DEVELOPMENT OF A LANDMARK-BASED MOTION DETECTION SYSTEM FOR ENHANCED FROZEN SHOULDER REHABILITATION

THANAWAT SRIKAEWSIEW

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Telecommunication and Computer Engineering
Suranaree University of Technology
Academic Year 2023

การพัฒนาระบบตรวจจับการเคลื่อนไหวด้วยเทคนิคแลนด์มาร์กสำหรับการกายภาพผู้ป่วยโรคไหล่ติด

นายธนวรรรต สีแก้วสิ่ว

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมโทรคมนาคมและคอมพิวเตอร์
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2566

DEVELOPMENT OF A LANDMARK-BASED MOTION DETECTION SYSTEM FOR ENHANCED FROZEN SHOULDER REHABILITATION

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for a Master's Degree.

Thesis Examining Committee

....

(Assoc. Prof. Dr. Paramate Horkaew)

Chairperson

....

(Asst. Prof. Dr. Sarunya Kanjanawattana)

Member (Thesis Advisor)

....

(Dr. Gun Bhakdisongkhram)

Member

....

(Dr. Keerachart Suksut)

Member

....

(Assoc. Prof. Dr. Yupaporn Ruksakulpiwat)

Vice Rector for Academic Affairs and

Quality Assurance

....

(Assoc. Prof. Dr. Pornsiri Jongkol)

Dean of Institute of Engineering

นาย ธนวรรต สีแก้วสิ่ว: การพัฒนาระบบตรวจจับการเคลื่อนไหวด้วยเทคนิคแลนด์มาร์ก สำหรับการกายภาพผู้ป่วยโรคไหล่ติด (DEVELOPMENT OF A LANDMARK-BASED MOTION DETECTION SYSTEM FOR ENHANCED FROZEN SHOULDER REHABILITATION)
อาจารย์ที่ปรึกษา : ผศ. ดร. ศรัญญา กาญจนวัฒนา, 71 หน้า.

คำสำคัญ: โรคไหล่ติด/การตรวจจับท่าทาง/การฟื้นฟูสมรรถภาพทางกาย/เทคนิคแลนด์มาร์ก/คอมพิวเตอร์วิทัศน์

งานวิจัยนี้นำเสนอการพัฒนาระบบตรวจจับการเคลื่อนไหวตามจุดสังเกตเพื่อเพิ่มประสิทธิภาพการฟื้นฟูโรคไหล่ติด เป้าหมายของการศึกษาคือการทำให้การฟื้นฟูสมรรถภาพโรคไหล่ติดมีความแม่นยำ และมีประสิทธิภาพมากขึ้น โดยใช้เทคโนโลยีล้ำสมัย เช่น การจับการเคลื่อนไหว การตรวจจับใบหน้า คอมพิวเตอร์วิทัศน์สำหรับการวัดระดับ และความคล้ายคลึงของโคไซน์ การวิจัยเกี่ยวข้องกับการจัดหาอาสาสมัครที่มีสุขภาพดีเพื่อประเมินความแม่นยำของระบบในการวัดประสิทธิภาพด้วยการใช้การวิเคราะห์เฟรมหลัก การวัดระดับมุม และคะแนนความคล้ายคลึง ระบบนี้มีจุดมุ่งหมายเพื่อระบุความเบี่ยงเบนในการเคลื่อนไหว และช่วยอำนวยความสะดวกในการปรับแต่งกลยุทธ์การฟื้นฟูสมรรถภาพสำหรับผู้ป่วยแต่ละราย วิธีการประกอบด้วยการคัดเลือกผู้เข้าร่วม การทำท่ากายภาพบำบัด การรวบรวมข้อมูลผ่านการวัดมุม และการวิเคราะห์เปรียบเทียบโดยใช้ระบบที่พัฒนาขึ้น คำถามในการวิจัยมุ่งเน้นไปที่การมีส่วนร่วมของระบบในการวัดความสม่ำเสมอของผู้ป่วยต่อท่าทางในการรักษา การระบุความเบี่ยงเบนในการเคลื่อนไหว และการประเมินผลกระทบของการบูรณาการทางเทคโนโลยีต่อประสิทธิภาพการรักษา ระบบที่นำเสนอสามารถที่จะเพิ่มความแม่นยำในการกายภาพบำบัดสำหรับผู้ป่วยไหล่ติด และปรับปรุงผลลัพธ์การฟื้นฟูให้มีประสิทธิภาพมากยิ่งขึ้น

สาขาวิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนักศึกษา............................

ปีการศึกษา 2566 ลายมือชื่ออาจารย์ที่ปรึกษา............................

THANAWAT SRIKAEWSIEW: DEVELOPMENT OF A LANDMARK-BASED MOTION DETECTION SYSTEM FOR ENHANCED FROZEN SHOULDER REHABILITATION THESIS ADVISOR: ASSISTANT PROFESSOR DR. SARUNYA KANJANAWATTANA , Ph.D. 71 PP.

Keywords: Frozen Shoulder/Motion Detection/Physical rehabilitation/Landmark-based System/Computer Vision

This research presents the development of a landmark-based motion detection system for enhanced frozen shoulder rehabilitation. The goal of the study was to make frozen shoulder rehabilitation more accurate and useful by using cutting edge technologies like motion capture, face detection, computer vision for degree measurement, and cosine similarity. The research involved the recruitment of healthy volunteers to assess the system's accuracy in measuring efficacy. By utilizing key frame analysis, angle degree measurements, and similarity scores, the system aimed to identify movement deviations and tailor rehabilitation strategies for individual patients. The methodology included participant recruitment, physical therapy pose execution, data collection through angle measurements, and comparative analysis using the developed system. The research questions focused on the system's contribution to measuring patient adherence to therapeutic postures, identifying movement deviations, and evaluating the impact of technological integration on treatment efficacy. The proposed system sought to enhance physical therapy precision for frozen shoulder patients, ultimately improving their rehabilitation outcomes.

School of Computer Engineering — Student's Signature..

Academic year 2023 — Advisor's Signature..

ACKNOWLEDGEMENTS

The author gratefully acknowledges the advice, support, and guidance given by Asst. Prof. Dr. Sarunya Kanjanawattana during the course of this research. Special note of thanks to Assoc. Prof. Dr. Paramate Horkaew, Dr. Gun Bhakdisongkhram and Dr. Keerachart Suksut for participating as members of the examination committee for this thesis.

The author wishes to express sincere gratitude to Dulyawat Wiriyaphong, who is currently pursuing a Master's degree in Biomedical Innovation Engineering, for his invaluable guidance throughout the rehabilitation process. His expertise as a rehabilitator significantly contributed to the clinical experiment, including his roles as an instructor in the instructional videos, an evaluator of volunteer outcomes, and the primary measurer for the study. This thesis would not have achieved its present quality without his advice on clinical procedures and methods. Additionally, his extensive knowledge of frozen shoulder and exercise postures was instrumental.

Special thanks are also due to all the volunteers who participated in this experiment. Their dedication and performance were crucial in supporting this research.

Furthermore, the author gratefully acknowledges the support provided by the OROG Scholarships (External Grants and Scholarships for Graduate Students). This funding was essential for the successful completion of this study.

Thanawat Srikaewsiew

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

Table | Page

LIST OF FIGURE

CHAPTER I

INTRODUCTION

1.1 Motivations of the study

In the ever-evolving landscape of healthcare, where technological innovations are reshaping the way we approach medical challenges, the realm of musculoskeletal disorders stands as a significant focal point. These disorders, with their intricate nuances and often debilitating impacts, beckon for refined rehabilitation methodologies. As we navigate this era of transformative healthcare, the imperative to address musculoskeletal issues becomes increasingly pronounced (Ahmad et al., 2022; Briggs et al., 2020).

Recent strides in healthcare underscore the potential advantages of integrating cutting-edge technologies into rehabilitation practices. Among these, motion detection systems emerge as powerful tools, capable of unraveling the complexities of patient movements and expanding the horizons of remote healthcare delivery (Cooper & Cooper, 2019; Albahri et al., 2018).

Amid this backdrop, frozen shoulder (Dias et al., 2005) takes center stage—an ailment characterized by pain and restricted joint mobility. Its intricacies demand a departure from conventional rehabilitation (Challoumas et al., 2020) approaches toward a more personalized paradigm (Choi et al., 2018).

It is within this narrative that the proposed landmark-based comparative analysis within motion detection systems (Srikaewsiew et al., 2022) steps into focus, positioning itself not just as an innovation but as a catalyst for nuanced assessments and tailored therapeutic interventions.

As the challenges posed by musculoskeletal disorders (Fernandes et al., 2018) persist, and the potential of technology continues to expand, this study seeks to contribute to the ongoing paradigm shift in musculoskeletal rehabilitation.

It aims to delve into the realm of personalized and technology-enhanced healthcare, where the scrutiny of patient movements against instructional videos, utilizing landmark-based analysis, becomes not just a method but a key to unlocking precision in treatment.

This study, driven by the dynamic interplay of healthcare and technology, poaspires to integrate advanced systems into rehabilitation practices, offering a path toward precision and personalization in the treatment of musculoskeletal disorders—particularly frozen shoulder. As we embark on this exploration, the synthesis of medical science and technological prowess becomes a symphony, promising to harmonize the complexities of musculoskeletal rehabilitation.

1.2 Advancements in frozen shoulder rehabilitation

Frozen shoulder (Dias et al., 2005), or adhesive capsulitis (Tasto & Elias, 2007), stands as a challenging musculoskeletal condition, causing pain, stiffness, and restricted motion in the shoulder joint. This condition profoundly affects daily life (Lyne et al., 2022), impeding routine activities and causing persistent discomfort. Traditional management involves a combination of physical therapy, exercises, and, in severe cases, surgical interventions, with physical therapy playing a key role in alleviating symptoms (Pandey & Madi, 2021).

Despite established treatments, optimizing rehabilitation for frozen shoulder patients remains a challenge due to varying individual responses. This necessitates exploring innovative solutions, and recent technological advancements in motion detection systems (Roggio et al., 2021) show promise for enhancing precision in rehabilitation.

The proposed thesis, " Development of a Landmark-Based Motion Detection System for Enhanced Frozen Shoulder Rehabilitation" seeks to contribute to the evolution of frozen shoulder rehabilitation. By developing a motion detection system grounded in landmark techniques, the research aims to address individual patient nuances, optimize interventions, and improve overall outcomes.

Through a comparative analysis of patient and therapist movements, leveraging advanced technologies such as cosine similarity (Abdulghani et al., 2023; Srikaewsiew et al., 2022) and facial expression modeling (Revina & Emmanuel, 2018), the thesis aims to establish a comprehensive understanding of the rehabilitation process. Integrating key frame analysis, angle degree measurements, and similarity scores offers a holistic approach to evaluating patient progress, allowing for a more nuanced and tailored therapeutic regimen.

The exploration of landmark-based motion detection in frozen shoulder rehabilitation aligns with the trajectory of technological advancements in healthcare. The proposed research responds to the imperative of providing a personalized, effective, and data-driven approach to address the multifaceted challenges posed by frozen shoulder. This thesis endeavors to contribute to the evolution of rehabilitation strategies, fostering improved patient outcomes and enhancing the quality of life for individuals grappling with the constraints of frozen shoulder.

1.3 Landmark-based motion detection

Landmark-based motion detection represents a transformative approach in healthcare (Fried et al., 2023), particularly in the realm of rehabilitation (Latreche et al., 2023). This innovative methodology utilizes anatomical landmarks to precisely track and analyze movements, offering a nuanced understanding of the dynamics involved in therapeutic processes.

In the context of frozen shoulder rehabilitation, a condition characterized by pain and restricted motion, traditional treatment approaches often face challenges in tailoring interventions to individual patient needs. Landmark-based motion detection systems present a potential solution by providing a more personalized and data-driven approach to physical therapy.

These systems, leveraging advanced technologies, enable a detailed examination of specific anatomical points during movement. By incorporating techniques such as cosine similarity and facial expression modeling, researchers can gain insights into patient progress with unprecedented precision.

The proposed thesis, " Development of a Landmark-Based Motion Detection System for Enhanced Frozen Shoulder Rehabilitation" positions itself at the intersection of technology and rehabilitation. By focusing on landmark-based motion detection, the research aims to pioneer a more advanced and tailored approach to frozen shoulder therapy.

Through a comparative analysis of patient and therapist movements, the thesis intends to contribute to the evolution of rehabilitation strategies. Key frame analysis, angle degree measurements, and similarity scores form integral components of this holistic approach, promising to redefine the standards of precision in therapeutic regimens.

The exploration of landmark-based motion detection not only aligns with the current trajectory of technological advancements in healthcare but also responds to the imperative of providing a personalized, effective, and data-driven approach to address the multifaceted challenges posed by conditions like frozen shoulder. This research endeavors to establish landmark-based motion detection as a cornerstone in enhancing precision and efficacy across various realms of physical therapy, fostering improved patient outcomes and quality of life.

1.4 Purpose of the research

The primary objective of this research was to pioneer advancements in frozen shoulder rehabilitation through the development and implementation of a Landmark-Based Motion Detection System. The overarching goal was to address the limitations of traditional rehabilitation approaches by introducing a more precise, personalized, and data-driven methodology. Specifically, the research aimed to:

1.4.1 Develop a prototype system for frozen shoulder patients using motion detection techniques based on landmark analysis. This involved comparing user-generated videos with original recordings conducted by physical therapy experts.

1.4.2 To evaluate the effectiveness of the prototype system in practical applications, specifically assessing its accuracy in motion detection, angle measurement, and comparison with instructor videos.

1.4.3 Explore the integration of computer technology into physical therapy practices, employing techniques such as face detection, motion capture, degree measurement via computer vision, and similarity comparisons.

1.4.4 Investigate and experiment with methods aimed at facilitating effective at-home physical therapy for frozen shoulder patients, with a focus on self-administration.

1.5 Scope of the research

1.5.1 Development of Landmark-Based Motion Detection System

The research centered on designing and implementing a motion detection system based on landmark techniques specifically tailored for comparing physical poses captured in user videos with those demonstrated by experts and measure the angles of specific body parts in each physical pose.

1.5.2 Comparative Analysis of Movements

The scope included a comparative analysis of movements between frozen shoulder patients and therapists, utilizing key frame analysis, angle degree measurements, and similarity scores to assess the precision and effectiveness of the motion detection system.

1.5.3 Engage with Healthy Volunteers

The research encompassed the recruitment of fourteen healthy volunteers, aged between 20 and 50 years, with careful consideration given to maintaining gender balance. Participants provided informed consent before engaging in the research procedures, thereby demonstrating their voluntary involvement. The study aimed to deliver detailed explanations regarding the experiment, including associated risks and terms, to ensure comprehensive understanding and ethical adherence throughout the research process.

1.5.4 Technological Integration and Evaluation

The research involved the integration of advanced technologies, such as cosine similarity and facial expression modeling, into the motion detection system. The focus was on evaluating the impact of technological integration on the accuracy and comprehensiveness of the analysis in the context of frozen shoulder rehabilitation.

1.6 Research questions

1.6.1 How does the Landmark-Based Motion Detection System contribute to precisely measuring and improving the frozen shoulder patient's adherence to therapeutic postures during rehabilitation interventions?

1.6.2 To what extent can the landmark-based comparative analysis identify and address specific movement deviations in frozen shoulder patients, leading to tailored and more effective rehabilitation strategies?

1.6.3 What is the impact of incorporating cosine similarity and facial emotion modeling in the motion detection system on the accuracy and thoroughness of measuring patient progress and treatment efficacy in frozen shoulder rehabilitation?

1.6.4 How does the developed motion detection system provide precise, quantifiable data for redefining frozen shoulder rehabilitation standards and advancing rehabilitation practices?

1.7 Contributions of the research

1.7.1 Precision Enhancement in Rehabilitation Practices

The research contributes by introducing a Landmark-Based Motion Detection System designed for frozen shoulder therapy. This technology enhances the precision of rehabilitation interventions, providing detailed and accurate analyses of patient movements during therapy sessions.

1.7.2 Tailored and Effective Rehabilitation Strategies

By leveraging landmark-based comparative analysis, the study pioneers personalized rehabilitation interventions. The system identifies individual patient responses and needs, allowing for tailored strategies that address specific movement deviations, thereby increasing the effectiveness of frozen shoulder therapy.

1.7.3 Integration of Advanced Technologies

The research integrates cosine similarity and facial expression modeling into the motion detection system, enhancing the accuracy and comprehensiveness of the analysis for patient progress and treatment efficacy in frozen shoulder rehabilitation.

CHAPTER II

LITERATURE REVIEW

2.1 Frozen shoulder

Frozen shoulder, medically known as adhesive capsulitis, represents a challenging musculoskeletal condition characterized by pain, stiffness, and restricted mobility within the shoulder joint. Epidemiological studies have indicated a prevalence ranging from 2% to 5% in the general population, with a notably higher occurrence among individuals aged 40 to 60 years. This condition is often associated with various risk factors, including diabetes mellitus, thyroid disorders, prior shoulder trauma or surgery, and a higher incidence in females (Dias et al., 2005; de la Serna et al., 2021).

The pathophysiology of frozen shoulder involves a multifaceted interplay of inflammatory, fibrotic, and contractile processes within the glenohumeral joint capsule and surrounding soft tissues. The clinical course of frozen shoulder typically progresses through distinct phases, starting with a painful phase characterized by increasing pain and stiffness, followed by an adhesive phase marked by significant loss of shoulder mobility, and concluding with a recovery phase where mobility gradually improves (Dias et al., 2005).

Diagnosis of frozen shoulder relies primarily on clinical evaluation, including history-taking and physical examination. However, imaging modalities such as X-rays, ultrasound, and magnetic resonance imaging (MRI) are often utilized to confirm the diagnosis and rule out other shoulder pathologies (Dias et al., 2005).

In terms of treatment modalities, a multimodal approach is usually adopted. Conservative treatments play a pivotal role and may include physical therapy interventions aimed at improving range of motion and reducing pain through specific exercises and manual techniques. Pharmacological interventions, such as corticosteroid injections, can provide symptomatic relief, particularly during the painful phase of the condition (Mertens et al., 2022).

