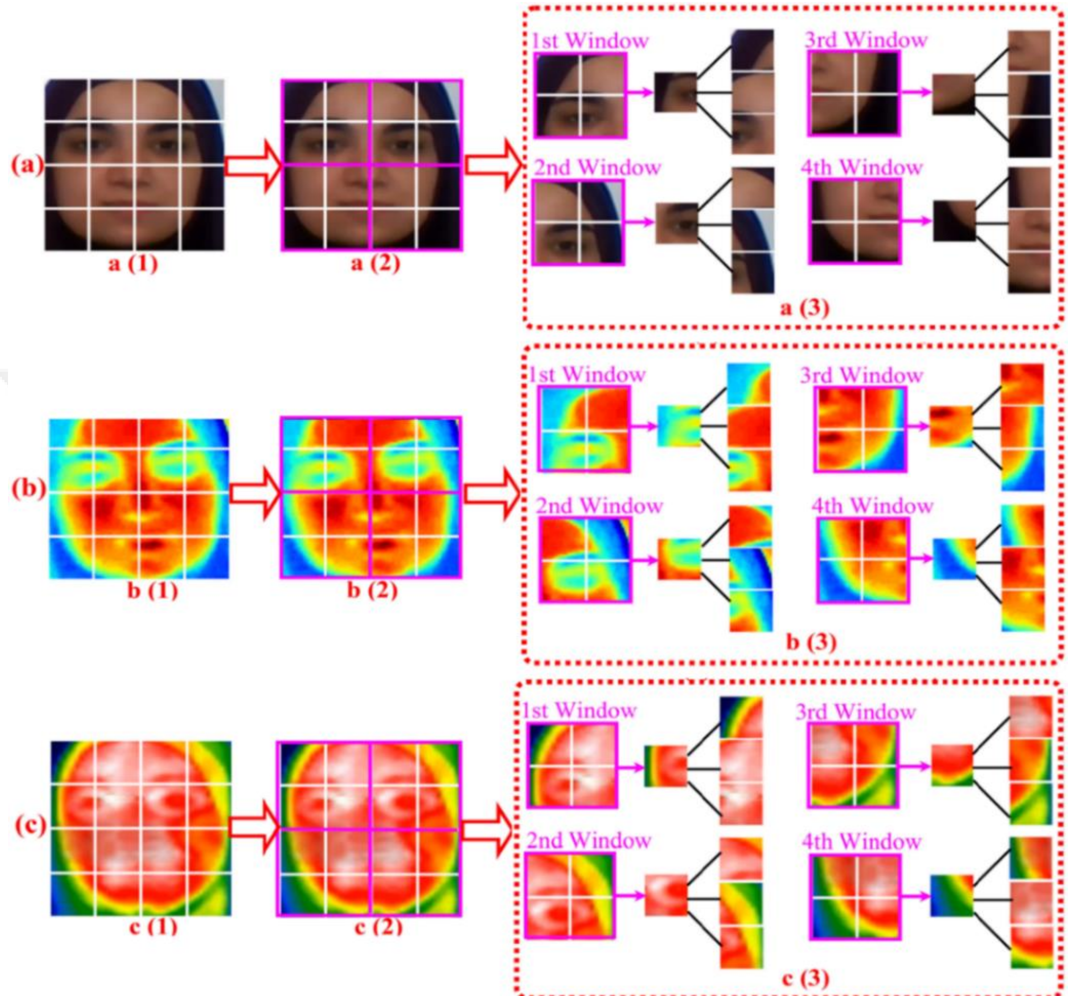


Figure 4.5. Visualization of the proposed $w\_MH_m^{SA}$ architecture applied to a) partitioned RGB image and its complementary b) LAB, and c) mapped thermal image modalities.

After applying $w\_MH_m^{SA}$ to the shifted feature map, a reverse circular shifting operation is used to return the feature map to its original alignment, as illustrated in Figure 4.6. This approach significantly enhances the architecture's ability to capture contextual dependencies and improves the accuracy of temperature feature extraction. To further enhance clarity and understanding, detailed illustrations of the cyclic shift operations within the $sw\_MH_m^{SA}$ framework are provided. Within this framework, the mechanism expands the four original windows of W-MSA to nine, as depicted in Figures 4.7 (a1), (b1), and (c1). The numerical labels from 1 to 9 correspond to the specific windows within this expanded layout. Following the subdivision by $sw\_MH_m^{SA}$, variations in window sizes require modifications to the window configuration.

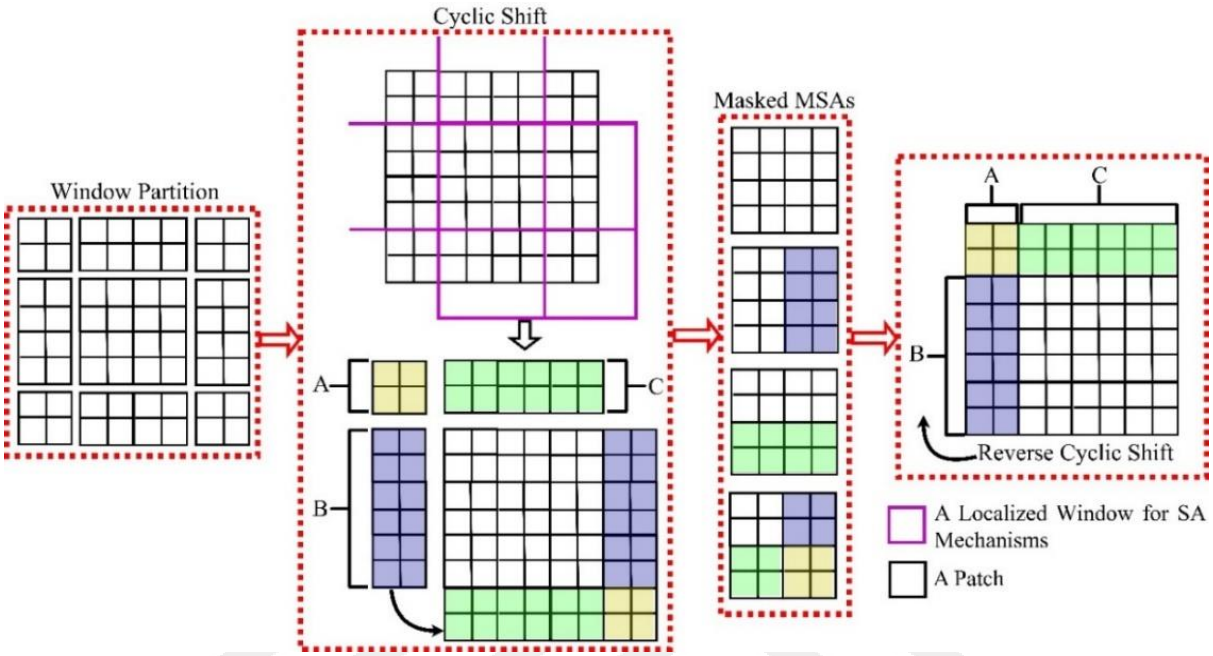


Figure 4.6. Illustration of the proposed $sw\_MH_m^{SA}$ architecture.

To implement this adjustment, the first row of the window is shifted to the fourth row, as shown in (a2), (b2), and (c2), while the fourth row is moved to the first row, as illustrated in (a3), (b3), and (c3). Subsequently, the first column is shifted to the fourth column, as depicted in (a4), (b4), and (c4), with the fourth column then replacing the first column, as demonstrated in (a5), (b5), and (c5). Finally, as illustrated in (a6), (b6), and (c6), the four masked multi-head self-attention operations conducted within the "larger" windows yield an equivalent computational effort to that of nine multi-head self-attention operations within the "smaller" windows. This process, as previously described, results in a notable reduction in computational overhead.

Following the detailed computational procedure established for $sw\_MH_m^{SA}$, the calculation of $\overline{f}^{l+1}$ is carried out in the subsequent stage, as depicted in the second block of Figure 4. 4 (a), and is defined as:

$$\overline{f}^{l+1} = sw\_MH_m^{SA}\left(\overline{\overline{f}}^{l}\right) + LN\left(\overline{\overline{f}}^{l}\right) \tag{4.54}$$

To finalize the computations for the second layer of the proposed transformer block, $\overline{\overline{f}}^{l+1}$ is again directed to the 2-layer MLP head module and processed as follows:

$$\bar{\bar{f}}^{l+1} = \delta\left(W_1^o LN\left(\bar{f}^{l+1}\right)\right)W_2^o + \bar{f}^{l+1} \tag{4.55}$$

where W_1^o and W_2^o represent the learnable parameters associated with the second layer of the MLP.

After computing the two successive layers within the initial transformer block, the resulting feature map is propagated through the subsequent transformer blocks, from the second to the fourth, as illustrated in Figure 4.3. Herein, these blocks utilize consistent calculation procedures. As the architecture's depth increases, the number of tokens is progressively reduced via patch merging layers, thereby constructing a hierarchical representation for each image modality.

Upon completing the computations across all transformer blocks, the dimensionalities of the flattened output features for each block are denoted as $T_1 \times C$, $T_2 \times 2C$, $T_3 \times 4C$ and $T_4 \times 8C$ alongside $T_1 = \frac{H}{4} \times \frac{W}{4}$, $T_2 = \frac{H}{8} \times \frac{W}{8}$, $T_3 = \frac{H}{16} \times \frac{W}{16}$, and $T_4 = \frac{H}{32} \times \frac{W}{32}$ (Please refer to Table 4.7 in the "Results" sub-section for detailed descriptions of the other primary parameters). These features are subsequently fused and passed on to the next section of the proposed architecture for further processing.

4.3. Development of A Multi-Modal Fusion Architecture

To further propagate the global contextual information, an advanced tripled self-and-mutual attention (TSMA) mechanism is introduced to capture intra-modal and multi-modal interactions across three modalities, as depicted in the lower section of Figure 4.3. As illustrated in Figure 4.9, the TSMA mechanism comprises two primary attention components: one for processing single-modal data and another for multi-head mutual attention to handle multi-modal data. In this framework, the initial inputs to the first TSMA layer are denoted as $tf_m^4 \in \left\{\bar{\bar{f}}_{rgb}, \bar{\bar{f}}_{lab}, \bar{\bar{f}}_{thermal}\right\}$. In this context, a projection operation is first applied to the tf_m^4 features to derive the corresponding Q_4^m, K_4^m, and V_4^m as:

$$Q_4^m = W_{Q_4}^m LN\left(tf_m^4\right) \tag{4.56}$$

$$K_4^m = W_{K_4}^m LN\left(tf_m^4\right) \tag{4.57}$$

$$V_4^m = W_{V_4}^m LN\left(tf_m^4\right) \tag{4.58}$$

with

$$Q_4^m = \left\{ Q_4^{rgb}, Q_4^{lab}, Q_4^{thermal} \right\}, \tag{4.59}$$

$$K_4^m = \left\{ K_4^{rgb}, K_4^{lab}, K_4^{thermal} \right\}, \tag{4.60}$$

$$V_4^m = \left\{ V_4^{rgb}, V_4^{lab}, V_4^{thermal} \right\} \tag{4.61}$$

and $W_{Q_4}^m = \left\{ W_{Q_4}^{rgb}, W_{Q_4}^{lab}, W_{Q_4}^{thermal} \right\}$, (4.62)

$$W_{K_4}^m = \left\{ W_{K_4}^{rgb}, W_{K_4}^{lab}, W_{K_4}^{thermal} \right\}, \tag{4.63}$$

$$W_{V_4}^m = \left\{ W_{V_4}^{rgb}, W_{V_4}^{lab}, W_{V_4}^{thermal} \right\} \tag{4.64}$$

Subsequently, the mutual "scaled dot-product attention" is computed by utilizing Q s from one modality and K s from other modalities, integrated with the multi-head self-attention results from the preceding section. Specifically, the Q from the RGB modality and the K s from the LAB and mapped thermal modalities of tf_m^4 are considered and calculated from the last 4th transformer block as follows:

$$MH_{rgb}^{MA1_4}\left(Q_4^{rgb}, K_4^{lab}, V_4^{lab}\right) = \sigma\left(\frac{Q_4^{rgb}\left(K_4^{lab}\right)^T}{\sqrt{d}} + B\right)V_4^{lab} \tag{4.65}$$

$$MH_{rgb}^{MA2_4}\left(Q_4^{rgb}, K_4^{thermal}, V_4^{thermal}\right) = \sigma\left(\frac{Q_4^{rgb}\left(K_4^{thermal}\right)^T}{\sqrt{d}} + B\right)V_4^{thermal} \tag{4.66}$$

Then, Q s from the LAB image modality is integrated with K s from the RGB and mapped thermal image modalities, and the computations are conducted as follows:

$$MH_{lab}^{MA1_4}\left(Q_4^{lab}, K_4^{rgb}, V_4^{rgb}\right) = \sigma\left(\frac{Q_4^{lab}\left(K_4^{rgb}\right)^T}{\sqrt{d}} + B\right)V_4^{rgb} \tag{4.67}$$

$$MH_{lab}^{MA2_4}\left(Q_4^{lab}, K_4^{thermal}, V_4^{thermal}\right) = \sigma\left(\frac{Q_4^{lab}\left(K_4^{thermal}\right)^T}{\sqrt{d}} + B\right)V_4^{thermal} \tag{4.68}$$

In the final stage, Q s derived from the mapped thermal image modality are systematically combined with K s from the RGB and LAB image modalities, and related computations are performed as follows:

$$MH_{thermal}^{MA1_4}\left(Q_4^{thermal}, K_4^{rgb}, V_4^{rgb}\right) = \sigma\left(\frac{Q_4^{thermal}\left(K_4^{rgb}\right)^T}{\sqrt{d}} + B\right)V_4^{rgb} \tag{4.69}$$

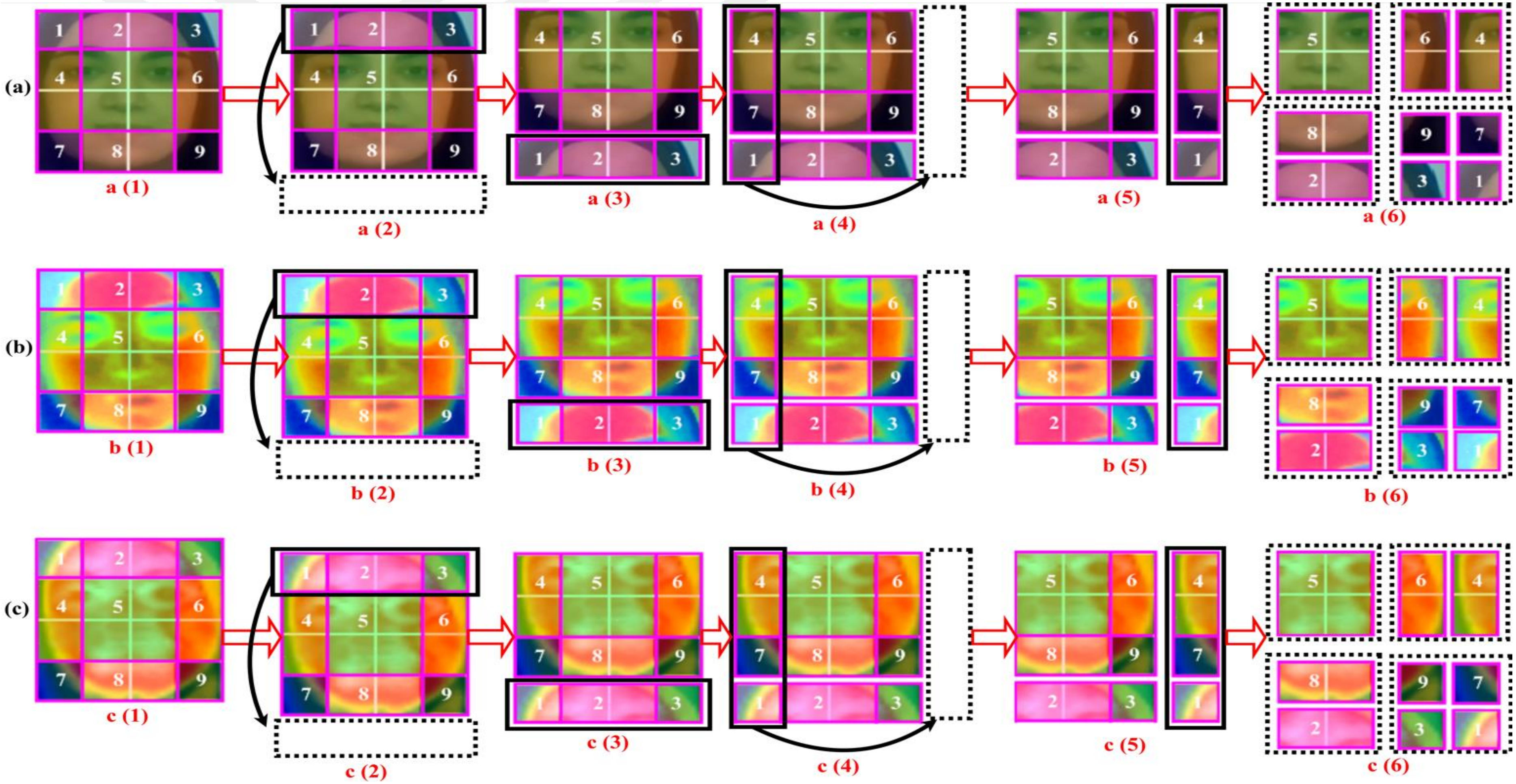


Figure 4.7. Calculation of shift configurations within $sw\_MH_m^{SA}$ for a) partitioned RGB facial images alongside its complementary b) LAB, and c) mapped thermal image modalities.

$$MH_{thermal}^{MA2_4}\left(Q_4^{thermal}, K_4^{lab}, V_4^{lab}\right) = \sigma\left(\frac{Q_4^{thermal}\left(K_4^{lab}\right)^T}{\sqrt{d}} + B\right)V_4^{lab} \tag{4.70}$$

and from here,

$MH_{mm}^{MA_4} = \left\{MH_{rgb}^{MA1_4}, MH_{rgb}^{MA2_4}, MH_{lab}^{MA1_4}, MH_{lab}^{MA2_4}, MH_{thermal}^{MA1_4}, MH_{thermal}^{MA2_4}\right\}$ is obtained. (4.71)

With this final computation, the enhanced multi-modal features of tf_m^4 are derived and subsequently processed as follows:

$$f_{m_4}^{'} = MH_{mm}^{MA_4} + tf_m^4 \tag{4.72}$$

where $f_{m_4}^{'} \in \left\{f_{rgb_4}^{'}, f_{lab_4}^{'}, f_{thermal_4}^{'}\right\}$ (4.73)

and $f_{m_4}^{''} = \delta\left(W_1^{fh} LN\left(f_{m_4}^{'}\right)\right)W_2^{fh} + f_{m_4}^{'}$ (4.74)

where W_1^{fh} and W_2^{fh} denote the unknown parameters associated with the first layer of $f_{m_4}^{''}$, and

$$f_{m_4}^{''} \in \left\{f_{rgb_4}^{''}, f_{lab_4}^{''}, f_{thermal_4}^{''}\right\} \tag{4.75}$$

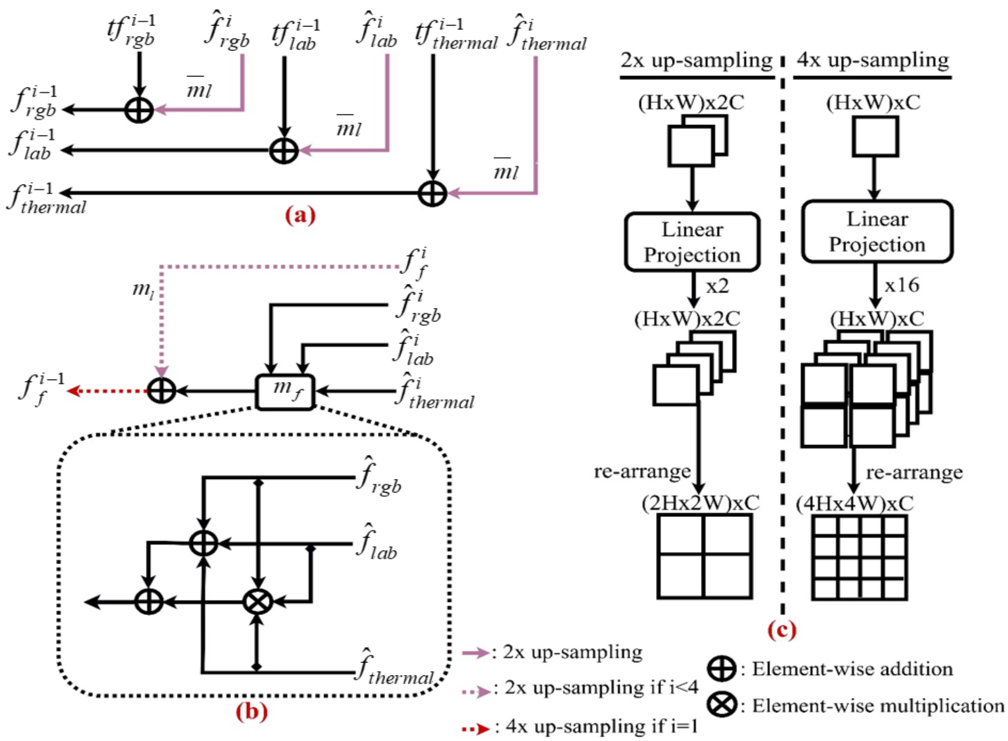


Figure 4.8. Visual depiction of a) the proposed multi-scale integration module with, b) m_f, and c) patch expanding modules respectively.

Finally, f''_{m_4} undergoes processing through a concatenation operation, followed by a normalization layer and an MLP head module, to yield the final fused features f^4_m for the input tf^4_m. This process is expressed as:

$$f^4_m = \delta\left(W^{fo}_1 LN\left(f''_{m_4}\right)\right)W^{fo}_2 \tag{4.76}$$

where W^{fo}_1 and W^{fo}_2 represent the unknown parameters associated with the second layer of f''_{m_4} and

$$f^4_m \in \left\{f^4_{rgb}, f^4_{lab}, f^4_{thermal}\right\} \tag{4.77}$$

Likewise, the other

$$f^3_m \in \left\{f^3_{rgb}, f^3_{lab}, f^3_{thermal}\right\}, \tag{4.78}$$

$$f^2_m \in \left\{f^2_{rgb}, f^2_{lab}, f^2_{thermal}\right\}, \tag{4.79}$$

$$f^1_m \in \left\{f^1_{rgb}, f^1_{lab}, f^1_{thermal}\right\} \tag{4.80}$$

final fused features are derived using exactly the same computational procedures. Details of these computations are omitted in this section to conserve space within the paper.

Since the features learned at the higher levels of the encoder contain more detailed temperature information, incorporating this higher-level temperature information into the lower-level features can effectively guide the feature fusion process, thereby reducing interference from irrelevant regions. Additionally, simple feature summation or concatenation methods are limited to linear fusion of the input image modalities' features, and thus fail to integrate multi-modal features effectively. To address these limitations, multi-scale integration (m_f) modules are introduced, utilizing a top-down approach to achieve more effective and robust feature fusion. Thus, the m_f process is carried out for each determined final fused feature and resulting in:

$$m_f\left(f^i_{rgb}, f^i_{lab}, f^i_{thermal}\right) = f^i_{rgb} \oplus f^i_{lab} \oplus f^i_{thermal} \oplus \left(f^i_{rgb} \otimes f^i_{lab} \otimes f^i_{thermal}\right) \tag{4.81}$$

Subsequently, the fused features from higher layers are introduced into adjacent lower layers through skip connections, thereby enhancing the fusion process, as depicted in Figures 4.8 (a) and (b).

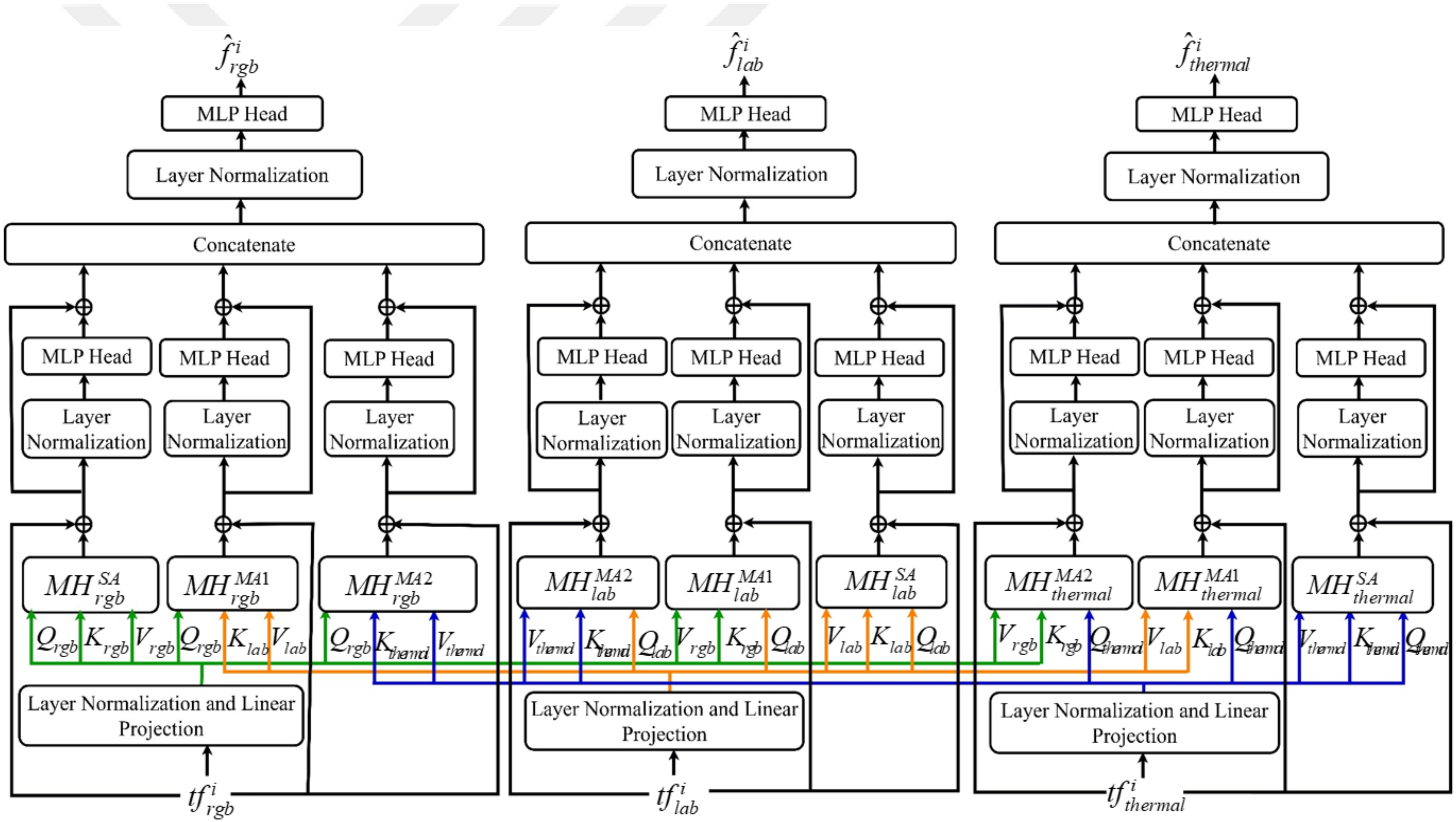


Figure 4.9. Schematic representation of the proposed TSMA architecture.

These high-level fused features are then up-sampled using a patch expanding layer before being integrated into the corresponding lower layers as follows:

$$f_f^3 = m_f\left(f_{rgb}^4, f_{lab}^4, f_{thermal}^4\right) \tag{4.82}$$

$$f_f^2 = m_f\left(f_{rgb}^3, f_{lab}^3, f_{thermal}^3\right) \oplus s_{\times 2}^{up}\left(f_f^3\right) \tag{4.83}$$

$$f_f^1 = m_f\left(f_{rgb}^2, f_{lab}^2, f_{thermal}^2\right) \oplus s_{\times 2}^{up}\left(f_f^2\right) \tag{4.84}$$

$$f_f^0 = s_{\times 4}^{up}\left(m_f\left(f_{rgb}^1, f_{lab}^1, f_{thermal}^1\right) \oplus s_{\times 2}^{up}\left(f_f^1\right)\right) \tag{4.85}$$

where $s_{\times 2}^{up}$ and $s_{\times 4}^{up}$ indicates 2× and 4× up-sampling processes.

Once f_f^0 is obtained, it is fed into a final estimation layer $\bar{y}$ with an additional multi-level guidance (m_l) fusion operation as:

$$\bar{y} = \sigma\left(W_1^{ft} LN\left(s_{\times 2}^{up}\left(f_f^0\right)\right)\right) W_2^{ft} + s_{\times 2}^{up}\left(f_f^0\right) \tag{4.86}$$

Then, training error e is computed as:

$$e = y - \bar{y} \tag{4.87}$$

Following this, the cost function J to be minimized is calculated as:

$$J = \min\left(ee^T\right) \tag{4.88}$$

Finally, the next section presents a swarm intelligence-based optimization of the proposed learning architecture in equation (86).

4.4. Cost Function Optimization

This section introduces a conventional GWO and subsequently explains the proposed M-GWO for optimizing the cost function of the developed architecture. Herein, the M-GWO iteratively chooses the optimum unknown parameters that reduces the minimum cost function in Equation (88).

4.4.1. Conventional GWO

The conventional GWO draws inspiration from the hierarchical hunting behaviour of grey wolves, characterized by a structured social hierarchy comprising alpha (α), beta (β), delta (δ), and omega (ω) wolves. The alpha wolves lead and manage the pack, the beta wolves assist in leadership, the delta wolves guard the pack and support weaker members, while the omega wolves fulfil the remaining roles. This hierarchy enhances social cohesion and ensures effective pack management during hunting and defence. The GWO optimization process imitates this behaviour by initializing a swarm of wolves and iteratively identifying the top-performing α, β, and δ wolves as leaders. The remaining wolves, classified as ω, encircle these leaders to explore the search space and identify optimal solutions. The positions of the omega wolves are updated based on the average positions of the leading wolves, enhancing exploration capabilities and optimization efficiency. Figure 4.10 (b) illustrates a 2D positional space and the selection of neighbouring wolves, demonstrating the impact of these organizational rules on the wolves' positioning and their interaction strategies.

The mathematical formulations used in updating the ω wolves according to the location of the prey (p) are as follows:

$$v_{i,j}^{t+1} = v_{p,j}^{t} - A_{p,j}^{t} D_{p,j}^{t} \tag{4.89}$$

with

$$D_{p,j}^{t} = \left| C_{p,j}^{t} v_{p,j}^{t} - v_{i,j}^{t} \right| \tag{4.90}$$

and

$$v_{i,j}^{t} = W_{i,j}^{T} \left(u_{i,j}^{b} - l_{i,j}^{b} \right) + l_{i,j}^{b} \tag{4.91}$$

$$A_{p,j}^{t} = 2 r_1 \mathrm{a} - \mathrm{a} \tag{4.92}$$

$$C_{p,j}^{t} = 2 r_2 \tag{4.93}$$

where

$$\mathrm{a} = 2\left(1 - \frac{t}{\max(t)} \right) \tag{4.94}$$

In equation (90), $v_{i,j}^{t}$ denotes the position of the i^{th} grey wolf in the j^{th} dimension in the t^{th} iterations. Herein, W_T represents the unknown parameters of the developed estimation architecture, which are updated during the optimization process. Furthermore, u^b and l^b represent the upper and lower bounds of the parameter space being explored. Additionally, t in equation (4.94) indicates the current iteration and "a" is staring from 2 when $\mathrm{t} = 0$ and decaying to 0 when $\mathrm{t} = \max(t)$, while $v_{p,j}^{t}$ in equation (4.90) is the position of the prey in the j^{th} dimension in the t^{th} iteration. Likewise, $\max(t)$ represents the maximum number of iterations, r_1 and r_2 are two random numbers uniformly generated in the interval $[0,1]$.

Depending on the values of ‘‘ a ’’, if the $|A|<1$ status is present, the grey wolf is considered to be in close proximity to the prey and is ready to attack. In the case of $|A|>1$, the grey wolf moves away from the prey with the intention of searching for new prey. The $C^t_{p,j}$, another important control parameter, serves as the exploration component of the algorithm and is assigned random values within the range $[0,2]$ of u^b and l^b. This parameter enhances the algorithm's ability to exhibit random behaviour, thereby reducing the likelihood of the optimization process becoming trapped in local optima. Specifically, if $|C^t_{p,j}|>1$, the algorithm exhibits increased randomness, which aids in escaping local optima. Conversely, if $|C^t_{p,j}|<1$, the extent of random behaviour is diminished.

After encircling the p, the wolves must update their positions based on their relative locations. Specifically, each ω wolf adjusts its position in relation to the positions of the α, β, and δ wolves, as detailed in the following formulations:

$$D^t_{\alpha,j}=\left|C^t_{\alpha,j}v^t_{\alpha,j}-v^t_{i,j}\right| \tag{4.95}$$

$$D^t_{\beta,j}=\left|C^t_{\beta,j}v^t_{\beta,j}-v^t_{i,j}\right| \tag{4.96}$$

$$D^t_{\delta,j}=\left|C^t_{\delta,j}v^t_{\delta,j}-v^t_{i,j}\right| \tag{4.97}$$

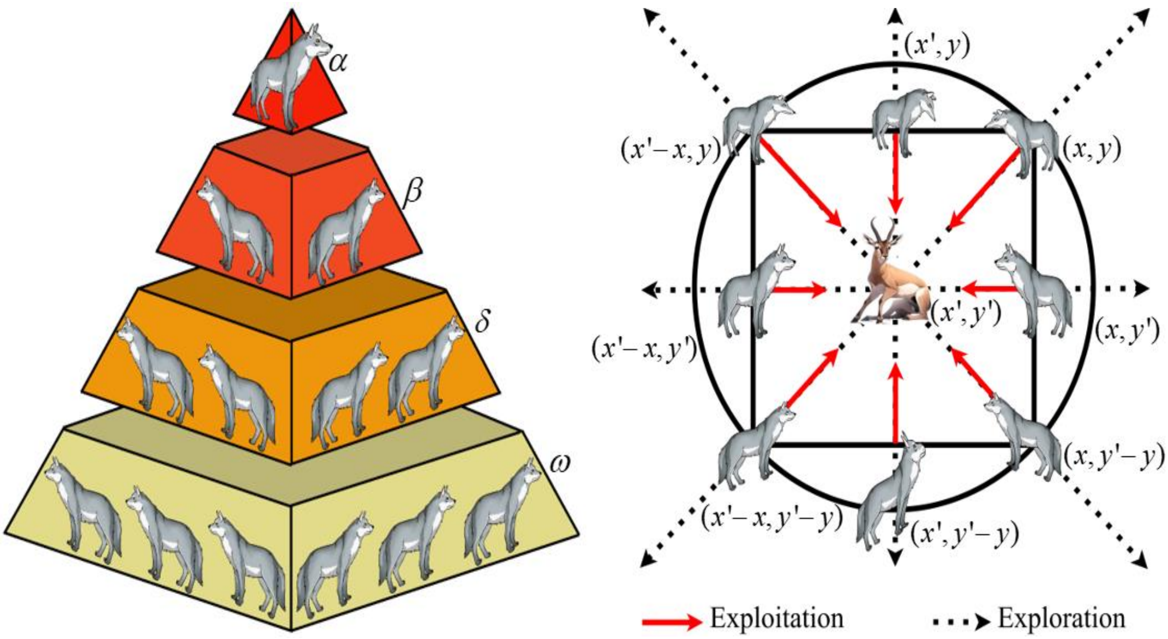


Figure 4.10. Wolf pack social hierarchy and their 2D positions in the conventional GWO.

Thus, the final state of the updated omega wolves is determined as follows:

$$v^{t+1}_{1,j}=v^t_{\alpha,j}-A^t_{\alpha,j}D^t_{\alpha,j} \tag{4.98}$$

$$v_{2,j}^{t+1} = v_{\beta,j}^{t} - A_{\beta,j}^{t} D_{\beta,j}^{t} \tag{4.99}$$

$$v_{3,j}^{t+1} = v_{\delta,j}^{t} - A_{\delta,j}^{t} D_{\delta,j}^{t} \tag{4.100}$$

with

$$v_{i,j}^{t+1} = \left(\frac{v_{1,j}^{t} + v_{2,j}^{t} + v_{3,j}^{t}}{3} \right) \tag{4.101}$$

However, in the conventional GWO, search agents do not possess velocity, resulting in abrupt and unrefined movements as they adjust their positions based on guide agents. As agents continuously change their acceleration towards new guides, this can cause discontinuities in their trajectories, resulting in potential drifts. These drifts may cause the agents to miss promising positions within the search space, increasing the risk of premature convergence and undermining the effectiveness of the optimization process.

4.4.2. M-GWO

To overcome these limitations, the M-GWO introduces initial random velocities for the ω wolves, allowing them to move more fluidly towards the leading α, β, and δ wolves. ach agent is assigned both a velocity and a position in the search space. The inclusion of velocity enables the wolves to maintain smoother transitions and maintain momentum as they explore the space. The M-GWO calculates the velocity for each search agent in the i^{th} dimension as follows:

$$V_{\alpha,j}^{t+1} = p_{tv}\left(\text{sgn}\left(A_{\alpha,j}^{t}\right)\left|V_{\alpha,j}^{t}\right|\right) + A_{\alpha,j}^{t} D_{\alpha,j}^{t} \tag{4.102}$$

$$V_{\beta,j}^{t+1} = p_{tv}\left(\text{sgn}\left(A_{\beta,j}^{t}\right)\left|V_{\beta,j}^{t}\right|\right) + A_{\beta,j}^{t} D_{\beta,j}^{t} \tag{4.103}$$

$$V_{\delta,j}^{t+1} = p_{tv}\left(\text{sgn}\left(A_{\delta,j}^{t}\right)\left|V_{\delta,j}^{t}\right|\right) + A_{\delta,j}^{t} D_{\delta,j}^{t} \tag{4.104}$$

where $V_{\alpha,j}^{t}$, $V_{\beta,j}^{t}$ and $V_{\delta,j}^{t+1}$ represent the initial random velocity of a ω wolf in the j^{th} dimension along a given dimension when wolves α, β, and δ wolves are designated to attract the search agent during the update process. Additionally, sgn signifies the sign function, while p_{tv} s a tuning parameter that acts as the inertia weight, facilitating a balanced progression from exploration to exploitation. and is calculated iteratively as follows:

$$p_{tv} = \max\left(p_{tv}\right) - \left(\max\left(p_{tv}\right) - \min\left(p_{tv}\right)\right)\left(\frac{t-1}{\max\left(t\right)-1}\right) \tag{4.105}$$

Furthermore, $A_{\alpha,j}^{t}$, $A_{\beta,j}^{t}$, and $A_{\delta,j}^{t}$ represent the momentum components for the search entities, and are calculated as follows:

$$A_{\alpha,j}^{t} = (2r_1 - 1)a^2 \quad (4.106)$$

$$A_{\beta,j}^{t} = (2r_2 - 1)a^2 \quad (4.107)$$

$$A_{\delta,j}^{t} = (2r_3 - 1)a^2 \quad (4.108)$$

where r_1, r_2, and r_3 are the randomly sampled numbers are determined in the range of $[0,1]$as in conventional approach. Additionally, "a" is a modified parameter that decreases linearly over time and is successively computed as follows:

$$a = \max(a) - (\max(a) - \min(a))\left(\frac{t-1}{\max(t)-1}\right) \quad (4.109)$$

As indicated by equation (4.109), the parameter "a" undergoes a transition from its initial value of $\sqrt{2}$ at the commencement of the algorithm to 0 in the concluding iteration. Nevertheless, as indicated in equations (4.106) through (4.108), this parameter "a" is raised to the power of 2, signifying that a^2 evolves from 2 to 0 over the course of the iterations. The incorporation of this squared term in these equations significantly enhances the M-GWO's capability to modulate the interplay between exploration and exploitation throughout the optimization process. Specifically, if $a^2 \geq 1$, it predominantly engages in the exploration phase with $a^2 \geq a$, and a is positive.

Within the context of the proposed M-GWO, a^2surpasses the conventional value of "a" expected in the standard GWO during the exploration phase. This means that $A \gg 1$, thereby enabling the search agents to more thoroughly explore the search space during this phase. Conversely, if $a^2 \leq 1$, the algorithm transitions into the exploitation phase with $a^2 \leq a$, and $a \leq 1$. In this phase, the parameter is significantly lower than what would typically be anticipated for "a" in the conventional GWO, signifying that $A \ll 1$, and thus, the exploitation phase is more prominently emphasized in the proposed M-GWO.

Subsequent to this, the parameters $D_{\alpha,j}^{t}$, $D_{\alpha,j}^{t}$, and $D_{\alpha,j}^{t}$ represent the adjusted distances between the targeted wolf and the leading wolves, respectively. These distances, calculated for each dimension, are expressed as follows:

$$D_{\alpha,j}^{t} = \left| C_{\alpha,j}^{t} v_{\alpha,j}^{t} - v_{i,j}^{t} \right| \quad (4.110)$$

$$D_{\beta,j}^{t} = \left| C_{\beta,j}^{t} v_{\beta,j}^{t} - v_{i,j}^{t} \right| \quad (4.111)$$

$$D_{\delta,j}^{t}=\left|C_{\delta,j}^{t}v_{\delta,j}^{t}-v_{i,j}^{t}\right| \tag{4.112}$$

where $v_{\alpha,j}^{t}$, $v_{\beta,j}^{t}$, and $v_{\delta,j}^{t}$ are the positions of the α, β, and δ wolves in the j^{th} dimension with the t^{th} iteration, $v_{i,j}^{t}$ is the position of the i^{th} wolf in the j^{th} dimension in the t^{th} iteration.

The $C_{\alpha,j}^{t}$, $C_{\beta,j}^{t}$, and $C_{\delta,j}^{t}$ coefficients also play a critical role in the optimization process by being multiplied by each leading wolf to either emphasize or de-emphasize their influence. This adjustment is particularly important due to the inherent uncertainties in the cost function associated with each leading wolf, especially during the early stages of the optimization process. These coefficients account for these uncertainties, thereby enhancing the algorithm's ability to effectively navigate the exploration phase. Thus, a new formulation is introduced for these coefficients, allowing them to be computed as follows:

$$C_{\alpha,j}^{t}=1+\left(2r_4-1\right)p_{tc}^{2} \tag{4.113}$$

$$C_{\beta,j}^{t}=1+\left(2r_5-1\right)p_{tc}^{2} \tag{4.114}$$

$$C_{\delta,j}^{t}=1+\left(2r_6-1\right)p_{tc}^{2} \tag{4.115}$$

where r_4, r_5, and r_6 represent uniformly distributed random numbers generated within the range of [0, 1], and p_{tc} is an additional control parameter adaptively determined as follows:

$$p_{tc}=\max\left(p_{tc}\right)-\left(\max\left(p_{tc}\right)-\min\left(p_{tc}\right)\right)\left(\frac{t-1}{\max\left(t\right)-1}\right) \tag{4.116}$$

As illustrated in equation (4.116), the parameter p_{tc} decreases linearly as the iterations progress. Initially, the coefficients are stochastically generated within the [0, 1] interval, but this range gradually narrows throughout the iterations. The adjustment of these ranges accelerates towards the final values, where no uncertainty is accounted for in the cost function of the leading search agents. This refinement is achieved by raising the parameterp_{tc} to the power of 2. Since p_{tc} remains within $[0,1]$ throughout the iterations, p_{tc}^2 is consistently less than p_{tc}. Consequently, the coefficients C approach $[1,1]=\{1\}$ more closely in the final iterations compared to the earlier ones. This methodology significantly augments the exploitation proficiency of the M-GWO while simultaneously bolstering its exploration effectiveness by integrating the velocity component into the update mechanism of the search agents. Finally, the wolves' next positions can be updated as follows:

$$v_{1,j}^{t+1}=v_{\alpha,j}^{t}-V_{\alpha,j}^{t+1} \tag{4.117}$$

$$v_{2,j}^{t+1} = v_{\beta,j}^{t} - V_{\beta,j}^{t+1} \tag{4.118}$$

$$v_{3,j}^{t+1} = v_{\delta,j}^{t} - V_{\delta,j}^{t+1} \tag{4.119}$$

From here, the new $v_{i,j}^{t+1}$ in the location of the i^{th} wolf in the j^{th} dimension is determined by averaging the three updated locations provided in equations (4.117)–(4.119) as follows:

$$v_{i,j}^{t+1} = \left(\frac{v_{1,j}^{t+1} + v_{2,j}^{t+1} + v_{3,j}^{t+1}}{3} \right) \tag{4.120}$$

Following the determination of the new updates, the final enhancement implemented in the M-GWO compared to the conventional GWO involves integrating an elitism mechanism into the proposed algorithm as follows:

$$v_{i,j}^{t+1} = \begin{cases} v_{i,j}^{t} & \text{if } J \le \bar{J} \\ \left(\dfrac{v_{1,j}^{t+1} + v_{2,j}^{t+1} + v_{3,j}^{t+1}}{3} \right) & \text{otherwise} \end{cases} \tag{4.121}$$

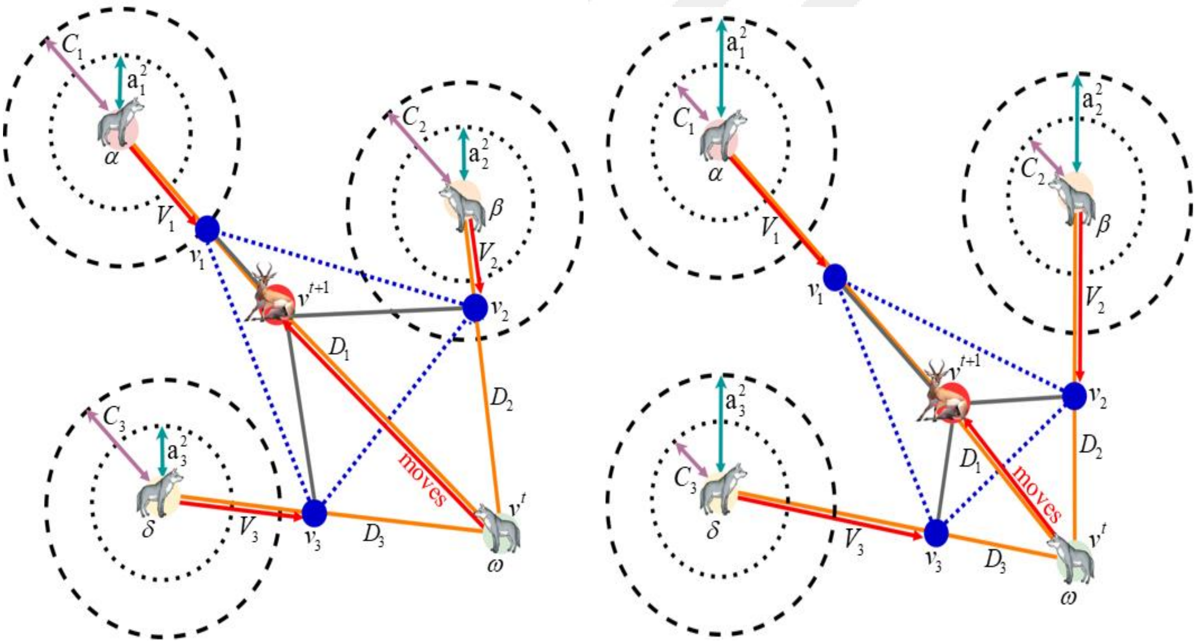


Figure 4.11. The trajectory of a representative ω wolf within the M-GWO during the exploration (left) and exploitation (right) phases.

This elitism mechanism compares the new positions of wolves to their best positions encountered so far. If the cost function at a wolf's new position yields a better result than its existing optimal position, the wolf adopts this new position as its updated best. Conversely, if the new position does not improve the cost function, the wolf retains its current optimal position, replacing the newly calculated one. This process ensures that wolves progressively improve over the course of iterations with the modified "a" parameter and the introduction of velocity adjustments governed by the p_{tv} and p_{tc} control parameters. These modifications contribute to enhanced optimization performance, facilitating more efficient convergence towards optimal solutions.

The detailed pseudo-code of the proposed M-GWO is presented with the following Algorithm 4.1 with the associated main parameter settings provided in Table 4.1.

Table 4.1. The main parameters of the proposed M-GWO.

Parameters	Values	Parameters	Values	Parameters	Values	Parameters	Values
J_t	Inf	$\min(a)$	0	$\min(p_{tv})$	0.3	$\min(p_{tc})$	0
α	500	$\max(a)$	$\sqrt{2}$	$\max(p_{tv})$	0.8	$\max(p_{tc})$	1
$\max(v)$	$0.1(u^b-l^b)$	δ	50	$\min(r_4,r_5,r_6)$	0	$\min(r_4,r_5,r_6)$	1
$\min(v)$	$-\max(v)$	u^b	$-\frac{\eta}{\max(W_{i,j}^T)}$	l^b	$-u^b$	η	[0,2]

Algorithm 4.1. The proposed M-GWO.

Input: v: Initialized population with v_α, v_β and v_δ grey wolves as unknown parameters, u^b: Upper bound , l^b: Lower bound, V: Initialized velocities as V_α, V_β and V_δ of v, p_{tv}: Velocity inertia tuning parameter, p_{tc}: Velocity cognition tuning parameter , a: Scaling factor, A: Decay factor, C : Adjustment factor, D: Distance , J_t: Initialized cost function, $\bar{J}_t$: Best cost function, δ: Size of the population, α: Number of iterations.

Output: Optimized v as the best solution

Initialization of the population:

for $i=1$ to v

1. Generate a random v_i with defined u_i^b and u_i^l as in equation (4.91)
2. Generate random velocities V_i of v_i
3. Generate "a$_i$", A and, C
4. Generate $p_{tv,i}$ and $p_{tc,i}$
5. Generate the cost function J_i for each $v_{a,i}$ as in equation (4.88)

endfor

Main Optimization Loop:

for $j=1$ to α

6. Obtain the average cost function as

$$J_a = J(j)/\sum_{k=1}^{j} J(k) \tag{4.122}$$

7. Sort and identify $v_{\alpha,j}$ as the best search agent based on the obtained J_a
8. Sort and identify $v_{\beta,j}$ as the second-best search agent based on the obtained J_a

9. Sort and identify $v_{\delta,j}$ as the third-best search agent based on the obtained J_a

Update Positions:

for $m = 1$ to δ

10. Compute A_m as in equation (4.106) - (4.108) with a_m in equation (4.109)
11. Compute $p_{tv,m}$ and $p_{tc,m}$ parameters as in equations (4.105) and (4.116)
12. Compute $C^t_{\alpha,m}$, $C^t_{\beta,m}$, and $C^t_{\delta,m}$ as in equations (4.113) - (4.115)
13. Calculate $D_{\alpha,m}$, $D_{\beta,m}$, and $D_{\delta,m}$ as in equations (4.110) - (4.112)
14. Calculate the $V_{\alpha,m}$, $V_{\beta,m}$, and $V_{\delta,m}$ as in equations (4.102) - (4.104)
15. Calculate $v_{1,m}$, $v_{2,m}$, $v_{3,m}$ final states as in equations (4.117) - (4.119)
16. Update each a_m, $A_{\alpha,m}$, $A_{\beta,m}$, $A_{\delta,m}$, $C_{\alpha,m}$, $C_{\beta,m}$, $C_{\delta,m}$, $p_{tv,m}$ and $p_{tc,m}$ parameters
17. Update the position of v_m based on the averaging equations of (4.117) - (4.119) as in equation (4. 120)
18. Compare the updated v_m to the $\bar{J}_m$ as

if $\bar{J}_m > J_m$

19. Update $\bar{J}_m = J_m$ with equation (4.122)

endif

endfor

20. Initialize the optimized v as a best solution

endfor

With the provision of the relevant pseudo-code, this section is now complete. The implementation of the proposed algorithm will be thoroughly examined in the next section.

4.5. Implementation of the Proposed Algorithm to HRs

This sub-section briefly introduces the utilized HR with the designed graphical user interface, and the development of a dialogue model utilizing Bayesian decision-tree-guided finite state machines (BDT-FSMs) for the implementation of the proposed temperature estimation algorithm.

4.5.1. An Overview of Alpha Mini HR and Designed Graphical User Interface

In this section, the proposed algorithm has been implemented on the Alpha Mini HR as illustrated in Figure 12 respectively. Herein, the Alpha Mini HR is distinguished by its compact and portable design, encompassing interactive features, expressive capabilities, and voice interactions. Enhanced with 4G LTE connectivity and advanced human face and body recognition technologies, the Alpha Mini significantly enhances the human experience. Its functionality is facilitated by 14 servo motors, enabling dynamic actions, while the expressive

LCD eyes convey emotions, rendering it a delightful tool for tracking critical physiological indicators.

Moreover, a sophisticated personalized graphical user interface has been specifically developed leveraging the Alpha Mini HR's Python Software Development Kit (SDK) as shown in Figure 4.14. This interface encompasses screens displaying body temperature analyses derived from the HR's camera, as well as an additional screen from a secondary computer that evaluates the surrounding environment during these assessments. The primary objective is to perpetually monitor the realization of HRIs and predictions from a different computational perspective. This tertiary screen is also shared with the designated doctor via their IP address.

