https://doi.org/10.31436/iiumej.v26i1.3455

DESIGN AND IMPLEMENTATION OF A DEEP LEARNING BASED HAND GESTURE RECOGNITION SYSTEM FOR REHABILITATION INTERNET-OF-THINGS (RIOT) ENVIRONMENTS USING MEDIAPIPE

NURUL HANIS MOHD DHUZUKI[1], AHMAD ANWAR ZAINUDDIN[1*], NUR ANIS SOFEA KAMARUL ZAMAN[2], ALIN NUR MAISARAH AHMAD RAZMI[2], WONDERFUL SHAMMAH KAITANE[3], ASMARANI AHMAD PUZI[1], MOHD NAQIUDDIN JOHAR[4], MASLINA YAZID[5], NOR AZLIN MOHD NORDIN[6], SHAHRUL NAIM SIDEK[7], HASAN FIRDAUS MOHD ZAKI[7]

[1]*Department of Computer Science, Kulliyyah of Information & Communication Technology, International Islamic University Malaysia, Gombak, Malaysia*
[2]*Department of Information System, Kulliyyah of Information & Communication Technology, International Islamic University Malaysia, Gombak, Malaysia*
[3]*MILA University, Putra Nilai, Nilai, Malaysia*
[4]*Physiotheraphy Unit, Rehabilitation Department, Hospital Putrajaya, Malaysia*
[5]*Department of Rehabilitation Medicine, Hospital Shah Alam, Malaysia*
[6]*Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Malaysia*
[7]*Department of Mechatronics Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Gombak, Malaysia*

**Corresponding author: anwarzain@iium.edu.my*

(Received: 1 October 2024; Accepted: 25 December 2024; Published online: 10 January 2025)

ABSTRACT: Frequent hospital visits for hand rehabilitation exercises, such as strengthening and opposition exercises, present significant challenges, especially for patients in remote areas. This paper addresses this problem by developing a Rehabilitation Internet-of-Things (RIOT) system that utilizes MediaPipe with its pre-trained Deep Learning (DL) to deliver real-time feedback during hand rehabilitation exercises alongside Web Assembly (WASM) for efficient processing. The system's objective is to provide precise, real-time tracking of hand movements, enabling patients to perform exercises at home by maintaining an optimal distance between the camera and hand placement, ensuring ideal room lighting conditions across IoT devices such as mobile phones' front cameras and webcams, while healthcare professionals remotely monitor their progress. The methodology involves the integration of MediaPipe for detecting hand landmarks and adaptive sensitivity algorithms to ensure reliable recognition across different environments, such as varying lighting and hand positions. Future work could incorporate additional deep-learning models like CNNs and RNNs to enhance gesture classification accuracy. Several limitations, including latency and distance sensitivity, are addressed in this system with edge computing alongside adaptive algorithms. The key contributions of this research are as follows: First, developing a real-time and cost-effective solution for remote stroke rehabilitation. Second, accuracy is improved by integrating MediaPipe with deep learning techniques. Lastly, latency issues and accuracy challenges at extended distances are alleviated by employing innovative calibration methods and adaptive adjustments. Initial trials demonstrate promising results, though further testing is required under real-world conditions to validate the system's effectiveness fully.

ABSTRAK: Perjalanan yang kerap ke hospital untuk latihan pemulihan tangan, seperti latihan rawatan fisioterapi telah memberikan cabaran yang besar bagi pesakit yang tinggal di pedalaman. Sistem Pemulihan Internet Benda (RIOT) menggunakan MediaPipe bersama

https://doi.org/10.31436/iiumej.v26i1.3455

Deep Learning (DL) yang telah dilatih untuk memberikan maklum balas masa nyata semasa latihan pemulihan tangan, serta Web Assembly (WASM) untuk pemprosesan yang cekap, sebagai penyelesaian. Tujuan sistem ini adalah untuk menyediakan penjejakan pergerakan tangan yang tepat dalam masa nyata, yang mampu dijalankan latihan di rumah dengan pemantauan pegawai perubatan untuk meneliti kemajuan mereka dari jarak jauh. Metodologi melibatkan penyatuan MediaPipe untuk mengesan titik penting pada tangan dan algoritma kepekaan suaian untuk memastikan pengiktirafan yang boleh dipercayai dalam pelbagai persekitaran, seperti pencahayaan dan kedudukan tangan. Lonjakan bagi kajian in adalah dapat menggabungkan model DL seperti CNNs dan RNNs untuk meningkatkan ketepatan dan penyusunan isyarat. Sistem ini juga dapat mengurangkan masalah masa pendam dan perubahab jara dengan melaksanakan edge computing dan penyesuaian algoritma. Sumbangan utama kajian ini termasuklah sistem masa nyata yang kos efektif untuk pemulihan strok jarak jauh, peningkatan ketepatan melalui gabungan MediaPipe dan model DL, dan pengurangan masalah masa pendam dan ketepatan jarak yang lebih jauh melalui tentuukur dan suaian algoritma. Percubaan awal telah menunjukkan hasil yang bagus. Walau bagaimanapun, ujian lanjut masih perlu dibuat dalam dunia sebenar untuk menjamin keberkesanan sistem secara keseluruhan.

KEYWORDS: *Rehabilitation Internet-of-Things (RIOT), MediaPipe, Deep Learning (DL), hand gesture recognition, Artificial Intelligence (AI).*

1. INTRODUCTION

Deep Learning (DL), a transformative branch of artificial intelligence (AI), has significantly advanced various domains by enabling machines to learn from vast datasets and make autonomous decisions. From a healthcare perspective, DL can analyze patient data to tailor personalized treatment plans and monitor progress effectively. For example, studies have demonstrated that DL algorithms can enhance the accuracy of gesture recognition systems, which are necessary for rehabilitation exercises, by processing and interpreting visual data in real-time [1]. The ability of DL to handle large datasets makes it particularly suitable for applications in medical imaging and patient monitoring using the Internet of Things (IoT), where it can uncover insights that traditional methods may overlook [2].

For stroke patients who require continuous exercise, IoT and DL combinations such as gesture recognition systems have emerged as tools for providing real-time feedback on patient movements. This is essential for effective therapy, and the accuracy of these systems is further enhanced through a robust calibration process, which adjusts for varying conditions such as lighting, hand positioning, and distance from the camera [3]. This shows that the integration of IoT and DL techniques has significantly improved the accuracy and reliability of gesture recognition systems, allowing for the detection of subtle movements that may indicate progress or areas needing improvement [2].

In addition, MediaPipe is an innovative framework developed by Google that facilitates real-time analysis and feedback and simplifies the implementation of complex machine-learning tasks, enabling developers to create applications that can detect and interpret human gestures with high accuracy and practicality. The combination of MediaPipe and its pre-trained DL models into rehabilitation technologies allows the development of user-friendly interfaces, thereby enhancing engagement and motivation during therapy sessions at home [4]. The system also uses Web Assembly (WASM) to ensure efficient processing, enabling it to run on various devices, including low-power edge devices.

RIOT utilizes IoT and AI to enable interactions between patients and healthcare professionals during remote therapy procedures to showcase advancements in modern technologies. This connectivity enables healthcare professionals to monitor patient progress with prescribed exercises and make data-driven decisions to enhance treatment outcomes [4]. The gathered data can be analyzed to create personalized rehabilitation plans tailored to each patient's specific needs, which can then be verified by healthcare providers based on rehabilitation scoring, ultimately enhancing overall treatment efficacy [5]. Rehabilitation scoring will assist the therapists in tracking the patient's progress records.

However, a study by Kelly et al. [6] the system addresses challenges such as latency and accuracy by incorporating a robust calibration process and proposing solutions like edge computing and adaptive sensitivity algorithms. The current work has focused on calibrating the system based on environmental factors, latency, and accuracy. Through the investigation, it has been shown that accuracy is influenced by latency and other environmental factors related to recognition and tracking. As the field of smart healthcare advances, RIOT is positioned to play a role in bridging the gap between patients and healthcare providers, ensuring that rehabilitation services are accessible and effective. To align with the efficacy and accuracy in RIOT development, the system's decision-making can be elevated by focusing on tuning the Convolutional Neural Networks (CNNs) for enhanced spatial analysis and Recurrent Neural Networks (RNNs) for dynamic gesture sequence recognition, which allows the system to adapt more effectively to the user's progress.

By offering a real-time, cost-effective, and adaptive solution for remote stroke rehabilitation, the RIOT system aims to greatly enhance the accessibility and quality of rehabilitation services for patients in remote areas. Initial trials demonstrate promising results in terms of recognition accuracy and real-time feedback, though further testing is required to validate the system fully under real-world conditions. Initial explorations of this work were already reported in [7], [8].

