Computer Systems Science & Engineering
DOI: 10.32604/csse.2023.034431
Article

Tech Science Press

A Novel Machine Learning–Based Hand Gesture Recognition Using HCI on IoT Assisted Cloud Platform

Saurabh Adhikari[1], Tushar Kanti Gangopadhayay[1], Souvik Pal[2,3], D. Akila[4], Mamoona Humayun[5], Majed Alfayad[6] and N. Z. Jhanjhi[7,*]

[1]School of Engineering, Swami Vivekananda University, India
[2]Department of Computer Science and Engineering, Sister Nivedita University (Techno India Group) Kolkata, West Bengal, India
[3]Sambalpur University, Sambalpur, India
[4]Department of Computer Applications, Saveetha College of Liberal Arts and Sciences, SIMATS Deemed to be University, Chennai, India
[5]Department of Information Systems, College of Computer and Information Sciences, Jouf University, KSA
[6]College of Computer and Information Sciences, Jouf University, Sakaka, 72341, Saudi Arabia
[7]School of Computer Science (SCS), Taylor's University, Subang Jaya, 47500, Malaysia
*Corresponding Author: N. Z. Jhanjhi. Email: noorzaman.jhanjhi@taylors.edu.my
Received: 16 July 2022; Accepted: 23 November 2022

Abstract: Machine learning is a technique for analyzing data that aids the construction of mathematical models. Because of the growth of the Internet of Things (IoT) and wearable sensor devices, gesture interfaces are becoming a more natural and expedient human-machine interaction method. This type of artificial intelligence that requires minimal or no direct human intervention in decision-making is predicated on the ability of intelligent systems to self-train and detect patterns. The rise of touch-free applications and the number of deaf people have increased the significance of hand gesture recognition. Potential applications of hand gesture recognition research span from online gaming to surgical robotics. The location of the hands, the alignment of the fingers, and the hand-to-body posture are the fundamental components of hierarchical emotions in gestures. Linguistic gestures may be difficult to distinguish from nonsensical motions in the field of gesture recognition. Linguistic gestures may be difficult to distinguish from nonsensical motions in the field of gesture recognition. In this scenario, it may be difficult to overcome segmentation uncertainty caused by accidental hand motions or trembling. When a user performs the same dynamic gesture, the hand shapes and speeds of each user, as well as those often generated by the same user, vary. A machine-learning-based Gesture Recognition Framework (ML-GRF) for recognizing the beginning and end of a gesture sequence in a continuous stream of data is suggested to solve the problem of distinguishing between meaningful dynamic gestures and scattered generation. We have recommended using a similarity matching-based gesture classification approach to reduce the overall computing cost associated with identifying actions, and we have shown how an efficient feature extraction method can be used to reduce the thousands of single gesture information to four binary digit gesture codes. The findings from the simulation support the accuracy, precision, gesture recognition, sensitivity, and efficiency rates. The Machine Learning-based Gesture Recognition Framework (ML-GRF) had an

accuracy rate of 98.97%, a precision rate of 97.65%, a gesture recognition rate of 98.04%, a sensitivity rate of 96.99%, and an efficiency rate of 95.12%.

Keywords: Machine learning; gesture recognition framework; accuracy rate; precision rate; gesture recognition rate; sensitivity rate; efficiency rate

1 Introduction

A computer's ability to recognize human body language begins with the identification of hand gestures [1]. Human-computer interaction (HCI) applications including android TV controls, interactive media, telesurgery, and holograms all benefit from this technology. Hand gesture recognition can be utilized for various purposes, including translating sign language [2]. To represent human communication and sentiments, sign language hand movements are constructed in a complicated manner [3]. Furthermore, the hand's positioning and posture in connection with its body are crucial to these manual expressions' foundation. All of these complementing primitives should be considered in a succession of frames in an efficient recognition system [4]. Although these frames are time-dependent, it is impossible to evaluate the blocks in Euclidean space because of this time dependency [5]. Most current recognition methods merely take into account the hand's localized configuration [6]. Depending on the system, either a segmented hand area is received as input or a hand recognition preprocessing phase is conducted utilizing skin pigmentation models or colored mittens [7]. This kind of system works well for basic alphabets and numbers, but it doesn't work as well for true sign language movements since it relies on the global setup [8]. When compared to other systems, this one takes into account each finger's unique setup. There are some HCI applications where these systems have worked well, but when it comes to recognizing real sign language gestures, they have not [9].

Recently, there seems to have been a marked increase in interest in the construction of intuitive natural user interfaces, which is a promising area of study [10]. Such interfaces must be imperceptible to the user, enabling them to engage with an application in an informal style without the need for sophisticated and expensive equipment [11]. They must provide natural engagement with the user and be flexible to the user's preferences without the need for complicated calibration processes. At the same time, they must perform their functions in real-time, with extreme accuracy, and with sufficient robustness to withstand background noise [12]. Despite the variety and complexity of these needs, researchers are nevertheless faced with substantial obstacles. Hand gestures may be regarded as an intuitive and easy means of communication between humans and robots since they are a potent mode of interhuman communication. An effective natural user interface must consequently have the capacity to identify hand movements in real-time [13].

Data gathering, hand positioning, hand feature identification, and action recognition based on the detected characteristics are the main components of a hand recognition system [14]. Conventional data collection methods include colored cameras, which have already been used for gesture identification. Computer vision and machine learning experts are working hard to improve their systems' ability to recognize hand movements [15]. Hand gesture recognition is a critical component of the physical-digital interface [16]. It has progressed from physical touch to virtual gesture-based interactions throughout the years. Automatic systems, intelligent machines, and simulated games are only a few uses for a hand gesture detection system [17]. Using a hand-recognition system in a smaller device requires a lower digital storage capacity and simpler and faster processing speed. The main goal of this method has been to take all of these things into account so that the planned system can be used despite all of the system's limitations [18].

The intricacy of the backdrop, the blurriness of the picture, and the angle of the motion all contribute to the difficulty of hand gesture detection [19]. Developers are focusing on reducing the amount of space needed for the system's functions as more and more digital information is required to be kept in the system [20]. This backs up the idea that low-resolution pictures that take up less space could be used to recognize gestures [21].

The following is the rest of the document. A literature review on gesture recognition is discussed in Section 2 of this paper. Details of the suggested approach are outlined in Section 3. Furthermore, Section 4 presents the results of the investigations that were carried out utilizing the proposed approach. The summary of findings is presented in Section 5.

1.1 Motivation

Creating a reliable hand gesture recognition system based on HCI is difficult. Therefore, systems reliant on vision-based hand gesture recognition typically necessitate a blend of expertise in application-specific programming, computational methods, and machine learning strategies. Deploying vision-based gesture-based systems in real-world settings is difficult for a number of reasons, including application-specific requirements, stability, and the accuracy of the camera sensor and lens characteristics. We have tried to overcome these issues in hand gesture recognition systems using HCI and machine learning approaches on the IoT Assisted Cloud Platform. The major goal of the suggested approach is to create a hand gesture recognition system that is simple, reliable, and efficient [22]. There are no classifiers in the suggested strategy since classifiers take time to train and their performance also rises. The geometrical structure of the hand is used to extract the information, not its hue, shape, or intensity. This guarantees the approaches' dependability and efficiency in low-light situations. As a result, the hand motion does not have to be in an upright posture. The suggested approach has an additional benefit.