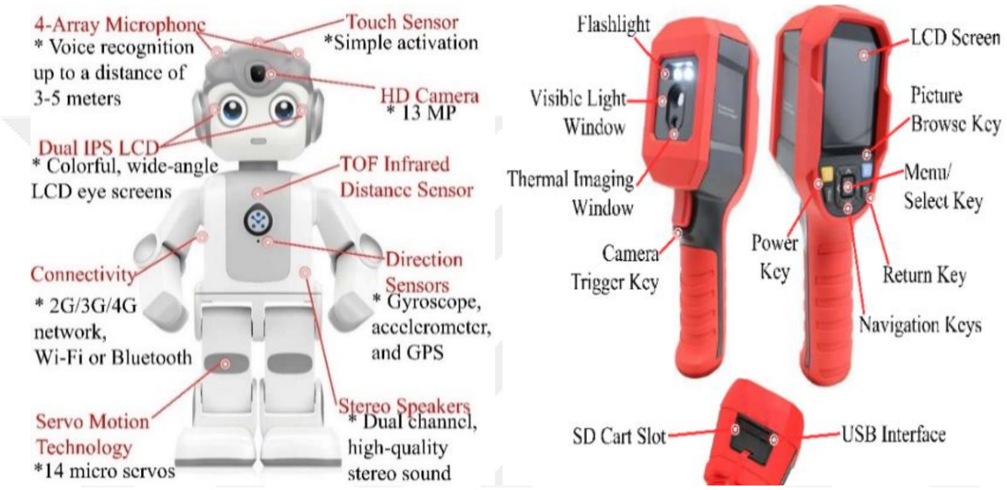


Figure 4.12. Left: Alpha Mini HR. Right: The UNI-T UTi220K Thermal Imager.

Furthermore, within this interface, body temperature values averaged at 5-minute intervals, including their start and end times with date information, are meticulously recorded in an Excel file. This file is subsequently integrated into the relevant interface to facilitate system monitoring for users. These averaged values are dispatched to the designated doctor's email address and are also communicated through phone messages at 30-minute intervals. In the event of an emergency, the Alpha Mini HR is capable of promptly contacting and informing the doctor or another assigned caregiver. In extreme situations, it can call the emergency number "ALO 112" using the SIM card inserted in its 4G SIM card slot. The current location of the patient is also displayed in real-time on a Google Map integrated into a separate screen within this interface. Herein, the locations of individuals shown in the map in Figure 4.14 have been zoomed out by a factor of 4x to safeguard their privacy. This location information is continuously updated and is retained for immediate transmission to relevant individuals in case of an emergency to ensure prompt intervention.

4.5.2. Development of A Dialogue Model for HRs using BDT-FSMs

The development of a dialogue model for HRs is crucial in the temperature estimation process as it enhances the HR's ability to interact effectively with subjects, gather accurate information, and provide precise feedback. By integrating advanced dialogue systems, HRs can ask clarifying questions, understand context, and respond appropriately to user inputs with designed behaviours. This interaction allows for the collection of detailed and relevant temperature data, improving the accuracy of temperature estimations. Furthermore, the dialogue model facilitates personalized and adaptive communication, ensuring that subjects are engaged and informed throughout the temperature monitoring process.

The utilization of BDT-FSMs in the development of dialogue models for HRs signifies a significant stride in enhancing HRI capabilities. In this framework, each dialogue state is represented as a node in an FSM, and transitions between states are guided by a Bayesian decision tree where a HR interacts with users by asking questions and processing yes or no answers. The decision tree assesses the likelihood of various responses based on historical data and context, updating its probabilities using Bayesian inference to update the HR's belief and make informed transitions. This allows the HR to handle uncertainties and variabilities in user interaction, making the dialogue more robust and adaptive. Through the implementation of BDT-FSMs, the dialogue model ensures a systematic and organized approach to HR behaviours, facilitating effective and context-aware interactions with the individuals. Formally, the transition from state S_t to the subsequent state S_{t+1} given an action A_t (a question Q_t and taken reaction D_t as detailed in Table 4.3) and an observation o_t (a yes or no answer) can be mathematically expressed using the following equations:

$$P\left(S_{t+1} \mid S_t, A_t, o_t\right) = \frac{P\left(o_t \mid S_{t+1}\right) P\left(S_{t+1} \mid S_t, A_t\right)}{P\left(o_t \mid S_t, A_t\right)} \tag{4.123}$$

with a belief update

$$b_{t+1}\left(S_{t+1}\right) = \eta P\left(o_t \mid S_{t+1}\right) \sum_{S_t} P\left(S_{t+1} \mid S_t, A_t\right) b_t\left(S_t\right) \tag{4.124}$$

where $P\left(o_t \mid S_{t+1}\right)$ is the probability of observing o_t given that the system is in state S_{t+1}. It indicates the likelihood of receiving the observation o_t if the system transitions to the state S_{t+1}. The $P\left(S_{t+1} \mid S_t, A_t\right)$ denotes the transition probability from state S_t to next state S_{t+1} given action A_t, capturing the dynamics of state transitions influenced by the action taken. The $P\left(o_t \mid S_t, A_t\right)$ represents the probability of o_t given the current state S_t and A_t. The term $b_t\left(S_t\right)$

refers to the belief state at time t, and η is a normalization factor applied to ensure that the updated beliefs are valid probabilities. This process iteratively refines the system's belief about the current state based on new evidence, allowing the dialogue system to make more informed decisions as it gathers more information from user responses. These formulas enable the dialogue system to maintain a probabilistic understanding of the user's state and adapt its questioning strategy dynamically, optimizing the interaction flow based on the accumulated evidence from user responses.

As depicted in Figure 4.13, the proposed BDT-FSMs applied to the Alpha Mini HR has 6 phases as. In the first phase, it initiates a greeting dialogue with the related human within the range of its microphone and speaker sensors and tries to engage with the individual. Once the “Initialization” stage is completed, the designed dialog enters the “Face Detection” phase. In this case Alpha Mini HR starts the “Face Tracking (FT)” module. Following the completion of the “Face Detection” phase, the designed dialog transitions to the “Temperature Estimation” phase. Upon estimating a temperature, the Temperature Tracking (TT) module enables the Alpha Mini HR to continuously monitor the individual's temperature. In this context, the system records the average body temperature data collected at 15-minute intervals and transmits this information as brief messages to the email and phone of the doctor and the relevant caregiver. If the estimated temperature disappears, the dialogue manager reverts to the “Temperature Estimation” stage via the S3 statement. After this stage is completed, the dialogue manager proceeds to the “Interpretative Analysis of the Estimated Temperature” phase to interpret estimated temperature statements. If temperature values are missed, the dialogue manager again reverts to the “Temperature Estimation” stage using the S4 statement. The Alpha Mini HR then interprets each estimated temperature value by progressing through the "I1" to "I9" stages for specific temperature ranges as outlined in the following Table 4.2. These ranges are specifically selected to assess the HR's engagement performance. Finally, upon completing this analysis, the Alpha Mini HR concludes the interaction with the "Termination" phase.

In each interpretation case, a BDT structure is employed to guide FSMs in initiating a social interaction with the relevant human counterpart. For instance, as depicted in Figure 4.15 (a), if the interpreted temperature values align with the ranges specified in cases "I1" and "I2", the Alpha Mini HR proceeds directly to the "D1" state to inform the relevant individual. It then proceeds to the "Q1" state to request permission for further assistance and awaits his/her response. During this interval, the Alpha Mini HR continues to assess and analyse the individual's body temperature through the activation of states "I1" and "I2". If the response from state "Q1" is negative ("No (N)"), the Alpha Mini HR promptly moves to the "D7" state, where it continues exclusive temperature analysis. It then advances to the "Q4" state. If the answer remains negative, the system proceeds to the "D9" state and remains in a standby mode. However, in the event of an affirmative response ("Yes (Y)"), the Alpha Mini HR initiates communication with the doctor and caregiver from the "D8" state, subsequently updating them on the situation. In an alternative scenario within the "Q1" state, if the response is affirmative, the HR transitions swiftly to the "D2" state and promptly engages with ChatGPT 3.5 to explore strategies for managing very low body temperature. Subsequently, it conveys the identified recommendations to the relevant individual.

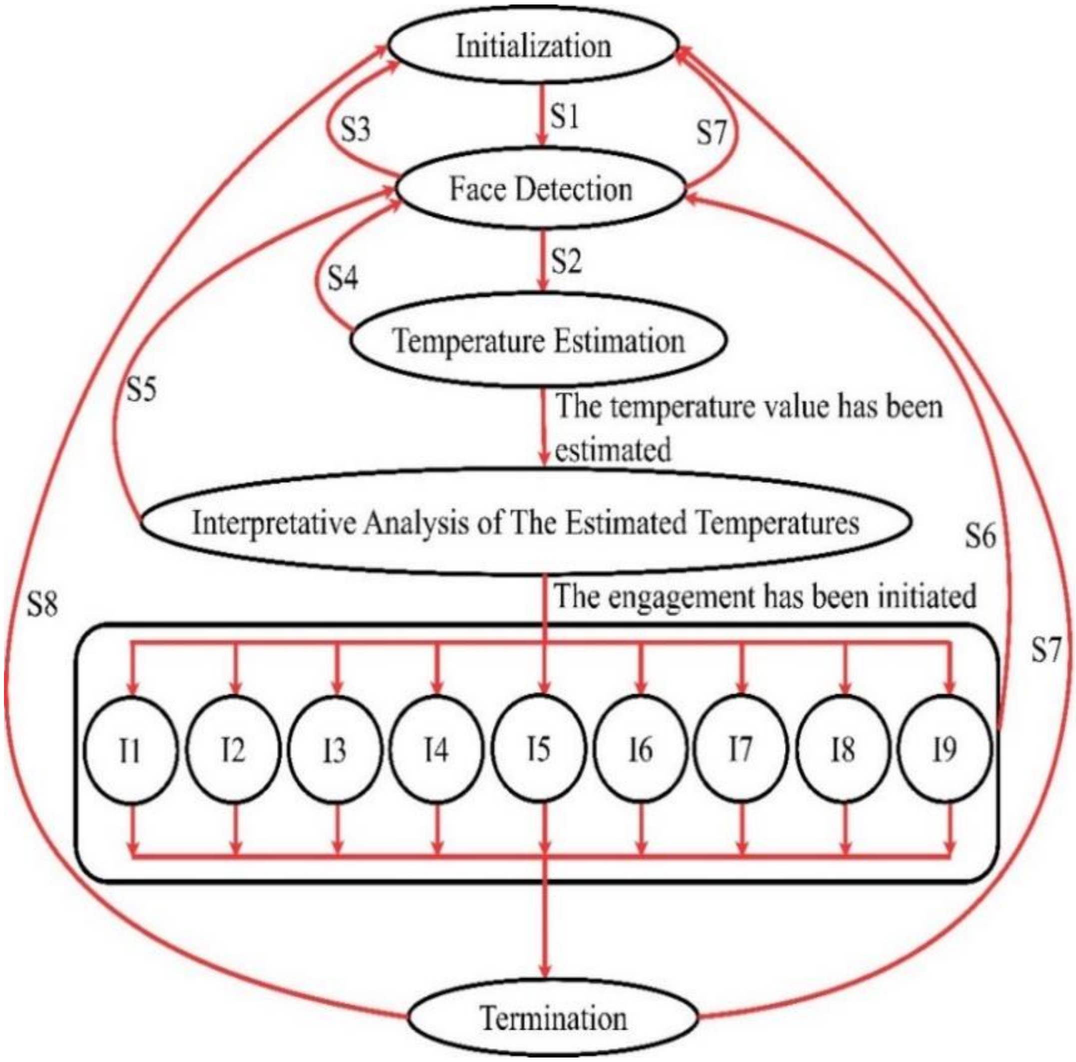


Figure 4.13. Illustration of the proposed BDT-FSMs-based dialogue model.

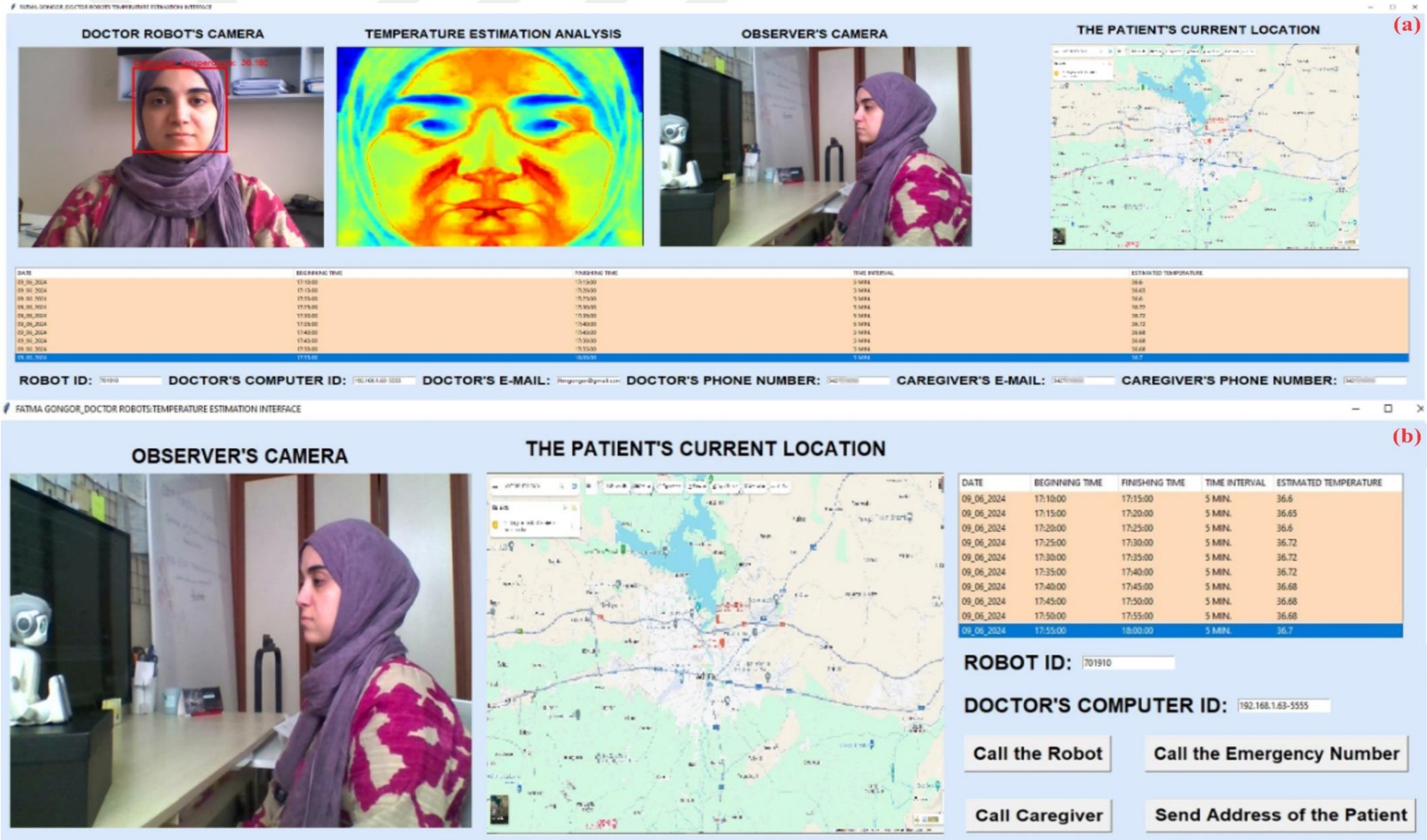


Figure 4.14. a) Personalized Alpha Mini HRI and b) Observer interfaces.

Upon providing the recommendations, the Alpha Mini HR advances to the "Q2" state. If the response remains negative, the system proceeds to the "D6" state and continues to wait. Conversely, if the response is affirmative, the Alpha Mini HR initiates further research with ChatGPT 3.5 in the "D3" state and subsequently provides additional advice to the individual. Following this interaction, it queries the individual again with the "Q3" statement before entering a final waiting state. If the response remains negative, the system progresses to the "D4" state. Alternatively, if the response is affirmative, the Alpha Mini HR moves to the "D5" state before entering a final waiting state.

In cases where the interpreted temperature value corresponds to the ranges indicated in case 'I3,' as depicted in Figure 4.15 (b), the Alpha Mini HR transitions to the "D10" state to notify the relevant individual. Herein, to enhance sociability, the HR personalizes its feedback by periodically addressing the subject by name throughout the sessions. The HR 's social presence is further reinforced by tracking the subject's attendance to improve adherence to the experimental protocol. For instance, if a subject misses two sessions, the HR provides feedback such as, "You didn’t come to the last experiment in session(x). I hope everything is all right!".

Table 4.2. Temperature range definitions for BDT-FSMs.

Case	Temperature Ranges	Definitions
I1	$T \leq 35.5°C.$	Hypothermia- Very Low Fever
I2	$35.6°C. \leq T \leq 36.0°C.$	Low Temperature
I3	$36.1°C. \leq T \leq 37.2°C.$	Normal Temperature
I4	$37.3°C. \leq T<38.0°C.$	Sub-Grade Fever
I5	$38.0°C. \leq T \leq 38.5°C.$	Mild Fever
I6	$38.6°C. \leq T \leq 39.0°C.$	Moderate Fever
I7	$39.1°C. \leq T \leq 39.9°C.$	High Fever
I8	$40.0°C. \leq T \leq 41.4°C.$	Very High Fever
I9	$41.5°C. \leq T \leq 45.0°C.$	Hyperthermia-Severe Fever

After providing a general overview, it queries whether the individual desires a more detailed report using the "Q5" state. If the response from "Q5" is negative, the Alpha Mini HR promptly moves to the "D11" state. In this state, it monitors for any sudden changes to inform the individual accordingly or switches to standby mode if no changes occur. In an alternative scenario within the "Q5" state, if the response is affirmative, the HR swiftly transitions to the "D12" state and requests additional advice to enhance health quality through the "Q6" state. If the response remains negative, the system continues to the "D11" state and remains in a waiting mode. Conversely, if the response is affirmative, the Alpha Mini HR initiates further consultation with ChatGPT 3.5 in the "D13" state and subsequently provides additional advice to the individual in the "D21" state. Following this interaction, it inquires whether the provided recommendations have been beneficial to the individual with the "Q7" state. If the response to "Q7" is affirmative, the Alpha Mini HR transitions to the "D22" state to conclude the engagement. Conversely, a negative response progresses to the "D14" state to request further advice using the "Q8" state. If the response in "Q8" is affirmative, the Alpha Mini HR engages in additional discussion with ChatGPT 3.5 in the "D15" state, exploring alternative strategies for healthier living in the "D22" state beyond previous advice. Finally, a negative response in the "Q8" state returns the system to the "D11" state before entering a final waiting phase.

If the interpreted temperature values fall within the ranges specified in cases 'I4,' 'I5,' and 'I6,' as depicted in Figure 4.15 (c), the Alpha Mini HR reacts accordingly. Upon estimation of body temperature within case "I4," it transitions directly to state "D16." Similarly, for temperatures falling within case "I5," it shifts to state "D17," and for case "I6," it moves to state "D18." Following these transitions, it promptly notifies the relevant doctor and caregiver, and initiates measures to comfort the corresponding individual if his/her body temperature exceeds normal thresholds.

Subsequently, the Alpha Mini HR queries the individual using the "Q1" state regarding his/her preference for receiving further recommendations until responses are received from both the doctor and caregiver. If the response from state "Q1" is negative, the Alpha Mini HR promptly moves to the "D6" state and again asks the corresponding individual for another attempt to call to the doctor and caregiver with “Q4” state. If the answer remains negative, the system proceeds to the "D4" state and remains in a standby mode. Conversely, if the response is affirmative, the Alpha Mini HR proceeds with another communication to the doctor and caregiver from state "D8," providing them with updates on the situation as necessary. In an alternative scenario within the "Q1" state, if the response is affirmative, the Alpha Mini HR swiftly transitions to state "D23" and promptly engages with ChatGPT 3.5 to explore strategies for managing high body temperature values in cases "I4," "I5," and "I6."

Following this consultation, the identified recommendations are relayed to the pertinent individual within the framework of the "D24" state. Upon providing the recommendations, the Alpha Mini HR advances to the "Q9" state. If the response remains negative, the Alpha Mini HR proceeds to the "Q8" state. A positive response in "Q8" prompts further dialogue between the HR and ChatGPT 3.5 within the "D25" state, focusing on alternative strategies for promoting healthier lifestyles beyond previously provided advice in the "D26" state.

Conversely, a negative response in "Q8" results in the system reverting to the "D4" state to conclude the engagement. Finally, upon the interpretation of temperature values falling within the specified ranges of cases "I7," "I8," and "I9" as depicted in Figure 4.15 (d), the Alpha Mini HR promptly engages in communication with the assigned doctor and caregiver to facilitate urgent intervention within the "D28" state. During this process, it aims to pacify and comfort the individual experiencing high fever to the fullest extent possible.

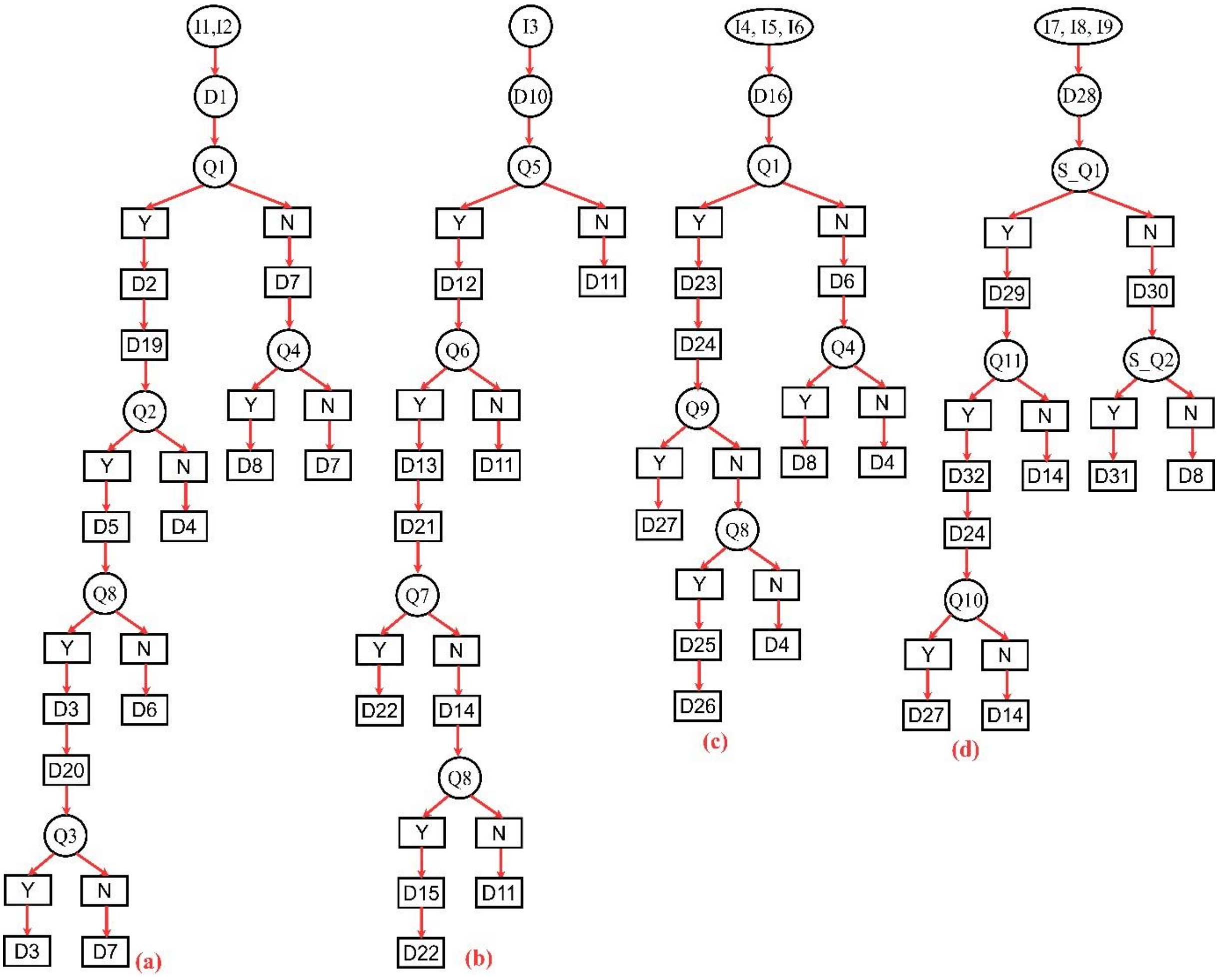


Figure 4.15. BDT structures for case classifications: (a) I1 and I2, (b) I3, (c) I4, I5, and I6, (d) I7, I8, and I9.

Following this initial action, the HR verifies the successful outreach and feedback from the concerned doctor and caregiver based on the outcomes of the communication attempts within the "S_Q1"state. In cases where communication with the relevant doctor and caregiver is not possible, the Alpha Mini HR promptly transitions to the "D30" state to initiate contact with the emergency number "ALO 112". During this phase, it evaluates the efficacy of establishing communication and receiving feedback from the emergency personnel within the "S_Q2" state. If the initial attempts are unsuccessful, the Alpha Mini HR proceeds to the "D8" state to make further efforts to reach the respective doctor and caregiver. If the initial attempts in the "S_Q2" state yield a positive response, the system advances to the "D31" state and prompts the individual to directly communicate his/her situation to the emergency team. It aims to reassure the individual of the immediate availability of help and encourages him/her to clearly convey his/her specific circumstances to facilitate effective and timely assistance from the emergency team.

Table 4.3. Descriptions of the states and statements in the proposed BDT-FSMs-based dialogue model.

States	Statements
S1	“Hi, I am Mini. I am here to assist you throughout the day by conducting and monitoring one of the essential health metrics, your body temperature. I will record this information in my system and periodically send updates to your doctor's email and mobile phone, thereby aiding in the maintenance of your physical health. Furthermore, a graphical user interface (GUI) has been designed to visually track my analyses, enabling both you and your doctor to monitor the data. If you wish, you can also follow your metrics through this interface. My interface is currently active and ready for me to commence the analysis. I am very excited to accompany you throughout the day. Now, let us begin the analysis.”
S2	“I have successfully detected your face and am now capable of measuring your real-time body temperature directly from your face.”
S3	“Sorry, I am currently unable to detect your face. Would you kindly position yourself within the range of my camera's coverage? Otherwise, I will be unable to conduct the analysis.”
S4	“Sorry, I am unable to estimate your body temperature at this moment. I will attempt this procedure once more now.”
S5	“Sorry, the interpreted temperature statement has been lost. I will now proceed to reinitiate the "Temperature Estimation" process.”
S6	“The engagement has not been executed due to the inability to identify the relevant interpretation case. I will make another attempt to analyse it.”
S7	“Due to an interruption in the analysis process, I will now proceed to restart the analysis. Please wait for a brief moment.”
S8	“Monitoring your body temperature is essential for me to oversee your physical well-being. Consequently, I will persist in conducting analyses throughout the day. Please be assured that I will maintain continuous presence and support.”
D1	“Your body temperature is significantly below normal levels. Although I report your temperature readings to the doctor regularly, as a precautionary measure, I will now

reach out to her via email and phone. Furthermore, I will notify your caregiver. Please stay composed; we will navigate through this situation together.”

D2 Various recommendations obtained via ChatGPT 3.5 for I1 and I2 cases.

D3 Different additional recommendations from D2 are acquired via ChatGPT 3.5.

D4 "I am deeply sorry that I could not be of further assistance to you. Therefore, let us await the response from the doctor and caregiver. If this situation persists and we receive no feedback, I will call the "ALO 112" number to request assistance from them."

D5 "I am delighted to hear that I have been able to assist you and that your body temperature has begun to rise."

D6 “Alright, if you do not need any additional suggestions, let us continue to wait.”

D7 “Okay, I understand. In that case, let us continue to wait.”

D8 “Alright, I am trying out to reach to your doctor and caregiver once more. Let's proceed with the call and await the result.”

D10 “According to the measurements conducted thus far, your body temperature values are observed to fall within the normal range. All parameters are in order, and these measurements are promptly being transmitted to your doctor.”

D11 “Alright, then I will continue taking measurements. If there is any sudden change, I will inform you again; for now, I am switching to silent mode.”

D12 Comprehensive numerical analysis results of the estimated temperature are presented verbally.

D13 Various recommendations obtained via ChatGPT 3.5 for I3 case.

D14 “I am very sorry that I couldn't provide you with helpful information.”

D15 Different additional recommendations from D13 are acquired via ChatGPT 3.5.

D16 "According to the measurements I have taken, there is an increase in body temperature within the "Sub-Degree Fever" category. Although I regularly report your temperature readings to the doctor, as a precautionary measure, I will now contact her via email and phone. Additionally, I will inform your caregiver. Please remain calm; we will navigate through this situation together."

D17 “Based on the measurements I have taken; your body temperature has risen to the "Mild Fever" category. Although I routinely report your temperature readings to the doctor, as a precaution, I will now contact her via email and phone. Additionally, I will inform your caregiver. Please remain calm; we will manage this situation together.”

D18 “Based on the measurements I have taken; your body temperature has risen to the "Moderate Fever" category. Although I routinely report your temperature readings to the doctor, as a precaution, I will now contact her via email and phone.

Additionally, I will inform your caregiver. Please remain calm; we will manage this situation together.”

D19 “I am pleased to have been able to assist you. I hope these recommendations will help increase your body temperature.”

D20 “I am pleased to have been able to assist you with these recent recommendations. I hope they will help increase your body temperature.”

D21 “I am pleased to have been able to assist you. I hope these recommendations will help improve your quality of life. ”

D22 “I am pleased to have been able to assist you with these latest recommendations. I hope these recommendations will help improve your quality of life. I am now transitioning to silent mode to continue my analyses.”

D23 Various recommendations obtained via ChatGPT 3.5 for I4, I5, and I6 cases.

D24 “I am pleased to have been able to assist you. I hope these recommendations will help decrease your body temperature.”

D25 Different additional recommendations from D23 are acquired via ChatGPT 3.5.

D26 “I am pleased to have been able to assist you with these recent recommendations. I hope they will help decrease your body temperature. I am now transitioning to silent mode to continue my analyses.””

D27 "I am delighted to hear that I have been able to assist you and that your body temperature has begun to decrease."

D28 “According to my current analysis results, your body temperature readings show a significant deviation from the normal range, indicating potentially dangerous levels. I am promptly notifying your doctor and caregiver and providing them with your current location to enable fast intervention. Rest assured; proactive measures are being implemented promptly to address this situation effectively.”

D29 “I have contacted your doctor and caregiver and provided them with all the detailed information. They will undertake all requisite supplementary measures on your behalf. Stay calm! We will bring your body temperature down to normal levels. I will accompany you throughout this process.”

D30 “I am unable to reach your doctor or caregiver. Stay calm; I am proceeding to the next step and calling the emergency number "ALO 112". They will definitely get back to us.”

D31 “Please provide a description of your situation directly to the emergency team; they are currently attentive and ready to assist you.”

D32 Various recommendations obtained via ChatGPT 3.5 for I7, I8, and I9 cases.

Q1 “Would you like me to offer you an additional recommendation until we receive responses from the doctor and your caregiver?”

Q2	“Have the recommendations I offered helped to increase your body temperature?”
Q3	" Have these latest recommendations I offered helped to increase your body temperature? "
Q4	"Would you prefer me to attempt another call to your doctor and caregiver?"
Q5	“If you desire, I am capable of providing you with a comprehensive overview of the numerical outcomes from your most recent measurements. Would you prefer to receive this information?”
Q6	“Additionally, would you be interested in receiving recommendations aimed at promoting a healthier lifestyle and enhancing your overall quality of life?”
Q7	“Did the recommendations I provided contribute to adopting healthier lifestyle practices and enhancing your overall well-being?”
Q8	“Would you desire an additional recommendation?”
Q9	“Have the recommendations I offered helped to decrease your body temperature?”
Q10	" Have these latest recommendations I offered helped to decrease your body temperature? "
Q11	“Would you like me to offer you some recommendations to lower your body temperature while awaiting assistance?”
S_Q1	“Has there been any response from the doctor or caregiver?”
S_Q2	“Has the emergency number "ALO 112" been contacted/reached?”

In an alternative scenario within the "S_Q1" state, upon successfully establishing communication with the relevant doctor and caregiver, the Alpha Mini HR promptly moves to the 'D29' state. In this state, the HR is attempting to reassure the individual that it has successfully contacted their doctor and caregiver, providing them with all necessary information. It assures the individual to remain calm, stating that the medical professionals will take all necessary steps to address the situation and reduce their body temperature to normal levels. After this state, HR passes to the “Q11” state to provide suggestions to help lower the individual's body temperature and managing his/her condition temporarily. If the response from state "Q11" is negative, the Alpha Mini HR promptly moves to the "D14" state and express its apology for being unable to offer useful or beneficial information to the individual. If the response is affirmative, the HR transitions swiftly to the "D32" state and promptly engages with ChatGPT 3.5 to explore strategies for managing very high body temperature. Subsequently, it communicates the identified recommendations to the relevant individual and expresses satisfaction at being able to assist within the "D24" state.

Upon providing the recommendations, the Alpha Mini HR progresses to the "Q10" state to solicit feedback on the effectiveness of the suggestions in reducing the individual's body temperature. If the feedback is negative, the system returns to the "D14" state to apologize for its inability to provide useful information. Conversely, if the response in the "Q10" state is affirmative, the Alpha Mini HR transitions to the "D27" state to express happiness at successfully aiding the individual, thereby concluding the interaction. This sequence underscores the HR's commitment to offering valuable assistance and its positive reaction to achieving favourable outcomes for the individual. With the successful development of the BDT-FSMs-based dialogue model, this section concludes. The following section will provide a comprehensive evaluation of the proposed algorithm, critically assessing its performance, accuracy, and overall effectiveness.

4.6. Simulation and Experimental Results

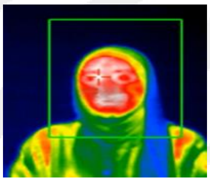
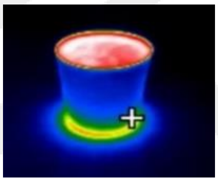
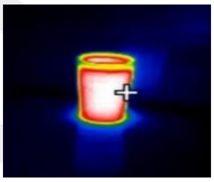
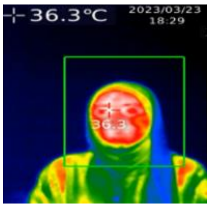
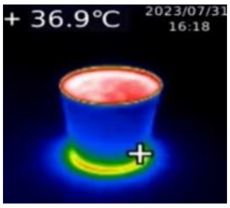
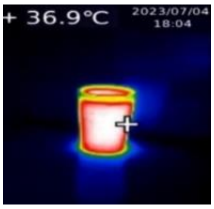
This sub-section evaluates the proposed algorithm both in the simulation and experimental environments. Initially, the experimental setting is introduced and then the results are assessed respectively.

4.6.1. Simulation and Experimental Settings

In order to assess the reliability and effectiveness of the proposed algorithm, initially a self-collected dataset called the Featured Temperature Measurement Dataset (FTM-DS) is created. This dataset is compiled through measurements conducted in a laboratory environment using sunlight as the only light source. Representative images from this dataset are presented in Table 4.4. As illustrated in Figure 4.16, it provides a detailed analysis of the FTM-DS by illustrating the distribution of image counts across different temperature ranges for various object types and imaging modalities. Specifically, the dataset encompasses three distinct categories: RGB images, thermal images, and thermal images with numeric values, each corresponding to human faces, iron cups, and glass cups, each characterized by distinct emissivity values of 0.96, 0.90 and 0.97, respectively. The temperature ranges are meticulously defined at intervals of 0.01°C, with each range containing a total image count of 5000 per temperature value. The y-axis of the figure represents these temperature ranges in Celsius, while the x-axis denotes the cumulative number of images. The numeric values displayed atop each bar offer precise image counts for each category, thereby enriching the understanding of the dataset's composition and providing a robust foundation for subsequent analysis.

A closer inspection of Figure 4.16 reveals a comprehensive and systematic generation of an extensive image dataset aimed at enhancing methodologies and algorithms pertinent to temperature measurement and analysis, particularly in the context of assistive health-care HRs. For instance, images for human face analysis are concentrated between the temperature ranges of 35.5°C and 39.5°C, whereas images for other objects span a broader range from 35.5°C to 45.5°C. This structured approach not only ensures a balanced distribution of data across varying temperature ranges but also fosters the development of advanced algorithms by offering a plentiful and diverse dataset.

Table 4.4. Representative images from the relevant FTM-DS dataset.

Type	Human Face / Emissivity= 96	Iron cup / Emissivity =90	Glass cup / Emissivity =97
RGB Image			
Thermal Image			
Thermal Image with Numerical Values			

Ultimately, this extensive dataset contributes significantly to a deeper understanding of temperature variations and supports ground-breaking research in vital health values analysis and monitoring, thereby bolstering the capabilities of assistive health-care technologies. During the data creation process, the UTi220K Thermal Imager is employed to capture the ground truth temperatures. It is positioned in front of human faces, iron cups, and glass cups to record the actual temperatures, thereby validating the accuracy of the proposed algorithm. As depicted in Figure 4.12, it features a high-resolution display with a measurement range of 30°C to 45°C and an accuracy of ±5°C or ±5%. This device includes PC software for automatic high-temperature data recording and imaging, complemented by an alarm buzzer for immediate alerts

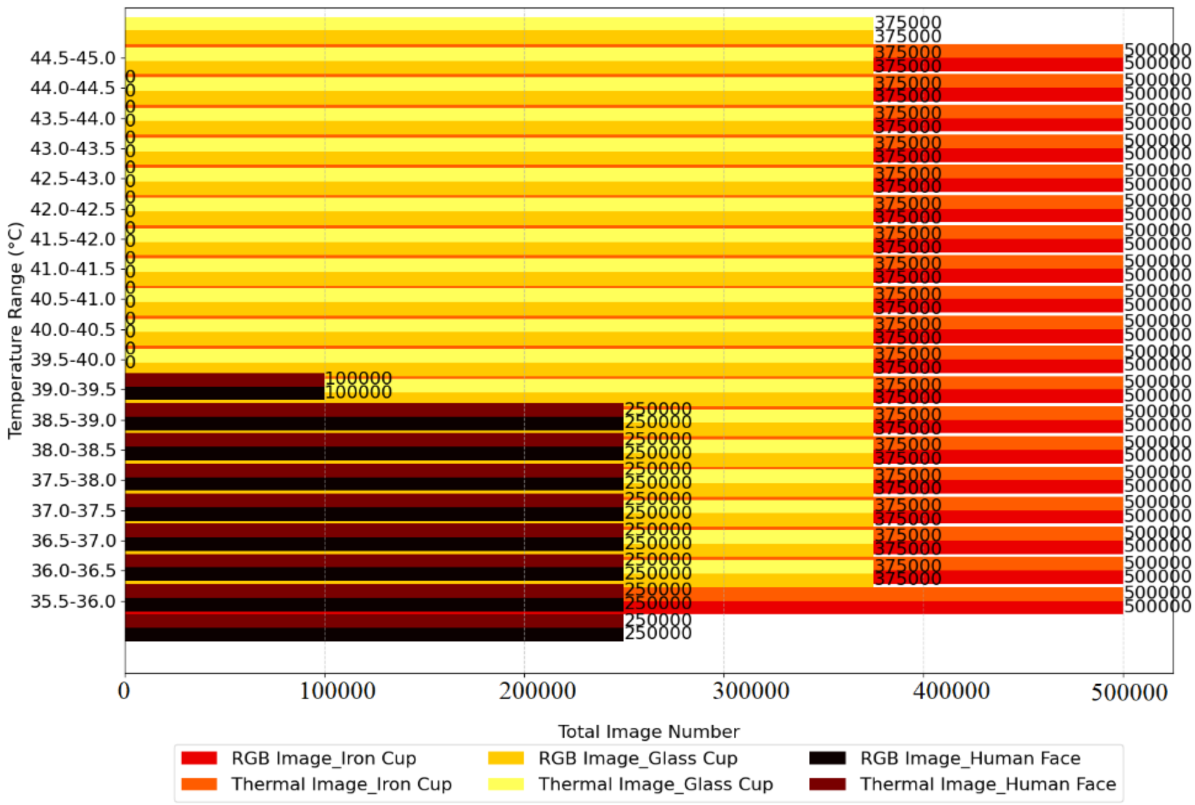


Figure 4.16. Illustration of the correlation between the total number of images and the corresponding temperature ranges for images of both human faces, iron and glass cups in the FTM-DS.

The UTi220K supports Region of Interest (ROI) mode, enhancing targeted temperature analysis. It also provides convenient photo and video storage via a micro-SD card, USB image transmission, and real-time image transfer capabilities. Equipped with a 1/4" tripod mounting hole and a Type-C USB interface, this thermal imager is both versatile and user-friendly, making it suitable for a variety of professional applications. The high-temperature alarm function further ensures safety and efficiency in different environments, making the UTi220K an essential tool for precise and reliable temperature measurements in experimental settings.

The evaluation of this dataset involved 15 sessions focused on cups, each with a duration of 10 minutes, and 10 sessions examining human faces, each lasting 5 minutes. In evaluating the proposed estimation algorithm for human faces, the study included 10 non-native English speakers (6 males, 4 females) aged between 19 and 55 years (Mean = 33.9, Standard Deviation = 13.01). Recruitment was conducted within the authors' personal networks. All participants had no visual, auditory, cognitive, or physical impairments and signed consent forms prior to participation. Throughout the experimental sessions, an experimenter was present in the laboratory for safety reasons but remained concealed behind a screen to avoid influencing the participant’s interaction with the HR. The experimenter only intervened in cases where the HR became unresponsive or experienced a connection failure, allowing participants the option to repeat the interaction. To prevent participants from being influenced by others' interactions and to avoid delays or queues, a structured schedule was implemented for participant attendance.

To investigate the effect of distance, two additional sessions were conducted at distances of 0.4, 0.6, 0.8, 1.0, and 1.2 cm, with 0.2 cm as the baseline distance from the HR camera. Five Canon EOS 1100D 18-55 DC professional cameras, matching the megapixel value of the HR's camera, were rented to capture data from these specified distances. Furthermore, the evaluation of the proposed algorithm's estimation performance across the designated dataset is conducted systematically through the application of performance metrics, including accuracy (m_a), precision (m_p), recall (m_r), and F1-Score (m_{F1}). This assessment also encompasses various variants of the proposed algorithm, with particular emphasis on the confusion matrix results of the variant demonstrating the highest accuracy rates. Within this framework, m_a is meticulously defined as the ratio of correctly predicted test samples, thereby serving as a metric that reflects the overall effectiveness of the proposed estimation algorithm, as delineated by the following equation.

$$m_a = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \tag{4.125}$$

where t_p represents the count of true positive samples, t_n refers to the categorization of true negative samples, f_p is used to signify false positive samples, and f_n is designated to indicate the occurrences of false negative samples. The m_p is quantitatively represented as the ratio of correctly predicted positive samples to the total number of predicted positive samples. Conversely, m_r is defined as the ratio of correctly predicted positive samples to the total number of true positive samples as articulated in the following equation.

$$m_p = \frac{t_p}{f_p + t_p},\ m_r = \frac{t_p}{f_n + t_p} \tag{4.126}$$

The m_{F1} represents the harmonic mean of precision and recall, thereby offering a unified metric that harmonizes both performance indicators. Its computation is performed as follows.

$$m_{F1} = 2\left(\frac{m_p m_r}{m_p + m_r}\right) \tag{4.127}$$

Following the performance analyses, the proposed algorithm was implemented on the Alpha Mini HR to validate its applicability and reliability in real-world settings. This quantitative approach allows for a rigorous assessment of the acceptability and effectiveness of the HRs in intervention contexts. To evaluate participants' experiences, preferences, task performance, and perceived personalization of the developed HR, the study utilized a 9-point Likert scale and a modified Unified Theory of Acceptance and the Use of Technology (UTAUT) questionnaire (Likert, 1932; Heerink et al., 2010). The Likert scale quantified subjective feedback across multiple dimensions, including "accuracy of temperature analysis," "completeness of interaction," "desire for additional interaction," "willingness for future experiences," "interest in real-world doctor robots," "enjoyment of the interaction," and "overall experience." On the other hand, the modified UTAUT questionnaire focused on constructs such as utility, usefulness, sociability, trust, social presence, and safety, as detailed in Table 4.5.

Table 4.5. The modified UTAUT questionnaire for evaluating subject perceptions of the personalized HR condition in terms of attractiveness (A), perceived usefulness (U), perceived utility (PU), perceived safety (S), ease of use (EU), perceived trust (PT), perceived sociability (PS) and social presence (SP).

Establishment	No	Question
U	1	I consider that a HR is an effective instrument for measuring and monitoring critical physiological indicators.
	2	I believe that my interaction with the HR was comfortable.
	3	I appreciated the verbal encouragement of the HR during fluctuations in my temperature readings.
	4	I am satisfied with the performance of the HR.
	5	I believe that the HR effectively adapts to my needs.
PU	1	I believe that utilizing and interacting with the HR is beneficial for monitoring my health condition and enhancing my self-care awareness.
	2	I believe that the role of the HR is crucial in the development of critical physiological indicator monitoring.
	3	I believe that using the HR enhances my commitment to monitoring my critical physiological indicators and performing necessary interventions effectively.
S	1	I feel secure during the sessions while interacting with the HR.
	2	I consider it easy to provide information to the HR.
EU	1	I believe that the HR is easy to use.
	2	I believe that the HR's suggestions were clear and its interventions were executed in a timely manner.
PT	1	The HR instilled a sense of confidence in me.

	2	I followed the recommendations provided by the HR owing to my confidence in its guidance.
	3	I find the utilization of the HR during the sessions to be highly beneficial and enjoy employing it.
	4	It instilled confidence in me that the HR provides guidance derived from the ongoing monitoring of my critical physiological indicators throughout the sessions.
PS	1	I consider the HR as a congenial and personalized medical companion in my daily life.
	2	I find interacting with the HR to be a pleasant experience.
	3	I believe the HR exhibits a pleasant demeanour and appears to be quite agreeable.
SP	1	While engaging with the HR, I felt as though I were conversing with a real human.
	2	It sometimes felt as though the HR was both genuinely directing its attention towards me and actively looking at me.
	3	I can imagine the HR as a living creature.
	4	I frequently consider that the HR is not a real person.
	5	Sometimes it seems like the HR has real emotions and feelings.

Additionally, supplementary questions were integrated into the modified UTAUT questionnaire to provide a thorough justification and acquire in-depth insights into the perceived personalization features, as detailed in Table 4.5. This enhanced questionnaire was administered following the final session with each HR condition. Open-ended questions were also employed to enable participants to freely articulate their perceptions of the HR. Audio recordings were also captured via the HR’s microphones and subsequently processed for speech recognition analysis. Participants were informed through consent forms that their interactions would be recorded, with an option to consent to the use of their videos and images for academic purposes.

Table 4.6. Supplementary questions integrated into the UTAUT questionnaire for evaluating subject perceptions of the personalized HR condition in terms of perceived usefulness (U), perceived utility (PU), perceived enjoyment (PE), perceived adaptivity (PA), perceived sociability (PS), social presence (SP) and attitude (A).

Establishment	No	Question
U	1	I feel motivated to attend the experimental sessions.
	2	I feel actively engaged in the experiments.
	3	I believe that the HR has contributed positively to my advancement in health monitoring.

	4	The feedback I receive from the HR about my session performance and health status inspires me to incorporate it into my everyday life.
PU	1	The HR accurately recognizes me and estimates my temperature value.
	2	The HR accurately remembers my previous sessions and estimated temperature values.
	3	The HR accurately tracks all of my session values.
PE	1	I am pleased that the HR accurately recognizes me.
	2	I am pleased that the HR uses my name and pronounces it correctly.
	3	I am pleased and feel secure receiving regular updates about my health status from the HR, and I appreciate that the HR shares this information with both my doctor and caregiver.
	4	I am pleased that the HR remembers me.
	5	I am pleased to collaborate with the HR.
PA	1	I feel that the HR personalizes its interactions.
PS	1	I believe that the HR has a thorough understanding of me.
SP	1	I perceive that the HR exhibits a distinct personality.
	2	I feel obligated to attend the sessions due to the HR's comments on my absence.
A	1	I feel a sense of attachment to the HR.

Upon discussing the final experimental setting, the ensuing subsection offers a comprehensive account of the obtained results and performs a thorough analysis of the pertinent findings.

4.6.2. Quantitative Analysis Results

This section delves into the quantitative assessment of the proposed personalized temperature estimation algorithm and its various adaptations. The primary focus is on evaluating the accuracy and effectiveness of these methods through rigorous analysis. Specifically, the evaluation is divided into two key sub-sections: the accuracy analysis of the personalized temperature estimation algorithm and its variants, and the assessment of a single subject's temperature data to gauge the performance of the implemented variant.

4.6.2.1. Accuracy analysis of the personalized temperature estimation algorithm and its variants

Initially, a detailed analysis is provided on the mean accuracy performance of the proposed personalized temperature estimation algorithm (PSIHVT) and its variants for the 'Human Face' and ‘Iron Cup’ classes within the FTM-DS as shown in Figure 4.17. In this analysis, the 'Iron Cup' category is specifically highlighted in conjunction with the 'Human Face' category to illustrate temperature values exceeding 39.5°C.

As depicted in Figure 4.17, all architectures showed a substantial increase in accuracy during the initial stages of training, followed by slight fluctuations after 10 epochs until the end of training. Specifically, by the end of training at 50 epochs, the highest mean accuracy value reached in PSIHVT-HFB2 for the "Human Face" category and PSIHVT-ICB2 in the "Iron Cup" category. This consistent trend across different PSIHVT variants indicates robust performance in temperature estimation for both the 'Human Face' and 'Iron Cup' classes.

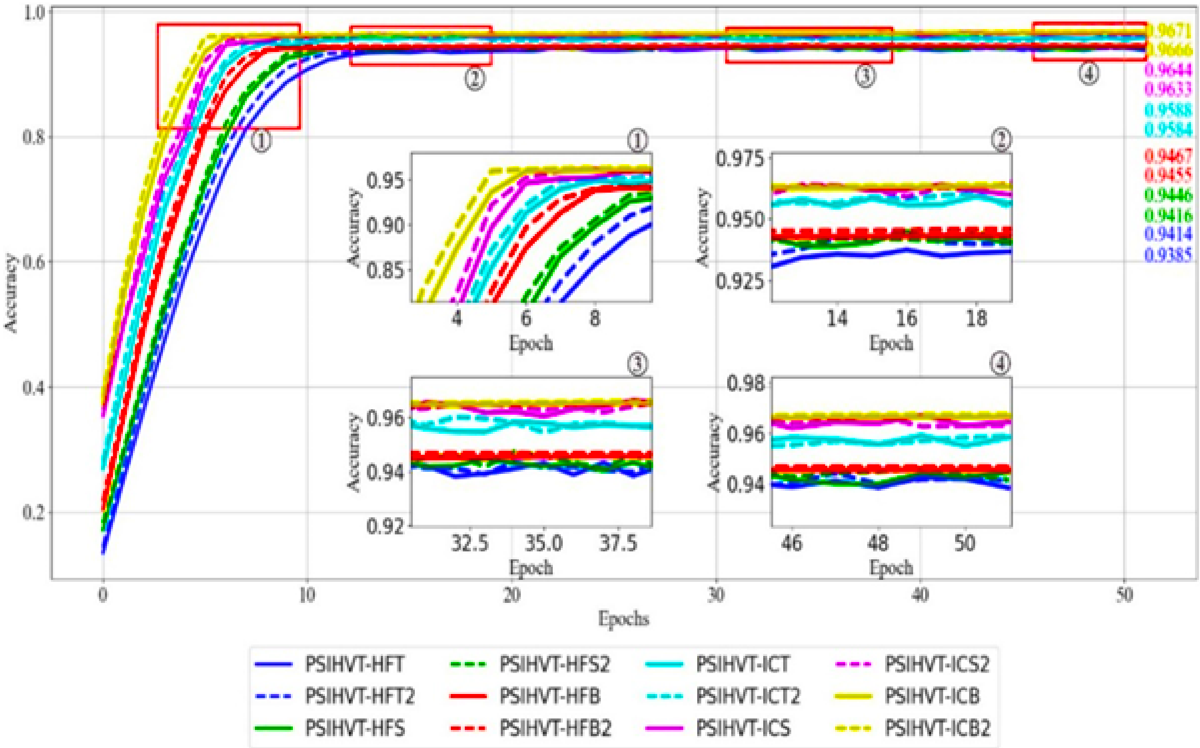


Figure 4.17. Mean accuracy performance of PSIHVT variants for the 'Human Face' and 'Iron Cup' classes within the FTM-DS.

The insets in Figure 4.17 provide zoomed-in views of specific epoch ranges, showing more detailed accuracy performance for selected regions. The first inset shows accuracy progression between 4 to 8 epochs, the second from 14 to 18 epochs, the third from 32.5 to 37.5 epochs, and the fourth from 46 to 50 epochs. This detailed analysis reveals that the proposed PSIHVT models maintain robust performance across different temperature ranges and classes, highlighting their efficacy in personalized temperature estimation tasks.

Table 4.7 provides detailed information about the PSIHVT variants and their parameter descriptions for the 'Human Face' and 'Iron Cup' classes within the FTM-DS. The variants include PSIHVT-HFT, PSIHVT-HFT2, PSIHVT-HFB, PSIHVT-HFB2, PSIHVT-HFS, PSIHVT-HFS2, PSIHVT-ICT, PSIHVT-ICT2, PSIHVT-ICB, PSIHVT-ICB2, PSIHVT-ICS, and PSIHVT-ICS2, each with specific configurations in terms of resolution and the number of blocks in four stages. These variants are further characterized by the number of heads (Head), the window size (M), and the number of channels (C) at each stage. Herein, the PSIHVT variants come in two different resolutions: 224×224 and 384×384, allowing for flexibility in application depending on the desired accuracy and computational resources. For example, PSIHVT-T and PSIHVT-T2 both have 224×224 and 384×384 resolutions respectively, with the same block configurations in each stage: 2/3/7/96 in Stage 1, 2/6/7/192 in Stage 2, 6/12/7/384 in Stage 3, and 2/24/7/768 in Stage 4. The PSIHVT-S and PSIHVT-S2 variants share a similar pattern, maintaining the same block configurations but differing in resolution. The PSIHVT-B and PSIHVT-B2 variants, which likely represent a more advanced configuration, have slightly higher numbers of blocks in the window sizes and channels, particularly noticeable in Stage 2 (2/8/7/256) and Stage 4 (2/32/7/1024).