Surgical interventions, including manipulation under anesthesia and arthroscopic release, may be considered for refractory cases where conservative measures have failed to provide adequate relief (Dias et al., 2005).

Psychological factors, including stress and anxiety, may also influence the onset and progression of frozen shoulder. Moreover, the prognosis of frozen shoulder varies among individuals, with some experiencing persistent limitations in shoulder mobility despite treatment efforts (Rizk & Pinals, 1982; Dias et al., 2005).

Furthermore, the research has identified additional risk factors associated with the development of adhesive capsulitis, particularly in high-risk populations such as neurosurgical patients. Bruckner and Nye (1981) conducted a prospective study focusing on neurosurgical patients, revealing several significant risk factors, including impairment of consciousness, hemiparesis, duration of post-operative intravenous infusion, age, and depressive personality. Routine treatment with corticosteroids post-operatively did not prevent capsulitis (Bruckner & Nye, 1981).

In conclusion, frozen shoulder poses a multifaceted clinical challenge, underscoring the importance of a thorough comprehension of its epidemiology, pathophysiology, diagnostic criteria, and treatment modalities. As advancements in technology and rehabilitation continue to evolve, there is a growing need for innovative approaches to optimize therapeutic outcomes for individuals with frozen shoulder. The development of a landmark-based motion detection system, as proposed in this thesis, holds promise for enhancing precision in physical therapy interventions. By leveraging technological advancements to refine rehabilitation strategies, we can strive towards improving patient outcomes and quality of life in individuals affected by this debilitating condition.

2.2 Landmark-based motion detection and pose estimation

Landmark-based motion detection techniques, encompassing various technologies to track and analyze movement by identifying specific anatomical landmarks on the body, have emerged as crucial tools in rehabilitation practices. Traditionally, these methods relied on markers placed on the body, but recent advancements in computer vision have introduced markerless techniques (Desmarais et al., 2021), such as Google's Mediapipe Blazepose (Bazarevsky et al., 2020), which employ machine learning algorithms to detect landmarks directly from video data.

In their research, Tharatipyakul and Pongnumkul (2023) conducted a systematic review focusing on deep learning-based pose estimation as a means of providing feedback for physical movement. Their study encompassed an extensive examination of 20 articles, specifically addressing pose estimation, movement assessment, and augmented feedback utilizing deep learning techniques. The authors meticulously categorized and analyzed the methodologies and outcomes presented in the selected articles, employing a comprehensive approach to evaluate pose estimation methods, movement assessment techniques, and classifications of augmented feedback derived from existing literature in motor learning. Their investigation revealed a predominant reliance on deep learning methodologies, notably Convolutional Neural Networks (CNN), for pose estimation tasks. They identified diverse approaches for movement assessment, ranging from mathematical formulas and rule-based methods to machine learning algorithms. Augmented feedback mechanisms predominantly manifested in visual and verbal forms, encompassing various modalities such as numbers, words, phrases, videos, images, and animations. Through their rigorous review process, the authors shed light on the current state of research in this domain, pinpointing both strengths and limitations within the existing literature. Their comprehensive analysis offered valuable insights into the application of deep learning techniques for pose estimation and augmented feedback in physical movement contexts, while also identifying avenues for future research and development.

Pauzi et al. (2021) developed a system for estimating human movement using Mediapipe Blazepose. The system tracks body movements from video sources and superimposes labelled skeleton joints onto the individual's body. This technology has wide-ranging applications, particularly in physically demanding work environments and the sports industry, where precise movement tracking is essential. The authors employed deep learning techniques, specifically utilizing the Mediapipe Blazepose algorithm and the PoseNet dataset, tailored for detecting and estimating movements prone to causing bodily injury during heavy workloads. To evaluate the system's accuracy, the authors compared it with IMU-based motion capture, revealing differences in accuracy within a 10% range. Despite this discrepancy, the proposed system aims to accurately identify and label skeleton joints on individuals' bodies. Additionally, it is designed to calculate movement velocity and joint angles, crucial factors in assessing the risk of both short- and long-term injuries. Through their research, Pauzi et al. contribute to the advancement of movement estimation technology, providing a foundation for enhanced injury prevention and movement analysis in various fields.

Singh, Kumbhare, and Arthi (2021) explored real-time human pose detection and recognition using MediaPipe technology. They introduced a framework capable of detecting human actions in real-time, even under diverse conditions and viewing angles. This framework utilized MediaPipe Holistic, which integrated pose, face, and hand landmark detection models. By parsing real-time video feed frames, they extracted 501 landmarks, exporting them as coordinates to a CSV file. These coordinates were then used to train a custom multi-class classification model, employing machine learning algorithms such as random forest, linear regression, ridge classifier, and gradient boosting classifier. The aim was to understand the relationship between body language poses and corresponding classes. Through this research, Singh et al. aimed to advance human action recognition technology for more accurate and efficient real-time detection and recognition of human poses.

The landmark detection for human pose estimation was conducted by Srikaewsiew et al (2022). The study utilized the MediaPipe framework with the BlazePose GHUM Heavy model to extract skeletal and joint data from each frame of the dance videos. Specifically, the upper portion of the body, including anatomical points such as the shoulders, elbows, and wrists, was the focus of landmark detection. By mapping these points and representing their significance as the names of body parts, the researchers were able to calculate the similarity between the vectors of each body part using the evaluation techniques of Cosine similarity, Euclidean distance, and Angular difference. This approach enabled the team to effectively analyze the similarity of posture in each frame between the instructor and the trainee, ultimately leading to the determination of the most effective method for evaluating human motion in the context of instructor-led dances.

However, applying landmark-based motion detection to frozen shoulder rehabilitation poses significant challenges. The complexity of the shoulder joint's motion and the variability of movement patterns across individuals with frozen shoulder present considerable hurdles. The intricate nature of shoulder biomechanics complicates the accurate detection and tracking of landmarks, particularly during dynamic movements.

Moreover, factors such as clothing, body composition, and patient positioning can further hinder landmark detection, leading to potential inaccuracies in motion analysis. Despite these challenges, Mediapipe offers a promising solution. Leveraging convolutional neural networks (CNNs) and pose estimation algorithms, Mediapipe enables real-time detection and tracking of key landmarks on the human body in video streams. This capability facilitates objective and quantifiable assessments of shoulder mobility, allowing clinicians to monitor progress and tailor treatment plans accordingly. By providing immediate feedback during exercise sessions, Mediapipe promotes adherence to prescribed rehabilitation protocols and enhances patient engagement in the recovery process.

Furthermore, its versatility extends to tele-rehabilitation (Gava et al., 2022) and remote monitoring (Erickson et al., 2023), enabling patients to participate in supervised rehabilitation sessions from home. This accessibility facilitates continuous monitoring of progress and adjustment of treatment plans as needed, ultimately improving patient outcomes and quality of care. Despite ongoing challenges, ongoing research and innovation in this field hold promise for optimizing the clinical utility of landmark-based motion detection in frozen shoulder rehabilitation, leading to improved outcomes and enhanced patient care.

2.3 Face expression recognition

Research in the field of facial expression recognition has been ongoing for several years, with notable contributions from various disciplines such as computer science and computer engineering. A recent study by Di Luzio et al. (2023) introduced a randomized deep neural network for emotion recognition, incorporating landmark detection. Utilizing the Extended Cohn-Kanade dataset (CK+) and Mediapipe technology, the authors extracted 468 face landmarks and employed a combination of randomized convolutional and Long Short-Term Memory (LSTM) layers to achieve over 90% accuracy in recognizing five emotions. Similarly, Hangaragi et al. (2023) proposed a face detection and recognition system using a face mesh and deep neural network, demonstrating superior accuracy compared to existing methods. Hamester et al. (2015) presented a 2-channel convolutional neural network for facial expression recognition, surpassing previous approaches in terms of accuracy on the JAFFE dataset.

Assari and Rahmati (2011) focused on non-intrusive driver drowsiness detection using facial expression recognition, achieving remarkable accuracy rates. Additionally, Munasinghe (2018) developed a method for facial expression recognition using facial landmarks and a random forest classifier, demonstrating promising results on the Extended Cohn-Kanade (CK+) database. Overall, these studies underscore the potential of deep learning techniques and landmark detection in advancing facial expression recognition technology for various applications, from affective computing to human-machine interaction

CHAPTER III

METHODOLOGY

3.1 Overview of the development of the landmark-based motion detection for enhanced physical therapy precision system

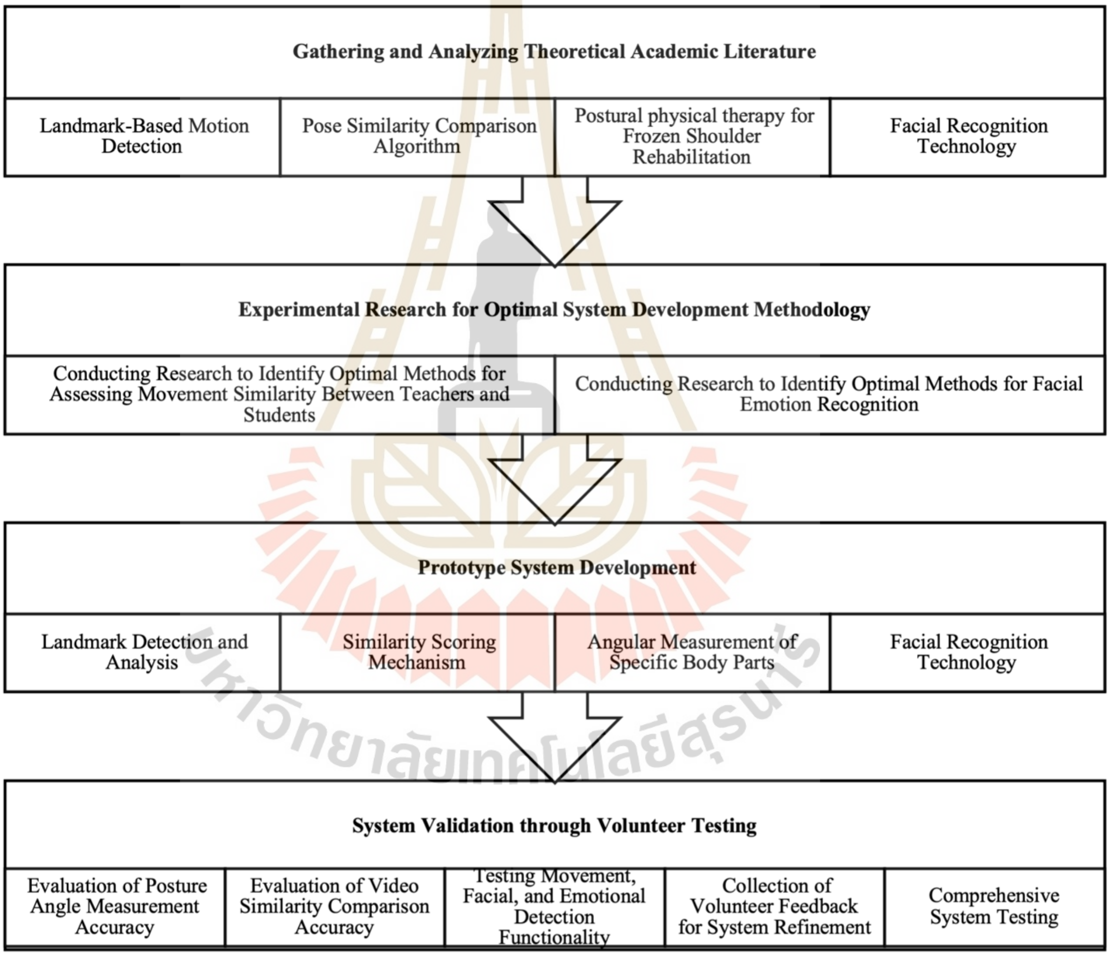


Figure 3.1 Segmented process chart of development of the Landmark-Based-Motion Detection for Enhanced Physical Therapy Precision System.

From the Figure 3.1, The methodology encompassed the gathering and analysis of theoretical academic literature on various topics, including Landmark-Based Motion Detection, Pose Similarity Comparison Algorithm, Postural physical therapy for Frozen Shoulder Rehabilitation, and Facial Recognition Technology. This phase involved scrutinizing past studies and scholarly works to establish a foundation for further research.

Following this, an experimental research approach was adopted to develop an optimal system. This involved conducting research to identify the best methods for assessing movement similarity between teachers and students (Srikaewsiew et al., 2022), as well as optimal techniques for facial emotion recognition (Srikaewsiew & Kanjanawattana, 2024). The aim was to refine existing methodologies and techniques based on empirical findings.

Subsequently, the prototype system was developed, incorporating components such as landmark detection and analysis, a similarity scoring mechanism, angular measurement of specific body parts, and facial recognition technology. This phase involved the implementation of theoretical concepts into practical solutions.

The system underwent validation through volunteer testing, wherein the accuracy of posture angle measurement, video similarity comparison, and detection functionalities (movement, facial, and emotional) were evaluated. Volunteer feedback was collected to refine the system further, and comprehensive testing was conducted to ensure its effectiveness.

Throughout the process, adherence to academic standards and rigorous methodology was paramount, ensuring the reliability and validity of the research outcomes.

3.2 Landmark-based motion detection and posture similarity score computation

In the Landmark-Based Motion Detection and Posture Similarity Score Computation methodology employed, the process commenced with data collection and pose estimation, wherein researchers gathered data by displaying a tutorial video while recording the user's motion through a smartphone. Human pose estimation was conducted on the recorded data utilizing MediaPipe, a machine learning framework designed for media applications. This estimation facilitated the extraction of x and y coordinates of the trainee based on pose landmarks, Mediapipe provides landmarks as illustrated in Figure 3.2, facilitating subsequent analysis.

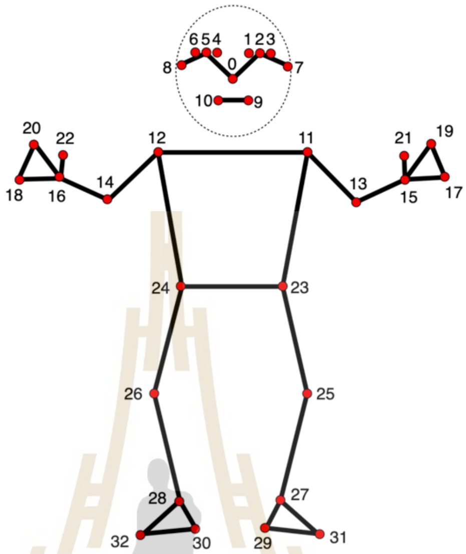


Figure 3.2 Mediapipe pose landmarks index (Bazarevsky et al., 2020).

Subsequently, attention was directed towards landmark detection, wherein researchers focused on identifying and comparing variations in landmark joints of the human body. Specific points situated in the upper body were targeted for analysis, encompassing joints such as the left shoulder, right shoulder, left elbow, right elbow, left wrist, and right wrist. These chosen landmark joints served as representative markers for significant body parts pertinent to motion analysis.

The methodology advanced to calculating posture similarity scores, aiming to discern the resemblance in motion between instructors and trainees based on their body joints in each frame. Utilizing the most effective technique identified through preliminary research (Srikaewsiew et al., 2022), which involved cosine similarity techniques, computed the posture similarity score, a numerical representation of the likeness in posture between individuals. This process involved calculating the posture similarity score for every frame based on the designated landmark joints, facilitating a comprehensive assessment of motion congruence.

Through these meticulously executed steps, successfully detected landmark joints, and computed posture similarity scores, thereby advancing the understanding and application of motion analysis techniques within academic research.

3.3 Measuring the angle of motion in a specific body region

The methodology employed in the study involved the utilization of joint angle rotations as a fundamental technique. This process entailed measuring the angles formed between adjacent body segments or joints, providing crucial insights into the orientation and movement of specific body parts. The shoulder was selected for analysis. Their initial positions in a reference frame were determined, and the movement of these joints over consecutive frames was tracked to capture motion sequences. Angles between adjacent joints at each time step were then calculated to represent joint angle rotations, utilizing arctangent functions, as illustrated in Figure 3.3, the formula demonstrates how the vector is calculated.

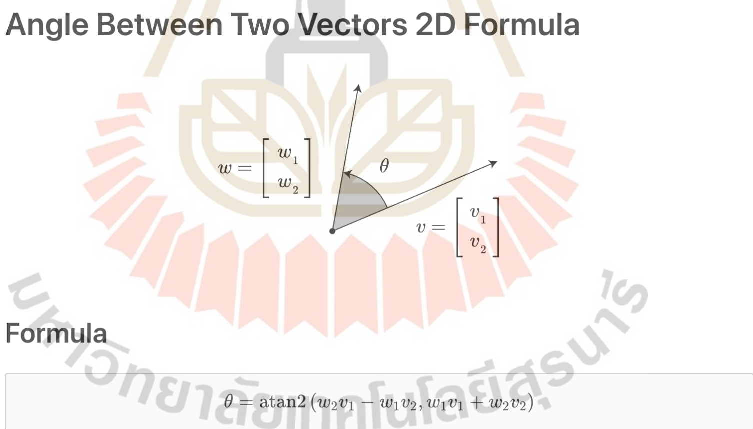


Figure 3.3 Determining the Angle Between Two 2D Vectors (Bruns, 2017).

3.4 Facial recognition

This methodology outlines a comprehensive approach to facial expression recognition utilizing a combination of Convolutional Neural Networks (CNNs) and a landmark-based technique. The landmark-based method involves the transformation of image data into coordinate representations of landmark facial points, employing the MediaPipe library for facial landmark detection. Specifically, 468 facial coordinates are extracted along both the x and y axes from facial images, as depicted in Figure 3.4. This approach leverages the capability of CNNs, a deep learning architecture renowned for its proficiency in learning intricate patterns and spatial dependencies from extensive datasets, to analyze and interpret facial data effectively.

Data preprocessing involved several steps. Facial landmark extraction utilized tools like the MediaPipe library to identify and extract key facial points such as the outer edges of the mouth, nose, and eyes. Subsequently, data normalization was performed to scale facial coordinate values appropriately, reducing variations in the data and facilitating effective pattern learning by the CNN model. Feature engineering enhanced the model's ability to recognize and classify emotions accurately by transforming raw facial coordinate data into more meaningful features capturing spatial relationships between facial landmarks.