1.2 Main Contributions

This section deals with the main contributions of the manuscript. The followings are the distinguishing characteristics of this research:

- Complex gestures are broken down into simple ones so that they can be recognized as simple sequences of basic gestures;
- The average jerk is an effective method for reducing thousands of single gestures to four binary-digit gesture codes.
- Also, it is suggested to use a similarity-matching-based approach to gesture classification to lower the overall cost of computing when figuring out what an action is.

2 Literature Survey

Kirishima et al. [23] investigated the Quadruple Visual Interest Point Strategy, where visual features are calculated from constantly changing areas of interest in a given image sequence that is not assumed to be shifted or flipped in any way. Each visual characteristic is referred to as a "visual interest point," and a probability density function is given to each one before the selection is made. To solve the second issue, we devised a selective control mechanism that allows the recognition system to monitor and manage its self-loading capabilities. The Quadruple Visual Interest Point Strategy (QVIPS) lays greater emphasis on spatially and temporally dependable visual interest points in image sequences provided as belonging to the same gesture class. The QVIPS framework can now recognize a wide range of gestures because it doesn't limit movement categories, motion orientations, or differences in how people make hand gestures.

Poularakis et al. [24] suggested a Reliable Hand Gesture Recognizer over continuous streaming digits and letters based on maximum cosine similarity and fast Nearest Neighbor (NN) algorithms for three core challenges of action recognition: standalone recognition, gesture authentication, and gesture identification over continuous streaming data. At least for basic trajectories like digits and characters, the authors have shown the experimental findings indicating that Maximum Cosine Similarity (MCS) can provide greater recognition accuracy at little computing cost. Digits are easier to read than letters, and body sensors are better than cameras at detecting them.

Plouffe et al. [25] discovered a Kinect-based natural gesture user interface that can identify and monitor real-time hand motions. The user's hand is assumed to be the nearest item in the scene to the camera, hence the interest space corresponding to the hands is initially divided. To speed up the scanning process, a new algorithm has been devised to detect the first pixel on the hand's contour inside this area. This pixel serves as the starting point for a directed search algorithm that identifies the full hand shape. Fingertip locations are found by using a technique called k-curvature, and dynamic time warping is utilized to pick action candidates and detect actions by comparing an observed motion to prepared reference motions. It also works the same way for identifying common signs both in motion and at rest and for the sign language alphabet.

Chen et al. [26] provided an innovative and robust gesture detection system that is unaffected by changes in ambient lighting or the backdrop. Hand gesture recognition systems utilize common vision sensors, such as Complementary Metal Oxide Semiconductor (CMOS) cameras, as wearable sensing devices in modern systems. The performance of these cameras is severely degraded by their use due to environmental restrictions such as illumination fluctuation and crowded backgrounds. Neuromorphic vision sensors with a microsecond sampling rate, high dynamic range, and reduced latency are used in this system. The sensor's output is a series of asynchronous events instead of discrete frames. Five high-frequency active Light-Emitting Diode (LED) markers (ALMs), each representing a finger or palm, are monitored accurately in the temporal domain using a constrained spatiotemporal particle filter technique to analyze the visual input.

Pramudita et al. [27] recommended using a spatial variety as a solution to the issue of inaccurate hand motion characteristics. To achieve spatial variety, an array radar was used as a sensor arrangement. In this research, an array of four Continuous Wave (CW) radars has been suggested as a contactless sensor for hand motions. Hand gesture Doppler responses were detected by each CW radar using peak detection based on cross-correlation. It becomes a characteristic of each hand motion when the timing position is employed. The HB 100 is used as a CW radar component in the CW radar array that operates at 10 GHz. Tests done with distance sensors that are 50 cm apart show that the hand motion characteristic can be distinguished with 96.6% accuracy. The Doppler effect can be used to tell the difference between hand gesture pairs that move in opposite directions, and basic data processing is needed to figure out what different hand gestures mean.

Zhou et al. [28] presented ultrasonic hand gesture identification using 40 kHz ultrasonic waves and an innovative classification technique. One transmitter, three receivers, and highly digitalized front-end hardware are used in this system. The RDM (Range Doppler Map) streams of gestures are collected by a coherent ultrasonic pulse train system with moving target identification capabilities. Each frame's RDM characteristics are mapped to the template sequences of each class using dynamic time warping in advance during training and inference. It is trained for each class using a two-class random forest that can detect whether or not a testing sample belongs to that class. Furthermore, the test sample's probability of belonging to various groups is compared. Tests demonstrate that the suggested classifier trained on six people has greater validation accuracy than the rivals when it comes to leaving one person out. A PC can recognize eight 3D motions in 37 milliseconds with a 93.9% accuracy rate. It has a model size of 700 KB. User-independent embedded applications are better served by this technology.

Vishwakarma et al. [29] suggested a convenient and successful technique to recognize hand motions from extremely low-quality photos. Processing digital photos have traditionally centered on enhancing low-resolution photographs for aesthetic and functional purposes. For recognition, images with low resolutions are examined. Webcam, cell phone, or low-cost camera images are analyzed methodically to determine the number of raised fingers. The detection of hand gestures from low-resolution photos is based on simple geometric notions of the hand. In addition to saving a lot of storage space, hand gesture detection from low-resolution photos decreases system processing time. This approach is ideal for situations when the background sources are not dynamic.

Gadekallu et al. [30] discussed one-hot encoding, which converts categorical data to binary. Then, a crow search algorithm (CSA) finds the ideal hyper-parameters for convolutional neural network training datasets. Removing extraneous parameters improves hand gesture classification. The model's 100% accuracy in training and testing indicates it's better than state-of-the-art models. Reddy et al. [31] discussed machine learning for text categorization. They've employed text records with SVM to learn about a system and see how therapeutic exchanges flow.

Abavisani et al. [32] used numerous modalities to train 3D-CNNs to recognize dynamic hand movements. They created an SSA loss to ensure that features from different networks have the same content. The "focal regularization parameter" they created evened out this loss to prevent negative knowledge transfer. Our system improves the accuracy of recognition at test time for unimodal networks and works best on datasets of hand gestures that change over time.

Ahmed et al. [33] optimized the CNN. They tried different things until they got the greatest accuracy-to-speed balance for the training dataset. Using an Impulse Radio (IR) radar sensor and a Convolutional Neural Network, they can recognize finger-counting hand movements. The proposed strategy was accurate in the real world. Chen et al. [34] have developed a CNN approach that enhances classification accuracy and reduces model parameters. Model testing used the Ninapro DB5 and Myo datasets. Gesture recognition was categorized effectively.

Oudah et al. [35] discussed the pros and weaknesses of hand gesture approaches. It also makes a table showing how well these approaches function, focusing on computer vision techniques that deal with similarity and difference points; hand segmentation; classification algorithms and their defects; quantity and types of movements; dataset; detection range (distance), and camera type. A summary of the studies cited in the Literature Review has been tabulated in Table 1.

Table 1: A summary of the studies cited in the literature review

Paper cited	Methods used	Key findings	Limitations
Gadekallu et al. [30]	Crow Search algorithm (CSA)	Compared to other methods, our performance evaluation shows that it works 100% of the time during training and testing, even though it only takes 16 min to train.	If this study were to be improved, it would benefit from the addition of high-quality multi-modal formulations.
Reddy et al. [31]	SVM (Support Vector System)	In order to better see how therapy sessions progress, a system learning approach was used. The accuracy rating of the SVM model is lower than that of current technology.	This paper's category correctness is inadequate to minimize the professional advisor's weight of the rectification guide.