Table 4.7. The PSIHVT variants and their main parameter descriptions for the 'Human Face' and 'Iron Cup' classes within the FTM-DS.

Class	PSIHVT Variant	Resolution	Blocks in Stage 1 / Head / M / C	Blocks in Stage 2 / Head / M / C	Blocks in Stage 3 / Head / M / C	Blocks in Stage 4 / Head / M / C
Human	PSIHVT-T	224×224	2 / 3 / 7 / 96	2 / 6 / 7 / 192	6 / 12 / 7 / 384	2 / 24 / 7 / 768
	PSIHVT-T2	384×384				
	PSIHVT-S	224×224	2 / 3 / 7 / 96	2 / 6 / 7 / 192	18 / 12 / 7 / 384	2 / 24 / 7 / 768
Iron Cup	PSIHVT-S2	384×384				
	PSIHVT-B	224×224	2 / 4 / 7 / 128	2 / 8 / 7 / 256	18 / 16 / 7 / 512	2 / 32 / 7 / 1024
	PSIHVT-B2	384×384				

Herein, Table 4.7 complements the analysis of the mean accuracy performance depicted in Figure 4.17, showcasing how different configurations of the PSIHVT variants are employed to achieve high accuracy in temperature estimation tasks. The consistent increase in accuracy over epochs correlates with the detailed block configurations in the table, emphasizing the architectures' adaptability and precision across different classes and resolutions. Furthermore, the subsequent Table 4.8 provides a comparative analysis of the implemented PSIHVT variants, evaluating their performance across four critical metrics: m_a, m_p, m_r, and m_{F1}. Each PSIHVT variant is presented with its corresponding values for these metrics, facilitating a detailed comparison of their performance.

Within the "Human Face" class, the PSIHVT-HFB2 variant exhibits the highest performance metrics, achieving an m_a of 0.9467, m_p of 0.9457, m_r of 0.9465, and an m_{F1} of 0.9461. The

PSIHVT-HFS and PSIHVT-HFS2 variants also demonstrate strong performance. Specifically, PSIHVT m_r -HFS2 attains an m_a of 0.9446, m_p of 0.9436, m_r of 0.9444, and an m_{F1} of 0.9440, slightly surpassing the PSIHVT-HFS variant, which records an m_a of 0.9416, m_p of 0.9407, m_r of 0.9414, and an m_{F1} of 0.9411. Conversely, PSIHVT-HFT registers the lowest performance metrics, with an m_a of 0.9385, a m_p of 0.9375, a m_r of 0.9383, and an m_{F1} of 0.9379, indicating its relatively lower efficacy. The PSIHVT-HFT2 variant shows a notable improvement over PSIHVT-HFT, achieving an m_a of 0.9414, a m_p of 0.9404, a m_r of 0.9412, and an m_{F1} of 0.9408. Despite these improvements, it still lags behind other PSIHVT variants. Overall, the observed incremental enhancements in the variants, particularly from HFB to HFB2 and from HFS to HFS2, underscore the significance of iterative advancements in the proposed architecture.

In the "Iron Cup" class, the PSIHVT-ICB2 variant excels across all metrics, with an m_a of 0.9671, m_p of 0.9670, m_r of 0.9672, and an m_{F1} of 0.9671. This variant's superior performance indicates its reliability in estimation, making it the most effective among the IC variants. The PSIHVT-ICB variant performs exceptionally well with an m_a of 0.9666, m_p of 0.9662, m_r of 0.9665, and an m_{F1} of 0.9663, but it is still slightly outperformed by PSIHVT-ICB2. Similarly, the PSIHVT-ICT and PSIHVT-ICT2 variants show strong performance, with PSIHVT-ICT2 having a slightly better overall score than PSIHVT-ICT. Specifically, PSIHVT-ICT achieves an m_a of 0.9584, m_p of 0.9574, m_r of 0.9582, and an m_{F1} of 0.9578, while PSIHVT-ICT2 shows slightly higher m_a and m_p but maintains the same m_r and m_{F1}, indicating marginal improvement.

Table 4.8. The results of the performance metrics for the proposed PSIHVT variants.

PSIHVT Variants	m_a	m_p	m_r	m_{F1}
PSIHVT-HFT	0.9385	0.9375	0.9383	0.9379
PSIHVT-HFT2	0.9414	0.9404	0.9412	0.9408
PSIHVT-HFS	0.9416	0.9407	0.9414	0.9411
PSIHVT-HFS2	0.9446	0.9436	0.9444	0.9440
PSIHVT-HFB	0.9455	0.9445	0.9453	0.9449
PSIHVT-HFB2	**0.9467**	**0.9457**	**0.9465**	**0.9461**
PSIHVT-ICT	0.9584	0.9574	0.9582	0.9578
PSIHVT-ICT2	0.9588	0.9578	0.9586	0.9582
PSIHVT-ICS	0.9633	0.9630	0.9631	0.9629
PSIHVT-ICS2	0.9644	0.9634	0.9642	0.9638
PSIHVT-ICB	0.9666	0.9662	0.9665	0.9663
PSIHVT-ICB2	**0.9671**	**0.9670**	**0.9672**	**0.9671**

The PSIHVT-ICS and PSIHVT-ICS2 variants also demonstrate incremental advancements in performance, with PSIHVT-ICS2 achieving notable metrics: an m_a of 0.9644, m_p of 0.9634, m_r of 0.9642, and an m_{F1} of 0.9638. The results show incremental improvements across the PSIHVT variants, particularly in the transition from ICT to ICB, with the latter consistently outperforming the former. This indicates that the ICB2 variant effectively leverages the hierarchical and swarm intelligence-based architecture to achieve optimal results in the estimation task, making it the most robust variant among those tested.

Furthermore, the confusion matrices for the variants with the highest accuracy rates are provided for further analysis. For example, Figure 4.18 presents the confusion matrix for the PSIHVT-HFB2 variant, illustrating precise temperature analysis from human faces within the range of 35.5 to 39.0 °C. The x-axis represents the predicted temperatures, while the y-axis shows the actual temperatures with each cell indicating the percentage of predictions for each true temperature value. A key feature of this confusion matrix is the dominance of high values along the main diagonal, which extends from the top-left to the bottom-right corner. These diagonal values represent instances where the predicted temperature matches the actual temperature, and their high percentages reflect the proposed architecture's accuracy.

For instance, an actual temperature of 35.5°C is correctly predicted 94.6% of the time, indicating strong predictive performance. Similar high percentages are observed for other temperatures, underscoring the implemented variant's reliability in accurately estimating temperatures within this range. Therein, misclassifications predominantly occur near the diagonal, indicating minor prediction errors typically involving temperatures close to the actual values. For instance, an actual temperature of 35.6°C is predicted as 35.7°C only 0.4% of the time. This pattern suggests that the variant's errors are minor, generally involving adjacent temperature values rather than significant discrepancies. Thus, the implemented variant demonstrates high accuracy within a narrow range, with minimal large errors, as values further from the diagonal approach zero. Due to the limited availability of human subjects with temperatures exceeding 39.0 °C, supplementary confusion matrix results for the PSIHVT-ICB2 variant are presented as depicted in Figure 4.20. These results highlight the variant's accuracy in estimating temperatures across the specific cases within the range of 35.5 to 45.0 °C, as previously detailed in Table 4.2 of Section 3.2.

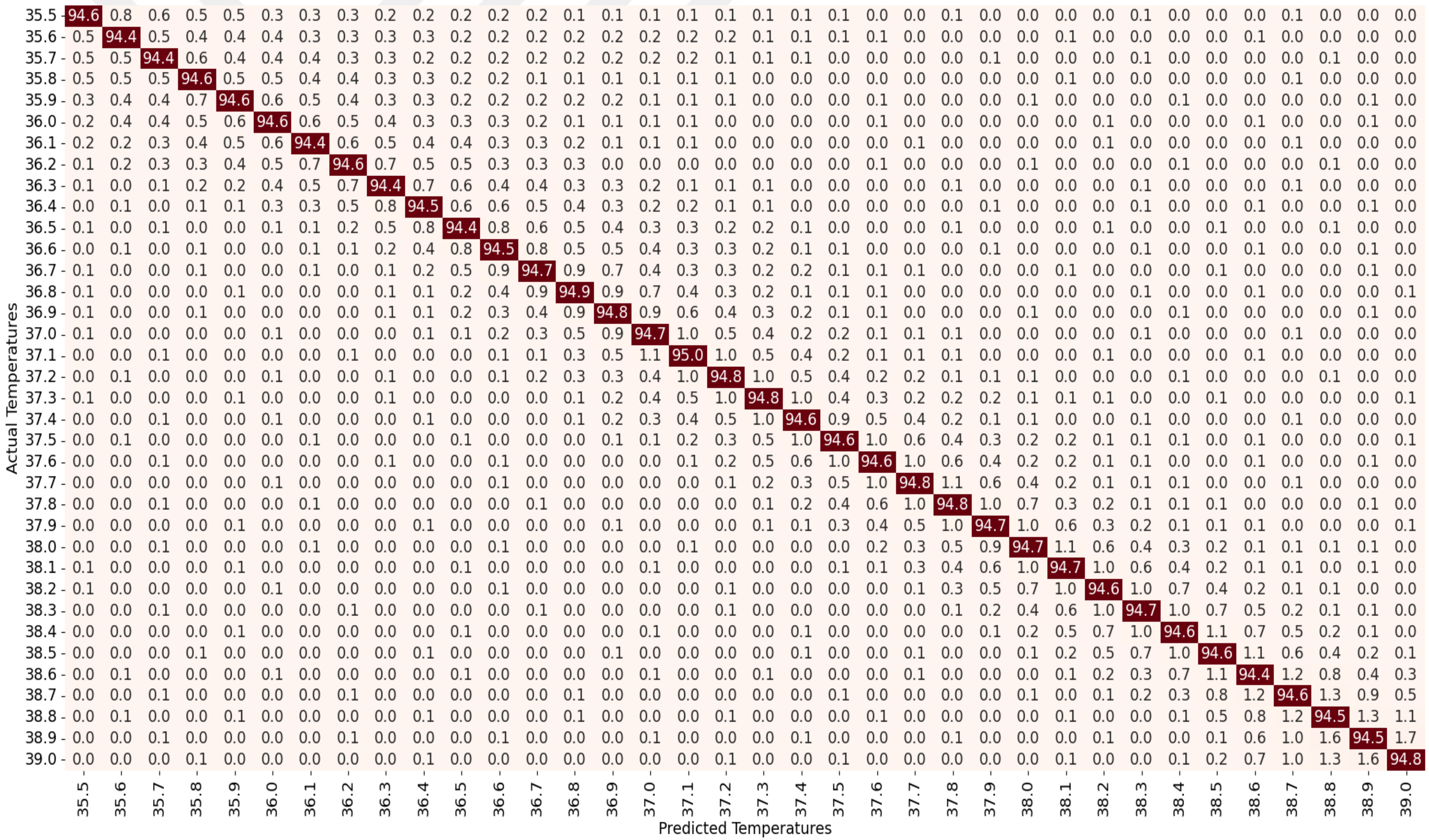


Figure 4.18. Confusion matrix evaluation of the implemented PSIHVT-HFB2 variant for the "Human Face" class.

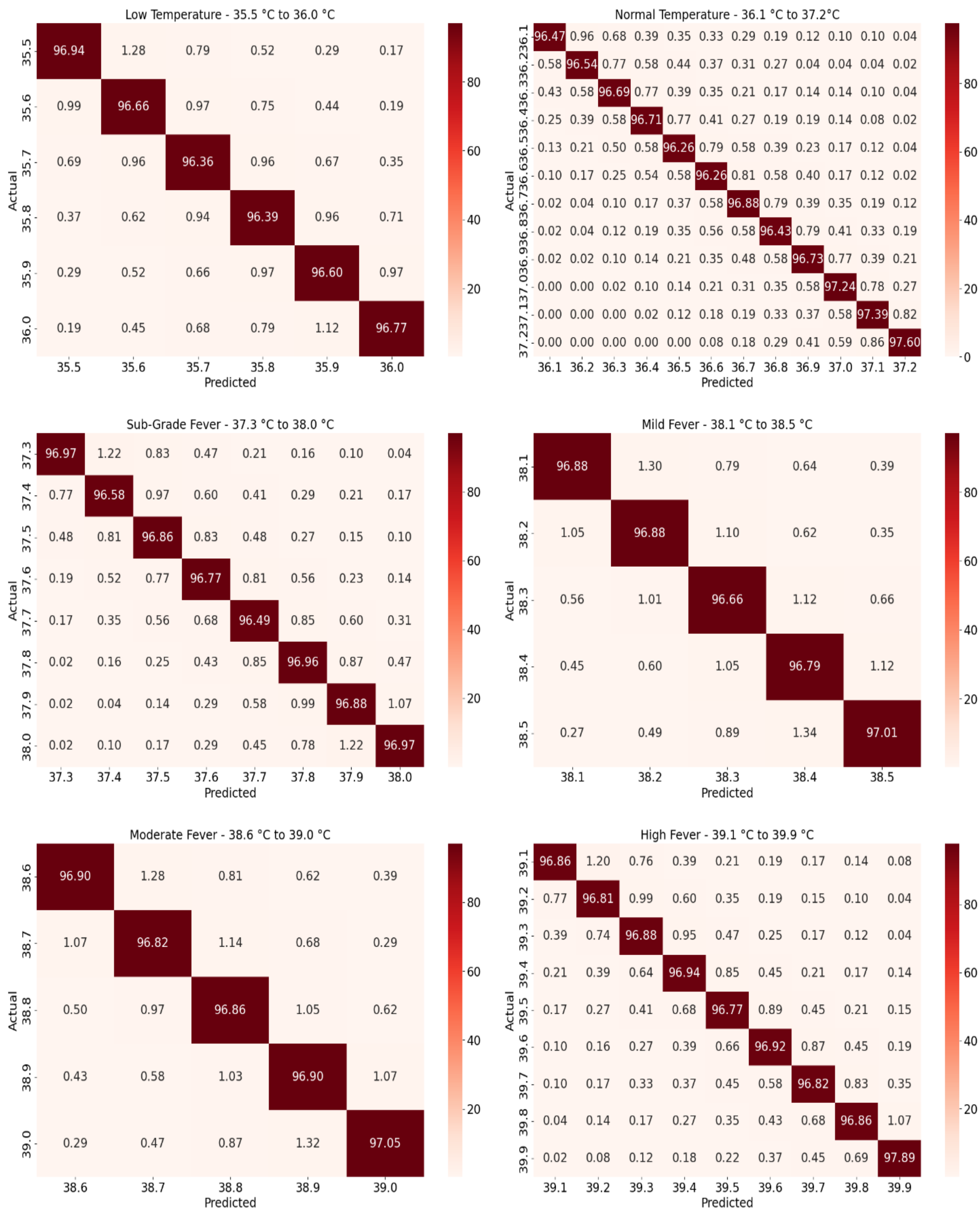


Figure 4.19. Confusion matrix evaluation of the implemented PSIHVT-ICB2 variant for the "Iron Cup" class within the FTM-DS specifying cases a) I1 and I2, b) I3, c) I4, d) I5, e) I6, and f) I7 as detailed in Table 4.2.

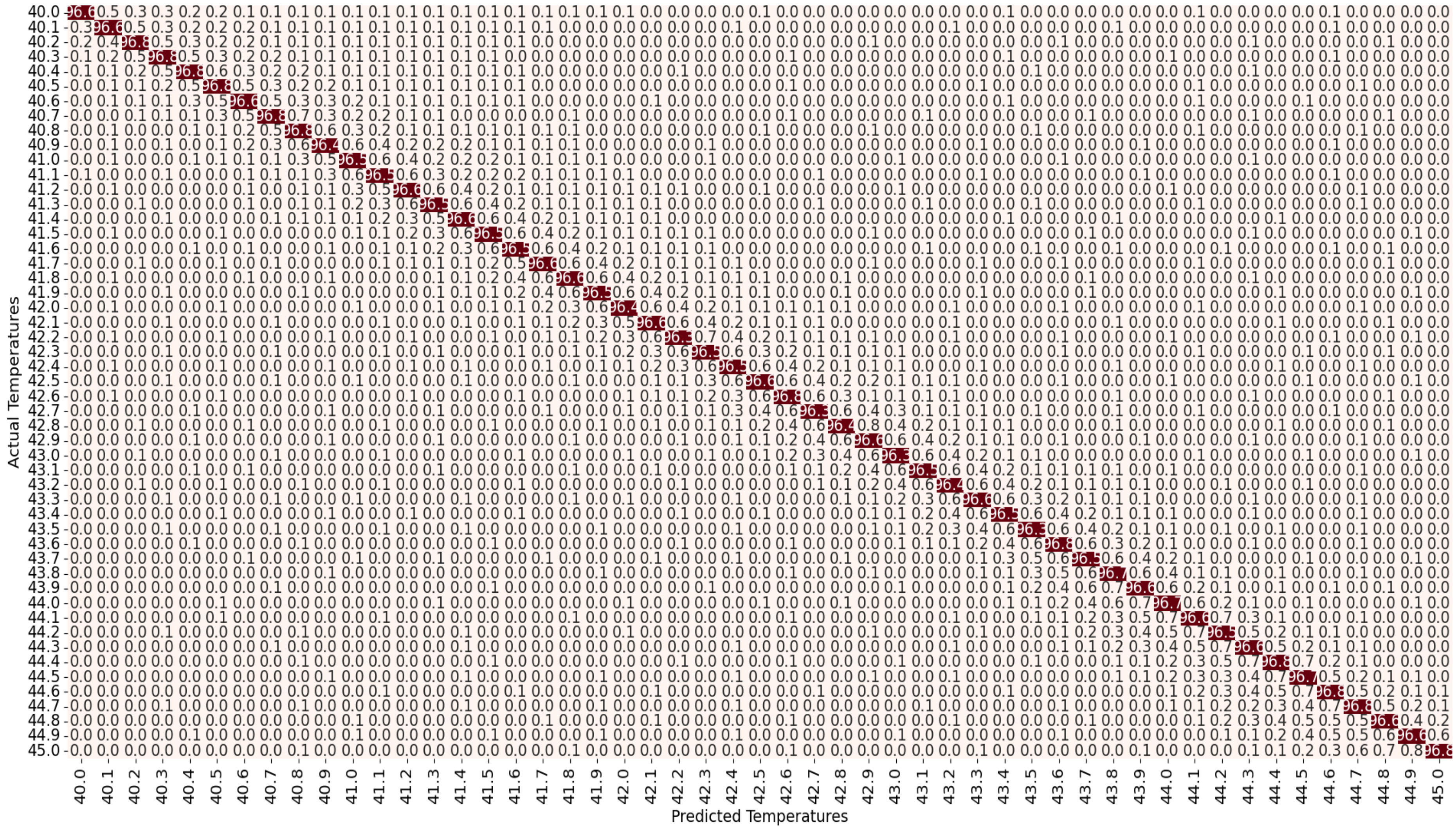


Figure 4.20. Confusion matrix evaluation of the implemented PSIHVT-ICB2 variant for the "Iron Cup" class within the FTM-DS specifying cases I8 and I9 as detailed in Table 4.2.

This analysis facilitates a deeper understanding of the performance of the PSIHVT-ICB2 variant. For instance, in the I1 and I2 cases, the confusion matrices demonstrate high accuracy, particularly in distinguishing between very low and low temperatures as illustrated in Figure 4.19 (a). This indicates that the architecture is proficient in recognizing subtle differences within this narrow temperature range, which is crucial for accurate low-temperature detection. In the I3 case, the confusion matrix also exhibits high accuracy, especially along the diagonal, signifying that the implemented variant accurately predicts these temperatures with minimal misclassifications as shown in Figure 4.19 (b). The accuracy values exceed 96%, underscoring the effectiveness of the implemented variant in these relatively stable temperature ranges. Notably, the highest accuracy values, reaching up to 97.00%, are observed in these categories, reflecting the variant's precision in identifying low and normal temperatures.

As the temperature range extends into fever categories (I4 to I7), confusion matrices show strong performance with minor increases in misclassifications near boundaries, yet accuracy remains above 96%. This indicates that the PSIHVT-ICB2 variant effectively handles a broad range of temperatures with high precision. In the very high fever range (I8), the matrix shows a strong diagonal pattern, indicating accurate temperature predictions with minimal misclassifications. Similarly, in the hyperthermia-severe fever range (I9), the matrix also demonstrates high accuracy with clear diagonal predictions, underscoring the algorithm's reliability and its importance in accurate severe fever estimation and timely medical intervention.

4.6.2.2. Assessment of the implemented variant via temperature analysis of a single subject

Following the analysis of relevant confusion matrices, instantaneous temperature measurements from a single subject are presented to evaluate the effectiveness of the implemented variant. Specifically, temperature data for Subject 1, a 20-year-old male engineering student of normal weight, were recorded over a 2000-second interval during the third session of his experiment. As detailed in the "Simulation and Experimental Settings" section, this subject has no visual, auditory, cognitive, or physical disabilities. Notably, He underwent rhinoplasty surgery two days before the experiments, and his body temperature fluctuates periodically in response to postoperative symptoms. Following discharge from the hospital, the subject 1 continued his treatment under the supervision of the HR. Lastly, it is important to note that Subject 1 refused to allow the visual or audio-visual dissemination of the collected data.

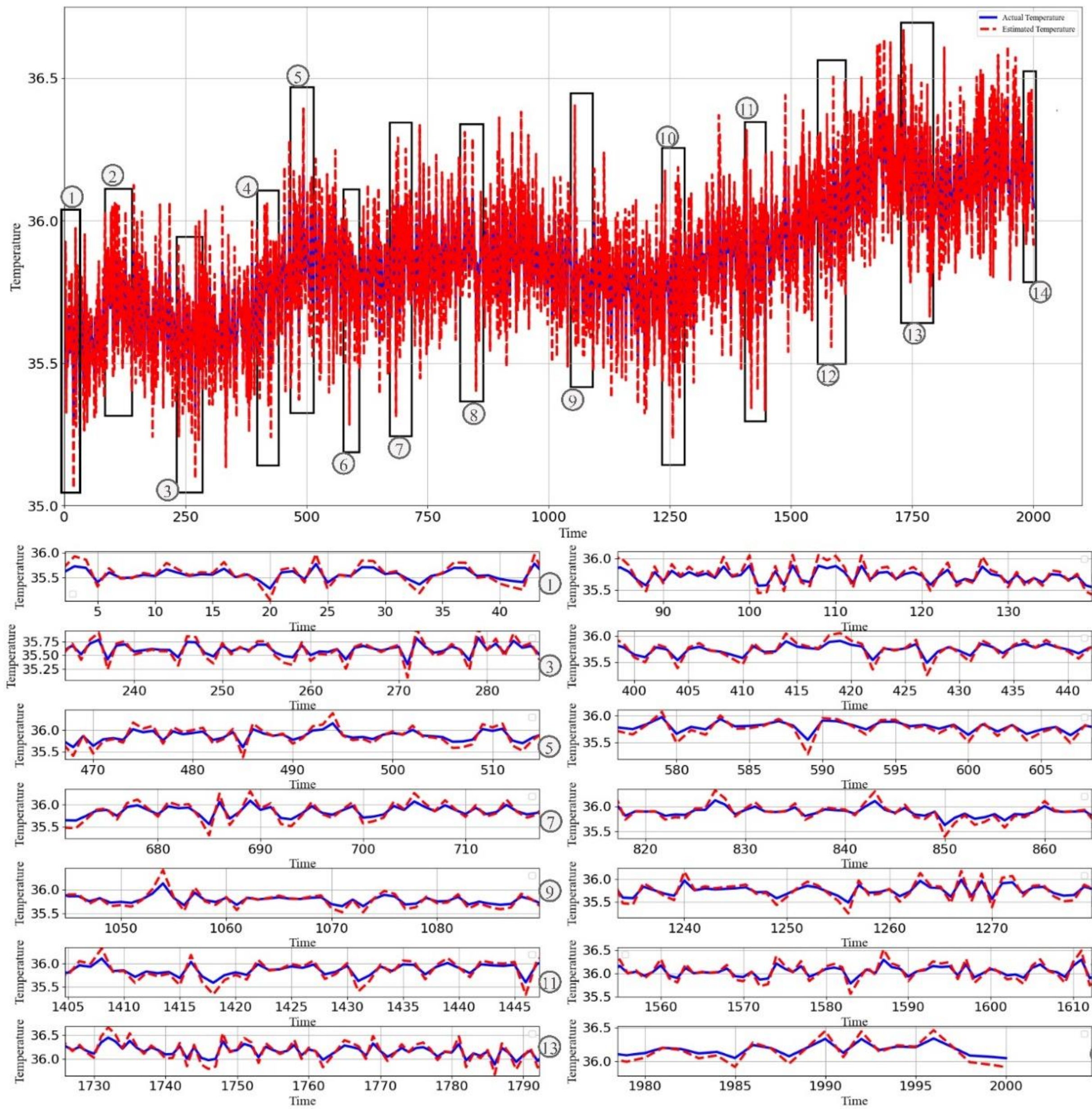


Figure 4.21. Instantaneous temperature measurements for subject 1 during session 3 to evaluate the effectiveness of the implemented variant.

As depicted in Figure 4.21, the top graph illustrates the overall temperature trend for Subject 1 throughout the duration of the experiment. The red dots represent actual temperature readings, while blue lines indicate estimated temperatures. Significant fluctuations are highlighted with numbered rectangles, which denote specific time intervals where notable changes are observed. The bottom subgraphs of Figure 4.21 provide detailed insets for each highlighted segment, allowing a closer examination of temperature variations within these intervals.

Figure 4.22. Instantaneous temperature measurements for subject 1 during session 5 to evaluate the effectiveness of the implemented variant.

Furthermore, to assess the performance of the temperature estimation variant at higher temperatures above 36.5°C, measurements from the fifth session of the same subject were analysed as depicted in Figure 4.22. The experiment started with a temperature of 36.5°C, which increased to 37.9°C within the first 100 seconds, followed by a decrease to 37.0°C over the next 200 frames. Herein, the temperature initially starts at 35.5°C and fluctuates through various phases over time.

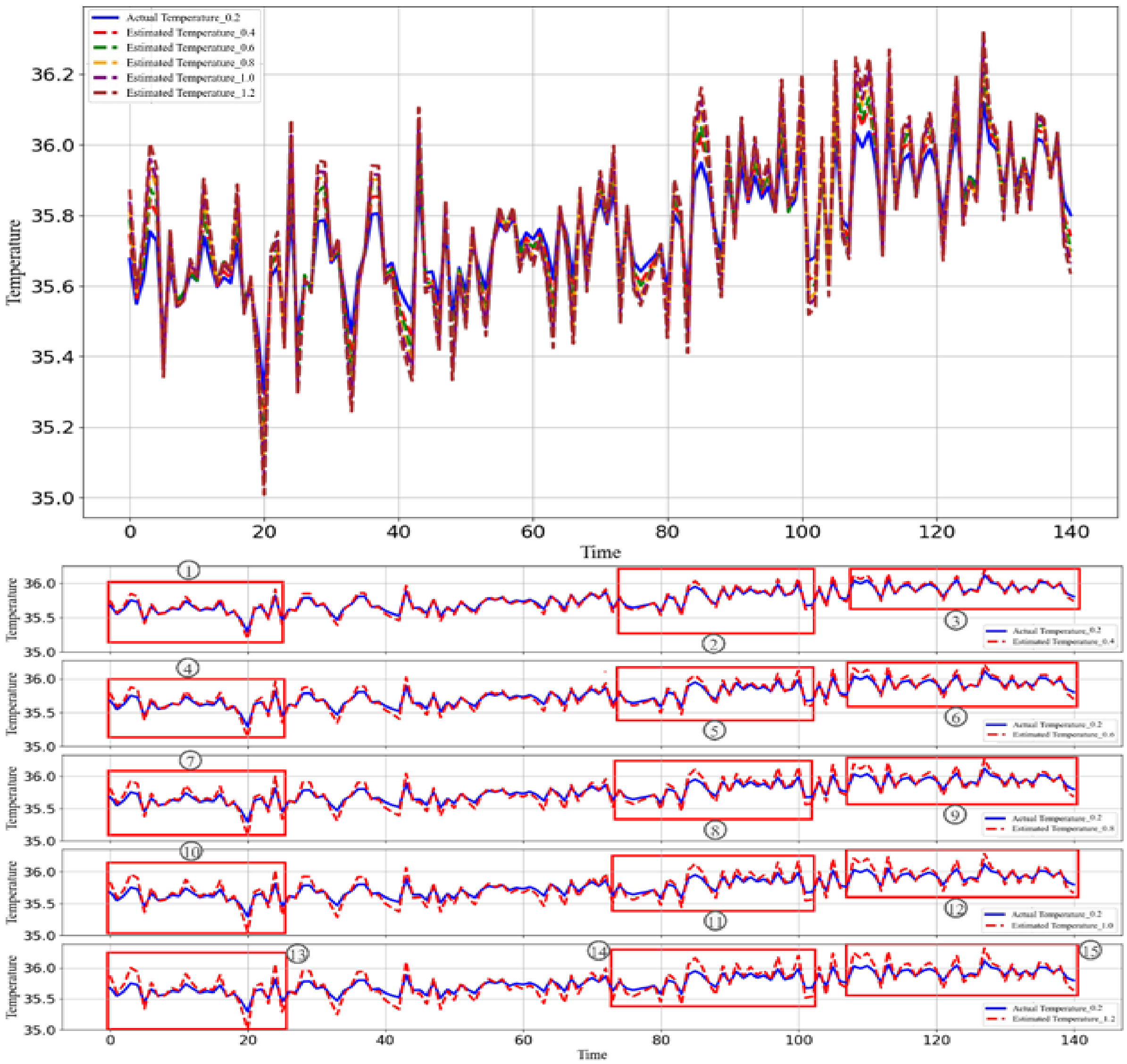


Figure 4.23. Assessing the impact of varying distances on the implemented variant in Case 1 and 2.

It increases to 35.7°C, decreases to 35.6°C, then rises to 35.9°C before decreasing again. After reaching multiple peaks and troughs, including drops to 35.7°C and 35.85°C, and rises to 35.95°C and 36.05°C, the temperature stabilizes at 36.20°C from 1900 seconds onward. The zoomed-in subplots offer a detailed perspective on temperature fluctuations for subject 1, revealing specific trends such as an initial rise from 35.5°C to 35.7°C within the first 40 seconds and variations between 800-860 seconds. These subplots highlight both increases and decreases in temperature over short intervals and demonstrate the alignment between actual and estimated temperatures, showcasing the reliability of the employed temperature estimation variant. The temperature then rose to 39.0°C, before decreasing to 38.0°C. It fluctuated between 38.0°C and 36.3°C in a series of linear increases and decreases before stabilizing at 36.3°C.

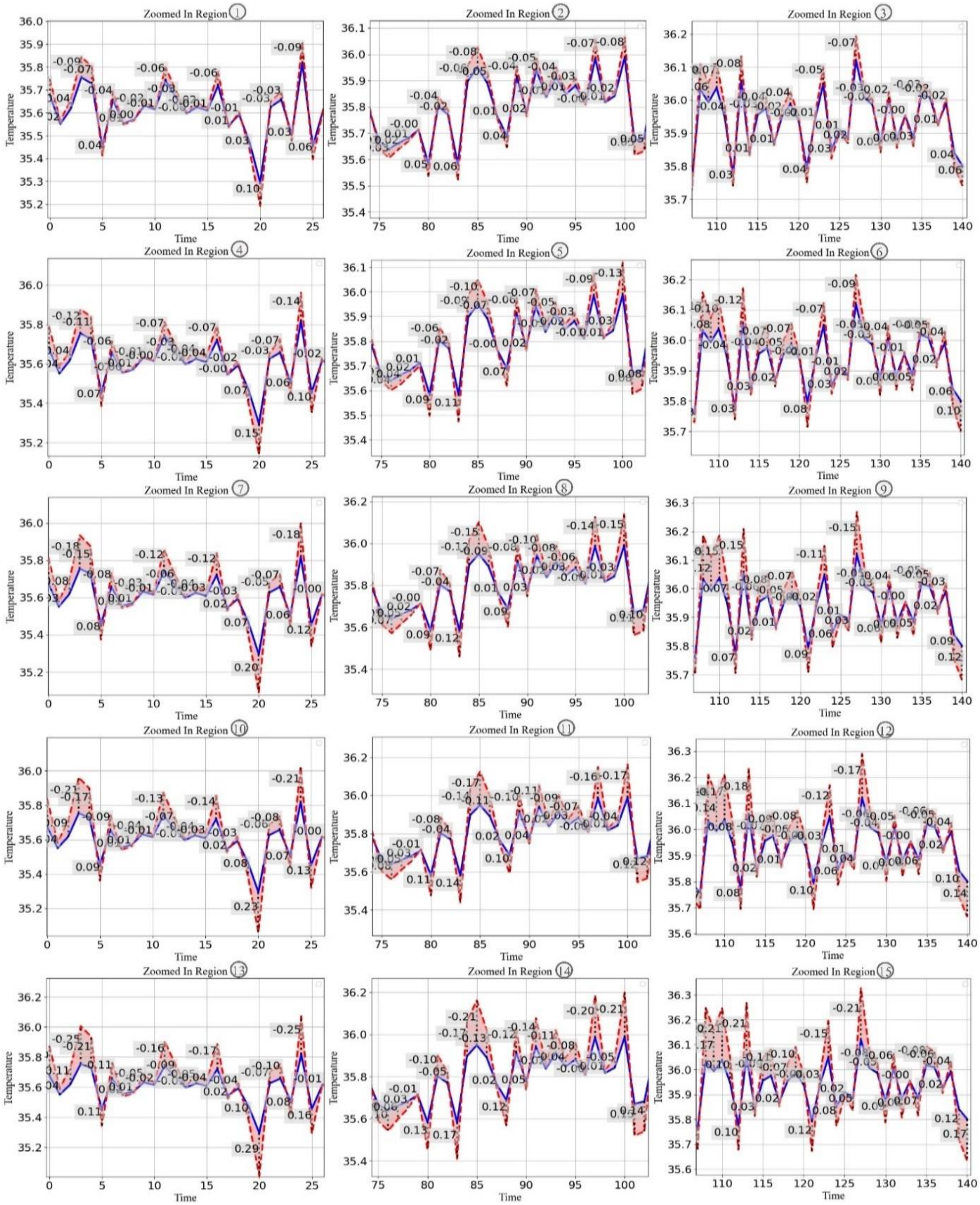


Figure 4.24. Error analysis of the designated red boxes in Figure 4.23 for Case 1 and 2.

This detailed analysis illustrates the variant's effectiveness in tracking and predicting temperature changes across a range of higher values. Upon assessing the performance of the applied variant, it is evident that the accuracy of detecting elevated temperatures is particularly effective. The data reveal a precise linear increase up to 39.0°C, followed by subsequent fluctuations and eventual stabilization. This observation implies that the subject possesses an effective capacity for tolerating and regulating elevated temperatures.

Furthermore, the variant proves to be proficient in monitoring temperature responses, accurately recording measurements even at temperatures exceeding 39°C. The sustained accuracy in capturing these temperature ranges, coupled with the detailed linear progression, underscores the reliability effectiveness of the measurement system throughout the experimental procedure. To assess the impact of distance on the same subject, further experiments were performed in session 11 at distances of 0.4, 0.6, 0.8, 1.0, and 1.2 cm, using a baseline distance of 0.2 cm as outlined in the specific cases previously detailed in Table 4.1 of Section 3.2.

For example, Figure 4.23 depicts the temperature estimation performance across Case 1 and Case 2. The top graph compares actual temperatures (blue) with estimated temperatures at various distances (0.2 to 1.2) in different colours. The bottom section includes zoomed-in sub-graphs that focus on specific regions of the top graph, highlighting areas with critical temperature variations. These detailed views, labelled ① through ⑮, show that the estimation variant consistently maintains accuracy even during rapid temperature changes. The red boxes highlight particularly notable intervals, underscoring the variant's capability to reliably detect both gradual and sudden temperature shifts.

Figure 4.24 provides an error analysis by quantifying the estimation errors within the selected red boxes from Figure 4.23. Each sub-graph details numerical error values for the corresponding zoomed-in regions. The analysis reveals that the effectiveness of the implemented variant diminishes as the distance from the baseline increases. At shorter distances (e.g., 0.4 cm and 0.6 cm), the estimation errors are relatively small, typically below 0.1 degrees in many regions. However, as the distance increases to 1.0 cm and 1.2 cm, the errors become more pronounced, often exceeding 0.15 degrees in several regions.

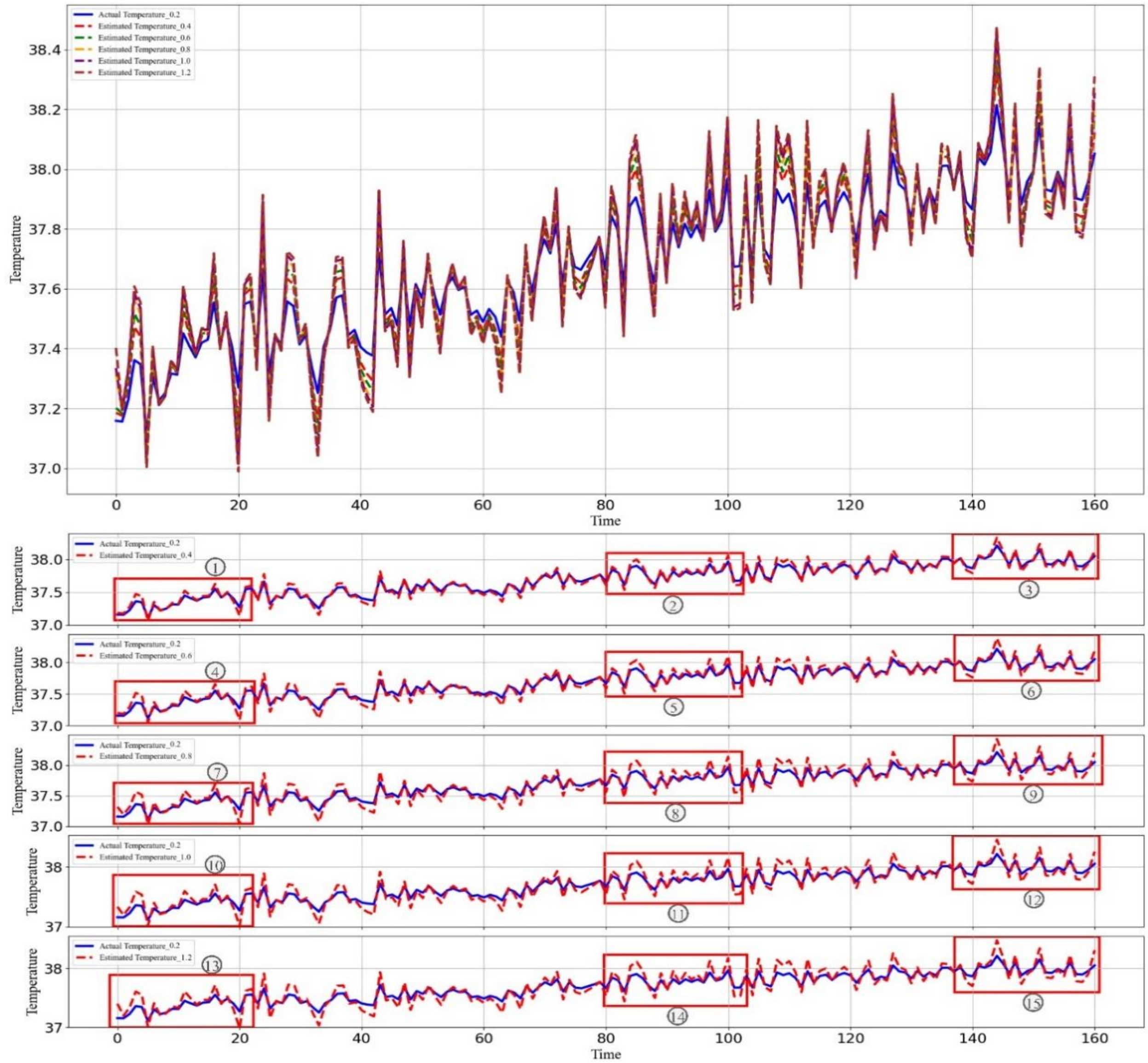


Figure 4.25. Assessing the impact of varying distances on the implemented variant in Case 4.

Similarly, Case 3 exhibited patterns comparable to those observed in Cases 1 and 2; however, the specific details of this case have been omitted for brevity. To elucidate the influence of varying distances at higher temperatures, Case 4 has been thoroughly examined with the findings depicted in Figure 4.25. The observed temperature values span from approximately 37.3°C to 38.9°C, exhibiting a general ascending trend over time. Specifically, Region ① features temperature oscillations between 37.2°C and 37.4°C, indicating a fluctuation of 0.2°C. In contrast, Region ④ demonstrates a more pronounced increase, with temperatures ranging from 37.6°C to 37.9°C, revealing a variation of 0.3°C. These sharp spikes and dips indicate transient events likely caused by external factors or subject activity, as noted in previous cases.

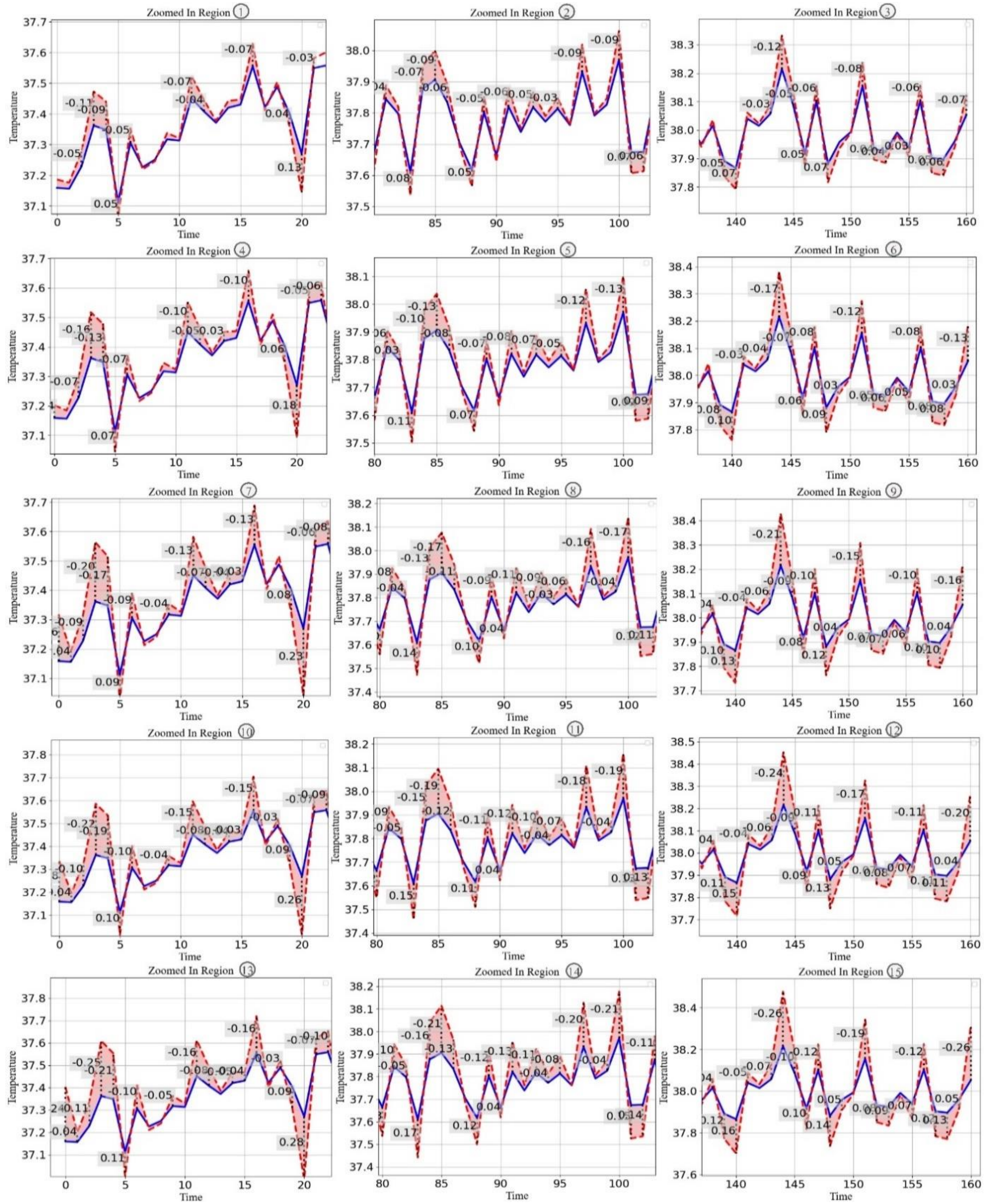


Figure 4.26. Error analysis of the designated red boxes in Figure 4.25 for Case 4.

As depicted in Figure 4.26, the subplots reveal the error dynamics within various regions of Figure 4.25, demonstrating that errors are neither constant nor uniformly distributed throughout the time series. For example, in Region ①, errors range from 0.0°C to 0.1°C, indicating minor discrepancies. In Region ②, the error values vary between 0.05°C and 0.12°C, demonstrating slightly larger deviations. Region ④ presents the highest inaccuracies, with errors escalating to between 0.15°C and 0.25°C.

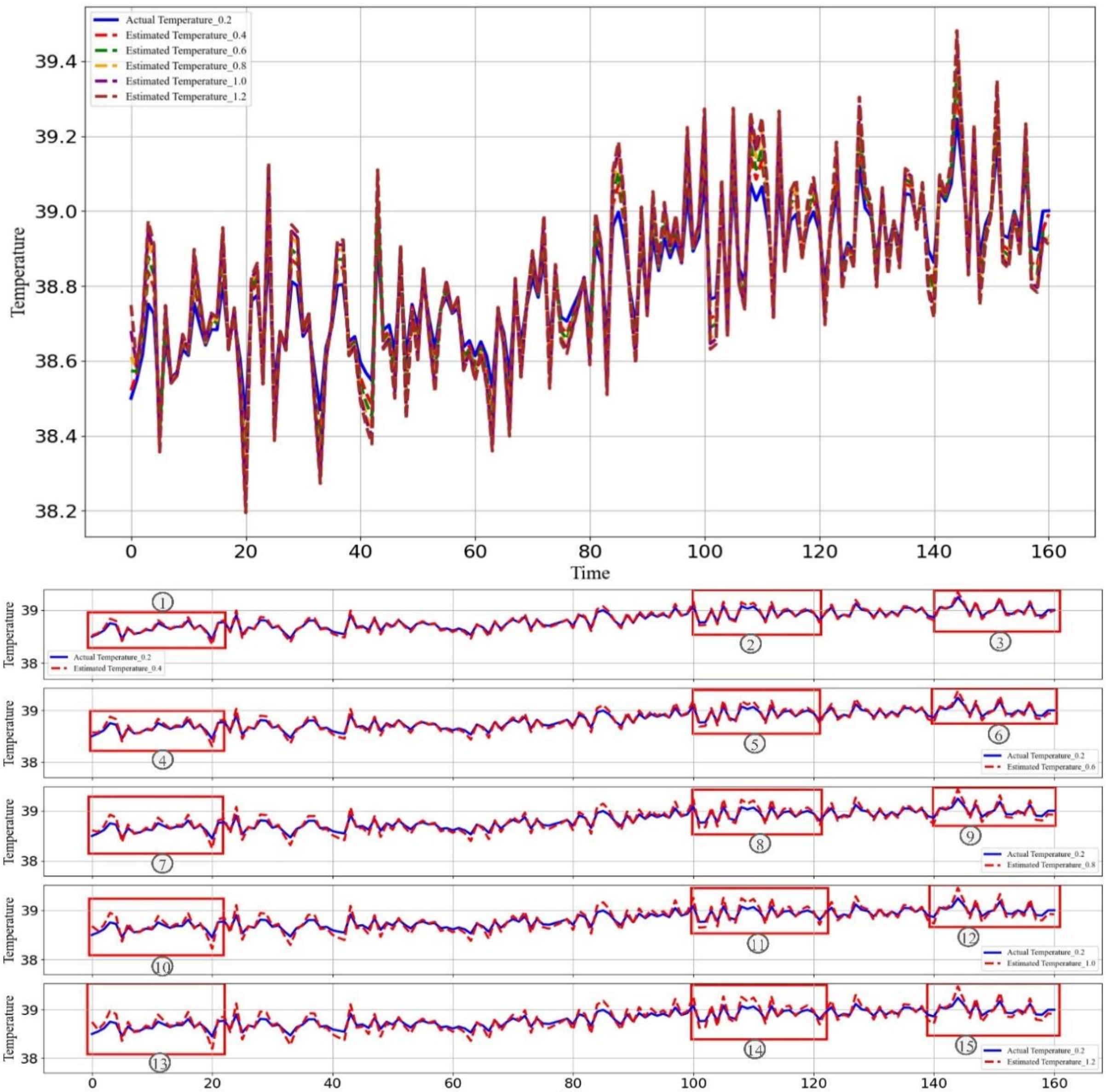


Figure 4.27. Assessing the impact of varying distances on the implemented variant in Case 6.

These findings highlight a pattern where discrepancies between estimated and actual temperatures become increasingly pronounced at greater distances, particularly within the highlighted regions of the error figure. Similar findings to those observed in Case 4 were also noted in Case 5; hence, the detailed specifics of Case 5 have been excluded for conciseness.

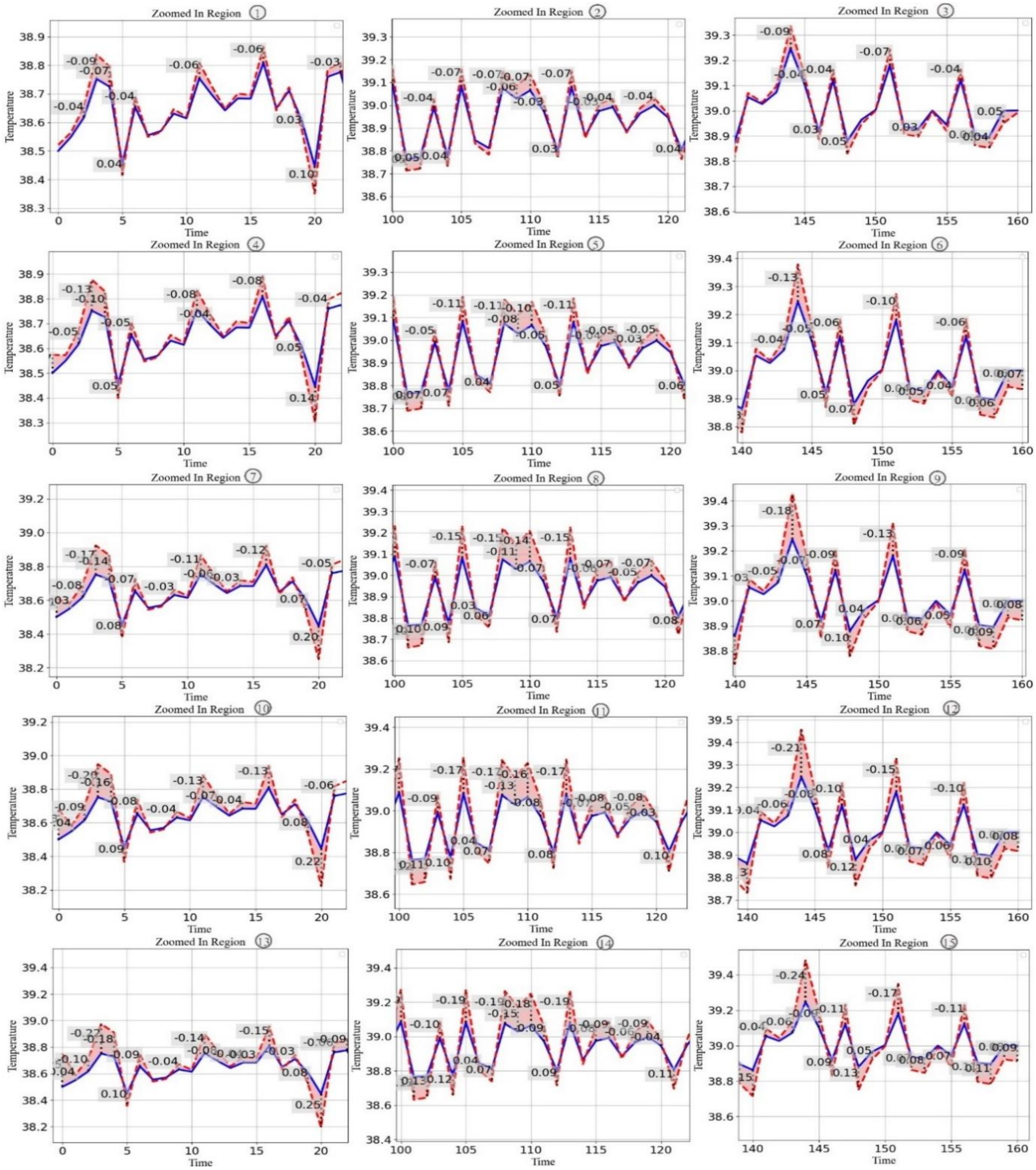


Figure 4.28. Error analysis of the designated red boxes in Figure 4.27 for Case 6.