Figure 3.4 Mediapipe Face Landmark (Google Developers, 2020).

Data augmentation techniques, such as rotation, flipping, or adding noise to the facial landmark data, were applied to increase the diversity and size of the training dataset, improving the generalization and robustness of the CNN model. The preprocessed facial landmark data was then split into training and testing sets, with the former used to train the CNN model on the facial expression recognition task and the latter to evaluate the model's performance on unseen data. Input preparation involved formatting and structuring the preprocessed facial landmark data for effective learning and feature extraction by the CNN architecture to make accurate predictions about the emotions expressed in images.

The combined CNN with landmark-based method was trained on large-scale datasets like The Delaware Pain Database (Mende-Siedlecki et al., 2020) and UTKFace, enabling it to capture both global and local facial data, thus enhancing its ability to recognize and identify various emotions. Performance evaluation revealed high accuracy, precision, recall, and F1 score values, demonstrating the effectiveness of the integrated approach in accurately identifying emotions from facial expressions (Srikaewsiew & Kanjanawattana, 2024).

3.5 System validation through volunteer testing

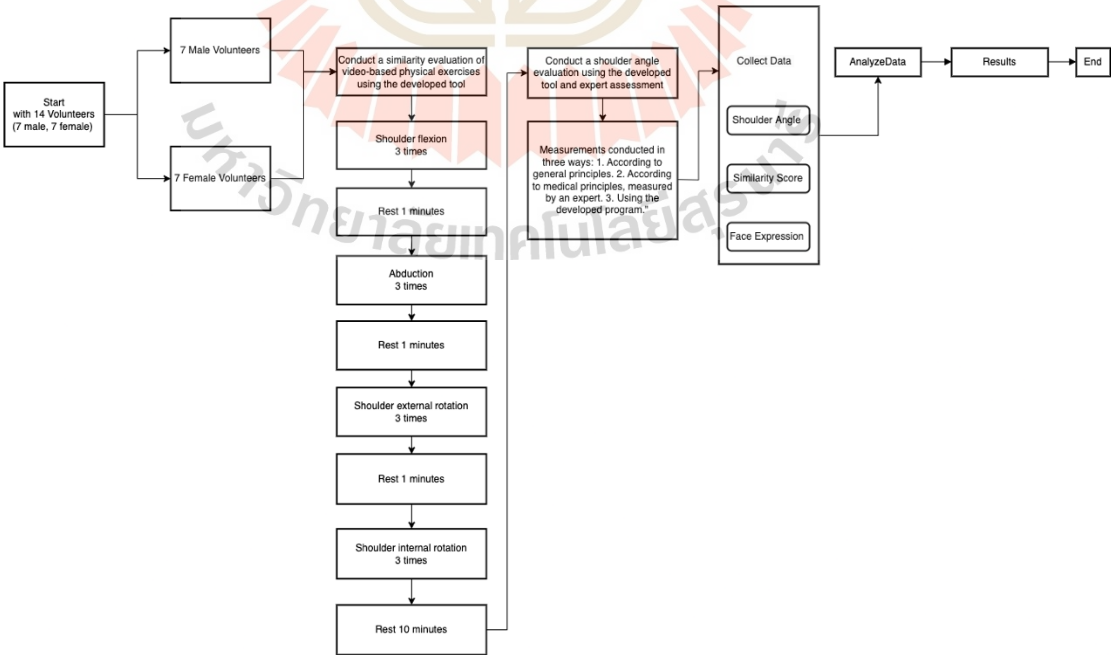


Figure 3.5 Participates testing diagram.

From Figure 3.5, the methodology adopted for participant involvement in the experiment commenced with the recruitment of fourteen healthy volunteers aged between 20 and 50 years, ensuring an equal gender distribution with seven males and seven females. Prior to participation, all individuals provided informed consent by signing a consent form, acknowledging their voluntary involvement in the experimental procedures. This study was conducted in accordance with the guidelines set forth by the Human Research Ethics Committee (EC-66-27). Following consent, participants received detailed explanations regarding the experiment, including associated risks and terms, delivered by the researchers. Adequate compensation was provided to participants as warranted.

Subsequently, the experiments commenced with the utilization of the developed system, during which participants' physical activity videos were recorded for system integration. Participants were divided into gender-specific groups and guided through a series of physical therapy poses, including shoulder flexion (Figure 3.6), abduction (Figure 3.7), shoulder external rotation (Figure 3.8), and shoulder internal rotation (Figure 3.9), under the supervision of researchers and medical professionals. Each pose was performed three times, with data collection facilitated by measuring the designated points using the system and expert physical therapists. One-minute breaks were implemented between each pose to ensure participant comfort and well-being. Upon completion of all poses, participants provided feedback on their experiences, including usability and suggestions for improvements.

The shoulder angle evaluation phase of the experiment employed a comprehensive approach to assess the accuracy and reliability of the developed system. This phase incorporated three distinct measurement methodologies:

1. General principles-based assessment: Shoulder angles were measured using standard angle measurement methods. This approach provided a baseline measurement following common practice. However, it is important to note that these principles do not conform to traditional medical measuring methods or rules.

2. Expert medical evaluation: A qualified medical professional, specifically an experienced physical therapist or orthopedic specialist, conducted measurements based on clinical expertise and medical principles. This method offered a gold

standard for comparison, leveraging years of clinical experience and specialized knowledge.

3. Assessment with Developed Program: The custom-designed program, central to this study, was utilized to measure shoulder angles. This novel approach aimed to validate the efficacy and accuracy of the developed system against established methods.

The triangulation of these measurement techniques allowed for a robust comparison between traditional methods and the innovative approach proposed in this study. Participants underwent evaluation using all three methods for each shoulder movement: flexion, abduction, external rotation, and internal rotation. This multi-faceted approach facilitated a comprehensive analysis of the developed system's performance in relation to established clinical and biomechanical standards.

Data collected from these three measurement methods were systematically recorded for subsequent statistical analysis. The comparative evaluation aimed to assess the concordance between the developed program and expert measurements, as well as to identify any significant deviations from established norms. This rigorous methodology ensures a thorough validation process for the newly developed shoulder angle evaluation tool, potentially contributing to advancements in biomechanical assessment techniques within physical therapy and sports medicine domains.

Upon conclusion, participants were allowed to depart at their convenience. Subsequently, video clips were imported into the system for comparative analysis, where measures such as the similarity between practitioners and instructors and facial expressions during physical therapy were assessed. The angles obtained were compared against traditional measurements recorded by medical professionals. Experimental results were meticulously documented, concluding the experimental phase.

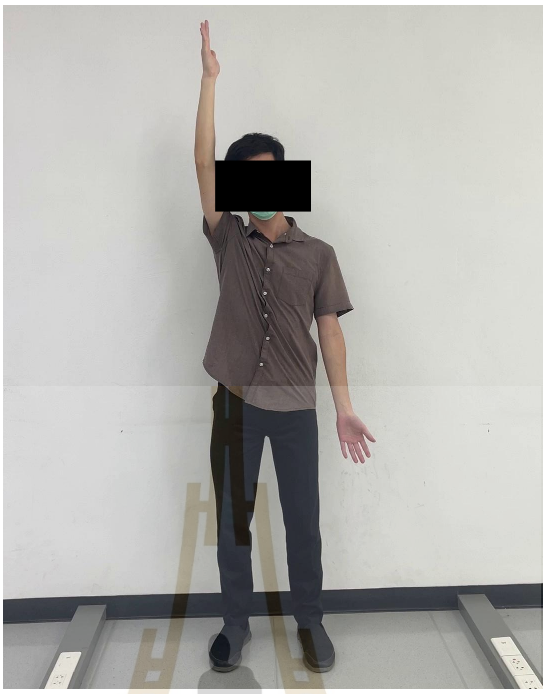


Figure 3.6 Shoulder flexion.

Figure 3.7 Abduction.

Figure 3.8 Shoulder external rotation

Figure 3.9 Shoulder internal rotation.

3.6 System application

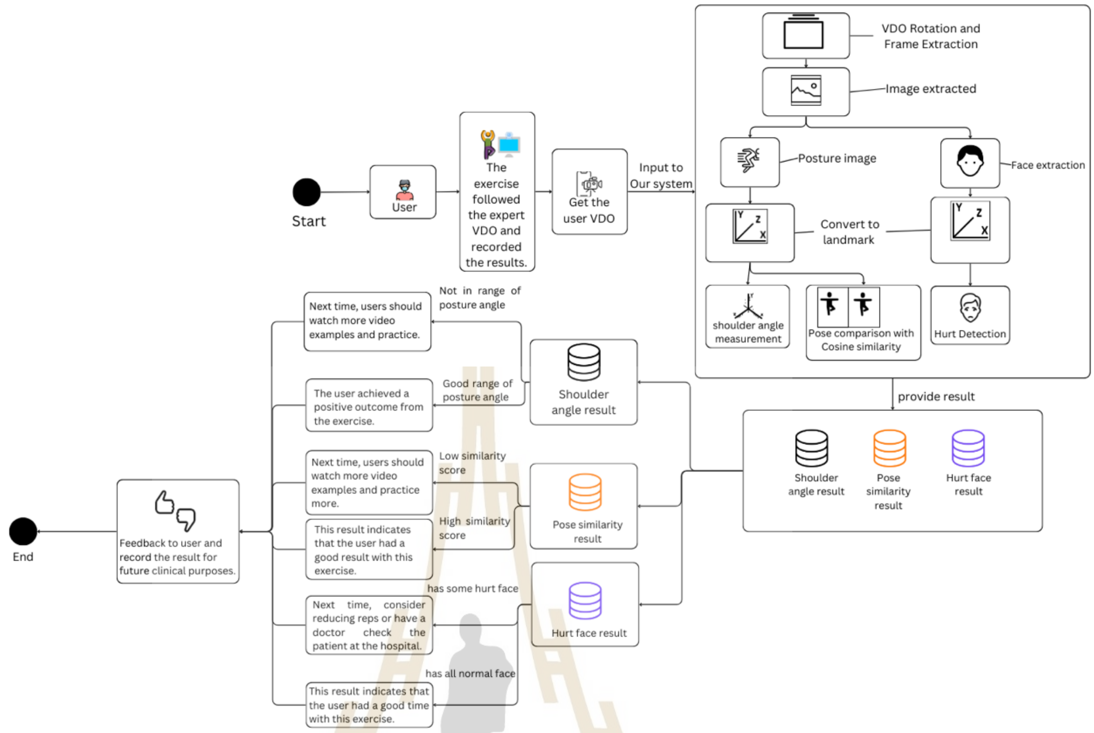


Figure 3.10 System Application Diagram.

Figure 3.10 shows an application for evaluating exercise performance through video analysis. Users follow an expert's video, record their performance, and the system processes the video by rotating it and extracting key posture images and facial features. These images are converted into landmarks to measure shoulder angles and detect expressions of pain. The system uses cosine similarity to compare the user's poses with the experts, providing detailed results on shoulder angle accuracy, pose similarity, and facial expressions. The results include feedback and recommendations, such as additional practice for low scores or consulting a healthcare professional if discomfort is detected. This feedback is documented for future clinical reference, enhancing the user's exercise performance and well-being.

CHAPTER IV
RESULTS AND DISCUSSIONS

4.1 Experimental setting

The experiment was conducted on June 14, 2024, at the SIRINDHORN WITSAWAPHAT building, 4th floor, involving 14 healthy volunteers aged between 20 and 50 years, with an equal gender distribution of seven males and seven females. The setting utilized equipment such as an iPhone 12 Pro Max, MacBook Air 2019, a goniometer, a projector display, and a projector. Some environmental conditions were controlled in the experiment: the same room was used for all participants, white lights were turned on when daylight was insufficient and turned off when daylight was bright, and the distance between the cameraman and the dancer was controlled at 1.62 meters. The experiment focused on exploring methods for at-home physical therapy for frozen shoulder patients.

4.2 The experimental procedure

The experiment involving volunteers was conducted in three stages. Initially, participants followed a video demonstrating exercises that included shoulder flexion, abduction, shoulder external rotation, and shoulder internal rotation. These exercises were supervised by researchers and medical professionals, with each participant completing three repetitions of each exercise followed by a one-minute rest period. The process continued until all volunteers had completed the exercises. Throughout these exercises, researchers utilized a Smartphone to record videos, which were subsequently analyzed using a developed program.

In the second stage, researchers employed a goniometer (Figure 4.1) for assessing shoulder angles through expert medical evaluation and general principles-based assessment, alongside evaluation facilitated by a developed program. Each exercise underwent evaluation at three distinct stages: initial posture, midpoint posture, and peak posture. The postures and stages of the poses are depicted in Figures 4.2 to 4.5.

During the final stage, researchers administered a survey to volunteers via a Google Form. The survey included questions asking participants to rate their discomfort level on a scale from 0 to 5, where 0 indicated no discomfort (normal) and 5 indicated severe discomfort (very painful).

Upon completion of the volunteer participation, the experiment concluded, and researchers proceeded to collate all gathered data for subsequent analysis to derive conclusions for their thesis.

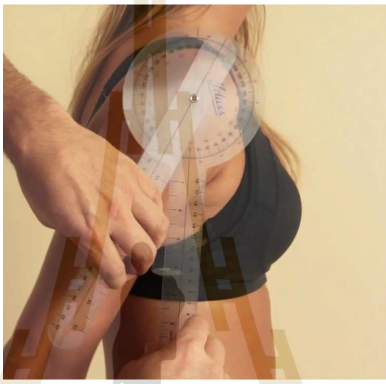
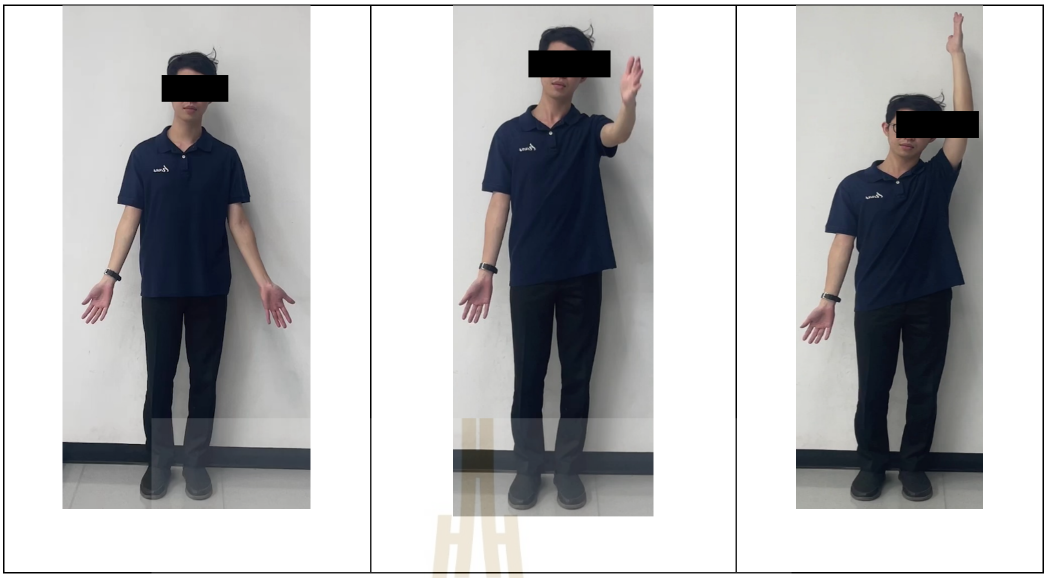


Figure 4.1 Goniometer and How to use a Goniometer to measure Range of Motion (The Goniometer, 2012).



a) b) c)

Figure 4.2 The postures and stages of the shoulder flexion: (a) initial posture, (b) midpoint posture, and (c) peak posture.

a) b) c)

Figure 4.3 The postures and stages of the abduction: (a) initial posture, (b) midpoint posture, and (c) peak posture.

a) b) c)

Figure 4.4 The postures and stages of the shoulder external rotation: (a) initial posture, (b) midpoint posture, and (c) peak posture.

a) b) c)

Figure 4.5 The postures and stages of the shoulder internal rotation: (a) initial posture, (b) midpoint posture, and (c) peak posture.

4.3 The results and discussion of participants following a demonstrated exercise video

The experiment involves the evaluation of 14 volunteer videos using frame-by-frame analysis, as depicted in Figure 4.6. Initially, an Excel file was created by the researchers where the first column contained images of the teacher's exercise, comprising approximately 20 frames each. The subsequent columns contained frames extracted from volunteer videos, also approximately 20 frames per video.

Subsequently, expert evaluators assessed the similarity between the frames by assigning scores, referred to as expert scores. The results of this similarity assessment were categorized into a rubric: scores of 0 to 49 indicated non-similarity, while scores of 50 to 100 indicated similarity.

Following expert evaluation, the evaluators returned the results to the researchers, who then incorporated these alongside cosine similarity scores. This additional data was presented in two columns: the first column contained cosine similarity scores, and the second column contained cosine similarity evaluations. The cosine similarity evaluations were derived from the cosine similarity scores using a cutoff threshold of 97.5.

Upon completion of the frame-by-frame evaluation by the experts and researchers, each video yielded a final evaluation result. These results were subsequently employed for comparison using a confusion matrix in the subsequent section of the study.

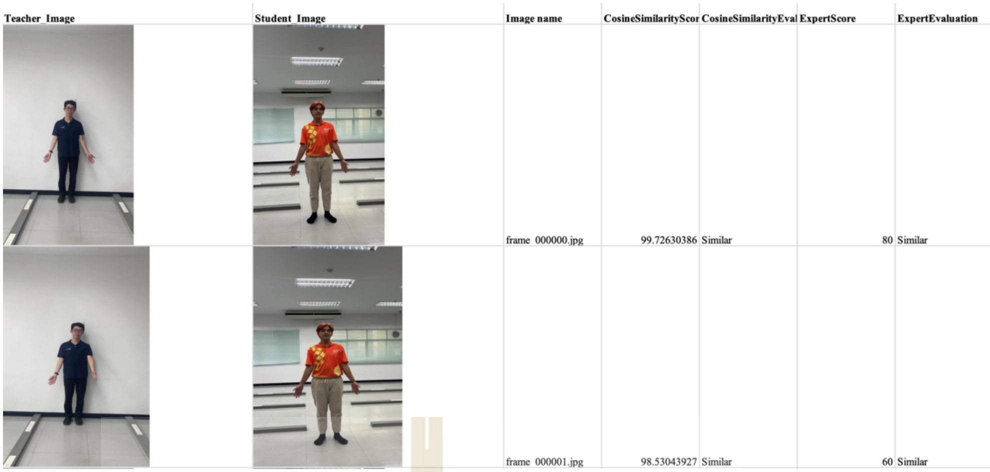


Figure 4.6 Illustrates an example of a complete Excel file used for evaluation.