(Continued)

Table 1 (continued)

Paper cited	Methods used	Key findings	Limitations
Abavisani et al. [32]	Unimodal 3D convolutional neural networks (3D-CNNs)	This framework outperforms state-of-the-art methods on a wide range of dynamic hand gesture recognition datasets while simultaneously increasing the test-time recognition accuracy of unimodal networks.	Can the method of Unimodal Dynamic Hand-Gesture Recognition be utilized for multimodal learning?
Ahmed et al. [33]	Convolutional Neural Network-based technique	The proposed approach achieved a level of accuracy suitable for practical use.	This Model only considers finger counts for one hand. This is the disadvantage of this strategy.
Chen et al. [34]	Convolution Neural Network (CNN) model	On both the Ninapro DB5 and the Myo Datasets, the classification accuracy of gesture recognition performed admirably.	The classification accuracy of both traditional machine learning and EMGNet is lower than that of the Myo Dataset.
Oudah et al. [35]	A review on hand gesture techniques and introduces their merits and limitations under different circumstances.	This paper gives a table that summarizes the results of different approaches, with a focus on computer vision methods that deal with issues of similarity and difference, hand segmentation technique, classification algorithms and their limits, gesture count and variety, dataset, detection range (distance), and camera type.	Due to the disparity in accuracy across the categorization algorithms, this method lacks some motions. Due to the matching dataset, utilizing a high number of the dataset requires more time.

3 Proposed Methodology

Many interrelated components and joints form a complicated anatomical framework that provides a total of 27 degrees of freedom (DOFs) in the human hand. As a consequence, a thorough grasp of human anatomy is essential in determining what kinds of positions and motions are comfortable for users. There should be no confusion about the differences between various hand motions and postures. Hand posture is a gesture position in which no actions are involved. A hand posture, for example, is the act of creating a fist and holding it in a certain position. As opposed to this definition, a hand gesture is a dynamic action that includes a series of hand postures linked by continuous movements over a short period, such as gesturing a goodbye. As a result of this composite quality of hand gestures, the issue of gesture identification may be divided into two levels: the low level of hand position recognition and the high level of sign language recognition. IoT-integrated sensor devices transmit data via Bluetooth or a serial interface, which is then analyzed in real time. Additionally, offline acquisition of pressure sensor data may be collected using the records of prior online

sessions. Each occurrence can and will prompt the generation of data by intelligent devices and sensors. This information can then be sent back to the primary application through the network. At this step, it is necessary to determine which standard will be used to create the data and how it will be transferred across the network. Typically, MQTT, HTTP, and CoAP are used as the common protocols for returning this data. Each of these protocols facilitates the transmission of information or updates from a single device to a central point. In the Internet of Things, devices submit data to the core application, which then consumes, transmits, and makes use of the data. Data may be transferred in real-time or in groups at any time, depending on the device, the network, and the available power. The Device Layer, Communication Layer, IT Edge Layer, Event Processing Layer, and Client Communication Layer are utilized to collect data. For Internet of Things applications, these time-series data must be accurate. Fig. 1 illustrates that it is possible to train a system utilizing IoT-integrated sensor input data for real-time gesture detection and evaluate alternative algorithms under the same settings using recorded sessions' blocks with sensor nodes.

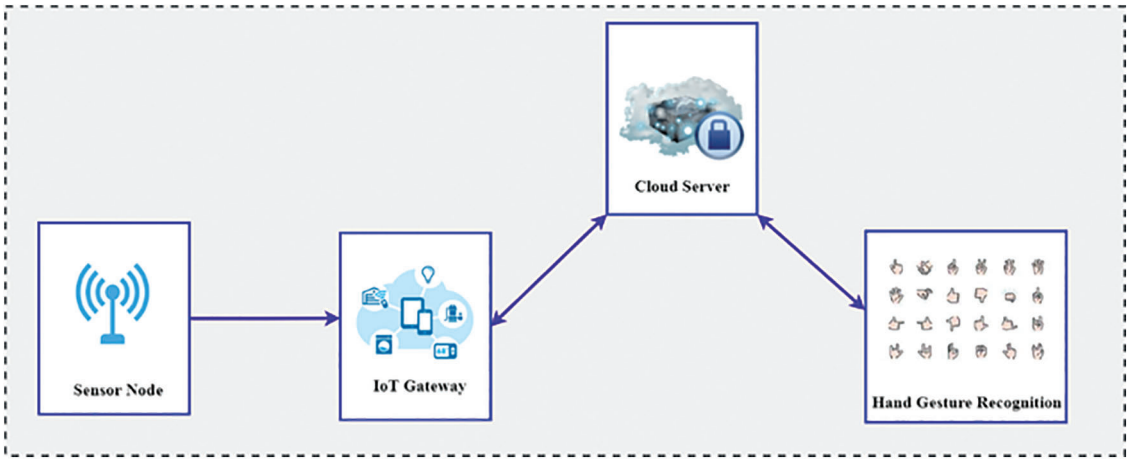

Figure 1: IoT block with sensor nodes

Machine Learning-based hand gesture recognition systems use video cameras to capture the movement of the hands to recognize the gestures. The input video is deconstructed into a collection of characteristics that consider individual frames. Some types of filtering may also be applied to the frames to eliminate any extraneous information and emphasize the important components. Examples include hand separation from other body parts as well as backdrop elements in a scene. It is possible to distinguish between various postures by looking at the solitary hands. Different postures of the solitary hands may be detected. A gesture recognizer may be trained against any conceivable language since gestures are nothing more than a series of hand postures linked together by continuous movements. It is now possible to describe hand gestures in terms of how they are constructed, just as sentences are composed of words. With the use of machine learning, this article explains how to recover the state of a 27-degree-of-freedom hand model from standard grayscale photos at up to ten frames per second (fps). Trigonometric functions that simulate joint mobility and perspective picture projection make the inverse mapping non-linear. Trigonometric functions that simulate joint mobility and perspective picture projection make the inverse mapping non-linear. Changing the settings produces a different picture, but it does so in an orderly fashion. A local linearity assumption might be made in this case. It is possible to solve non-linear equations using iterative approaches, such as Newton's method, that presuppose local linearity. This process of determining the best match for a particular frame is then repeated for the following frame, and so on. At each phase, the technique may be thought of as a sequence of hypotheses and tests, where a new hypothesis of model parameters is developed to reduce mis-correspondence in a certain direction. Using these parameters, the

picture is then compared to the model. Fig. 2 depicts the basic framework for ML-based gesture recognition. In our envisioned ML-based Gesture Recognition Framework, sensor nodes collect and interpret environmental input. The collected data are used for gesture detection and adaptation. The next stage for machine learning algorithms is to determine how to leverage the concepts of feature constellations to distinguish between the various database classes. Using the resulting classifier, it is now feasible to examine recognition results.