Finally, to assess the influence of distance in Case 6, the relevant results are thoroughly presented in Figure 4.27. Herein, the actual temperature varying between 38.5°C and 39.4°C. The lower graphs in the same figure detail the error values corresponding to each distance parameter over time. As in previous experiments, the highlighted regions offer an in-depth analysis of error trends, thereby affirming the minimal discrepancies observed between the actual and estimated temperatures. More specifically, Figure 4.28 presents an enhanced view of specific regions from Figure 4.27, illustrating the nuanced error values associated with each distance parameter. In Region ①, corresponding to a distance parameter of 0.2, the error values over the time span from 0 to 20 are recorded as 0.06, 0.08, 0.07, and 0.04°C, with the maximum error being 0.08°C and the minimum error being 0.04°C. Region ②, with a distance parameter of 0.4, reveals errors between 0.07 and 0.05°C within the time interval from 105 to 120, where the maximum error is 0.07°C and the minimum is 0.05°C. For Region ③, with a distance parameter of 0.6, the recorded errors from 145 to 160 are 0.09, 0.08, and 0.06°C, showing a maximum error of 0.09°C and a minimum of 0.06°C. The errors associated with distance parameters of 0.8, 1.0, and 1.2 are similarly depicted in the zoomed-in regions of the error figure.

Collectively, these errors range from approximately -0.15 to 0.15 degrees Celsius. Despite occasional minor deviations, these discrepancies are promptly rectified in subsequent time points, demonstrating the variant's efficacy in adapting to temperature fluctuations. Upon concluding the analysis of all cases, it is evident that Case 6 yields more precise results compared to the other cases. This increased accuracy may be attributed to a reduction in random bodily movements of the subject and diminished influence of environmental factors. The pattern observed across all cases highlights the significant impact of distance on the accuracy of the implemented variant. Specifically, closer proximity to the baseline measurement correlates with more reliable temperature estimates, demonstrating that all cases exhibit sensitivity to distance. This finding aligns with expectations, and future refinements will focus on addressing and minimizing the adverse effects of distance.

4.6.3. Qualitative Analysis Results

While quantitative analysis enables the assessment of the overall effects of the proposed variant, it does not offer detailed information regarding the underlying reasons for user perception and the monitoring of personalized HR events. Specifically, it does not elucidate how the personalized HR assesses events at specific time intervals and intervenes accordingly. Therefore, an additional case study conducted in Session 12 of Subject 1 aimed to provide a detailed examination of each robotic event, allowing for an analysis of how interactions between the personalized HR and the subject occurred and varied throughout time. During that session the subject had a higher temperature value compared to the previous ones, crossing a critical threshold on two instances which resulted in a call to doctor shown red dotted points in Figure 4.29.

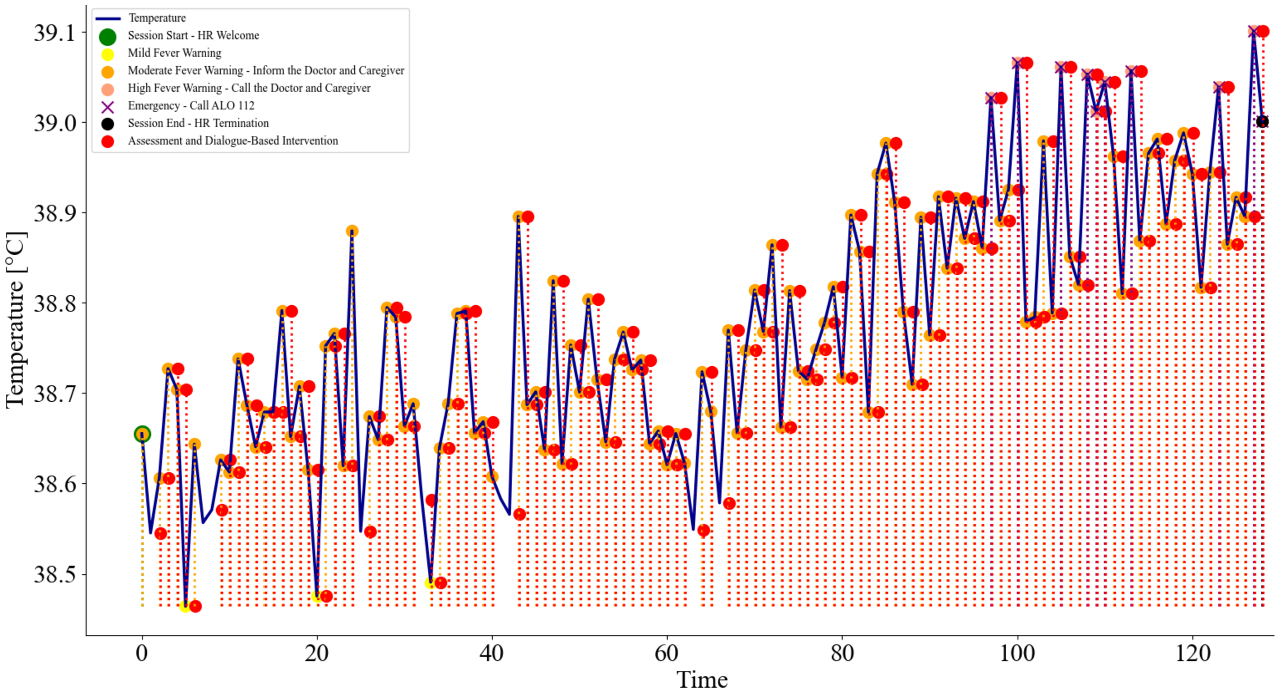


Figure 4.29. Estimated patient temperatures and corresponding HR events during the "critical high temperature" session.

The provided graph illustrates the subject's temperature fluctuations over a specific period, along with corresponding HR interventions and alerts. Initially, as subject's temperature shows a mild fever, the HR issues a "Mild Fever Warning" (yellow circles), alerting the caregiver to monitor his condition closely. As time progresses, the temperature spikes to moderate levels (orange circles), prompting the HR to notify both the doctor and the caregiver. This pattern of moderate fever warnings and subsequent temperature dips indicates that initial interventions may be having some effect, though the underlying condition causing the fever is persistent.

Despite the HR's timely interventions, subject's temperature continues to escalate, reaching high fever levels (red circles), where the HR issues a "High Fever Warning" and advises immediate medical consultation. Eventually, the temperature reaches a critical point where the HR issues an

"Emergency" alert (purple crosses), directing the medical team to call emergency services (ALO 112). Throughout this period, the HR also marks the session start (green circles) and end (black circles) to document intervention timelines. The HR's sophisticated real-time monitoring ensure that each spike in fever is met with appropriate and escalating responses, ultimately aiming to stabilize the subject's condition through continuous assessment and dialogue-based interventions (red dotted lines).

Upon the intervention, the subject reported to the HR and related doctor and caregiver, feeling dizzy and very tired and continued the session with ice bag to progressively decrease the temperature value, as required. Following this session, the subject's temperature gradually decreased and stabilized at a normal level as a result of the implemented interventions. Furthermore, following the completion of the relevant session, the UTAUT questionnaire is administered to Subject 1. The results of the questionnaire indicated that the subject perceived both the HR and the experiment very positively, with 93.2% of the questions receiving a response of strong agreement and the remainder receiving agreement. Subject 1's response to the open-ended questions in the UTAUT further corroborate this positive perception. For instance, he stated, "I would recommend using the HR as it is a great help during my high temperature management process." Additionally, he noted, "Occasionally, I experienced delays in the HR's responses, which could be improved. Beyond this issue, I would not propose any modifications to the current system. The HR interacts positively with me, assisting both myself and the medical staff, and serves as an effective collaborator".

After analysing the UTAUT and open-ended questionnaire responses for the subject 1, the HR interacts with another subject, as illustrated in Figure 4.30. In this session, the subject 2 has normal temperature values and no health issues, and has also consented to share her visual information. As shown in Figure 4.30, number 1 indicates the real-time temperature information obtained from the thermal camera at position 2. The number 3 represents the HR involved in the interaction, while number 4 shows the observer interface displaying the real-time temperature information obtained from the subject. This allows the observer to monitor the subject's interaction with the HR without interfering with the experiments. Finally, number 5 illustrates the HR interface displaying the real-time temperature values extracted from the subject.

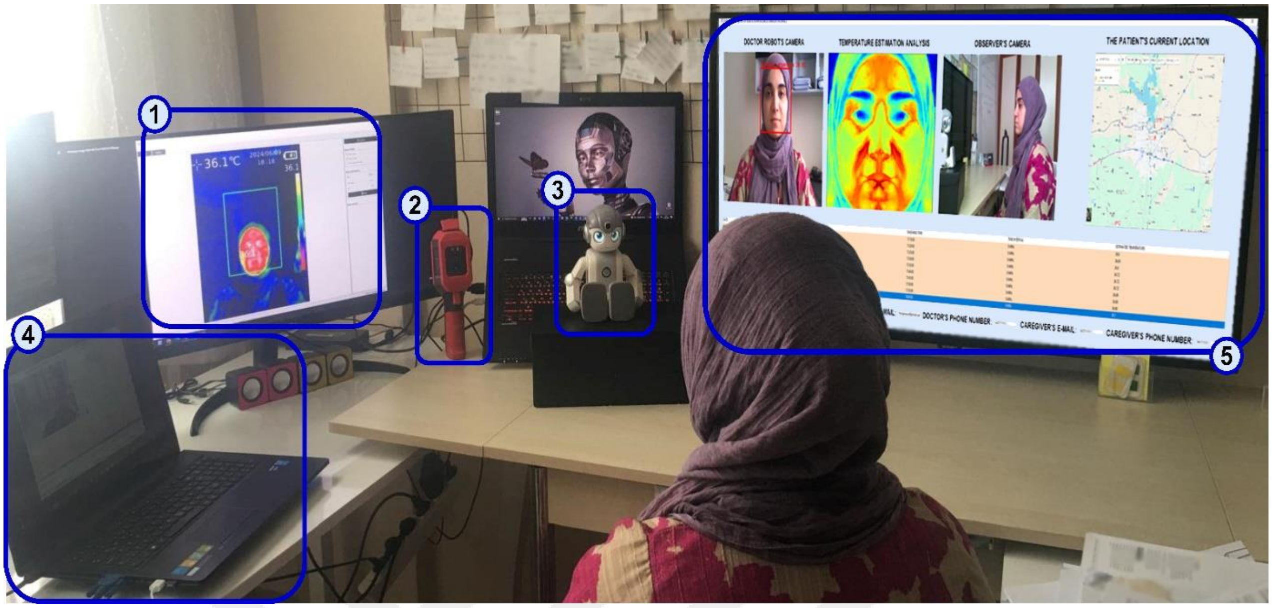


Figure 4.30. Experimental configuration for subject 2 and HR engagement in accordance with estimated temperature values.

As illustrated in Figure 4.30, during the initial phase of this experiment, the HR introduces itself and gathers temperature data from the subject, subsequently presenting the results.

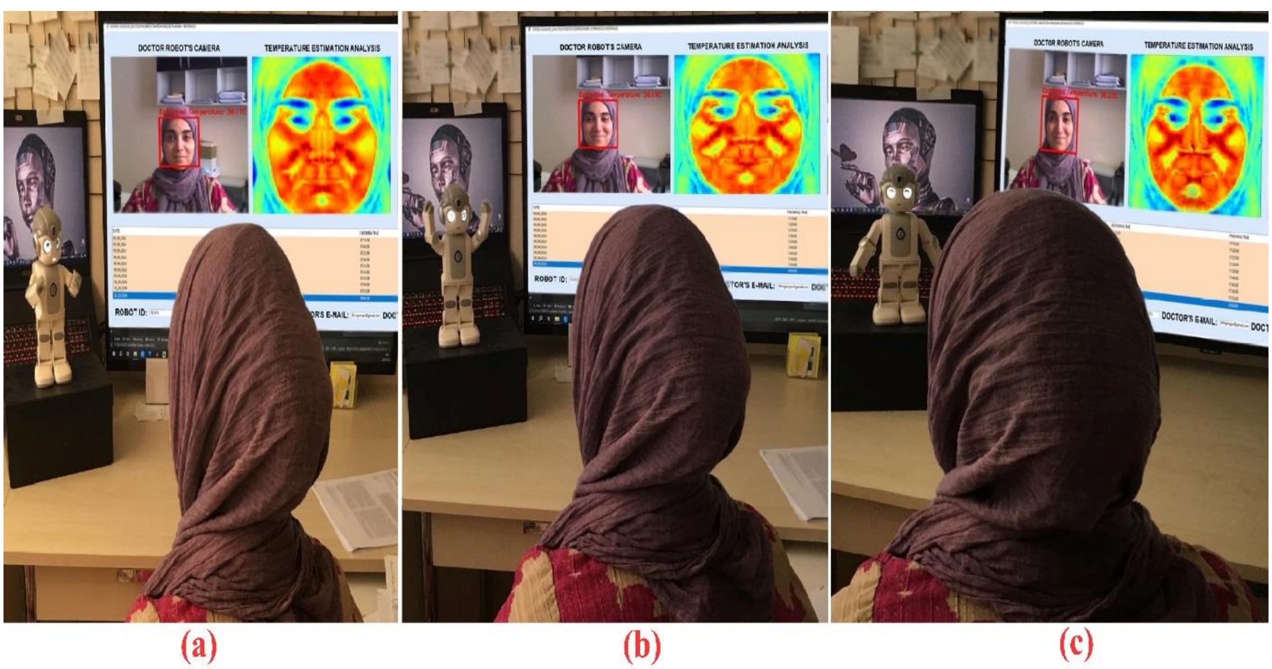


Figure 4.31. HR engagement with the subject 2 in accordance with estimated temperature values.

Since the results fall within the normal temperature range, it prompts the subject to offer additional suggestions for maintaining a healthy lifestyle as depicted in Figure 4.31(a). Upon receiving positive feedback, the HR transitions to a standing position and begins to provide suggestions utilizing its ChatGPT integration and gestural movements as shown in Figure 4.31(b). Finally, after providing the relevant advice, the HR inquiries about the perceived effectiveness of the suggestions and awaits a response. Upon the subject expressing her appreciation, the HR displays its satisfaction for contributing positively to the subject's well-being as illustrated in Figure 4.31(c).

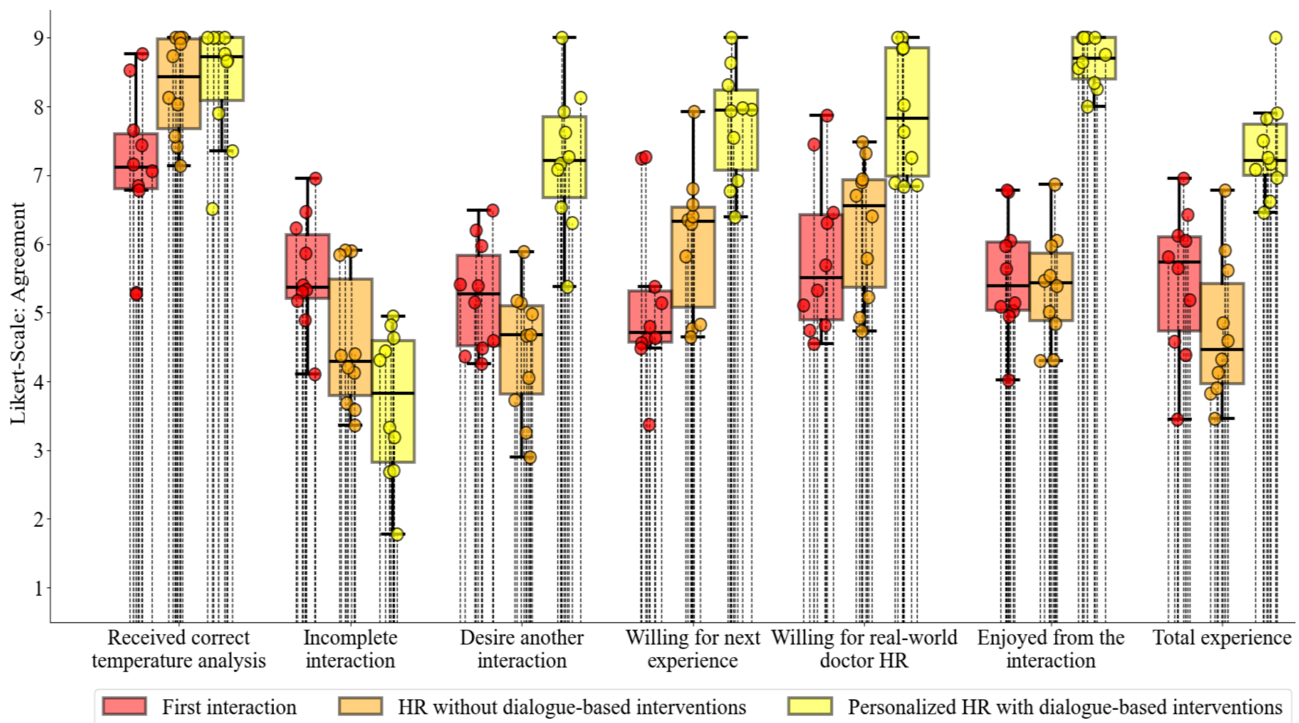


Figure 4.32. The Likert-scale agreement results for 10 subjects.

Following the completion of this last session, a Likert-scale questionnaire is administered to all 10 subjects to gauge overall user perceptions and assess the effects of the dialogue-based interventions. The following Figure 4.32 illustrates the Likert-scale agreement levels (1-9) across seven interaction metrics for three distinct conditions: First Interaction (red), HR without dialogue-based interventions (orange), and Personalized HR with dialogue-based interventions (yellow).

For the "Received correct temperature value," Base HR and Personalized HR have medians around 8.5 and 9, respectively, with low variability, while the First Interaction median is about 7. "Incomplete interaction" and "Desire another interaction" show higher variability, especially for

Personalized HR, but with higher medians (around 8) compared to the First Interaction (around 6). The "Willing for next experience" and "Willing for real world doctor HR" metrics reveal a clear advantage for Personalized HR, with medians near 8.5-9, while the First Interaction lags at medians of 5. "Enjoyed from the interaction" and "Total experience" again highlight the superior performance of the personalized HR, with medians of 8.5-9 and minimal variability, compared to lower medians and higher variability for the other conditions. Overall, “Personalized HR with dialogue-based interventions" consistently achieves the highest agreement scores and least variability, indicating a more favourable and consistent user experience

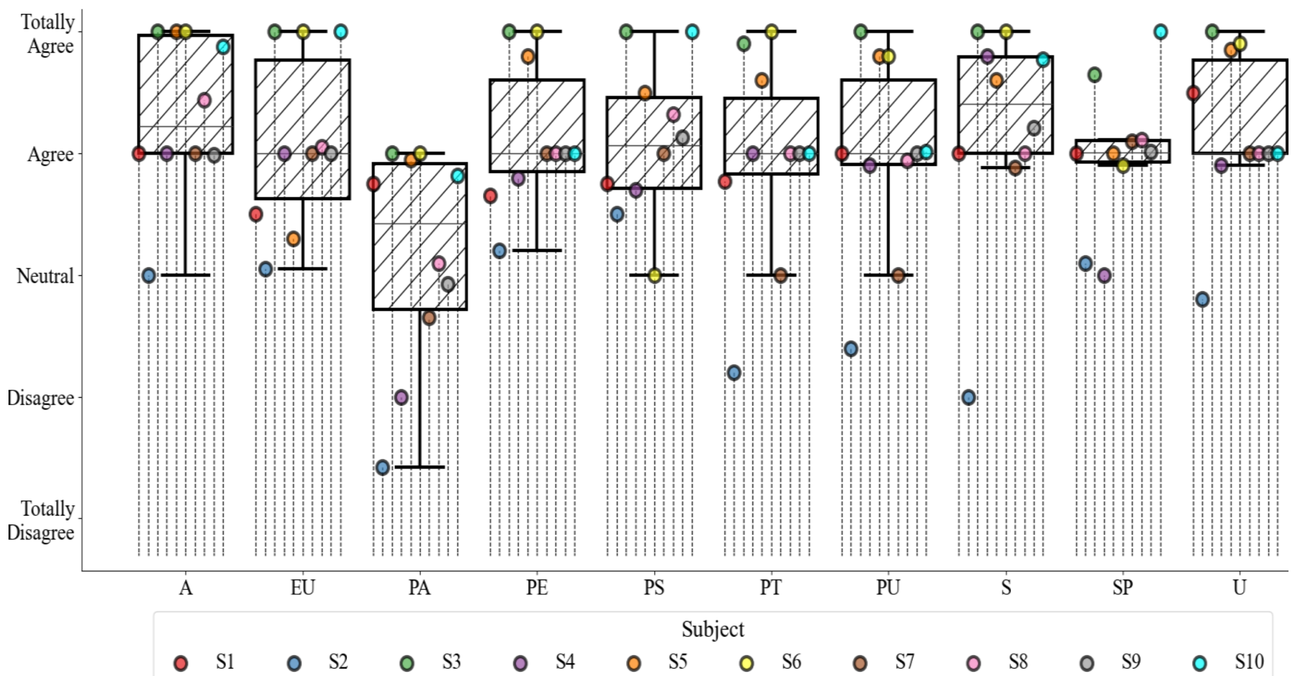


Figure 4.33. The modified UTAUT questionnaire results for 10 subjects to measure the general efficiency of personalized HR events.

Furthermore, the modified UTAUT questionnaire with supplementary questions is also administered to these 10 subjects to measure the general efficiency of personalized HR events. The results indicate a generally positive perception of the personalized HR among patients. The constructs such as U with a mean (M) of 4.20 and a standard deviation (SD) of 0.65, S with M = 4.23 and SD = 0.80, and PS with M = 4.09 and SD = 0.71, were rated highly by the patients. This suggests that the patients found the HR to be useful, safe, and sociable, contributing to an overall positive user experience. The high scores in these areas reflect the HR's effectiveness in providing valuable and reliable interactions that patients appreciate. The A received a mean score of around 4.00 with a standard deviation of 0.70, indicating a generally positive but somewhat varied perception of the HR's appeal. The PU scored a mean score of 4.15 with a standard deviation of 0.68 would suggest that the HR is perceived as practically useful, with a relatively consistent agreement among subjects. The PT had a mean score of 4.15 and a standard deviation of 0.60, reflecting a high and consistent level of confidence in the HR's reliability and effectiveness.

While the general perception is positive, certain constructs such as PA and SP showed more neutral responses. The mean for PA was 3.17 with an SD of 0.94, and for SP, the mean was 3.89 with an SD of 0.70. These results suggest that while the HR was seen as generally adaptive and present, there is room for improvement in these areas. Enhancing the HR's ability to adapt to individual patient needs and increasing its perceived presence could further improve user satisfaction and engagement. Constructs related to the HR's interactive and enjoyable aspects, such as PE EU, also received positive feedback with means of 4.05 and 4.08, and standard deviations of 0.64 each. These findings indicate that the subjects found the HR to be easy to use and enjoyable, which are critical factors for sustained engagement and adherence. The results align with the overall positive sentiment towards the HR, highlighting the importance of maintaining high levels of usability and enjoyment to ensure continued patient interaction and satisfaction Overall, the high scores across most constructs confirm that the personalized HR was well-received by the subjects. These findings validate the HR's potential to provide effective, engaging, and personalized therapeutic support.

4.7. Discussions

In this section, a new four-stage personalized temperature estimation algorithm is introduced along with its implementation in HRs. The initial stage employs the Viola-Jones algorithm for precise face detection across RGB and its complementary LAB and mapped thermal image modalities. In the subsequent stage, a modified Grey Wolf Optimizer-based hierarchical vision transformer architecture with shifted-window mechanism is applied to each modality for the extraction of relevant thermal features. These extracted features are then processed through a tripled multi-modal fusion architecture to accurately estimate human temperature. In the final stage, the entire algorithm is integrated into HRs by encompassing a custom graphical user interface and a dialogue model based on Bayesian decision-tree guided finite state machines.

The effectiveness of the developed algorithm is meticulously validated through an in-depth analysis of each decision point across all stages with a focus on evaluating temperature estimation processes derived from the learned human faces. Furthermore, an extensive assessment of both the computational load and accuracy of the algorithm is conducted to demonstrate its overall efficacy and performance. By introducing a detailed new framework, this paper offers researchers an opportunity to advance their understanding of this rapidly evolving domain. It serves as a valuable resource for academic endeavours, offering in-depth insights and facilitating future advancements in assistive robotics within HRI-based health-care applications.

Future research will focus on enhancing the proposed algorithm by mitigating the impact of distance on temperature measurement accuracy. Additionally, the algorithm will be implemented with a larger participant cohort, particularly focusing on individuals living alone or in care facilities. Moreover, the algorithm's capabilities will be expanded to include the monitoring of other critical physiological indicators, such as heart rate and blood pressure measurements derived from facial analysis, thereby contributing to a more comprehensive array of health assessments.

5. DESIGN AND DEVELOPMENT OF A SPEECH-BASED EMOTION ANALYSIS ALGORITHM FOR QUARANTINED COVID-19 PATIENTS

This section outlines a three-stage speech-based emotion analysis algorithm and demonstrates its application on the NAO HR. Initially, the algorithm extracts key speech features using MFCC. These features are then classified into emotional statements through a LSTM-RNN. The final stage involves assessing engagement through decision tree structures.

5.1. Feature Extraction using MFCC Technique

To identify the most informative speech features for emotion analysis, feature extraction is crucial. This section employs the MFCC technique, which captures perceptual characteristics of human hearing, to extract key speech features. The process begins by applying a first-order difference equation to the speech samples $(s_n,\ n=1,\ldots,N)$ to enhance higher frequencies and increase the signal's energy as:

$$x(n)=s(n)-\alpha s(n-1) \tag{5.1}$$

wherein, α represents the pre-emphasis coefficient, which should be in the range $0\le\alpha<1$.

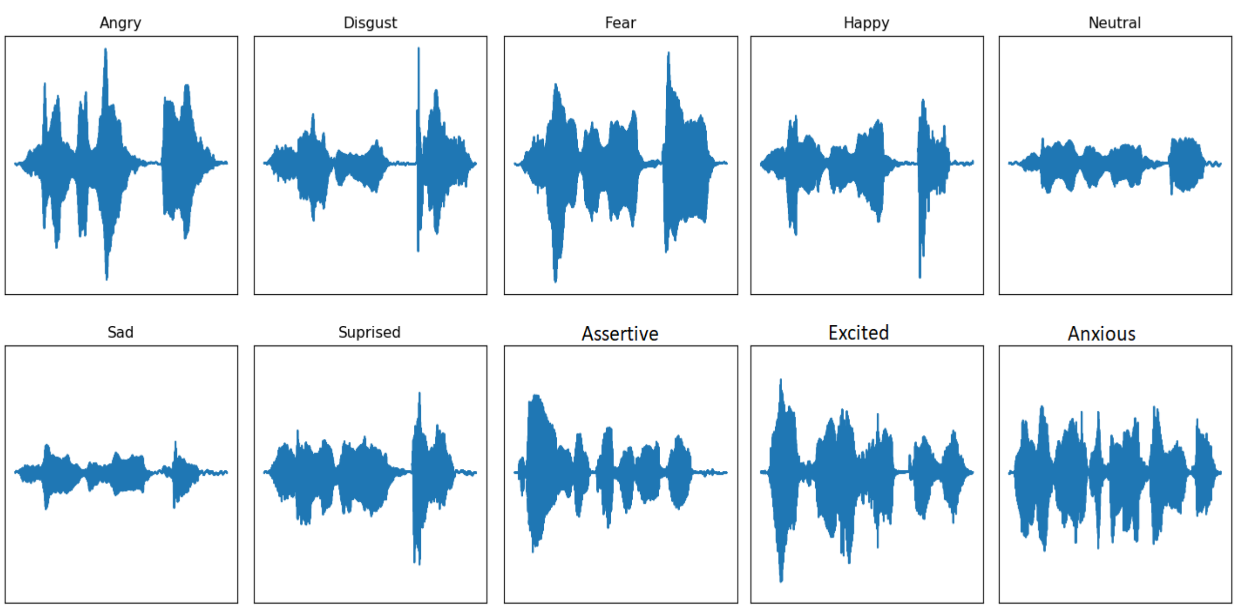


Figure 5.1. Pre-emphasized speech signals for 10 emotional statements of one subject.

For each of the 10 emotional statements (angry, disgust, fear, happy, neutral, sad, surprised, assertive, excited, and anxious) of a subject, the pre-emphasized speech signals are depicted in Figure 5.1. Subsequently, these emphasized signals are processed frame-by-frame, with each frame ranging from 20 ms to 30 ms. To address discontinuities due to segmentation and overlapping frames, windowing is applied to these frames, as shown in Figure 5.2. A key consideration is that signals with reduced discontinuities must have their ends sufficiently smoothed to seamlessly connect with the beginning of the previous frame.

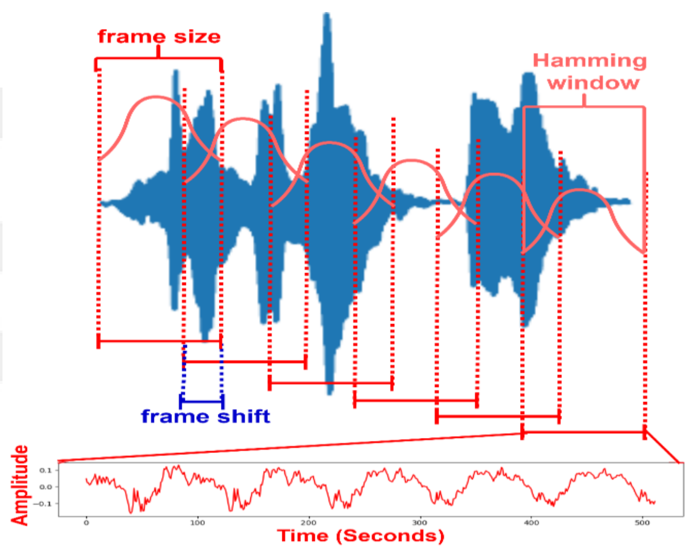


Figure 5.2. Illustration of framing and windowing processes of the emphasized speech signal.

In the subsequent step, a Hamming window is applied to taper the framed speech signal to zero at the beginning and end of each frame using the following equation:

$$w(n)=\begin{cases} 0.5-0.5\cos\left(\dfrac{2\pi n}{N-1}\right) & 0\le n\le N-1 \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

In this context, windowing involves multiplying the window function $w(n)$ by the framed speech signal $x(n)$ to obtain the windowed speech signal $s_w(n)$ as follows:

$$s_w(n)=x(n)w(n) \tag{5.3}$$

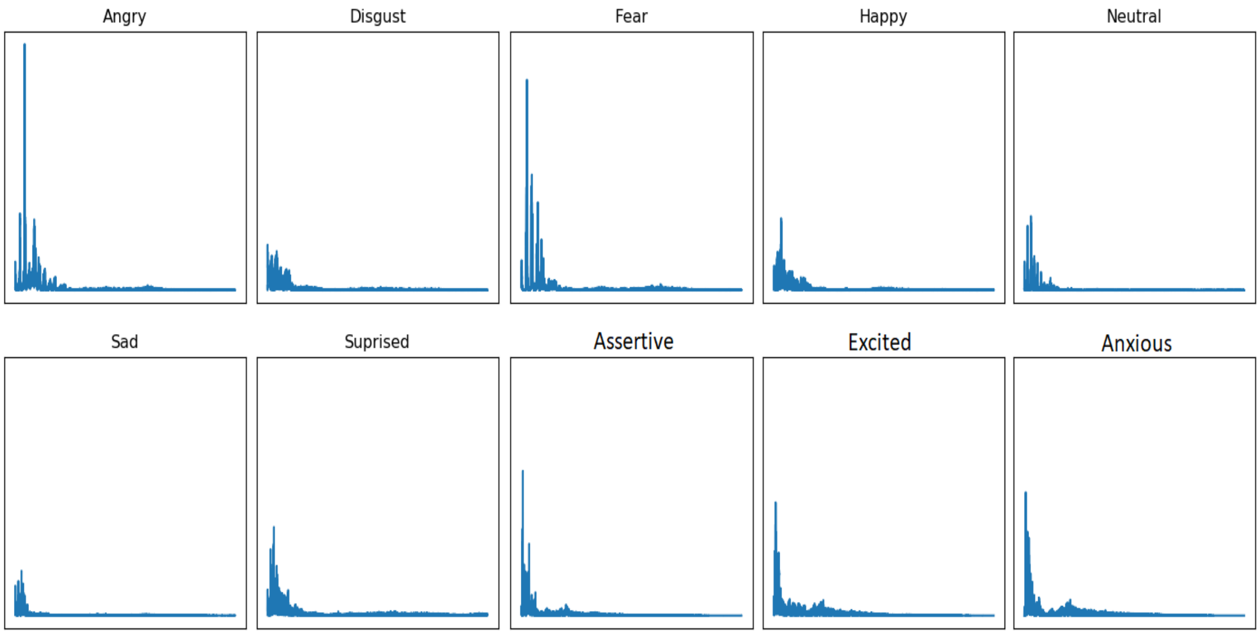


Figure 5.3. DFT results of windowed speech signals for 10 emotional statements of one subject in Figure 5.1.

The DFT of the windowed speech signal is computed using the following equation:

$$s_w(k) = \sum_{n=0}^{N-1} s_w(n) e^{\frac{-j2\pi kn}{N}} \tag{5.4}$$

where $1 \le k \le N$.

The goal of this step is to convert each frame of N windowed samples from the time domain to the frequency domain. Figure 5.3 displays the DFT results of the windowed speech signals for 10 emotional statements from one subject. Subsequently, the power spectrum of the DFT is calculated as follows:

$$s_e = \left| s_w(k) \right|^2 \tag{5.5}$$

The power spectrum obtained from the DFT is subsequently binned through correlation with triangular filters, aligning it with the frequency resolution of human auditory perception, as depicted in Figure 5.4. Binning involves multiplying the power spectrum coefficients by the triangular filter gains and summing the results to derive the Mel-Spectral coefficients. The frequency scale is converted from Hertz to the Mel-Scale using filter banks, which are linearly spaced at lower frequencies and logarithmically spaced at higher frequencies. This approach is designed to capture phonetically relevant features of speech, reflecting human auditory perception.

Following this, a filter bank is constructed with U triangular filters of unit height, evenly spaced on the Mel-Scale, as follows:

$$f_{mel} = 2595\log_{10}\left(1+\frac{f}{700}\right) \tag{5.6}$$

where f_{mel} represents the Mel-scale pitch corresponding to the actual frequency f in Hertz.

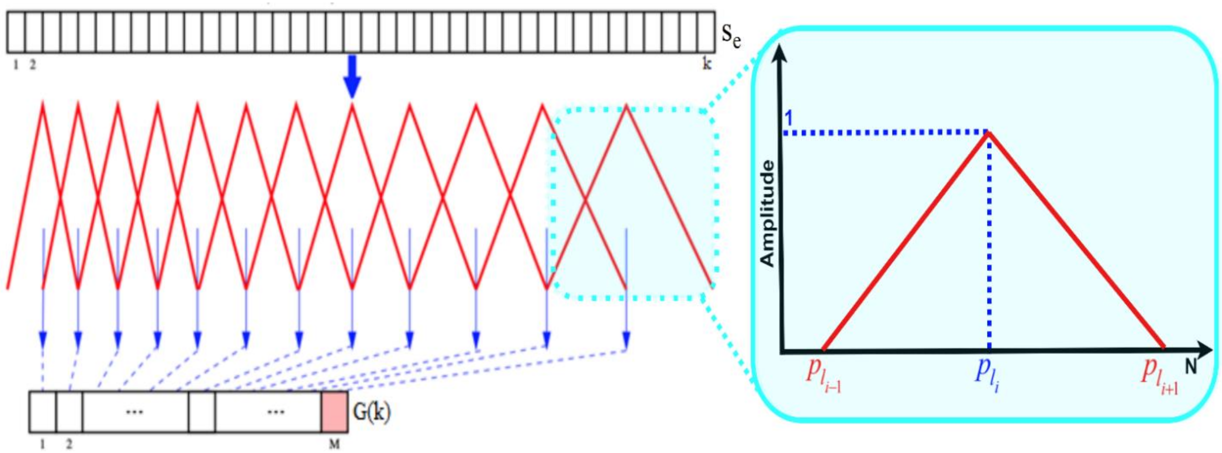


Figure 5.4. Illustration of the binning process and $\zeta_i(k)$ filter response of a typical Mel-Scale filter.

The filter response $\zeta_i(k)$ of the i^{th} filter in the bank is defined by four constraints as follows:

$$\zeta_i(k) = \begin{cases} 0 & p \le p_{l_{i-1}} \\ 0 & p \ge p_{l_{i+1}} \\ \dfrac{p-p_{l_{i-1}}}{p_{l_i}-p_{l_{i-1}}} & p_{l_{i-1}} \le p \le p_{l_i} \\ \dfrac{p_{l_{i+1}}-p}{p_{l_{i+1}}-p_{l_i}} & p_{l_i} \le p \le p_{l_{i+1}} \end{cases} \tag{5.7}$$

where $1 \le i \le U$, $\{p_{l_i}\}_{i=0}^{U+1}$ denotes the limit points of the corresponding filters, and k represents the coefficient index within N.

The filter bank limit points $\{p_{l_i}\}_{i=0}^{U+1}$ are uniformly distributed on the Mel-Scale, as described by the following equation.

$$p_{l_i} = \left(\frac{N}{F_s}\right) f_{mel}^{-1}\left[f_{mel}(f_{low}) + \frac{i\{f_{mel}(f_{high}) - f_{mel}(f_{low})\}}{U+1}\right] \tag{5.8}$$

where N denotes the sample size, F_s represents the sampling frequency, f_{high} and f_{low} are the high and low frequency limits of the filter bank.

Additionally, f_{mel}^{-1} denotes the reciprocal of f_{mel} as illustrated by the subsequent equation.

$$f_{mel}^{-1}\left(f_{mel}\right)=700\left[10^{\frac{f_{mel}}{2595}}-1\right] \tag{5.9}$$

where $k=0,1,\ldots,k-1$ and $n=0,1,2,\ldots,N/2$.

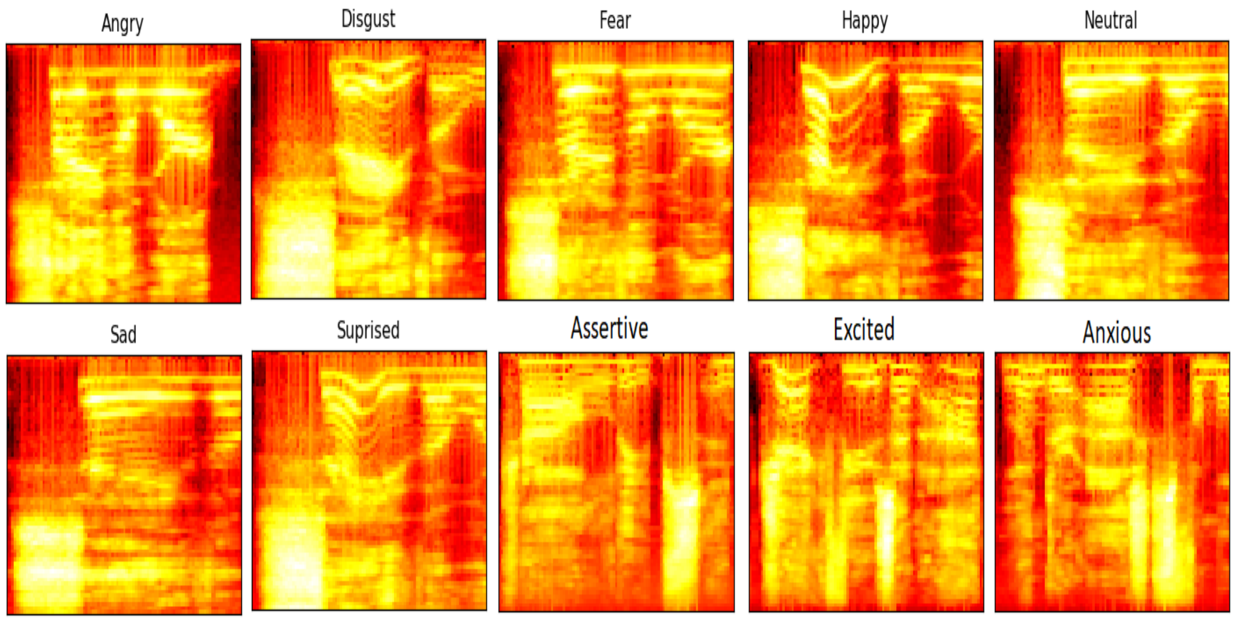


Figure 5.5. Extracted filter bank coefficients derived from $G(k)$ for 10 emotional statements of one subject in Figure 5.1.

The Mel-Spectral coefficients $G(k)_{i=0}^{U}$ are derived by performing a weighted summation of the corresponding filter responses $\zeta_i(k)$ and energy spectrum $s_e(k)$ as follows:

$$G(k)=\sum_{n}^{N/2}\zeta_i(k)s_e(k) \tag{5.10}$$

Subsequently, the logarithm of the Mel-Spectral coefficients $G(k)$ is applied to smooth out undesired ripples in the spectrum, as follows:

$$e_k=\log G(k) \tag{5.11}$$

Finally, the Discrete Cosine Transform (DCT) is applied to the logarithm of the Mel-Cepstrum e_k to extract the corresponding coefficients, as follows:

$$C_m = \sqrt{\frac{2}{U}} \sum_{j=0}^{U-1} e_k \cos\left(m \left(\frac{2j-1}{2} \right) \frac{\pi}{U} \right) \tag{5.12}$$

where $0 \le m \le d-1$, and d represents the desired number of cepstral coefficients.

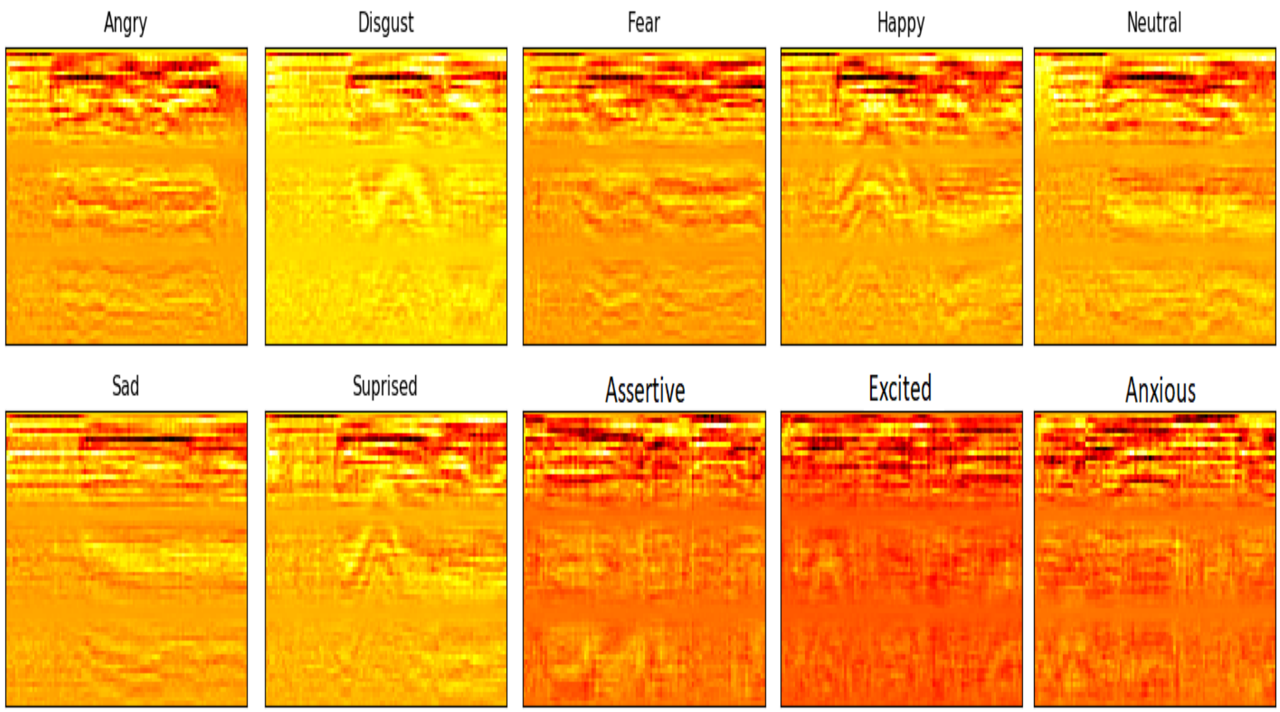


Figure 5.6. Extracted Mel-Cestrum coefficients derived from DCT for 10 emotional statements of one subject in Figure 5.1.

In this study, the initial 19 cepstral coefficients are preserved, and their first-order derivatives are calculated for each speech signal. This results in a feature vector comprising 38 elements (19 coefficients and their 19 derivatives), which is then used in the subsequent classification step.

5.2. Classification using LSTM-RNN Structure

In the final stage of the proposed algorithm, emotion classification is conducted using an LSTM-RNN. Specifically, it employs a deep RNN with two layers, each containing 100 LSTM cells, to effectively classify emotions from speech features extracted via the MFCC technique. The input layer, as depicted in Figure 5.7, receives 38 units of speech features, including coefficients and their derivatives, at each time step. The subsequent layers comprise the LSTM structure, where each unit incorporates input g_t^i , forget g_t^f and output gate g_t^o .

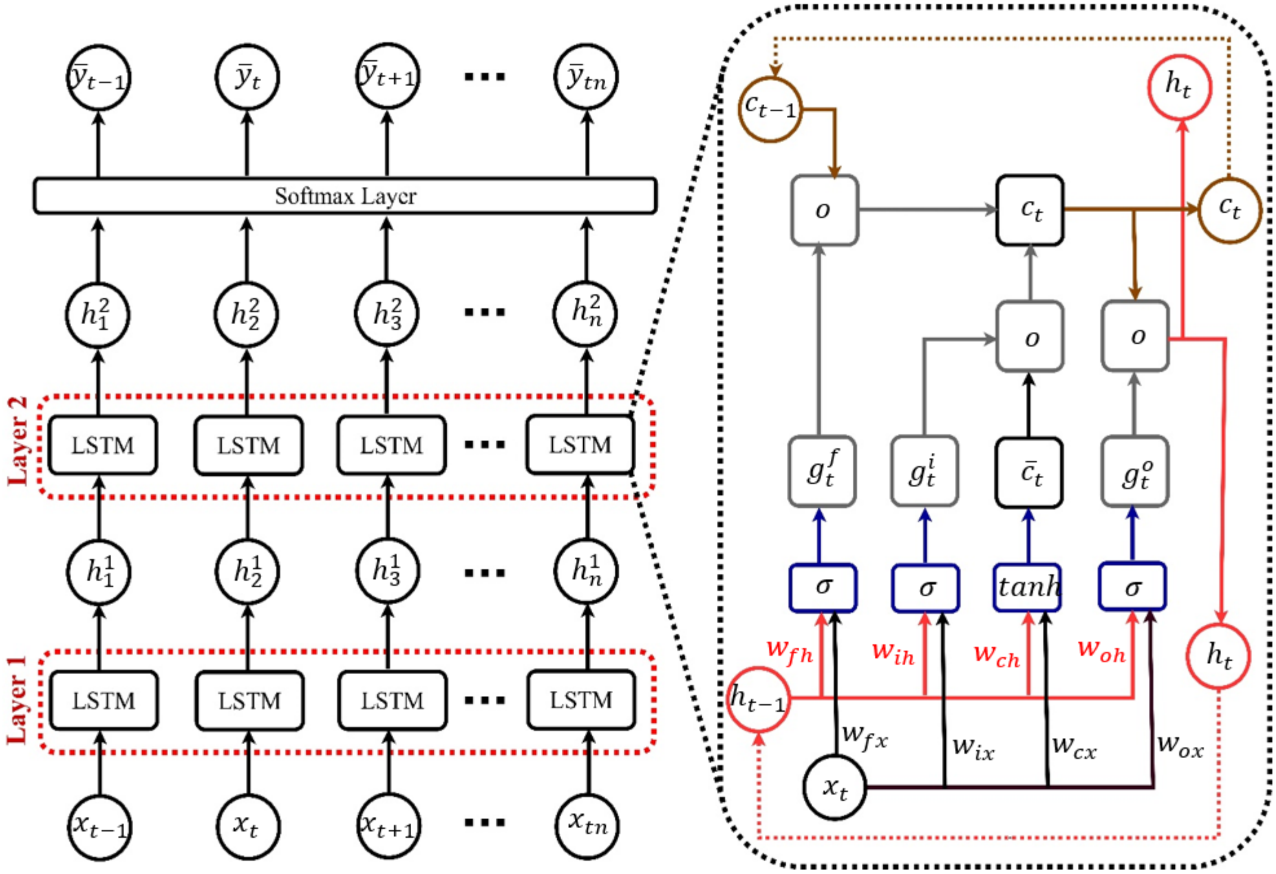


Figure 5.7. The LSTM-RNN structure for classification.

In each cell state c_t, the gates receive the current feature input x_t, the hidden state h_{t-1} at previous moment and the state information c_{t-1} of the internal memory of the cell c_t respectively. The final layer is a SoftMax output layer that classifies 10 emotional states. During the calculation process of each LSTM unit, the value of the candidate memory cell $\bar{c}_t$ is initially determined using the following equation.

$$\bar{c}_t = \gamma\left(w_{cx}x_t + w_{ch}h_{t-1} + b_c\right) \tag{5.13}$$

Herein, γ denotes the hyperbolic tangent activation function, b_c represents the bias term of the respective cell, x_t is the input to the current cell, and h_{t-1} is the output of the previous cell, w_{cx} and w_{ch} unknown parameters are utilized for the input with hidden state of related cell respectively.

An input gate g_t^i is then introduced to regulate the amount of information from $\bar{c}_t$ those flows into the new memory cell c_t as follows:

$$g_t^i = \sigma\left(w_{ix}x_t + w_{ih}h_{t-1} + b_i\right) \tag{5.14}$$

where σ represents the sigmoid activation function, b_i denotes the bias term of the related input gate, and w_{ix},w_{ih} are unknown parameters used for the input and hidden state of the input gate, respectively. Additionally, a forget gate g_t^f s introduced to regulate the amount of information in the previous memory cell c_{t-1} that should be discarded or retained.

$$g_t^f = \sigma\left(w_{fx}x_t + w_{fh}h_{t-1} + b_f\right) \tag{5.15}$$

where b_f represents the bias term of the forget gate, and w_{fx} and w_{fh} are unknown parameters associated with the input and hidden state of the forget gate, respectively. Information from the previous hidden state h_{t-1} and current input x_t is processed through the sigmoid activation function. Subsequently, the old cell state c_{t-1} is updated to the new cell state c_t as follows:

$$c_t = g_t^i\bar{c}_t + g_t^f c_{t-1} \tag{5.16}$$

Finally, the output gate g_t^o is introduced to decide which portion of the memory cellc_t should be passed to the final hidden state h_t as follows:

$$g_t^o = \sigma\left(w_{ox}x_t + w_{oh}h_{t-1} + b_o\right) \tag{5.17}$$

$$h_t = g_t^o\gamma\left(c_t\right) \tag{5.18}$$

where b_o denotes the bias term of the output gate, while w_{ox} and w_{oh} are unknown parameters associated with the input and hidden state of the output gate. The outputs from the LSTM units are then fed into a softmax layer ϕ to compute the predicted emotional statement $\bar{y}$ across the class labels as follows:

$$\bar{y}_t = \phi\left(w_{hy}h_t + b_y\right) \tag{5.19}$$

Thus, the error between the real outputs y and estimated outputs $\bar{y}$ is calculated as follows:

$$\varepsilon = y - \bar{y} \tag{5.20}$$

To minimize the error difference and accurately classify each emotion statement from the detected speech features, the relationship between changes in unknown parameters and their error gradients must be defined relative to the derivatives of each LSTM parameter for the previous time step $t-1$. Initially, the derivative of the error function with respect to the current memory cell c_t is calculated as follows:

$$\delta c_t = \frac{\partial \varepsilon}{\partial c_t} = \frac{\partial \varepsilon}{\partial c_t} + \frac{\partial \varepsilon}{\partial h_t}\frac{\partial h_t}{\partial c_t} = \delta c_t + \delta h_t g_t^o \left[1-\gamma^2(c_t)\right] \tag{5.21}$$

Then, for simplicity, the linear transformations of each updated memory cell $\overline{c}_t$ and all gates are defined respectively as follows:

$$\mathrm{Z}(\overline{c}_t) = w_c \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_c \tag{5.22}$$

$$\text{with } w_c = \left[w_{ch} \,|\, w_{cx}\right] \tag{5.23}$$

$$\mathrm{Z}(g_t^i) = w_i \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_i \tag{5.24}$$

$$\text{with } w_i = \left[w_{ih} \,|\, w_{ix}\right] \tag{5.25}$$

$$\mathrm{Z}(g_t^f) = w_f \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_f \tag{5.26}$$

$$\text{with } w_f = \left[w_{fh} \,|\, w_{fx}\right] \tag{5.27}$$

$$\mathrm{Z}(g_t^o) = w_o \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_o \tag{5.28}$$

$$\text{with } w_o = \left[w_{oh} \,|\, w_{ox}\right] \tag{5.29}$$

Subsequently, each transformed parameter is individually derived as follows:

$$\delta\mathrm{Z}(\overline{c}_t) = \frac{\partial \varepsilon}{\partial \mathrm{Z}(\overline{c}_t)} = \frac{\partial \varepsilon}{\partial c_t}\frac{\partial c_t}{\partial \overline{c}_t}\frac{\partial \overline{c}_t}{\partial \mathrm{Z}(\overline{c}_t)} = \delta c_t g_t^i \left[1-\overline{c}_t^2\right] = \left(\delta c_t + \delta h_t g_t^o \left[1-\gamma^2(c_t)\right]\right) g_t^i \left(1-\overline{c}_t^2\right) \tag{5.30}$$

$$\delta\mathrm{Z}(g_t^i) = \frac{\partial \varepsilon}{\delta \mathrm{Z}(g_t^i)} = \frac{\partial \varepsilon}{\partial c_t}\frac{\partial c_t}{\partial g_t^i}\frac{\partial g_t^i}{\partial \mathrm{Z}(g_t^i)} = \delta c_t \overline{c}_t g_t^i \left(1-g_t^i\right) = \left(\delta c_t + \delta h_t g_t^o \left[1-\gamma^2(c_t)\right]\right) \overline{c}_t g_t^i \left(1-g_t^i\right) \tag{5.31}$$

$$\delta\mathrm{Z}(g_t^f) = \frac{\partial \varepsilon}{\partial \mathrm{Z}(g_t^f)} = \frac{\partial \varepsilon}{\partial c_t}\frac{\partial c_t}{\partial g_t^f}\frac{\partial g_t^f}{\partial \mathrm{Z}(g_t^f)} = \delta c_t c_{t-1} g_t^f \left(1-g_t^f\right) = \left(\delta c_t + \delta h_t g_t^o \left[1-\gamma^2(c_t)\right]\right) c_{t-1} g_t^f \left(1-g_t^f\right) \tag{5.32}$$

$$\delta\mathrm{Z}(g_t^o) = \frac{\partial \varepsilon}{\partial \mathrm{Z}(g_t^o)} = \frac{\partial \varepsilon}{\partial h_t}\frac{\partial h_t}{\partial g_t^o}\frac{\partial g_t^o}{\partial \mathrm{Z}(g_t^o)} = \delta h_t \gamma(c_t) g_t^o \left(1-g_t^o\right) \tag{5.33}$$

Finally, the gradient update is performed with respect to each derivative of the relevant unknown parameters and bias terms at time step t as follows:

$$\delta w_c^t = \delta Z(\bar{c}_t)\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix}^T, \ \delta b_c = \delta Z(\bar{c}_t) \tag{5.34}$$

$$\delta w_i^t = \delta Z\left(g_t^i\right)\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix}^T, \ \delta b_i = \delta Z\left(g_t^i\right) \tag{5.35}$$

$$\delta w_f^t = \delta Z\left(g_t^f\right)\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix}^T, \ \delta b_f = \delta Z\left(g_t^f\right) \tag{5.36}$$

$$\delta w_o^t = \delta Z\left(g_t^o\right)\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix}^T, \ \delta b_o = \delta Z\left(g_t^o\right) \tag{5.37}$$

The gradients for δc_{t-1} and δh_{t-1} for the previous time step $t-1$ are also computed as follows:

$$\delta c_{t-1} = \frac{\partial \varepsilon}{\partial c_{t-1}} = \frac{\partial \varepsilon}{\partial c_t}\frac{\partial c_t}{\partial c_{t-1}} = \delta c_t g_t^f = \left(\delta c_t + \delta h_t g_t^o\left[1-\gamma^2\left(c_t\right)\right]\right)g_t^f \tag{5.38}$$

$$\begin{aligned}\delta h_{t-1} &= \frac{\partial \varepsilon}{\partial h_{t-1}} = \frac{\partial \varepsilon}{\partial Z(\bar{c}_t)}\frac{\partial Z(\bar{c}_t)}{\partial h_{t-1}} + \frac{\partial \varepsilon}{\partial Z\left(g_t^i\right)}\frac{\partial Z\left(g_t^i\right)}{\partial h_{t-1}} + \frac{\partial \varepsilon}{\partial Z\left(g_t^f\right)}\frac{\partial Z\left(g_t^f\right)}{\partial h_{t-1}} + \frac{\partial \varepsilon}{\partial Z\left(g_t^o\right)}\frac{\partial Z\left(g_t^o\right)}{\partial h_{t-1}} \\ &= w_{ch}^t \delta Z(\bar{c}_t) + w_{ih}^t \delta Z\left(g_t^i\right) + w_{fh}^t \delta Z\left(g_t^f\right) + w_{oh}^t \delta Z\left(g_t^o\right)\end{aligned} \tag{5.39}$$

After minimizing the error difference and accurately classifying each emotion based on the detected speech features, the following section presents a comprehensive analysis of the implementation of the proposed algorithm on the NAO HR.