The design and implementation of the solution were delivered in this paper as arranged as follows: After this introduction, Section 2 will elaborate on the study of related and previous research, while the ML and DL elaboration in Section 3 entails how the system is developed. Next, Section 4 further explains the software used and shows the experimental results after testing the system. Section 5 breaks down the system's plans, as this paper concludes.

2. LITERATURE REVIEW

This section will elaborate on the role of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) in the RIOT system for hand gesture recognition by exploring the integration of those technologies within the system, their significance in improving patient rehabilitation results, and recent advancements in gesture recognition. Also, MediaPipe's role as a pre-trained model uses deep learning while leaving the possibility of future CNN/RNN integrations. Since MediaPipe handles gesture detection, it is important to be clarified, and CNNs/RNNs could be added in future work.

2.1. Introduction of Machine Learning and Deep Learning: An Overview of AI, ML, and DL in RIOT System

Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) are integral components of the RIOT system for hand gesture recognition, automating the detection process, and delivering real-time feedback for stroke rehabilitation. MediaPipe, a framework developed by Google, plays a pivotal role in this system, leveraging a pre-trained deep-learning

https://doi.org/10.31436/iiumej.v26i1.3455

model to detect hand landmarks accurately and efficiently in real time. This pre-trained model enables the RIOT system to conduct gesture recognition without requiring additional ML or DL frameworks during deployment.

While MediaPipe serves as the primary tool for hand gesture recognition, the system's future enhancements may include advanced deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to improve the detection of more complex movements. Integrating these DL techniques would further enhance the system's ability to adapt to patient-specific needs during rehabilitation exercises.

2.1.1. Machine Learning and Deep Learning in Hand Gesture Recognition

Machine Learning (ML) and Deep Learning (DL) play a crucial role in contemporary gesture recognition systems, particularly in the realm of rehabilitation. MediaPipe uses deep neural networks to recognize hand gestures in real-time [9]. This framework ensures high accuracy in detecting hand landmarks and movements, which is essential for rehabilitation exercises.

Firstly, in Machine Learning (ML), according to Figure 1, feature extraction is a process that requires manual intervention from domain experts who identify and design relevant features from structured data. This manual feature engineering is essential because the performance of ML models, like Support Vector Machines (SVM) and Decision Trees, heavily depends on the quality of these features [10]. Research suggests that effective feature engineering can greatly improve model accuracy and efficiency, highlighting its significance in traditional ML workflows [11].

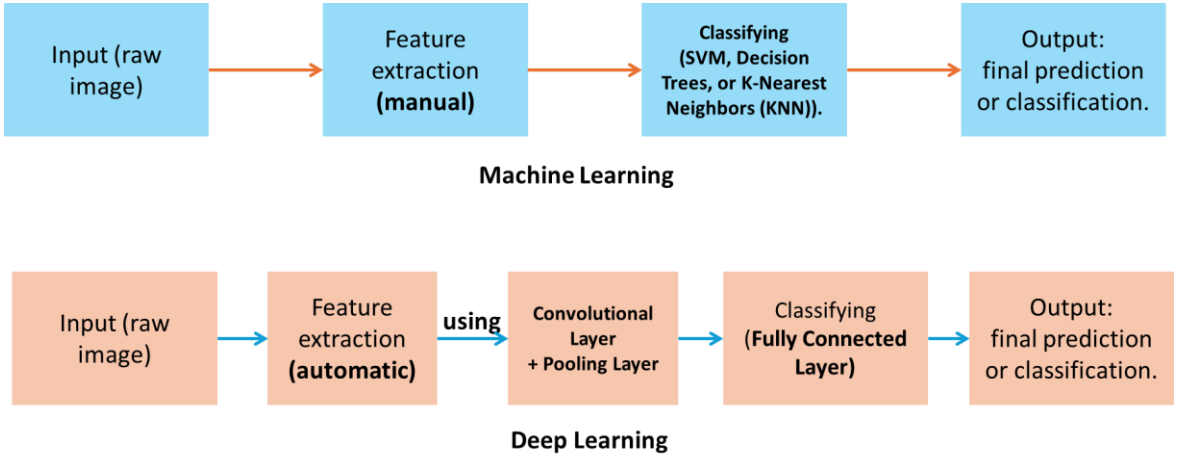

Figure 1. Key differences between ML and DL were presented in their flows and architecture before their final output.

In contrast, Deep Learning (DL) automates the feature extraction process, allowing models to learn directly from raw, unstructured data without the requirement for manual feature design (see Figure 1). This is accomplished through deep architectures that utilize layers such as Convolutional and Pooling layers to identify patterns automatically. The automatic nature of feature learning in DL enables it to excel in complex tasks, such as image and audio processing, where manual feature extraction would be impractical [12]. Thus, the fundamental distinction between ML and DL lies in their approach to feature handling, with ML depending on human expertise and DL leveraging automated learning processes. Figure 2 shows the visualization of machine learning and deep learning model architecture. Each model has its own structure for producing the output. For example, the architecture of CNN and RNN displays more layers

before reaching the output layer than SVM and Decision Tree. Here, it is evident that deep learning models are more complex than machine learning models.

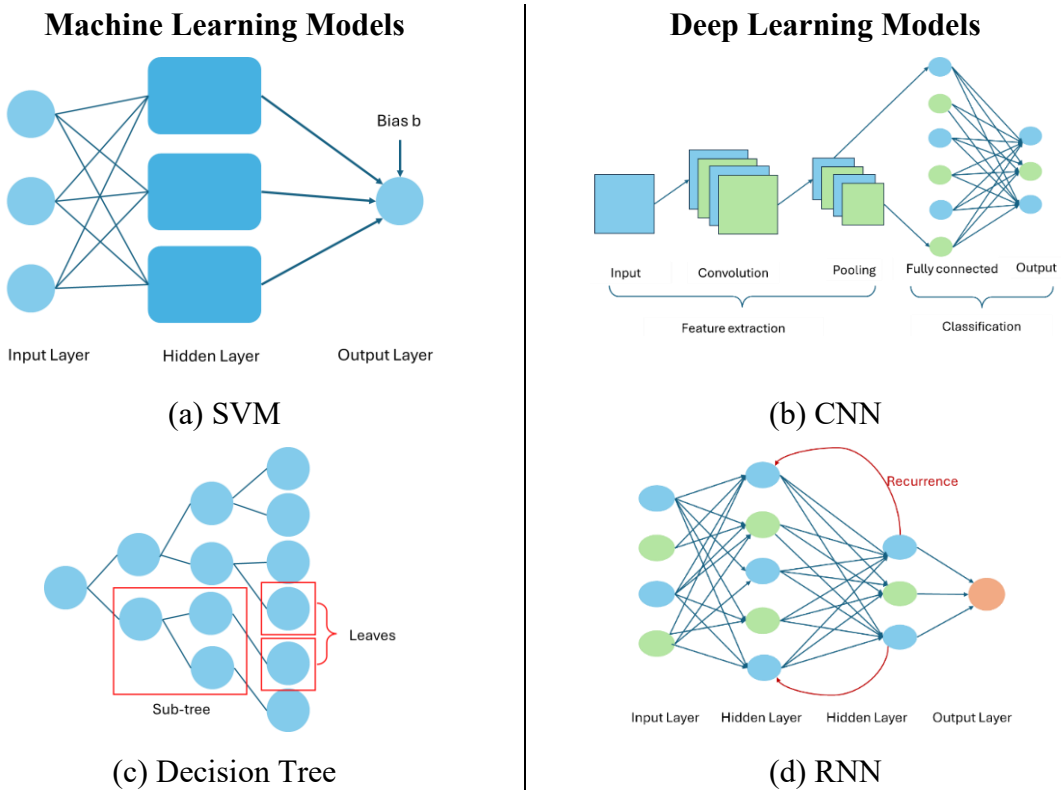

Figure 2 Examples of architecture diagrams of ML and DL models.

In addition, Figure 3 shows a step-by-step process of hand gesture recognition using the MediaPipe framework, which relies on CNN for accurate detection and classification [13]. The process begins with a video stream input, where each frame is analyzed for the presence of a hand. The model accurately locates the hand by extracting features like edges and contours. Subsequently, pooling layers are employed to decrease the dimensionality of the data while retaining information.

As a result, ML models are typically shallow and involve simpler algorithms. In contrast, DL models rely on multiple layers (deep architectures) to process and classify the data, making them better suited for more complex tasks.

