

Article

Wearable-Sensors-Based Platform for Gesture Recognition of Autism Spectrum Disorder Children Using Machine Learning Algorithms

Uzma Abid Siddiqui ¹, Farman Ullah ^{1,*}, Asif Iqbal ², Ajmal Khan ¹, Rehmat Ullah ³, Sheraz Paracha ¹, Hassan Shahzad ¹ and Kyung-Sup Kwak ^{2,*}

- ¹ Electrical and Computer Engineering Department, COMSATS University Islamabad Attock Campus, Attock 43600, Pakistan; uzmaabidsiddiqui@gmail.com (U.A.S.); drajmal@ciit-attock.edu.pk (A.K.); sheryparacha666@gmail.com (S.P.); hshahzad837@gmail.com (H.S.)
- ² Department of Information and Communication Engineering, Inha University, Incheon 22212, Korea; asifsoul@inha.ac.kr
- ³ Department of Computer Systems Engineering, University of Engineering and Technology, Peshawar 25000, Pakistan; rehmatkttk@nwfpuest.edu.pk
- * Correspondence: farmankttk@ciit-attock.edu.pk (F.U.); kskwak@inha.ac.kr (K.-S.K.)



Citation: Siddiqui, U.A.; Ullah, F.; Iqbal, A.; Khan, A.; Ullah, R.; Paracha, S.; Shahzad, H.; Kwak, K.-S. Wearable-Sensors-Based Platform for Gesture Recognition of Autism Spectrum Disorder Children Using Machine Learning Algorithms. *Sensors* **2021**, *21*, 3319. <https://doi.org/10.3390/s21103319>

Academic Editors: Emad Shihab and Omid Sarbishei

Received: 28 February 2021

Accepted: 7 May 2021

Published: 11 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Autistic people face many challenges in various aspects of daily life such as social skills, repetitive behaviors, speech, and verbal communication. They feel hesitant to talk with others. The signs of autism vary from one individual to another, with a range from mild to severe. Autistic children use fewer communicative gestures compared with typically developing children (TD). With time, the parents may learn their gestures and understand what is occurring in their child's mind. However, it is difficult for other people to understand their gestures. In this paper, we propose a wearable-sensors-based platform to recognize autistic gestures using various classification techniques. The proposed system defines, monitors, and classifies the gestures of the individuals. We propose using wearable sensors that transmit their data using a Bluetooth interface to a data acquisition and classification server. A dataset of 24 gestures is created by 10 autistic children performing each gesture about 10 times. Time- and frequency-domain features are extracted from the sensors' data, which are classified using k-nearest neighbor (KNN), decision tree, neural network, and random forest models. The main objective of this work is to develop a wearable-sensor-based IoT platform for gesture recognition in children with autism spectrum disorder (ASD). We achieve an accuracy of about 91% with most of the classifiers using dataset cross-validation and leave-one-person-out cross-validation.

Keywords: wearable sensors; autism spectrum disorder (ASD); stereotype movements; gestures; machine learning; KNN; decision tree; random forest; neural network

1. Introduction

Autism spectrum disorder, commonly called autism, is defined as a variety of disorders, which include challenges with social rules, difficulty in verbal and non-verbal communication, and restricted or repetitive actions [1]. Each child with ASD has their own specific needs and a collection of habits and behaviors that can hinder their day-to-day tasks. As ASD is a heterogeneous neurodevelopmental disorder, its symptoms appear during the early ages, normally in the first two or three years of life [2]. It is an intricate neurobehavioral condition that makes social interactions problematic for such individuals. Not all the disorders in this spectrum are equally severe; some appear to be a minor handicap, while others are serious and can practically disrupt the whole lifestyle of the affected person. Children with ASD struggle to communicate with others. Reading visual emotions is tricky for them and they usually struggle to understand what other people feel and think.

Autistic people face many challenges in their daily lives in areas such as social skills, repetitive behaviors, speech, and nonverbal communication, and experience feelings of hesitation. They use fewer communicative gestures compared with typically developing children (TD), so they struggle to convey their ideas or thoughts with words, gestures, or facial expressions. Usually, people with autism develop strange behaviors and, in some cases, they can be dangerous to themselves and to the people around them [3]. Due to the impairment in their speaking ability, their children's physical abilities may be weakened, which can lead to risk [3]. These disorders are more common in boys than girls, with a ratio of about 4:1 [4]. To effectively communicate with people from an early age and initiate social interactions, the child must be able to understand verbal and non-verbal messages [5]. Gestures are the universal and most important feature of our communication [6]. Verbal gestures involve the use of our voice and co-verbal gestures involve hand and arm movements. They simplify linguistic content, emphasize our point of view, regulate the flow of speech, and maintain the audience's attention toward the speaker. Although there is no rigid rule about gesture categories, conventional gestures (CG) have well-established premises [7]. These are conversational and deliberate actions that enable direct and accurate verbal translation so that they are easily understood even without spoken help. Because conventional gestures contribute to communication and are a pre-cursor to verbal language, their recognition in children presenting neurodevelopmental disorders is crucial. Over the course of time, their parents can grow accustomed to their gestures and understand what is occurring in their child's mind, but others may find these gestures difficult to interpret.

Due to nonverbal communication or repetitive speaking, ASD children have difficulties conveying their message and other people struggle to understand their gestures. Sign-language, or hand-speak, has become a popular method of communicating for those who cannot verbally speak. However, sign language is difficult to understand for normal people. A gesture translator is needed to communicate with ASD people. The Internet of Things (IoT) is providing new and emerging capabilities through continuous connectivity. This concept generally revolves around the rapidly expanding environments of computing capabilities and network connectivity to sensors, objects, devices, and items that efficiently exchange digital data with each other without external assistance, i.e., human involvement. The main objective of the IoT is to design, operate, deploy, and transform physical devices through data acquisition, predictive analytics, smart networking, deep optimization, and some other related solutions [8]. Wearable devices provide opportunities for innovative services in health sciences along with predictive health monitoring by persistently acquiring the data of the wearer [9]. Wearable sensors provide reliable and accurate information about human gestures and behaviors to ensure a safe and secure living environment [10]. Gesture recognition is required for the development of various operations such as feedback from acquired data, tracking physical fitness, health monitoring, and self-control/management of a wearable device [11].

Our proposed platform can effectively recognize gestures. Gesture recognition is divided into different steps, the first of which is the collection of data by a body-worn sensor. A tri-axial accelerometer and gyroscope are used with a sampling frequency of 50 Hz. Then, for the removal of noise and unreliable data, pre-processing is performed. The data are then used to extract various time- and frequency-domain features such as entropy, standard deviation, mean, and root mean square values. The features data with labels are then used for classification.

The following were the main objectives in this study:

- Constructing a wearable-sensors-based platform to acquire and recognize ASD children's gestures.
- Extracting various features from the gestures data and comparing performance to select features for efficient recognition.
- Comparing performance using various machine learning algorithms to increase recognition accuracy.

In the literature related to ASD gesture recognition, the Flex sensor and switch sensors have mostly been used. The limitation of these sensors is that they only have on and off switch status. So, the limited nature of gestures that do not involve much variation in movement is not discussed in the literature. In order to cope with these challenges, the novel contributions in this paper are as follows:

- Since ASD is a special body condition, both medically and physically, we did not use the data of normal people to train the supervised machine learning algorithm for the gestures recognition of ASD. Instead, we collected a novel dataset of 24 physical activities from 10 children who had mild and moderate levels of ASD.
- We performed features extraction on the acquired data using various statistical measures from both the time and frequency domains.
- For gestures recognition on the novel dataset, we evaluated several classifiers and selected the one that produced the most accurate cumulative result.
- We conducted Raspberry-PI-based real-time gesture monitoring of ASD to facilitate the communication between ASD and normal people.

