Computer Vision-Based Approach for Breast Cancer Rehabilitation Evaluation: A Survey

Muriati Muda¹, Azim Zaliha Abd Aziz²

¹University College TATI, Teluk Kalong, 24000, Kemaman, Terengganu, Malaysia, ²Universiti Sultan Zainal Abidin, Tembila Campus, 22200, Besut, Terengganu, Malaysia

ARTICLE INFO

Article history:
Received May 28, 2023
Revised June 02, 2023
Accepted June 05, 2023

Keywords:
Computer Vision,
Human Action Recognition,
Motion Evaluation,
Machine Learning,
Home Rehabilitation System,
Breast Cancer

Conflict of Interest:
None

Funding:
None

Corresponding Author: Muriati Muda, University College TATI, Teluk Kalong, 24000, Kemaman, Terengganu, Malaysia. Tel. +60-19-9834367. E-mail: muriati@uctati.edu.my

© Muriati Muda, Azim Zaliha Abd Aziz, This is an open-access article under the CC BY-SA 4.0 international license.

1. Introduction

Breast cancer is still the most prevalent cancer among women worldwide, and a cancer diagnosis can significantly impact a person's perspective on health and life. Patients with breast cancer often struggle with physical lethargy, pain, breast sensitivity, and difficulty concentrating, which can lead to poor quality of life and potentially affect cancer survival rates if not addressed with proper support (Fong & Cheah, 2016). These issues are also prevalent in Malaysia, where breast cancer is the most common cancer among women, with 21,634 cases diagnosed between 2012-2016 (Azizah et al., 2019).

Rehabilitation is one of the main steps towards recovery, especially after surgery. Early rehabilitation treatment focuses on upper limb exercise to improve muscle strength (Klein et al., 2021; Zhou et al., 2021). Home-based therapy is commonly used together with clinic-based rehabilitation to increase the program's flexibility and convenience. A medical professional will create a patient's unique rehabilitation plan, which will include a list of advised physical activities. The patient must carry out the prescribed physical activity and return to the clinic regularly to have their progress evaluated. Physical activity has been suggested to help increase the survival of individuals with cancer. Regularly performing physical exercise will help improve their Quality of Life (QoL).

According to the American Cancer Society (2021), patients who engage in physical activity have a lower risk of cancer, recurrence compared to those who are inactive. In addition, physical activity can improve several...
aspects of physical health, including aerobic fitness, muscular strength, flexibility, body weight status, and composition, bone health, quality of life, vitality, and sleep. It can also reduce symptoms of fatigue, depression, and anxiety (Roberts et al., 2017; Nurnazahiah et al., 2020). Furthermore, engaging in physical activity after a breast cancer diagnosis has been linked to a reduced risk of death from the disease, suggesting that it may play a vital role in survival (Mctiernan et al., 2019; Pekmezi et al., 2012; Barbaric et al., 2010).

However, studies show that up to 70% of breast cancer survivors do not meet the minimum recommended physical activity guidelines for optimal health benefits (Irwin et al., 2004; Lynch et al., 2010). In Malaysia, only 39% of breast cancer survivors engage in a sufficient level of physical activity (Rufa’i et al., 2019). Several barriers have been identified as factors that limit the ability to be physically active, such as fear of exercise due to lack of knowledge and confidence regarding safety, inability to maximize healing benefits due to lack of supervision and evaluation of home exercise, and the need to pay large amounts of money to Physical Therapists (PT) or Occupational Therapists (OCT), which can increase the financial burden on families (Pekmezi et al., 2012; Smith-Turchyn et al., 2016; Sulaiman et al., 2017; Lavallée et al., 2019; Frikkel et al., 2020). Therefore, there is a demand for the development of a system that can monitor patient movement and automatically analyze and evaluate movement performance, which would be crucial in helping patients enrolled in home-based rehabilitation programs.