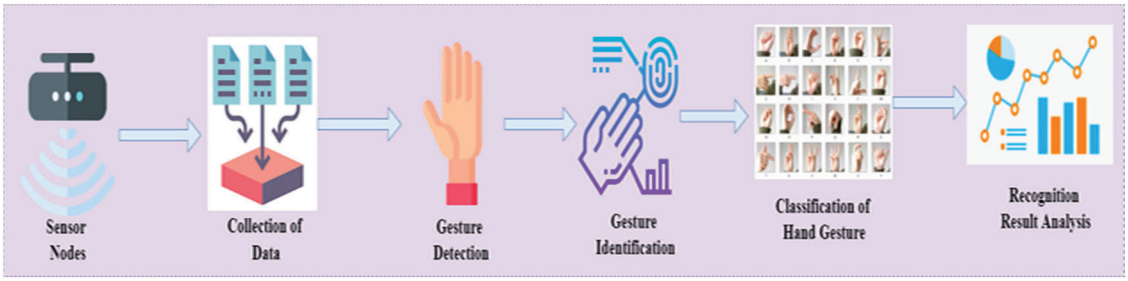

Figure 2: Basic framework for ML-based gesture recognition framework (ML-GRF)

Detection techniques that are sensitive to a complex backdrop tend to be the most common. The difficulties of distinguishing skin-colored items from one another and the sensitivity to illumination conditions make color-based hand detection problematic. Using shape models necessitates enough contrast between the item and the backdrop. It is the goal of data segmentation to demarcate the signal stream into active and inactive portions. The static actions are represented by the inactive segments when the hand is motionless. Furthermore, the active portions refer to the periods of time whenever the hand seems to be in motion that encompass expressive actions and the transition from one activity to another. Wrist pressure has been shown to correlate with the amount of muscular action. A minor shift in the form of the wrist occurs as the hand moves, due to the movement of tendons on the wrist. There will be a change in sensor data if such changes are recognized by the wrist pressure sensors. It is thus possible to separate active portions from inactive portions using a Machine Learning-based Gesture Recognition Framework in real time. The architecture for the proposed ML-GRE Framework is demonstrated in Fig. 3.

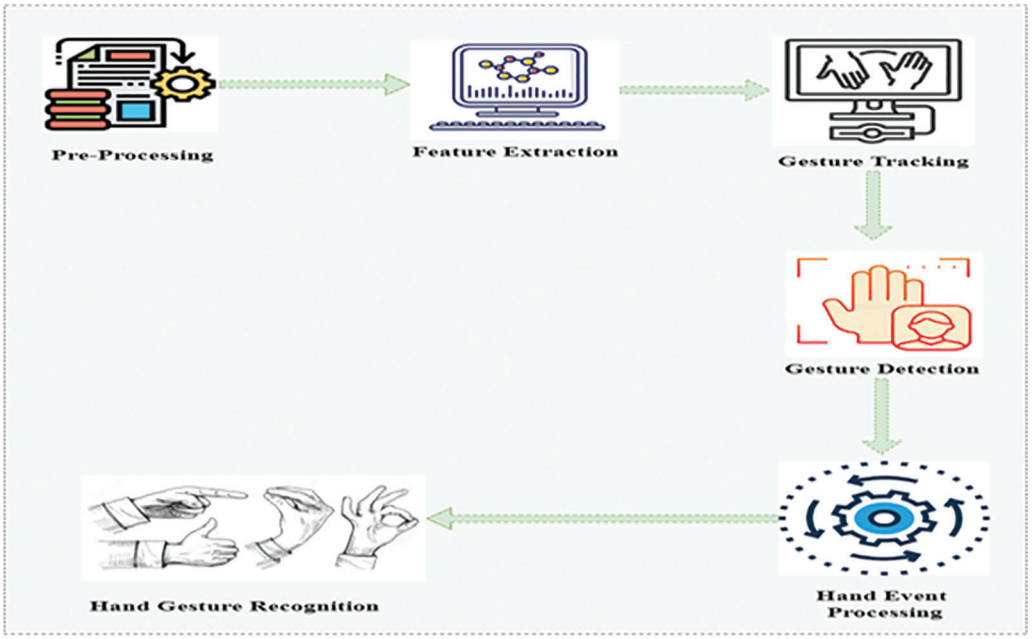

Figure 3: Architecture for the proposed framework

There are five static wrist motions, which are hand-motion states where the hand does not move. Stationary movements on the hand include the following: palm leftwards, palm upwards, palm right, palm downwards, and a slight hold. Active gestures, such as the ones made with three fingers, may be thought of as hand movements. All of the dynamic movements are small-scale, such as the percussion of the forefinger, clicking the forefinger twice, and clicking the forefinger and middle finger. It is necessary to conduct all dynamic motions immediately following the palm-flat condition.

Four wrist-mounted pressure sensors attached to the user collect pressure signal waveforms from four distinct channels. The pressure data stream is first processed using an overlapping sliding-window technique, with band c denoting the size of the window and a step value, respectively. This utilizes the term "B" with a sample rate of 300 to capture the whole motion or action, and the sample rate of 8 has been set to detect pressure changes with minimal delay.

The present window's highest peak to valley pressure rate for the channel is determined in Eq. (1),

$$T_{max} = \max_{k}(\max_{l}(PT_{k,l}) - \min_{l}(PT_{k,l})) \tag{1}$$

where T_max is the present window's highest peak-to-peak valley pressure rate and $PT_{k,l}$ symbolizes the k^{th} value of the channel l and k ranging from 1 to 4 and, l ranging from 1 to 300 in the present window. As a result, T_{max} can reliably monitor changes in pressure, even initial pressure varies. Finally, a threshold limit value is employed to distinguish between the active and idle or inactive portions as follows:

$$Condition = \begin{cases} active\ portion & T_{max} > threshold \\ inactive\ portion & T_{max} \leq threshold \end{cases} \tag{2}$$

Threshold may be used to identify whether the present window is active or idle based on a user-experience parameter as given in Eq. (2). It is recorded as an active portion when the present window has a T_{max} greater than its Threshold. Additionally, if the T_{max} of the present window is lower than the Threshold, the window is labeled as an idle portion in the recording.

There seems to be a correlation between gestures and muscular actions. Compared to other static gestures, Palm Flat's pressure signals are more consistent and less erratic than other motions with a more relaxed muscle. The dynamic movements are similarly initiated from the Palm-Flat state and all of them are minor motions, implying that signal modifications are also minimal. To identify the small differences in the signal induced by dynamic actions, a very small threshold is required. When there is a lot of noise around, it can be harder to tell when a stationary action is about to change. A dynamic threshold-based approach that instantly determines the threshold depending on the gesture condition of the last instant L_{final} is presented.

$$Threshold = \begin{cases} B_s & L_{final} = Plam\_Flat \\ B_h & others \end{cases} \tag{3}$$

The threshold only affects how sensitively a judgment is made about whether the present window belongs to an active or idle state. When the status of the gesture changes, the threshold is altered in real time as well. Using Eq. (3), when a flat motion is identified, the threshold is altered from a large threshold to a small one to identify dynamic gestures. To recognize static gestures using inactive portions, it is necessary to extract features that accurately describe the pressure changes induced by distinct static gestures and that are resistant to transitions. The wearable device's limitations necessitated the implementation of two simple features: the spatial feature and the temporal feature. Specifically, the spatial characteristic is intended to describe the spatial patterns of wrist stress, which is illustrated in the following equation:

$$a_k = \frac{\underset{k}{\text{mean}}\left(PT_{k,l}\right)}{\max\limits_{k}\left(\underset{k}{\text{mean}}\left(PT_{k,l}\right)\right)} \tag{4}$$