The rest of the paper is organized as follows: Section 2 briefly introduces the related literature work. The methodology is explained in Section 3, and the results and discussion in Section 4. Finally, we conclude the paper in Section 5.

2. Background and Related Work

In this section, we briefly focus on the background and the related work conducted regarding the platform used for autism activity and gesture recognition, placement of sensors on the body, extracting features from the acquired sensors data, and the performance comparison analysis of classifiers.

In the literature, some studies focused on the ASD subject, their types of gesture used, and how they behave while communicating with others. Autistic children face difficulties in conveying their thoughts to others. They use fewer communicative gestures compared with typically developing (TD) children [7]. Over time, their parents may grow accustomed to their gestures and understand what their child's is thinking, but it often remains hard for other people to understand their gestures. Their sign language is normally different from those who are deaf and mute, making it even more difficult to grasp their thoughts. Almost all TD children undergo the same procedure of gestures development [12]. Autistic children may either have an unusual behavioral pattern or have a stereotypic behavioral display [13]. Stereotypic behaviors are abnormal gestures that humans make without having an obvious function or purpose [14]. This kind of behavioral pattern elevates the activation level of children with autism. An autistic person may show some of the classic stereotypic behaviors, e.g., hand flapping, head-banging, body rocking, and top spinning [15].

Human gestures can be detected by ambient-environmental sensors or body-worn sensors. Ambient sensors are installed in offices or homes and are stable in nature, so they cannot perform outside the installed area. These kinds of sensor systems are stationary; hence, they are bound in their specific areas. These systems are not known to have the best efficiency as their observation is strictly bound to a limited area. These types of sensors are highly application-specific. Due to this issue, wearable or mobile monitoring sensors are usually preferred to acquire data continuously and effectively. Mobile monitoring sensors can be used to acquire data remotely and accurately. Despite these many benefits, mobile-sensors-based acquisition has some drawbacks as well [16]. One of them is that most of the time, the smartphone is placed in some pocket position, which reduces the efficiency of recognition of certain activities. Wearable sensors are used to overcome the localization problem. Modern smartphones and smartwatches are equipped with sensors. Gyroscope, accelerometer, magnetometer, temperature, and sound sensors have been used for activity detection [17]. Microphones and web cameras have been used for gestures and expression recognition [18]. A bone conducting speaker, a heads-up display, and wearable glasses were used for recognizing facial gestures [19]. A multisensor accelerometer was

used for the detection of stereotypical motor movements (SMMs), which include complex hand movements, body rocking, and mouthing [20]. The Kinect and Flex sensors with a camera have been used to recognize head and hand movements [21–25]. Force sensitive resistor sensors (FSRs) were used to identify multiple gait cycles during walking [26]. Wi-Fi and GPS systems were used for movement recognition [27]. Leap motion device was used to record gesture movements of users for American sign language in virtual reality by generating a 3D hand model [28]. An electromyography sensor (EMG) was used to acquire hand gestures from 15 people. The hand movements included open and closed hand, neutral, victory sign, wrist flexion, tap, and wrist extension. [29]. The Myo Arm band was used to collect gestures data for Hand Cricket between two persons [30].

Sensor placement on human body considers the relative position of the body and the orientation of the sensor. Studies showed that the location of a sensor on a part of the body depends on the activities to be recognized. Accuracy is proportional to the number of sensors used: more sensors lead to higher accuracy and less sensors result in lower accuracy. Researchers have placed sensors on various parts of the body to recognize different activities, such as the wrist, ankle, and chest [31–33].

For gesture recognition, various features are extracted from data sensed by sensors. Feature extraction is an important and difficult step in activity recognition. The features that have been extracted by different researchers from the acquired sensors data related to different activities are as follows: the duration of activity, mean, variance, standard deviation, median absolute deviation, zero-crossing, cross-correlation, autocorrelation, maximum peaks, total peaks, average of all peaks, area of signal magnitude, energy of signal, frequency component with prime magnitude, mean of the frequency components, power spectral entropy, root mean square, fast Fourier transform, etc. The most commonly adopted classifiers include the k-nearest neighbor (KNN), random forest (RF), multilayer perceptron (MLP), support vector machine (SVM), decision tree, and artificial neural network (ANN) etc. Estrada used KNN and decision trees to recognize static gestures, and the dynamic time warping (DTW) algorithm for the recognition of dynamic gestures [24]. Sombandith et al. [25] used the histogram of oriented gradients and correlation coefficients to recognize the hand gestures of the Lao alphabet sign language. Gonçalves proposed the detection of stereotype movements of autistic people using the dynamic time warping algorithm [22]. Rad used a convolutional neural network for the detection of SMM using a accelerometer sensor [20]. Comprehensive details and a comparison of the proposed technique with techniques reported in the literature are tabulated in Table 1. We investigated gestures or activity recognition based on the type of activities performed, the data set, and the machine learning algorithms applied.

Table 1. A summary of the literature and related works about gestures and activities recognition of normal and autistic people.

Ref. No	Sensors	Activities	Features	Algorithms and Accuracy
[2]	Moto 360 smartwatch	Flapping, painting, and sibbing	Discrete cosine transform, FFT, variance, bi-spectrum, z transform, entropy	Simple tree, complex tree, linear and gaussian SVM, boosted and bagged ensemble trees Accuracy: 96.7%
[34]	ECG, accelerometer, gyroscope, magnetometer	Walking, climbing stairs, frontal elevation of arms, knees bending, cycling, jogging, running, jump front and back, sitting, relaxing	Mean, standard deviation, and correlation	Mean prediction rate 99.69%, HMM 89.98%, DBN 92.01%, RNN 99.69%
[35]	Not mentioned	9 uniform hand gestures	Not mentioned, total 576 features extracted	SVM 98.72%

Table 1. Cont.

Ref. No	Sensors	Activities	Features	Algorithms and Accuracy
[36]	Gyroscope, accelerometer	Hand movements, body movements	Publicly available dataset features	Convolutional neural network 87.1%, KNN 66.1%, SVM 77.1%, fully CN 88%
[37]	Not mentioned	Static and dynamic unistroke hand gestures	Not mentioned	SVM 97.95%
[38]	Accelerometer, magnetometer, gyroscope	Jogging, walking, cycling jumping, running, jump-rope	Mean, standard dev, kurtosis, skewness, range, correlation, spectral energy, spectral entropy, peak frequencies, and cross-spectral densities	SVM 26%, DT 93.24%, KNN 96.07%, RF 97.12%, Naïve Bayes 76.47%
[39]	Accelerometer, strain sensor	Walking, eating	Mean value, standard dev, percentiles, and correlation frequency domain (energy, entropy)	DT 93.15%
[40]	Camera	Gestures of alphabets	Not mentioned	KNN 94.49%
[41]	Flex sensor, accelerometer, camera,	Malaysian sign language gestures	Not mentioned	General algorithm for the data-glove detection system 78.33–5%
[42]	Camera	24 Fingerspelling static gestures	Not mentioned	KNN classifier 87.38%, Logistic regression 84.32%, naïve Bayes classifier 84.62%, support vector machine (SVM) 91.35%
[43]	Leap Motion Sensor	Gestures for greetings, possessive adjectives, colors, numbers, names, etc.	Not mentioned	Hidden Markov models (HMM) 87.4%, KNN+DTW 88.4%
[44]	Accelerometer	Cycling, sedentary, ambulation	Mean, standard deviation, acceleration range	SVM from 88.5% to 91.6%
[45]	Not mentioned	ASL alphabets and basic hand shapes	The number of fingers, the width and height of the gesture, the distance between the hand fingers, etc.	Type-2 Fuzzy HMM (T2FHMM) 100% accuracy for uniform hand images and 95.5% for cluttered hand images
[24]	Flex sensor	Patterns representing: • Letters/Words • Numbers	Not mentioned	K-nearest neighbor decision tree dynamic time warping average accuracy = 90%
[46]	QA screening method using mobile app	Not mentioned	Age, sex, ethnicity, country of residence, etc.	RIPPER 80.95%, C4.5 82.54%
[47]	Not mentioned Dataset taken from UCL Machine Learning repository	Common attributes like age, nationality, sex, etc.	Not mentioned	SVM 98.30%, KNN 88.13%, CNN 98.30% ANN 98.30%, naïve Bayes 94.91%, LR 98.30%