5.3. Implementation of the Proposed Algorithm to NAO HR

This section provides an overview of implementing the proposed algorithm on the NAO HR. Herein, it manages the speech-based emotion analysis algorithm through four stages. In the first stage, NAO HR uses its four directional microphones and speakers to initiate a dialogue with the human, aiming to enhance mental health as a social-interactive companion. After completing the "Initialization" stage, the system transitions to the "Speech Detection" stage, where NAO HR activates the "Speech Tracking" (ST) module.

Figure 5.8 depicts the following abbreviations and their meanings: "ST" denotes "Speech Tracking"; "S1" represents the statement “Hello, my name is Mehmet. I am here to accompany you and protect your physical and mental health!”; "S2" indicates “I can hear and analyse your voice”; "S3" means “Sorry, I cannot hear your voice. Can you speak again, please?”; "S4" refers to “Sorry, I cannot detect your emotional statement. I am attempting this again”; "S5" stands for “Sorry, I lost the interpreted emotional statement. I will now perform the emotion identification process again"; "S6" represents "The engagement did not execute as the relevant condition was not identified. I will try to analyse it again"; "E1" through "E3" denote "Engagement 1," "Engagement 2," and "Engagement 3," respectively; while "C1," "C2," and "C3" signify "Condition 1," "Condition 2," and "Condition 3," respectively.

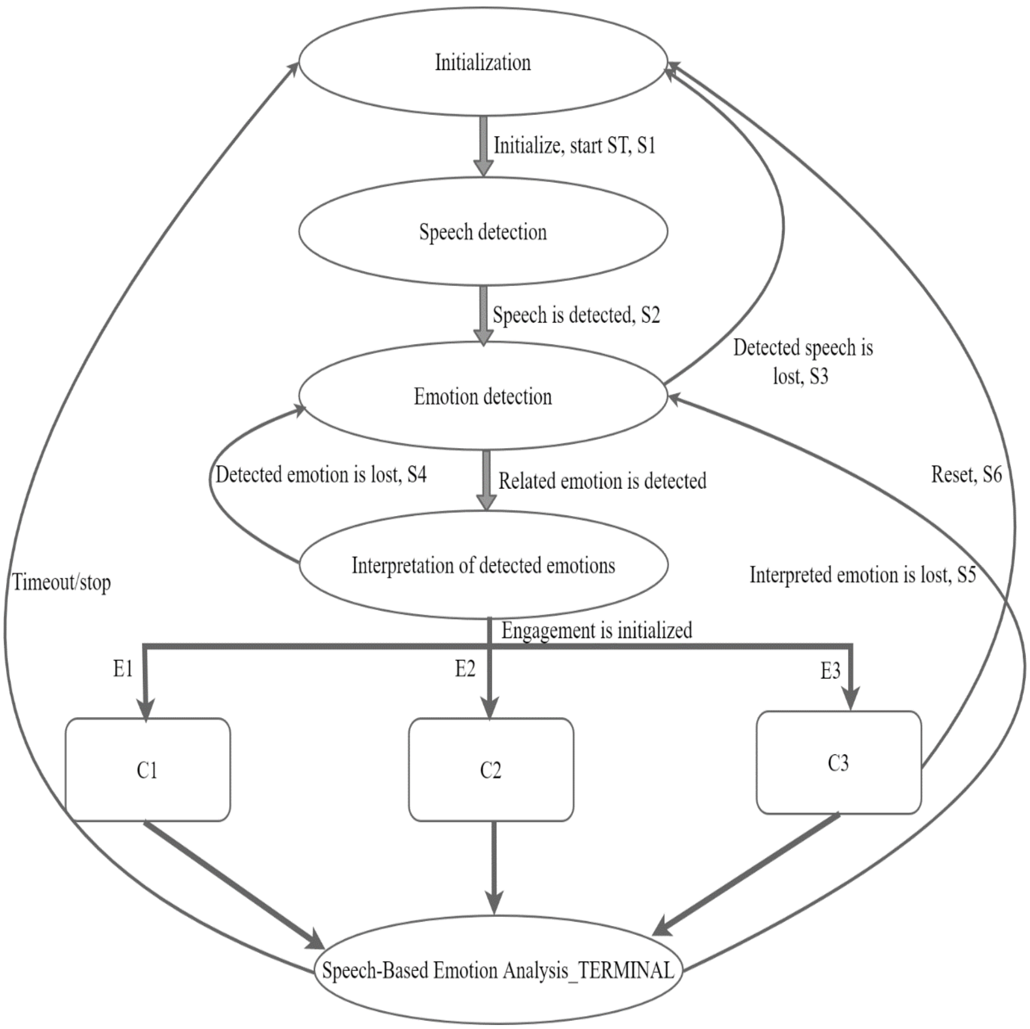


Figure 5.8. Dialogue model of the proposed speech-based emotion analysis algorithm.

Once speech is detected, the "ST" module allows the NAO HR to track the human voice. Initially, NAO HR introduces itself (S1) and inquiries about the person's feelings and emotional state (S2). It then proceeds to the "Emotion Detection" phase to identify the individual's current emotional state. If the voice is no longer detected, the dialogue manager reverts to the "Speech Detection" phase with the S3 statement. Upon successful detection, the dialogue manager moves to "Interpretation of Detected Emotions" to analyze the identified emotions. If emotions are not detected, the dialogue manager reverts to the "Emotion Detection" phase with the S4 statement. Subsequently, NAO HR interprets emotions at stages "C1," "C2," and "C3" for "sad," "angry," and "afraid/anxious" respectively, focusing on these emotions to evaluate HR's engagement performance.

During each interpretation stage, a decision tree is utilized to facilitate social dialogue with the individual. Upon detecting a 'sad' emotion, it transitions to the "C1" stage. As depicted in Figure 5.9, the dialogue includes various responses: "A1" offers reassurance and support; "A2" proposes a traditional Turkish folk dance; "A3" expresses satisfaction at improving the person's mood; "A4" conveys empathy; "A5" shares a lighthearted comment; "A6" provides a neutral response; "A7" offers to leave the person alone if they prefer; "A8" emphasizes the importance of mental health; "A9" thanks the person for any positive change; "Q1" through "Q6" are queries designed to further engage the individual and assess their mood. Each statement is assigned an identifier for accurate emotion interpretation and response.

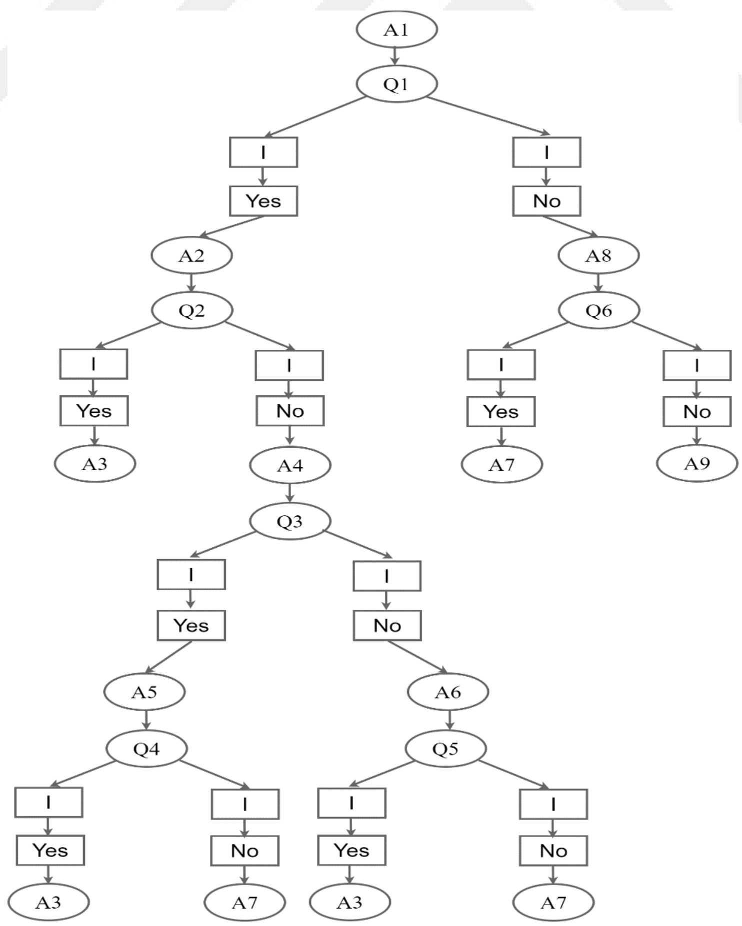


Figure 5.9. Decision tree structure for “C1” stage.

In this process, NAO HR initially transitions to the "A1" state, delivering motivational statements to uplift the individual. It then moves to the "Q1" state, seeking permission to entertain the person and awaits their response. While waiting, NAO HR continues to analyze the individual's speech by activating the "I" state. During the decision phase,

---

If the received answer is "No"

NAO HR immediately enters the "A8" and “Q6” states

If the received answer is "No"

NAO HR immediately enters the "A3"

If the received answer is "Yes"

NAO HR immediately enters the "A7"

If the received answer is "Yes"

NAO HR immediately enters the "A2" state and takes some actions to perform traditional Turkish folk dance with integrated music. After completing this state, it passes to the “Q2” statement

If the received answer is "Yes"

NAO HR immediately enters the "A3"

If the received answer is "No"

NAO HR immediately enters the "A4" and “Q3” statements

If the received answer is “Yes”

NAO HR immediately enters the "A5" and “Q4” statements

If the received answer is “Yes”

NAO HR immediately enters the "A3" statement

If the received answer is “No”

NAO HR immediately enters the "A7" statement

If the received answer is “No”

NAO HR immediately enters the "A6" and “Q5” statements

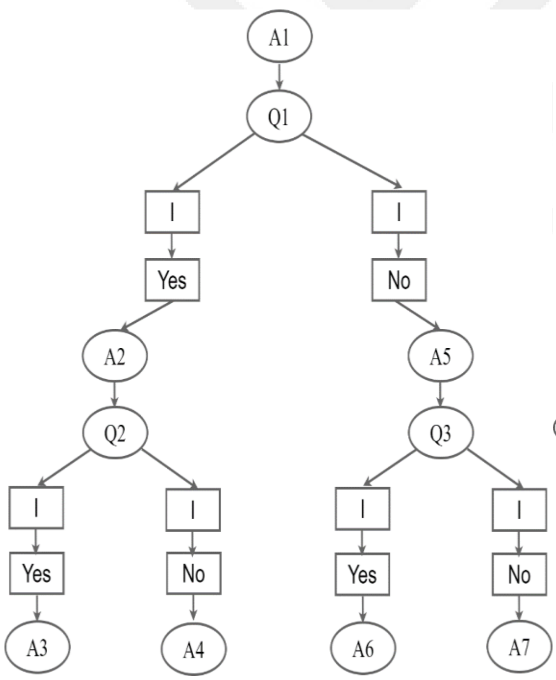
If the received answer is “Yes”

NAO HR immediately enters the "A7" statement

If the received answer is “No”

NAO HR immediately enters the "A9"and then “A2” statements

In the event that the NAO HR detects an 'angry' emotion, it transitions to the "C2" stage as illustrated in Figure 11. In this stage: “A1” advises the individual to calm down for health reasons; “A2” guides them through a calming breathing exercise; “A3” expresses satisfaction if the individual feels better; “A4” apologizes if the calming attempt was unsuccessful; “A5” offers further assistance to improve their mood; “A6” provides the option to be left alone with an invitation to re-engage; “A7” thanks them for any positive change. Additionally, “Q1” through “Q3” are queries designed to assess the individual's need for help and their current state.



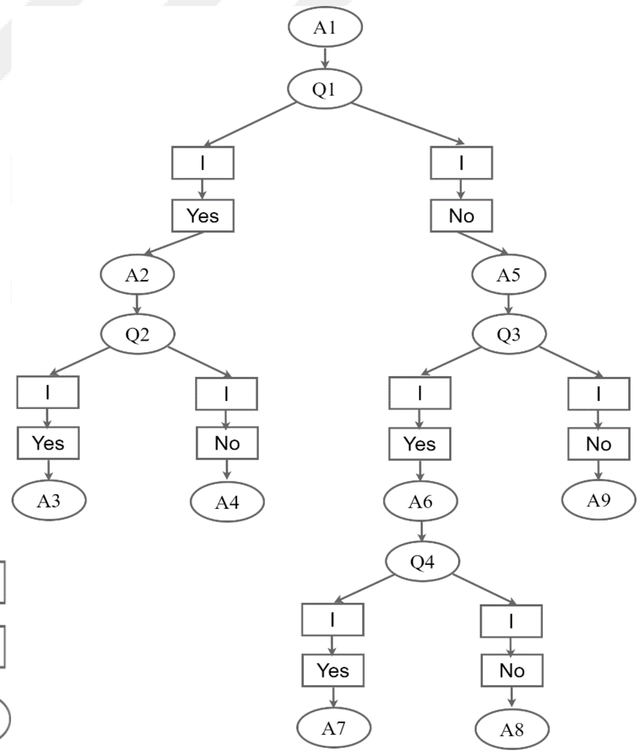


Figure 5.10. Decision tree structure for “C2” and “C3” stages.

In this scenario, NAO HR initially enters the "A1" state, aiming to calm the angry individual with motivational statements. It then transitions to the "Q1" state, seeking permission to assist and awaiting a response. During the decision phase,

If the received answer is "No"

NAO HR immediately enters the "A5" and “Q3” states

If the received answer is "No"

NAO HR immediately enters the "A7" statement and then passes to “A2” statement

If the received answer is "Yes"

NAO HR immediately enters the "A6"

If the received answer is "Yes"

NAO HR immediately enters the "A2" state and shows some exercise actions with its arms, then passes to the “Q2” statement

If the received answer is "Yes"

NAO HR immediately enters the "A3"

If the received answer is "No"

NAO HR immediately enters the "A4" statement

As the NAO HR detects an emotion of 'fear and anxiety', it transitions to the "C3" stage. In this stage, "A1" offers reassurance and emphasizes that there is nothing to fear, highlighting that the individual is under medical care and supported. "A2" begins to narrate a comforting story to distract and soothe the individual. "A3" celebrates any improvement in mood. "A4" expresses sympathy if the efforts were unsuccessful. "A5" emphasizes the importance of interaction for effective support. "A6" acknowledges understanding of the situation. "A7" indicates readiness to proceed with further actions if needed. "A8" offers to leave the individual alone but remains available if they change their mind. "A9" thanks the individual for any positive change. "Q1" through "Q4" are questions designed to assess the individual's willingness to receive help, evaluate their response to the story, and offer further assistance if needed.

In this scenario, NAO HR first enters the "A1" state to calm the anxious individual with motivational statements. It then moves to the "Q1" state, seeking permission to assist and awaiting the response. In the decision phase,

---

If the received answer is "Yes"

NAO HR immediately enters the "A2" and “Q2” states

If the received answer is "No"

NAO HR immediately enters the "A4" state

If the received answer is "Yes"

NAO HR immediately enters the "A3" state

If the received answer is "No"

NAO HR immediately enters the "A5" and “Q3” states

If the received answer is "No"

NAO HR immediately enters the "A9" statement then passes to “A2” statement

If the received answer is "Yes"

NAO HR immediately enters the "A6" and “Q4” statements

If the received answer is “Yes”

NAO HR immediately enters the "A7" statement

If the received answer is “No”

NAO HR immediately enters the "A8" statement

After explaining the last scenario, the subsequent section finally provides a detailed examination of the simulation and experimental results of the proposed algorithm.

5.4. Simulation and Experimental Results

This sub-section assesses the proposed algorithm through both simulation and experimental environments. It begins with an introduction to the experimental setup, followed by an evaluation of the results.

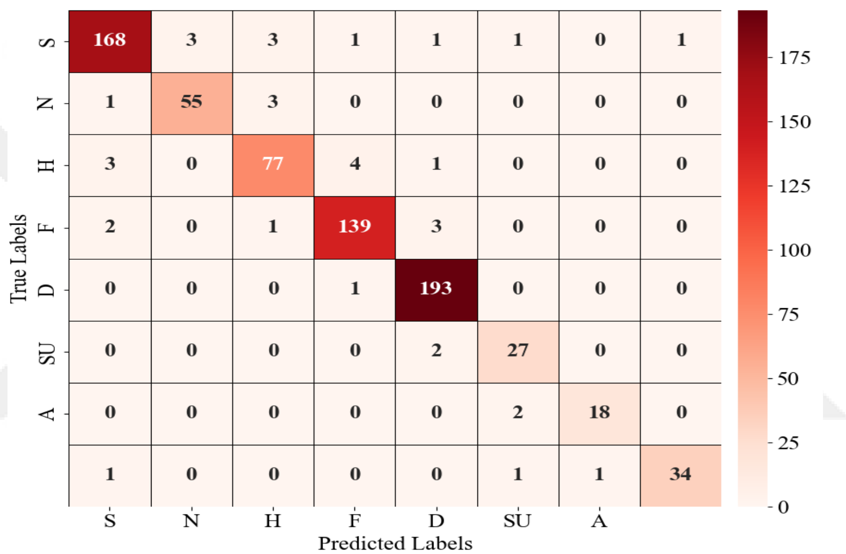
5.4.1. Simulation and Experimental Settings

To assess the reliability and effectiveness of the proposed algorithm, two databases are used: the Berlin Database of Emotional Speech (EMO-DB) and the Surrey Audio-Visual Expressed Emotion (SAVEE-DB). The EMO-DB consists of 800 speech samples from 10 speakers (5 men and 5 women) expressing eight emotions: happiness, sadness, anger, fear, boredom, disgust, surprise, and neutral. The SAVEE-DB includes recordings from four male actors, each expressing seven emotions: happiness, sadness, anger, fear, disgust, surprise, and neutral. Performance metrics such as accuracy (m_a), precision (m_p), recall (m_r), and F1-Score (m_{F1}), along with confusion matrices, are used to quantitatively assess the classification capabilities of the proposed algorithm. Following this evaluation, the algorithm is integrated into NAO HR's interface via "Choregraphe" to demonstrate its applicability and reliability in a natural setting.

5.4.2. Results

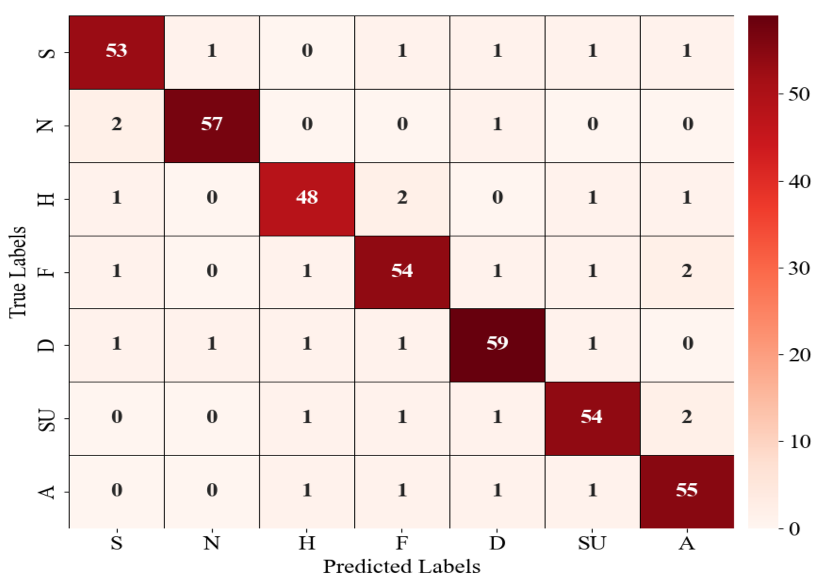
After detailing the simulation and experimental settings, the confusion matrices for each dataset are presented as follows:

Table 5.1. The confusion matrix evaluation of the proposed algorithm on EMO-DB Dataset.



True Labels \ Predicted Labels	S	N	H	F	D	SU	A	
S	168	3	3	1	1	1	0	1
N	1	55	3	0	0	0	0	0
H	3	0	77	4	1	0	0	0
F	2	0	1	139	3	0	0	0
D	0	0	0	1	193	0	0	0
SU	0	0	0	0	2	27	0	0
A	0	0	0	0	0	2	18	0
	1	0	0	0	0	1	1	34

Table 5.2. The confusion matrix evaluation of the proposed algorithm on SAVEE Dataset.



True Labels \ Predicted Labels	S	N	H	F	D	SU	A
S	53	1	0	1	1	1	1
N	2	57	0	0	1	0	0
H	1	0	48	2	0	1	1
F	1	0	1	54	1	1	2
D	1	1	1	1	59	1	0
SU	0	0	1	1	1	54	2
A	0	0	1	1	1	1	55

In the confusion matrices, the labels A, F, H, S, N, B, D, and SU denote anger, fear, happiness, sadness, neutral, boredom, disgust, and surprise, respectively. The performance of the proposed algorithm is then analyzed using these confusion matrices, with m_a, m_p, m_r, and m_{F1} metrics detailed in Table 5.3 and Table 5.4.

Table 5.3. Classification results of the proposed algorithm on EMO-DB.

Emotions \ Metrics	m_a	m_p	m_t	m_{f_1}
A	0.95	0.96	0.94	0.94
F	0.93	0.94	0.93	0.93
H	0.91	0.91	0.90	0.90
S	0.96	0.95	0.95	0.95
N	0.99	0.96	0.99	0.97
B	0.93	0.90	0.93	0.91

The results indicate that the average m_a, m_p, m_r, and m_{F1} metrics are 0.93, 0.94, 0.93, and 0.93, respectively.

Table 5.4. Classification results of the proposed algorithm on SAVEE-DB.

Emotions \ Metrics	m_a	m_p	m_r	m_{F1}
S	0.92	0.91	0.91	0.91
N	0.95	0.95	0.95	0.95
H	0.91	0.92	0.90	0.90
F	0.90	0.90	0.90	0.90
D	0.92	0.92	0.92	0.92
SU	0.90	0.91	0.90	0.90
A	0.93	0.90	0.93	0.91

Although Table 5.4 shows a slight decrease in these rates, the classification performance remains effective. Overall, the proposed algorithm demonstrates satisfactory performance, closely aligning with actual emotional statements. Additionally, the speech-based emotion analysis algorithm has been integrated into NAO HR's interface via Choregraphe to facilitate social-interactive dialogue with a quarantined patient, as illustrated in the following figure.

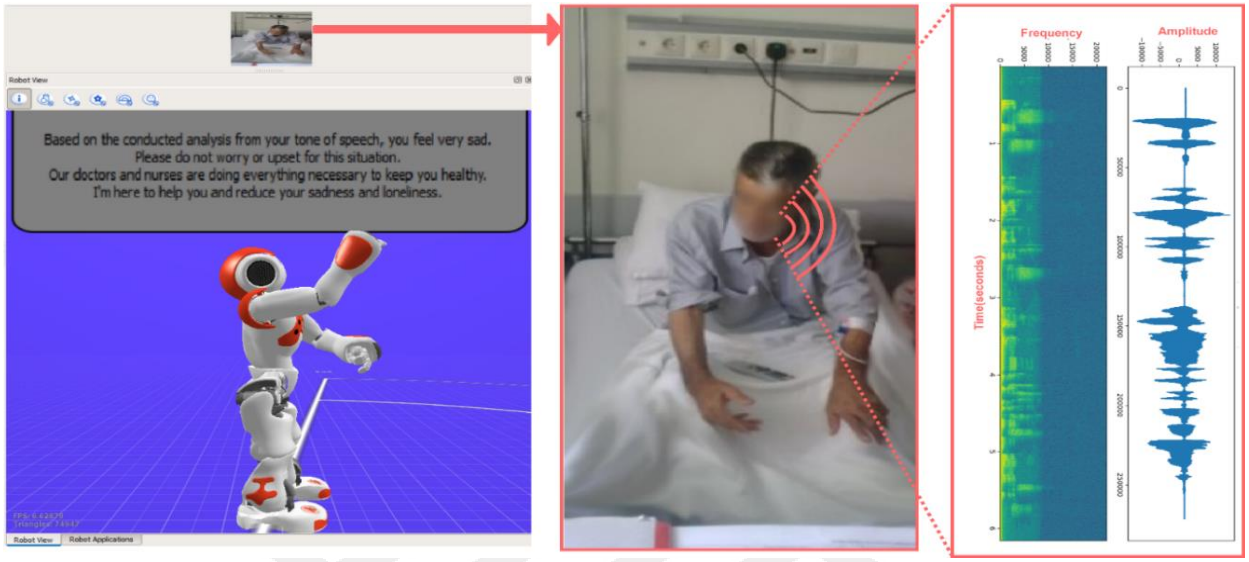


Figure 5.11. Speech-based emotion analysis application 1 using Choregraphe.

As depicted in Figure 5.11, during the initial stage of the social dialogue, NAO HR detects the subject's emotional state from their voice and aims to improve their mood.

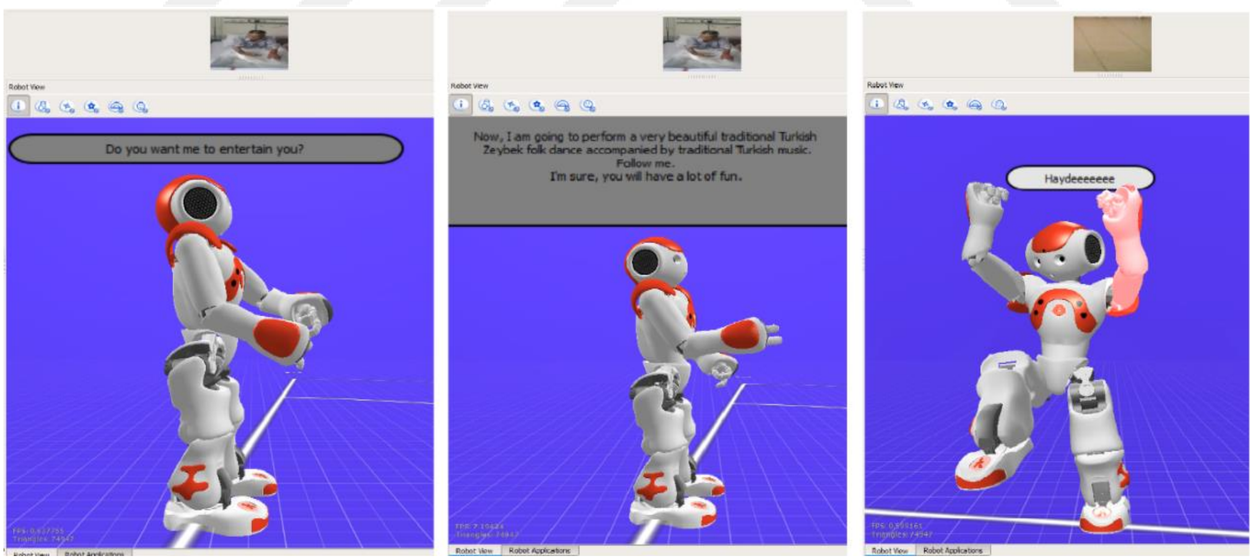


Figure 5.12. Speech-based emotion analysis application 2 using Choregraphe.

In this instance, NAO HR accurately detects the patient's emotion as "sad" with a 91.5% accuracy rate. Following this, NAO HR requests permission to improve the patient's mood using motivational phrases. Once approval is granted, NAO HR performs the traditional Turkish folk dance 'Zeybek,'. The robot's joints were meticulously designed and modeled for this performance.

After completing the dance, NAO HR inquiries about the patient's feelings, as shown in the following figure.

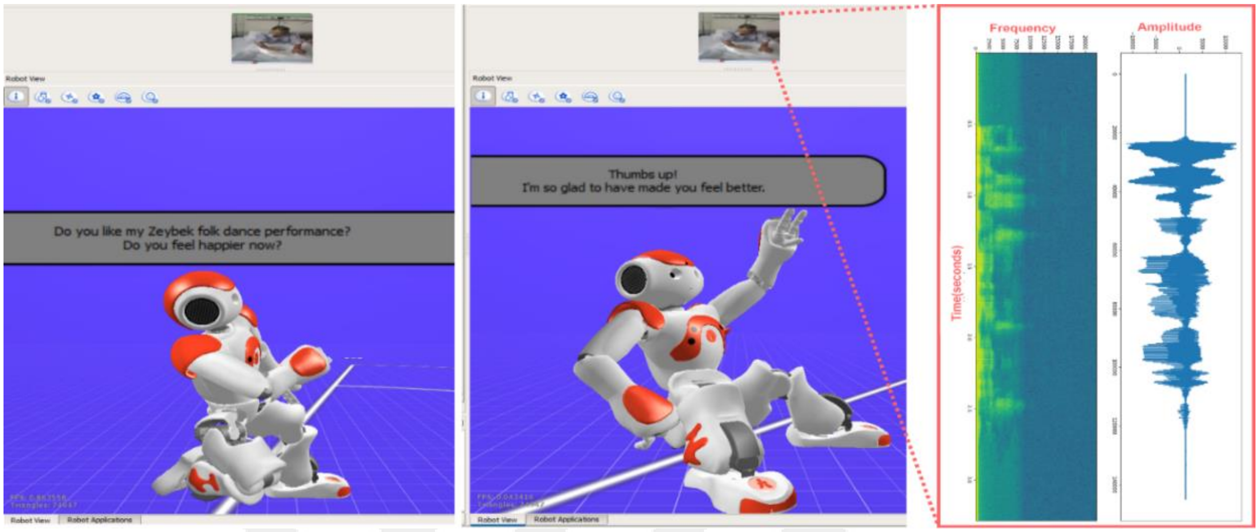


Figure 5.13. Speech-based emotion analysis application 3 using Choregraphe.

As shown in Figure 5.13, NAO HR re-evaluates the patient's speech tone and compares it with the previously detected emotional state.

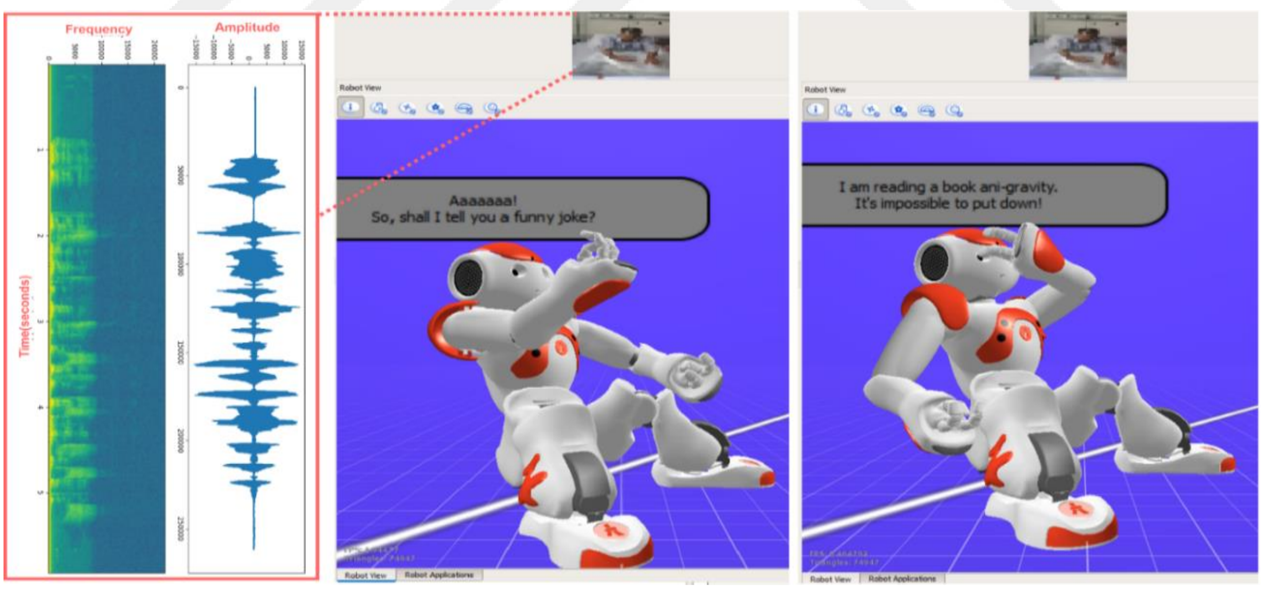


Figure 5.14. Speech-based emotion analysis application 4 using Choregraphe.

In response to a positive reaction from the patient, NAO HR expresses its happiness through various gestures and facial expressions. If the response is negative, NAO HR promptly offers an alternative, such as telling a joke, based on the patient's permission, as illustrated in the following figure. Upon completing this performance, if the patient provides positive feedback, NAO HR again displays its satisfaction with the progress, as shown in the subsequent figure.

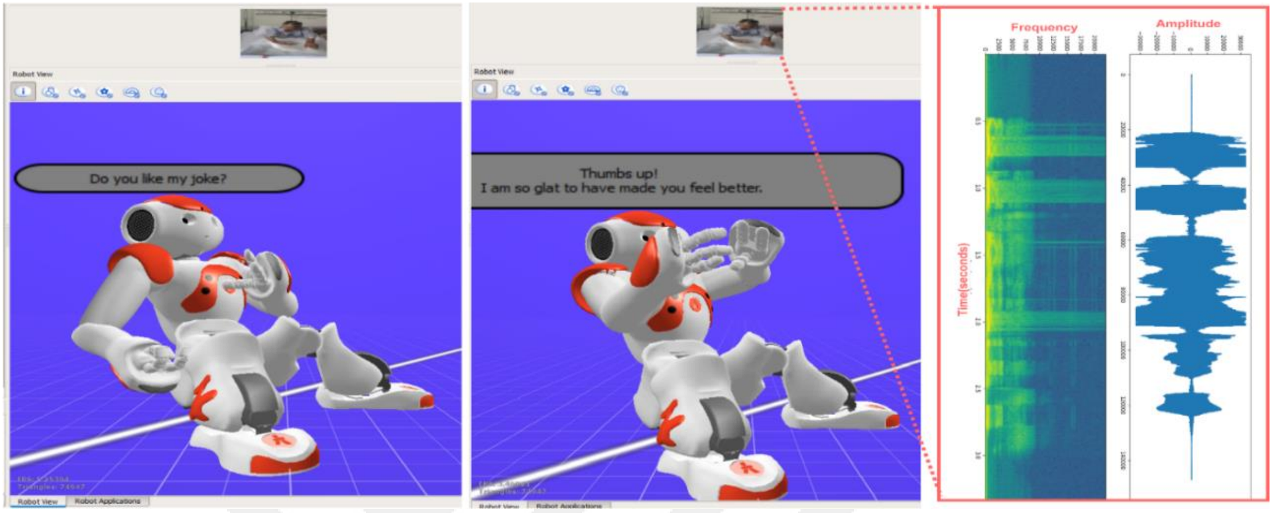


Figure 5.15. Speech-based emotion analysis application 5 using Choregraphe.

If NAO HR continues to receive a negative response from the patient despite the previous performances, it proposes a third alternative by playing a folk song, provided it has the patient's approval, as illustrated in the following figure.

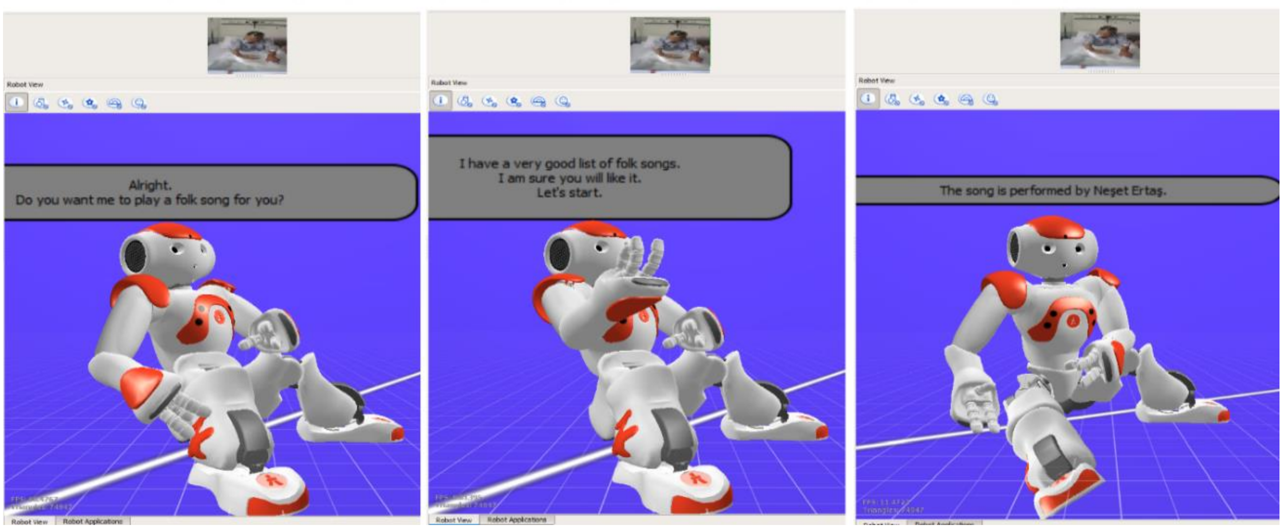


Figure 5.16. Speech-based emotion analysis application 6 using Choregraphe.

As illustrated in Figure 5.16, NAO HR provides the patient with background information about the song and the singer. The playlist has been curated based on the patient's age, with a focus on the "folk" song category for enhanced engagement. After delivering the folk song performance, NAO HR inquiries about the patient's current emotional state, as shown in Figure 5.17.

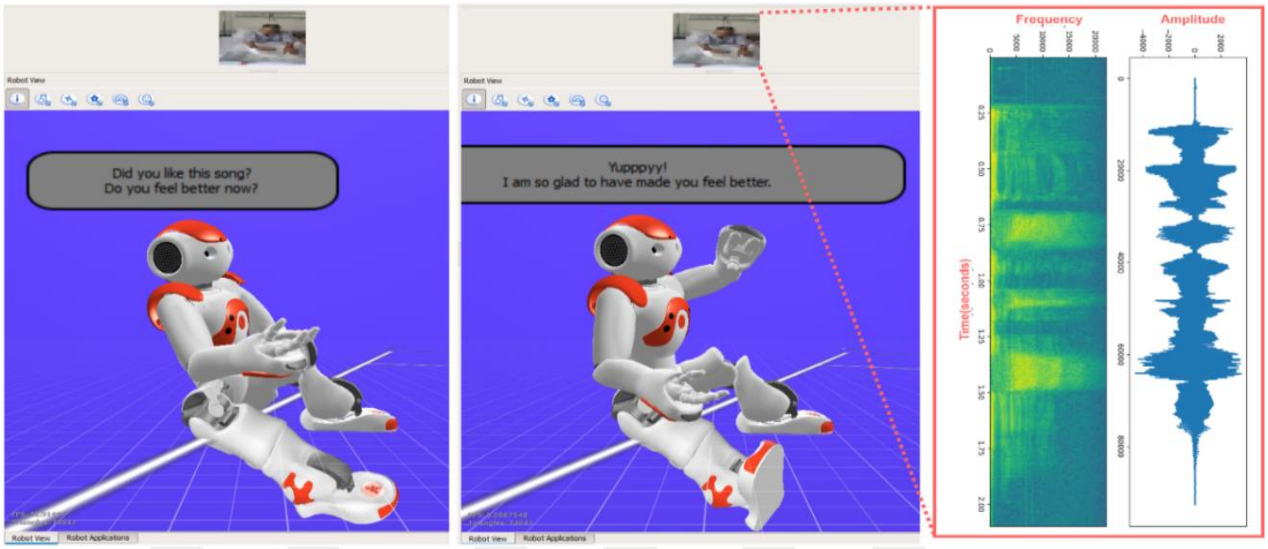


Figure 5.17. Speech-based emotion analysis application 7 using Choregraphe.

If the NAO HR receives positive feedback from the patient, it promptly expresses its joy at the successful outcome of the interaction. Following a comprehensive evaluation of the simulation and experimental results, the concluding section provides an in-depth analysis of the proposed algorithm's effectiveness.

5.5. Discussions

In this section, an effective speech-based emotion analysis algorithm is presented and implemented to a HR in real time. The developed algorithm begins by extracting key speech features using the MFCC technique. It then classifies these features into emotional statements using an LSTM-RNN structure. Finally, it facilitates interaction with the user through decision tree-based dialogues. Developing the speech-based emotion analysis algorithm significantly enhances HRI by enabling HRs to accurately detect and respond to emotional states, thereby improving the quality of interactions and providing personalized assistance. This capability is particularly valuable in healthcare and therapeutic settings, where robots can offer tailored emotional support and comfort. By addressing and altering negative stereotypes of robots, the algorithm showcases them as empathetic and responsive entities, contributing to increased safety and well-being. The algorithm also contributes to increased efficacy across various applications, reduces the need for human intervention in emotional support tasks, and promotes advancements in robotics technology, ultimately leading to more effective and cost-efficient solutions.

6. DESIGN AND DEVELOPMENT OF A FACIAL MASK-WEARING CONTROL AND REGULATION ALGORITHM

This section introduces a four-stage algorithm for detecting the absence of or improper wearing of facial masks using HRs. Initially, images are captured by the HR's camera. The Viola–Jones algorithm then detects faces and masks. Subsequently, geometric-based distance measurements assess the spatial relationship between the face and mask models. These measurements are then used to label and evaluate mask-wearing conditions. In the final stage, a three-layer neural network classifies the results. The algorithm has been successfully implemented on the NAO HR, and its performance has been thoroughly analyzed.

6.1. Face and Facial Mask Detection with The Viola-Jones Algorithm

In the first stage of the algorithm, face and facial mask detection is performed using the Viola-Jones algorithm, as depicted in Figure 6.1.

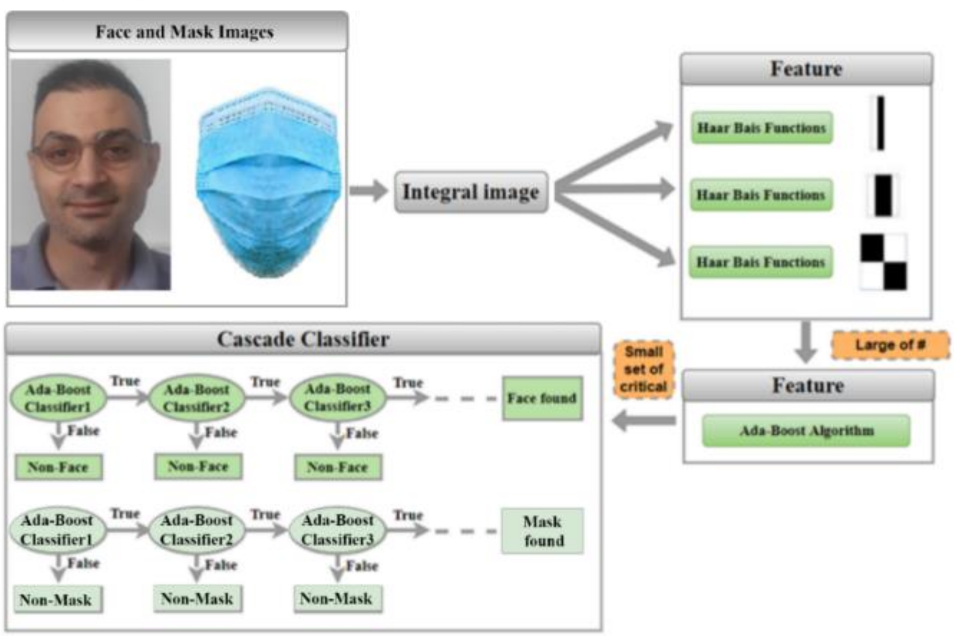


Figure 6.1. Illustration of the Viola-Jones algorithm.

For a more detailed explanation of the algorithm, please refer to the third and fourth sections.

6.2. Geometric-Based Distance Measurements

Following the face and facial mask detection stage, the detected face and mask models are mapped to a coordinate system. The centres of the face and its key features, including the eyes, nose, and mouth, are defined using bounding boxes, as illustrated in Figure 6.2.

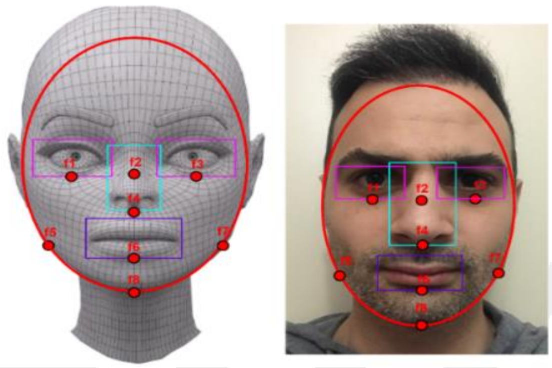


Figure 6.2. Placement of facial feature points in the geometric model.

Once the centers of each facial feature are identified, eight facial feature points are placed in the corresponding regions of the face. Similarly, the center of the detected facial mask (marked with a blue point) and its boundary are defined using a blue rectangle. Six feature points are then assigned to the detected areas of the facial mask model, as shown in Figure 6.3.

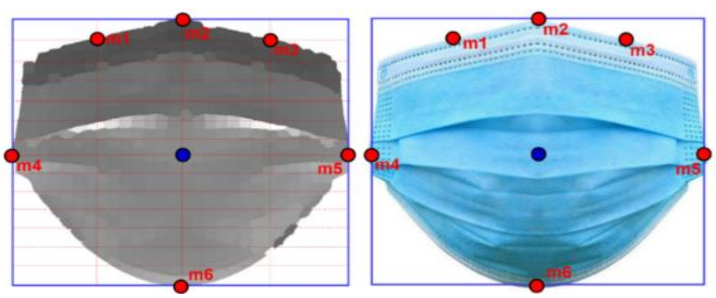


Figure 6.3. Placement of the facial mask model's feature points in the geometric model.

The process involves identifying key feature points on both the detected face and the facial mask model, which are then matched to assess various mask-wearing scenarios. This method ensures that the mask aligns with specific facial regions for each scenario. The study defines five matching types: (1) correct mask coverage (nose, mouth, and chin), (2) mask covering mouth and chin only (nose exposed), (3) mask covering only chin (nose and mouth exposed), (4) no mask covering chin, mouth, and nose (face fully exposed), and (5) no mask present at all. For each scenario, an objective is formulated to determine the transformation $\{T,t\}$ and matching matrix $\{m_{ij}\}$ that align the mask's feature points $\{M_i\}$ of the determined facial mask model M with the face's feature points $\{F_j\}$of detected face F as:

$$\varepsilon(m,t,T)=\sum_{i=1}^{I}\sum_{j=1}^{J}\left(m_{ij}\left\|M_i-t-TF_j\right\|^2+r(T)-\alpha\sum_{i=1}^{I}\sum_{j=1}^{J}m_{ij}\right) \tag{6.1}$$

In this context, α serves as a threshold parameter that biases the objective function towards the relevant matching process and measures the deviation from the associated points. Equation (6.1) incorporates the following constraints:

1) $$\forall_i \sum_{j=1}^{J} m_{ij} \leq 1 \tag{6.2}$$

2) $$\forall_j \sum_{i=1}^{I} m_{ij} \leq 1 \tag{6.3}$$

$\forall_{ij} m_{ij} \in \{0,1\}$ and

3) $$m_{ij}=\begin{cases}1 & \text{If feature point of } M_i \text{ matches to feature point of } F_j\\ 0 & \text{Otherwise}\end{cases} \tag{6.4}$$

The constraints imposed on matrix $\{m_{ij}\}$ ensure that each feature point within the facial mask model corresponds uniquely to a feature point detected in the face image.

4) $$r(T)=\gamma\left(s^2+t_x^2+t_y^2\right) \tag{6.5}$$

Herein, $r(T)$regularizes the transformation by penalizing large values of the scaling factor s, as well as the translation components t_x and t_y along the x and y directions. The term γ functions as the regularization parameter for $r(T)$. Subsequently, T is decomposed into s, θ, t_x and t_y as follows:

$$T=S(s)R(\theta)C_1(t_x)C_2(t_y) \tag{6.6}$$

and is structured as:

$$T=\underbrace{\begin{bmatrix}e^s & 0\\ 0 & e^s\end{bmatrix}}_{S(s)}\underbrace{\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}}_{R(\theta)}\underbrace{\begin{bmatrix}e^{t_x} & 0\\ 0 & e^{-t_x}\end{bmatrix}}_{C_1(t_x)}\underbrace{\begin{bmatrix}\cosh(t_y) & \sinh(t_y)\\ \sinh(t_y) & \cosh(t_y)\end{bmatrix}}_{C_2(t_y)} \tag{6.7}$$

Following a thorough explanation of the transformation and matching processes, the final step involves performing the matching for each facial mask scenario to assign labels. For masks covering the nose, mouth, and chin, six feature points from both the face and the mask are matched, as depicted in Figure 6.4.

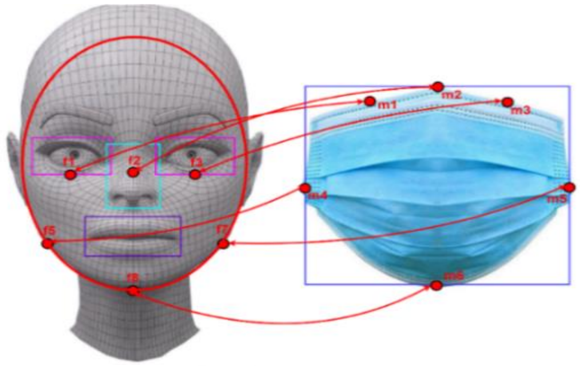


Figure 6.4. Matching the facial mask-covering nose, mouth, and chin regions.

In Figure 6.4, f1, f2, f3, f4, f5 and f5 denote facial feature points on the detected face, while m1, m2, m3, m4, m5 and m6 represent the corresponding feature points on the facial mask model. The details of each matched feature point and the labelled matching results are provided in Table 6.1.

Table 6.1. Matching process for the correct facial mask-covering.