The average mean of the kth channel is represented as $\underset{k}{\text{mean}}\left(PT_{k,l}\right)$ in the present window. To indicate the difference between the present pressure and reference value of pressure G_k that may be determined from historical pressure data, the temporal characteristic d_k is employed in Eq. (4). Using Minimum-Maximum Normalization, the result values are then transferred to the range [0,1] as shown:

$$d_k = \frac{\underset{k}{\text{mean}}' - \underset{k}{\text{mean}}\left(\underset{k}{\text{mean}}'\right)}{\max\limits_{k}\left(\underset{k}{\text{mean}}'\right) - \min\limits_{k}\left(\underset{k}{\text{mean}}'\right)} \tag{5}$$

From Eq. (5), $\left(\underset{k}{\text{mean}}'\right) = \underset{k}{\text{mean}}\left(PT_{k,l}\right) - G_k$ defined as a difference between a channel kth mean and the reference rate of the kth channel. The average mean value of the four channels represented as $\underset{k}{\text{mean}}\left(\underset{k}{\text{mean}}'\right)$. G_k presents the pressure reference and is equal to $\frac{\sum_{k=1}^{UV} PT_{k,l}}{UV}$. Pressure relationships may be reflected in multiple directions since sensors are located on the front, backside, right, and left sides of the wrist using d_k value.

A variety of factors, including hand shapes, positions, and other scene elements, affect the strain or pressure on the wrist. Since gestures might have different absolute values in various contexts, this could lead to performance reduction when directly employed for action recognition. But when motions are switched around, the pattern of how pressure changes tend to stay the same.

To acquire an initial reference pressure value, the user must hold the palm in a flat position for nearly a moment during the starting process. In addition, the recommended solution is more user-friendly since no additional gesture data must be gathered during the setup phase. In a real-world application, the client can shift their hands or change the hand positions by using this system, which leads to a variation in wrist pressure distribution for a certain action. Thus, the reference pressure value is no longer appropriate. To deal with the current application problem, an adaptive pressure-parameter update method is shown to lessen the effect of scene changes.

First, we must separate dynamic actions. We need to distinguish between dynamic actions and other sorts of intrusive actions by transforming stationary actions and non-included motions to identify active portions. The active segment's signal energy might be utilized to separate dynamic motions from others since they all include small-scale activity. When the total energy of the current signal window goes over a certain limit, the current segment is recognized as the repetitive segment and is taken out of consideration.

Small-scale movements such as one- and two-finger clicking make it difficult to distinguish between single- and double-click motions for dynamic gestures. A method called dynamic time warping is used to compare two different time sequences of varying lengths. The greater the similarity between the time signals, the lesser the distance. A data pretreatment procedure is used before the use of the temporal wrapping algorithm, which makes the active segment signal more efficient and easier to recognize. Due to the high complexity of computing, a downscaling technique was made to simplify things while still keeping enough signal integrity for recognition. After normalizing the data from several channels, we keep the primary changing pattern of every channel and restrict the effect of other factors, like starting pressures as,

$$PT_{k,l,m} = \left(PT_{k,l} - PT_{k,1}\right) / \max_{l}\left(\left(PT_{k,l} - PT_{k,1}\right)\right) \tag{6}$$

where $PT_{k,1}$ denotes the beginning value of channel k and max l, and $PT_{k,l} - PT_{k,1}$ indicates the highest fluctuation in amplitude value compared with the beginning value of the channel k. The initial pressure is represented by Eq. (6). The flow diagram for starting pressures is depicted in Fig. 4, which can be found below (6). A temporal warping technique is then used to calculate the distance between previously collected gesture templates, an active data segment, and the one with a reduced distance, indicating that signal patterns are much more similar, which is chosen as the recognition outcome. This procedure is repeated until a recognition outcome has been determined.

$$F_0 = HTP\left(M_{active}, E_{single}\right)\ F_1 = HTP\left(M_{active}, E_{dual}\right) \tag{7}$$

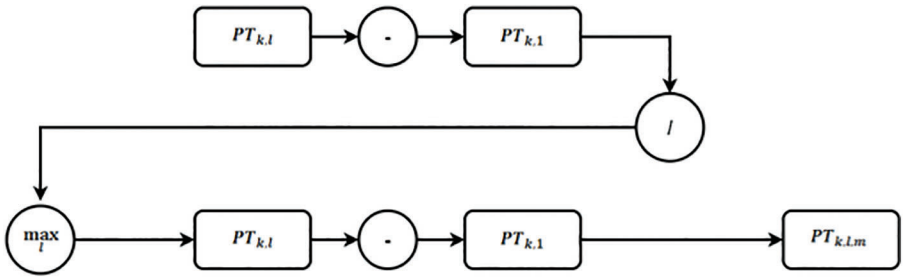

Figure 4: Flow diagram for starting pressures

M_{active} refers to the single-click motion. E_{single} describes the double-click gesture. The term HTP is referred to as temporal wrapping. There are two types of gesture templates: one for single-clicking and a second one for clicking twice. Fig. 5 illustrates the path flow for Eq. (7).

$$Action = \begin{cases} P_{single} & F_0 < F_1 \\ P_{dual} & F_0 \geq F_1 \end{cases} \tag{8}$$

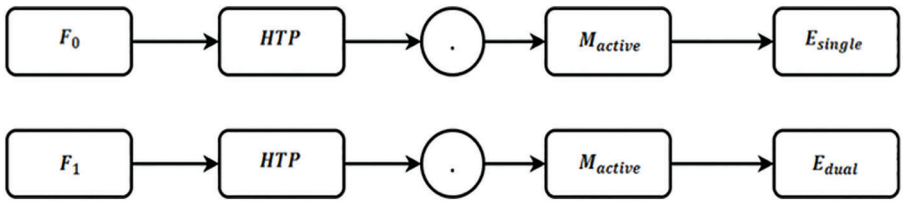

Figure 5: Path diagram for gesture template

A threshold-based approach is suggested to discriminate between the fore finger percussive action, the middle finger, and the fore finger percussive action is computes using Eq. (8). The maximum value $\max_{k}\left(PT_{k,l} - PT_{k,1}\right)$ of the initial active portion is computed to recognize the range of single-finger click actions and other actions because dual-finger clicks are greater than single-finger clicks.

$$Action = \begin{cases} PGK_1 \max_{k}\left(PT_{k,l} - PT_{k,1}\right) < R_{select} \\ PGK_2 \max_{k}\left(PT_{k,l} - PT_{k,1}\right) \geq R_{select} \end{cases} \tag{9}$$

As in Eq. (9), PGK_1 and PGK_2 depicts the single-finger click actions and dual-click actions respectively. R_{select} denotes the threshold for recognition. And our investigations have shown that a value of 400 is optimal. The complexity of the gesture is determined by the combination of the two actions. The fundamental gesture is detected by obtaining a feature from the average jerk among the beginning and ending points of the hand movement. To assess the performance of many representative techniques systematically, two tasks are used: gesture categorization in segmented data and gesture spotting and identification in continuous data. As part of our research, our empirical investigation gives a detailed look at how the input modality is chosen and how the domain adapts to different situations.