3. Proposed Wearable-Sensors-Based Platform for Gesture Recognition of Autism Spectrum Disorder Children

In this section, we describe the proposed platform, the acquisition of data from sensors, the construction of feature vectors, and the classifiers used for gesture recognition. Figure 1 shows the complete architecture of the proposed framework, which consists of two main parts.

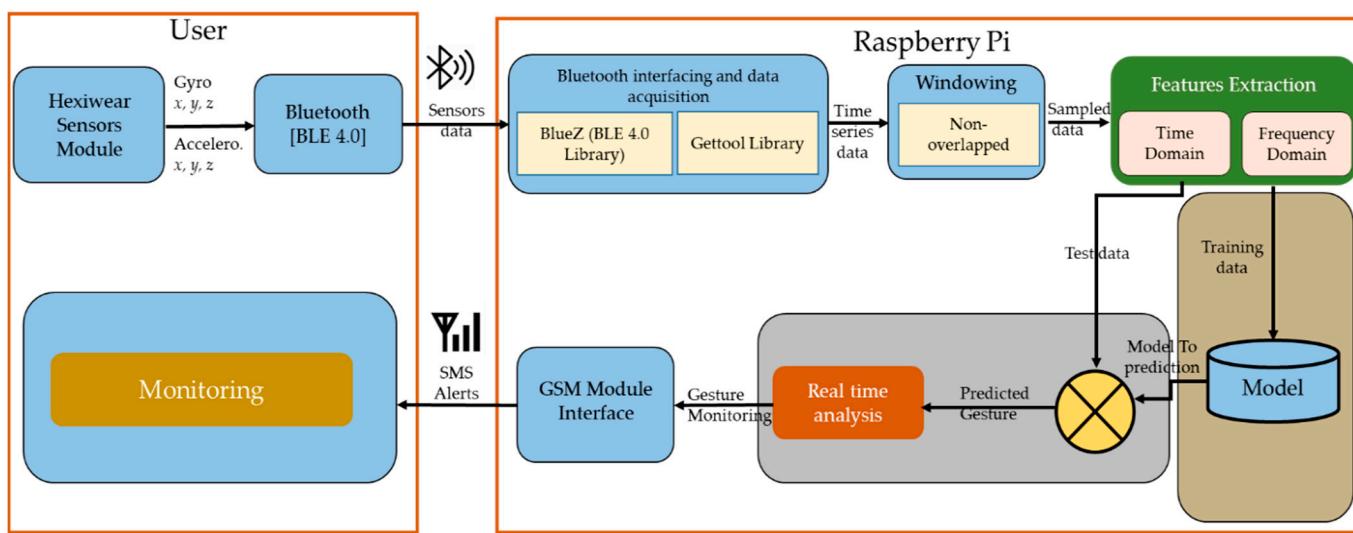


Figure 1. The proposed architecture for wearable-sensors-based platform for the gesture recognition of autism spectrum disorder children using machine learning algorithms.

The first part is based on the acquisitioning of data from sensors and its pre-processing, and the other part consists of the recognition of the processed data using different machine learning algorithms. A single Hexiwear sensor module consisting of both an accelerometer and a gyroscope was installed at the writing position of either the right or left hand. Sensor placement on a child mainly includes the relative position of the activity being performed and the orientation of the sensor. The proposed methodology is explained in the following subsections.

3.1. Data Collection

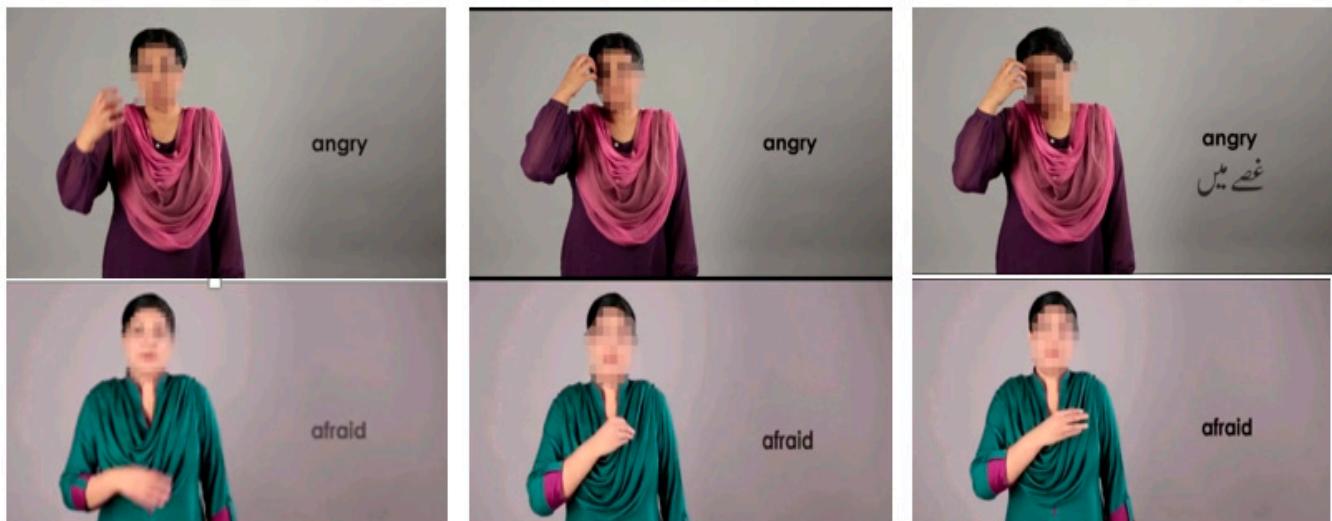
In the literature related to ASD gesture recognition, the Flex sensor or switch sensors are most often used. These sensors only have on and off switch status. So, few gestures that do not have movement variations are mentioned in the literature. In this study, we collected data from the sensors installed on the wrist of the child. The sensor was worn by the user and data were measured through a specific Bluetooth range of approximately 100 m. The data through sensors were collected at different sampling rates based on sensor type. We configured the sensor by fixing the accelerometer and gyroscope sampling frequency or sampling rate to 50 Hz to precisely capture the hand movement of the gesture. Figure 2 shows a sequence of pictures for two gestures. The gesture signal was digitized and the acquired data points comprise a timestamp and three axes values for the accelerometer and gyroscope. Table 2 shows the sensors configuration that was used for collecting data. The data set comprised different records gathered from the 10 subjects using two sensors. Each subject produced gestures for 24 activities and data were gathered for 3 seconds; this process was repeated 7–12 times. Table 3 shows the complete set of activities for which gestures of ASD children were recorded.