4.3.1 Shoulder flexion Result

	Expert Result	Cosine Similarity Result
	Similar	Not Similar
	Similar	Similar
	Similar	Not Similar
	Not Similar	Not Similar
	Similar	Not Similar
	Similar	Similar
Shoulder flexion	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar

Table 4.1 Comparison between Expert Result and Cosine Similarity of Shoulder Flexion Result

	Predicted Similar	Predicted Not Similar
Similar	10	3
Not Similar	0	1

Table 4.2 Confusion Matrix between Expert Result and Cosine Similarity of Shoulder Flexion Result

The analysis of shoulder flexion data aimed to compare the similarity results obtained from expert evaluations with those derived from a cosine similarity algorithm. The data show as Table 4.1. It included 14 instances, with experts labeling 13 as "Similar" and 1 as "Not Similar," while the cosine similarity algorithm labeled 10 as "Similar" and 4 as "Not Similar." The performance metrics were calculated using a confusion matrix (Table 4.2), revealing a precision of 1.0, recall of approximately 0.769, accuracy of about 0.786, and an F1-score of 0.870. These results indicate that the cosine similarity algorithm is highly precise, correctly identifying similarities 100% of the time when it makes such predictions. However, its recall value suggests it misses some instances identified as similar by experts. The overall accuracy demonstrates a good agreement between the algorithm and expert judgments, while the F1-score reflects a balanced consideration of precision and recall. Although the cosine similarity algorithm shows promise with excellent precision and good accuracy, its lower recall suggests the need for further adjustments to enhance its sensitivity to expert-identified similarities.

4.3.2 Abduction Result

	Expert Result	Cosine Similarity Result
Abduction	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Not Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar

Table 4.3 Comparison between Expert Result and Cosine Similarity of Abduction Result

	Predicted Similar	Predicted Not Similar
Similar	13	1
Not Similar	0	0

Table 4.4 Confusion Matrix between Expert Result and Cosine Similarity of Abduction Result

The analysis of abduction data aimed to compare similarity results obtained from expert evaluations with those derived from a cosine similarity algorithm. In this data (Table 4.3), experts labeled all 14 instances as "Similar," while the cosine similarity algorithm labeled 13 instances as "Similar" and 1 instance as "Not Similar." The performance metrics were calculated using a confusion matrix (Table 4.4), which revealed a precision of 1.0, recall (sensitivity) of approximately 0.929, accuracy of about 0.929, and an F1-score of approximately 0.963.

These results indicate that the cosine similarity algorithm is highly precise, correctly identifying similarities 100 percentage when it makes such predictions. The recall value shows that the algorithm missed only one instance identified as similar by experts, indicating a high level of sensitivity. The overall accuracy demonstrates strong agreement between the algorithm and expert judgments, while the F1-score reflects a balanced consideration of precision and recall.

4.3.3 Shoulder external rotation

	Expert Result	**Cosine Similarity Result**
	Similar	Similar
	Similar	Not Similar
	Similar	Not Similar
	Similar	Similar
	Similar	Similar
	Similar	Not Similar
Shoulder external rotation	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar

Table 4.5 Comparison between Expert Result and Cosine Similarity of Shoulder external rotation Result

	Predicted Similar	**Predicted Not Similar**
Similar	11	3
Not Similar	0	0

Table 4.6 Confusion Matrix between Expert Result and Cosine Similarity of Shoulder external rotation Result

The analysis of shoulder external rotation data compared expert evaluations with results from a cosine similarity algorithm (Table 4.5). Experts labeled all 14 instances as "Similar," while the algorithm labeled 11 as "Similar" and 3 as "Not Similar." The performance metrics (Table 4.6) show a precision of 1.0, a recall of 0.786, an accuracy of 0.786, and an F1-score of 0.88. These results indicate that the cosine similarity algorithm is highly precise but has room for improvement in recall. While the algorithm reliably identifies similarities, it missed a few instances recognized by experts.

4.3.4 Shoulder internal rotation

	Expert Result	Cosine Similarity Result
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
Shoulder Internal rotation	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar

Table 4.7 Comparison between Expert Result and Cosine Similarity of Shoulder internal rotation Result

	Predicted Similar	Predicted Not Similar
Similar	14	0
Not Similar	0	0

Table 4.8 Confusion Matrix between Expert Result and Cosine Similarity of Shoulder internal rotation Result

The analysis of shoulder internal rotation data (Table 4.7) compared expert evaluations with results from a cosine similarity algorithm. In this dataset, both experts and the cosine similarity algorithm labeled all 14 instances as "Similar." The cofussion metrics (Table 4.8) derived from this perfect agreement are as follows: a precision of 1.0, a recall (sensitivity) of 1.0, an accuracy of 1.0, and an F1-score of 1.0. These results indicate that the cosine similarity algorithm perfectly matches the expert evaluations, identifying all instances accurately without any errors.

4.3.5 Example of an individual experiment with interesting results

	Expert Result	Cosine Similarity Result
Shoulder flexion (M005)	Similar	Similar
	Similar	Similar
	Not Similar	Not Similar
	Similar	Similar
	Similar	Similar
	Similar	Not Similar
	Not Similar	Not Similar
	Similar	Similar
	Not Similar	Not Similar
	Similar	Similar
	Similar	Similar
	Not Similar	Not Similar
	Not Similar	Not Similar
	Not Similar	Not Similar
	Similar	Not Similar
	Similar	Similar
	Not Similar	Not Similar
	Not Similar	Not Similar
	Similar	Similar

Table 4.9 Comparison between Expert Result and Cosine Similarity of M005's Shoulder flexion

	Predicted Similar	**Predicted Not Similar**
Similar	9	2
Not Similar	0	8

Table 4.10 Confusion Matrix between Expert Result and Cosine Similarity of M005's Shoulder flexion

The results presented in Table 4.9 indicate two discrepancies between the predictions made by the system and those made by the experts. The researcher observed that the participant did not adjust their posture promptly, leading to a slight delay between posture changes. This delay caused the system to judge the posture as incorrect, despite the posture beginning to resemble the subsequent correct pose. In contrast, experts awarded partial points if certain aspects of the posture were similar, resulting in discrepancies in some frames. Analyzing these discrepancies using the confusion matrix (Table 4.10) reveals slight inconsistencies in the judgment of shoulder flexion for participant M005.

	Expert Result	Cosine Similarity Result
	Similar	Similar
	Not Similar	Not Similar
	Similar	Not Similar
	Similar	Similar
	Similar	Similar
	Not Similar	Not Similar
	Similar	Similar
	Not Similar	Not Similar
Abduction (M005)	Similar	Similar
	Similar	Similar
	Not Similar	Not Similar
	Not Similar	Not Similar
	Not Similar	Not Similar
	Similar	Similar
	Similar	Similar
	Not Similar	Not Similar
	Not Similar	Not Similar
	Similar	Similar

Table 4.11 Comparison between Expert Result and Cosine Similarity of M005's Abduction

	Predicted Similar	Predicted Not Similar
Similar	9	1
Not Similar	0	8

Table 4.12 Confusion Matrix between Expert Result and Cosine Similarity of M005's Abduction

Table 4.11 shows that M005's poses were generally well executed, with only one error observed in a single frame. The researcher noted that the participant was unable to lift their arms to the appropriate height. Except for the arm-lifting aspect, the participant's postures matched those of the experts in every other respect. This discrepancy led the system to judge the pose differently for that particular frame, whereas the experts awarded a low similarity score despite recognizing the similarity. The conflicting results are illustrated in Table 4.12, which compares the expert judgments with the system's assessments using a confusion matrix.

	Expert Result	Cosine Similarity Result
Shoulder external rotation (M003)	Similar	Similar
	Similar	Similar
	Similar	Similar
	Not Similar	Not Similar
	Similar	Similar
	Similar	Similar
	Not Similar	Not Similar
	Similar	Similar
	Not Similar	Not Similar
	Similar	Not Similar
	Similar	Not Similar
	Similar	Similar
	Not Similar	Not Similar
	Similar	Similar
	Similar	Not Similar
	Not Similar	Not Similar
	Not Similar	Not Similar

Table 4.13 Comparison between Expert Result and Cosine Similarity of M003's Shoulder external rotation

	Predicted Similar	Predicted Not Similar
Similar	8	3
Not Similar	0	6

Table 4.14 Confusion Matrix between Expert Result and Cosine Similarity of M003's Shoulder external rotation

Table 4.13 shows that the results of M003's poses across various frames were largely incorrect. The researcher observed that this participant frequently performed the poses incorrectly and more slowly than the instructor. Notably, this movement occurred in a system blind spot, which may have contributed to the system's incorrect predictions, making errors more likely compared to the first two movements. However, experts also determined that this participant made numerous mistakes, aligning with the system's overall judgment. Despite this, the system's judgments did not fully match those of the experts. The results are detailed in Table 4.14, which compares the expert judgments with the system's assessments using a confusion matrix.

	Expert Result	Cosine Similarity Result
Shoulder Internal rotation (M006)	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Not Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Similar	Similar
	Not Similar	Not Similar
	Similar	Similar
	Not Similar	Not Similar
	Similar	Similar

Table 4.15 Comparison between Expert Result and Cosine Similarity of M006‘s Shoulder internal rotation

	Predicted Similar	Predicted Not Similar
Similar	14	1
Not Similar	0	2

Table 4.16 Confusion Matrix between Expert Result and Cosine Similarity of M006‘s Shoulder internal rotation

Table 4.15 shows that the judgment results for this specific pose differ from the overall video judgment results, where the system and experts had 100% agreement. The researcher believes that one reason for the high accuracy in this pose is its similarity to a previous pose, which provided the volunteers with a better understanding. Additionally, this pose is easy to follow because it goes against the principles of body movement, resulting in slower and more controlled motions, making it easier for participants to mimic. In Table 4.16, which presents the confusion matrix comparing the judgments of experts and the system, there is only one discrepancy in the decision.

4.3.6 Discussion

The study aimed to develop and evaluate a prototype system for frozen shoulder patients, leveraging motion detection techniques based on landmark analysis to compare user-generated videos with original recordings by physical therapy experts. The analysis of different shoulder postures—flexion, abduction, external rotation, and internal rotation—provided valuable insights into the system's effectiveness and areas for improvement. For shoulder flexion, the cosine similarity algorithm labeled 10 out of 14 instances as "Similar" and 4 as "Not Similar," compared to the expert evaluations which labeled 13 as "Similar" and 1 as "Not Similar." The performance metrics, including a precision of 1.0, recall of 0.769, accuracy of 0.786, and an F1-score of 0.870, demonstrate high precision but lower recall, indicating that while the algorithm is excellent at correctly identifying similarities, it misses some instances identified by the expert. In shoulder abduction, the algorithm labeled 13 out of 14 instances as "Similar" and 1 as "Not Similar," while experts labeled all instances as "Similar." The resulting metrics—precision of 1.0, recall of 0.929, accuracy of 0.929, and an F1-score of 0.963—show strong algorithm reliability, though it missed one instance. For shoulder external rotation, the algorithm labeled 11 instances as "Similar" and 3 as "Not Similar," compared to all "Similar" labels by experts. The metrics, with a precision of 1.0, recall of 0.786, accuracy of 0.786, and an F1-score of 0.88, indicate high precision but the need for better sensitivity to match expert-identified similarities. In shoulder internal rotation, both the algorithm and experts labeled all 14 instances as "Similar," achieving perfect metrics—precision, recall, accuracy, and F1-score of 1.0—demonstrating exceptional algorithm performance in this posture.

This analysis maps directly to the study's objectives. The prototype system effectively utilized motion detection techniques, comparing user-generated videos with expert recordings, fulfilling objective 1.4.1. The evaluation of the system's effectiveness in practical applications, specifically in motion detection and angle measurement, showed high precision and accuracy, though with varying degrees of success across different postures, addressing objective 1.4.2. The study also explored the integration of computer technology into physical therapy practices, using motion capture and similarity comparisons, validating the feasibility of these technologies in aiding physical therapy, thus achieving objective 1.4.3. Lastly, the findings indicate that the prototype system can facilitate effective at-home physical therapy for frozen shoulder patients, enabling reliable feedback on posture correctness, aligning with objective 1.4.4. However, it is vital to note that expert evaluations can be subjective and may introduce biases, especially when relying on a single expert. Future research should involve multiple experts to reduce bias and increase the robustness of comparison analysis. This would ensure a more accurate and reliable system for at-home therapy for frozen shoulder patients.

4.4 Evaluation of face detection model results and discussion

This experiment investigates a group of 14 healthy volunteers aged between 20 and 50 years, comprising an equal distribution of seven males and seven females. The study involves capturing video footage of participants prior to the experimental intervention, specifically while they follow a demonstrated exercise video. Using face detection techniques, the research team extracts facial images from the video. These images are then processed to identify facial landmarks using the Mediapipe framework. The landmark data is subsequently fed into a developed classification model to determine whether the participants exhibit a "hurt" or "normal" facial expression. The model used for face classification is derived from the work of Srikaewsiew and Kanjanawattana (2024), which has demonstrated outstanding accuracy with a score of 0.95. Consequently, the results obtained are highly accurate. Additionally, the classification results from the face detection model are consistent with the feedback collected from the participants via a Google Form, further validating the model's accuracy and reliability in this context.

	Predicted Normal	Predicted Hurt
Normal	1120	0
Hurt	0	0

Table 4.17 Confusion Matrix between Feedback Result (Actual) and Developed Model Result (Predict)

4.4.1 Results explanation

From Table 4.17, the confusion matrix reveals that the classification model has achieved perfect performance on the Normal class but has no performance metrics for the Hurt class due to the absence of actual Hurt instances in the dataset. Specifically, the model correctly predicted all 1120 Normal instances, resulting in an accuracy of 100%. Both the precision and recall for the Normal class are 100%, indicating flawless classification for this category. However, the metrics for the Hurt class, including precision, recall, and F1 score, are undefined because the dataset contains no actual Hurt instances.

4.4.2 Discussion

The fundamental purpose of this experiment was to examine the integration of computer technology into physical therapy processes, primarily focused on the development of a prototype system for frozen shoulder patients utilizing motion detection techniques based on landmark analysis. By capturing and analyzing videos of participants following a demonstrated exercise video, the study tried to test the performance of face detection and classification systems. The experiment effectively showed the possibility of employing these methods in a real environment. The classification techniques, based on the work of Srikaewsiew and Kanjanawattana (2024), demonstrated well exact results, matching with participant feedback obtained via Google Forms. This alignment highlights the potential of computer-assisted physical therapy to enhance the accuracy and efficacy of home-based rehabilitation exercises.

The findings of this experiment indicated that the face detection and classification model could accurately identify "normal" and "hurt" expressions in the participants, obtaining a perfect classification rate for the normal class. This accuracy illustrates the model's robustness and the dependability of the landmark extraction technique utilizing Mediapipe. The full agreement between the model's results and participant comments further verifies the model's performance. However, the confusion matrix revealed that there were no actual instances of the "hurt" class in the dataset, which limits the evaluation of the model's efficacy in distinguishing this specific condition. This was due to the fact that the participants reported no hurt feelings in any posture, as reflected in their feedback on Google Forms. Consequently, the metrics for the hurt class, including precision, recall, and F1 score, are undefined because the dataset contains no actual hurt instances. Despite this, the good accuracy for the typical class indicates the system's potential effectiveness in actual applications.

One important problem noted during the experiment was the model's reduced confidence when identifying images of people wearing glasses. This illustrates that the presence of spectacles might interfere with the landmark detection process, resulting in lower classification accuracy. This issue highlights a need for extra development of the model to handle such changes in facial expression effectively. Moreover, the absence of hurt incidents in the dataset underlines the requirement for a more balanced dataset to thoroughly assess the model's effectiveness across diverse situations. Future study should focus on addressing these difficulties to enhance the model's generalizability and provide dependable performance across different participant characteristics.

4.5 Evaluation of shoulder angle results and discussion

This experiment investigates shoulder angles in a cohort of 14 healthy volunteers aged between 20 and 50 years, with an equal distribution of seven males and seven females. The study examines four specific shoulder poses: Shoulder Flexion, Abduction, Shoulder External Rotation, and Shoulder Internal Rotation, evaluated at three stages—initial posture, midpoint posture, and peak posture.

The shoulder angles were measured across these poses using both general actual angles and clinical actual angles, evaluated by experts. Additionally, predictive angles were generated using a concurrently developed program. The aggregated results from these measurements are presented and compared in Table 4.18.

4.5.1 Results explanation

From the Table 4.18, it reveals notable patterns when comparing general actual angles, clinical actual angles, and predicted angles across both genders. For shoulder flexion, females exhibit a range from an initial angle of 16.43° (SD 3.92°)

to a finishing angle of 172.57° (SD 5.94°), while males demonstrate a similar trend from 17.00° (SD 1.83°) to 170.00° (SD 5.23°). Predicted angles consistently underestimate these positions, with females starting at 9.57° (SD 1.27°) and males at 8.29° (SD 1.38°), suggesting potential limitations in the predictive model's accuracy for initial joint positions. Clinical actual angles closely align with general actual angles, indicating robust consistency across measurement methodologies.

In abduction movements, females display angles ranging from 16.29° (SD 2.27°) to 174.14° (SD 4.34°), and males from 17.71° (SD 0.95°) to 174.00° (SD 4.28°). Predicted angles once again indicate lower starting positions (9.14° for females, 8.16° for males) but comparable finishing angles. Notably, shoulder external rotation exhibits more pronounced variability, with females ranging from 90.14° (SD 2.23°) to 175.00° (SD 3.64°) and males from 92.43° (SD 2.37°) to 174.71° (SD 3.82°). Predicted angles diverge notably, particularly at the start (84.57° for females, 89.43° for males) and mid-range, reflecting challenges in accurately predicting these movements.