2.1.2. Gesture Recognition Algorithm in Existing Products

Accurate and real-time gesture detection is essential for hand rehabilitation systems. MediaPipe's pre-trained deep learning model currently serves as the backbone of the RIOT system, detecting and analyzing hand gestures in real-time. However, AI-based algorithms, such as LDA (Linear Discriminant Analysis), SVM, CNN, and LSTM, are often employed to recognize dynamic hand gestures. Researchers have developed systems that combine these techniques to achieve high accuracy in recognizing hand movements. For example, a researcher at Qingdao University of Science and Technology [15]. The LDA approach was used to create a wrist rehabilitation robot that recognizes five different types of movements with an accuracy rate of over 90% [20]. Other successful researchers from Nanchang University's SVM model were able to identify four distinct types of hand motions with an average success rate of 99.3%.

https://doi.org/10.31436/iiumej.v26i1.3455

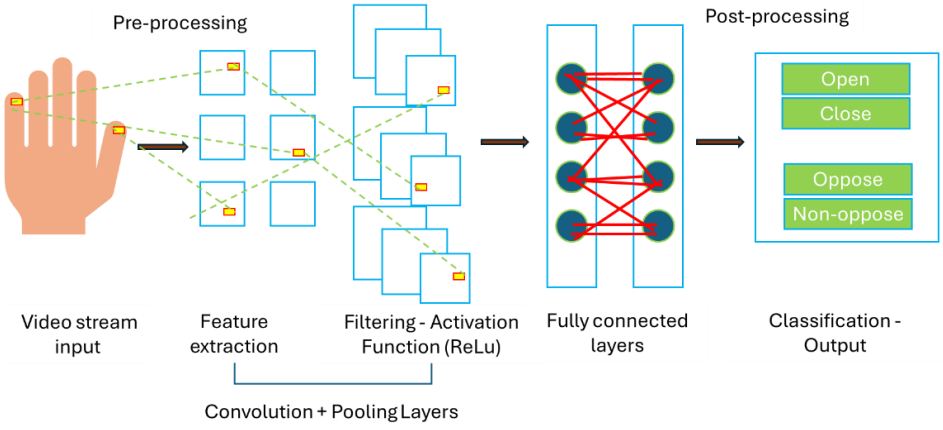

Figure 3. Hand gesture recognition process using MediaPipe framework, which integrates CNN for effective hand detection and classification.

By integrating CNNs and LSTMs as the future version of RIOT, the system could recognize more complex movements and provide real-time feedback to patients during therapy sessions. These studies demonstrate the effectiveness of combining different AI techniques for improved gesture recognition in rehabilitation. Lastly, in a study regarding combining RNN and LSTM algorithms, Zhang Jianxi has created a hand rehabilitation robot that recognizes nine movements with an average accuracy of 91.44% [16].

2.2. Significance and Contribution of Deep Learning in the RIOT System

Deep learning can contribute to real-time feedback in gesture recognition, especially for RIOT environments. Currently, in the environment, MediaPipe processes hand gesture data in real-time, ensuring that patients receive immediate feedback during rehabilitation exercises. The real-time feedback loop improves patient engagement and therapeutic outcomes during remote therapy sessions. It is a continuous adaptive support by DL models to eliminate the delays often associated with traditional rehabilitation methods, thereby creating an interactive environment that is necessary for effective therapy [17]. Moreover, RIOT could implement CNNs to process visual data in the future. At the same time, LTSMs handle sequential data, making these systems instantaneously recognize gestures and adapt to changing patterns in rehabilitation exercises.

Besides the algorithms themselves, it is believed that reducing latency in gesture recognition is key to maintaining a smoother feedback process within the rehabilitation tool and improving the user experience. Delays may cause interruptions and adversely affect the session and patients' progress. By refining advanced algorithms that learn from user interactions, such systems can tailor their responses according to the needs of patients and enable a better rehabilitation process. With the further development and refinement of DL architectures, including attention mechanisms and transformer models, additional gains in the responsiveness of the recognizing system can be expected. These advancements will enable more intricate interpretations of gestures, expanding the range of rehabilitation exercises that can be efficiently monitored and adjusted in real-time [18-19]. In short, DL's contribution to real-time feedback mechanisms is not only significant but also central to the future of rehabilitation technologies.

2.3. Latest Development in Gesture Recognition Technology

Gesture recognition has become immensely popular for rehabilitation based on its improvement in accuracy and adaptability. Traditional models, such as transformer-based

architecture, were replaced by more advanced models aiming to enhance gesture recognition performance. This includes the hybrid of CNN and LSTM for spatial and temporal dimensions in hand gesture study, which are essential for accurately interpreting complex gestures [14,17]. However, MediaPipe remains a highly practical framework for real-time gesture recognition, offering superior performance to older methods like Microsoft Kinect and Leap Motion.

Recent breakthroughs within hardware technologies like edge computing and low-latency networks have further made it possible to deploy these improved models for real-time rehabilitation. For example, technological advances such as these have allowed gesture recognition systems to perform seamlessly with no interruptions in remote environments and provide immediate feedback, thereby supporting users [21]In this feature, AI in gesture recognition not only enhances the accuracy of the systems but also makes them adaptable to the various needs of different patients under rehabilitation. Given AI-driven gesture recognition systems' ability to respond dynamically to patient progress, this feature will become one of the most important ways of creating personalized rehabilitation. As AI-driven rehabilitation tools evolve, MediaPipe's integration with more advanced models could improve system adaptability, providing patients with real-time feedback tailored to their needs.

Furthermore, the research was carried out to develop new AI architectures and learning paradigms that continue pushing gesture recognition's limits beyond what was previously possible. For instance, exploring multi-modal data input, such as the combination of visual and inertial sensor data, provides perspectives for developing more robust recognition systems capable of functioning in diverse environments. This ongoing research and development in AI signals bright prospects for gesture recognition technologies, particularly in the enhancement of rehabilitation practices [22].

2.4. AI-Driven Smart Healthcare System

RIOT system represents a significant advancement in smart healthcare. By imposing AI-driven gesture recognition, healthcare providers can achieve continuous monitoring and develop adaptive rehabilitation tools that are both efficient and accessible [21]. This connectivity allows therapists to track patient progress remotely, monitor therapy exercises in real time, and ensure adherence to therapeutic guidelines, thus enhancing the overall effectiveness of rehabilitation programs [17].

Additionally, the system incorporates edge computing to reduce latency and ensure smooth therapy sessions. This low-latency system allows for seamless interaction between the patient and the rehabilitation tool, eliminating delays that could negatively impact patient progress. Advanced adaptive sensitivity algorithms could further enhance the system's ability to recognize gestures in varying environmental conditions to make it more robust.

Here, the data collected through RIOT systems can be collected to be recorded for healthcare providers' rehabilitation scoring assessment on the patient's progress, which improves patient outcomes but also facilitates a more proactive approach to rehabilitation, where guidance can be made based on provided instructions for an immediate patient response [23].

2.5. Evaluation Metrics and Latency Measurement

Metrics such as accuracy, precision, recall, F1-score, and latency were used to calculate the implementation of the hand gesture recognition system developed in RIOT. During the simulated rehabilitation sessions, the effectiveness of identifying gestures and providing real-time feedback was investigated.

https://doi.org/10.31436/iiumej.v26i1.3455

First, accuracy was measured of how well the system correctly recognizes gestures [24] and was calculated in Eq. (1). Where TP (True Positive) and FP (False Positives) are represented as correctly detected gestures, incorrectly detected gestures, and missed gestures, respectively. Meanwhile, Precision Eq. (2) measures the proportion of correctly identified gestures out of all detected gestures to understand the accuracy of the system in prediction [25]. Moreover, Recall Eq. (3), which is also known as sensitivity, identifies true gestures out of all possible gestures [26]. Next, the balance between precision and recall was defined in Eq. (4) as the harmonic mean of the two metrics that can spot uneven class distribution during evaluation.

Besides the metrics mentioned above, latency is also an essential factor in real-time gesture recognition systems, particularly in rehabilitation contexts where immediate feedback is essential for the user [27]. Latency is measured in milliseconds (ms) and can be defined in two ways depending on the application, and Eq. (5) displays the time taken by the system to provide feedback and gesture inputs. Subsequently, in the experiment, the console window of the RIOT website inspection browser was used to retrieve timestamps and perform latency calculations [28] for every test, as displayed in Figure 4.