Table 2. Sensors configuration for the collection of data from autistic children.

Sensors	Sampling Frequency (Hz)	Quantization Levels (Bits)	Range
Accelerometer	50	16	$\pm 16 \text{ gs}$
Gyroscope	50	16	$\pm 2000^\circ/\text{s}$

Table 3. Information about the gestures which are recorded for data collection.

Gesture	Label	Gesture	Labels
Good Morning	G1	Angry	G13
Good Afternoon	G2	Bulb	G14
Good Night	G3	Cricket	G15
Good Bye	G4	Fan off	G16
Thank you	G5	Fan on	G17
Please	G6	Switch	G18
Yes	G7	Milk	G19
No	G8	Need eraser	G20
Wow	G9	Need pencil	G21
Hello	G10	Need toilet	G22
Sleep	G11	Need water	G23
Afraid	G12	School book	G24

**Figure 2.** Pictorial overview of the sequence of images for showing the hand movements when performing the gestures.

3.2. Features Extraction and Selection

The collected data were limited to a window size of three seconds, a set of features was extracted from the window, and a specific label was given to this features set, which was then used for learning purposes to construct a trained model. We extracted several features in both the time and frequency domains, inspired by the literature and [48]. The details and formulation of some of the features are given below, and the overall features vector processing is shown in Figure 3.

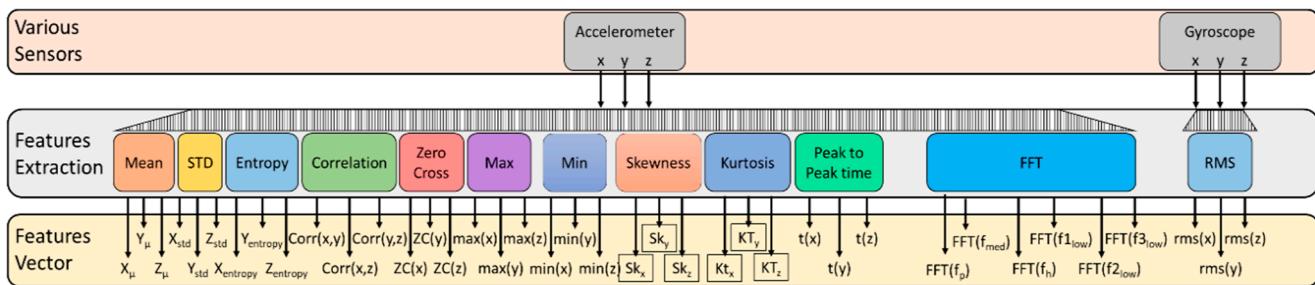


Figure 3. Features vector processing to convert the time-series sensors data into statistical measures in terms of the time- and frequency-domain features.

- **Mean:** We found the mean value for the accelerometer (x, y, z), and magnetometer (x, y, z).

$$\mu = \frac{1}{N} \sum_{i=0}^{N-1} x_i \quad (1)$$

- **Standard Deviation:** We calculated the spread in the sensors data around the mean as,

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (x_i - \mu)^2} \quad (2)$$

- **Entropy:** Entropy was used to differentiate between the gestures of a static nature, i.e., low movement, and activities having higher variation.

$$\text{Entropy} = -\frac{1}{N} \sum_{i=0}^{N-1} p_i \log p_i \quad (3)$$

- **Cross-correlation** was used to help differentiate between activities with variation.

$$\text{Corr} = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y} \quad (4)$$

$$\text{where } \text{Cov}(x, y) = \frac{\sum_{i=0}^{N-1} (x_i - \mu_x)(y_i - \mu_y)}{N-1}$$

- **Zero-crossing (ZC):** Zero-crossing is the number of times the signal crosses zero and its sign changes. We considered ZC for the accelerometer along three axes. Mathematically, it can be written as:

$$ZC = \text{COUNT}(\{(x_i > 0) \text{ AND } (x_{i+1} < 0)\} \text{ OR } \{(x_i < 0) \text{ AND } (x_{i+1} > 0)\}), 0 \leq i \leq N-1 \quad (5)$$

- **Maximum Value:** We calculated the maximum value of the accelerometer (x, y, z).

$$\text{Acc}_{\max} = \max(x_{\phi}), 0 \leq i \leq N-1 \quad (6)$$

- **Skewness:** The coefficient of skewness is a measure of the degree of symmetry in the variable distribution. It was calculated for every axis of the accelerometer.

$$\text{Skewness} = \frac{\sum_{i=0}^{N-1} (x_i - \mu_x)^3}{(N-1)\sigma_x^3} \quad (7)$$

- **Kurtosis:** The coefficient of kurtosis is a measure of the degree of tail in the variable distribution.

$$\text{Kurtosis} = \frac{\sum_{i=0}^{N-1} (x_i - \mu_x)^4}{(N-1)\sigma_x^4} \quad (8)$$

- **Fast Fourier Transform:** Equation (9) was used to find the fast Fourier transform (FFT) of acceleration data. We considered six frequency domain features based on the FFT of the acceleration data. The six features were the FFT magnitude: peak_f, low_f1, low_f2, low_f3, med_f, and high_f.

$$H(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi(k\frac{n}{N})} \quad (9)$$

3.3. Classification Algorithms for the Proposed Work

The task of this recognition system is the labelling of the recorded gesture from G1 to G24. For this, we used different supervised machine learning algorithms commonly known as classifiers. The process consisted of two parts. In the first phase, the classification algorithm found the relationship between the features and their corresponding labels using the training data to generate a model. Then, in the second part, the model was tested by providing new input features that are unknown to the model and then the model-output labels were compared with the actual labels to determine the classification accuracy of the algorithm. In this study, we used four classifiers, KNN, DT, RF, and the back-propagation model of a neural network. The details of the algorithms are explained in the following subsections.

3.3.1. The K-Nearest Neighbor Algorithm

The KNN algorithm is known as a lazy method of learning, which means that learning (finding the relationship between input features and their labels) does not start until a testing input is used. The algorithm only finds the k labels from the training data that are similar to the testing input [49]. These k samples and their corresponding labels are then used to predict the label for the new testing input. The closeness was found in terms of Euclidean and Manhattan distances between the new sample and every sample present in the training set in this paper. Equations (10) and (11) were used to find these closeness distances, respectively.

$$D(x, y) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \quad (10)$$

$$D(x, y) = \sum_{i=0}^n |x_i - y_i| \quad (11)$$

3.3.2. The Decision Tree Algorithm

Decision tree (DT) is a supervised Learning algorithm mostly used to solve classification problems [50]. The main idea is to create a tree for all the data, and process a single outcome at every leaf node or minimize the error at every leaf node. In this structure, internal nodes represent the features of a dataset, branches represent the decision criteria, and leaf nodes represents the outcome. The algorithm uses entropy (E) and the Gini Index (G)-based information gain (I) to select the root node and leaf node. If a number of classes are represented by C, an attribute by A, and V represents the possible values in the attribute A, then the following equations can be used to find the E, G, and I of entropy, respectively.

$$E(C) = - \sum_{i=1}^C p_i \log_2 p_i \quad (12)$$

$$G(C) = 1 - \sum_{i=1}^C p_i^2 \quad (13)$$

$$I(C, A) = E(C) - \sum_{V \in \text{values}(A)} \frac{|C_v|}{|C|} E(A) \quad (14)$$

The algorithm maximizes the information gain value, and the node having the highest gain splits first. The algorithm we used is shown in Figure 4.