Internal rotation angles demonstrate a narrower range and higher variability, with females ranging from 91.14° (SD 2.23°) to 26.86° (SD 9.29°) and males from 91.14° (SD 2.19°) to 32.57° (SD 16.89°). Predicted angles reveal significant disparities, especially at the finishing position (10.86° for females, 16.86° for males), highlighting the complexity of accurately predicting these intricate movements.

4.5.2 General Actual Angle vs. Predicted Angle

In shoulder flexion, both genders consistently exhibit general actual angles lower than predicted. Notably, females' initial actual angle (16.43°) contrasts markedly with their predicted angle (9.57°), while males show a higher initial actual angle (17.00°) compared to their predicted angle (8.29°).

For abduction, a significant disparity exists at the initial posture, with general actual angles (16.29° for females, 17.71° for males) exceeding predicted angles (9.14° and 8.16° respectively).

External rotation generally shows lower general actual angles than predicted, particularly evident at the midpoint (e.g., females: actual 131.71° vs. predicted 133.00°).

Internal rotation displays varied results, with instances of both higher and lower general actual angles compared to predicted angles across different postures.

4.5.3 Clinical Actual Angle vs. Predicted Angle

Clinical actual angles in shoulder flexion tend to closely align with predicted angles compared to general actual angles. For example, females' clinical peak angle (171.29°) approaches the predicted angle (173.71°), contrasting with their general actual angle (172.57°).

In abduction, clinical measurements closely approximate predicted angles, notably in midpoint and peak postures. For instance, males' clinical peak abduction (172.86°) closely matches the predicted angle (174.86°).

External rotation clinical angles exhibit smaller variations from predicted angles compared to general actual angles.

Clinical angles in internal rotation generally align closer to predicted angles than general actual angles, particularly in midpoint and peak postures.

Table 4.18 Comparison of Shoulder Joint Angles: General Actual, Clinical Actual, and Predicted Angles for Females and Males

	General actual angle				Clinical actual angle				Predict angle			
	Female		Male		Female		Male		Female		Male	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Shoulder flexion(initial posture)	16.43	3.92	17.00	1.83	17.00	3.92	16.43	1.13	9.57	1.27	8.29	1.38
Shoulder flexion(midpoint posture)	90.86	2.98	89.71	1.70	90.71	2.98	91.14	3.29	88.14	6.67	89.43	3.60
Shoulder flexion(peak posture)	172.57	5.94	170.00	5.23	171.29	5.94	167.57	4.54	173.71	4.46	171.79	5.48
Abduction(initial posture)	16.29	2.27	17.71	0.95	16.86	2.27	16.71	1.70	9.14	1.86	8.16	0.40
Abduction(midpoint posture)	93.14	2.29	91.71	1.60	91.29	2.29	91.43	2.51	94.29	3.40	91.57	3.87
Abduction(peak posture)	174.14	4.34	174.00	4.28	172.86	4.34	172.86	5.98	173.71	3.68	174.86	3.24
Shoulder external rotation(initial posture)	90.14	2.23	92.43	2.37	89.43	2.23	93.00	3.65	84.57	11.47	89.43	13.40
Shoulder external rotation(midpoint posture)	131.71	3.13	139.14	6.47	129.86	3.13	140.00	8.23	133.00	9.15	145.29	6.73
Shoulder external rotation(peak posture)	175.00	3.64	174.71	3.82	175.29	3.64	174.71	3.82	171.57	5.32	170.14	4.53
Shoulder internal rotation(initial posture)	91.14	2.23	91.14	2.19	90.43	2.23	92.14	3.93	86.57	6.60	83.43	6.27
Shoulder internal rotation(midpoint posture)	58.71	7.64	57.00	12.62	61.00	7.64	57.00	11.28	60.43	11.80	51.57	15.44
Shoulder internal rotation(peak posture)	26.86	9.29	32.57	16.89	26.00	9.29	33.00	16.69	10.86	7.40	16.86	23.94

4.5.5 Discussion

The study successfully developed and evaluated an initial system for patients with frozen shoulder, using motion detection techniques based on landmark analysis. This method aimed to enhance the accuracy of motion identification and angle measurement by comparing user-generated videos with recordings by physical therapy experts. The integration of computer technologies, such as facial detection and motion capture using computer vision, enabled precise degree measurements and facilitated comparisons with expert videos. This methodology represents significant progress in leveraging technology to improve physical therapy procedures, particularly for aiding patients with frozen shoulder in performing exercises at home.

The experiment focused on analyzing shoulder angles across various postures among a group of 14 healthy volunteers. The results revealed considerable discrepancies among the measured angles in shoulder flexion, abduction, external rotation, and internal rotation motions, notably between general actual, clinical actual, and predicted angles. While clinical actual angles closely agreed with predicted angles, general actual angles typically revealed initial variations that resolved towards similar peak angles. This gap was notably obvious in external and internal rotation movements, showing difficulty in specifically measuring these complex motions using current computer models.

Despite the favorable outcomes, the study revealed several challenges and recommendations. One notable issue was the impact of clothing edges on MediaPipe recognition accuracy, leading to erroneous angle estimates. Participants wearing loose clothing obscured body parts crucial for measurement, highlighting the importance of form-fitting attire in future studies to avoid detection errors. Additionally, the study noted the limitations of having only two researchers oversee measurements, which was insufficient for ensuring consistent accuracy and rapid troubleshooting. Future research could benefit from deploying additional personnel to enhance oversight and maintain measurement precision throughout the evaluation process.

CHAPTER V

CONCLUSIONS

5.1 Conclusion

This study represents a significant advancement in the development and evaluation of a landmark-based motion detection system tailored for improving frozen shoulder rehabilitation. Integrating motion analysis, facial expression recognition, and shoulder angle measurement, our multi-faceted approach underscores the potential of advanced technologies in physical therapy.

The comparative analysis between user-generated videos and expert recordings yielded promising outcomes, demonstrating high precision in identifying similarities across shoulder postures. Particularly, the cosine similarity algorithm achieved notable accuracy in abduction and internal rotation exercises, aligning closely with expert evaluations. However, variability in recall rates across different movements suggests areas for refinement to enhance sensitivity to expert-identified similarities.

Facial expression recognition, leveraging the MediaPipe framework and a custom classification model, exhibited impressive accuracy in detecting normal expressions, which corroborates participant feedback on its potential for real-time assessment of patient comfort during exercises. Challenges such as the absence of "hurt" expressions in the dataset and issues with glasses wearers underscore the necessity for more diverse training data and robust feature extraction methods.

Analysis of shoulder angle measurements revealed nuanced patterns across general actual, clinical actual, and predicted angles. While the system demonstrated promise in approximating clinical measurements, particularly in flexion and abduction movements, discrepancies in external and internal rotation angles underscore the challenges in accurately modeling these motions. In summary, this research successfully developed a prototype system integrating motion detection, facial recognition, and angle measurement technologies to enhance frozen shoulder rehabilitation. These findings provide valuable insights into technology-assisted physical therapy, paving the way for personalized and effective at-home rehabilitation programs.

5.2 Limitations

Several limitations of the present study must be acknowledged:

5.2.1 Sample Size and Diversity: The study was conducted with a limited number of participants, potentially limiting the generalizability of the findings.

5.2.2 Environmental Factors: The impact of various environmental conditions, such as lighting and background, on the system's performance was not comprehensively explored.

5.2.3 Clothing Interference: Loose clothing was found to interfere with accurate landmark detection, potentially skewing results.

5.2.4 Limited Personnel: Having only two researchers during the assessment phase limited the ability to provide comprehensive supervision and immediate troubleshooting.

5.2.5 Single Expert Evaluation: Relying on a single expert for clinical measurements may have introduced potential bias or limited the robustness of the comparison.

5.2.6 Focus on Static Postures: While the study examined three stages of movement, it may not fully capture the dynamics of continuous motion.

5.3 Suggestions for future research

Based on the findings and limitations of this study, the following suggestions are proposed for future research:

5.3.1 Expanded Participant Pool: Future studies should include a larger, more diverse group of participants to enhance the generalizability of findings.

5.3.2 Dynamic Movement Analysis: Develop methods to assess continuous shoulder movements rather than focusing solely on static postures.

5.3.3 Multi-Expert Validation: Incorporate assessments from multiple clinical experts to establish a more robust ground truth for comparisons.

5.3.4 Environmental Testing: Evaluate the system's performance under various lighting conditions and backgrounds to assess its reliability in different settings.

5.3.5 Clothing Standardization: Develop and test standardized clothing protocols to minimize interference with landmark detection.

5.3.6 Enhanced Personnel Training: Increase the number of trained personnel involved in data collection and assessment to improve accuracy and troubleshooting capabilities.

5.3.7 Integration of Additional Metrics: Explore incorporating other relevant measurements (e.g., range of motion, movement speed) to provide a more comprehensive analysis of shoulder function.

5.3.8 Long-term Reliability Testing: Conduct longitudinal studies to assess the system's consistency and reliability over time.

5.3.9 Application-Specific Refinement: Tailor the system for specific applications (e.g., post-operative rehabilitation, sports-specific movements) and evaluate its effectiveness in these contexts.

5.3.10 Machine Learning Enhancements: Explore the use of advanced machine learning techniques to improve the system's accuracy, particularly for challenging initial and transitional movements.

REFERENCE

Abdulghani, M. M., Ghazal, M. T., & Salih, A. B. M. (2023). Discover human poses similarity and action recognition based on machine learning. Bulletin of Electrical Engineering and Informatics, 12(3), 1570-1577.

Ahmad, K. A. B., Khujamatov, H., Akhmedov, N., Bajuri, M. Y., Ahmad, M. N., & Ahmadian, A. (2022). Emerging trends and evolutions for Smart city healthcare systems. Sustainable Cities and Society, 80, 103695.

Albahri, O. S., Zaidan, A. A., Zaidan, B. B., Hashim, M., Albahri, A. S., & Alsalem, M. A. (2018). Real-time remote health-monitoring Systems in a Medical Centre: A review of the provision of healthcare services-based body sensor information, open challenges and methodological aspects. Journal of medical systems, 42, 1-47.

Assari, M. A., & Rahmati, M. (2011, November). Driver drowsiness detection using face expression recognition. In 2011 IEEE international conference on signal and image processing applications (ICSIPA) (pp. 337-341). IEEE.

Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M. (2020). Blazepose: On-device real-time body pose tracking. arXiv preprint arXiv:2006.10204.

Briggs, A. M., Shiffman, J., Shawar, Y. R., Åkesson, K., Ali, N., & Woolf, A. D. (2020). Global health policy in the 21st century: challenges and opportunities to arrest the global disability burden from musculoskeletal health conditions. Best Practice & Research Clinical Rheumatology, 34(5), 101549.

Bruckner, F. E., & Nye, C. J. S. (1981). A prospective study of adhesive capsulitis of the shoulder (‘frozen shoulder’) in a high risk population. QJM: An International Journal of Medicine, 50(2), 191-204.

Bruns, K. (2017). Angle between two vectors 2d [Online Image]. In https://wumbo.net/formulas/angle-between-two-vectors-2d/. https://wumbo.net/formulas/angle-between-two-vectors-2d/main-360-210.light.svg

Challoumas, D., Biddle, M., McLean, M., & Millar, N. L. (2020). Comparison of treatments for frozen shoulder: a systematic review and meta-analysis. JAMA network open, 3(12), e2029581-e2029581.

Choi, Y., Nam, J., Yang, D., Jung, W., Lee, H. R., & Kim, S. H. (2019). Effect of smartphone application-supported self-rehabilitation for frozen shoulder: a prospective randomized control study. Clinical rehabilitation, 33(4), 653-660.

Cooper, R. A., & Cooper, R. (2019). Rehabilitation Engineering: A perspective on the past 40-years and thoughts for the future. Medical engineering & physics, 72, 3-12.

De la Serna, Daniel, Santiago Navarro-Ledesma, Fany Alayón, Elena López, and Leo Pruimboom. "A comprehensive view of frozen shoulder: a mystery syndrome." Frontiers in Medicine 8 (2021): 638.

Desmarais, Y., Mottet, D., Slangen, P., & Montesinos, P. (2021). A review of 3D human pose estimation algorithms for markerless motion capture. Computer Vision and Image Understanding, 212, 103275.

Di Luzio, F., Rosato, A., & Panella, M. (2023). A randomized deep neural network for emotion recognition with landmarks detection. Biomedical Signal Processing and Control, 81, 104418.

Dias, R., Cutts, S., & Massoud, S. (2005). Frozen shoulder. Bmj, 331(7530), 1453-1456.

Erickson, B. J., Shishani, Y., & Gobezie, R. (2023). Remote Patient Monitoring of Postoperative Rehabilitation. Physical Medicine and Rehabilitation Clinics, 34(2), 489-497.

Fernandes, G., Gonçalves, D. A., & Conti, P. (2018). Musculoskeletal disorders. Dental Clinics, 62(4), 553-564.

Fried, I., Hoelscher, J., Akulian, J. A., Pizer, S., & Alterovitz, R. (2023, October). Landmark Based Bronchoscope Localization for Needle Insertion Under Respiratory Deformation. In 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 6593-6600). IEEE.

Gava, V., Ribeiro, L. P., Barreto, R. P. G., & Camargo, P. R. (2022). Effectiveness of physical therapy given by telerehabilitation on pain and disability of individuals with shoulder pain: a systematic review. Clinical rehabilitation, 36(6), 715-725.

Google Developers. (2020). Mediapipe face landmark [Online Image]. In *https://developers.google.com/mediapipe/solutions/vision/face_landmarker*. https://storage.googleapis.com/mediapipe-assets/documentation/mediapipe_face_landmark_fullsize.png

Hamester, D., Barros, P., & Wermter, S. (2015, July). Face expression recognition with a 2-channel convolutional neural network. In 2015 international joint conference on neural networks (IJCNN) (pp. 1-8). IEEE.

Hangaragi, S., Singh, T., & Neelima, N. (2023). Face detection and Recognition using Face Mesh and deep neural network. Procedia Computer Science, 218, 741-749.

Latreche, A., Kelaiaia, R., Chemori, A., & Kerboua, A. (2023). Reliability and validity analysis of MediaPipe-based measurement system for some human rehabilitation motions. Measurement, 214, 112826.

Lyne, S. A., Goldblatt, F. M., & Shanahan, E. M. (2022). Living with a frozen shoulder–a phenomenological inquiry. BMC Musculoskeletal Disorders, 23(1), 1-7.

Mende-Siedlecki, P., Qu-Lee, J., Lin, J., Drain, A., & Goharzad, A. (2020). The Delaware Pain Database: a set of painful expressions and corresponding norming data. Pain reports, 5(6).

Mertens, M. G., Meeus, M., Verborgt, O., Vermeulen, E. H., Schuitemaker, R., Hekman, K. M., ... & Struyf, F. (2022). An overview of effective and potential new conservative interventions in patients with frozen shoulder. Rheumatology International, 42(6), 925-936.

Munasinghe, M. I. N. P. (2018, June). Facial expression recognition using facial landmarks and random forest classifier. In 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS) (pp. 423-427). IEEE.

Pandey, V., & Madi, S. (2021). Clinical guidelines in the management of frozen shoulder: an update!. Indian journal of orthopaedics, 55(2), 299-309.

Pauzi, A. S. B., Mohd Nazri, F. B., Sani, S., Bataineh, A. M., Hisyam, M. N., Jaafar, M. H., ... & Mohamed, A. S. A. (2021). Movement estimation using mediapipe blazepose. In Advances in Visual Informatics: 7th International Visual Informatics Conference, IVIC 2021, Kajang, Malaysia, November 23–25, 2021, Proceedings 7 (pp. 562-571). Springer International Publishing.

Revina, I. M., & Emmanuel, W. S. (2021). A survey on human face expression recognition techniques. Journal of King Saud University-Computer and Information Sciences, 33(6), 619-628.

Rizk, T. E., & Pinals, R. S. (1982, May). Frozen shoulder. In Seminars in arthritis and rheumatism (Vol. 11, No. 4, pp. 440-452). WB Saunders.

Roggio, F., Ravalli, S., Maugeri, G., Bianco, A., Palma, A., Di Rosa, M., & Musumeci, G. (2021). Technological advancements in the analysis of human motion and posture management through digital devices. World Journal of Orthopedics, 12(7), 467.

Singh, A. K., Kumbhare, V. A., & Arthi, K. (2021, June). Real-time human pose detection and recognition using mediapipe. In International Conference on Soft Computing and Signal Processing (pp. 145-154). Singapore: Springer Nature Singapore.

Srikaewsiew, T., & Kanjanawattana, S. (2024). Comparative Analysis of Facial Expression Recognition: Image-Based vs. Landmark-Based Approaches. In The 9th International Conference on Advanced Technology Innovation 2024 (ICATI2024). (Manuscript waiting for publication).

Srikaewsiew, T., Khianchainat, K., Tharatipyakul, A., Pongnumkul, S., & Kanjanawattana, S. (2022, December). A Comparison of the Instructor-Trainee Dance Dataset Using Cosine similarity, Euclidean distance, and Angular difference. In 2022 26th International Computer Science and Engineering Conference (ICSEC) (pp. 235-240). IEEE.

Tasto, J. P., & Elias, D. W. (2007). Adhesive capsulitis. Sports medicine and arthroscopy review, 15(4), 216-221.

Tharatipyakul, A., & Pongnumkul, S. (2023). Deep Learning-based Pose Estimation in Providing Feedback for Physical Movement: a Review. Preprints. https://doi.org/10.20944/preprints202306.0395.v1

The Goniometer. (2012). https://www.google.co.th/url?sa=i&url=https%3A%2F%2Fwww.youtube.com%2Fwatch%3Fv%3DZUF7tpkVAIY&psig=AOvVaw3PSYyLPH4XoemuAHI_NEjI&ust=1719379174626000&source=images&cd=vfe&opi=89978449&ved=0CBEQjRxqFwoTCPiN59qA9oYDFQAAAAAdAAAAABA.