Label	Facial Mask-Wearing Situation	Matched Feature Points
cw	**Correct Wearing Condition:**	f1 and m1
	Nose, mouth and chin regions are covered	f2 and m2
		f3 and m3
		f5 and m4
		f7 and m5
		f8 and m6

where *cw* utilized to label the "Correctly Worn" condition. Subsequently, the facial feature points f4, f5, f7 and f8 are matched with the facial mask feature points m2, m4, m5 and m6 to identify masks covering the mouth and chin regions, as illustrated in Figure 6.5.5.

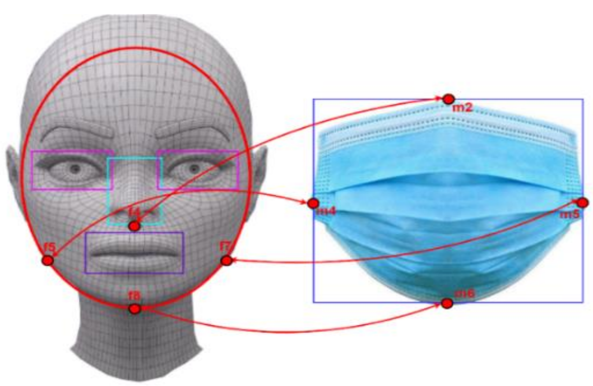


Figure 6.5. Matching process for the facial mask-covering mouth and chin regions.

where f4, f5, f7, and f8 represent the facial feature points on the detected face. Subsequently, the matched feature points and the labelled matching results are detailed in Table 6.2.

Table 6.2. Representation of the matched feature points and labeled facial mask-wearing condition.

Label	Facial Mask-Wearing Situation	Matched Feature Points
$iw1$	**Incorrect Wearing Condition 1:** Mouth and chin regions are covered	f4 and m2 f5 and m4 f7 and m5 f8 and m6

where $iw1$ is used to label the "Incorrect Wearing Condition 1". In the case of a facial mask covering only the chin region, the facial feature points f5, f6, f7 and f8 are matched with the mask feature points m2, m4, m5 and m6 as shown in Figure 6.6.

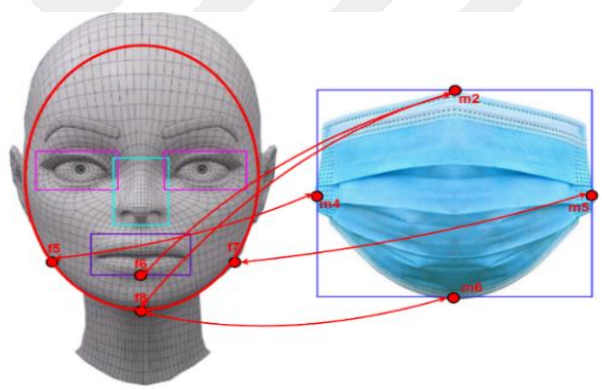


Figure 6.6: Matching process for the facial mask-wearing only chin region.

Subsequently, the matched feature points and the labelled matching results are presented in Table 6.3. Herein, $iw2$ is used to label the "Incorrect Wearing Condition 2".

Table 6.3. Representation of the matched feature points and labeled facial mask-wearing situation.

Label	Facial Mask-Wearing Situation	Matched Feature Points
$iw2$	**Incorrect Wearing Condition 2:** Only chin region is covered	f5and m4 f6 and m2 f7 and m5 f8 and m6

For the scenario where the facial mask does not cover the chin, mouth, or nose regions, the facial feature points f5, f7 and f8 are matched with the mask feature points m2, m4, m5 and m6 as illustrated in left side of Figure 6.7.

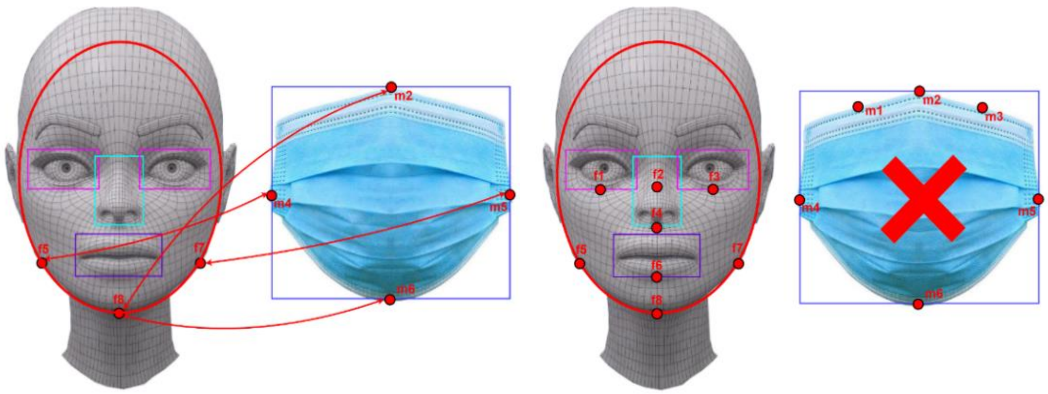


Figure 6.7. Matching process for the facial mask not covering chin, mouth, and nose regions and another matching process for the absence of a facial mask.

To indicate the absence of the facial mask, no facial feature points are matched with any mask feature points, as shown on the right side of Figure 6.7.

Table 6.4. Representation of the matched feature points and labeled facial mask-wearing situation.

Label	Facial Mask-Wearing Situation	Matched Feature Points
$iw3$	**Incorrect Wearing Condition 3:** Nose, mouth and chin regions are not covered	f5and m4 f7 and m5 f8 and m2 f8 and m6

Subsequently, the matched feature points and the labelled matching results are listed in Table 6.4 . Herein, $iw3$ is used to label the "Incorrect Wearing Condition 3".

Following the completion of the relevant measurements, the subsequent section presents the final stage of the proposed algorithm.

6.3. Classification

Once the face and facial mask matching processes are completed and each mask-wearing scenario is labelled using a geometric-based technique, the final stage of the algorithm involves classifying the mask-wearing conditions with a 3-layer NN, as depicted in Figure 6.8. In this stage, the labelled data for each mask-wearing scenario is fed through the input layer x to hidden layer h and then output layer o for prediction. The input data is formatted as a vector, as follows:

$$x = [x_{cw} \quad x_{iw1} \quad x_{iw2} \quad x_{iw3} \quad x_{am}] \tag{6.8}$$

where x_{cw}, x_{iw1}, x_{iw2}, $x_{iw3}$and x_{am}present the input data. Then, the output vector of the hidden layer is formed as follows:

$$h = [h_1 \quad h_2 \quad h_3 \quad h_4 \quad h_5] \tag{6.9}$$

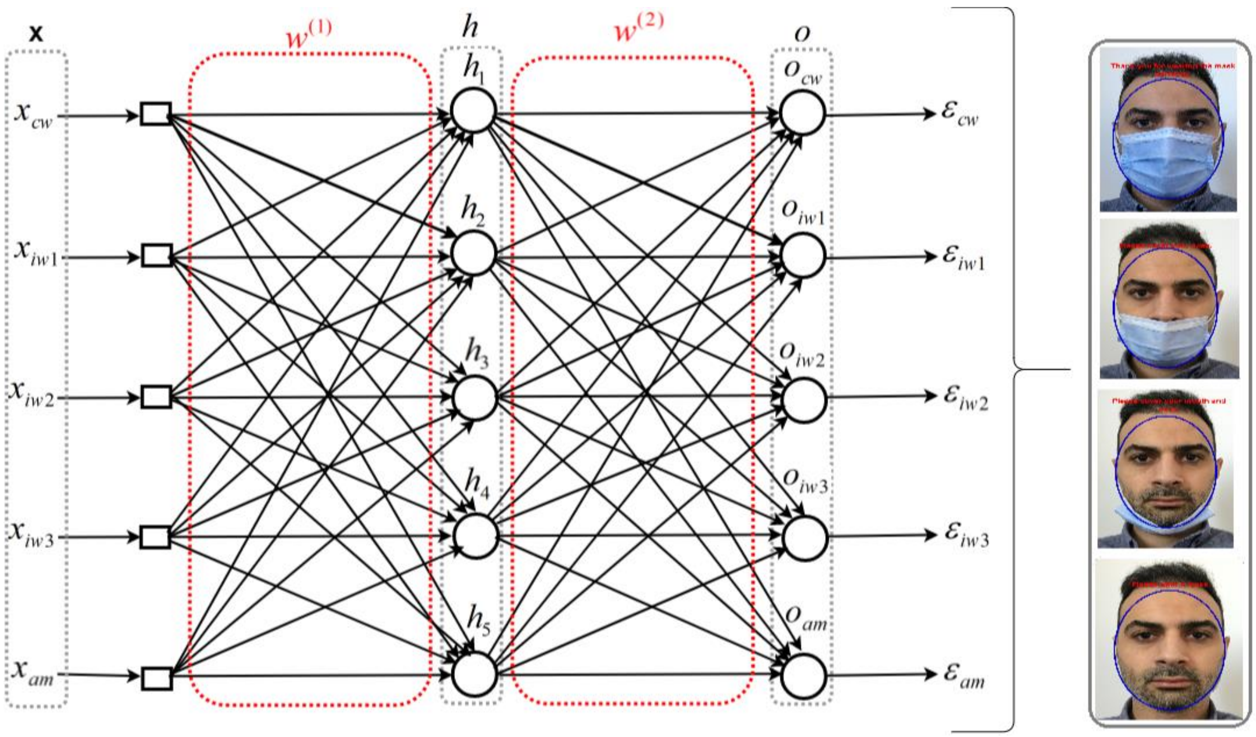


Figure 6.8. Illustration of the NN algorithm.

Herein h_1, h_2, h_3, $h_4$and h_5 represent each hidden node, respectively. The predicted outputs of the NN are:

$$o = [o_{cw} \quad o_{iw1} \quad o_{iw2} \quad o_{iw3} \quad o_{am}] \tag{6.10}$$

where o_{cw}, o_{iw1}, o_{iw2}, o_{iw3} represent the predicted outputs for the correctly worn mask and various incorrect wearing conditions.

Additionally, the unknown parameters between the input and hidden layers of the NN are denoted by $w^{(1)}$, while those between the hidden and output layers are labelled as $w^{(2)}$:

$$w^{(1)} = w^{(2)} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} & w_{15} \\ w_{21} & w_{22} & w_{23} & w_{24} & w_{25} \\ w_{31} & w_{32} & w_{33} & w_{34} & w_{35} \\ w_{41} & w_{42} & w_{43} & w_{44} & w_{45} \\ w_{51} & w_{52} & w_{53} & w_{54} & w_{55} \end{bmatrix} \tag{6.11}$$

Furthermore; ε_{cw}, ε_{iw1}, ε_{iw2}, ε_{iw},ε_{am} error values between the predicted and actual outputs for each facial mask-wearing scenario are quantified using the specified metrics. The actual output vector is defined as follows:

$$y = [y_{cw} \quad y_{iw1} \quad y_{iw2} \quad y_{iw3} \quad y_{am}] \tag{6.12}$$

As illustrated in Figure 6.9, the network's objective is to assess the predicted output values against the actual output values, aiming to minimize their discrepancies in order to accurately classify each condition.

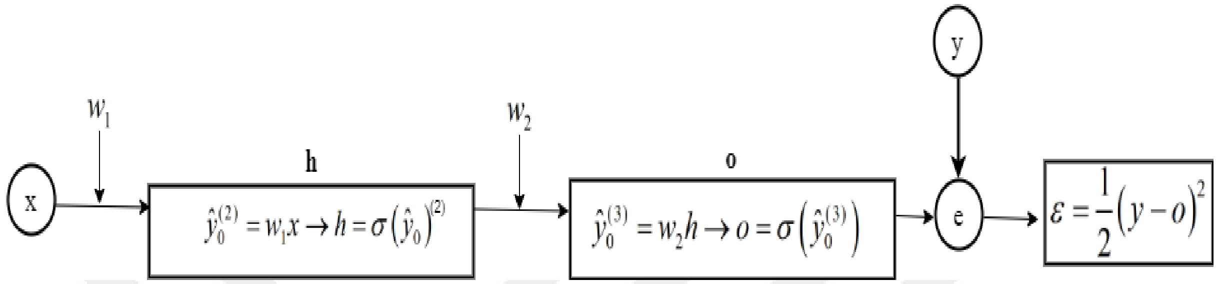


Figure 6.9. The expanded illustration of the simple k^{th}layer NN.

To minimize error differences and classify each facial mask-wearing condition, it is essential to define the relationship between changes in unknown parameters and their error gradients concerning each NN parameter.

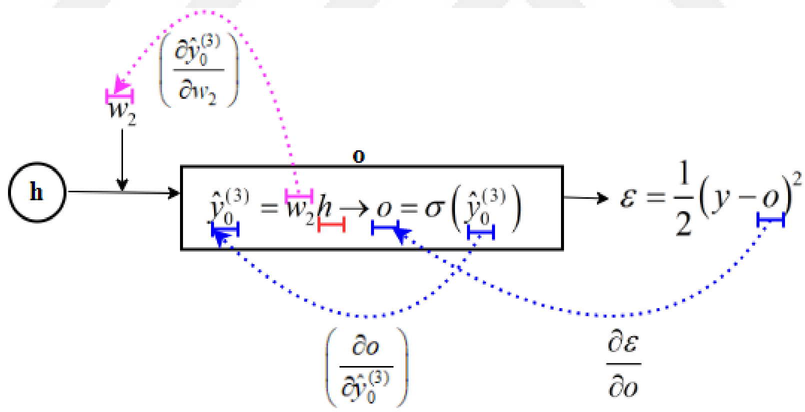


Figure 6.10. Gradient computation for output to hidden layer unknown parameter w_2 .

The calculation of error gradients begins with determining the output to hidden layer unknown parameters, as illustrated in order to minimize the error difference and classify each facial mask-wearing, the relationship between the unknown parameter changes and their error gradients should be defined with respect to each NN parameters. In order to compute the error gradients, initially, calculation of the output to hidden layer unknown parameters are formed. It is evident that the $\sigma(w_2 h)$ value produces output layer activation through a weighted sum involving the $\hat{y}_o^{(3)}$ hidden-to-output parameter. Here, σ denotes the activation function, and o is derived from $\sigma\left(\hat{y}_o^{(3)}\right)$ and compared to the desired output values y to estimate the error gradient of the output, with e as the target ε. The corresponding ε value is calculated as follows:

$$\varepsilon = \frac{1}{2}(y - o)^2 \tag{6.13}$$

In Figure 6.10, the reverse arrows illustrate the path from w_2 to ε for calculating the gradient $\frac{\partial \varepsilon}{\partial w_2}$ using the chain rule, as follows:

$$\frac{\partial \varepsilon}{\partial w_2} = \frac{\partial \varepsilon}{\partial o} \frac{\partial o}{\partial \hat{y}_o^{(3)}} \frac{\partial \hat{y}_o^{(3)}}{\partial w_2} \tag{6.14}$$

where

$$\hat{y}_o^{(3)} = w_2 h \tag{6.15}$$

and $o = \sigma\left(\hat{y}_o^{(3)}\right) = \frac{1}{1+e^{-\hat{y}_o^{(3)}}}$ (6.16)

Subsequently, the intermediate partial derivatives in Equation (2) can be readily computed as follows:

$$\frac{\partial \varepsilon}{\partial o} = -(y - o) = -e \tag{6.17}$$

$$\frac{\partial o}{\partial \hat{y}_o^{(3)}} = \sigma'\left(\hat{y}_o^{(3)}\right) = o(1 - o) \tag{6.18}$$

$$\frac{\partial \hat{y}_o^{(3)}}{\partial w_2} = \sigma(w_2 h) \tag{6.19}$$

where $h = \sigma\left(\hat{y}_o^{(2)}\right) = \frac{1}{1+e^{-\hat{y}_o^{(2)}}}$ (6.20)

Substituting Equation (6.20) into Equation (6.14) results in:

$$\frac{\partial \varepsilon}{\partial w_2} = -e\sigma'\left(\underbrace{\hat{y}_o^{(3)}}_{\Delta^{(2)}}\right)\sigma(w_2 h) = -\Delta^{(2)}\sigma(w_2 h) \tag{6.21}$$

Equation (6.21) generally represents the product of the error at the post-synaptic neuron, the derivative of the activation function at the post-synaptic neuron's value, and the signal from the pre-synaptic neuron. To compute the gradients for unknown parameters between the hidden and input layers, Figure 6.11 illustrates the pathway from an input x to an output o, incorporating the intermediate variables generated according to the activation function.

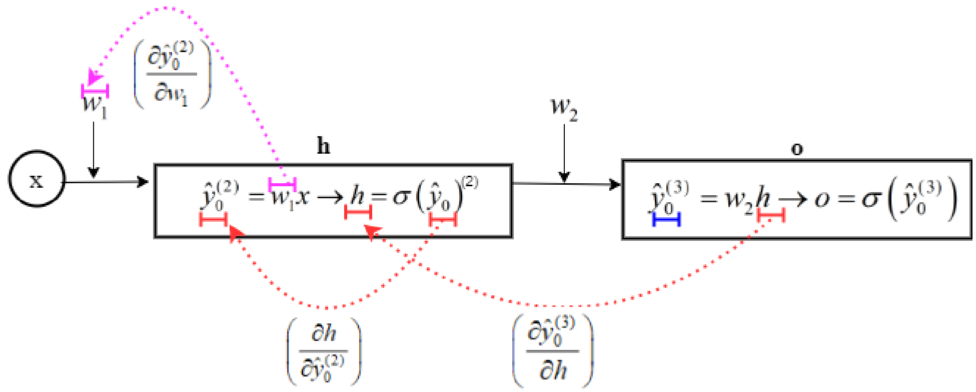


Figure 6.11. Gradient computation for input to hidden layer unknown parameter w_1.

In Figure 6.11, the reverse arrows indicate the variable directions used to compute the target error derivative $\frac{\partial \varepsilon}{\partial w_1}$. The following equation is thus determined:

$$\frac{\partial \varepsilon}{\partial w_1} = \frac{\partial \varepsilon}{\partial h} \frac{\partial h}{\partial \hat{y}_o^{(2)}} \frac{\partial \hat{y}_o^{(2)}}{\partial w_1} \tag{6.22}$$

where $\hat{y}_o^{(2)} = \sum_{i=0}^{n} w_1 x$ (6.23)

The following figure illustrates the expanded variable path needed to calculate $\frac{\partial \varepsilon}{\partial h}$, derived from the chain rule, providing a comprehensive view of the entire process.

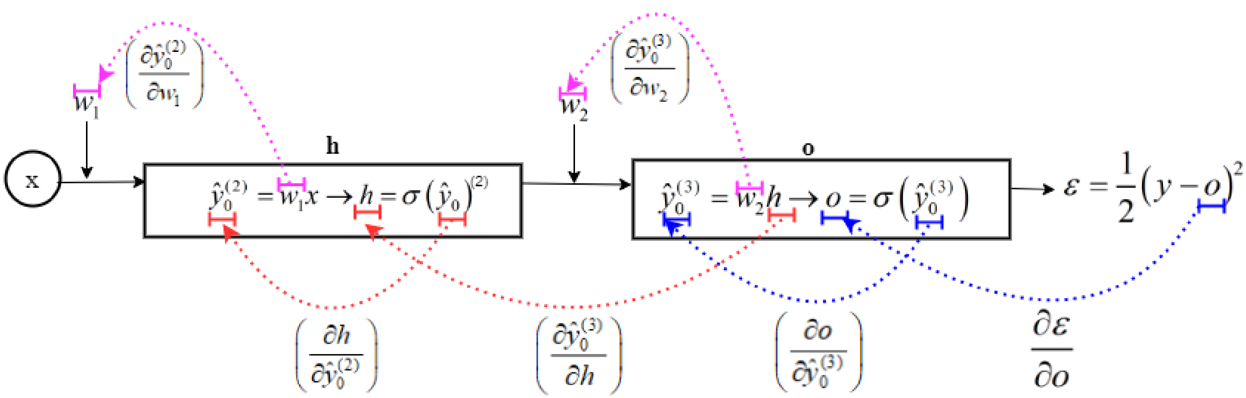


Figure 6.12. Whole picture of the gradient computation.

Since ε is a function of the error gradients at each of output neuron, the path to h traverses through p as indicated by the dashed arrows.

$$\frac{\partial \varepsilon}{\partial h} = \frac{\partial \varepsilon}{\partial \hat{y}_o^{(3)}} \frac{\partial \hat{y}_o^{(3)}}{\partial h} \tag{6.24}$$

The substitution of Equation (6.25) into Equation (6.11) yields:

$$\frac{\partial \varepsilon}{w_1} = \left\{\frac{\partial \varepsilon}{\partial o}\frac{\partial o}{\partial \hat{y}_o^{(3)}}\frac{\partial \hat{y}_o^{(3)}}{\partial h}\right\}\sigma'\left(\hat{y}_o^{(2)}\right)\sigma(x) = \left\{\underbrace{-e_o\sigma'\left(\hat{y}_o^{(3)}\right)}_{\Delta^{(2)}}w_2\right\}\sigma'\left(\hat{y}_o^{(2)}\right)x = -\left(\Delta^{(2)}w_2\right)\sigma'\left(\hat{y}_o^{(2)}\right)x \tag{6.25}$$

The $\Delta^{(2)}$ values have already been computed for each output neuron and are then back-propagated to a hidden neuron via the unknown parameters connecting hidden neurons to output neurons. To determine the contribution of the hidden neuron to the output error, the scaled errors $\Delta^{(2)}$ are propagated back through the unknown parameters w_2. In other words, the error contribution of the h^{th} hidden node is:

$$e_h = \Delta^{(2)}w_2 \tag{6.26}$$

and

$$\Delta^{(1)} = e_h\sigma'\left(\hat{y}_o^{(2)}\right) \tag{6.27}$$

we therefore obtain

$$\frac{\partial \varepsilon}{\partial w_1} = \Delta^{(1)}x \tag{6.28}$$

The derivation of updates for unknown parameters in the output-to-hidden layer is now straightforward.

$$w_2^{new} = w_2 + \Delta w_2 = w_2 + \eta\left(-\frac{\partial \varepsilon}{\partial w_2}\right) \tag{6.29}$$

Herein, η represents a learning rate parameter within the range $0 < \eta \leq 2/x^T x$. For the unknown parameters from the hidden layer to the input layer:

$$w_1^{new} = w_1 + \Delta w_1 = w_1 + \eta\left(-\frac{\partial \varepsilon}{\partial w_1}\right) \tag{6.30}$$

Finally, the derivation of the fundamental NN structure is completed using the following equation to learn and classify each facial mask-wearing scenario:

$$= w_1 + \eta\sigma_h x \tag{6.31}$$

Furthermore, the training process of the neural network algorithm is summarized by the following pseudo-code:

Algorithm 6.1. NN classifier for m facial mask-wearing situations and n Outputs.

Input: Training input value for each matched facial mask-wearing conditions$x \in \mathbb{R}^{m\times l}$, where l is the length of the each m number of the input data, labelled real output$y \in \mathbb{R}^{l\times n}$, where n is the number of the output, randomly initialized unknown parameter matrix/vector $w \in \mathbb{R}^{m\times n}$, weighted sum of the predicted output node $\hat{y}_o^{(3)} \in \mathbb{R}^{n\times l}$, weighted sum of the hidden node $\hat{y}_o^{(2)} \in \mathbb{R}^{n\times l}$ and estimated output $o \in \mathbb{R}^{n\times l}$, estimated output of hidden layer $h \in \mathbb{R}^{n\times l}$, the training error for hidden layer $\Delta^{(1)} \in \mathbb{R}^{l\times n}$, the training error for output layer $\Delta^{(2)} \in \mathbb{R}^{l\times n}$ the parameter learning rate η, store matrix w_s, store error e_s, error stopping threshold e_t, stored estimated output o_s.

Output: The terminal value of the trained unknown parameters w, store learning error e_s, store output of the predicted facial mask wearing conditions o_s

for i to iteration

for j to l

1. Calculate the weighted sum of hidden node $\hat{y}_{o,1}$ as: $\hat{y}_o^{(2)}(:,j) = w^{(1)T}x(:,j)$
2. Apply a threshold σ to the output from the related hidden nodes as: $h(:,j) = \sigma\left(\hat{y}_o^{(2)}(:,j)\right)$
3. Calculate the weighted sum of the output node as: $\hat{y}_o^{(3)}(:,j) = w^{(2)T}h(:,j)$
4. Apply a threshold σ to the output as: $o(:,j) = \sigma\left(\hat{y}_o^{(3)}(:,j)\right)$
5. Determine the error for the related output layer $\varepsilon(:,j) = \frac{1}{2}\left(y - o(:,j)\right)^2$
6. Calculate gradient for the output layer $\Delta^{(2)}(:,j) = o(:,j)(1-o(:,j))\varepsilon(:,j)$
7. Determine the error for the related hidden layer $\varepsilon h(:,j) = w^{(2)}(:,j)\Delta^{(2)}(:,j)$
8. Calculate gradient for hidden layer $\Delta^{(1)}(:,j) = h(:,j)(1-h(:,j))\varepsilon h(:,j)$
9. Update and store the unknown parameter for hidden layer
   $w_{new}^{(1)} \to w^{(1)} + \eta\Delta^{(1)}x(:,j)^T$
   $w_s = \left[w_s; reshape\left(w_{new}^{(1)}, 1, []\right)\right]$
10. Update and store the unknown parameter for output layer
   $w_{new}^{(2)} \to w^{(2)} + \eta\Delta^{(2)}h(:,j)^T$
   $w_{s2} = \left[w_{s2}; reshape\left(w_{new}^{(2)}, 1, []\right)\right]$

end j

11. Store the error and the saturated output
    $e_s = [e_s; reshape(\varepsilon, 1, [])]$
    $o_s = [o_s; reshape(o, 1, [])]$
    If $e(:,j) < e_t$ then
    break
    end if

end i

Following the provision of the relevant pseudo-code, the subsequent section offers a detailed explanation of the implementation of the proposed algorithm for NAO HR.

6.4. Implementation of The Proposed Algorithm to NAO HR

This section provides a concise overview of implementing the proposed algorithm on the NAO HR.

6.4.1. Dialogue Model for NAO HR

In this study, the NAO HR executes the Facial Mask-Wearing Control (FMWCC) algorithm in five stages. Initially, the robot engages in dialogue with individuals within its camera's range during the initialization phase. Upon completion of this phase, the system progresses to the facial detection stage, employing the Viola-Jones Algorithm to activate the face and facial mask tracking modules, FT and MT, respectively.

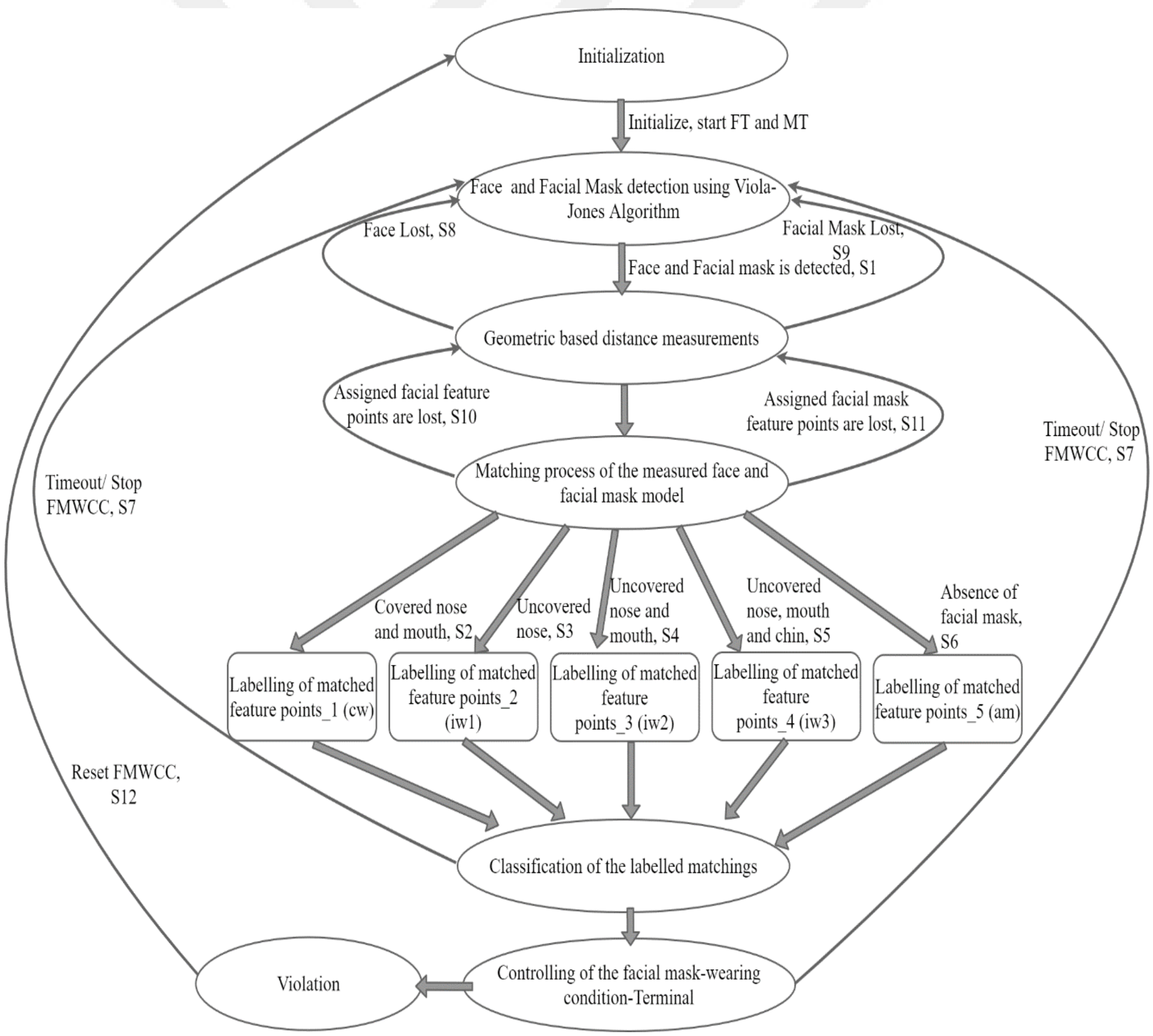


Figure 6.13. Dialogue model of the FMWCC algorithm.

In Figure 6.13, the abbreviations represent various functions and statements of the system: “FT” denotes Face Tracking, “MT” refers to Facial Mask Tracking, “FD” indicates Face Detection, “MD” signifies Facial Mask Detection. The statements include: “S1” introduces the robot, “S2” thanks for proper mask use, “S3” advises on mask usage, “S4” emphasizes covering both nose and mouth, “S5” instructs against mask placement under the chin, “S6” urges immediate mask use, “S7” indicates inability to analyze mask wearing, “S8” notes the face is not detected, “S9” mentions the mask is not detected, “S10” highlights missing facial feature points, “S11” indicates missing mask feature points, and “S12” reports the user to the relevant authority.

As both a face and facial masks are detected, the “FT” and “MT” modules allow the NAO HR to track them. Following detection, the NAO HR introduces itself (S1) and proceeds to the “Geometric-Based Distance Measurements” stage to assess both face and mask dimensions. If either the face or mask is no longer detected, the dialogue manager initiates statements S8 and S9 to return to the “Face and Facial Mask Detection using the Viola-Jones Algorithm” stage. Upon completion of this stage, the dialogue manager advances to the “Matching Process of the Measured Face and Facial Mask,” where it matches and labels the mask usage. If the assigned feature points are missing, the dialogue manager issues statements S10 and S11 to revert to the “Geometric-Based Distance Measurements” stage.

NAO HR then classifies each condition by advancing to the “Classification of the Labelled Matchings” stage using the obtained matching results. In this stage:

If the nose, mouth and chin regions of the detected human face are covered properly
  NAO HR enters to the “S2” statement.
If the warning is not taken into account by the violator
   NAO HR sends a photo of the violator to the relevant officials via “Violation” stage and enters to the “S12” statement.
     If the nose region of the detected human face is uncovered
       NAO HR enters to the “S3” statement.
     If the warning is not taken into account by the violator
       NAO HR sends a photo of the violator to the relevant officials via “Violation” stage and enters to the “S12” statement.
     If the nose and mouth regions of the detected human face are uncovered
       NAO HR enter to the “S4” statement.

If the warning is not taken into account by the violator

NAO HR sends a photo of the violator to the relevant officials via "Violation" stage and enters to the "S12" statement.

If the nose, mouth, and chin regions of the detected human face are uncovered

NAO HR enter to the "S5" statement.

If the warning is not taken into account by the violator

NAO HR sends a photo of the violator to the relevant officials via "Violation" stage and enters to the "S12" statement.

If there is no any facial mask over the detected human face

NAO HR enter to the "S6" statement.

If the warning is not taken into account by the violator

NAO HR sends a photo of the violator to the relevant officials via "Violation" stage and enters to the "S12" statement.

If the tracked face is lost, the dialogue manager presumes the individual has left the scene and prompts a return to the "Face and Facial Mask Detection using Viola-Jones Algorithm" stage with the "S7" statement. After presenting the results of the "FMWCC" algorithm, the system transitions back to the "Initialization" stage.

Following a comprehensive explanation of the implementation of the proposed algorithm for the NAO HR, the subsequent section presents the simulation and experimental results of the proposed algorithm.

6.5. Simulation and Experimental Results

This section assesses the proposed algorithm in both simulation and experimental environments. It begins with an introduction to the experimental setup, followed by an evaluation of the results.

6.5.1. Simulation and Experimental Settings

To evaluate the effectiveness of the proposed algorithm, a self-collected dataset was created. This dataset includes five sessions, each lasting 15 seconds and involving 5 subjects. In the first session, subjects were instructed to wear their facial masks correctly, covering the nose, mouth, and chin. In the second session, they were asked to cover only the mouth and chin. The third session required covering only the chin. In the fourth session, subjects were instructed to position the masks under their chin. Finally, in the fifth session, subjects were asked to remove the masks completely from their faces.

6.5.2. Results

Following the explanation of the simulation and experimental settings, the accuracy of the proposed algorithm's classification performance is analyzed as the first experiment. This analysis involves evaluating the algorithm's performance for a single subject across the five sessions, with each session tested in five separate trials.

Table 6.5. Matching results for 1 person for five trials.

Matched Conditions	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
cw	98.01	98.00	98.00	98.03	98.00
iw1	98.00	98.00	98.02	98.01	98.01
iw2	98.03	98.01	98.00	98.01	98.01
iw3	98.01	98.01	98.01	98.03	98.01
am	99.00	98.00	98.03	98.03	98.01

As shown in Table 6.5, although there are minor variations between classified values across trials, these differences do not impact the overall results of the proposed algorithm. Following this numerical analysis, the algorithm was implemented on NAO HR's interface via Choregraphe to validate its applicability and reliability. In the subsequent experiment, NAO HR's detection performance was evaluated with a single subject wearing a facial mask correctly (session 1). Once the subject approached NAO HR in this condition, the system promptly checked the facial mask and thanked the subject for proper mask use, as illustrated in Figure 6.14.

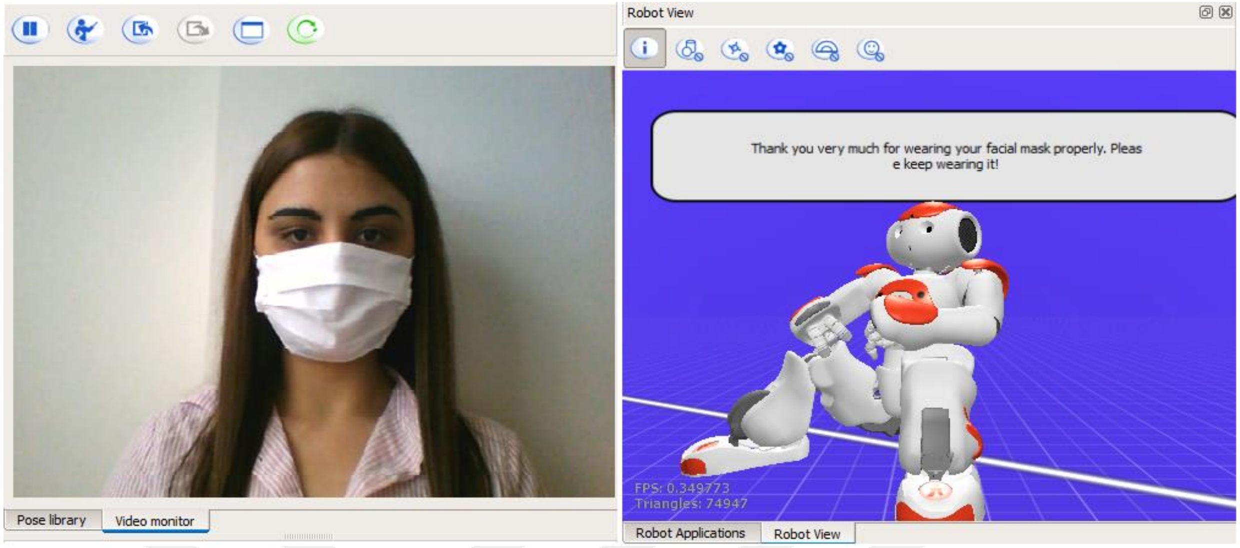


Figure 6.14. Session 1: Implementation of the FMWCC algorithm via Choreographe in response to the "cw" condition.

The subsequent experiment evaluates the NAO HR's detection efficacy under the uncovered nose condition (session 2).

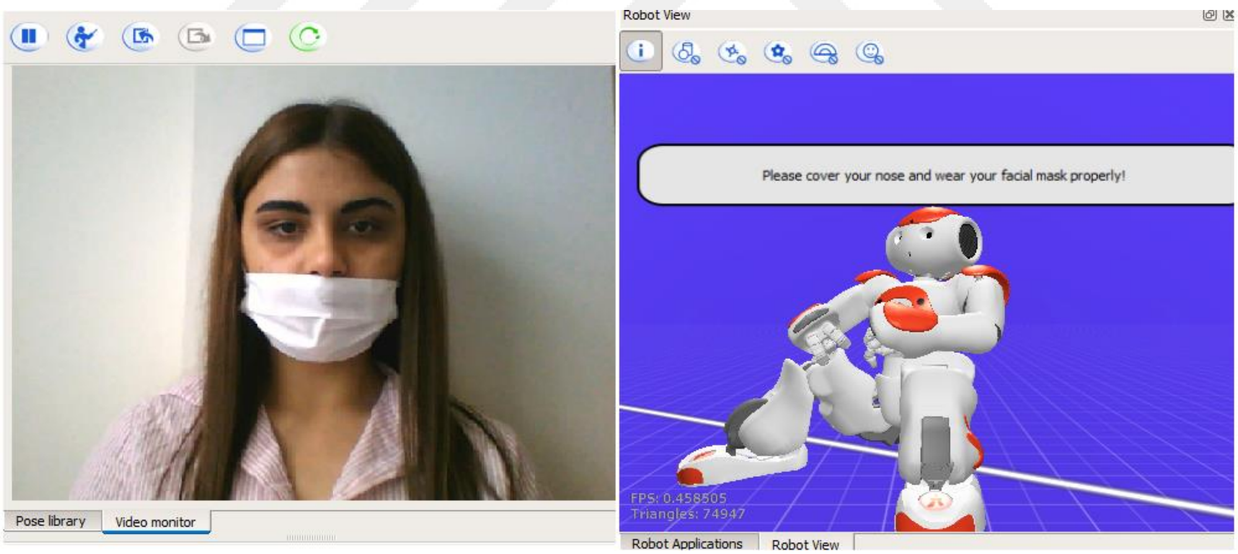


Figure 6.15. Session 2: Implementation of the FMWCC algorithm via Choreographe in response to the "$iw1$" condition.

If the subject approaches the HR under this condition, the system promptly assesses the facial mask status and alerts the subject to correct its placement.

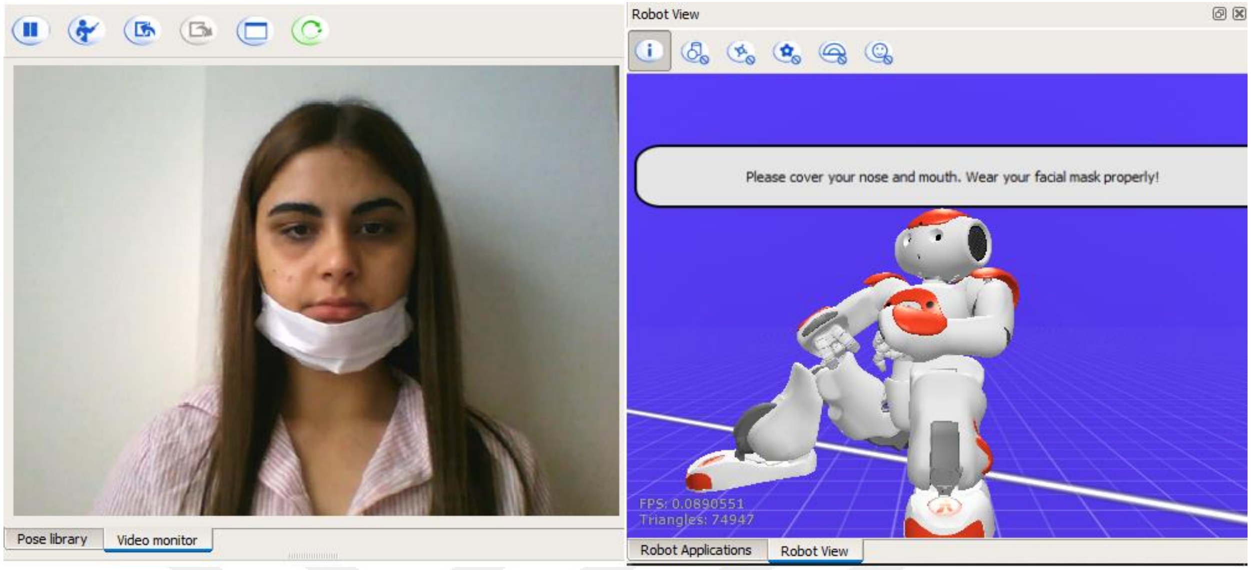


Figure 6.16. Session 3: Implementation of the FMWCC algorithm via Choreographe in response to the "$iw2$" condition.

Then, the NAO HR's performance in detecting uncovered nose and mouth scenarios (session 3) is evaluated, as depicted in Figure 6.16.

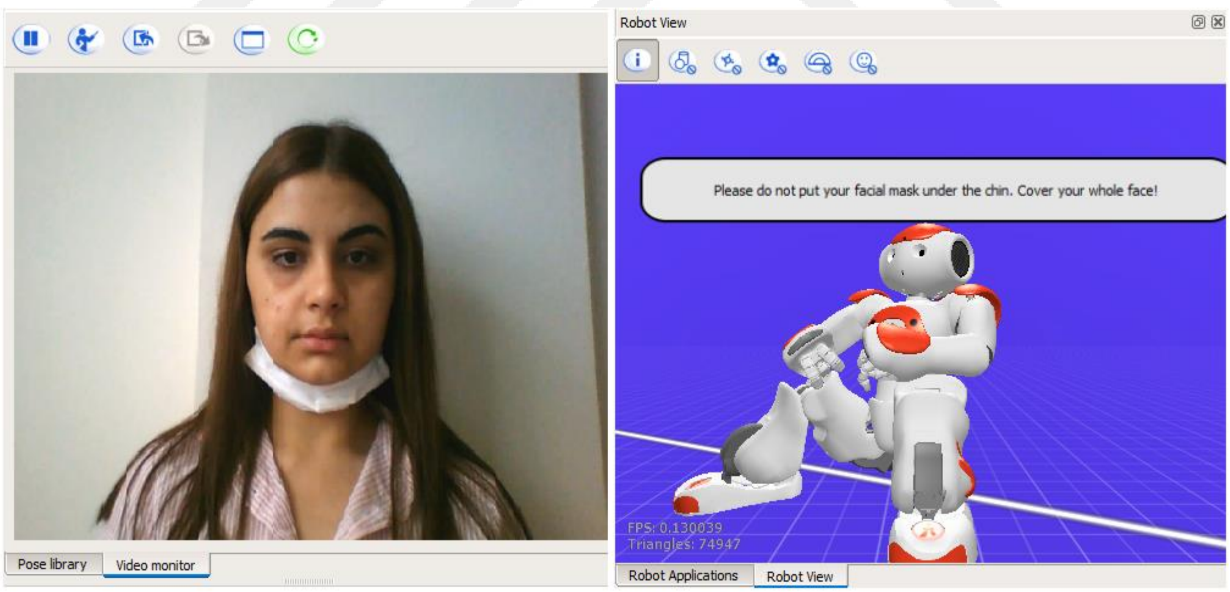


Figure 6.17. Session 4: Implementation of the FMWCC algorithm via Choreographe in response to the "$iw3$" condition.

In these instances, the NAO HR promptly adjusts the facial mask and issues an immediate warning to the subject to ensure proper mask usage. In the event that the subject presents with an uncovered nose, mouth, and chin (session 4), as illustrated in Figure 6.17, NAO HR adjusts its facial mask and promptly issues a warning to the subject to ensure proper mask usage. In the final experiment, the NAO HR's performance in detecting the absence of a facial mask (session 5) is assessed. As illustrated in Figure 6.18, NAO HR promptly adjusts the subject's facial mask and issues an immediate warning to ensure its proper use. If the subject disregards the warning, they are classified as a violator.

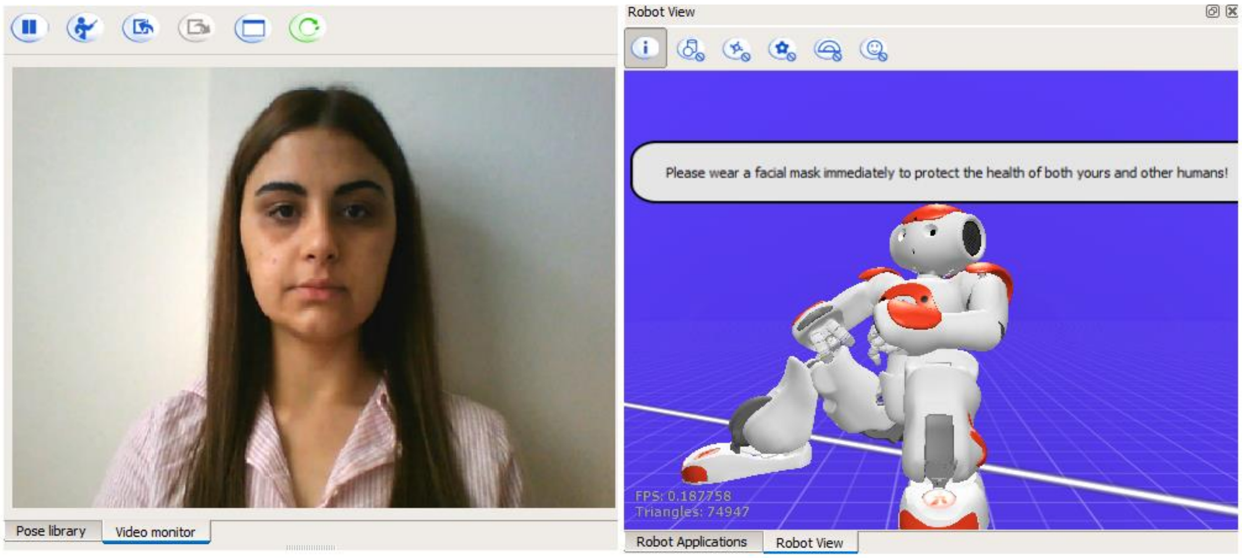


Figure 6.18. Session 5: Implementation of the FMWCC algorithm via Choreographe in response to the “*am*” initial condition.

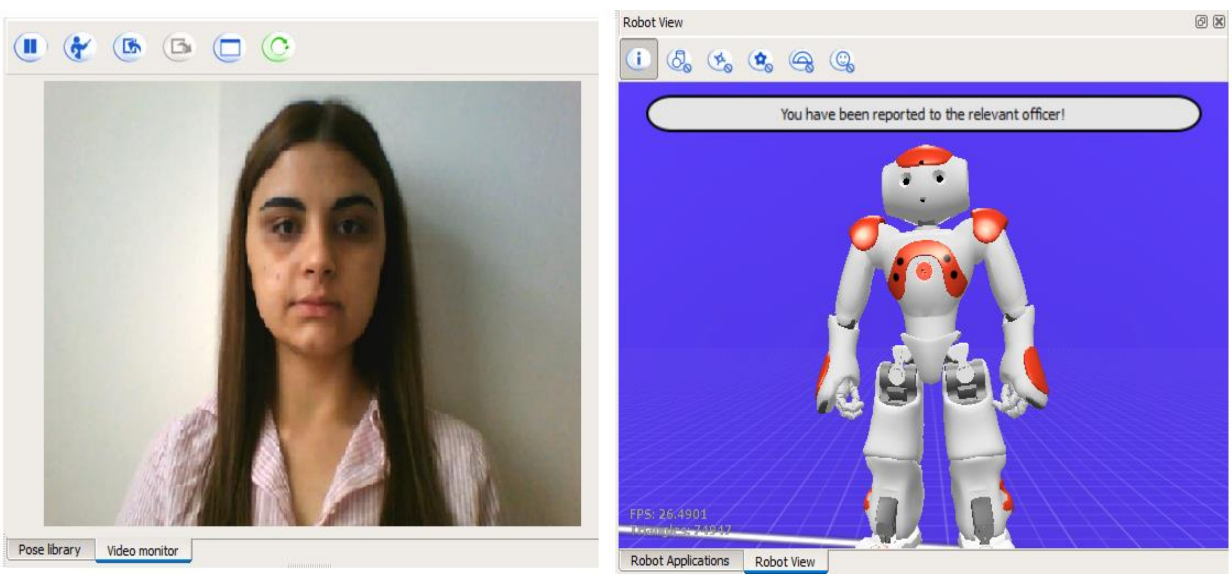


Figure 6.19. Session 5: Implementation of the FMWCC algorithm via Choreographe in response to the “*am* warning” condition.

The NAO HR then transmits the violator's photo to the relevant authorities via Wi-Fi and notifies the subject of this action, as depicted in Figure 6.19. After conducting a thorough evaluation of both the simulation and experimental results, the final section presents a detailed analysis of the effectiveness of the proposed algorithm.

6.6. Discussions

This section introduces a four-stage protective algorithm designed for HRs to mitigate the spread of COVID-19 by monitoring facial mask use. The algorithm first employs the Viola-Jones method to detect faces and masks, followed by geometric distance measurement to assess critical distances between the face and mask. It then matches these measurements to evaluate mask-wearing conditions. In the final stage, a three-layer neural network classifies the results. The algorithm's reliability is rigorously validated by analysing terminal decisions across all stages, assessing its accuracy in detecting various mask-wearing violations, including uncovered nose, nose and mouth, chin, and absence of the mask.

Integrating this algorithm into HRs designed as doctors provides substantial benefits for health-care settings. HRs equipped with this algorithm can autonomously monitor and enforce mask-wearing protocols, enhancing infection control measures in clinical environments. By ensuring that patients and staff adhere to mask-wearing guidelines, they contribute to reducing the risk of disease transmission. Additionally, the algorithm’s real-time detection and classification capabilities enable prompt corrective actions and accurate reporting of compliance issues. This integration not only improves adherence to health protocols but also enhances the overall safety and efficiency of healthcare operations, making robots valuable assets in managing public health crises and maintaining a safe environment.

7. PHYSICAL EXERCISE OF GERIATRIC PATIENTS WITH ASSISTIVE HEALTH-CARE HRS

The final section of the thesis presents a four-stage physical exercise algorithm and details its application to HRs. The first stage utilizes the Viola-Jones algorithm for precise detection and tracking of the human body. In the second stage, geometric distance measurements are applied to analyze human actions, focusing on exercises for elderly individuals. The third stage involves learning these actions using an LSTM-RNN integrated with a genetic algorithm. The final stage implements an FSM dialogue model for structured, user-friendly interaction. Although the algorithm is demonstrated on Alpha Mini and Alpha 1Pro HRs, it is applicable to other HRs as well.

7.1. Human Body Detection using Viola-Jones Algorithm

This sub-section offers a brief overview of the Viola-Jones algorithm for human body detection. Initially, the algorithm segments the input image into parallelogram rectangles to improve feature extraction efficiency. Typically, two, three, or four rectangles are used to capture the relevant human body features, as illustrated in the following figure.

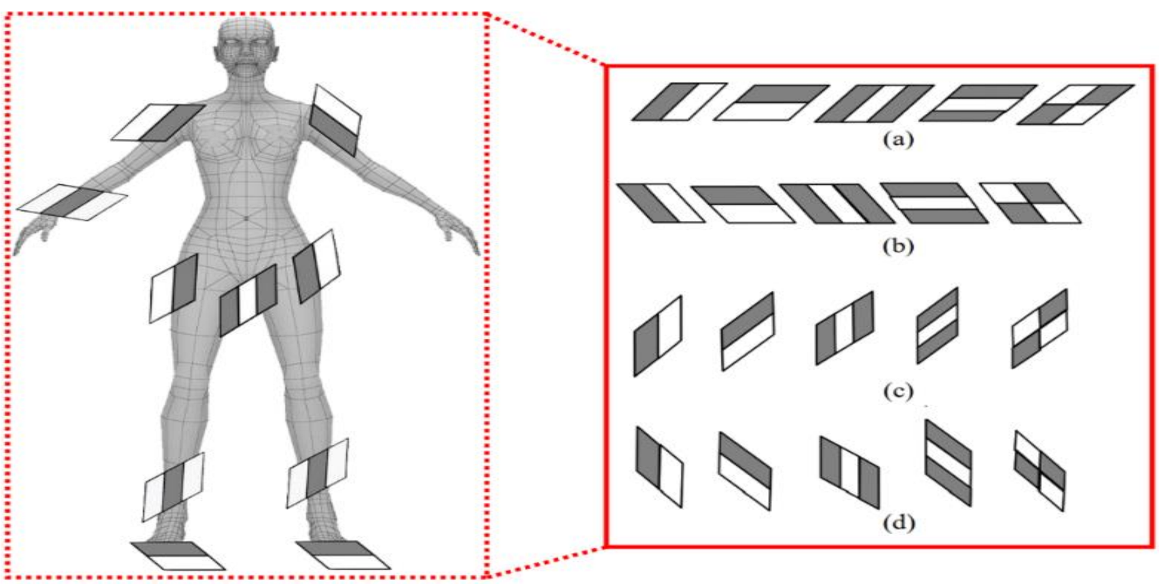


Figure 7.1. Rectangle-based feature collection with two, three, and four rectangles.