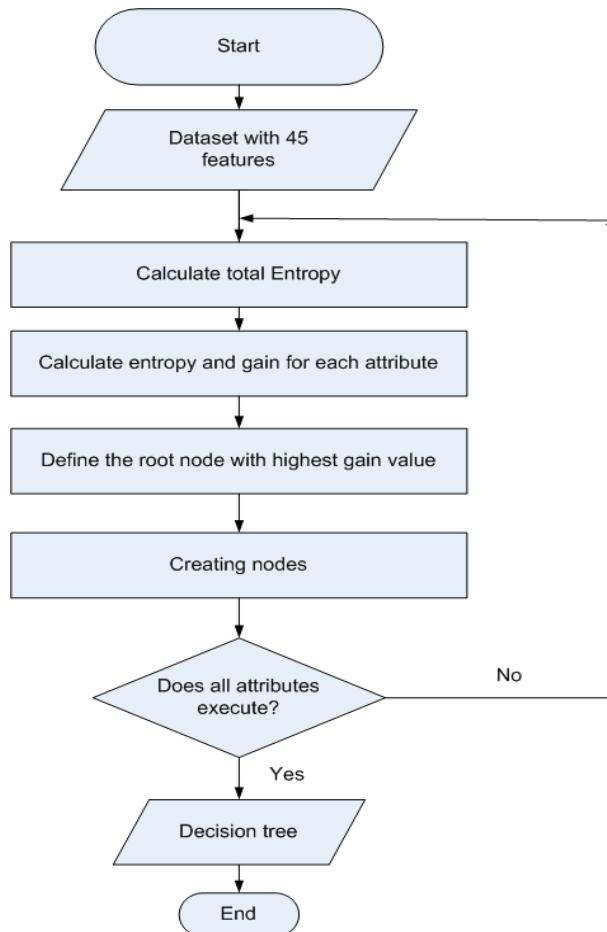


Figure 4. The decision tree algorithm used for the classification of the ASD children's gestures.

3.3.3. The Random Forest Algorithm

Random forest [51] is type of classification that works by building multiple decision trees (weak learners) and finally identifying the decision made by the majority of weak learners. Normally, pruning of the decision trees is used to avoid over-fitting. Pruning is basically a trade-off between accuracy and complexity. No pruning results in high complexity, larger time consumption, and higher resource utilization. Random forest has the same parameters as a decision tree classifier. However, it grows each tree on an independent bootstrap sample from the training data. At each node, a subset of variables is randomly selected from all possible variables (independently for each node) and the best split is found on the selected subset variables. After the forest is formed, the trees are voted or averaged to obtain predictions.

3.3.4. Back-Propagation-Based Neural Networks Algorithm

The back-propagation model [52] is the core of the neural network training process. It is a method of fine-tuning the weights of a neural net based on the error rate obtained in the previous iteration. Tuning the weights properly ensures the model is reliable by increasing its generalization and reduces the error rates. The feature vector acts as the input to the neural network. Different activation functions are compared to generate the

output and then the error is calculated for the back-propagation procedure. The following equations show how the process starts, error is calculated, and the backward propagation with corresponding weight adjustments. Equation (15) calculates the forward value of the input to the output.

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{145} \\ w_{21} & w_{22} & \dots & w_{245} \\ \vdots & \vdots & \vdots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{n45} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{45} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \quad (15)$$

Error calculation:

$$\delta = \phi'(v)e \quad (16)$$

We used the sigmoid function as our activation function, which is given by Equation (17).

$$\phi(v) = \frac{1}{1 + e^{-v}} \quad (17)$$

$$\phi'(v) = \phi(v)(1 - \phi(v)) \quad (18)$$

Error propagation:

$$e^{(k)} = W^T \delta \quad (19)$$

$$\delta^{(K)} = \phi'(v^{(k)})e^{(k)} \quad (20)$$

Weight adjustment:

$$\Delta w_{ij} = \alpha \delta_i x_j \quad (21)$$

$$w_{ij}^{new} = w_{ij}^{old} \quad (22)$$

where x is input, e is the error, v is the product of weights and corresponding inputs, W is the weight matrix, b is the bias vector of the node, $\phi(v)$ represents the activation function, $\phi'(v)$ shows its derivative, α is the learning rate, and i and j are the output and input node numbers, respectively.

4. Simulation Results and Discussion

In this section, we briefly introduce the sensors' response in time-series, describe the data set, and compare the performance comparison of the machine learning algorithms using cross-validation, i.e., dividing the whole dataset into percentage of training and testing, and leave-one-person-out cross-validation (LOOCV). Nine ASD children were used for training and one for testing.

4.1. Sensors Response and Dataset Description

Figure 5 shows the time-series response of the accelerometer and gyroscope for three activities G12 (Afraid), G13 (Angry), and G17 (FAN ON). The sampling period was 50 Hz and the graph shows each gesture was performed six times by the ASD child. Figure 5 depicts that each gesture was performed in 3 s including the variation in hand gesture movement. So, a window of 3 s was used to extract the features for each gesture from sensors data and label it. Table 4 describes the complete data set of the ASD children who performed each gesture 7–12 times so the records for each gesture varied from 83–103 records.