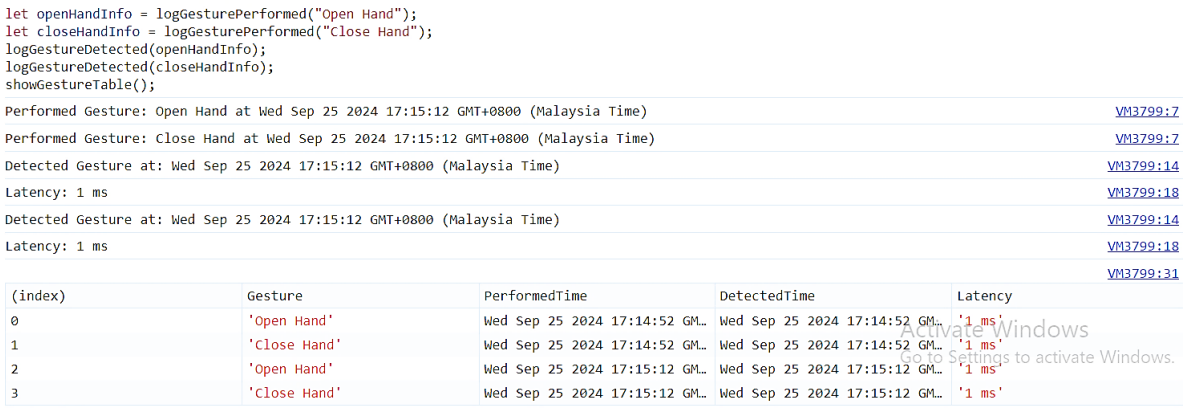

Figure 4. The real-time time taken was displayed in a table from the console window as outputs to calculate the gesture's latency.

In this experiment, latency was examined as one of the factors measuring the delay between executed gestures and gesture detection time. As illustrated below, Eq. (1), Eq. (2), Eq. (3), Eq. (4), and Eq. (5) represent the formulas for the measurements.

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP+FP+FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{2}$$

$$\text{Recall (or Sensitivity)} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

$$\text{F1Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \tag{4}$$

$$\text{Latency (ms)} = \text{Response Time} - \text{Input Time} \tag{5}$$

$$\text{Latency (ms)} = \text{Gesture Detected Time} - \text{Gesture Performed Time} \tag{6}$$

All these equations were implemented to observe the Design-of-Experiment factors' outcomes, which will be discussed in the next section.

https://doi.org/10.31436/iiumej.v26i1.3455

3. METHODOLOGY

The methodology for this investigation focuses on integrating MediaPipe, Web Assembly, and Deep Learning technologies to achieve real-time hand gesture recognition in the RIOT system. This chapter elaborates on the theoretical and practical implementation of these technologies, including the post-calibration process and experimental design factors that affect system accuracy.

3.1. System Design Overview

The RIOT system utilizes MediaPipe, a real-time framework for hand landmark detection and deep learning models to facilitate hand gesture recognition for rehabilitation. The system's design incorporates WASM for web-based operation, enabling lightweight execution of complex models without requiring significant computational resources from the client side.

MediaPipe captures real-time hand gestures through a camera device, such as a webcam, and detects 21 key points. MediaPipe Hands uses pre-trained machine learning models, likely developed using deep learning techniques. However, the MediaPipe framework itself does not require external deep learning frameworks such as TensorFlow or PyTorch during its deployment, as its models are pre-optimized for real-time performance [9].

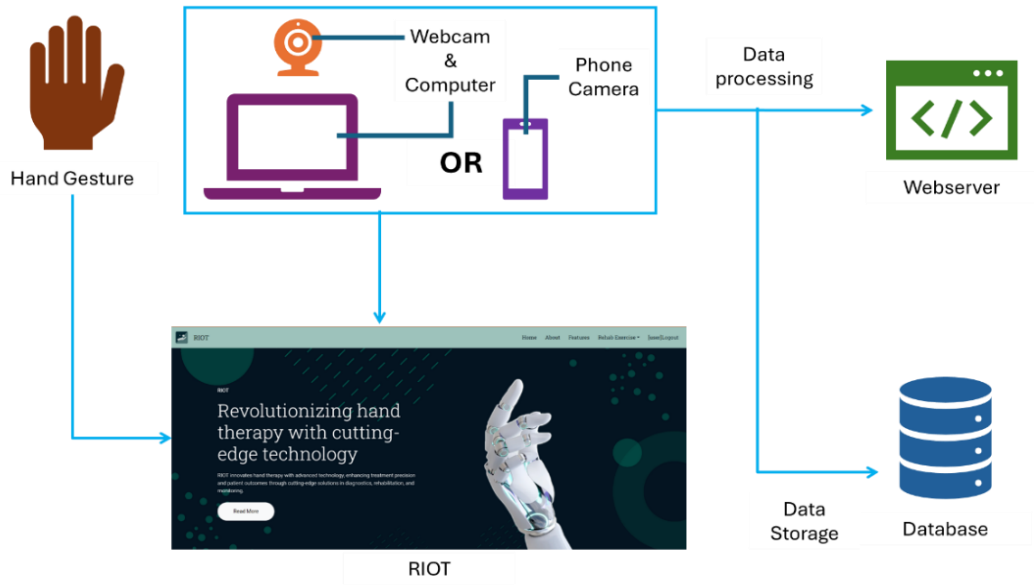

Figure 5. The RIOT system utilizes MediaPipe and WASM with IoT devices to detect and recognize exercise gestures in real time while efficiently managing data.

The website's interactivity was created for two types of users: the patient's guardian or the patient themselves and the administrators. Both users were required to sign up and log in before gaining full access to RIOT features. Since product was successfully built, the next step of the research is optimizing and tuning the system to ensure the optimal conditions for the exercises included in RIOT.

3.2. Calibration of Hand Gesture Recognition

Calibration is crucial to ensure precise hand gesture recognition in the RIOT system. The following section outlines the calibration process and the post-calibration adjustments.

https://doi.org/10.31436/iiumej.v26i1.3455

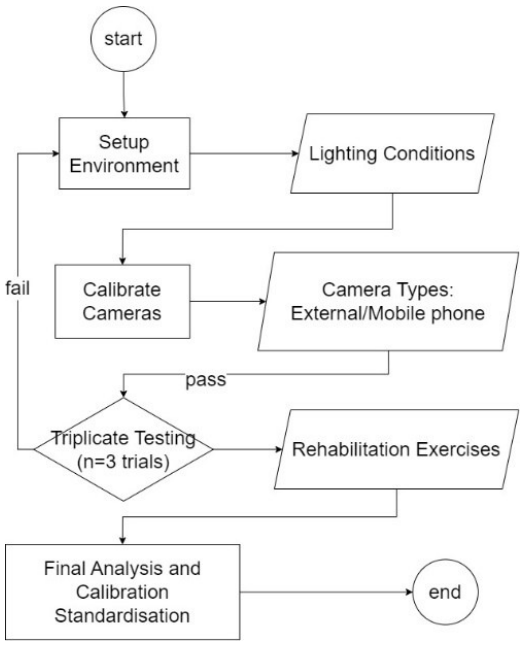

Figure 5. Flowchart of the post-calibration research displaying the process of testing the RIOT website.

Figure 5 reveals the flow of the calibration process, where the user's gestures are captured at various hand positions and distances at the setup, followed by factor adjustments to ensure accurate tracking. The study was repeated three times to ensure thorough analysis. The process began with setting up environmental factors, followed by adjusting camera variables for testing.

3.2.1. Post-Calibration Adjustment

The initial calibration (pre-calibration) phase was focusing on setting up the system to ensure precise hand gesture recognition. This involves utilizing a Design of Experiment (DoE) methodology to test and optimize various factors like camera type, distance, and lighting conditions. Pre-calibration ensures that the gesture recognition system functions reliably, laying the groundwork for more precise rehabilitation assessments in stroke patients, as conducted in prior research [7]. The current focus is on post-calibration adjustments, which ensure that the system maintains high accuracy during real-time rehabilitation sessions. These adjustments involve compensating for varying environmental conditions, such as lighting, hand position, and distance from the camera.

The adaptive sensitivity algorithms implemented in the post-calibration phase dynamically adjust the system's recognition thresholds in response to changes in the user's environment. For example, if the user moves farther from the camera or if the lighting conditions change, the system recalibrates its detection sensitivity to maintain consistent accuracy. This ensures that the hand gestures are recognized effectively throughout the rehabilitation session.