APPENDIX A

LIST OF PUBLICATIONS

LIST OF PUBLICATIONS

Srikaewsiew, T., & Kanjanawattana, S. (2024). Comparative Analysis of Facial Expression Recognition: Image-Based vs. Landmark-Based Approaches. In The 9th International Conference on Advanced Technology Innovation 2024 (ICATI2024). (Manuscript waiting for publication).

Srikaewsiew, T., Khianchainat, K., Tharatipyakul, A., Pongnumkul, S., & Kanjanawattana, S. (2022, December). A Comparison of the Instructor-Trainee Dance Dataset Using Cosine similarity, Euclidean distance, and Angular difference. In 2022 26th International Computer Science and Engineering Conference (ICSEC) (pp. 235-240). IEEE.

The 9th International Conference on Advanced Technology Innovation 2024 (ICATI2024)

Comparative Analysis of Facial Expression Recognition: Image-Based vs. Landmark-Based Approaches

Thanawat Srikaewsiew [1], Sarunya Kanjanawattana [2,*]

[1] *School of Computer Engineering, Institute of Engineering, Suranaree University of Technology, Thailand*

[2] *School of Computer Engineering, Institute of Engineering, Suranaree University of Technology, Thailand*

Received 15 June 20xx; received in revised form 05 August 20xx; accepted 10 September 20xx

Abstract

Facial expression recognition plays a crucial role in human-computer interaction. This study aimed to compare image-based and landmark-based learning methods to gain a deeper understanding of these techniques. Various algorithms, including Convolutional Neural Networks (CNN), Support Vector Machines (SVM), Random Forest Classification (RFC), Logistic Regression Classification (LRC), and Gradient Boosting Classifier (GBC), were leveraged to investigate their performance aspects. Image-based learning, implemented by CNN, specialized in acquiring global and local facial data, while landmark-based learning focused on key facial points. The results demonstrated that CNN achieved an outstanding accuracy of 0.95, particularly with the landmark technique. SVM displayed proficiency with landmarks, while GBC and RFC exhibited robust results. LRC, known for its efficiency in training, varied in performance. Datasets from The Delaware Pain Database and UTKFace were utilized in this study to provide insights into the specifics of face emotion recognition. The implications of the findings extended beyond the study's primary emphasis, suggesting promise for applications such as assisting in physical rehabilitation. Future initiatives aimed to leverage these findings to enhance safety and facilitate rapid injury diagnosis during physical therapy sessions. This study not only advanced our understanding of facial emotion identification but also carried practical implications for the development of emotionally intelligent systems.

Keywords: Facial expression recognition, Face landmark, Machine learning, Image classification, Computer Vision

1. Introduction

Facial expression recognition [1] is a captivating field at the intersection of computer vision [2], artificial intelligence [3], and psychology [4], attempting to figure out the complex language of human emotions expressed through facial expressions.

Facial expressions have long been recognized as a basic method for subtle communication [5], helping humans to show their emotions, motives, and perceptions. Knowing how to interpret these expressions come effortlessly to humans however remain an important obstacle for machines [6]. As our reliance on human-computer interaction intensifies, the requirement for emotionally intelligent technologies that are capable of recognizing and responding to human emotions grows more and more obvious. The image-based method depends on the transformative capabilities of deep learning [7], with Convolutional Neural Networks (CNN) at its vanguard. CNN have changed computer vision tasks by effectively learning patterns and spatial

* Corresponding author. E-mail address: Sarunya.k@sut.ac.th

dependencies from huge quantities of data. By teaching CNN on large-scale datasets, such as The Delaware Pain Database [8] and UTKFace, image-based models become efficient at capturing both global and local data from facial images, enabling them to recognize and identify hidden facial data illustrative of various emotions. In contrast, face landmark-based learning [9] concentrates on key facial points, or landmarks, such as the outer edges of the mouth, nose, and eyes. These landmarks serve as instructive representations of facial expressions, regardless of particular facial features, poses, or illumination conditions. Landmark-based methods, frequently utilizing geometric features or handcrafted descriptors, represent the spatial relationships between these facial landmarks, providing robust and comprehensible models for emotion classification.

This research focused on a comprehensive comparative analysis of two different techniques for facial expression recognition: image-based learning [10] and landmark-based learning [11]. By leveraging the abilities of CNN [12], Logistic Regression Classification (LRC) [13], Support Vector Machines (SVM) [14], Random Forest Classification (RFC) [15], and Gradient Boosting Classifier (GBC) [16], we provided recommendations for choosing the most suitable approach to obtain accurate and reliable emotion recognition.

This research aimed to study the advantages and limitations of both image-based and landmark-based learning techniques by conducting comprehensive comparisons. Utilizing the rich datasets from the Delaware Pain Database and UTKFace, we attempted to fully analyze the effectiveness of various classification models, including LRC, SVM, RFC, and GBC, when applied to both methods. By researching the strengths and weaknesses of each methodology, we aim to provide researchers and developers with helpful recommendations for choosing the most appropriate technique for facial expression recognition in various fields. The knowledge achieved from this study demonstrates an opportunity to advance emotionally intelligent systems [17], enabling them to enhance human-computer interaction [18], virtual reality experiences [19], and mental health diagnostics [20].

2. Literature review

Face expression recognition has been researched for several years in the areas of computer science and computer engineering. We briefly reviewed an academic work released on the topic.

In their introductory study, Di Luzio et al. presented a randomized deep neural network for emotion recognition with landmark detection [21]. They applied The Extended Cohn-Kanade dataset (CK+) and Mediapipe to extract 468 face landmarks. The model combined a randomized convolutional layer with an Long Short-Term Memory (LSTM) layer, receiving over 90% accuracy for five emotions: disgust, fear, happiness, sadness, and surprise. This work improved emotion recognition and demonstrated the possibilities of deep learning and landmark detection in affective computing and human-machine interaction. To overcome the limitations in the field of face recognition, Hangaragi et al. proposed face detection and recognition using a face mesh and deep neural network [22]. This study utilized a Labeled Faces in the Wild (LFW) dataset and images captured in real-time by using Face Mesh to reconstruct the complete face with face landmarks and a deep learning model. They compared this model to 3DMM (LFW), 3DDFA (LFW), and 3DMM-CNN (LFW) and got a superior result in terms of accuracy. The proposed model achieved 94.23% accuracy, as well as The model detected and recognized faces in various illuminations and non-frontal images efficiently, which other existing algorithms failed to do.

In [23], Hamester et al. proposed a 2-Channel CNNs for recognizing facial expressions. The design included two channels, one to handle the raw image data and the other for processing the output of a Convolutional Autoencoder (CAE). The CAE was trained in an unsupervised fashion to extract features from the input image, which were subsequently merged with the raw image data in the second channel. The combined features then got transmitted into a completely connected layer for classification. They evaluated this method on the JAFFE dataset while comparing it to previously released methods. They demonstrated that their technique surpassed these methods in the area of accuracy. The algorithm that was suggested obtained an average accuracy of 95.8% with a standard deviation of 1.6 on the JAFFE dataset. Face expression can provide additional stability in our daily lives. Assari et al. attempted to utilize a non-intrusive technique for detecting driver drowsiness through facial expressions [24]. The proposed method employed an infrared light-sensitive camera to capture images of the driver's face, which were processed in order to detect facial features such as eye expressions, mouth openness, and eyebrow elevation. If any of these states were stable for a certain length of time, a notification appeared displaying the driver's drowsiness status. The suggested method was tested in a real-life driving case using images obtained under various lighting conditions and from multiple individuals with varied appearances. The photos went through processing at a frame rate of 20 frames per second with a resolution of 360 x 240 pixels. The outcomes demonstrated that the proposed method was efficient at recognizing drowsiness in drivers, with an outstanding level of accuracy and a low false-positive rate. The method was also compared to several other non-intrusive drowsiness detection methods and proved to perform better in both accuracy and robustness. In [25], Munasinghe et al. presented a technique that involves recognizing facial landmarks and employing them to calculate a feature vector representing emotion in the face. The feature vector is obtained by calculating the distance among sets of landmarks and normalizing them to eliminate facial size differences. Once the feature vector was calculated, it served as input to a random forest classifier that was trained to classify expressions. The researchers applied the dlib library to locate 68 facial landmarks as well as the scikit-learn library to implement the random forest classifier. The accuracy of the method was evaluated via the Extended Cohn-Kanade (CK+) database, which was a commonly used facial expression database. This study used a total of 225 different poses coming from 104 individual people, with 156 poses implemented as training datasets and 69 poses used as testing datasets. The outcomes of the study demonstrated that the proposed technique was successful in correctly detecting emotions, with an average accuracy rate of 90%. The accuracy rates for individual expression were 79% for anger, 95% for happiness, 89% for sadness, and 96% for surprise. The suggested method outperformed current methods, acquiring an average performance rate of 90% compared to 80% for Omer et al. [26] and 72–100% for Akram et al. [27].

Based on the above research, it was obvious that the topic could be properly divided into two primary technique: an image-based technique and face-landmark technique. Additionally, it was essential to note that the source of both datasets came from image sources. Moreover, it was important to acknowledge that the model commonly applied CNN, SVM, many different algorithms belonging to the Tree family, as well as other fundamental machine learning techniques. As a result, the primary goal of this study was a comprehensive study and comparative analysis of the performance demonstrated through the two techniques and commonly used machine learning classification. This research effort was particularly focused on finding an effective method for facial expression and making recommendations for developers in the future.

3. Machine learning techniques

The following section provides a succinct overview of the machine learning algorithms compared in this study.

3.1. Convolutional Neural Networks (CNN)

CNN is a deep learning model that has significantly transformed the field of computer vision. It is primarily designed to process and evaluate grid-like data, such as photographs. CNNs are built using convolutional layers that convolve input with learnable filters, capturing local patterns hierarchically. Subsequent pooling layers down-sample the representations while keeping crucial characteristics. This architecture intrinsically respects spatial connections, which is vital in visual data interpretation. CNNs commonly employ non-linear activations and regularization approaches to boost representation learning and control over-fitting. With several layers, they exhibit increasing feature abstraction. Fully connected layers towards the end merge these features for categorization or other tasks. Overall, CNNs have automated feature engineering and delivered state-of-the-art performance across image-related applications, making them a popular choice for academic research in the field of computer vision.

3.2. Random Forest Classification (RFC)

The Random Forest technique was initially introduced in 2001 and has since become a widely used ensemble learning approach for tasks involving regression as well as classification. This algorithm is particularly effective due to its use of multiple decision trees that utilize both bootstrapped aggregation (bagging) and random feature selection. These techniques assist to minimize overfitting and enhance predictive capabilities. In classification assignments, the final judgment is made based on a majority vote, while regression predictions are based on the average of outcomes. RFC is a fantastic choice for high-dimensional data and can even determine feature importance, making it an immensely valuable tool for data analysis.

3.3. Logistic Regression Classification (LRC)

Logistic Regression is a frequently used binary classification approach in machine learning. It is a probabilistic classification approach that assesses the chance that an instance belongs to a certain class (typically 0 or 1). Despite its name, it is not a regression technique. It employs the logistic function to build the relationship between input attributes and the likelihood of the target class, which transforms any input to a number between 0 and 1. By setting a threshold (typically 0.5), predictions are made: values over the threshold are classed as one class, and those below it as the other. It is a basic yet efficient strategy for problems with virtually linear connections between characteristics and classes. Regularization techniques can be applied to prevent over-fitting.

3.4. Support Vector Machine (SVM)

SVM is a commonly used generalized linear classification algorithm that may also be applied to regression situations [28]. Its name originates from its capacity to maximize the geometric margin while minimizing classification mistakes, making it a popular choice for many applications. To do this, SVMs apply Structural Risk Minimization (SRM), which helps to optimize the separation between distinct classes of data points. This is done by constructing parallel hyperplanes on either side of the decision boundary, with a greater margin assuring better generalization of the model.

3.5. Gradient Boosting Classifier (GBC)

Gradient boosting is a commonly used methods for machine learning that could get effective used for both classification and regression purposes. This approach functions by adding decision trees iteratively into a model, where each tree rectifies the mistakes of the preceding trees. This method repeats until the model achieves an acceptable level of accuracy. One of the main benefits of gradient boosting is being able to deal with complicated datasets and find non-linear relationships between features. However, it is sensitive to overfitting, so it requires comprehensive adjustment of hyperparameters.

4. Experiment

4.1. Dataset

In this study, we collected image data from publically available sources for research usage, including The Delaware Pain Database and UTKFace, two of which are human image databases that have collected a huge number of photographs. It contained many emotional characteristics we selected faces with normal emotions. and hurt emotions, totaling 1200 images, divided between 800 images for training, 200 images for testing, and 200 images for further tests. The examples of image data are as illustrated in Fig. 1 and Fig. 2.

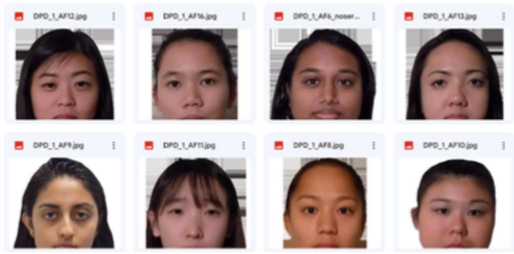


Fig. 1 Example of normal face image dataset from The Delaware Pain Database and UTKFace

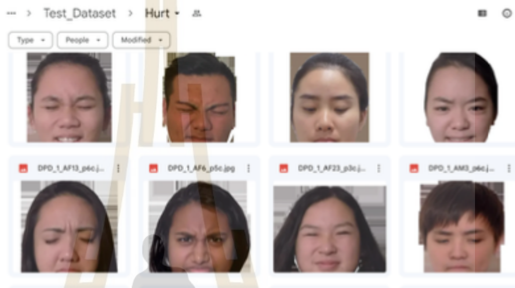


Fig. 2 Example of hurt face image dataset from The Delaware Pain Database and UTKFace

4.2. Image-based method

In the Image-based method experiment, we chose five algorithms including CNN, LRC, SVM, RFC, and GBC. In terms of LRC, SVM, RFC, and GBC, we utilized the library of scikit-learn to employ these techniques, although these algorithms may not work directly with images. Therefore, it is important for converting images into numerical data. To manipulate data to suit the algorithm, we employed the Histogram of Oriented Gradients (HOG) feature extraction technique.

In our study, various functions were constructed, including the extraction of features. Each image underwent processing by converting it to gray scale, scaling it, and extracting HOG features. The "get_data" function was devised to create training and testing datasets, achieved by iterating over image folders, extracting features, and assigning labels. Feature scaling was applied to normalize the data. Once the data was obtained, LRC, SVM, RFC, and GBC models were developed. These models were trained using the provided training data.

In the CNN the Keras package was employed for binary image categorization. The network architecture comprised convolutional layers designed to capture hierarchical features within an image. The initial convolution layer utilized 16 filters with a size of (3,3) and a step of 1. A Rectified Linear Unit (ReLU) activation function was applied after each convolutional layer. Maximum pooling was used to reduce spatial dimensions while retaining essential information. This process continued with subsequent convolution layers, using 32 and 16 filters, respectively. A flattening layer converted the 3D output into a 1D vector, facilitating the transition to a fully connected layer.

The first dense layer featured 256 neurons with a ReLU activation function, capturing complex high-level characteristics from the hierarchical representation in the previous convolutional layer. The final dense layer comprised a single neuron with a sigmoid activation function, suitable for binary classification. Sigmoid activation produced an output between 0 and 1, representing the probability of the input image belonging to a positive class. The model architecture is illustrated in Fig. 3.

For model training and optimization, the code utilized the Adam optimizer, an adaptive learning rate method. The binary cross-entropy loss function was chosen to quantify the difference between the expected probability and the actual binary label. Model performance was evaluated using accuracy as the metric. This comprehensive architecture was designed to autonomously learn and extract information from input images, progressing from low-level features in the initial layer to high-level features in the fully connected layer.

All methods underwent final evaluation, assessing model performance on a test set using measures such as precision, recall, and F1 score, with confusion matrices providing insight into the results. Classification tasks also included observing the elapsed time for training the model.

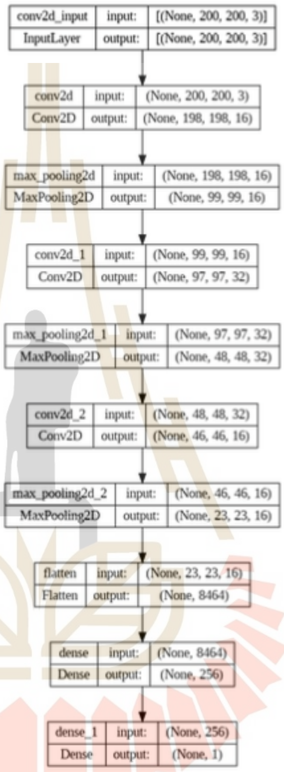


Fig. 3 Image-Based CNN Model Architecture

4.3. Landmark-Based Method

In the landmark-based method experiment, the image data needed transformation into coordinates representing points on the face. The MediaPipe library was employed for this purpose, extracting 468 facial coordinates in both the x and y axes. This transformation was applied to all images in both the training and test sets. To leverage the acquired information effectively, specified algorithms, including CNN, LRC, SVM, RFC, and GBC, were employed.

While LRC, SVM, RFC, and GBC could be directly implemented using the sklearn library to create models, constructing a CNN for binary image classification required the use of TensorFlow and Keras. Initial data handling involved loading training and testing datasets into Pandas DataFrames. The features and corresponding labels were separated, and the features were standardized using the StandardScaler from scikit-learn. For numerical processing, label encoding converted target classes into numeric values. The feature data was then reshaped into a 3D format suitable for the subsequent CNN.

The CNN model was defined using the Keras Sequential API and included a 1D convolutional layer with 32 filters and a kernel size of 3, followed by max-pooling and flattening operations. Two fully connected layers followed, incorporating ReLU activation in the first layer and a sigmoid activation in the second, suitable for binary classification. The model was compiled using the Adam optimizer with binary crossentropy loss. The architecture of the model is depicted in Fig. 4. Model training occurred within a loop, continuously doubling the number of epochs until a predefined accuracy threshold was reached. The training and validation sets were employed during this process.

Finally, all methods underwent evaluation. Model performance was assessed on a test set using measures such as precision, recall, and F1 score, with confusion matrices providing insight into the results. The elapsed time for training the model was also observed in the classification process.