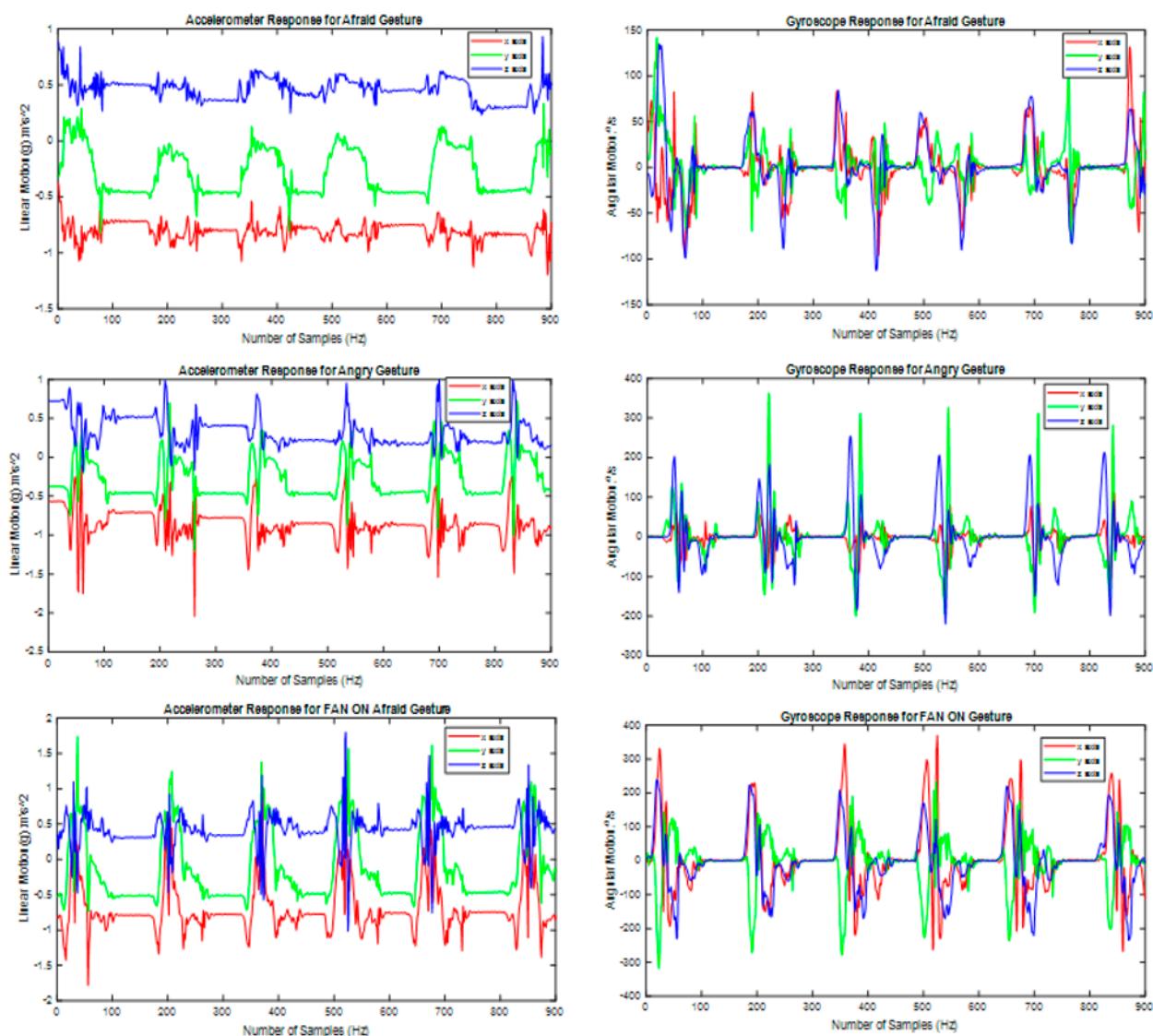


Figure 5. Sensors response for gestures performed by the ASD children (performed six times).

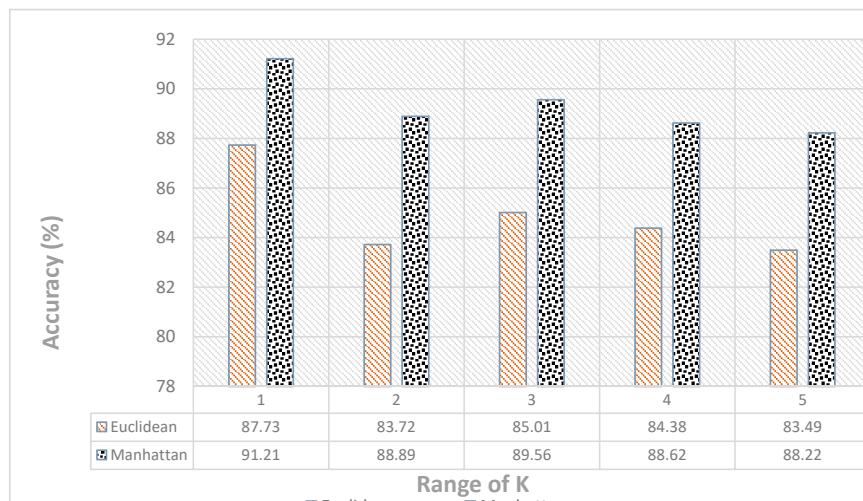
Table 4. ASD children data set description.

Gestures Label	No. Records	Gestures Label	No. Records
G1	99	G13	100
G2	85	G14	99
G3	100	G15	90
G4	90	G16	89
G5	86	G17	97
G6	90	G18	99
G7	100	G19	99
G8	98	G20	90
G9	93	G21	89
G10	80	G22	97
G11	100	G23	90
G12	103	G24	78

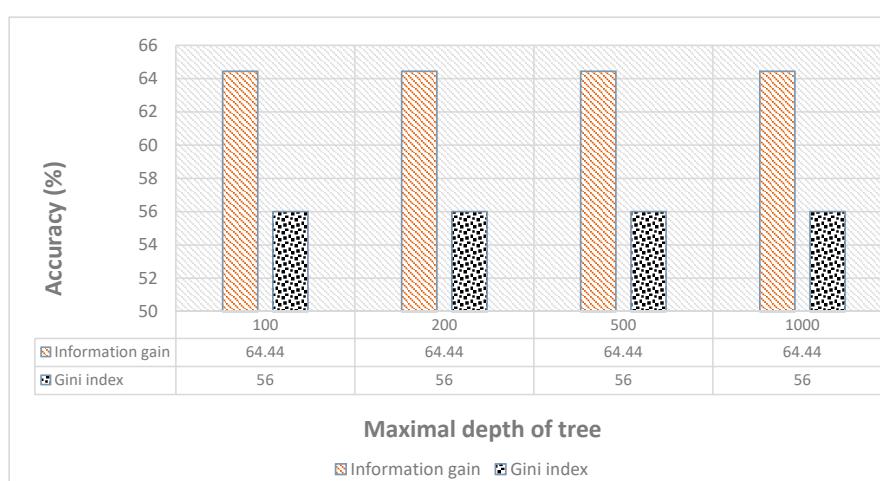
4.2. Individual Classifier Performance Comparison Using Data Cross-Validation

For different classifiers, we used different evaluating functions. For KNN, we used 10 folds for validation and comparison was performed on the basis of distances, i.e., Euclidean and Manhattan. For DT and RF, we calculated both the information gain and Gini index for the different number of trees and variable depths of trees. For the neural network, we compared the results on both the single layer and double layers with a variable number of neurons and learning rates.

Figure 6 shows the complete individual comparison of all the classifiers using different parameters. Figure 6a compares the performance of the KNN classifier, which clearly shows that the Manhattan distance performed better compared with the Euclidean distance. We used 10-fold cross-validation, i.e., 90% of the data set was used for training and 10% for testing. The accuracy indicates the average of the 10-fold cross-validation. Figure 6 compares the performance of decision tree with varying depths of the decision tree; the data were split into 90% training and 10% testing. The accuracies were found by changing number of trees and the tree depth to evaluate the accuracy. The results are summarized in Figure 6c for random forest. The classifiers performed with accuracy of about 91%.

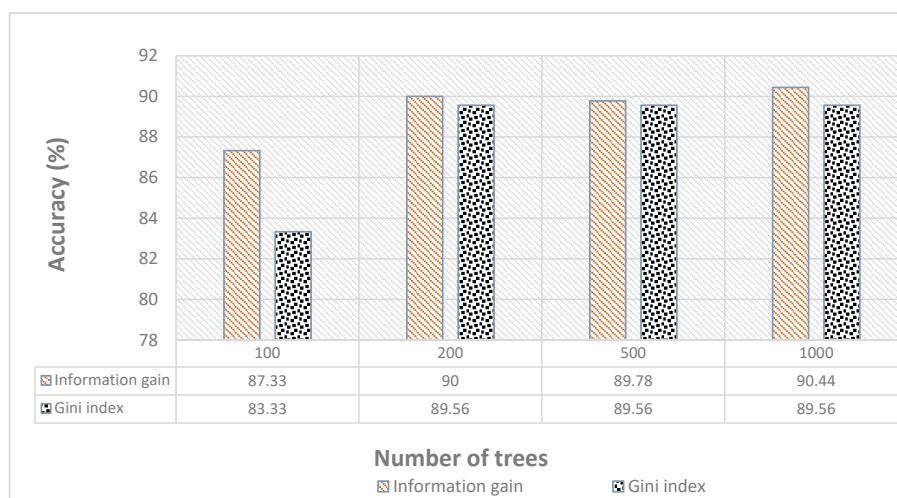


(a) Performance Comparison of k-nearest neighbor using 10-fold cross-validation.