Reliable motion capture sensors and advancements in machine learning provide an opportunity for the development of such a system. With the emergence of vision-based rehabilitation systems, there is now a more standardized and objective way to evaluate exercise rehabilitation in healthcare. The development of an intelligent motion analysis system using a vision camera is a challenging task that has garnered increasing attention in recent years. The idea is to obtain vital information about the human body through visual means, eliminating the need for physical contact sensors. Currently, more advanced methods have been suggested to analyse motion using a skeleton-based method.

2. Computer Vision-based Approach for Rehabilitation Evaluation

A study conducted by Debnath (Debnath et al., 2022) revealed that a rehabilitation system typically consists of an activity program that provides feedback. The research in the field of computer vision-based rehabilitation systems can be broadly divided into three major parts that cover all important characteristics: (i) primary data collection, (ii) feature extraction and representation, and (iii) feature comparison (Debnath et al., 2022).

2.1 Primary Data Collection

Primary data collection involves self-collected motion data captured by motion device sensors. In traditional computer motion analysis, body-worn sensors, also known as market-based approaches, were used to track human movement. The sensors typically include an accelerometer and gyroscope and are placed at strategic points on the body to collect inertial data. However, data synchronization is a common issue with this type of sensor data, as it can vary in start and finish times depending on the activity's speed. The data may also include noise and differences when performed by different people (W. Zhang et al., 2020). While several researchers have found this type of sensor to provide accurate movement data, patients may feel burdened by wearing the device, making the routine use of motion analysis systems impractical.

A practical and effective solution is to use vision sensors or markerless motion analysis systems. By implementing vision-based sensors, patients only need a computing device with one or more cameras attached to perform the exercises, making data collection affordable and straightforward. RGB cameras are commonly used vision devices to capture movement data. An RGB image contains red, green, and blue bands in the visible spectrum, and object detection is performed based on colour. However, this method has limited ability due to the lack of joint position information and limited range to track the whole human body. It is also susceptible to noise and background interference, resulting in insufficient recognition accuracy (Ding et al., 2020; Debnath et al., 2022).

With the release of Microsoft Kinect in 2010, skeleton tracking became a viable option for researchers, as it provided real-time 3D positions of a player's joints. This tool was widely utilized in tracking human movement across various applications due to its adaptability and affordability, making it an attractive alternative to other high-cost motion-tracking systems such as Vicon and OptiTrack (Ryselis et al., 2020). Numerous studies have yielded promising results in terms of accuracy and the system's ability to identify posture and motion (Bonnechère et al., 2018; Diaz-Monterrosas et al., 2018; Tölgysessy et al., 2021).

There are three types of Kinect cameras used for capturing skeleton data, namely Kinect V1, Kinect V2, and Azure Kinect, as reported by Liao, Vakanski, Xian, et al. (2020) and Tölgysessy et al. (2021). By producing 640x480 RGB and 320x240 depth images at a 30 Hz frame rate, the Kinect V1 can create RGB-D data, which is a combination of both data streams. The Kinect V1 can capture 3D coordinates for 15 to 20 body joints. In 2013, Kinect V2 was released, which added 5 joints around the hands and neck, totalling 25 body joints. It
offers an improved version with a resolution of 1920X1080 RGB and 512X424 depth images at a frame rate of 30 Hz, providing greater accuracy with low measurement errors. The latest Kinect sensor, Azure Kinect, was released in 2019, supporting multiple depth-sensing modes and 32 body joints. This technology is intended mainly for those who are interested in artificial intelligence applications, such as businesses and developers. Skeleton data has simple and distinct characteristics and is less likely to be affected by appearance factors. This allows for faster recognition processes and has garnered the attention of many researchers (Hai & An, 2016; Capecci et al., 2018; Yu & Xiong, 2019; Osgouei et al., 2020). In addition to the Kinect camera, skeleton data can also be acquired using RGB cameras and deep learning methods (Shi et al., 2020; Lee & Ahn, 2020). The OpenPose algorithm has been used for real-time detection of skeleton joints using RGB cameras. It can detect the skeleton of multiple persons in real-time by utilizing Part Affinity Fields, providing 18 body joints of the human skeleton.