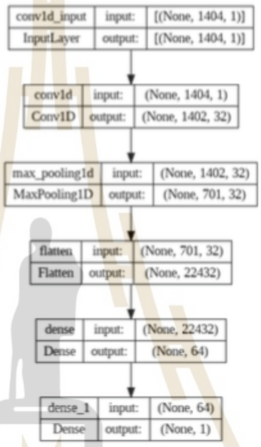


Fig. 4 Landmark-Based CNN Model Architecture

5. Experimental results

In this study, various machine learning models were explored using different techniques and feature sets. The results, presented in Table 1, indicate that the SVM exhibited excellent performance with the landmark technique, achieving an accuracy of 0.8550, precision of 0.8876, recall of 0.8550, and an F1 score of 0.8519. However, the SVM model with the image-HOG-scaling strategy showed slightly lower metrics, with an accuracy of 0.8200, precision of 0.8363, recall of 0.8200, and an F1 score of 0.8178.

GBC demonstrated robust results, particularly with the landmark technique, where it achieved an accuracy of 0.8850, precision of 0.8965, recall of 0.8850, and an F1 score of 0.8842. The GBC model with image-HOG-scaling also performed well, achieving an accuracy of 0.8700, precision of 0.8724, recall of 0.8700, and an F1 score of 0.8698.

RFC displayed good accuracy and precision, particularly with the landmark approach, reaching 0.8950 and 0.9018, respectively. However, the image-HOG-scaling technique significantly reduced performance to an accuracy of 0.8600 and precision of 0.8601. The RFC model consistently maintained recall and F1 scores above 0.8600.

LRC demonstrated efficiency in terms of training time, with the landmark technique requiring only 0.106s. The model achieved an accuracy of 0.8850, precision of 0.8939, recall of 0.8850, and an F1 score of 0.8843. However, the image-HOG-scaling technique resulted in a drop in performance, with an accuracy of 0.7900, precision of 0.8047, recall of 0.7900, and an F1 score of 0.7874.

The CNN outperformed other models, especially with the landmark method, achieving an accuracy of 0.9500, precision of 0.9541, recall of 0.9541, and an F1 score of 0.9541. The image approach similarly yielded strong results, with an accuracy of 0.8700, precision of 0.8558, recall of 0.8900, and an F1 score of 0.8725, albeit with a slightly longer training time of 88.89s. These findings offer a comprehensive overview of the models' performance across different methods and feature sets, facilitating the selection of the most suitable methodology for the specific goal.

Table 1. Machine Learning Model Performance

Model	Techniques	Training Time	Accuracy	Precision	Recall	F1 score
SVM	landmark	0.24s	0.8550	0.8876	0.8550	0.8519
SVM	image-hog-scaling	0.19s	0.8200	0.8363	0.8200	0.8178
GBC	landmark	31.72s	0.8850	0.8965	0.8850	0.8842
GBC	image-hog-scaling	20.93s	0.8700	0.8724	0.8700	0.8698
RFC	landmark	1.34s	0.8950	0.9018	0.8950	0.8946
RFC	image-hog-scaling	1.36s	0.8600	0.8601	0.8600	0.8600
LRC	landmark	0.11s	0.8850	0.8939	0.8850	0.8843
LRC	image-hog-scaling	0.09s	0.7900	0.8047	0.7900	0.7874
CNN	landmark	2.01s	0.9500	0.9541	0.9541	0.9541
CNN	image-hog-scaling	88.89s	0.8700	0.8558	0.8900	0.8725

6. Discussion

The study underscored the critical importance of thoughtful machine learning model selection and feature extraction methods to achieve optimal classification results. Notably, the SVM exhibited sensitivity to the applied techniques, with the landmark approach outperforming the image-HOG-scaling method. This suggests that landmark features significantly contribute to the SVM model's ability to discern patterns within the dataset.

GBC showcased robust performance across both landmark and image-HOG-scaling techniques, indicating its adaptability to different feature sets. While the model's longer training period, especially with the landmark technique, suggests a more intricate learning process, the superior accuracy, precision, recall, and F1 scores achieved justify this investment in time.

The RFC technique consistently delivered strong results across both landmark and image-HOG-scaling feature sets. Maintaining high precision, recall, and F1 scores while training quickly highlights its effectiveness and dependability with diverse data.

LRC demonstrated efficiency in terms of training time, particularly with landmarks. However, the more noticeable decrease in performance with image-HOG-scaling implies that LRC may be more impacted by feature selection, especially with intricate datasets.

The CNN emerged as the top-performing experimental model, excelling in accuracy, precision, recall, and F1 scores, especially with the landmark technique. The longer training time associated with the image technique emphasizes the computational demands of CNNs, but the substantial improvement in performance justifies the investment. These findings emphasize the nuanced considerations necessary in choosing appropriate models and feature extraction methods to achieve optimal results in machine learning classification tasks.

7. Conclusions and Future Work

This study aims to assess the effectiveness of image-based and landmark-based learning approaches in face emotion recognition. The results indicate that the CNN outperformed other models, particularly when employing the landmark technique, achieving an accuracy of 0.9500, precision of 0.9541, recall of 0.9541, and an F1 score of 0.9541. The image-based strategy also yielded favorable results, with an accuracy of 0.8700, precision of 0.8558, recall of 0.8900, and an F1 score of 0.8725, albeit with a somewhat longer training time of 88.89s.

For future studies, the insights gained from this work will be applied in a program aimed at assisting with the physical rehabilitation of Frozen Shoulder. This application is designed to enhance safety and identify injuries during physical therapy.

References

[1] I.M. Revina and W.R.S. Emmanuel, "A survey on human face expression recognition techniques," Journal of King Saud University - Computer and Information Sciences, vol. 33, no. 6, pp. 619-628, 2021.

[2] D. Canedo and A. J. R. Neves, "Facial Expression Recognition Using Computer Vision: A Systematic Review," Applied Sciences, vol. 9, no. 21, Nov 2019, pp. 4678.

[3] S. Kanjanawattana, P. Kittichaiwatthana, K. Srivisut, and P. Praneetpholkrang, "Deep Learning-Based Emotion Recognition through Facial Expressions," Journal of Image and Graphics, vol. 11, no. 2, Jun 2023, pp. 140–145.

[4] S. Schindler, C. Tirloni, M. Bruchmann, and T. Straube, "Face and emotional expression processing under continuous perceptual load tasks: An ERP study," Biological Psychology, vol. 161, Apr 2021, pp. 108056.

[5] R. W. Buck, V. J. Savin, R. E. Miller, and W. F. Caul, "Communication of affect through facial expressions in humans," Journal of Personality and Social Psychology, vol. 23, no. 3, 1972, pp. 362–371.

[6] C. L. Lisetti and D. J. Schiano, "Automatic facial expression interpretation: Where human-computer interaction, artificial intelligence and cognitive science intersect," Pragmatics & Cognition, vol. 8, no. 1, 2000, pp. 185–235.

[7] C. Affonso, A. L. D. Rossi, F. H. A. Vieira, and A. C. P. de L. Ferreira de Carvalho, "Deep learning for biological image classification," Expert Systems with Applications, vol. 85, 2017, pp. 114–122.

[8] P. Mende-Siedlecki, J. Qu-Lee, J. Lin, A. Drain, and A. Goharzad, "The Delaware Pain Database: a set of painful expressions and corresponding norming data," Pain reports, vol. 5, no. 6, 2020.

[9] U. Sharma, K. N. Faisal, R. R. Sharma, and KV Arya, "Facial Landmark-Based Human Emotion Recognition Technique for Oriented Viewpoints in the Presence of Facial Attributes," SN Computer Science, vol. 4, no. 3, 2023, pp. 273.

[10] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," International journal of Remote sensing, vol. 28, no. 5, 2007, pp. 823-870.

[11] M. Bodini, "A review of facial landmark extraction in 2D images and videos using deep learning," Big Data and Cognitive Computing, vol. 3, no. 1, 2019, pp. 14.

[12] R. Chauhan, K. K. Ghanshala, and RC Joshi, "Convolutional neural network (CNN) for image detection and recognition," in 2018 first international conference on secure cyber computing and communication (ICSCCC), 2018, pp. 278-282.

[13] J. M. Hilbe, Logistic regression models. CRC press, 2009.

[14] M. A. Chandra and SS Bedi, "Survey on SVM and their application in image classification," International Journal of Information Technology, vol. 13, 2021, pp. 1-11.

[15] N. M. Abdulkareem and A. M. Abdulazeez, "Science and Business," International Journal, vol. 5, no. 2, 2021, pp. 128-142.

[16] Z. He, D. Lin, T. Lau, and M. Wu, "Gradient boosting machine: a survey," arXiv preprint arXiv:1908.06951, 2019.

[17] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," IEEE transactions on pattern analysis and machine intelligence, vol. 23, no. 10, 2001, pp. 1175-1191.

[18] M. S. Bartlett, G. Littlewort, I. Fasel, and JR Movellan, "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction.," in 2003 Conference on computer vision and pattern recognition workshop, vol. 5, 2003, pp. 53-53.

[19] E. Bekele, D. Bian, J. S. Peterman, S. Park, and N. Sarkar, "Design of a virtual reality system for affect analysis in facial expressions (VR-SAAFE); application to schizophrenia," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 6, 2016, pp. 739-749.

[20] Y. S. Lee and W. H. Park, "Diagnosis of depressive disorder model on facial expression based on fast R-CNN," Diagnostics, vol. 12, no. 2, 2022, pp. 317.

[21] F. Di Luzio, A. Rosato, and M. Panella, "A randomized deep neural network for emotion recognition with landmarks detection," Biomedical Signal Processing and Control, vol. 81, 2023, pp. 104418.

[22] S. Hangaragi, T. Singh, and N. Neelima, "Face detection and Recognition using Face Mesh and deep neural network," Procedia Computer Science, vol. 218, 2023, pp. 741-749.

[23] D. Hamester, P. Barros, and S. Wermter, "Face expression recognition with a 2-channel convolutional neural network," in 2015 international joint conference on neural networks (IJCNN), 2015, pp. 1-8.

[24] M. A. Assari and M. Rahmati, "Driver drowsiness detection using face expression recognition," in 2011 IEEE international conference on signal and image processing applications (ICSIPA), 2011, pp. 337-341.

[25] MINP Munasinghe, "Facial expression recognition using facial landmarks and random forest classifier," in 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), 2018, pp. 423-427.

[26] Ö. F. Söylemez, B. Ergen, and N. H. Söylemez, "A 3D facial expression recognition system based on SVM classifier using distance based features," in 2017 25th Signal Processing and Communications Applications Conference (SIU), 2017, pp. 1-3.

[27] A. Alsubari, DN Satange, and RJ Ramteke, "Facial expression recognition using wavelet transform and local binary pattern," in 2017 2nd International Conference for Convergence in Technology (I2CT), 2017, pp. 338-342.

[28] V. Vapnik, The nature of statistical learning theory. Springer science & business media, 1999.

A Comparison of the Instructor-Trainee Dance Dataset Using Cosine similarity, Euclidean distance, and Angular difference

Thanawat Srikaewsiew
Department of Computer Engineering
Institute of Engineering
Suranaree University of Technology
Nakhon Ratchasima, Thailand
M6500986@g.sut.ac.th

Khatadet Khianchainat
Department of Computer Engineering
Institute of Engineering
Suranaree University of Technology
Nakhon Ratchasima, Thailand
M6500962@g.sut.ac.th

Atima Tharatipyakul
Data Science and Analytics Research Group
National Electronics and Computer Technology Center (NECTEC)
Pathumthani, Thailand
atima.tharatipyakul@gmail.com

Suporn Pongnumkul
Data Science and Analytics Research Group
National Electronics and Computer Technology Center (NECTEC)
Pathumthani, Thailand
suporn.pongnumkul@nectec.or.th

Sarunya Kanjanawattana
Department of Computer Engineering
Institute of Engineering
Suranaree University of Technology
Nakhon Ratchasima, Thailand
sarunya.k@sut.ac.th

Abstract*—The COVID-19 outbreak has restricted most outdoor activities, leads to increasing interest in exercise at home with online trainers. One issue of online exercise technology is the safety since improper motion might result in injury. As a basis to prevent improper motion, methods for evaluating the motion similarity between an instructor and a trainee are essential. Cosine similarity, Angular difference, and Euclidean distance are three general ways for the motion evaluation. This study aimed to determine the most effective way for analyzing the similarity of human motion on the dataset of instructor-led dances. We first experimented with the data to find the appropriate cut-off value for classifying posture into two classes based on the similarity score. Confusion matrix, precision, recall, F1-score, accuracy of the results were then used to compare the efficiency. We discovered that Cosine similarity had the highest accuracy, 82.77 percent at cut-off 93.

***Keywords*— Human Pose Estimation; Cosine similarity; Angular difference; Euclidean distance; Human Motion Similarity**

I. Introduction

Dancing is a set of motions that correspond to music. People use dancing not only for entertainment but also for exercise. For instance, primary school students practice dancing in their physical education class. During COVID-19, human behaviors have shifted to a new norm. All outdoor or group activities, such as dance and aerobics, have been restricted. As a result, there is increasing interest in a method of exercise using online technology, such as exercising at home with an online instructor [1] or online video [2]. Similar to traditional exercise, improper motion may result in trainee injury. To prevent damage in the context of online exercise technology, it is essential to recognize the appropriate evaluation techniques for measuring and comparing the degree of difference, displacement, and velocity between instructors and trainees [3].

Data comparison is a typical technique for distinguishing between similar and distinct data. It is utilized in a variety of fields, including business [4], marketing [5], education [6] and research [7]. Comparisons are required to examine data at all levels efficiently. Careful selection of data or techniques can lead to productive outcomes [8], [9]. There are several techniques for comparing data, including Euclidean distance, Cosine similarity, and Angular difference, among others.

In previous studies, Cosine similarity was used to evaluate similarities between face verification differences [10], word differences in natural language processing [11], and non-speaker information reduction [12]. Various applications of Euclidean distance include obtaining skeletons [13], estimating the distance between syntactically related phrases [14], and mapping the pixel distance between two images [15], and etc. Angular difference is utilized in image registration [16], feature process extraction [17], and out-of-step generator identification [18], among other applications. The evaluation results depend on how exactly the data can be compared. Human motion is challenging to be compared because there are numerous measurements for this kind of data [19].

In this study, we examined evaluation techniques for measuring the similarity of data on human motion through the similarity of posture in each frame. The simple approaches of Euclidean distance, Cosine similarity, and Angular difference were employed to evaluate the similarity of human motion in current scientific research. This work's datasets included dance video from trainees and footage from an exercise instructor obtained from YouTube.

II. Related Works

A. Pose estimation

When comparing human motions, the first and most crucial step is to record the motion characteristics. Today's technology has changed to make it easier to detect human motion through pose estimation, which is a computer vision technique to estimate the spatial locations of key body joints of a person from an image or a video. One of the pose estimation strategies employed by the researcher is to bring MediaPipe, in which Pauzi [20] discussed. Mediapipe Blazepose [21] is a form of algorithm that can be used to detect human motion accurately in real-time.

In this research, the MediaPipe was utilized for extracting skeletal and joint data from each frame.

B. Motion similarity

Motion is the continual change of postures. Valcik et al [22] addressed this concept by discovering that human motion has several components. From the numerous components, a variety of approaches are utilized to retrieve the data for comparison. One approach is in the category of Position Features [23] [22], which is based on the premise of extracting distinct postures in each frame. This approach will not be affected by the speed of the motion or the surrounding frame. The techniques in this area include, for example, Joint Angle Rotations, Distance-Based Pose Features, and Relational Features.

In this research, we used Joint Angle Rotations. This strategy relies on retrieving and comparing the differences in the landmark joints of human. The comparison does not have to be performed on every frame since the difference between each consecutive frames could be minor. This notion is found in Choi et al. [24], where only distinct moves were indexed to save time in operation, and in Ferrari et al. [25], where only one representative frame were selected.

C. Integrating human and machine in evaluation

Motion evaluation is challenging. Evaluation by human takes time and can lead to bias, whereas automated evaluation by machine may be inaccurate, i.e., contradict with human evaluation, due to the method used or model's integrity. One way to address this problem is to integrating human and machine. For example, Wright et al. [26] integrated the labels from human annotators with the results from an image classifier, resulting in increasing in accuracy.

Based on this concept, we designated the similarity of dancing between instructors and trainees as the target for accurate model prediction during model developing and testing phases. Note that the developing phase is a process to determine the cut-off values and parameters from the prepared dataset (developing data). The testing phase is a process to use the cut-off value and apply with unseen dataset (test dataset). A confusion matrix was utilized to evaluate and compare the accuracy of each technique.

D. Efficient pose estimation for limited camera angles

R. S. Hiremath et al. [27] assessed the effectiveness of learners by comparing their dances to those of experts. There were effective methods for gathering data of human motion by detecting the upper and lower portions of the human body independently. The Ferrari research [25] focuses on detecting motion specifically in the upper body. The result is an excellent performance.

In this research, we collected data by assumed that users may use a web camera attached to a laptop in their home, resulting in limited camera angles. Thus, we applied the similar approach of focusing on scoring the motion of the upper part of the body.

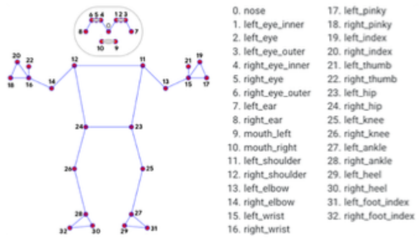


Fig. 1. Pose landmarks. Source: Adapted from [28]

Fig. 2. Screenshot of our application, displaying a tutorial video with 100% opacity and a trainee video feed from a web camera with 80% opacity to facilitate motion learning.

III. Data Source

We collected data using our web application [2], which could display an online tutorial video while showing and recording a user motion from web camera beside the tutorial video to facilitate motion learning, as shown in Fig. 2. Two authors individually followed a video taken from YouTube (https://youtu.be/b0aX6b5lc3M), which instructs an easy dance motion, and recorded the motion. Then, human pose estimation was performed on the recordings using MediaPipe, a machine learning framework for media, with BlazePose GHUM Heavy model. The result contained the x and y coordinates of the trainee based on the pose landmark (Fig. 1) and the timestamp of each frame. Each video recording was considered as one dataset, which was use in different phases as described in the next section.