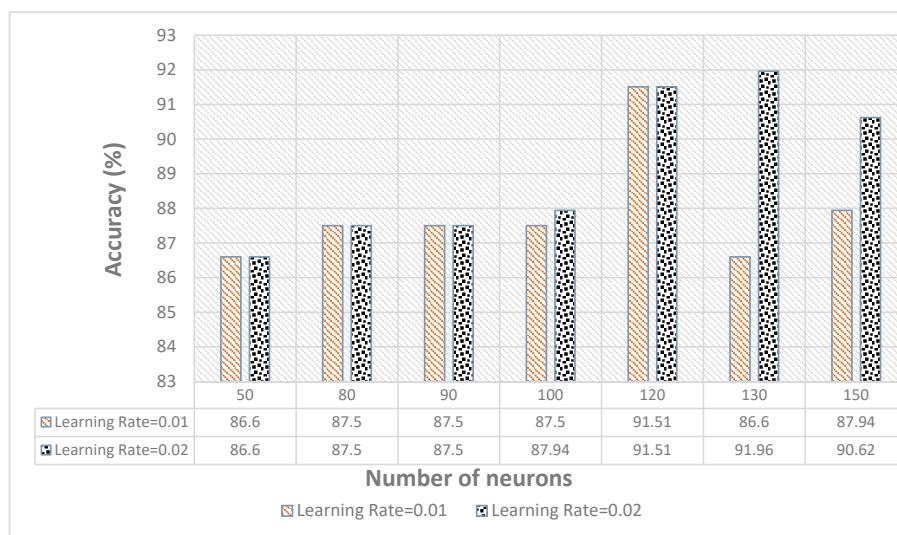


(b) Performance comparison of decision tree varying tree depths.

Figure 6. Cont.



(c) Performance comparison of random forest varying the number of trees.



(d) Performance comparison of single-layer back-propagation neural network using different learning rates.

Figure 6. Individual performance comparison of all the classifiers: (a) KNN with different distances applied, (b) DT (c) RF, and (d) single-layer NN.

The results for the single-layer network are summarized in Figure 6d with varying numbers of neurons and using the sigmoid activation function. It achieved the highest accuracy of 91.96% with 130 neurons. At each iteration, the learning rate helps to find the step size needed to minimize the loss function in order to obtain the best parameter that produces the highest accuracy. In the Figure 6d, the learning rate of 0.02 achieved the highest accuracy.

Figure 7 shows the confusion matrices for the different algorithms. In most of the algorithms, each individual gesture was recognized with accuracy more than 85%. Due to the similar nature of some gestures, some showed high correlation to each other such G1 and G3, as shown in Figure 7a,c. G3 showed low precision and recall values and had high correlation with G1 due to the similar variation in hand gesture movement. G22, G23, and G24 received the highest accuracy for almost every classification algorithm. These gestures involve angular hand movements that are mostly different from the other gestures.

Actual Predicted \	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	G20	G21	G22	G23	G24	Class Precision	
G1	89	1	6	0	2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	91.8%	
G2	0	79	2	0	0	1	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	90.8%	
G3	3	1	89	2	0	0	1	1	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	79.5%	
G4	1	0	2	83	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	91.2%	
G5	0	1	1	1	80	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	96.4%
G6	0	3	0	0	0	76	1	2	0	1	0	0	1	0	0	0	0	1	0	0	2	1	0	2	92.7%	
G7	1	1	0	0	0	0	93	0	1	0	0	3	0	0	0	0	0	1	0	0	0	0	0	0	86.9%	
G8	0	0	2	2	0	1	0	87	0	0	1	1	0	0	0	0	1	0	0	2	0	0	0	1	92.6%	
G9	0	0	0	0	0	0	3	1	85	1	0	1	0	2	0	0	0	0	0	0	0	0	0	0	91.4%	
G10	0	0	0	0	0	0	1	0	2	75	0	0	0	0	0	0	0	0	0	1	1	0	0	0	92.6%	
G11	1	0	1	0	0	1	1	0	0	0	94	0	0	0	0	0	1	0	0	1	0	0	0	0	93.1%	
G12	0	0	0	0	0	0	2	0	2	0	0	98	0	0	0	0	0	1	0	0	0	0	0	0	90.7%	
G13	0	0	1	0	0	0	0	0	0	0	0	90	0	2	5	0	0	0	2	0	0	0	0	0	90.9%	
G14	1	0	0	0	0	0	2	0	1	1	2	0	0	91	0	0	0	1	0	0	0	0	0	0	95.8%	
G15	1	0	2	0	0	0	0	1	0	0	0	0	1	1	76	6	1	0	0	0	0	0	0	0	90.5%	
G16	0	0	0	0	0	0	0	1	0	0	0	0	4	0	4	78	0	0	0	1	1	0	0	0	83.0%	
G17	0	0	0	1	0	0	0	0	1	1	0	0	0	0	1	89	0	0	0	1	0	0	0	3	85.6%	
G18	0	0	0	0	0	1	1	0	0	0	0	2	0	1	0	91	2	0	0	0	0	0	0	0	91.9%	
G19	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	2	95	0	0	0	0	0	96.9%	
G20	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	2	2	0	0	81	1	0	0	1	91.0%	
G21	0	0	2	1	0	1	0	1	0	1	0	2	0	0	0	0	2	0	0	1	77	0	1	0	91.7%	
G22	0	0	1	0	0	0	0	0	1	0	3	0	0	0	0	0	1	0	0	0	1	90	0	0	98.9%	
G23	0	1	1	0	0	0	0	0	0	1	0	0	1	0	1	1	1	0	0	0	83	0	0	0	97.6%	
G24	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	76	91.4%	
Class Recall	89.9	92.9	89.0	92.2	93.0	84.4	93.0	88.8	91.4	93.8	94.0	95.1	90.0	91.9	84.4	87.6	91.8	91.9	96.0	90.0	86.5	92.8	92.2	96.2		
	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%		

(a) Confusion matrix of Manhattan distance with 10-fold cross-validation.

Actual Predicted \	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	G20	G21	G22	G23	G24	Class precision
G1	6	0	5	0	1	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	37.50%
G2	0	4	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	57.14%
G3	1	0	0	4	0	1	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	40.00%
G4	1	0	0	3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	60.00%
G5	0	2	0	1	6	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	54.55%
G6	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%
G7	0	1	0	0	0	0	7	0	1	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	58.33%
G8	1	0	0	1	0	0	0	6	0	0	1	0	0	0	1	0	0	0	0	0	1	1	0	0	50.00%
G9	0	0	0	0	0	0	1	0	6	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	75.00%
G10	0	0	0	0	1	0	0	0	2	6	2	0	0	0	0	0	0	0	0	1	2	0	0	0	42.86%
G11	0	0	0	0	0	0	0	2	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	75.00%
G12	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%
G13	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	1	88.89%
G14	0	0	0	0	0	0	0	0	1	0	0	0	6	0	0	0	2	0	0	0	0	0	0	0	66.67%
G15	1	0	1	1	0	0	1	0	0	0	0	0	0	5	1	0	0	0	1	0	0	0	0	0	45.45%
G16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	100.00%
G17	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	7	0	0	0	0	0	0	0	1	70.00%
G18	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	5	1	0	0	0	0	0	0	0	71.43%
G19	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	7	0	0	0	1	0	0	0	0	63.64%
G20	1	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	1	1	5	0	0	0	0	0	41.67%
G21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	1	0	0	87.50%
G22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	100.00%
G23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	100.00%
G24	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	6	60.00%
Class recall	60%	50%	40%	33.33%	66.67%	88.89%	70%	60%	66.67%	75%	60%	70%	80%	60%	55.56%	77.78%	70%	50%	70%	55.56%	77.78%	90%	44.44%	75%	
	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%		

(b) Confusion matrix of the decision tree for 10% testing data only.

Figure 7. Cont.