Collecting datasets for rehabilitation studies can be challenging due to issues such as patient access, safety concerns, and ethical considerations. Consequently, the datasets collected are often small. To address this challenge, some researchers have used alternative strategies such as using healthy individuals to mimic patients, introducing noise to create varied data, and utilizing publicly available datasets (Liao et al., 2020; Zhang et al., 2021). Publicly available datasets typically contain motion data from healthy individuals and are used for benchmarking purposes. Examples of such datasets include the University of Idaho-Physical Rehabilitation Movement Data (UI-PRMD) and IntelliRehabDS (IRDS) (Vakanski et al., 2018; Miron et al., 2021). Using a Vicon optical tracker and a Microsoft Kinect sensor, the UI-PRMD dataset has collected a variety of exercises that are typical in physical rehabilitation programs. Meanwhile, the IRDS dataset consists of multiple repetitions of nine different gestures performed by 29 individuals, including both patients and healthy subjects. The gestures were captured using a Kinect camera.

2.2 Feature Extraction and Representation

The process of feature extraction involves identifying useful information by reducing the dimensions of the data. Since the raw data often comprises time-series sequences that are high-dimensional, including joint positions and orientations, it is not practical to use the data directly for analysis as it can be redundant and correlated (Debnath et al., 2022; Liao, Vakanski, Xian, et al., 2020).

A traditional approach to feature extraction is the handcrafted feature-based method, which has been widely used for a long time and has produced remarkable results. However, it is time-consuming due to the manual intervention required in the feature selection and validation process, which is highly dependent on the researcher's understanding of the key characteristics of a given motion, making domain knowledge essential. Handcrafted feature-based analysis can be divided into two categories: global feature extraction, which considers the image as a single vector and describes the image, and local feature extraction, which focuses on individual parts of the image (Girdhar et al., 2021).

2.3 Feature Comparison

Assessing patients typically depends on their clinical requirements. For some cases, a simple comparison of joint angle trajectories suffices, while more advanced comparison algorithms may be required for others. Traditional machine learning algorithms such as Random Forest (RF), Bayesian networks, Markov Models, and Support Vector Machines (SVM) have shown good performance when appropriate handcrafted features are used. In a study by Liao (2020), three evaluation methods were listed.

The first method is discrete movement where this approach classifies the output with binary class for valid or invalid movement. While this approach is highly accurate in distinguishing valid from invalid movement sequences, the relevance of this aspect is lower when developing a system to access rehabilitation performance. The second method is a rule-based approach that uses a set of rules defined by experts. The rules act as a standard to assess the movement. This approach is valuable for simpler exercises but more difficult for complex rehabilitation exercises. Some studies used fuzzy logic to generate the score to represent the quality of the exercises. In the third evaluation method of the template-based approach, the exercises were evaluated based on the variance between the training motion sequence executed by the patients and the template motion sequences of healthy subjects. This method is classified into Distance Function and Probability Density Function. Dynamic Time Wrapping (DTW) is widely regarded as the most popular distance function method for evaluating rehabilitation exercises. This is because DTW is well-suited for capturing unpredictability and time-shift in movement sequences, both of which are common in physical rehabilitation. However, this method cannot derive a model of rehabilitation data, and the distances are calculated between each pair of corresponding features in two-time series. Meanwhile for the probability density function method, Hidden Markov Model (HMM) is the most used model.

In recent years, deep learning has demonstrated remarkable performance in computer vision. A key advantage represent features using trainable feature extractors and performing deep processing using multiple hidden
layers (Girdhar et al., 2021). Currently, the most used deep learning methods with skeleton-based data are Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Graph Convolutional Networks (GNNs). However, when working with small datasets, deep learning methods may not be sufficient for accurate analysis (Debnath et al., 2022; Gjoreski et al., 2016). To address this issue, a combination of deep learning and handcrafted approaches has been proposed to improve analysis results (Gjoreski et al., 2016).