The system setup for the calibration process involves laptop webcams. Meanwhile, post-calibration, an external webcam, and a mobile phone camera were experimented with, and the laptop ran the RIOT software. The specified distance to position the hand is 20.00 cm to 70.00 cm away from the camera because this range ensures optimal focus and clarity of the hand movements, which is accurate for gesture recognition [29]. When movements are precisely tracked in rehabilitation settings, the feedback can significantly enhance patient performance

https://doi.org/10.31436/iiumej.v26i1.3455

and confidence. Also, by emphasizing that real-time feedback during exercises can simulate the presence of a physical therapist, thereby improving adherence and outcomes.

This distance is a primary factor that influences the accuracy of gesture recognition when the camera captures hand movements, and the system processes this data to provide real-time feedback during rehabilitation exercises.

3.3. Design of Experiment (DoE) Factors for Post-Calibration

Following the initial calibration process described in the previous study, this paper focuses on post-calibration adjustments to enhance the hand gesture recognition system across various devices, including external webcam and mobile phone cameras. These adjustments are necessary to maintain accuracy in real-world environments where lighting, camera angles, and distances vary significantly.

To ensure the system's robustness, a Design of Experiments (DoE) methodology was implemented to systematically evaluate and optimize the performance of the gesture recognition system. The following factors were analyzed in Table 1. Data was collected from both external webcams and mobile phone cameras under the mentioned conditions.

Table 1. DoE Factors affecting the accuracy of hand gesture recognition during data collection and analysis.

Factor	Levels/SI Unit	Objectives
Distance	5.0 cm increment from 20.00 cm to 70.00 cm.	To determine the optimal distance at which the camera accurately captures hand gestures. This factor was tested by placing the hand at different distances from the webcam and observing the system's recognition accuracy.
Types of Cameras	• External webcam. • Smartphone front camera.	To compare gesture recognition performance between a mobile phone camera and an external webcam.
Lighting conditions	Indoor under artificial light settings: • Dim. • Bright.	To assess how different ambient lighting conditions impact the accuracy of hand gesture detection.
Latency	Milliseconds (ms).	To measure the time delay (latency) between when a hand gesture is made and when the system recognizes it. This helps determine system responsiveness.

The DoE framework determined the optimal setup for each device type. The results were analyzed to identify the most reliable combinations of camera type, distance, lighting, and hand positioning for both mobile and desktop environments. During this stage, JavaScript was used to conduct part of the investigation, as shown in Figure 7.

https://doi.org/10.31436/iiumej.v26i1.3455

```
// Create an array to store the data for gestures
let gestureData = [];

// Function to log when the gesture is performed
function logGesturePerformed(gestureType) {
    let gesturePerformedTime = new Date();
    console.log("Performed Gesture: " + gestureType + " at " + gesturePerforme
    return {gestureType, gesturePerformedTime};
}
// Function to log when the gesture is detected by the system
function logGestureDetected(gestureInfo) {
    let gestureDetectedTime = new Date();
    console.log("Detected Gesture at: " + gestureDetectedTime);
    // Calculate latency in milliseconds
    let latency = gestureDetectedTime - gestureInfo.gesturePerformedTime;
    console.log("Latency: " + latency + " ms");
    // Add the data to the array
    gestureData.push({
        Gesture: gestureInfo.gestureType,
        PerformedTime: gestureInfo.gesturePerformedTime,
        DetectedTime: gestureDetectedTime,
        Latency: latency + " ms"
    });
}
// Function to display the data in a table format
function showGestureTable() {
    console.table(gestureData);
}
let openHandInfo = logGesturePerformed("Open/Oppose Hand");
let closeHandInfo = logGesturePerformed("Close/Non-Oppose Hand");
logGestureDetected(openHandInfo);
logGestureDetected(closeHandInfo);
showGestureTable();
```

Figure 6. The snippet source code for retrieving the time taken for the gesture actions for latency data from live simulated rehabilitation sessions was written in the console window to perform the calculation.

3.3.1. Potential for Future Deep Learning Integration

Future iterations of the RIOT system will explore the integration of CNNs and RNNs (LSTMs) to improve gesture recognition and enhance the rehabilitation experience. CNNs can provide more refined spatial analysis of hand gestures, while LSTMs can manage dynamic movement sequences, offering users more personalized and accurate feedback.

4. RESULTS AND DISCUSSION

This section evaluates the RIOT system, highlighting its performance across different test environments. This chapter is structured to cover the experimental setup, performance benchmarking results, and comparison with existing systems.

4.1. Experimental Setup

The experimental setup was designed to evaluate the performance of the RIOT system under various environmental conditions and with different camera types, including external webcams and mobile phone cameras. The system was tested with hands positioned between 20.0 cm and 70.0 cm from the camera, as shown in Figs 8(a) and 8(b), to simulate real-world rehabilitation exercises. Key variables such as lighting conditions (dim and bright) and camera type were systematically controlled to ensure a comprehensive performance evaluation.

https://doi.org/10.31436/iiumej.v26i1.3455

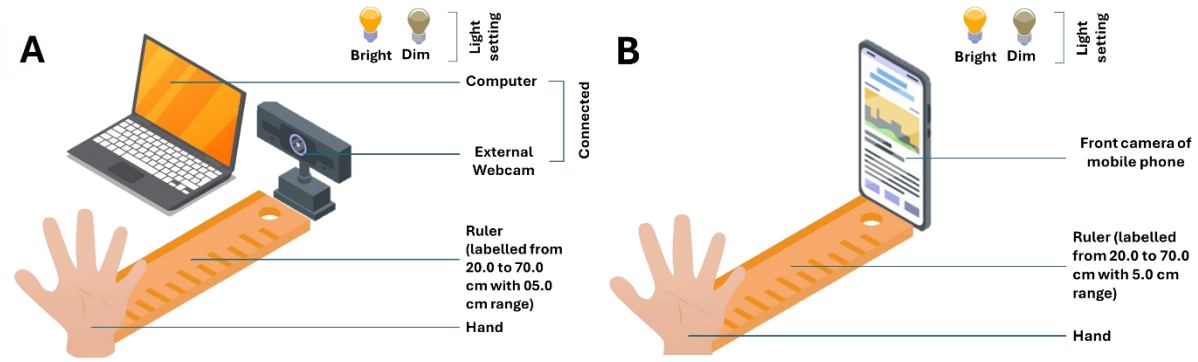

Figure 7. Setup for post-calibration was organized according to the discussed DoE elements.

All experiments were conducted in a controlled environment, ensuring consistency in lighting and camera placement. For example, the brightness of the light bulb and the distance are labeled every 5.0 cm increment gap. Finally, MediaPipe processed these data and integrated them into the RIOT system for gesture recognition. Then, the output of the experiments was analyzed for further assessment.

4.2. Results of Post-Calibration Adjustments

The results are categorized based on key factors that affect the system's performance. These factors include distance, camera type, and lighting conditions, which were systematically tested to assess the system's accuracy, precision, recall, F1 score, and latency. Table 2, Table 3, Table 4, and Table 5 were demonstrated as the brief findings of the experiment based on DoE factors.

4.2.1. Performance by Distance

The distance between the camera and the subject's hand significantly impacted the system's accuracy and latency. As the distance increased from 20 cm to 70 cm, there was a noticeable improvement in accuracy, precision, and F1 score. However, latency also increased at these distances, reflecting longer processing times. However, the trade-off is acceptable considering the significant gains in accuracy and F1 score. At 20 cm, the accuracy was 0.40, but at 60.00 cm, the accuracy increased to 0.97. The F1-score also increased from 0.43 at 20.00 cm to 1.00 at 60.00 cm, indicating an improved balance between precision and recall at greater distances. This trend aligns with the findings of a demonstration of gesture recognition accuracy improving with distance by achieving an average accuracy of 96.64% in their optimized model to reduce interference from background noise. Table 2 indicates results showing that the system performs optimally at 50.00 to 60.00 cm, where accuracy and F1-score reach their highest levels despite the increase in latency.

Similarly, the importance of distance in their real-time gesture recognition system is also noted in improved performance metrics at greater distances. The increase in latency with distance, while a drawback, is acceptable due to the significant gains in accuracy and F1 score, resulting in the notion that distance is a crucial factor in gesture recognition systems.

https://doi.org/10.31436/iiumej.v26i1.3455

Table 2. Comparison of all evaluation metrics and latency for gesture recognition by distances between the camera and patients' hand.