4 Results and Discussion

In a wide range of applications, legitimate hand gesture identification can be based on machine learning methods. In addition to assisting in the improvement of communication with deaf individuals, machine learning algorithms with IoT and Human-to-Computer Interactions (HCI) methods have also aided in the development of gesture-based signaling pathways, which are becoming more popular. The steps involved in gesture recognition are: a collection of data and preprocessing; classification and recognition of gestures; feature extraction and basic gesture recognition; basic encoding of gestures; and pattern matching for complicated gesture detection. This part includes a demonstration of the effectiveness of the suggested hand movement recognition method or technique via the use of two investigations: one on simple action detection and another on complicated action detection. The user-dependent and user-independent tests were conducted for both basic and complicated gesture detections to assess the degree to which the user relies on the system and its capabilities. The gesture detection technique comprises many processes, including accelerated acquisition and signal processing procedures, movement delineation and extraction of features, a classification model, basic gesture embedding, and similarity matching. It should be noted that the methods for basic gesture recognition never include basic gesture embedding or similarity matching; these two techniques are reserved for complex gesture detection. Individual variances resulted from gestures being produced at various frequencies, intensities, and sizes by various people at different times.

4.1 Accuracy Rate

In the investigation, the classification accuracy is tested to analyze the suggested hand gesture recognition method's performance. The guidelines for distinguishing between the different gestures are developed throughout the training stage. Finally, an algorithm based on the rules is used to forecast whether or not the testing gestures will be correctly identified. Distinct actions may be taught and differentiated throughout the training process. Finally, an algorithm based on the rules is used to forecast whether or not the testing picture will be identified. The accuracy rate comparing the existing method with the proposed framework is shown in Fig. 6.

A sequential or adaptive strategy may be used to automatically select the learning rate, which has a significant impact on the network's overall accuracy. The classification accuracy may be shown to alter significantly when a proper learning rate is used. However, even though the overall validation success rate and, consequently, the classification accuracy of the network are primarily determined by the learning rate, it is also necessary to determine the possibility of acquiring a lower validation loss than the previously selected value for training cycles by increasing the number of validation attempts. This is why it is important to train for a longer period of time to verify this feature. The success of a classification model can be measured in terms of its accuracy, which is calculated as the proportion of correct predictions to all predictions. It is the most common statistic used to test classifier models because it is easy to figure out and understand.

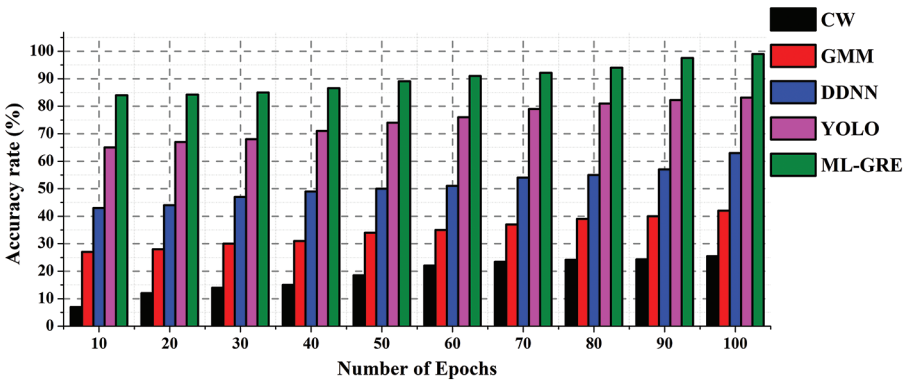

Figure 6: Accuracy rate (%)

4.2 Precision Rate

According to the results of this study, the suggested framework has high accuracy in the detection of human action target gestures and has the potential to significantly increase the precision of human action-based target gesture detection. The proposed algorithm takes advantage of this by collecting coordinate information from individual human skeletal joints using the Kinect interface device, and the variation in feature points between a static gesture and a motion gesture is calculated, followed by extraction of the node feature from the motion gesture, resulting in a reduction in the actual feature value. The departure from the target gesture in human motion enhances the detection precision of the target action. And the suggested algorithm is used to compare the recognition times of each joint point of the human motion target gesture of various algorithms, as well as the recognition time of the target gesture of the proposed method. So the proposed algorithm employs a supervised machine learning algorithm to construct the entire structure of an important network node, decreasing the placement of each network node by converting the entire placement of each main point into the local gesture detection area, thus further greatly reducing the detection time of every joint point in the hand gesture. Fig. 7 represents the precision rate (%).

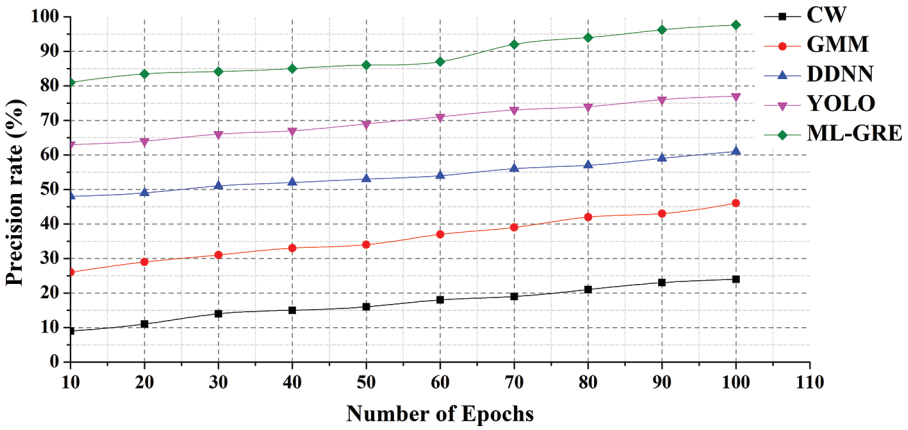

Figure 7: Precision rate (%)

4.3 Gesture Recognition Rate

There is evidence that the Machine Learning-based Gesture Recognition Framework suggested for gesture detection has greater average recognition accuracy than traditional classification methods. The ML-GRF system is designed to identify human actions such as dancing, strolling, washing, and other behaviors. When distinguishing minor actions, such as arm motions, the accuracy will be greatly reduced. In summary, our suggested approach is better at action recognition. Additionally, we compared the suggested mechanism to two additional baseline mechanisms to ensure its validity. A Machine Learning-based Gesture Recognition Framework result is based on the sensor with the highest matching probability, which is referred to as the maximum. The "average" signifies that the results are based on the average of the sensors, which means that the detected gesture type is one with the greatest average matching probability. The recognition rate (%) is depicted in Fig. 8.