3. Related Works

Liao et al. (2020) utilized a deep learning-based framework to evaluate the efficacy of physical rehabilitation exercises. The framework comprises performance metrics for measuring movement quality, scoring functions for converting these metrics into numerical scores, and deep neural network models that generate scores for input movements through supervised learning. The researchers tested the framework on a dataset of 10 rehabilitation exercises and found that the quality scores produced by the proposed approach were highly consistent with the actual quality scores of the movements. While the report does not mention any motion devices, the data employed in the study were based on skeleton-based approaches.

A study conducted by Zhang and colleagues (M. Zhang et al., 2021) proposed an action recognition method called DV-MHNet, which is based on a distance vector and a multi-high view adaptive network. The MH view adaptive networks have been developed to identify the most appropriate observation angle for different heights, collect comprehensive information on critical points from the current frame's image, and enhance the model's ability to accurately recognize movements across different elevations. To establish the global potential relationship of each key point, the DV mechanism is used to calculate the relative distance and orientation between different key points in the same frame and the same key points in different frames. By constructing a spatial-temporal graph convolutional network, the model learns the action characteristics. The proposed method was evaluated on two widely used action recognition benchmarks and achieved superior performance.

Ding and colleagues (Ding et al., 2020) have proposed a novel algorithm that utilizes multiple features and rules learning. The algorithm is based on skeleton information extracted from a Kinect sensor and defines angle and distance features based on the local relationship between joints and the global spatial location of joints. To enhance the classification performance, the rule learning method is employed in conjunction with Bagging and random subspace methods, generating diverse samples and features. The experimental results demonstrate the effectiveness of the proposed algorithm in recognizing various human postures. Furthermore, the rule-based learning method yields results with higher interpretability than traditional machine learning methods and CNNs.

To facilitate automated coaching within a virtual game context, Yu, and Xiong (2019) developed a Dynamic Temporal Warping (DTW) that can access physical rehabilitation exercises performed at home using Kinect technology. Using the DTW approach, the algorithm initially calculates the motion similarity between two-time series. Bone vectors and body orientation are applied as input by the algorithm, which also proposes an inventive technique to convert DTW distance into a performance score that is both straightforward and meaningful. The study's outcomes imply that the DTW-based algorithm is a practical means of automatically assessing an individual's performance while engaging in home-based rehabilitation exercises.

Capecci et al. (2018) proposed the use of the Hidden Semi-Markov Model (HSMM) for the evaluation of rehabilitation exercises. This approach extracts motion features related to clinical aspects from the trajectory of skeleton joints captured by an RGB-D camera and provides an output score for the subject. The HSMM method combines elements of rule-based and template-based approaches by using clinical descriptors as features and testing them through an HSMM model trained on an exemplary motion sequence. The effectiveness of the proposed solution was evaluated by analysing its correlation with both clinical evaluation and Dynamic Time Warping (DTW) algorithms. The results showed that the HSMM-based method had a better correlation with the physician's score than DTW.

Instead of utilizing a depth camera, Shi, and colleagues (Shi et al., 2020) have introduced a novel motion evaluation system that employs a light camera to obtain skeleton data for assessing human motion in healthcare rehabilitation. The pose estimation technique is utilized to extract the body skeleton from the image, and the joint trajectory feature is obtained by applying the Fourier transform to normalize the trajectory and extract the Fourier transform coefficients. A regression model is then established to evaluate the quality of motion. The experimental results indicate that the proposed method is effective.