Distance (cm)	Mean Accuracy (%)		Mean Precision (%)		Mean Recall (%)		Mean F1-score (%)		Mean Latency (ms)	
	Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation
20.0	0.40	0.32	0.33	0.47	1.14	0.58	0.43	0.37	0.54	0.42
25.0	0.44	0.37	0.67	0.36	1.74	1.21	0.51	0.45	0.61	0.50
30.0	0.50	0.36	0.58	0.24	0.93	0.35	0.57	0.42	0.64	0.41
35.0	0.53	0.37	0.58	0.30	1.14	0.63	0.60	0.41	0.68	0.44
40.0	0.59	0.32	0.63	0.42	1.59	1.36	0.65	0.38	0.74	0.40
45.0	0.75	0.30	0.46	0.40	1.98	1.39	0.80	0.32	0.91	0.36
50.0	0.82	0.19	0.71	0.21	1.69	1.06	0.93	0.18	1.01	0.18
55.0	0.86	0.13	0.63	0.28	2.03	1.35	0.98	0.03	1.05	0.08
60.0	0.97	0.03	0.67	0.36	1.34	1.06	1.00	0.00	1.02	0.10
65.0	0.89	0.10	0.50	0.44	1.99	1.25	1.00	0.01	1.08	0.12
70.0	0.89	0.10	0.54	0.50	1.27	0.43	0.98	0.07	1.03	0.07

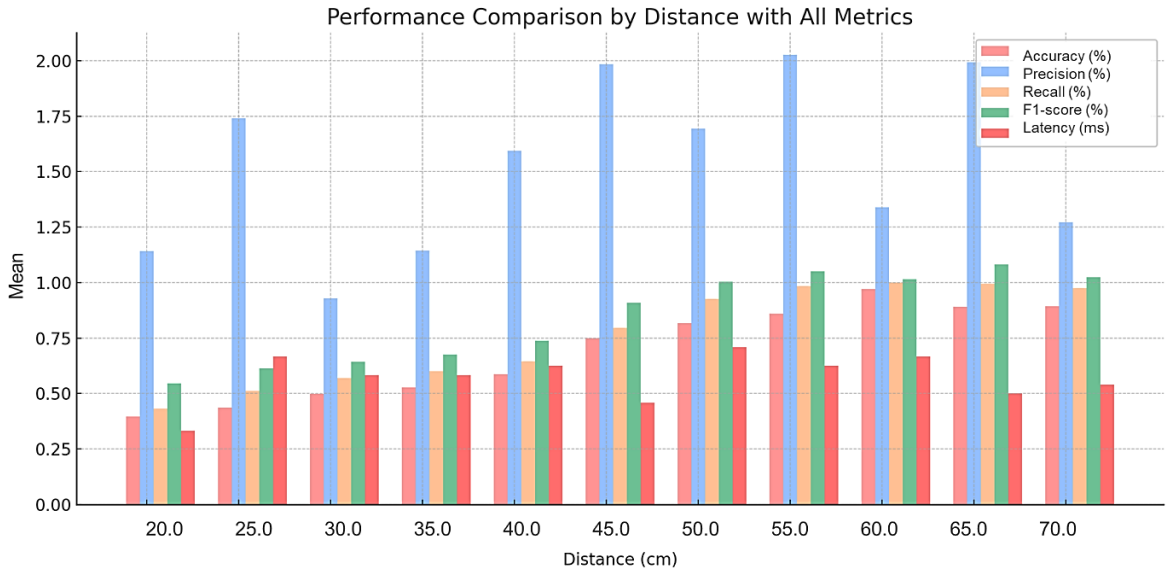

Figure 8. Various distances were determined in relation to the structure of the environmental factors of the post-calibration period. The bar chart effectively summarizes the findings across the metrics.

Accuracy and F1-score improved significantly at distances between 50.00 cm and 60.00 cm, where the system achieved the highest level of gesture recognition accuracy. The accuracy at this range is attributed to the effective capture of hand features and gestures, which diminishes at greater distances due to increased latency and reduced resolution.

Several scientific factors support the optimal distance of 60.00 cm for hand gesture recognition systems. At this distance, the system achieves high accuracy and F1-score, critical metrics for evaluating the performance of gesture recognition algorithms. As shown in Figure 8, the latency increases as distance increases, but it remains within an acceptable range, making the system responsive and practical for real-time hand gesture recognition. This is particularly true in the context of image recognition. As the distance increases, larger images must be processed, which can lead to increased computational demands and longer processing times.

https://doi.org/10.31436/iiumej.v26i1.3455

4.2.2. Performance by Camera Types

An external webcam and a mobile phone camera have performed some tests on gesture recognition for both exercises to assess their impact on accuracy and latency. Table 3 shows that both cameras were tested and exhibited certain patterns. The comparative analysis of camera types reveals that while the mobile phone camera provides slightly better accuracy (0.70) compared to the external camera (0.68), the external camera outperforms in terms of precision, recall, and F1-score and it is clear that precision and low latency are dominant, according to Figure 9.

Table 3. Evaluation metrics and latency are based on the types of cameras used to capture hand gestures for both stroke rehabilitation exercises.

Camera Type	Mean Accuracy (%)		Mean Precision (%)		Mean Recall (%)		Mean F1-score (%)		Mean Latency (ms)	
	Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation
External camera	0.68	0.31	0.47	0.35	1.59	1.08	0.79	0.35	0.85	0.36
Mobile phone camera	0.70	0.33	0.67	0.36	1.47	1.02	0.75	0.35	0.84	0.36

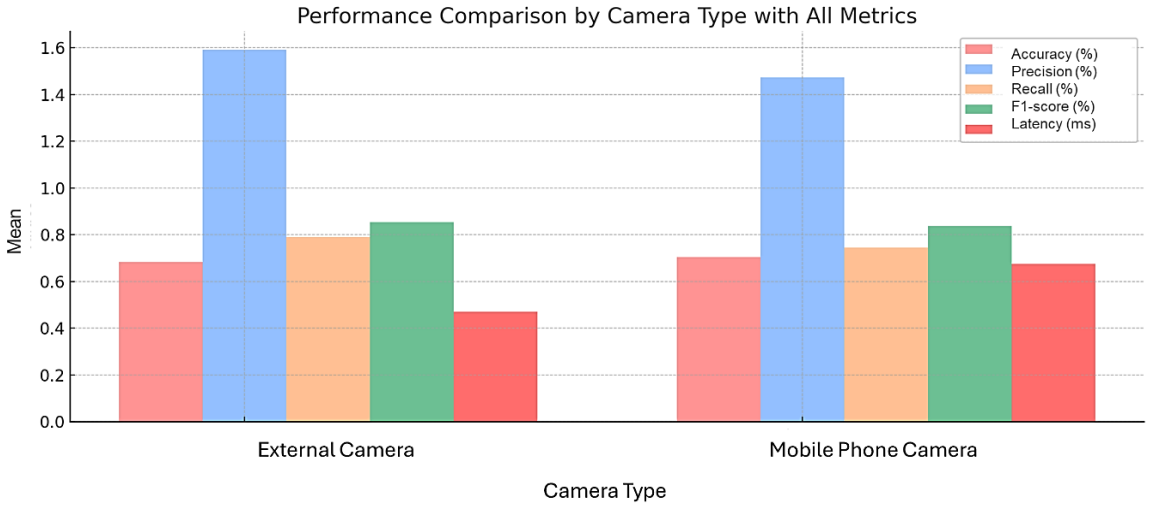

Figure 9. External camera and Mobile Phone Camera front camera present slight differences during the evaluation

These results specify that while the mobile phone camera provides marginally better accuracy, the external camera offers faster processing with higher precision, making it more efficient in real-time rehabilitation scenarios.

The system must respond promptly and prevent false positives, making it evident that an external camera is more suitable for this application. The findings align with research on the importance of camera quality in gesture recognition systems, highlighting that superior cameras result in improved recognition rates. Therefore, although both camera types are effective, the external camera's enhanced precision and reduced latency make it better suited for real-time rehabilitation scenarios.

https://doi.org/10.31436/iiumej.v26i1.3455

4.2.3. Performance by Lighting Conditions

The lighting conditions during the experiment played a role in determining the system's performance. Investigations were conducted under both bright and dim lighting to evaluate how ambient light affects accuracy, precision, and latency, as shown in Table 4.

Table 4. The simplified version of average mean percentages of metrics under lighting settings.

Lighting Condition	Mean Accuracy (%)		Mean Precision (%)		Mean Recall (%)		Mean F1-score (%)		Mean Latency (ms)	
	Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation
Bright	0.72	0.30	1.84	1.20	0.81	0.33	0.91	0.35	0.67	0.38
Dim	0.67	0.34	1.23	0.77	0.73	0.37	0.78	0.37	0.47	0.32

Referring to Figure 10, the system performed much better in bright lighting, with an accuracy of 0.72 and a precision of 1.84, compared to dim lighting, where accuracy dropped to 0.67 and a precision of 1.23. While dim lighting provided a slight speed advantage with lower latency (0.47 ms vs. 0.67 ms), this was not enough to compensate for the lower performance. Therefore, bright lighting is better for more accurate and precise gesture recognition. In addition to lighting, other factors, such as reinforcing what has already been found in other research, highlight how important it is to fine-tune these conditions for better real-world performance [30].