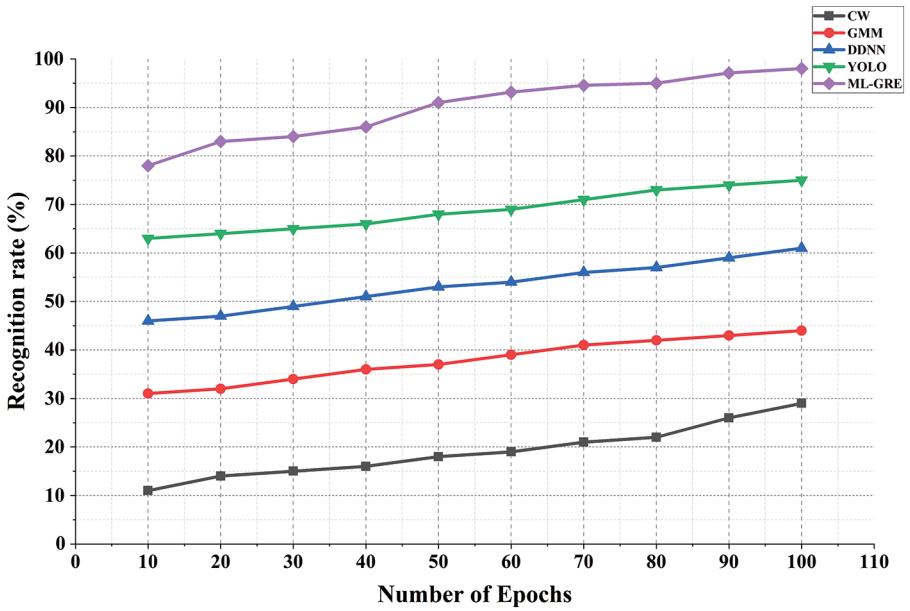

Figure 8: Recognition rate (%)

4.4 Sensitivity Rate

The ML-GRF may be affected by altering lighting and illumination conditions. Experiments are carried out to determine the effect of illumination fluctuation on the sensitivity of the hand motions identified using this method. For this reason, each individual was subjected to a variety of illumination settings, both natural and artificial. Different illumination settings were used for training and testing purposes. The sensitivity of gesture detection systems to backdrop variables has remained a problem. There are times when the hand is seen against a brighter backdrop and others when the hand is seen against a darker background, making it difficult to follow the hand precisely and recognize the action. Additionally, the difficulty is exacerbated since it may be impossible to tell the difference between the hand and its backdrop if their gray levels match. This is why the ML-GRF-based technique was tested against several backdrop circumstances, such as a uniformly dark background, a uniformly light background, and an ambiguous background. Fig. 9 represents the sensitivity rate (%).

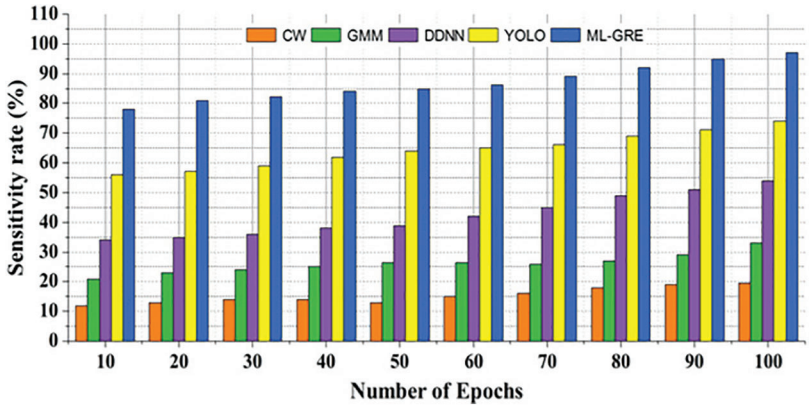

Figure 9: Sensitivity rate (%)

An experiment was conducted to investigate the effect of a change in camera and hand distance on the ability to reliably recognize hand actions. Experiments were done on three different scales using preset gestures. Variable illumination, size, and view angles or rotations are all factors that may affect the ML-GRF-based technique of recognition's ability to identify objects.

4.5 Efficiency Rate

Hand detection is the initial step in an uncontrolled environment for an end-to-end action recognition system. To a large extent, a system's total efficiency is determined by the accuracy of its projected gesture bounding boxes. Noise or information loss might occur if the bounding boxes are too large or too small, which could affect the succeeding stages of the system. Gesture recognition faces a number of key obstacles, including variable resolutions, positions, lighting, shadowing, and the accuracy-efficiency trade-off, which is the most important issue. Despite significant progress in uncontrolled gesture detection, efficient gesture detection with a low computing cost to scale the gesture recognition system while maintaining high efficiency remains an unsolved problem. Using training approaches that reduce the negative impacts of data label noise would boost efficiency and make the recognition more durable in more extreme environments with poor picture quality, variable gestures, and low light. The models may learn general gesture representations more efficiently by balancing datasets across racial backgrounds, as well as training techniques. The efficiency rate (%) is demonstrated in Fig. 10.

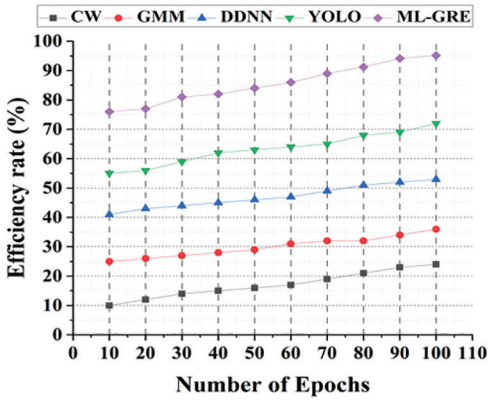

Figure 10: Efficiency rate (%)

5 Conclusion

This study combined various machine learning approaches to offer a novel system for identifying dynamic hand motions. For sophisticated sign language hand movements, a system that incorporates both localized hand structure characteristics and globalized body configuration characteristics has been suggested. For a long time, variations in illumination and dynamic backgrounds were disregarded as potential hinderances to recognition performance. In order to acquire empirical representations of body language, affective body language recognition should take cues from gesture recognition and agree on output spaces based on which to disclose vast amounts of marked and unmarked data. In this proposed approach, we separately address the two fundamental subproblems of gesture identification and verification in continuous data streams. The accelerated information is first evaluated using a basic moving average filter, and then a segmentation technique is used to automatically establish the starting and stopping positions of each input action. The common jerk is a useful trait for determining basic behaviors. After that, the identified fundamental gesture is encoded using the Johnson code. The distance between two adjacent gesture codes is calculated using the similarity metric, which translates to standardized similarity. Therefore, a gesture's gesture code can be determined by comparing it to one of the saved templates. Since motion analysis provides the typical gesture sequences that the recognition system detects, pre-training databases are not required. Using decomposition and similarity matching, this method gets a 98.97% accuracy rate, a 97.65% precision rate, a 98.04% gesture identification rate, a 96.99% sensitivity rate, and a 95.12% efficiency rate for gesture recognition.

Acknowledgement: The authors would like to acknowledge the Deanship of Scientific Research at Jouf University for their support.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] M. Rezende, S. G. M. Almeida and F. G. Guimaraes, "Development and validation of a Brazilian sign language database for human gesture recognition," *Neural Computing and Applications*, vol. 33, no. 16, pp. 10449–10467, 2021.

[2] M. Rinalduzzi, A. D. Angelis, F. Santoni, E. Buchicchio, A. Moschitta *et al.,* "Gesture recognition of sign language alphabet using a magnetic positioning system," *Applied Sciences*, vol. 11, no. 12, pp. 1–20, 2021.

[3] B. Li, J. Yang, Y. Yang, C. Li and Y. Zhang, "Sign language/gesture recognition based on cumulative distribution density features using UWB radar," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, no. 1, pp. 1–13, 2019.

[4] E. Jove, J. A. Mata, H. A. Moreton, J. L. Roca, D. Y. Marcos *et al.,* "Intelligent one-class classifiers for the development of an intrusion detection system: The MQTT case study," *Electronics*, vol. 11, no. 3, pp. 1–12, 2022.

[5] S. Nikolopoulos, I. Kalogeris and V. Papadopoulos, "Non-intrusive surrogate modeling for parametrized time-dependent partial differential equations using convolutional auto encoders," *Engineering Applications of Artificial Intelligence*, vol. 109, no. 1, pp. 1–21, 2022.