After extracting these features, the algorithm calculates the sum of features within each rectangle and across the entire image, iterating this process for various features and scales. To enhance computational efficiency, Summed Area Charts (C_R) are employed for the designated rectangles. Four distinct tables are created for different feature types. For instance, Figure 7.2 (a) shows the first feature type $C_R^{(1)}$of C_R, and its calculation is defined by:

$$C_R^{(1)}(x,y)=\sum_{j=1}^{y}\sum_{i=1}^{x+y-j}H(i,j) \tag{7.1}$$

where $H(i,j)$ denote the intensity of the human body image at pixel (i,j). The expansion of this equation involves calculating the intensity values for each pixel sequentially from left to right and top to bottom, as given by:

$$C_R^{(1)}(x,y) = C_R^{(1)}(x-1,y) + C_R^{(1)}(x+1,y-1) - C_R^{(1)}(x,y-1) + H(x,y) \tag{7.2}$$

The variables $C_R^{(1)}(x,0) = 0$ and $C_R^{(1)}(0,y) = C_R^{(1)}(1,y-1)$ are defined for all x and y values. The other three feature types of C_R are calculated in a similar manner.

$$C_R^{(2)}(x,y)=\sum_{j=1}^{y}\sum_{i=x+j-y}^{k}H(i,j)=C_R^{(2)}(x-1,y-1)+C_R^{(2)}(x+1,y)-C_R^{(2)}(x,y-1)+H(x,y) \tag{7.3}$$

$$C_R^{(3)}(x,y)=\sum_{i=1}^{x}\sum_{j=1}^{y+x-j}H(i,j)=C_R^{(3)}(x-1,y+1)+C_R^{(3)}(x,y-1)-C_R^{(3)}(x-1,y)+H(x,y) \tag{7.4}$$

$$C_R^{(4)}(x,y)=\sum_{i=x}^{k}\sum_{j=y+i-x}^{l}H(i,j)=C_R^{(4)}(x,y-1)+C_R^{(4)}(x+1,y+1)-C_R^{(4)}(x+1,y)+H(x,y) \tag{7.5}$$

where k and l represent the width and height of the image, respectively.

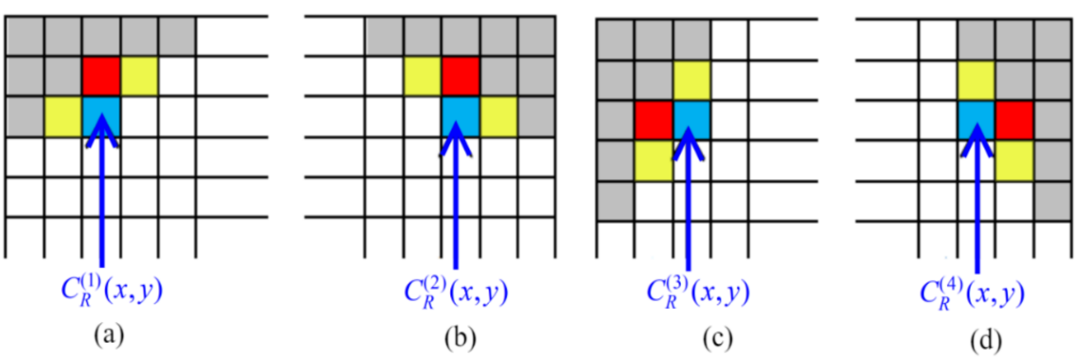


Figure 7.2. Calculation of C_R charts.

Subsequently, the cumulative sum of intensities within a defined rectangular area (IS_R) can be calculated using the predetermined C_R values, as illustrated in Figure 7.3.

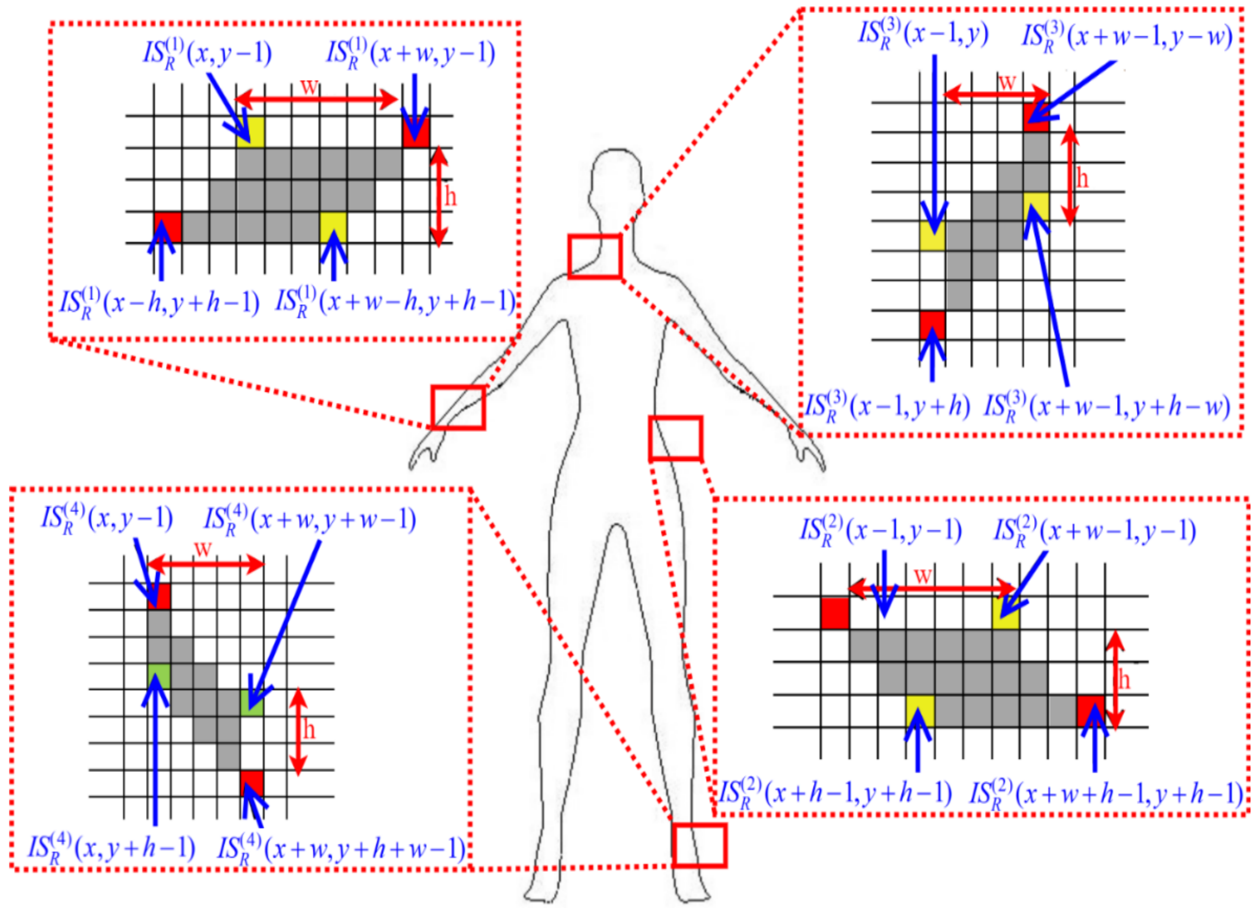


Figure 7.3. Calculation of IS_R.

For example, the calculation of IS_R for the initial feature set is conducted as follows:

$$IS_R^{(1)}(x,y,w,h)=C_R^{(1)}\left(x+w-h,y+h-1\right)+C_R^{(1)}\left(x-h,y+h-1\right)-C_R^{(1)}\left(x+w,y-1\right)-C_R^{(1)}\left(x,y-1\right) \tag{7.6}$$

Herein, (x, y) denotes the coordinates of the top-left corner of the image region, with w and h representing the width and height of the parallelogram region, respectively. The calculation of IS_R for the remaining C_R values follow a similar approach, as detailed below:

$$\begin{aligned}IS_R^{(2)}(x,y,w,h)=&C_R^{(2)}\left(x+h-1,y+h-1\right)+C_R^{(2)}\left(x+w-1,y-1\right)-C_R^{(2)}\left(x-1,y-1\right)\\&-C_R^{(2)}\left(x+w+h-1,y+h-1\right)\end{aligned} \tag{7.7}$$

$$\begin{aligned}IS_R^{(3)}(x,y,w,h)=&C_R^{(3)}\left(x+w-1,y+h-w\right)+C_R^{(3)}\left(x-1,y\right)-C_R^{(3)}\left(x-1,y+h\right)\\&-C_R^{(3)}\left(x+w-1,y-w\right)\end{aligned} \tag{7.8}$$

$$\begin{aligned}IS_R^{(4)}(x,y,w,h)=&C_R^{(4)}\left(x,y+h-1\right)+C_R^{(4)}\left(x+w,y+w-1\right)-C_R^{(4)}\left(x,y-1\right)\\&-C_R^{(4)}\left(x+w,y+h+w-1\right)\end{aligned} \tag{7.9}$$

After aggregating features within rectangles with varying intensities, the AdaBoost algorithm classifies these aggregated features as either weak or strong candidates for detecting the target human body. The categorized candidates are then compared with designated testing samples for each body component. Successful candidates advance through subsequent stages, leading to the identification of the respective human body components.

7.2. Human Action Characterization using Geometric-Based Distance Measurement

Following the human body detection phase, the detected body is mapped onto a coordinate system to analyze actions. Initially, the center of the detected body, along with its associated parts—such as the neck, shoulders, arms, torso, legs, and feet—is defined using the bounding box, as shown in Figure 7.4.

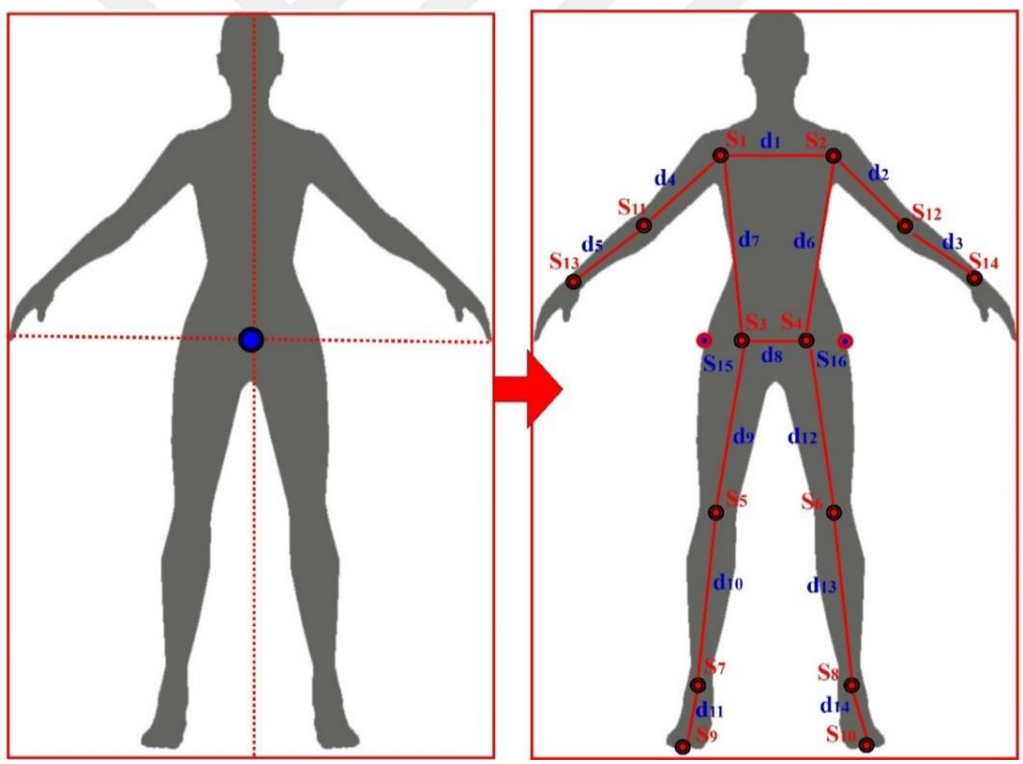


Figure 7.4. Placement of the human body points in the geometric frames.

Once the centre of the human body is identified, sixteen feature points are strategically placed to track the movements of the corresponding body parts, as detailed in Table 7.1. The distances and angles between these feature points are then accurately measured to assess the action states of the individuals.

Table 7.1. Description of the geometric-based distance measurements.

Point	Description	Dist.	Description	Dist.	Description
s_1	Right shoulder	d_1	Distance between s_1 and s_2	d_{17}	Distance between s_{11} and s_{15}
s_2	Left shoulder	d_2	Distance between s_2 and s_{12}	d_{18}	Distance between s_{13} and s_{15}
s_3	Right hip	d_3	Distance between s_{12} and s_{14}	d_{19}	Distance between s_9 and s_{10}
s_4	Left hip	d_4	Distance between s_1 and s_{11}	d_{20}	Distance between s_{11} and $s_{11}-1$
s_5	Right Knee	d_5	Distance between s_{11} and s_{13}	d_{21}	Distance between s_{12} and $s_{12}-1$
s_6	Left knee	d_6	Distance between s_2 and s_4	d_{22}	Distance between s_{14} and $s_{14}-1$
s_7	Right ankle	d_7	Distance between s_1 and s_3	d_{23}	Distance between s_{13} and $s_{13}-1$
s_8	Left ankle	d_8	Distance between s_3 and s_4	q_1	Angle between s_2, s_1, s_3
s_9	Right big toe	d_9	Distance between s_3 and s_5	q_2	Angle between s_{15}, s_1, s_{11}
s_{10}	Left big toe	d_{10}	Distance between s_5 and s_7	q_3	Angle between s_1, s_{11}, s_{13}
s_{11}	Right elbow	d_{11}	Distance between s_7 and s_9	q_4	Angle between s_1, s_2, s_{16}
s_{12}	Left elbow	d_{12}	Distance between s_4 and s_6		
s_{13}	Right wrist	d_{13}	Distance between s_6 and s_8	q_5	Angle between s_{16}, s_2, s_{12}
s_{14}	Left wrist	d_{14}	Distance between s_8 and s_{10}		
s_{15}	Left border hip	d_{15}	Distance between s_{12} and s_{16}	q_6	Angle between s_2, s_{12}, s_{14}
s_{16}	Right border hip	d_{16}	Distance between s_{14} and s_{16}		

In this context, the relative distance measurements d_{ij} are calculated along the x-y axes between two feature points $(s_{x,i}, s_{y,i})$ and $(s_{x,j}, s_{y,j})$ as follows:

$$d_{ij} = \begin{bmatrix} d_{x,ij} \\ d_{y,ij} \end{bmatrix} = \begin{bmatrix} s_{x,i} \\ s_{y,i} \end{bmatrix} - \begin{bmatrix} s_{x,j} \\ s_{x,j} \end{bmatrix} \tag{7.10}$$

and angle measurements q_{ijl} between three feature points $(s_{x,i}, s_{y,i})$, $(s_{x,j}, s_{y,j})$, and $(s_{x,l}, s_{y,l})$ are computed as follows:

$$q_{ijl} = \arccos\left(\frac{\left(\begin{bmatrix} s_{x,j} \\ s_{x,j} \end{bmatrix} - \begin{bmatrix} s_{x,i} \\ s_{y,i} \end{bmatrix} \right)\left(\begin{bmatrix} s_{x,j} \\ s_{x,j} \end{bmatrix} - \begin{bmatrix} s_{x,l} \\ s_{x,l} \end{bmatrix} \right)}{\sqrt{\left(s_{x,j} - s_{x,i}\right)^2 + \left(s_{y,j} - s_{y,i}\right)^2} \sqrt{\left(s_{x,j} - s_{x,l}\right)^2 + \left(s_{y,j} - s_{y,l}\right)^2}} \right) \tag{7.11}$$

A comparative analysis of the cumulative displacement of the wrist, elbow, and shoulder across successive frames during each repetition is performed using the following formula:

$$T_d = \sum_k \sum_a \left\| s^i_{la,k} - s^i_{la,k-1} \right\| - \sum_k \sum_a \left\| s^a_{ra,k} - s^a_{ra,k-1} \right\| \tag{7.12}$$

Herein, $s^i_{la,k-1}$ and $s^i_{ra,k-1}$ are 2D vectors representing the coordinates (s_x, s_y) of the left and right arms a in the previous frame $k-1$, while $s^i_{la,k}$ and $s^i_{ra,k}$ are 2D vectors for the current frame k. The evaluation of this outcome depends on the condition that if $T_d > 0$, it indicates that the left arm is performing the specified action; otherwise, the right arm is executing the action.

After performing these detailed calculations, the executed actions are classified into six fundamental physical exercises: stretching the left arm to the left side, raising the left arm upward, stretching the right arm to the right side, raising the right arm upward, simultaneously stretching both arms to their respective lateral sides, and simultaneously raising both arms upward. For each exercise, specific measurement constraints are defined to distinguish between complete, partially complete, and resting actions, as outlined in Table 7.2.

Table 7.2. Measurement constraints for each physical exercise include criteria for complete, partially complete, and resting actions.

Exercises	Complete Action Measurements	Partially Complete Action Measurements
Stretching left arm by lifting it towards the left side	$d_{15} \geq 55cm$, $d_{16} \geq 70cm$, $q_4 \geq 90^\circ$, $q_5 \geq 80^\circ$ $q_6 \geq 175^\circ$	$d_{15} < 55cm$, $d_{16} < 70$, $q_4 < 90^\circ$, $q_5 < 80^\circ$ $q_6 < 175^\circ$

Raising left arm upward direction	$d_{15} \geq 70cm$, $d_{16} \geq 95cm$, $q_4 \geq 90^\circ$, $q_5 \geq 125^\circ$, $q_6 \geq 175^\circ$	$d_{15} < 70cm$, $d_{16} < 95cm$, $q_4 < 90^\circ$, $q_5 < 125^\circ$, $q_6 < 175^\circ$
Stretching right arm by lifting it towards the right side	$d_{17} \geq 55cm$, $d_{18} \geq 70cm$, $q_1 \geq 90^\circ$, $q_2 \geq 80^\circ$, $q_3 \geq 175^\circ$	$d_{17} < 55cm$, $d_{18} < 70cm$, $q_1 < 90^\circ$, $q_2 < 80^\circ$, $q_3 < 175^\circ$
Raising right arm upward direction	$d_{17} \geq 70cm$, $d_{18} \geq 95cm$, $q_1 \geq 90^\circ$, $q_2 \geq 125^\circ$, $q_3 \geq 175^\circ$	$d_{17} < 70cm$, $d_{18} < 95cm$, $q_1 < 90^\circ$, $q_2 < 125^\circ$, $q_3 < 175^\circ$
Simultaneously stretching the left and right arms by lifting them towards their respective lateral sides	$d_{15}, d_{17} \geq 55cm$, $d_{16}, d_{18} \geq 70cm$, $q_1, q_4 \geq 90^\circ$, $q_2, q_5 \geq 80^\circ$, $q_3, q_6 \geq 175^\circ$	$d_{15}, d_{17} < 55cm$, $d_{16}, d_{18} < 70cm$, $q_1, q_4 < 90^\circ$, $q_2, q_5 < 80^\circ$, $q_3, q_6 < 175^\circ$
Simultaneously stretching both the left and right arms in an upward direction	$d_{15}, d_{17} \geq 70cm$, $d_{16}, d_{18} \geq 95cm$, $q_1, q_4 \geq 90^\circ$, $q_2, q_5 \geq 125^\circ$, $q_3, q_6 \geq 175^\circ$	$d_{15}, d_{17} < 70cm$, $d_{16}, d_{18} < 95cm$, $q_1, q_4 < 90^\circ$, $q_2, q_5 < 125^\circ$, $q_3, q_6 < 175^\circ$

Herein, the exercise routine begins from a seated position, where the left arm is stretched to the left side at a specified angle parallel to the shoulders. The arm is then raised upward and returned to the starting position. This sequence is repeated for the right arm. Figure 7.5 illustrates this exercise routine.

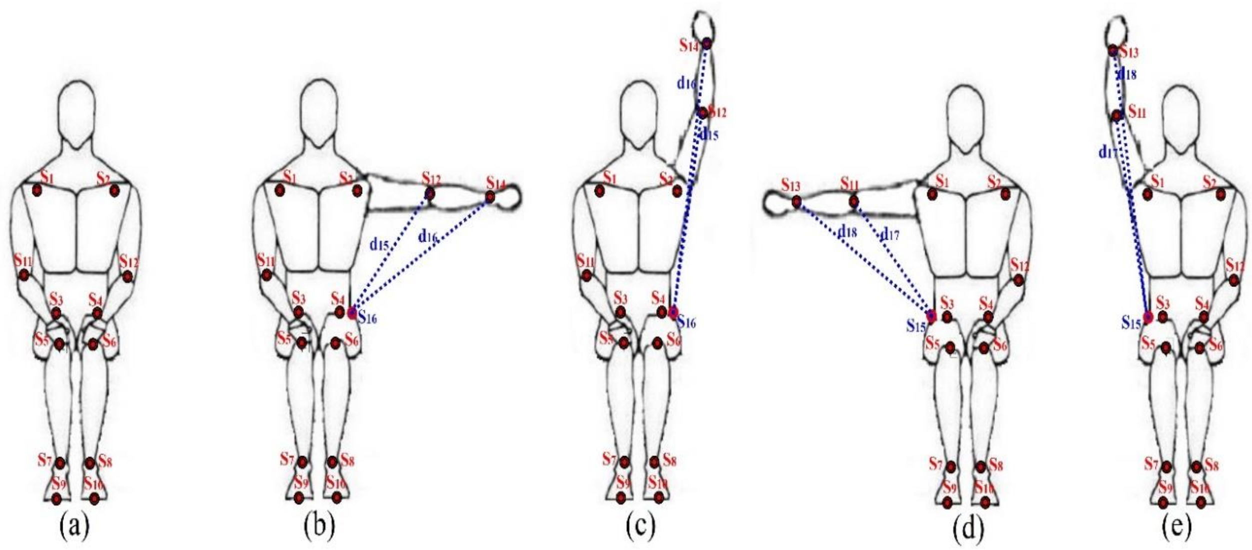


Figure 7.5. The first exercise routine: (a) Seated position, (b) Stretching left arm to the left, (c) Raising left arm upward, (d) Stretching right arm to the right, (e) Raising right arm upward.

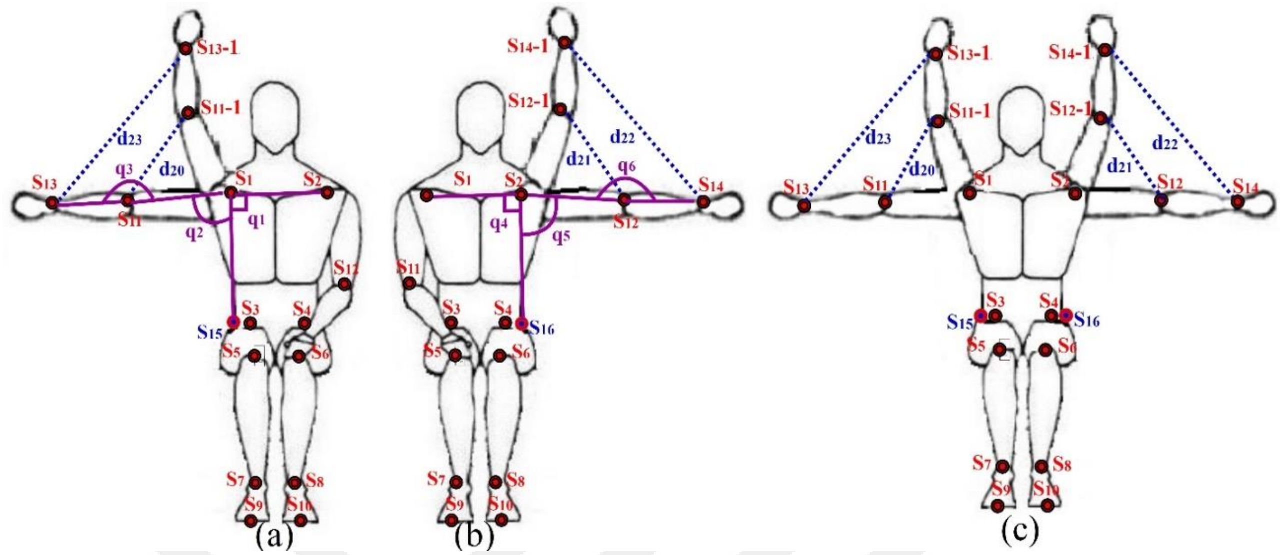


Figure 7.6. The visual depiction of alterations in distance and angular positions for (a) The left arm, (b) The right arm, and (c) Both arms.

Additionally, Figure 7.6 provides a visual depiction of the distance and angular changes throughout this exercise routine. The second exercise routine also begins from a seated position. Both the left and right arms are simultaneously stretched to their respective lateral sides. Next, both arms are raised simultaneously in an upward stretch. The routine concludes with another lateral stretch of both arms. An illustrative depiction of this routine is shown in the following figure.

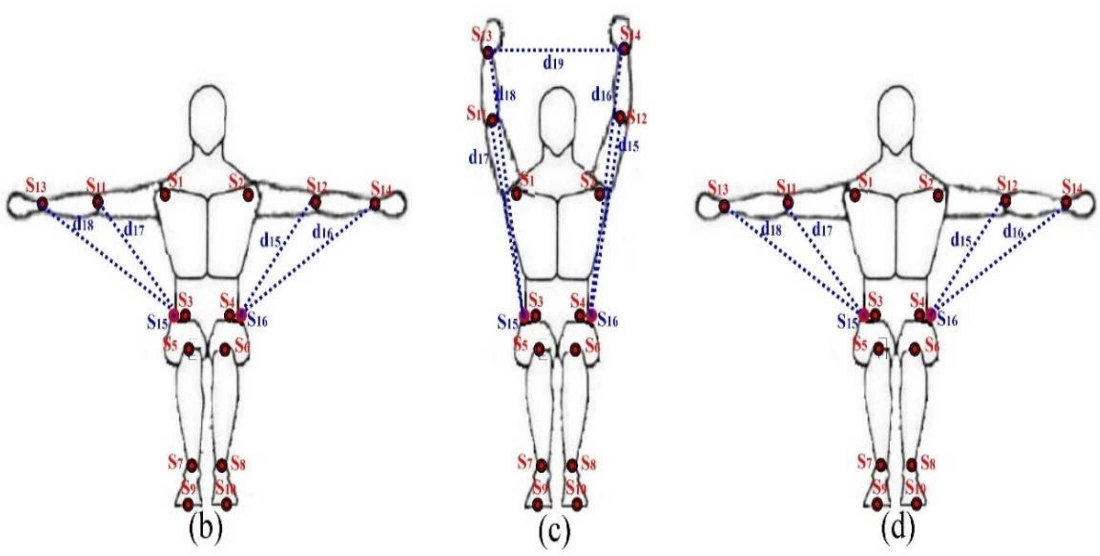


Figure 7.7. The second exercise routine: (a) Seated position, (b) Simultaneously stretching the left and right arms laterally, (c) Simultaneously stretching both arms upward, (d) Repeating the lateral stretch for both arms.

With the presentation of the latest exercise routine, this section is now complete. The final stage of the proposed algorithm will be thoroughly discussed in the subsequent section.

7.3. Human Action Learning with Genetic Algorithm-Based LSTM-RNN Structure

The final stage of the proposed algorithm involves optimizing physical exercise learning using an enhanced LSTM-RNN structure. This structure leverages a genetic algorithm to globally optimize the unknown parameters between neuron nodes, with a fitness function as the cost function, thereby improving classification performance.

7.3.1. The Overview of the LSTM-RNN Structure

The specialized design of LSTM-RNNs significantly improves their effectiveness in human action learning.

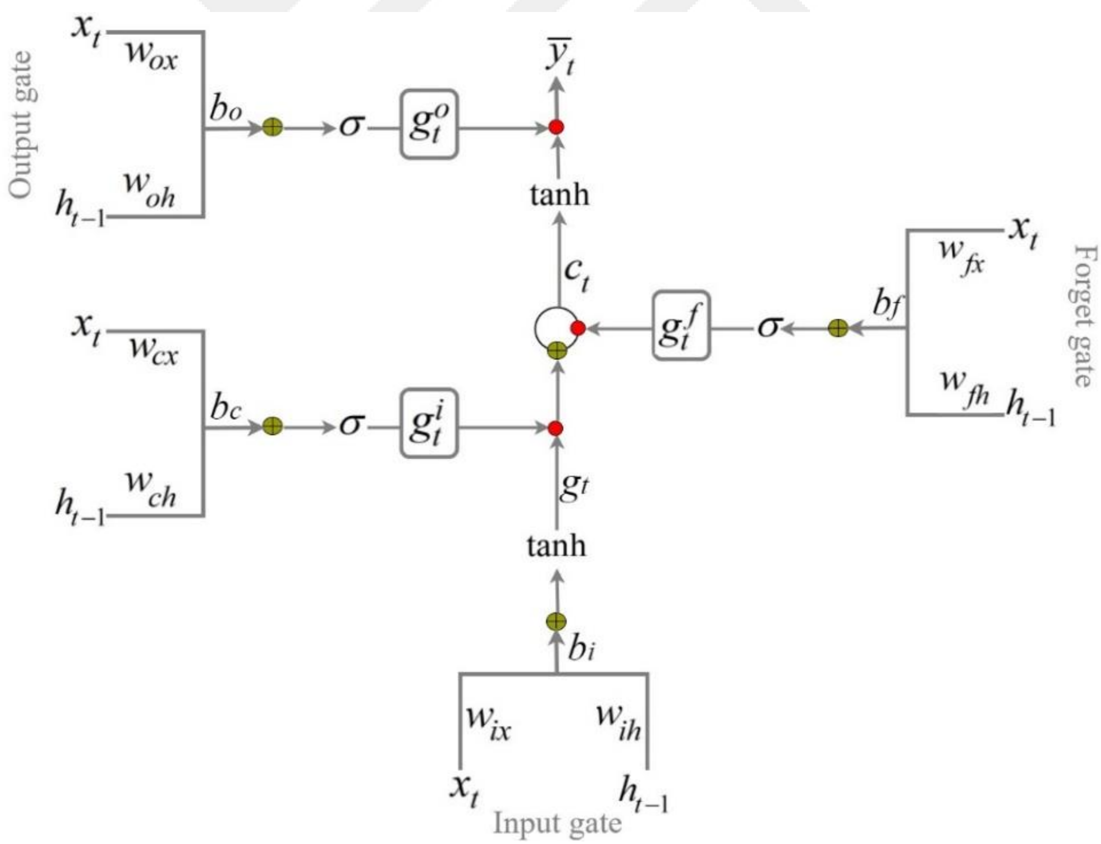


Figure 7.8. The LSTM- RNN structure.

As illustrated in Figure 7.8, the input layer of the LSTM-RNN unit receives the characterized human actions at each time step within the time frame, along with the corresponding unknown parameter vector.

$$x_t = [s_1 \quad \dots \quad s_{20} \quad d_1 \quad \dots \quad d_{23} \quad q_1 \quad \dots \quad q_6] \tag{7.13}$$

$$w_t = [w_{cx} \quad w_{ch} \quad w_{ix} \quad w_{ih} \quad w_{fx} \quad w_{fh} \quad w_{ox} \quad w_{oh}] \tag{7.14}$$

Each gate within the LSTM-RNN cell state processes the current feature input x_t, the previous hidden state h_{t-1}, and the state information c_{t-1} of the internal memory of the cell c_t. The final layer is a softmax output layer, which aids in learning human actions corresponding to the six defined states associated with physical exercises.

In each LSTM unit's computational procedure, the candidate memory cell value $\bar{c}_t$ is calculated using the following equation:

$$\bar{c}_t = \gamma(w_{cx}x_t + w_{ch}h_{t-1} + b_c) \tag{7.15}$$

Herein, γ represents the hyperbolic tangent activation function, b_c denotes the bias term for the cell, x_t is the current cell input, $h_{t-1}$is the output of the preceding cell, and w_{cx} and w_{ch} are the unknown parameters related to the input and hidden state of the cell. The input gate g_t^i then regulates the extent to which information in $\bar{c}_t$ is incorporated into the new memory cell c_t as follows:

$$g_t^i = \sigma(w_{ix}x_t + w_{ih}h_{t-1} + b_i) \tag{7.16}$$

where σ represents the sigmoid activation function, b_i denotes the bias term for the input gate, and w_{ix} and w_{ih} are the unknown parameters for the input and hidden state of the input gate. Subsequently, a forget gate $g_t^f$is introduced as:

$$g_t^f = \sigma\left(w_{fx}x_t + w_{fh}h_{t-1} + b_f\right) \tag{7.17}$$

In the given expression, b_f represents the bias term for the forget gate, and w_{fx} and w_{fh} are the unknown parameters for the input and hidden states of the forget gate. The information from h_{t-1} and x_t is processed by the sigmoid activation function, yielding values between 0 and 1. Values closer to 0 indicate that the corresponding information will be discarded, while values closer to 1 suggest that more information will be retained. The updated value of c_{t-1} is then integrated into the new cell state c_t as follows:

$$c_t = g_t^i\bar{c}_t + g_t^f c_{t-1} \tag{7.18}$$

In the subsequent phase, an output gate g_t^o is incorporated to specify the fraction of the memory cell c_t to be transmitted into the final hidden state h_t as:

$$g_t^o = \sigma(w_{ox}x_t + w_{oh}h_{t-1} + b_o) \tag{7.19}$$

and

$$h_t = g_t^o\gamma(c_t) \tag{7.20}$$

In the given equation, the variable b_o denotes the bias term for the output gate, while w_{ox} and w_{oh} denote the parameters of the output layer. Ultimately, the LSTM-RNN's computed outputs are inputted into a softmax layer ϕ to determine the predicted states for the six specified physical exercises across the class labels as:

$$\bar{y}_t = \phi\left(w_{hy}h_t + b_y\right) \tag{7.21}$$

Then, training error e_t is calculated as follows:

$$e_t = y_t - \bar{y}_t \tag{7.22}$$

Herein, y_t represents the actual output. Subsequently, the constrained cost function J_t to be minimized for optimal w_t is computed with minimum $w_{t\_\min}$ and maximum $w_{t\_\max}$ boundaries as:

$$J_t = \min(e_t e_t^T) \text{ subject to } w_{t\_\min} \le w_t \le w_{t\_\max} \tag{7.23}$$

The following section details the optimization of the LSTM-RNN learning model, as outlined in equation (7.21), through the application of a genetic algorithm.

7.3.2. Parameter Optimization with Genetic Algorithm

In this sub-section, the genetic algorithm is chosen to optimize the unknown parameters of LSTM-RNN networks for the human action learning task. This algorithm offers a systematic, evolution-inspired method for parameter refinement. The detailed pseudo-code is provided in Algorithm 7.1.

Algorithm 7.1. Learning model optimization with the genetic algorithm.

Input: w_t : Initialized unknown parameters, $w_{t\_\min}$: the lower limit of : w_t , $w_{t\_\max}$: the upper limit of w_t , J_t : Initialized cost function, $\overline{J}_t$: Best cost function, δ : Size of the population, β : Rate of the reproduction, α : Number of iterations, μ : Rate of crossover, γ : Mutation threshold, λ : Rate of the mutation.
Output: Optimized unknown parameters $w_t \in \Re^{1\times n}$
Initialization of the population
for $i = 1$ to δ

1. Generate a random population of w_t.

$w_t(i) = uniform\_random(w_{t\_\min}, w_{t\_\max}, n)$ (7.24)

2. Construct the input x_t in equation (7.13).
3. Calculate the gating units in equations (7.15)-(7.20).
4. Calculate estimated output $\bar{y}$ in equation (7.21).
5. Assess the constrained cost function $J_t(i)$ in equation (7.23).
6. Compare the obtained solution to the best cost function $\bar{J}_t$ as

if $\bar{J}_t > J_t$

update $\bar{w}_t = w_t(i)$ and $\bar{J}_t = J_t(i)$

endif

endfor

Main Loop:

for $j = 1$ to α

7. Obtain the average cost function as

$$J_{t\_avg} = J_t(j) / \sum_{k=1}^{j} J_t(k). \quad (7.25)$$

8. Calculate the reproduction probability as

$$p(j) = e^{-\beta J_t(j)}. \quad (7.26)$$

9. Perform roulette wheel selection-based crossover.

for $l = 1$ to $\delta/2$

Determine the first *parent* p_1 as

1. Generate random value as

$$r = rand \sum_{k=1}^{j} p(k). \quad (7.27)$$

2. Obtain cumulative probability as

$$c(k) = c(k-1) + p(k). \quad (7.28)$$

3. Determine the selected p_1 as

$$p_1(l) = \text{argmin}_k(r \le c(k)). \quad (7.29)$$

4. Repeat equations (7.26)-(7.28) for the second parent p_2.
5. Perform crossover for the first and second parents' unknown parameters w_{c_p1} and w_{c_p2} as

$$w_{c\_p1}(l) = \mu p_1(l) + (1-\mu) p_2(l) \quad (7.30)$$

$$w_{c\_p2}(l) = \mu p_2(l) + (1-\mu) p_1(l) \quad (7.31)$$

6. Save the results in w_{c_p}.

endfor

10. Perform the mutation process to w_{c_p}.

for $m = 1$ to δ

1. Apply Gaussian mutation to w_{c_p} parameters as

$$f = rand(\delta) < \gamma \quad (7.32)$$

$$gauss = randn(\delta) \tag{7.33}$$

$$w_{c\_p}(l) = w_{c\_p}(f+\lambda)gauss(f) \tag{7.34}$$

2. Apply the selected constraints on $w_{c\_p}(l)$ as

$$w_{c\_p} = \min(w_{c\_p}, w_{t\_\min}) \tag{7.35}$$

$$w_{c\_p} = \max(w_{c\_p}, w_{t\_\max}) \tag{7.36}$$

3. Repeat steps 1,2,3,4 and 5 respectively.

endfor

endfor

Alongside the provided pseudo-code, Table 7.3 details the parameters of the genetic algorithm.

Table 7.3. Parameters of genetic algorithm.

Parameters	Descriptions	Parameters	Descriptions
$J_t = \inf$	Initialized cost function	$\lambda = 0.3$	Rate of the mutation
$\beta = 1$	Rate of the reproduction	$\alpha = 500$	Number of iterations
$\gamma = 0.1$	Mutation threshold	$\delta = 100$	Size of the population

Following this, an initial population of diverse parameter configurations for the LSTM-RNN network is established. Each configuration's fitness is evaluated based on its predictive accuracy for human actions. Genetic operators—selection, crossover, and Gaussian mutation—are iteratively employed to evolve the population towards configurations with enhanced action estimation performance. This adaptive and evolutionary method mirrors natural selection and genetic inheritance principles, allowing the algorithm to identify parameter configurations that improve the network's capacity to understand and predict complex human body actions across various contexts.

After explaining the final stage of the proposed algorithm, the subsequent section provides a detailed account of its implementation on the NAO HR.

7.4. Implementation of The Proposed Algorithm to HRs

This section provides an overview of the utilized HRs and the implementation of the proposed physical exercise algorithm.

7.4.1. An Overview of Alpha 1Pro HR

In this study, the proposed algorithm has been implemented on the Alpha Mini and Alpha 1Pro HRs as depicted in Figure 7.9. The Alpha Mini HR is characterized by its compact, portable design, incorporating interactive features and voice capabilities. In contrast, the Alpha 1 Pro HR boasts a sophisticated design with 16 high-precision servo motors, allowing for a wide range of fluid and realistic movements essential for performing complex tasks and expressive gestures. It is also equipped with touch sensors and Bluetooth connectivity for enhanced interaction and external device communication.

Figure 7.9. Alpha 1Pro HR.

This research utilizes the UCode and AlphaRobot1s software to simulate physical exercises, enabling the Alpha Mini and Alpha 1 Pro robots to perform specific actions. Subsequently, OpenCV and Python are employed for further processing and analysis. The following section will focus on developing the dialogue model.

7.4.2. Dialogue Model for HRs

FSMs play a crucial role in developing dialogue systems for HRs, particularly in their interactions with geriatric patients. They offer a structured approach for outlining and defining different states and transitions within the dialogue model. Each state corresponds to a particular phase or condition of the interaction, and transitions govern how the HR moves from one state to another based on inputs or conditions.

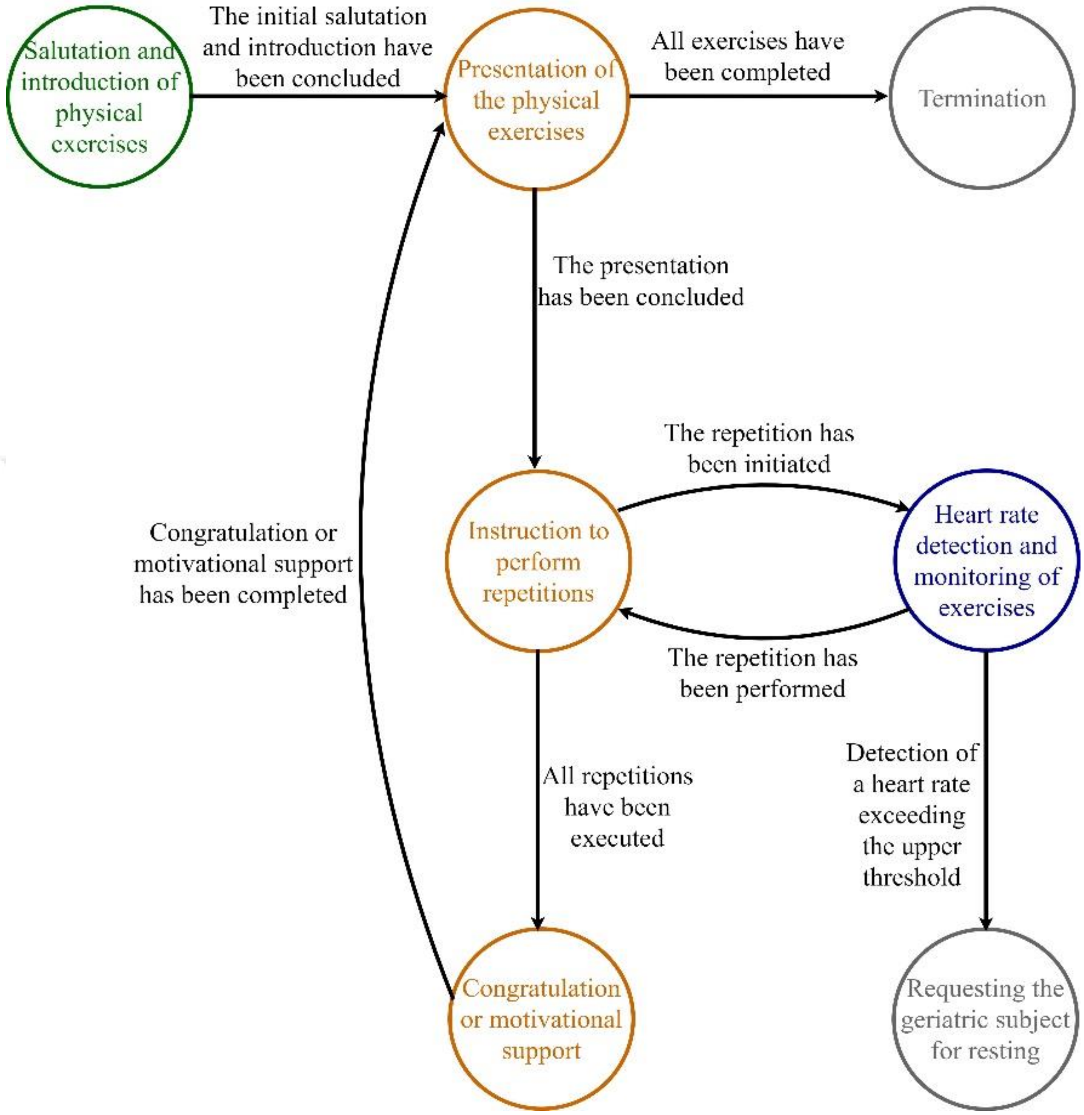
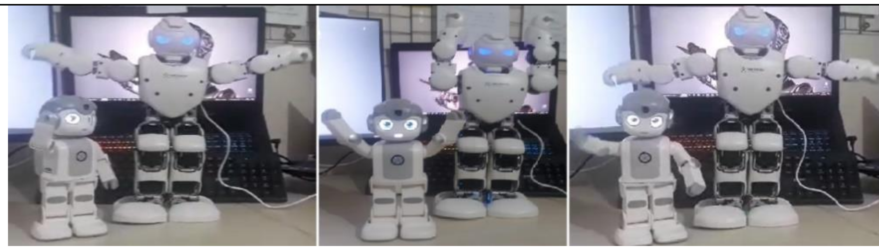


Figure 7.10. FSM dialogue model.

By employing FSMs, the dialogue model provides a systematic approach to HR behaviors, enabling effective and context-sensitive interactions with geriatric patients. This study assesses user states during physical exercise sessions to determine HR responses based on observed body actions, evaluating performance through user engagement as shown in Figure 7.10. HRs communicate through speech, body gestures, and changes in eye color. If no movement is detected, HRs prompt users to retry the exercises.

Table 7.4. HR behaviors during exercise sessions: Stage 1 is "Salutation," Stage 2 is "Introduction of the first exercise routine," Stage 3 is "Application of the second exercise routine," and Stage 4 is "Termination" of the session.

Stage	Non-Verbal Interaction
1	
	
2	
	
	

3	
4	

Additionally, if the user's heart rate exceeds the predefined upper threshold, the HR terminates the exercise session and advises the user to take a break. After each exercise, the HR provides congratulatory or encouraging feedback based on the user's performance. At the end of the physical exercise routine, the HR provides a final salutation to the users. The HR's total response time during exercise facilitation is approximately 25 seconds, covering the period needed to react to user states after each exercise.

Table 7.5. Verbal communication statements corresponding to each stage in Table 7.4.

Stage	Verbal Interaction
1	"Hello, my name is Alpha mini, and this is my friend Alpha 1Pro. Today, we will be guiding you through some exercises together. Throughout this process, if you feel tired, do not push yourself and take a break. So, are you ready to begin?"
2	"Initially, stretch your left arm by lifting it towards the left side in parallel to the shoulders as demonstrated by Alpha 1Pro. Following this, lift your left arm upward from the aforementioned position and then return to the initial position following the actions performed by Alpha 1Pro". "We will now perform these exercises five times. I have confidence in your ability to complete them! Let's begin. One, two, three, four and the final one! “Excellent job! We thoroughly enjoyed participating in these exercises with you." “Let's now repeat the same exercises for the right arm five times. You are doing great! Inform me if you experience fatigue. Otherwise, we will proceed to the next exercise routine”.

3	“Alright, now we are going to perform the second exercise routine. Are you ready? Follow the Alpha 1 Pro robot. Let's begin!”, “Now stretch your left and right arms by lifting them towards their respective lateral sides. Subsequently, raise both your left and right arms in an upward direction, and once more, simultaneously stretch your left and right arms by lifting them towards their respective lateral sides”. “Let's repeat these for five times!”
4	"You performed these exercises excellently! I wish you a great remainder of the day. Let's repeat this tomorrow. Goodbye for now!"

The successful completion of the dialogue model development thus leads to the evaluation phase.

7.5. Simulation and Experimental Results

This sub-section systematically evaluates the proposed physical exercise algorithm in both simulated and experimental settings. It begins with an introduction to the experimental setup and then provides a thorough analysis of the obtained results.

7.5.1. Simulation and Experimental Settings

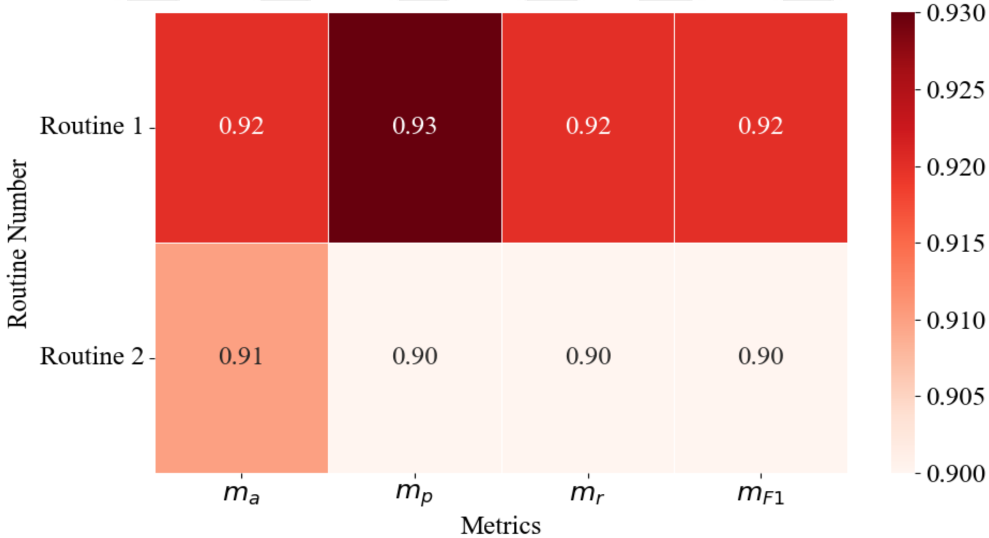
To evaluate the effectiveness of the proposed algorithm, a self-curated dataset is created. The assessment includes two distinct exercise routines, each lasting 5 minutes, with participation from three subjects across 30 sessions. The algorithm's performance is systematically analysed using metrics such as accuracy (m_a), precision (m_p), recall (m_r), and F1-Score (m_{F1}).

Following comprehensive performance analyses, the proposed algorithm is deployed on Alpha Mini and Alpha 1Pro HRs to validate its applicability and reliability in real-world settings. To evaluate subjects' experiences and preferences, the 5-point Likert scale (Likert, 1932) is used. This scale provides a structured method for quantifying subjective feedback on aspects such as enjoyment, perceived effectiveness, clarity of instructions, perceived difficulty, and overall satisfaction. This quantitative approach allows for a thorough assessment of HRs' acceptability and effectiveness in geriatric exercise routines. The insights gained from these evaluations inform future developments and improvements, aiming to better meet the specific needs and preferences of older adults.

7.5.2. Results

Following the description of simulation and experimental setups, the initial evaluation of the proposed algorithm's performance is based on m_a, m_p, m_r, and m_{F1} metrics, as detailed in Table 7.6. This table offers a comprehensive overview of the algorithm's notable learning outcomes. Analysis across two distinct exercise routines demonstrates significant results. In the first routine, the algorithm achieves an m_a of 0.92, indicating a high rate of correct predictions. The m_p of 0.93 suggests a low rate of false positives, while the m_r of 0.92 reflects its effectiveness in identifying true positives. The m_{F1} of 0.92 indicates a balanced performance between $m_p$and m_r. In the second routine, the algorithm maintains robust performance with an m_a of 0.91, $m_p$of 0.90, $m_r$of 0.90, and m_{F1} of 0.90. These findings highlight the algorithm's effectiveness in accurately categorizing exercise-related activities and its potential for supporting diverse physical exercise routines.

Table 7.6. The learning outcomes of the proposed algorithm on the generated dataset.



In addition to the experiments, the proposed physical exercise algorithm is implemented on Alpha Mini and Alpha 1Pro HRs to initiate a social-interactive dialogue with the subject, as shown in Figure 7.11. In the initial interaction phase, the Alpha Mini HR provides detailed instructions to the subject, who then begins the first exercise routine in collaboration with the Alpha 1Pro HR. Throughout each phase, the Alpha Mini HR supports and motivates the subject to improve performance. The subject successfully completes the exercise routine with a notable success rate of 91%.

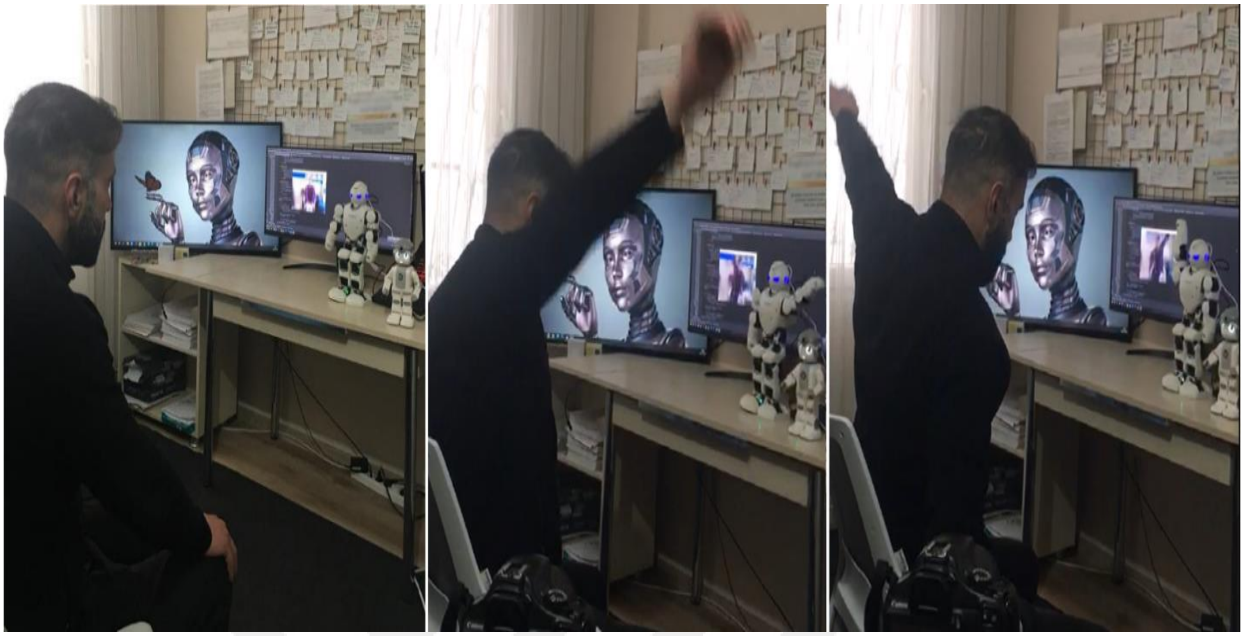


Figure 7.11. Application of the first physical exercise routine with Alpha Mini and 1 Pro HRs.