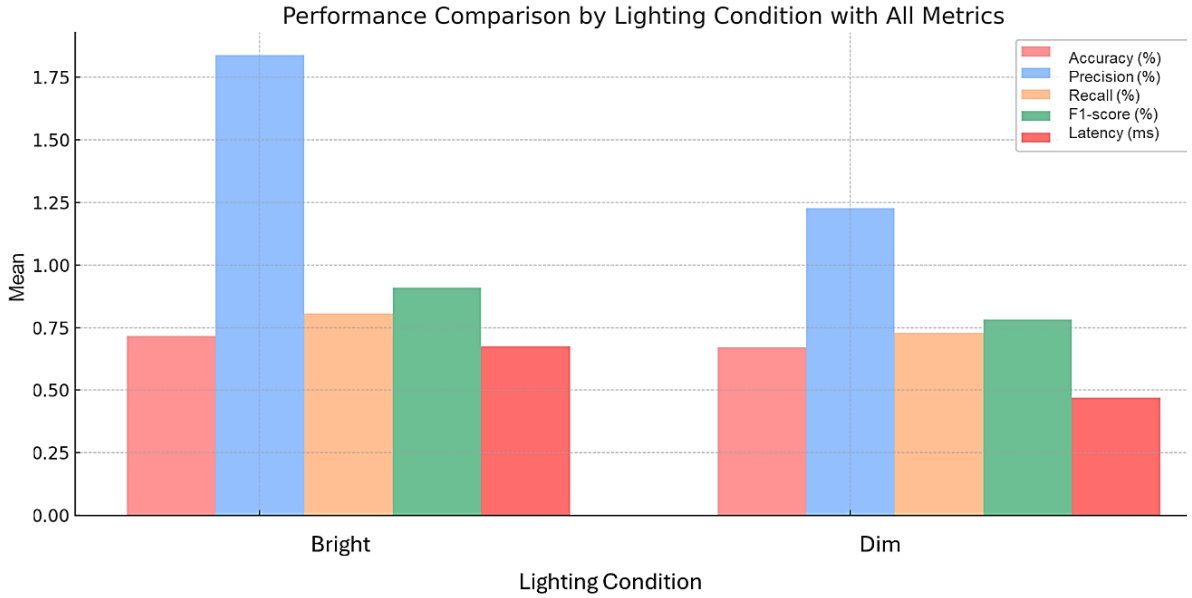

Figure 10. The chart illustrates the impact of bright and dim lighting conditions on the system.

4.2.4. Outline of Post-Calibration Outcomes

Post-calibration adjustments were necessary to optimize the hand gesture recognition system within the RIOT framework. The calibration process focused on refining the system's accuracy, precision, recall, F1-score, and latency across varying distances, camera types, and lighting conditions. In this segment, Figure 11 and Table 5 provide insights into the post-calibration effects. The results showed that distance and camera accuracies averaged 0.69, with

https://doi.org/10.31436/iiumej.v26i1.3455

precision at 1.53. However, recall and F1-score were lower with distance, and latency increased to 1.62 ms, reflecting longer processing times.

Table 5. The finalized average means of factors and their evaluations in hand gesture recognition through the RIOT website.

Factors	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Latency (ms)
Distances	0.69	1.53	0.59	0.36	1.62
Camera	0.69	1.53	0.77	0.85	0.57
Lighting	0.70	1.54	0.77	0.85	0.57

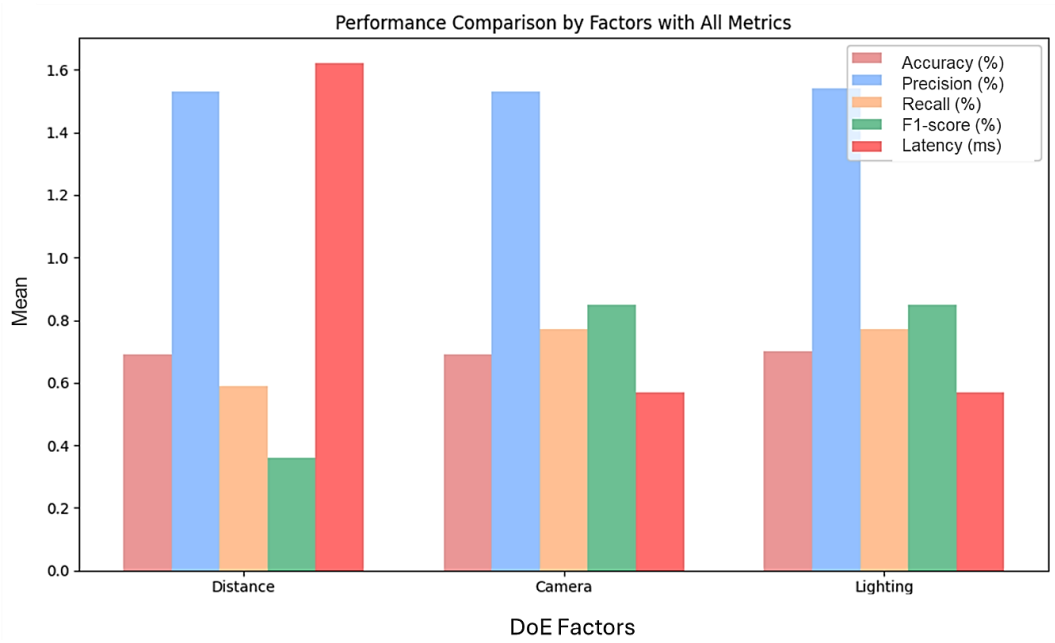

Figure 11. The grouped bar chart concluded the system's execution across three factors—distance, camera type, and lighting—evaluating metrics such as accuracy, precision, recall, F1-score, and latency.

Briefly, the optimal distance is 60.00 cm; the preference for external cameras due to their precision where it can avoid false positives, lower latency from the external camera, and the necessity of bright lighting conditions collectively enhance the system's reliability and effectiveness in real-world rehabilitation applications. These findings align with existing literature, confirming the validity of the post-calibration adjustments made to the system.

5. CONCLUSION

In conclusion, the RIOT system presents an innovative real-time, remote rehabilitation solution using pre-trained deep learning models like MediaPipe, recognizing CNNs and LSTMs to accurately detect hand gestures that can be considered for RIOT improvement in the future. Consequently, post-calibration adjustments have enhanced its performance across various conditions, making it suitable for at-home therapy. However, the system's effectiveness in real-world scenarios remains to be validated through clinical trials. Future work will focus on integrating adaptive learning and wearable technology and conducting broader testing to fully realize the system's potential in improving patient outcomes. As research advances in this

area, the potential for hand gesture recognition to transform rehabilitation practices and improve patient outcomes is immense. An edge computing approach can also impact healthcare, the environment, and scientific involvement in the computing world.

ACKNOWLEDGEMENT

This work is funded by the Ministry of Higher Education (MOHE) Malaysia under FUNDAMENTAL RESEARCH GRANT SCHEME (FRGS): FRGS23-307-0916; FRGS/1/2023/TK07/UIAM/02/2; Formulation of Associating 4D Skeletal-based Hand Gesture Recognition with Hand Rehabilitation Scoring of Stroke Patient for Rehabilitation Internet-of-things (RIOT). The assistance and resources provided by the National Medical Research Register (NMRR), particularly under NMRR ID-24-02136-NJQ, were instrumental in successfully completing this paper.

REFERENCES

[1] Y. J. Choo and M. C. Chang, "Use of Machine Learning in Stroke Rehabilitation: A Narrative Review," *Brain Neurorehabilitation*, vol. 15, no. 3, Nov. 2022, doi: 10.12786/bn.2022.15.e26.

[2] W. Zhang, C. Su, and C. He, "Rehabilitation Exercise Recognition and Evaluation Based on Smart Sensors With Deep Learning Framework," *IEEE Access*, vol. 8, pp. 77561–77571, 2020, doi: 10.1109/ACCESS.2020.2989128.

[3] X. Li and J. Zhong, "Upper Limb Rehabilitation Robot System Based on Internet of Things Remote Control," *IEEE Access*, vol. 8, pp. 154461–154470, 2020, doi: 10.1109/ACCESS.2020.3014378.