Although markerless motion sensors are currently the most used sensors, marker-based sensors are still in use. Pereira and colleagues (Pereira et al., 2019) conducted research that used two types of wearable sensors to detect exercise and monitor human motion. The first type is an inertial sensor equipped with triaxial accelerometers, gyroscopes, and magnetometers. The second type is a Body Area Network (BAN) equipped with an electromyography sensor and an accelerometer. These devices communicate wirelessly with a
smartphone through Bluetooth. In this study, data were obtained in a laboratory setting from seven healthy subjects with a defined protocol by a physiotherapist. To reduce undesired noise, the data were pre-processed using a low-pass filter. The researchers then used their novel segmentation algorithm, which is based on a syntactic approach and symbolic method of pattern and query search tasks. After segmentation, the main features were extracted for the next task of movement evaluation. The findings suggest that the combination of these two types of wearable sensors can be used to detect exercise repetitions, and the extracted features based on human posture can support the assessment of exercise performance.

Zhang and his team (W. Zhang et al., 2020) also concentrate on sensor-based rehabilitation exercise recognition, specifically on Smart Sensor-based Rehabilitation Exercise Recognition (SSRER). They employed a deep learning framework to evaluate the exercises. To address the issue of data synchronization, they utilized a convolutional neural network on a dynamic platform (D-CNN). They proposed a unique matrix for testing rehabilitation exercises, and the learned classifier was used to determine the best features. The findings demonstrated that SSRER was highly effective in recognizing exercise using sensor data.

Chae et al. (2020) designed a system for home rehabilitation that monitor the type and frequency of exercises performed by the user through a smartwatch and smartphone application. The system is equipped with a machine-learning algorithm. The algorithm was trained with a convolutional neural network to detect home exercises. Results indicated that the home care system, which utilizes a commercial smartwatch and machine learning model, can enhance participation in-home training, and improve functional scores in treating patients with chronic stroke. This approach has the potential to serve as a cost-effective tool for the home care treatment of stroke survivors in the future.

Zhu et al. (2019a) introduced a novel multipath Deep Learning model for classifying therapy exercises using sensor-based data. The proposed model combines Dynamic CNN (D-CNN) with State Transition Probability CNN (S-CNN) to address data alignment issues. The model is also able to discover the exercise's hidden states. To evaluate the exercises, a novel method is introduced by identifying the feature. During classification, the feature is extracted, and the distance of the selected feature is used as the evaluation score. The results show that the proposed model outperforms other models in classification accuracy, as well as the evaluation scores.

Several studies have focused on promoting rehabilitation among breast cancer patients. However, these studies are still in the early stages of development and lack proper design reporting. Most of these studies support personalized coaching and self-management systems through smart devices. Users are guided using illustrated instructions to perform exercises, and triaxial accelerometers and questionnaires are used to evaluate exercise performance (Harder et al., 2017; Monteiro-Guerra et al., 2020). While these solutions can solve several barriers, they lack efficient feedback on exercise accuracy using computer vision-based approaches. Table 1 and Table 2 list the rehabilitation systems and a summary of the related works discussed.