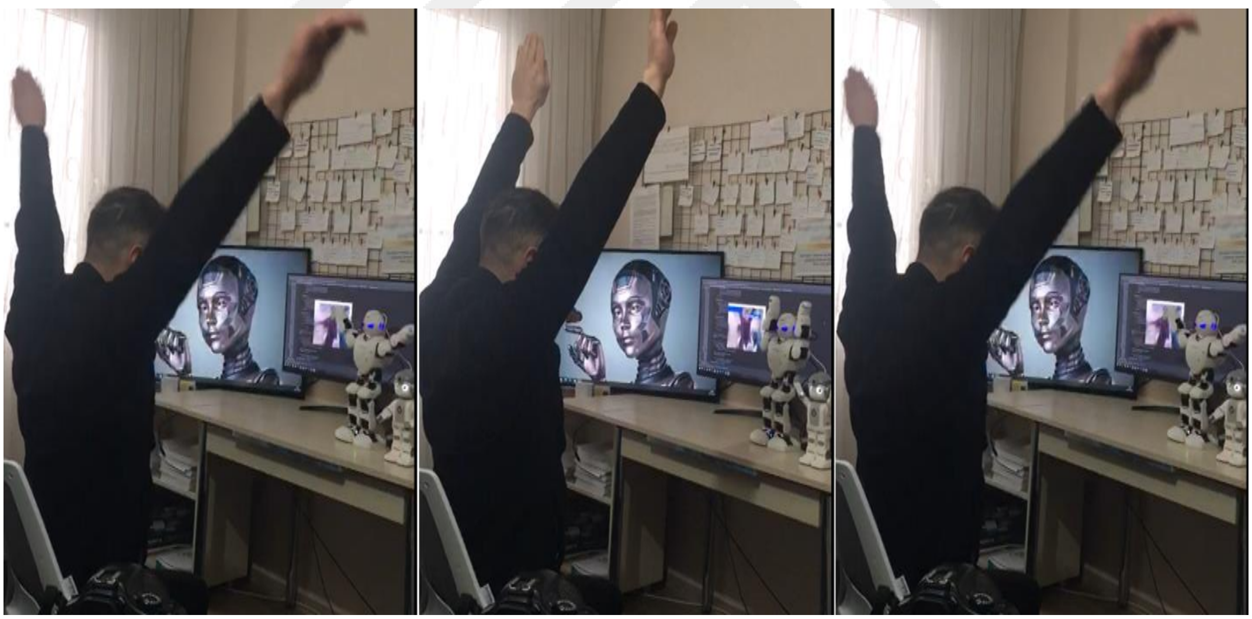
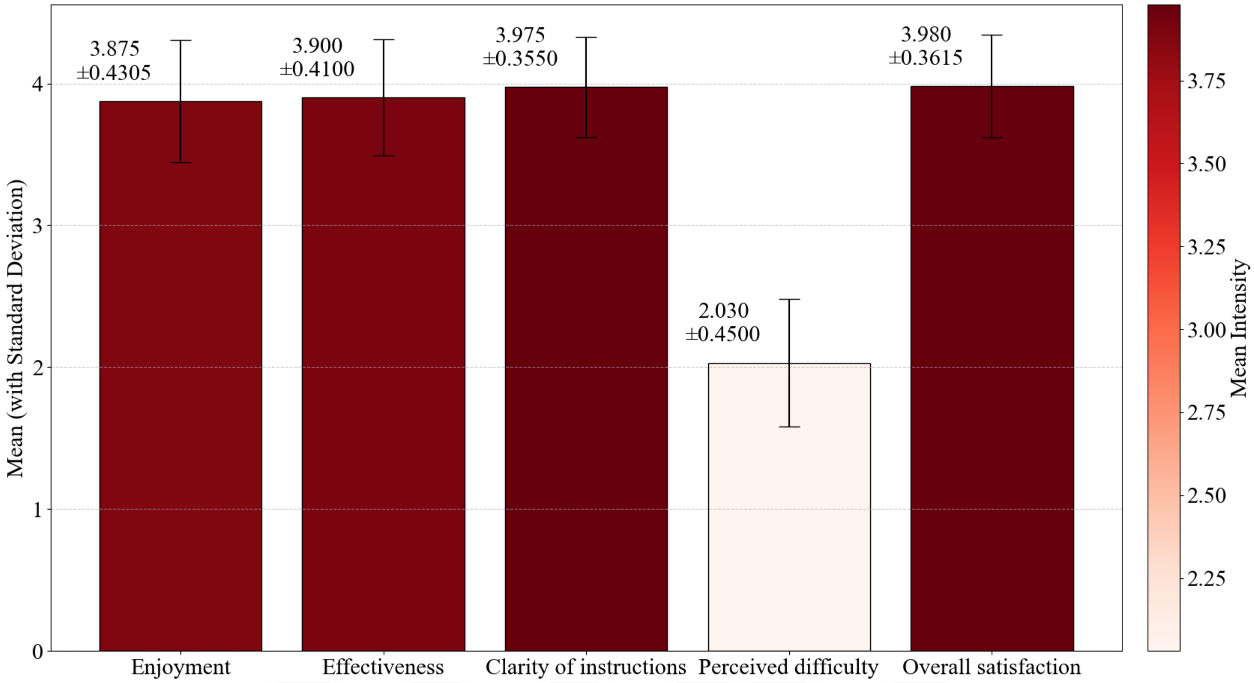


Figure 7.12. Application of the second physical exercise routine with Alpha Mini and 1 Pro HRs.

Furthermore, as illustrated in Figure 7.12, the HRs effectively encouraged the subject to participate in the exercises during the second routine. The subject achieved a precise success rate of 90% for the exercises. In the final analysis, the 5-point Likert scale responses from the subjects indicate a notably positive inclination toward the prescribed physical exercise routines, as shown in Table 7.7.

Table 7.7. Findings from participant assessments employing a 5-point Likert scale.



Herein, subjects reported an average enjoyment score of 3.875, reflecting substantial gratification with the exercise sessions. The perceived effectiveness received a mean score of 3.90, indicating strong confidence in the routine's ability to produce positive outcomes. Clarity of instructions achieved a high mean score of 3.975, suggesting that the instructions were clear and easily understood. The perceived difficulty had a mean score of 2.030, representing a manageable level of challenge. Overall satisfaction was notably high at 3.980, demonstrating strong overall contentment with the exercise routine. The low standard deviations across these metrics further highlight a consistent and positive response from the participants.

7.6. Discussions

In this last section of the thesis, a four-stage physical exercise algorithm is presented and implemented to HRs. The initial stage of the proposed algorithm utilizes the Viola-Jones algorithm to achieve accurate detection and continuous tracking of the human body within the perceptual framework of HR. Subsequently, geometric-based distance measurements are employed to characterize human actions, enabling the algorithm to comprehend physical movements relevant to targeted exercises for geriatric individuals. In the third stage, the characterized human actions are learned through an LSTM-RNN in conjunction with a genetic algorithm. The final stage involves the execution of a structured and user-friendly engagement utilizing an FSM dialogue model. While this algorithm is explicitly implemented on Alpha Mini and Alpha 1Pro HRs within the context of this research, its applicability extends to encompass all other HR models. The proposed algorithm's reliability is affirmed through a detailed analysis of decision-making at each stage and an evaluation of exercise routines based on learned actions. An in-depth assessment of its computational demands and accuracy further demonstrates its effectiveness. Future research will extend the algorithm's application to larger cohorts, particularly focusing on elderly individuals in various living conditions, and will broaden the range of exercise routines to enhance its applicability and impact.

8. CONCLUSION

This thesis advances a series of sophisticated algorithms designed to empower HRs to autonomously recognize critical health parameters, conduct comprehensive assessments, and implement dialog-based interventions, thereby positioning them as personal health-care providers in individuals' daily routines. The initial section offers an extensive literature review, exploring the latest advancements in assistive robotics and their health-care implications. The subsequent sections provide a detailed exploration of the design and implementation of the heart rate monitoring algorithm, followed by an in-depth analysis of the personalized temperature estimation algorithm. In response to the psychological impacts of social isolation and quarantine induced by the COVID-19 pandemic, the following section presents a speech-based emotion analysis algorithm specifically designed for patients in quarantine. This is followed by the development of algorithms for regulating facial mask-wearing, aimed at mitigating the effects of the pandemic. Finally, the thesis culminates in a case study that examines the application of HRs in facilitating physical exercise for geriatric patients. While the developed algorithms are expressly implemented on NAO, Alpha Mini and Alpha 1Pro HRs within the scope of this thesis, but its applicability extends to encompass all other HRs. The simulation and experimental results of each implemented algorithm highlight the significant role of assistive HRs in health-care, emphasizing their positive impact and effectiveness in advancing health interventions.

REFERENCES

Abubakar, S., Das, S. K., Robinson, C., Saadatzi, M. N., Cynthia Logsdon, M., Mitchell, H., Chlebowy, D., & Popa, D. O. (2020). ARNA, a Service robot for Nursing Assistance: System Overview and User Acceptability. *IEEE International Conference on Automation Science and Engineering*, 1408–1414.

Akalin, N., & Loutfi, A. (2021). Reinforcement learning approaches in social robotics. Sensors, 21(4), 1292.

Al-Taee, M. A., Kapoor, R., Garrett, C., & Choudhary, P. (2016). Acceptability of Robot Assistant in Management of Type 1 Diabetes in Children. *Diabetes Technology and Therapeutics*, *18*(9), 551–554.

Alemi, M., Ghanbarzadeh, A., Meghdari, A., & Moghadam, L. J. (2015). Clinical Application of a Humanoid Robot in Pediatric Cancer Interventions. *International Journal of Social Robotics*, *8*(5), 743–759.

Balasubramanian, S., Wei, H. R., Perez, M., Shepard, B., Koeneman, E., Koeneman, J., & He, J. (2008). Rupert: An exoskeleton robot for assisting rehabilitation of arm functions. *Virtual Rehabilitation*, 163–167.

Beer, J. M., Fisk, A. D., & Rogers, W. A. (2014). Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction. *Journal of Human-Robot Interaction*, *3*(2), 74.

Bell, T. E. (1985). Robots in the home: Promises. *IEEE Spectrum*, *22*(5), 51–55.

Bemelmans, R., Gelderblom, G. J., Jonker, P., & De Witte, L. (2012). Socially assistive robots in elderly care: a systematic review into effects and effectiveness. *Journal of the American Medical Directors Association*, *13*(2), 114-120.

Bemelmans, R., Gelderblom, G. J., Jonker, P., & de Witte, L. (2015). Effectiveness of Robot Paro in Intramural Psychogeriatric Care: A Multicenter Quasi-Experimental Study. *Journal of the American Medical Directors Association*, *16*(11), 946–950.

Beraldo, G., Menegatti, E., De Tommasi, V., Mancin, R., & Benini, F. (2019). A preliminary investigation of using humanoid social robots as non-pharmacological techniques with children. *IEEE Workshop on Advanced Robotics and Its Social Impacts*, 393–400.

Beran, T. N., Ramirez-Serrano, A., Vanderkooi, O. G., & Kuhn, S. (2013). Reducing children's pain and distress towards flu vaccinations: A novel and effective application of humanoid robotics. *Vaccine*, *31*(25), 2772–2777.

Bezerra, C. A. D., & Zampieri, D. E. (2004). Biped Robots: The State of Art. *International Symposium on History of Machines and Mechanisms*, 371–389.

Bhaumik, A. (2018). From AI to Robotics : Mobile, Social, and Sentient Robots. *From AI to Robotics*.

Capek, K. (1971). RUR (Robots Universales Rossum). Escuelas Profesionales Sagrado Corazón.

Carberry, J., Hinchly, G., Buckerfield, J., Tayler, E., Burton, T., Madgwick, S., & Vaidyanathan, R. (2011). Parametric design of an active ankle foot orthosis with passive compliance. *IEEE Symposium on Computer-Based Medical Systems*.

Card, S. K., Moran, T. P., & Newell, A. (1983). The Psychology of Human-Computer Interaction. CRC Press.

Chalvatzaki, G., Papageorgiou, X. S., & Tzafestas, C. S. (2017). Towards a user-adaptive context-aware robotic walker with a pathological gait assessment system: First experimental study. *IEEE*

International Conference on Intelligent Robots and Systems, 5037–5042.
Chen, Y., Hu, J., Peng, L., & Hou, Z. (2014). The FES-assisted control for a lower limb rehabilitation robot: simulation and experiment. *Robotics and Biomimetics*, *1*(1), 1–20.
Cheng, L., Chen, M., & Li, Z. (2018). Design and Control of a Wearable Hand Rehabilitation Robot. *IEEE Access*, *6*, 74039–74050.
Christoforou, E. G., Avgousti, S., Ramdani, N., Novales, C., & Panayides, A. S. (2020). The upcoming role for nursing and assistive robotics: Opportunities and challenges ahead. *Frontiers in Digital Health*, *2*, 585656.
Clarke, R. (2020). Asimov's laws of robotics: Implications for information technology: Part 1. *Machine Ethics and Robot Ethics*, 33–41.
Colombo, G., Joerg, M., Schreier, R., & Dietz, V. (2000). Treadmill training of paraplegic patients using a robotic orthosis. *Journal of Rehabilitation Research and Development*, *37*(6), 693–700.
Costa, N., Bezdicek, M., Brown, M., Gray, J. O., Caldwell, D. G., & Hutchins, S. (2006). Joint motion control of a powered lower limb orthosis for rehabilitation. *International Journal of Automation and Computing*, *3*(3), 271–281.
Csala, E., Németh, G., & Zainko, C. (2012). Application of the NAO humanoid robot in the treatment of marrow-transplanted children. *3rd IEEE International Conference on Cognitive Infocommunications* , 655–659.
Breazeal, C. (2002). Designing sociable robots. MIT press.
Newton, D. E. (2018). Robots. Bloomsbury Publishing.
Di Napoli, C., Ercolano, G., & Rossi, S. (2023). Personalized home-care support for the elderly: a field experience with a social robot at home. User Modeling and User-Adapted Interaction, 33(2), 405-440.
Ding, J., Lim, Y. J., Solano, M., Shadle, K., Park, C., Lin, C., & Hu, J. (2014). Giving patients a lift - The robotic nursing assistant (rona). *IEEE Conference on Technologies for Practical Robot Applications*.
Ding, M., Ikeura, R., Mukai, T., Nagashima, H., Hirano, S., Matsuo, K., Sun, M., Jiang, C., & Hosoe, S. (2012). Comfort estimation during lift-up using nursing-care robot - RIBA. *1st International Conference on Innovative Engineering Systems*, 225–230.
Ding, Y., & Tay, E. H. (2019). An interactive training system for upper limb rehabilitation using visual and auditory feedback. *PervasiveHealth: Pervasive Computing Technologies for Healthcare*, 54–58.
Dong, M., Yuan, J., & Li, J. (2022). A lower limb rehabilitation robot with rigid-flexible characteristics and multi-mode exercises. Machines, 10(10), 918.
Engelberger, J. F. (1988). Health-care robotics goes commercial: the 'HelpMate' experience. *Robotica*, *11*(6), 517–523.
Feil-Seifer, D., & Matarić, M. J. (2011). Socially assistive robotics. IEEE Robotics & Automation Magazine, 18(1), 24-31.
Feng, G., Zhang, J., Zuo, G., Li, M., Jiang, D., & Yang, L. (2022). Dual-modal hybrid control for an upper-limb rehabilitation robot. Machines, 10(5), 324.
Figueiredo, J., Félix, P., Santos, C. P., & Moreno, J. C. (2017). Towards human-knee orthosis interaction based on adaptive impedance control through stiffness adjustment. *IEEE International Conference on Rehabilitation Robotics*, *2017*, 406–411.
Firescu, V., Gaşpar, M. L., Crucianu, I., & Rotariu, E. (2022). Collaboration Between Humans and Robots in Organizations: A Macroergonomic, Emotional, and Spiritual Approach. Frontiers in Psychology,

13, 855768.
Fischinger, D., Einramhof, P., Papoutsakis, K., Wohlkinger, W., Mayer, P., Panek, P., Hofmann, S., Koertner, T., Weiss, A., Argyros, A., & Vincze, M. (2016). Hobbit, a care robot supporting independent living at home: First prototype and lessons learned. *Robotics and Autonomous Systems*, *75*, 60–78.
Fong, T. (2001). Collaborative Control : A Robot-Centric Model for Vehicle Teleoperation. *Aaai*, *Mishkin*, 198.
Fujita, M. (2001). AIBO: Toward the era of digital creatures. The International Journal of Robotics Research, 20(10), 781-794.
Gadde, P., Kharrazi, H., Patel, H., & MacDorman, K. F. (2011). Toward Monitoring and Increasing Exercise Adherence in Older Adults by Robotic Intervention: A Proof of Concept Study. *Journal of Robotics*.
Gamborino, E., Yueh, H. P., Lin, W., Yeh, S. L., & Fu, L. C. (2019). Mood estimation as a social profile predictor in an autonomous, multi-session, emotional support robot for children. In 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (pp. 1-6).
Gerłowska, J., Skrobas, U., Grabowska-Aleksandrowicz, K., Korchut, A., Szklener, S., Szczęśniak-Stańczyk, D., Tzovaras, D., & Rejdak, K. (2018). Assessment of perceived attractiveness, usability, and societal impact of a multimodal robotic assistant for aging patients with memory impairments. Frontiers in neurology, 9, 392.
Ghafurian, M., Hoey, J., & Dautenhahn, K. (2021). Social robots for the care of persons with dementia: a systematic review. *ACM Transactions on Human-Robot Interaction, 10*(4), 1-31.
Gombolay, M., Yang, X. J., Hayes, B., Seo, N., Liu, Z., Wadhwania, S. & Shah, J. (2018). Robotic assistance in the coordination of patient care. The International Journal of Robotics Research, 37(10), 1300-1316.
Gongor, F., & Tutsoy, O. (2022). Doctor Robots: Design and Implementation of a Heart Rate Estimation Algorithm. *International Journal of Social Robotics*, *14*(6), 1435–1461.
Goodrich, M. A., & Schultz, A. C. (2008). Human–robot interaction: a survey. Foundations and Trends® in Human–Computer Interaction, 1(3), 203-275.
Gonzalez, A., Garcia, L., Kilby, J., & McNair, P. (2021). Robotic devices for paediatric rehabilitation: a review of design features. *BioMedical Engineering OnLine*, *20*(1), 1-33.
Gordleeva, S. Y., Lobov, S. A., Grigorev, N. A., Savosenkov, A. O., Shamshin, M. O., Lukoyanov, M. V., Khoruzhko, M. A., & Kazantsev, V. B. (2020). Real-Time EEG-EMG human-machine interface-based control system for a lower-limb exoskeleton. *IEEE Access*, *8*, 84070–84081.
Görer, B., Salah, A. A., & Akın, H. L. (2017). An autonomous robotic exercise tutor for elderly people. *Autonomous Robots*, *41*(3), 657–678.
Granata, C., Pino, M., Legouverneur, G., Vidal, J. S., Bidaud, P., & Rigaud, A. S. (2013). Robot services for elderly with cognitive impairment: Testing usability of graphical user interfaces. *Technology and Health Care*, *21*(3), 217–231.
Granda, T. M., Kirkpatrick, M., Julien, T. D., & Peterson, L. A. (1990). Evolutionary role of humans in the human-robot system. *Proceedings of the Human Factors Society*, 664–668.
Graves, R. H. (2013). The triumph of an idea. The story of Henry Ford. Edizioni Savine.
Greczek, J., & Mataric, M. (2015). Toward personalized pain anxiety reduction for children. *AAAI Fall Symposium - Technical Report*, *FS-15-01*, 74–76.
Han, J. I., Lee, J. H., Choi, H. S., Kim, J. H., & Choi, J. (2022). Policy Design for an Ankle-Foot Orthosis

Using Simulated Physical Human–Robot Interaction via Deep Reinforcement Learning. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 30, 2186-2197.

Han, D., Mulyana, B., Stankovic, V., & Cheng, S. (2023). A Survey on Deep Reinforcement Learning Algorithms for Robotic Manipulation. Sensors, 23(7), 3762.

HansonRobotics. (2021). *Meet Grace, the health care robot COVID-19 created*. https://nypost.com/2021/06/09/meet-grace-the-healthcare-robot-covid-19-created/

Helm, M., Carros, F., Schädler, J., & Wulf, V. (2022). Zoomorphic robots and people with disabilities. Proceedings of Mensch und Computer, 431-436.

Henkemans, O. A. B., Bierman, B. P. B., Janssen, J., Looije, R., Neerincx, M. A., van Dooren, M. M. M., de Vries, J. L. E., van der Burg, G. J., & Huisman, S. D. (2017). Design and evaluation of a personal robot playing a self-management education game with children with diabetes type 1. *International Journal of Human Computer Studies*, *106*, 63–76.

Hillis, D., McCarthy, J., Mitchell, T. M., Mueller, E. T., Riecken, D., Sloman, A., & Winston, P. H. (2007). In honor of Marvin Minsky's contributions on his 80th birthday. *AI Magazine*, *28*(4), 103–110.

Housman, S. J., Le, V., Rahman, T., Sanchez, R. J., & Remkensrneyer, D. J. (2007). Arm-training with T-WREX after chronic stroke: Preliminary results of a randomized controlled trial. *IEEE 10th International Conference on Rehabilitation Robotics*, 562–568.

Irfan, B., Ramachandran, A., Spaulding, S., Glas, D. F., Leite, I., & Koay, K. L. (2019). Personalization in long-term human-robot interaction. *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 685-686).

Iwata, H., & Sugano, S. (2009). Design of human symbiotic robot TWENDY-ONE. *IEEE International Conference on Robotics and Automation*, 580–586.

Jamwal, P. K., Hussain, S., Ghayesh, M. H., & Rogozina, S. V. (2016). Impedance Control of an Intrinsically Compliant Parallel Ankle Rehabilitation Robot. *IEEE Transactions on Industrial Electronics*, *63*(6), 3638–3647.

Jeong, S., Logan, D. E., Goodwin, M. S., Graca, S., O'Connell, B., Goodenough, H., Anderson, L., Stenquist, N., Fitzpatrick, K., Zisook, M., Plummer, L., Breazeal, C., & Weinstock, P. (2015). A Social Robot to Mitigate Stress, Anxiety, and Pain in Hospital Pediatric Care. *10th Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, 103–104.

Jibb, L. A., Birnie, K. A., Nathan, P. C., Beran, T. N., Hum, V., Victor, J. C., & Stinson, J. N. (2018). Using the MEDiPORT humanoid robot to reduce procedural pain and distress in children with cancer: A pilot randomized controlled trial. Pediatric blood & cancer, 65(9), e27242.

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., Van Riemsdijk, M. B., & Sierhuis, M. (2014). Coactive Design: Designing Support for Interdependence in Joint Activity. *Journal of Human-Robot Interaction*, *3*(1), 43.

Kachouie, R., Sedighadeli, S., Khosla, R., & Chu, M. T. (2014). Socially assistive robots in elderly care: a mixed-method systematic literature review. *International Journal of Human-Computer Interaction*, *30*(5), 369-393.

Kahn, L. E., Zygman, M. L., Rymer, W. Z., & Reinkensmeyer, D. J. (2006). Robot-assisted reaching exercise promotes arm movement recovery in chronic hemiparetic stroke: A randomized controlled pilot study. *Journal of NeuroEngineering and Rehabilitation*, *3*, 12.

Kanal, V., Brady, J., Nambiappan, H., Kyrarini, M., Wylie, G., & Makedon, F. (2020). Towards a serious game based human-robot framework for fatigue assessment. *13th ACM International Conference on*

PErvasive Technologies Related to Assistive Environments.

Kang, K. Il, Freedman, S., Matarić, M. J., Cunningham, M. J., & Lopez, B. (2005). A hands-off physical therapy assistance robot for cardiac patients. *IEEE 9th International Conference on Rehabilitation Robotics*, 337–340.

Khan, M. N., Altalbe, A., Naseer, F., & Awais, Q. (2024). Telehealth-Enabled In-Home Elbow Rehabilitation for Brachial Plexus Injuries Using Deep-Reinforcement-Learning-Assisted Telepresence Robots. Sensors, 24(4), 1273.

Kyrarini, M., Lygerakis, F., Rajavenkatanarayanan, A., Sevastopoulos, C., Nambiappan, H. R., Chaitanya, K. K. & Makedon, F. (2021). A survey of robots in healthcare. *Technologies*, *9*(1), 8.

Kerstin Dautenhahn. (2004). Socially Intelligent Agents in Human Primate Culture. *Uniwersytet Śląski*, *7*(1), 45–72.

Kiguchi, K., & Hayashi, Y. (2012). An EMG-based control for an upper-limb power-assist exoskeleton robot. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *42*(4), 1064–1071.

Kim, S. (2022). Working with robots: human resource development considerations in human–robot interaction. Human Resource Development Review, 21(1), 48-74.

King, C. H., Chen, T. L., Jain, A., & Kemp, C. C. (2010). Towards an assistive robot that autonomously performs bed baths for patient hygiene. *IEEE/RSJ International Conference on Intelligent Robots and Systems- Conference Proceedings*, 319–324.

Klamroth-Marganska, V., Blanco, J., Campen, K., Curt, A., Dietz, V., Ettlin, T., Felder, M., Fellinghauer, B., Guidali, M., Kollmar, A., Luft, A., Nef, T., Schuster-Amft, C., Stahel, W., & Riener, R. (2014). Three-dimensional, task-specific robot therapy of the arm after stroke: A multicentre, parallel-group randomised trial. *The Lancet Neurology*, *13*(2), 159–166.

Krebs, H. I., Ferraro, M., Buerger, S. P., Newbery, M. J., Makiyama, A., Sandmann, M., Lynch, D., Volpe, B. T., & Hogan, N. (2004). Rehabilitation robotics: Pilot trial of a spatial extension for MIT-Manus. *Journal of NeuroEngineering and Rehabilitation*, *1*(1), 1–15.

Kulpa, E., Rahman, A. T., & Vahia, I. V. (2021). Approaches to assessing the impact of robotics in geriatric mental health care: a scoping review. *International Review of Psychiatry*, *33*(4), 424-434.

Kyriakidou, M., Padda, K., & Parry, L. (2017). Reporting robot ethics for children-robot studies in contemporary peer reviewed papers. In A World with Robots: International Conference on Robot Ethics, (pp. 109-117). Springer International Publishing.

Li, S., Huang, S., Huang, L., Shen, H., Liu, Y., & Xie, L. (2023). Design and optimization of a body weight support system for lower-limb rehabilitation robots considering vibration characteristics. Structural and Multidisciplinary Optimization, 66(12), 1-21.

Li, X., Yang, Q., & Song, R. (2020). Performance-Based Hybrid Control of a Cable-Driven Upper-Limb Rehabilitation Robot. *IEEE Transactions on Biomedical Engineering*, *68*(4), 1351–1359.

Li, Z., Huang, Z., He, W., & Su, C. Y. (2016). Adaptive impedance control for an upper limb robotic exoskeleton using biological signals. *IEEE Transactions on Industrial Electronics*, *64*(2), 1664–1674.

Littler, B. K. M., Alessa, T., Dimitri, P., Smith, C., & de Witte, L. (2021). Reducing negative emotions in children using social robots: systematic review. *Archives of Disease in Childhood*, *106*(11), 1095-1101.

Looije, R., Neerincx, M. A., & Cnossen, F. (2010). Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of*

Human Computer Studies, *68*(6), 386–397.

Looije, R., Neerincx, M. A., Peters, J. K., & Henkemans, O. A. B. (2016). Integrating Robot Support Functions into Varied Activities at Returning Hospital Visits: Supporting Child's Self-Management of Diabetes. *International Journal of Social Robotics*, *8*(4), 483–497.

Lu, S. C., Blackwell, N., & Do, E. Y. L. (2011). mediRobbi: An interactive companion for pediatric patients during hospital visit. *Lecture Notes in Computer Science*, *6762 LNCS*(PART 2), 547–556.

Lum, P. S., Lehman, S. L., & Reinkensmeyer, D. J. (1995). The Bimanual Lifting Rehabilitator: An Adaptive Machine for Therapy of Stroke Patients. *IEEE Transactions on Rehabilitation Engineering*, *3*(2), 166–174.

Luperto, M., Monroy, J., Renoux, J., Lunardini, F., Basilico, N., Bulgheroni, M., & Borghese, N. A. (2022). Integrating social assistive robots, IoT, virtual communities and smart objects to assist at-home independently living elders: the MoveCare project. International Journal of Social Robotics, 15(3), 517-545.

Mamun, K., Islam, F. M., Sharma, A., Hoque, A. S. M., & Szecsi, T. (2016). Patient condition monitoring modular hospital robot. Journal of Software, 11(8), 768-786.

Martinez-Martin, E., Costa, A., & Cazorla, M. (2019). PHAROS 2.0—A PHysical Assistant RObot System Improved. *Sensors, Vol. 19, Page 4531*, *19*(20), 4531.

McColl, D., & Nejat, G. (2013). Meal-Time with a Socially Assistive Robot and Older Adults at a Long-term Care Facility. *Journal of Human-Robot Interaction*, *2*(1), 152–171.

Miyake, T., Wang, Y., Yan, G., & Sugano, S. (2022). Skeleton recognition-based motion generation and user emotion evaluation with in-home rehabilitation assistive humanoid robot. In IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids) (pp. 616-621).

Miyoshi, T., Hiramatsu, K., Yamamoto, S. I., Nakazawa, K., & Akai, M. (2008). Robotic gait trainer in water: Development of an underwater gait-training orthosis. *Disability and Rehabilitation*, *30*(2), 81–87.

Moran, M. E. (2007). Evolution of robotic arms. *Journal of Robotic Surgery*, *1*(2), 103–111.

Mordoch, E., Osterreicher, A., Guse, L., Roger, K., & Thompson, G. (2013). Use of social commitment robots in the care of elderly people with dementia: A literature review. *Maturitas*, *74*(1), 14-20.

Muggleton, S. (2014). Alan Turing and the development of Artificial Intelligence. AI communications, 27(1), 3-10.

Mukai, T., Hirano, S., Nakashima, H., Kato, Y., Sakaida, Y., Guo, S., & Hosoe, S. (2010). Development of a nursing-care assistant robot RIBA that can lift a human in its arms. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 5996–6001.

Naganuma, M., Tetsui, T., Ohkubo, E., Kimura, R., & Kato, N. (2008). Trial of robot assisted rehabilitation using robotic pet. *Gerontechnology*, *7*(2).

Narayan, J., Kalita, B., & Dwivedy, S. K. (2021). Development of robot-based upper limb devices for rehabilitation purposes: a systematic review. *Augmented Human Research*, *6*, 1-33.

Narayanan, K. L., Krishnan, R. S., Son, L. H., Tung, N. T., Julie, E. G., Robinson, Y. H., & Gerogiannis, V. C. (2022). Fuzzy guided autonomous nursing robot through wireless beacon network. Multimedia tools and applications, 1-29.

Naseer, F., Khan, M. N., Nawaz, Z., & Awais, Q. (2023a). Telepresence Robots and Controlling Techniques in Healthcare System. Computers, Materials & Continua, 74(3).

Naseer, F., Khan, M. N., & Altalbe, A. (2023b). Telepresence Robot with DRL Assisted Delay Compensation in IoT-Enabled Sustainable Healthcare Environment. Sustainability, 15(4), 3585.

Neerincx, A., Leven, J., Wolfert, P., & de Graaf, M. M. (2023). The Effect of Simple Emotional Gesturing in a Socially Assistive Robot on Child's Engagement at a Group Vaccination Day. ACM/IEEE International Conference on Human-Robot Interaction (pp. 162-171).

Nef, T., & Riener, R. (2005). ARMin - Design of a novel arm rehabilitation robot. *IEEE 9th International Conference on Rehabilitation Robotics*, 57–60.

Nieto Agraz, C., Pfingsthorn, M., Gliesche, P., Eichelberg, M., & Hein, A. (2022). A survey of robotic systems for nursing care. *Frontiers in Robotics and AI, 9*, 832248.

Nishat, F., Hudson, S., Panesar, P., Ali, S., Litwin, S., Zeller, F., & Stinson, J. (2023). Exploring the needs of children and caregivers to inform design of an artificial intelligence-enhanced social robot in the pediatric emergency department. Journal of Clinical and Translational Science, 7(1), e191.

Nonami, K., Barai, R. K., Irawan, A., & Daud, M. R. (2014). Hydraulically actuated hexapod robots. M], Springer Japan.

Norman, D. A. (1986). Cognitive engineering. User centered system design, 31(61), 2.

Ohneberg, C., Stöbich, N., Warmbein, A., Rathgeber, I., Mehler-Klamt, A. C., Fischer, U., & Eberl, I. (2023). Assistive robotic systems in nursing care: a scoping review. *BMC nursing*, *22*(1), 1-15.

Okita, S. Y. (2013). Self-other's perspective taking: The use of therapeutic robot companions as social agents for reducing pain and anxiety in pediatric patients. *Cyberpsychology, Behavior, and Social Networking*, *16*(6).

Onishi, M., Luo, Z. W., Odashima, T., Hirano, S., Tahara, K., & Mukai, T. (2007). Generation of human care behaviors by human-interactive robot RI-MAN. *IEEE International Conference on Robotics and Automation*, 3128–3129.

Otten, A., Voort, C., Stienen, A., Aarts, R., Van Asseldonk, E., & Van Der Kooij, H. (2015). LIMPACT:A Hydraulically Powered Self-Aligning Upper Limb Exoskeleton. *IEEE/ASME Transactions on Mechatronics*, *20*(5).

Park, K. H., Lee, H. E., Kim, Y., & Bien, Z. Z. (2008). A steward robot for human-friendly human-machine interaction in a smart house environment. *IEEE Transactions on Automation Science and Engineering*, *5*(1), 21–25.

Pasek, A. (2014). Renaissance Robotics: Leonardo da Vinci's Lost Knight and Enlivened Materiality. Shift, 7.

Perez, J., Aguilar, J., & Dapena, E. (2020). MIHR: A Human-robot interaction model. *IEEE Latin America Transactions*, *18*(9), 1521–1529.

Perry, J. C., & Rosen, J. (2006). Design of a 7 degree-of-freedom upper-limb powered exoskeleton. *1st IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics*, 805–810.

Piccinini, G. (2004). The First computational theory of mind and brain: a close look at mcculloch and pitts's "logical calculus of ideas immanent in nervous activity". Synthese, 141, 175-215.

Pineau, J., Montemerlo, M., Pollack, M., Roy, N., & Thrun, S. (2003). Towards robotic assistants in nursing homes: Challenges and results. *Robotics and Autonomous Systems, 42*(3–4), 271–281.

Rajaraman, V. (2014). JohnMcCarthy—Father of artificial intelligence. Resonance, 19, 198-207.

Ren, Y., Park, H. S., & Zhang, L. Q. (2009). Developing a whole-arm exoskeleton robot with hand opening and closing mechanism for upper limb stroke rehabilitation. *IEEE International Conference on Rehabilitation Robotics*, 761–765.

Retto, J. (2017). Sophia, first citizen robot of the world. ResearchGate, URL: https://www. researchgate. net.

Romdhane, L., & Zeghloul, S. (2009). AL-JAZARI (1136–1206). In Distinguished Figures in Mechanism and Machine Science: Their Contributions and Legacies, Part 2 (pp. 1-21). Dordrecht: Springer Netherlands.

Rossi, S., Larafa, M., & Ruocco, M. (2020). Emotional and Behavioural Distraction by a Social Robot for Children Anxiety Reduction During Vaccination. *International Journal of Social Robotics*, *12*(3), 765–777.

Ryan, H., & Tsuda, S. (2015). History of and current systems in robotic surgery. *Essentials of Robotic Surgery*, 1–12.

Sahoo, S. K., & Choudhury, B. B. (2023). Evaluating Material Alternatives for low cost Robotic Wheelchair Chassis: A Combined CRITIC, EDAS, and COPRAS Framework. Jordan Journal of Mechanical & Industrial Engineering, 17(4).

Sato, M., Yasuhara, Y., Osaka, K., Ito, H., Dino, M. J. S., Ong, I. L., Zhao, Y., & Tanioka, T. (2020). Rehabilitation care with Pepper humanoid robot: A qualitative case study of older patients with schizophrenia and/or dementia in Japan. *Enfermeria Clinica*, *30*, 32–36.

Schabowsky, C. N., Godfrey, S. B., Holley, R. J., & Lum, P. S. (2010). Development and pilot testing of HEXORR: Hand exoskeleton rehabilitation robot. *Journal of NeuroEngineering and Rehabilitation*, *7*(1), 1–16.

Schmidtler, J., Knott, V., Hölzel, C., & Bengler, K. (2015). Human Centered Assistance Applications for the working environment of the future. *Occupational Ergonomics*, *12*(3), 83–95.

Scholtz, J. (2003). Theory and evaluation of human robot interactions. *36th Annual Hawaii International Conference on System Sciences*.

Sifeng, Z., Min, T., Zehao, Z., & Zhao, Y. (2016). Capturing the opportunity in developing intelligent elderly care robots in China challenges, opportunities and development strategy. IEEE Workshop on Advanced Robotics and its Social Impacts, 61-66.

Shaik, T., Tao, X., Xie, H., Li, L., Yong, J., & Dai, H. N. (2023). AI-Driven Patient Monitoring with Multi-Agent Deep Reinforcement Learning. arXiv preprint arXiv:2309.10980.

Sharkey, N., & Sharkey, A. (2008). Electro-mechanical robots before the computer. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 223(1), 235-241.

Shenoy, S., Hou, Y., Wang, X., Nikseresht, F., & Doryab, A. (2021). Adaptive humanoid robots for pain management in children. In Companion of the ACM/IEEE International Conference on Human-Robot Interaction (pp. 510-514).

Shibata, T., Wada, K., Ikeda, Y., & Sabanovic, S. (2009). Cross-cultural studies on subjective evaluation of a seal robot. *Advanced Robotics*, *23*(4), 443-458.

Shibata, T., & Wada, K. (2011). Robot therapy: a new approach for mental healthcare of the elderly–a mini-review. *Gerontology*, *57*(4), 378-386.

Shigemi, S. (2018). ASIMO and Humanoid Robot Research at Honda. *Humanoid Robotics: A Reference*, 55–90.

Smakman, M. H., Smit, K., Buser, L., Monshouwer, T., van Putten, N., Trip, T., & Tiel Groenestege, W. M. (2021). Mitigating children's pain and anxiety during blood draw using social robots. Electronics, 10(10), 1211.

Song, A., Pan, L., Xu, G., & Li, H. (2015). Adaptive motion control of arm rehabilitation robot based on impedance identification. *Robotica*, *33*(9), 1795–1812.

Qassim, H. M., & Wan Hasan, W. Z. (2020). A review on upper limb rehabilitation robots. *Applied*

Sciences, *10*(19), 6976.

Tamantini, C., Cordella, F., Lauretti, C., Di Luzio, F. S., Campagnola, B., Cricenti, L., & Zollo, L. (2023). Tailoring upper-limb robot-aided orthopedic rehabilitation on patients' psychophysiological state. IEEE Transactions on Neural Systems and Rehabilitation Engineering.

Tamura, T., Yonemitsu, S., Itoh, A., Oikawa, D., Kawakami, A., Higashi, Y., Fujimooto, T., & Nakajima, K. (2004). Is an Entertainment Robot Useful in the Care of Elderly People with Severe Dementia? *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *59*(1), 83–85.

Tapus, A., Ţăpuş, C., & Matarić, M. J. (2008). User-robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics*, *1*(2), 169–183.

Tanaka, K., Hayakawa, M., Noda, C., Nakamura, A., & Akiyama, C. (2022). Effects of artificial intelligence aibo intervention on alleviating distress and fear in children. Child and Adolescent Psychiatry and Mental Health, 16(1), 1-7.

Tiberio, L., Cesta, A., Cortellessa, G., Padua, L., & Pellegrino, A. R. (2012). Assessing affective response of older users to a telepresence robot using a combination of psychophysiological measures. *IEEE International Workshop on Robot and Human Interactive Communication*, 833–838.

Triantafyllidis, A., Alexiadis, A., Votis, K., & Tzovaras, D. (2023). Social Robot Interventions for Child Healthcare: A Systematic Review of the Literature. *Computer Methods and Programs in Biomedicine Update*, 100108.

Trost, M. J., Chrysilla, G., Gold, J. I., & Matarić, M. (2020). Socially-Assistive robots using empathy to reduce pain and distress during peripheral IV placement in children. *Pain Research and Management*.

Tsoi, Y. H., Xie, S. Q., & Graham, A. E. (2009). Design, modeling and control of an ankle rehabilitation robot. *Studies in Computational Intelligence*, *177*, 377–399.

Van Lam, P., & Fujimoto, Y. (2019). A Robotic Cane for Balance Maintenance Assistance. *IEEE Transactions on Industrial Informatics*, *15*(7), 3998–4009.

Van Bindsbergen, K. L., van Gorp, M., Thomassen, B. W., Merks, J. H., & Grootenhuis, M. A. (2022). Social robots in pediatric oncology: opinions of health care providers. *Journal of Psychosocial Oncology Research and Practice*, *4*(2), e073.

Vélez-Guerrero, M. A., Callejas-Cuervo, M., & Mazzoleni, S. (2021). Design, Development, and Testing of an Intelligent Wearable Robotic Exoskeleton Prototype for Upper Limb Rehabilitation. Sensors, 21(16), 5411.

Veneman, J. F., Kruidhof, R., Hekman, E. E. G., Ekkelenkamp, R., Van Asseldonk, E. H. F., & Van Der Kooij, H. (2007). Design and evaluation of the LOPES exoskeleton robot for interactive gait rehabilitation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *15*(3), 379–386.

Wang, S., Wang, L., Meijneke, C., Van Asseldonk, E., Hoellinger, T., Cheron, G., Ivanenko, Y., La Scaleia, V., Sylos-Labini, F., Molinari, M., Tamburella, F., Pisotta, I., Thorsteinsson, F., Ilzkovitz, M., Gancet, J., Nevatia, Y., Hauffe, R., Zanow, F., & Van Der Kooij, H. (2015). Design and Control of the MINDWALKER Exoskeleton. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *23*(2), 277–286.

Wu, Y. H., Wrobel, J., Cornuet, M., Kerhervé, H., Damnée, S., & Rrigaud, A. S. (2014). Acceptance of an assistive robot in older adults: A mixed-method study of human-robot interaction over a 1-month period in the living lab setting. *Clinical Interventions in Aging*, *9*, 801–811.

Wullenkord, R., & Eyssel, F. (2020). Societal and ethical issues in HRI. Current Robotics Reports, 1, 85-96.

Xu, W., Huang, J., & Cheng, L. (2018). A novel coordinated motion fusion-based walking-aid robot system. Sensors, 18(9), 2761.

Yanco, H. A., & Drury, J. (2004). Classifying human-robot interaction: An updated taxonomy. *IEEE International Conference on Systems, Man and Cybernetics*, *3*, 2841–2846.

Yang, C. Y., Lu, M. J., Tseng, S. H., & Fu, L. C. (2017). A companion robot for daily care of elders based on homeostasis. 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), 1401-1406.

Yeh, T. J., Wu, M. J., Lu, T. J., Wu, F. K., & Huang, C. R. (2010). Control of McKibben pneumatic muscles for a power-assist, lower-limb orthosis. *Mechatronics*, *20*(6), 686–697.

Yun, Y., Dancausse, S., Esmatloo, P., Serrato, A., Merring, C. A., Agarwal, P., & Deshpande, A. D. (2017). Maestro: An EMG-driven assistive hand exoskeleton for spinal cord injury patients. *IEEE International Conference on Robotics and Automation*, 2904–2910.

Zhang, H., Austin, H., Buchanan, S., Herman, R., Koeneman, J., & He, J. (2011). Feasibility studies of robot-assisted stroke rehabilitation at clinic and home settings using RUPERT. *IEEE International Conference on Rehabilitation Robotics*.

Zhang, F., & Demiris, Y. (2023). Visual-tactile learning of garment unfolding for robot-assisted dressing. IEEE Robotics and Automation Letters.

Zhou, J., Yang, S., & Xue, Q. (2021). Lower limb rehabilitation exoskeleton robot: A review. *Advances in Mechanical Engineering*, *13*(4), 16878140211011862.

APPENDICES

Appendix A: The Voluntary Participation Form (For all participants whose images were utilized in this thesis study)

ATÜ	T.C. ADANA ALPARSLAN TÜRKEŞ BİLİM VE TEKNOLOJİ ÜNİVERSİTESİ BİLGİLENDİRİLMİŞ GÖNÜLLÜ OLUR FORMU (ARAŞTIRMA/ANKET ÇALIŞMALARI İÇİN)	1/3

LÜTFEN BU DÖKÜMANI DİKKATLİCE OKUMAK İÇİN ZAMAN AYIRINIZ

Sayın..

Sizi Fatma GÖNGÖR' ün ***"ROBOT DOKTORLAR: TANIMLAMALAR, DEĞERLENDİRMELER VE DİYALOG TABANLI MÜDAHALELER (ROBOT DOCTORS: RECOGNITIONS, ASSESSMENTS, AND DIALOG-BASED INTERVENTIONS)"*** başlıklı doktora tez çalışmasına davet ediyoruz. Bu araştırma çalışmasına katılıp katılmama kararını vermeden önce, araştırmanın neden ve nasıl yapılacağını, bu araştırmanın gönüllü katılımcılara getireceği olası fayda ve rahatsızlıklarını bilmeniz gerekmektedir. Bu nedenle bu formun okunup anlaşılması büyük önem taşımaktadır. Aşağıdaki bilgileri dikkatlice okumak için zaman ayırınız. İsterseniz bu bilgileri aileniz ve/veya yakınlarınız ile tartışınız. Eğer anlayamadığınız ve sizin için açık olmayan şeyler varsa, ya da daha fazla bilgi isterseniz bize sorunuz. Bu araştırma / anket çalışmasına katılmayı kabul ettiğiniz takdirde, gerekli yerleri siz ve sorumlu araştırmacı tarafından doldurulmuş bu formun bir kopyası saklamanız için size verilecektir.

Araştırma çalışmasına katılmak tamamen **gönüllülük** esasına dayanmaktadır. Size verilen **formlarındaki** soruları yanıtlarken kimsenin baskısı veya telkini altında olmayın. Bu formlardan elde edilecek bilgiler tamamen araştırma amacı ile kullanılacaktır ve kimlik bilgileriniz kesinlikle gizli tutulacaktır.

Çalışmaya katılmama veya herhangi bir anda çalışmadan çıkma hakkında sahipsiniz. Her iki durumda da bir ceza veya hakkınız olan yararların kaybı kesinlikle söz konusu olmayacaktır.

Danışman
Prof. Dr. Önder TUTSOY

Tezin Amacı:
Bu tez, IR'lerin kritik sağlık parametrelerini otonom bir şekilde tanımlamasını, kapsamlı değerlendirmeler yapmasını ve diyalog tabanlı müdahaleler gerçekleştirmesini sağlayacak bir dizi sofistike algoritma geliştirmekte ve böylece bu robotları bireylerin günlük yaşamlarında kişisel sağlık hizmeti sunucuları olarak konumlandırmaktadır. Tezin ilk bölümü, yardımcı robotik alanındaki son gelişmeleri ve bunların sağlık hizmetlerine etkilerini ayrıntılı bir literatür incelemesi ile ele almaktadır. Devam eden bölümlerde, kalp atış hızı izleme algoritmasının tasarımı ve uygulanması detaylı bir şekilde incelenmekte, ardından kişiselleştirilmiş sıcaklık tahmini algoritmasının derinlemesine bir analizi sunulmaktadır. COVID-19 pandemisinin yol açtığı sosyal izolasyon ve karantinanın psikolojik etkilerine yanıt olarak, bir sonraki bölümde karantina altındaki hastalar için özel olarak geliştirilmiş bir konuşma bazlı duygu analizi algoritması sunulmaktadır. Bunu takiben, pandeminin

ATÜ	**T.C. ADANA ALPARSLAN TÜRKEŞ BİLİM VE TEKNOLOJİ ÜNİVERSİTESİ BİLGİLENDİRİLMİŞ GÖNÜLLÜ OLUR FORMU (ARAŞTIRMA/ANKET ÇALIŞMALARI İÇİN)**	**2/3**

hafifletilmesine yönelik olarak geliştirilen yüz maskesi takma kontrolü algoritması ele alınmaktadır. Tez, IR'lerin geriatrik hastalar için fiziksel egzersizlerin kolaylaştırılmasındaki rolünü inceleyen bir vaka çalışması ile sonlanmaktadır. Geliştirilen algoritmalar, bu tez kapsamında özellikle NAO, Alpha Mini ve Alpha 1Pro insansı robotlarda uygulanmış olmakla birlikte, tüm diğer IR'lere de uyarlanabilir niteliktedir. Uygulanan her bir algoritmanın simülasyon ve deney sonuçları, yardımcı IR'lerin sağlık hizmetlerindeki kritik rolünü vurgulamakta ve sağlık müdahalelerinin ilerletilmesinde sağladıkları olumlu etki ve etkinlikleri gözler önüne sermektedir..

Araştırmanın Nedeni: ☒ Bilimsel Araştırma ☒ Tez Çalışması

İzlenecek Olan Yöntem ve Yapılacak İşlemler:

Bu çalışmada sadece belli periyotlarda deneklerin yüz ve vücut görüntüleri (belli sayıda)alınacaktır.

Araştırmanın Süresi: 5 yıl

Katılımcı / Gönüllü Sayısı: 35

Size Getirebileceği Olası fayda ve Rahatsızlıklar:
Bu çalışmada denekler kullanılarak, gerçek yüz ve vücut ölçüm değerlerinin kullanılması, çalışmanın doğruluğunu test etme açısından fayda sağlamaktadır. Aksi takdirde elde edilen sonuçların gerçekliğini ve doğruluğunu test emek pek mümkün olmamakla birlikte herhangi bir anlam ifade etmez.

Araştırmanın Yapılacağı Yer(ler): Adana Alparslan Türkeş Bilim ve Teknoloji Üniversitesi

Araştırmaya Katılan Araştırıcılar:

Katılma ve Çıkma:
Bu araştırma çalışmasına katılmak tamamen gönüllülük esasına dayanmaktadır. Çalışmaya katılmama veya herhangi bir anda çalışmadan çıkma hakkına sahipsiniz. Ayrıca sorumlu araştırıcı gerek duyarsa sizi çalışma dışı bırakabilir. Çalışmaya katılmama, çalışmadan çıkma veya çıkarılma durumlarında bir ceza veya hakkınız olan yararların kaybı kesinlikle söz konusu olmayacaktır.

Masraflar:
Bu çalışmada sadece deneklerin yüz ve vücut görüntüleri alınıp, ölçümleri kullanıldığı için herhangi bir masraf söz konusu değildir.

Gizlilik:
Bu çalışmadan elde edilen bilgiler tamamen araştırma amacı ile kullanılacak ve kimlik bilgileriniz kesinlikle gizli tutulacaktır.

Ben ……………………………………… yukarıdaki metni okudum ve katılmam istenen tez çalışmasının amacını, gönüllü olarak üzerime düşen sorumlulukları tamamen anladım.

	T.C. ADANA ALPARSLAN TÜRKEŞ BİLİM VE TEKNOLOJİ ÜNİVERSİTESİ BİLGİLENDİRİLMİŞ GÖNÜLLÜ OLUR FORMU (ARAŞTIRMA/ANKET ÇALIŞMALARI İÇİN)	3/3

Çalışma hakkında soru sorma ve tartışma imkanı buldum ve tatmin edici yanıtlar aldım. Bu çalışmayı istediğim zaman ve herhangi bir neden belirtmek zorunda kalmadan bırakabileceğimi ve bıraktığım zaman herhangi bir olumsuzlukla karşılaşmayacağımı anladım.

Bu koşullarda söz konusu araştırma çalışmasına kendi rızamla, hiçbir baskı ve zorlama olmaksızın katılmayı (çocuğumun/vasimin bu çalışmaya katılmasını) kabul ediyorum.

GÖNÜLLÜNÜN (Kendi el yazısı ile)		İMZASI
ADI SOYADI		
ADRESİ		
TELEFON / FAKS		
TARİH		

VELAYET VEYA VESAYET ALTINDA BULUNANLAR İÇİN VELİ VEYA VASİNİN (Kendi el yazısı ile)		İMZASI
ADI SOYADI		
ADRESİ		
TELEFON / FAKS		
TARİH		

AÇIKLAMALARI YAPAN ARAŞTIRICININ (YAZARIN) (Kendi el yazısı ile)		İMZASI
ADI SOYADI		
TARİH		

	T.C. ADANA ALPARSLAN TÜRKEŞ BİLİM VE TEKNOLOJİ ÜNİVERSİTESİ BİLGİLENDİRİLMİŞ GÖNÜLLÜ OLUR FORMU (ARAŞTIRMA/ANKET ÇALIŞMALARI İÇİN)	4/3

VELAYET VEYA VESAYET ALTINDA BULUNANLAR İÇİN VELİ VEYA VASİNİN (Kendi el yazısı ile)		İMZASI
ADI SOYADI		
GÖREVİ		
TARİH		