IV. Experiment

The experiment included three steps: (A) calculating posture similarity score (using Cosine similarity, Euclidean distance, and Angular difference); (B) developing phase (determining the cut-off value of each technique to classify the posture into two classes based on the score); and (C) testing phase (testing and comparing the accuracy of each technique). We explain each step in the following subsections.

A. Calculating posture similarity score

The experiment investigated the similarity in motion between an instructor and a trainee using their body joints in each frame. As the motion involved mostly the upper body and we assumed limited camera angle, we used only points located in the upper portion of the body, including 11 (left shoulder), 12 (right shoulder), 13 (left elbow), 14 (right elbow), 15 (left wrist), and 16 (right wrist). We mapped two points that joined joints and represented their significance as the names of body parts (Table I).

TABLE I. POINT MAPPING AND MEANING

First Point	Second Point	Meaning
11 (left shoulder)	12 (right shoulder)	shoulder
11 (left shoulder)	13 (left elbow)	left upper arms
12 (right shoulder)	14 (right elbow)	right upper arms
13 (left elbow)	15 (left wrist)	left lower arms
14 (right elbow)	16 (right wrist)	right lower arms

For instance, if the first point was 11 and the second point was 12, these two points could be connected by a line to depict the "shoulder". Each body part was represented as a vector. The similarity between a vector pair of each body part of the instructor ($\vec{u}$) and the trainees ($\vec{v}$) was then determined using Cosine similarity, Euclidean distance, and Angular difference.

For the Cosine similarity technique, we used Equation (1).

$$\text{Cosine similarity}(\vec{u},\vec{v}) = \cos(\theta) = \left(\frac{(\vec{u}\cdot\vec{v})}{(||\vec{u}||\cdot||\vec{v}||)}\right) \quad (1)$$

Then, we normalized the result by transforming the 0 - 1 range to the 0 – 100 range by multiplying the result of (1) by 100.

Euclidean distance was calculated by Equation (2).

$$\text{Euclidean distance}(\vec{u},\vec{v}) = \sqrt{\sum_{i=1}^{k}(\vec{u}_i+\vec{v}_i)^2} \quad (2)$$

For Angular difference, we first calculated the angle of each body part of a human to the positive real axis using function numpy.angle(). Then, Equation (3) was used to calculate the angle difference between the instructor angle and the trainee angle, where θ_1 represents the instructor angle, and θ_2 represents the trainee angle. Finally, Equation (4) was used to transform the numbers from 2π to 0 to the range 0 to 100 to normalize the result obtained from Equation (3), where x is the normalized value.

$$\text{Angular difference }(\theta_1,\theta_2) = \theta_1-\theta_2 \quad (3)$$

$$x = 100-(((\text{Angular difference }(\theta_1,\theta_2))*100)/2\pi) \quad (4)$$

The similarity score of the posture of each frame is the average of the similarity between a vector pair of each body part.

B. Developing phase

The cut-off value is used to classify whether the motion of the trainee and instructor are similar. This step's objective was to determine the cut-off value that delivers the highest level of accuracy on one video (developing data) of each technique.

We manually examined and labeled the similarity of the instructor and trainee's dance of each frame into "similarity" and "not-similarity". Images of trainees and instructors were retrieved every 10 frames, totaling 504 frames.

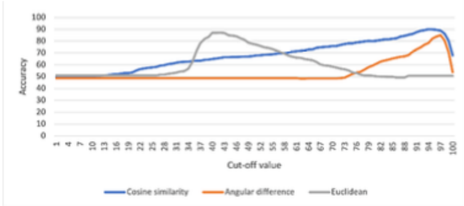


Fig. 3. A comparison of accuracy of cut-off values from 1 to 100.

Note that experiments were conducted so that if Trainee A was the one who followed the instructor, then Trainee B would be the one who determined the similarity of posture, and vice versa.

We created a Python program with the Grid search concept. The posture similarity score of each frame were calculated using three evaluation techniques (Cosine similarity, Euclidean distance, and Angular difference). Then, the frame that received the score below the cut-off value was classified as "not-similarity", otherwise it was classified as "similarity". For the Euclidean distance, the cut-off score was reversed, e.g., the score below the threshold was classified as "similarity". The process was repeated with the cut-off value ranging from 1 to 100, resulting in 100 sets of automated labels of each evaluation technique. We then compared the automated label sets with the manual label to examine their accuracy, as shown in Fig. 3. Finally, we obtained the cut-off value of each evaluation technique by selecting the one that produced the highest accuracy.

C. Testing phase

To evaluate the evaluation techniques, we compared the instructor's and trainees' motions in the test dataset. A confusion matrix is used to evaluate each technique. The confusion matrix is an essential tool for evaluating the predictive results projected by a model or machine learning method, since it determines the proportional relationship between what is expected (what the model predicts) and what really occurs.

We manually labeled the similarity of the instructor and trainee's dance of each frame into "similarity" and "not-similarity" in the same manner as the developing phase. We retrieved one frame per 10 frames, for a total of 470 frames.

We then used the cut-off value of each technique to automatically classify the frame into two classes based on their score. If the Cosine similarity score and the Angular difference are less than the cut-off value, the results are considered as "not-similarity", and vice versa. In contrast, if the Euclidean distance result is greater than the cut-off value, the results are considered as "not-similarity", and vice versa. Finally, we compared the automated label of each technique with the manual label and analyzed the performance of each technique using precision, recall, F1-score, accuracy, and confusion matrix.

TABLE II. THE CUT-OFF VALUES AND ACCURACIES OF EACH TECHNIQUE

Technique	Cut-off Value	Accuracy
Cosine similarity	93	90%
Euclidean distance	41	87%
Angular difference	96	85%

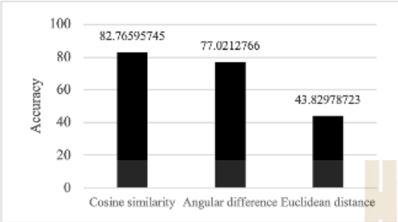


Fig. 4. Comparison of accuracy results in the test dataset.

V. Experimental Results

The experimental outcomes were described as follows:

A. The value that yields the most accurate outcomes for each technique throughout the developing phase.

Table II shows the selected cut-off value of each technique and their accuracy with the Developing data. It is determined that the cut-off value must be set as follows for the best accuracy: Cosine similarity had 90% similarity accuracy with a cutoff value of 93. The Angular difference was 85% accurate at the cutoff value of 96. At a cutoff value of 93, the Euclidean distance had 87% accuracy.

B. Comparison accuracy results during the testing phase

Fig. 4 shows the accuracy of each technique with the test dataset. The accuracy of Cosine similarity, Angular difference, and Euclidean distance were 82.77%, 77%, and 43.83%, respectively.

C. Comparison of accuracy of each technique

Fig. 5 compares the accuracies of each evaluation technique when applying to two datasets. For the developing phase, the accuracies of each technique were high and comparable. For the testing phase, Cosine similarity and Angular difference were not much different from those acquired during the developing phase. However, the accuracy of Euclidean distance was much lower compared to the result in the developing phase.

For this study, the Cosine similarity produced the most accurate results, with 82.77% accuracy.

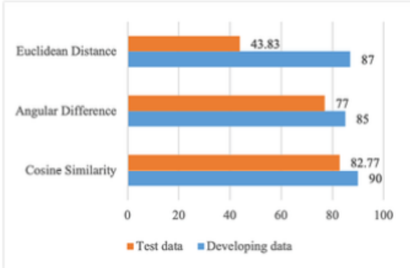


Fig. 5. Comparison of accuracy between the developing phase and the testing phase.

D. Detailed comparison of Cosine similarity, Angular difference and Euclidean distance.

Table III and Fig. 6 show the analysis of Cosine similarity results. Evidentially, the precision of Similarity was 0.917603, whereas the precision of Not-similarity was 0.709360. This indicates that the Cosine similarity can efficiently identify the trainee's motion in close proximity to the instructor's. The recall of the Cosine similarity were not notably different. Both values were relatively high at 0.805921 and 0.867470.

Table IV and Fig. 7 show the analysis of Angular difference. The precision offered by Angular difference for the Not-similarity class was 0.615009, whereas the result for the Similarity class was 0.949541. These outcomes are comparable to those produced using the Cosine similarity technique. The recall values were 0.933735 for Not-similarity and 0.680921 for Similarity.

Table V and Figure 8 show the analysis of the Euclidean distance. The technique revealed the most differences, with a recall result of Not-similarity equal to 0.993976 and Similarity equal to 0.134886.

Fig. 6. Confusion matrix of Cosine similarity.

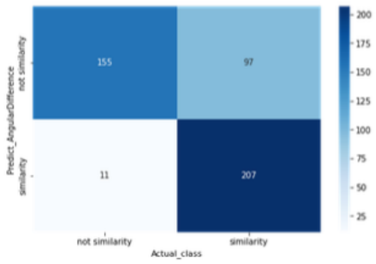


Fig. 7. Confusion matrix of Angular difference.

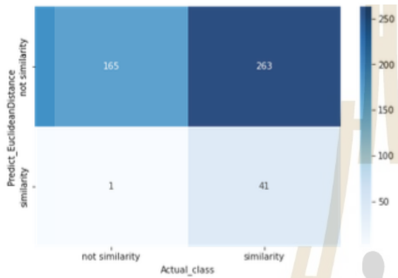


Fig. 8. Confusion matrix of Euclidean distance.

TABLE III. PRECISION , RECALL , F1-SCORE AND ACCURACY OF COSINE SIMILARITY

	Not-similarity	Similarity	Accuracy
Precision	0.709360	0.917603	0.827660
Recall	0.867470	0.805921	0.827660
F1-Score	0.780488	0.858144	0.827660

TABLE IV. PRECISION , RECALL , F1-SCORE AND ACCURACY OF ANGULAR DIFFERENCE

	Not-similarity	Similarity	Accuracy
Precision	0.615079	0.949541	0.770213
Recall	0.933735	0.680921	0.770213
F1-Score	0.741627	0.793103	0.770213

TABLE V. PRECISION , RECALL , F1-SCORE AND ACCURACY OF EUCLIDEAN DISTANCE.

	Not-similarity	Similarity	Accuracy
Precision	0.385514	0.976190	0.438298
Recall	0.993976	0.134868	0.438298
F1-Score	0.555556	0.236994	0.438298

VI. CONCLUSIONS AND DISCUSSION

This study compared Cosine similarity, Angular difference, and Euclidean distance as evaluation techniques for analyzing trainee-instructor dance data. We established the appropriate cut-off value for each technique, which resulted in the highest degree of accuracy during the testing phase, as follows: Cosine similarity with accuracy of 82.77% at a cut-off 93, Angular difference with accuracy of 77% at a cut-off value of 96, Euclidean distance with accuracy 43.83% at a cut-off value of 41. The most suitable evaluation technique for trainee-instructor dance data was the Cosine similarity.

This study has the limitation that the dataset contained only one motion of two people. In the future, we will explore methods to improve the accuracy of the evaluation techniques and methods to compare the evaluation techniques other than the confusion matrix. For instance, the accuracy of Euclidean distance in the testing phase was dramatically decrease. One possible reason is that the technique takes into account the vector length (i.e., the physique of the user). Normalizing the vector length should increase this technique's accuracy. In addition to evaluation techniques, we would like to explore features for facilitating analysis of the choreographies. For example, we may display the angle of the landmark joint in our web application and display suggestions for trainees to improve their performance.

ACKNOWLEDGMENT

This research was funded by National Electronics and Computer Technology Center (NECTEC), National Science and Technology Development Agency (NSTDA), Thailand.

REFERENCES

[1] M. C. Dorneich, B. O' Dwyer, A. R. Dolowitz, J. L. Styron, and J. Grogan, "Application exercise design for team - based learning in online courses," New Directions for Teaching and Learning, vol. 2021, no. 165, pp. 41-52, March 2021.

[2] A. Tharatipyakul, K. T. W. Choo, and S. T. Perrault, "Pose Estimation for Facilitating Movement Learning from Online Videos," Proceedings of the International Conference on Advanced Visual Interfaces, Sep. 2020, Published, doi: 10.1145/3399715.3399835.

[3] J. C. P. Chan, H. Leung, J. K. T. Tang, and T. Komura, "A Virtual Reality Dance Training System Using Motion Capture Technology," IEEE Transactions on Learning Technologies, vol. 4, no. 2, pp. 187–195, April 2011.

[4] M. Van Osnabrugge, "A comparison of business angel and venture capitalist investment procedures: An agency theory-based analysis," Venture Capital, vol. 2, no. 2, pp. 91–109, April 2000.

[5] H. McDonald and S. Adam, "A comparison of online and postal data collection methods in marketing research," Marketing Intelligence & Planning, vol. 21, no. 2, pp. 85–95, April 2003.

[6] M. Jain, A. Bansal, P. Tyagi, A. Jain, U. Tiwari, and N. Nashine, "Comparison and Evaluation of Attitude and Knowledge Towards the Management of Dental Injury in School Teachers Before and After Oral Health Education," International Journal of Clinical Pediatric Dentistry, vol. 11, no. 5, pp. 425–429, 2018.

[7] X. Zheng, M. Wang, and J. Ordieres-Meré, "Comparison of Data Preprocessing Approaches for Applying Deep Learning to Human Activity Recognition in the Context of Industry 4.0," Sensors, vol. 18, no. 7, p. 2146, July 2018.

[8] A. KumarPatidar, J. Agrawal, and N. Mishra, "Analysis of Different Similarity Measure Functions and Their Impacts on Shared Nearest Neighbor Clustering Approach" *International Journal of Computer Applications*, vol. 40, no. 16, pp. 1–5, February 2012.

[9] H. Tong and P. Dabas, "Cluster of time series models: an example" Journal of Applied Statistics, vol. 17, no. 2, pp. 187–198, January 1990.

[10] H. V. Nguyen and L. Bai, "Cosine Similarity Metric Learning for Face Verification," Computer Vision – ACCV 2010, pp. 709–720, 2011, doi: 10.1007/978-3-642-19309-5_55.

[11] B. Li and L. Han, "Distance Weighted Cosine Similarity Measure for Text Classification," Intelligent Data Engineering and Automated Learning – IDEAL 2013, pp. 611–618, 2013, doi: 10.1007/978-3-642-41278-3_74.

[12] H. Zeinali, A. Mirian, H. Sameti, and B. BabaAli, "Non-speaker information reduction from Cosine Similarity Scoring in i-vector based speaker verification," Computers & Electrical Engineering, vol. 48, pp. 226–238, November 2015.

[13] U. Montanari, "A Method for Obtaining Skeletons Using a Quasi-Euclidean Distance," Journal of the ACM, vol. 15, no. 4, pp. 600–624, October 1968.

[14] R. Ferrer i Cancho, "Euclidean distance between syntactically linked words," Physical Review E, vol. 70, no. 5, November 2004.

[15] P.-E. Danielsson, "Euclidean distance mapping," Computer Graphics and Image Processing, vol. 14, no. 3, pp. 227–248, November 1980.

[16] Y. Keller, Y. Shkolnisky, and A. Averbuch, "The angular difference function and its application to image registration," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 6, pp. 969–976, June 2005.

[17] X. Huang, H. Chen, and J. Gong, "Angular difference feature extraction for urban scene classification using ZY-3 multi-angle high-resolution satellite imagery," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 135, pp. 127–141, January 2018.

[18] Y. Cui, R. G. Kavasseri, and S. M. Brahma, "Dynamic State Estimation Assisted Out-of-Step Detection for Generators Using Angular Difference," IEEE Transactions on Power Delivery, vol. 32, no. 3, pp. 1441–1449, June 2017.

[19] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh, "Experimental comparison of representation methods and distance measures for time series data" *Data Mining and Knowledge Discovery*, vol. 26, no. 2, pp. 275–309, February 2012.

[20] A. S. B. Pauzi et al., "Movement Estimation Using Mediapipe BlazePose," Advances in Visual Informatics, pp. 562–571, 2021, doi: 10.1007/978-3-030-90235-3_49.

[21] Z. Min, "Human Body Pose Intelligent Estimation Based on BlazePose," 2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), Feb. 2022, Published, doi: 10.1109/eebda53927.2022.9745022.

[22] J. Valcik, "Similarity Models for Human Motion Data" Ph.D. Thesis, Faculty of Info., Masaryk Univ., Brno, 2016. [Online]. Available: https://is.muni.cz/th/wx926/thesis.pdf

[23] S. Agahian, F. Negin, and C. Köse, "An efficient human action recognition framework with pose-based spatiotemporal features," Engineering Science and Technology, an International Journal, vol. 23, no. 1, pp. 196–203, Feb. 2020, doi: 10.1016/j.jestch.2019.04.014.

[24] W. Choi, T. Ono, and K. Hachimura, "Body Motion Analysis for Similarity Retrieval of Motion Data and Its Evaluation," 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Sep. 2009, Published, doi: 10.1109/iih-msp.2009.174.

[25] V. Ferrari, M. Marin-Jimenez and A. Zisserman, "Pose search: Retrieving people using their pose", Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 1-8, June 2009.

[26] D. E. Wright *et al.*, "A transient search using combined human and machine classifications," *Monthly Notices of the Royal Astronomical Society*, vol. 472, no. 2, pp. 1315–1323, July 2017.

[27] R. S. Hiremath, S. Bhat, and H. R. Srikanth, "An automated evaluator for a classical dance — Bharatanatyam (Nritta)," *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, February 2017.

[28] "Pose," mediapipe. https://google.github.io/mediapipe/solutions/pose.html (accessed Oct. 07, 2022).

BIOGRAPHY

Thanawat Srikaewsiew was born in August 16, 1999 at Nakhon ratchasima, Thailand. He initially completed his primary school education at Phophat Wittaya School, and completed high school at Ratchasima Witthayalai School. He obtained Bachelor's Degree in Computer Engineering at Suranaree University of Technology. He then, in year 2022, continued his Master's degree in Computer Engineering at School of Computer Engineering Institute of Engineering at Suranaree University of Technology, Nakhon Ratchasima, Thailand.