Table 1. Related Works of Rehabilitation System

<table>
<thead>
<tr>
<th>Author</th>
<th>Primary Data/ Sensor</th>
<th>Methods</th>
<th>Finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Liao et al., 2020)</td>
<td>Skeleton/ Kinect</td>
<td>Deep Neural Network</td>
<td>Quality scores produced were highly consistent with the actual quality scores</td>
</tr>
<tr>
<td>(M. Zhang et al., 2021)</td>
<td>Skeleton/ public datasets (PKU-MMD and NTU-RGB+ D)</td>
<td>Distance Vector and Multi-high View Adaptive Network (DV-MHNet)</td>
<td>Achieved better performance</td>
</tr>
<tr>
<td>(Ding et al., 2020)</td>
<td>Skeleton/Kinect</td>
<td>Rule Learning Method Bagging approach</td>
<td>Demonstrate the effectiveness in recognizing various human postures</td>
</tr>
<tr>
<td>(Yu &amp; Xiong, 2019)</td>
<td>Skeleton/Kinect</td>
<td>Handcrafted Dynamic Time Warping (DTW)</td>
<td>Effective for automatically evaluating an individual's performance</td>
</tr>
<tr>
<td>(Capecci et al., 2018)</td>
<td>RGB-D/ skeleton/Kinect</td>
<td>Handcrafted Hidden Semi Markov Model (HSMM) Dynamic Time Warping (DTW)</td>
<td>HSMM-based method had a better correlation score</td>
</tr>
<tr>
<td>(Shi et al., 2020)</td>
<td>Skeleton/light camera</td>
<td>Relative Distance Fourier Transform</td>
<td>Effective to evaluate the quality of motion</td>
</tr>
</tbody>
</table>
Support Vector Regression (SVR) (Pereira et al., 2019)
Handcrafted Syntactic Approach
Supervised Learning Method
The combination of sensors can be used to detect exercise repetitions, and the extracted features based on human posture can support the assessment of exercise performance SSRER was highly effective in recognizing exercise using sensor data

Inertial data/ triaxial accelerometer, gyroscope and magnetometer, electromyography sensor

D-CNN S-CNN

Convolutional Neural Network

Able to enhance participation in home training and improve functional scores in treating patients

Outperforms in classification accuracy, and the evaluation scores are effective for practical applications

Table 2. Summary of Related Works of Rehabilitation System

<table>
<thead>
<tr>
<th>Author</th>
<th>Primary Data</th>
<th>Features Extraction and Representation</th>
<th>Features Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inertial Data</td>
<td>RGB Skeleton Handcrafted Feature-Based Approach Deep Learning-Based Approach Discrete Movement Score Approach Rule-Based Approach Template-Based Approach</td>
<td></td>
</tr>
<tr>
<td>(Liao et al., 2020)</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>(M. Zhang et al., 2021)</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>(Ding et al., 2020)</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>(Yu &amp; Xiong, 2019)</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>(Capecci et al., 2018)</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>(Shi et al., 2020)</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>(Pereira et al., 2019)</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>(W. Zhang et al., 2020)</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>(Chae et al., 2020)</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>(Zhu et al., 2019b)</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

5. Conclusion

Due to barriers such as fear of exercising, lack of supervision and evaluation for home exercises, and the financial burden that breast cancer patients face when trying to follow rehabilitation treatment, the development
of vision-based rehabilitation systems can be an effective solution to help these patients. This review aims to discuss the research area of human motion analysis in rehabilitation. To automate the evaluation of rehabilitation movements, the most critical step involves three phases: primary data collection, feature extraction and representation, and feature comparison. Various approaches have been discussed for data collection, including inertial data, RGB-D, and skeleton data using different methods and devices.

The most recent research in this field has utilized Kinect-based skeleton data, which has several advantages. Kinect provides a variety of data formats, including RGB videos, depth videos, and 3D joint positions of the skeleton, with clear and straightforward features compared to other methods. Additionally, the Kinect system is inexpensive and easy to use. However, there are several limitations to the Kinect system, such as its static position, which limits the view angle, and the need for efficient machine learning methods to extract and produce meaningful skeleton data for accurate assessment. Researchers have introduced various machine learning and deep learning approaches to improve this process, but there is still room for exploration and overcoming challenges.

Furthermore, researchers must address other challenges, such as how to work well with a limited set of data due to restricted access to the target population and how to generate accurate and informative assessments in real time. The implementation of other soft computing techniques, such as evolutionary algorithms and fuzzy logic, may help pave the way for solutions. Additionally, combining deep learning with a handcrafted approach is an interesting solution to be explored.

References


Nurnazahiah, A., Shahril, M. R., Nor Syamimi, Z., Aryati, A., Sulaiman, S., & Lua, P. L. (2020). Relationship of objectively measured physical activity and sedentary behaviour with health-related quality of