[4] S. H. Chae, Y. Kim, K.-S. Lee, and H.-S. Park, "Development and Clinical Evaluation of a Web-Based Upper Limb Home Rehabilitation System Using a Smartwatch and Machine Learning Model for Chronic Stroke Survivors: Prospective Comparative Study," *JMIR MHealth UHealth*, vol. 8, no. 7, p. e17216, Jul. 2020, doi: 10.2196/17216.

[5] on behalf of the Writing Expert Group of Expert Consensus on Clinical Application of IOT Medical Technology in the Rehabilitation of Chronic Obstructive Pulmonary Disease, the Respiratory Disease Rehabilitation Professional Committee of China Medical Education Association *et al.*, "Clinical guidelines on the application of Internet of Things (IOT) medical technology in the rehabilitation of chronic obstructive pulmonary disease," *J. Thorac. Dis.*, vol. 13, no. 8, pp. 4629–4637, Aug. 2021, doi: 10.21037/jtd-21-670.

[6] J. T. Kelly, K. L. Campbell, E. Gong, and P. Scuffham, "The Internet of Things: Impact and Implications for Health Care Delivery," *J. Med. Internet Res.*, vol. 22, no. 11, p. e20135, Nov. 2020, doi: 10.2196/20135.

[7] A. A. Zainuddin, N. H. M. Dhuzuki, A. A. Puzi, M. N. Johar, and M. Yazid, "Calibrating Hand Gesture Recognition for Stroke Rehabilitation Internet-of-Things (RIOT) Using MediaPipe in Smart Healthcare Systems," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 7, 2024, doi: 10.14569/IJACSA.2024.0150756.

[8] N. H. Mohd Dhuzuki, A. A. Zainuddin, S. I. Kamarudin, D. Handayani, K. Subramaniam, and Mohd. I. Mohd. Tamrin, "Web-Based Medical Information System for Stroke Rehabilitation Intemet-of-Things (RIOT) Patients: A Prototype," in *2024 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, Kota Kinabalu, Malaysia: IEEE, Aug. 2024, pp. 326–330. doi: 10.1109/IICAIET62352.2024.10729958.

[9] "MediaPipe Solutions guide | Google AI Edge," Google AI for Developers. Accessed: Sep. 20, 2024. [Online]. Available: https://ai.google.dev/edge/mediapipe/solutions/guide

[10] B. Saravi *et al.*, "Clinical and radiomics feature-based outcome analysis in lumbar disc herniation surgery," *BMC Musculoskelet. Disord.*, vol. 24, no. 1, p. 791, Oct. 2023, doi: 10.1186/s12891-023-06911-y.

[11] S. Rajayyan and S. M. M. Mustafa, "Prediction of dementia using machine learning model and performance improvement with cuckoo algorithm," *Int. J. Electr. Comput. Eng. IJECE*, vol. 13, no. 4, 2023, doi: 10.11591/ijece.v13i4.pp4623-4632.

[12] R. Magar, L. Ghule, J. Li, Y. Zhao, and A. B. Farimani, "FaultNet: A Deep Convolutional Neural Network for Bearing Fault Classification," *IEEE Access*, vol. 9, 2021, doi: 10.1109/access.2021.3056944.

[13] C. Lugaresi *et al.*, "MediaPipe: A Framework for Perceiving and Processing Reality".

[14] D. Parashar, "A Deep Learning-Based Approach for Hand Sign Recognition Using CNN Architecture," *Rev. Intell. Artif.*, 2023, doi: 10.18280/ria.370414.

[15] J. A. Díez, A. Blanco, J. M. Catalán, F. J. Badesa, L. D. Lledó, and N. García-Aracil, "Hand exoskeleton for rehabilitation therapies with integrated optical force sensor," *Adv. Mech. Eng.*, vol. 10, no. 2, p. 168781401775388, Feb. 2018, doi: 10.1177/1687814017753881.

[16] W. Zhang, T. Zhao, J. Zhang, and Y. Wang, "LST-EMG-Net: Long Short-Term Transformer Feature Fusion Network for sEMG Gesture Recognition," *Front. Neurorobotics*, 2023, doi: 10.3389/fnbot.2023.1127338.

[17] Z. Slimane, K. Lakhdari, and D. Souhila Korti, "Enhancing Dynamic Hand Gesture Recognition using Feature Concatenation via Multi-Input Hybrid Model," *Int. J. Electr. Comput. Eng. Syst.*, vol. 14, no. 5, pp. 535–546, Jun. 2023, doi: 10.32985/ijeces.14.5.5.

[18] I.-J. Ding and N. Zheng, "RGB-D Depth-Sensor-Based Hand Gesture Recognition Using Deep Learning of Depth Images With Shadow Effect Removal for Smart Gesture Communication," *Sens. Mater.*, 2022, doi: 10.18494/sam3557.

[19] Y. Huo, J. Shen, X. Chen, and K. Yu, "A dynamic gesture recognition method based on R(2+1)D-transformer network," in *Third International Conference on Computer Graphics, Image, and Virtualization (ICCGIV 2023)*, Y. Wang and A. J. Moshayedi, Eds., Nanjing, China: SPIE, Nov. 2023, p. 57. doi: 10.1117/12.3008203.

[20] T. Bao, S. A. R. Zaidi, S. Xie, P. Yang, and Z.-Q. Zhang, "A CNN-LSTM Hybrid Model for Wrist Kinematics Estimation Using Surface Electromyography," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021, doi: 10.1109/TIM.2020.3036654.

[21] M. Jaén-Vargas *et al.*, "A Deep Learning Approach to Recognize Human Activity Using Inertial Sensors and Motion Capture Systems," in *Frontiers in Artificial Intelligence and Applications*, A. J. Tallón-Ballesteros, Ed., IOS Press, 2021. doi: 10.3233/FAIA210196.

[22] J.-S. Kim, M.-G. Kim, and S.-B. Pan, "Two-Step Biometrics Using Electromyogram Signal Based on Convolutional Neural Network-Long Short-Term Memory Networks," *Appl. Sci.*, vol. 11, no. 15, p. 6824, Jul. 2021, doi: 10.3390/app11156824.

[23] G. Palanisamy and S. S. Thangaswamy, "An Efficient Hand Gesture Recognition Based on Optimal Deep Embedded Hybrid Convolutional Neural Network-long Short Term Memory Network Model," *Concurr. Comput. Pract. Exp.*, 2022, doi: 10.1002/cpe.7109.

[24] D. Copaci, J. Arias, M. Gómez-Tomé, L. Moreno, and D. Blanco, "sEMG-Based Gesture Classifier for a Rehabilitation Glove," *Front. Neurorobotics*, vol. 16, May 2022, doi: 10.3389/fnbot.2022.750482.

[25] C. Bagath Basha, "Enhancing Healthcare Data Security Using Quantum Cryptography for Efficient and Robust Encryption," *J. Electr. Syst.*, vol. 20, no. 5s, pp. 2070–2077, Apr. 2024, doi: 10.52783/jes.2544.

[26] A. Cisnal, J. Pérez-Turiel, J. Fraile, D. Sierra, and E. de la Fuente, "RobHand: A Hand Exoskeleton With Real-Time EMG-Driven Embedded Control. Quantifying Hand Gesture Recognition Delays for Bilateral Rehabilitation," *IEEE Access*, vol. 9, pp. 137809–137823, 2021, doi: 10.1109/ACCESS.2021.3118281.

[27] A. K. Panda, R. Chakravarty, and S. Moulik, "Hand Gesture Recognition using Flex Sensor and Machine Learning Algorithms," *2020 IEEE-EMBS Conf. Biomed. Eng. Sci. IECBES*, pp. 449–453, 2021, doi: 10.1109/IECBES48179.2021.9398789.

[28] S. Zhang *et al.*, "Real-Time and Accurate Gesture Recognition With Commercial RFID Devices," *IEEE Trans. Mob. Comput.*, vol. 22, pp. 7327–7342, 2023, doi: 10.1109/TMC.2022.3211324.

[29] K. Bell *et al.*, "Verification of a Portable Motion Tracking System for Remote Management of Physical Rehabilitation of the Knee," *Sensors*, vol. 19, no. 5, 2019, doi: 10.3390/s19051021.

[30] P. Lin, R. Zhuo, S. Wang, Z. Wu, and J. Huangfu, "LED Screen-Based Intelligent Hand Gesture Recognition System," *IEEE Sens. J.*, vol. 22, pp. 24439–24448, 2022, doi: 10.1109/JSEN.2022.3219645.