[6] S. Djosic, I. Stojanovic, M. Jovanovic, T. Nikolic and G. L. Djordjevic, "Fingerprinting-assisted UWB-based localization technique for complex indoor environments," *Expert Systems with Applications*, vol. 167, no. 1, pp. 1–14, 2021.

[7] E. G. Llano and A. M. Gonzalez, "Framework for biometric iris recognition in video, by deep learning and quality assessment of the iris-pupil region," *Journal of Ambient Intelligence and Humanized Computing*, vol. 1, no. 1, pp. 1–13, 2021.

[8] A. S. Dhanjal and W. Singh, "An automatic machine translation system for multi-lingual speech to Indian sign language," *Multimedia Tools and Applications*, vol. 81, no. 3, pp. 4283–4321, 2022.

[9] A. A. Barbhuiya, R. K. Karsh and R. Jain, "CNN based feature extraction and classification for sign language," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 3051–3069, 2021.

[10] A. Krestanova, M. Cerny and M. Augustynek, "Development and technical design of tangible user interfaces in wide-field areas of application," *Sensors*, vol. 21, no. 13, pp. 1–41, 2021.

[11] N. Pellas, S. Mystakidis and I. Kazanidis, "Immersive virtual reality in k-12 and higher education: A systematic review of the last decade scientific literature," *Virtual Reality*, vol. 25, no. 3, pp. 835–861, 2021.

[12] H. Park, Y. Lee and J. Ko, "Enabling real-time sign language translation on mobile platforms with on-board depth cameras," in *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, New York, USA, vol. 5, pp. 1–30, 2021. https://doi.org/10.1145/3463498.

[13] S. Tam, M. Boukadoum, A. C. Lecours and B. Gosselin, "Intuitive real-time control strategy for high-density myoelectric hand prosthesis using deep and transfer learning," *Scientific Reports*, vol. 11, no. 1, pp. 1–14, 2021.

[14] A. Nadeem, A. Jalal and K. Kim, "Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy markov model," *Multimedia Tools and Applications*, vol. 80, no. 14, pp. 21465–21498, 2021.

[15] N. Shah, N. Bhagat and M. Shah, "Crime forecasting: A machine learning and computer vision approach to crime prediction and prevention," *Visual Computing for Industry, Biomedicine, and Art*, vol. 4, no. 1, pp. 1–14, 2021.

[16] J. Leng, D. Wang, W. Shen, X. Li, Q. Liu *et al.,* "Digital twins-based smart manufacturing system design in industry 4.0: A review," *Journal of Manufacturing Systems*, vol. 60, no. 1, pp. 119–137, 2021.

[17] H. Kim, Y. -T. Kwon, H. -R. Lim, J. -H. Kim, Y. -S. Kim *et al.,* "Recent advances in wearable sensors and integrated functional devices for virtual and augmented reality applications," *Advanced Functional Materials*, vol. 31, no. 39, pp. 1–12, 2021.

[18] A. Madridano, A. A. Kaff, D. Martın and A. D. Escalera, "Trajectory planning for multi-robot systems: Methods and applications," *Expert Systems with Applications*, vol. 173, no. 1, pp. 1–14, 2021.

[19] H. Zhuang, Y. Xia, N. Wang and L. Dong, "High inclusiveness and accuracy motion blur real-time gesture recognition based on YOLOv4 model combined attention mechanism and DeblurGanv2," *Applied Sciences*, vol. 11, no. 21, pp. 1–19, 2021.

[20] V. Vakkuri, K. K. Kemell, M. Jantunen, E. Halme and P. Abrahamsson, "ECCOLA—A method for implementing ethically aligned AI systems," *Journal of Systems and Software*, vol. 182, no. 1, pp. 1–16, 2021.

[21] R. B. Burns, H. Lee, H. Seifi, R. Faulkner and K. J. Kuchenbecker, "Endowing a NAO robot with practical social-touch perception," *Frontiers in Robotics and AI*, vol. 86, no. 1, pp. 1–17, 2022.

[22] K. Bayoudh, F. Hamdaoui and A. Mtibaa, "Transfer learning based hybrid 2d3d CNN for traffic sign recognition and semantic road detection applied in advanced driver assistance systems," *Applied Intelligence*, vol. 51, no. 1, pp. 124–142, 2021.

[23] T. Kirishima, K. Sato and K. Chihara, "Real-time gesture recognition by learning and selective control of visual interest points," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 351–364, 2005.

[24] S. Poularakis and I. Katsavounidis, "Low-complexity hand gesture recognition system for continuous streams of digits and letters," *IEEE Transactions on Cybernetics*, vol. 46, no. 9, pp. 2094–2108, 2015.

[25] G. Plouffe and A. M. Cretu, "Static and dynamic hand gesture recognition in depth data using dynamic time warping," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 2, pp. 305–316, 2015.

[26] G. Chen, Z. Xu, Z. Li, H. Tang, S. Qu *et al.,* "A novel illumination-robust hand gesture recognition system with event-based neuromorphic vision sensor," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 2, pp. 508–520, 2021.

[27] A. A. Pramudita, L. Ukas and E. Dwar, "Contactless hand gesture sensor based on array of CW radar for human to machine interface," *IEEE Sensors Journal*, vol. 21, no. 13, pp. 15196–15208, 2021.

[28] F. Zhou, X. Li and Z. Wang, "Efficient high cross-user recognition rate ultrasonic hand gesture recognition system," *IEEE Sensors Journal*, vol. 20, no. 22, pp. 13501–13510, 2020.

[29] D. K. Vishwakarma, R. Maheshwari and R. Kapoor, "An efficient approach for the recognition of hand gestures from very low resolution images," in *Proc. CSNT*, Gwalior, India, pp. 467–471, 2015.

[30] T. R. Gadekallu, M. Alazab, R. Kaluri, P. K. R. Maddikunta, S. Bhattacharya *et al.,* "Hand gesture classification using a novel CNN-crow search algorithm," *Complex & Intelligent Systems*, vol. 7, no. 4, pp. 1855–1868, 2021.

[31] P. Reddy, S. Koppu, K. Lakshmanna, P. Venugopal and R. Poluru, "Automated category text identification using machine learning," in *Proc. 2020 Int. Conf. on Emerging Trends in Information Technology and Engineering (ICETITE)*, Vellore, India, pp. 1–4, 2020.

[32] M. Abavisani, H. R. V. Joze and V. M. Patel, "Improving the performance of Unimodal dynamic hand-gesture recognition with Multimodal training," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, pp. 1165–1174, 2019.

[33] S. Ahmed, F. Khan, A. Ghaffar, F. Hussain and S. H. Cho, "Finger-counting-based gesture recognition within cars using impulse radar with convolutional neural network," *Sensors*, vol. 19, no. 6, pp. 1429–1442, 2019.

[34] L. Chen, J. Fu, Y. Wu, H. Li and B. Zheng, "Hand gesture recognition using compact CNN via surface electromyography signals," *Sensors*, vol. 20, no. 3, pp. 672–686, 2020.

[35] M. Oudah, A. Al-Naji and J. Chahl, "Hand gesture recognition based on computer vision: A review of techniques," *Journal of Imaging*, vol. 6, no. 8, pp. 73–101, 2020.