Actual Predicted	G 1	G 2	G 3	G 4	G 5	G 6	G 7	G 8	G 9	G 10	G 11	G 12	G 13	G 14	G 15	G 16	G 17	G 18	G 19	G 20	G 21	G 22	G 23	G 24	Class precision	
G1	16	1	3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	76.19%	
G2	0	13	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	92.86%	
G3	3	0	15	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	75.00%	
G4	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G5	0	0	0	0	17	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	94.44%	
G6	0	1	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	94.12%	
G7	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G8	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	94.74%	
G9	0	1	0	0	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	94.74%	
G10	0	0	0	0	0	1	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	94.12%	
G11	0	0	0	1	0	0	1	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	88.89%	
G12	0	0	0	0	0	0	0	0	1	0	21	0	0	0	0	0	0	0	0	0	0	0	0	2	84.00%	
G13	0	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G14	0	0	0	0	0	0	0	0	1	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	95.00%	
G15	1	0	0	0	0	0	0	1	0	0	0	1	0	16	0	1	0	0	0	0	0	0	0	1	76.19%	
G16	0	0	0	0	0	0	0	0	0	0	0	0	0	1	18	1	0	0	0	0	0	0	1	0	85.71%	
G17	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	1	0	1	0	0	84.21%	
G18	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	94.74%	
G19	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	20	0	0	0	0	0	0	86.96%	
G20	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	94.74%	
G21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	0	0	100.00%	
G22	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	19	0	90.48%	
G23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	100.00%
G24	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	12	85.71%
Class Recall	80.0 %	76.47 %	75.0 %	88.89 %	100.0 %	88.8 %	95.0 %	90.0 %	89.4 %	100.0 %	80.0 %	100.0 %	90.0 %	95.0 %	88.89 %	100.0 %	84.21 %	90.0 %	100.0 %	100.0 %	94.4 %	100.0 %	88.8 %	75.0 %		

(c) Confusion matrix of the random forest for 80% training and 20% testing only.

Actual Predicted	G 1	G 2	G 3	G 4	G 5	G 6	G 7	G 8	G 9	G 10	G 11	G 12	G 13	G 14	G 15	G 16	G 17	G 18	G 19	G 20	G 21	G 22	G 23	G 24	Class Precision		
G1	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	87.5%		
G2	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%		
G3	0	0	10	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	90.9%		
G4	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%		
G5	0	0	0	0	7	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	100%		
G6	0	1	0	0	0	11	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	91.7%		
G7	0	0	0	0	0	0	5	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	71.4%		
G8	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%		
G9	0	0	0	0	0	0	0	0	8	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	100%		
G10	0	0	0	0	0	0	0	0	0	1	4	0	0	0	1	0	0	0	0	0	0	0	0	0	80.0%		
G11	0	0	1	0	0	0	0	1	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	81.8%		
G12	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	100%		
G13	0	0	0	0	0	0	0	0	0	0	0	0	8	0	1	0	0	0	0	0	0	0	0	0	88.9%		
G14	0	0	0	0	0	0	0	0	1	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	90.9%		
G15	0	0	0	0	0	0	0	0	0	0	1	0	0	0	17	0	0	0	0	0	0	0	0	0	94.4%		
G16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	100%		
G17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	1	0	0	1	0	70.0%		
G18	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	5	0	0	0	0	0	0	83.3%		
G19	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	11	0	0	0	0	0	91.7%		
G20	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0	0	92.9%		
G21	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	85.7%		
G22	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	10	0	0	0	0	90.9%	
G23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	100%	
G24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	100%
Class Recall	87.5% 90.9%	76.9% 100%	75.0% 100%	88.89% 100%	91.7% 100%	80.0% 100%	90.0% 100%	63.6% 100%	100% 100%	94.4% 100%	100% 100%	90.9% 90.9%	90.0% 90.0%														

(d) Confusion of the single-layer neural network with 130 neurons for 10% training only.

Figure 7. Confusion matrices of all the classifiers: (a) KNN, (b) DT (c) RF, and (d) single-layer NN.

4.3. Performance Comparison of the Classifiers

The overall comparison showed that the single-layer neural network produced the most accurate results. However, KNN algorithm with the Manhattan distance along with random forest also produced similar results for the data set, as shown in Figure 8. Figure 9 shows the precision and recall of the classifiers. Overall, the three classifiers, KNN using Manhattan distance, RF, and single-layer NN, had an accuracy of about 91%.

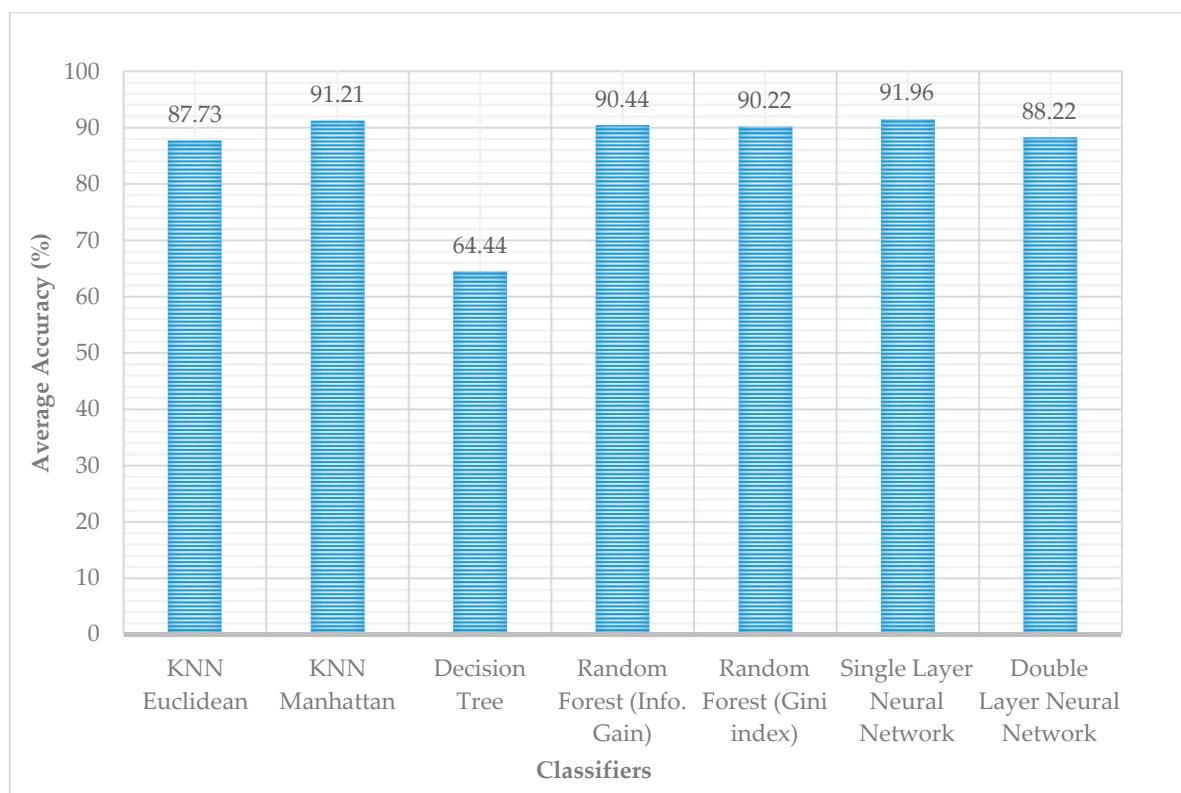


Figure 8. Performance comparison of different classifiers in terms of the accuracy of ASD gestures recognition.

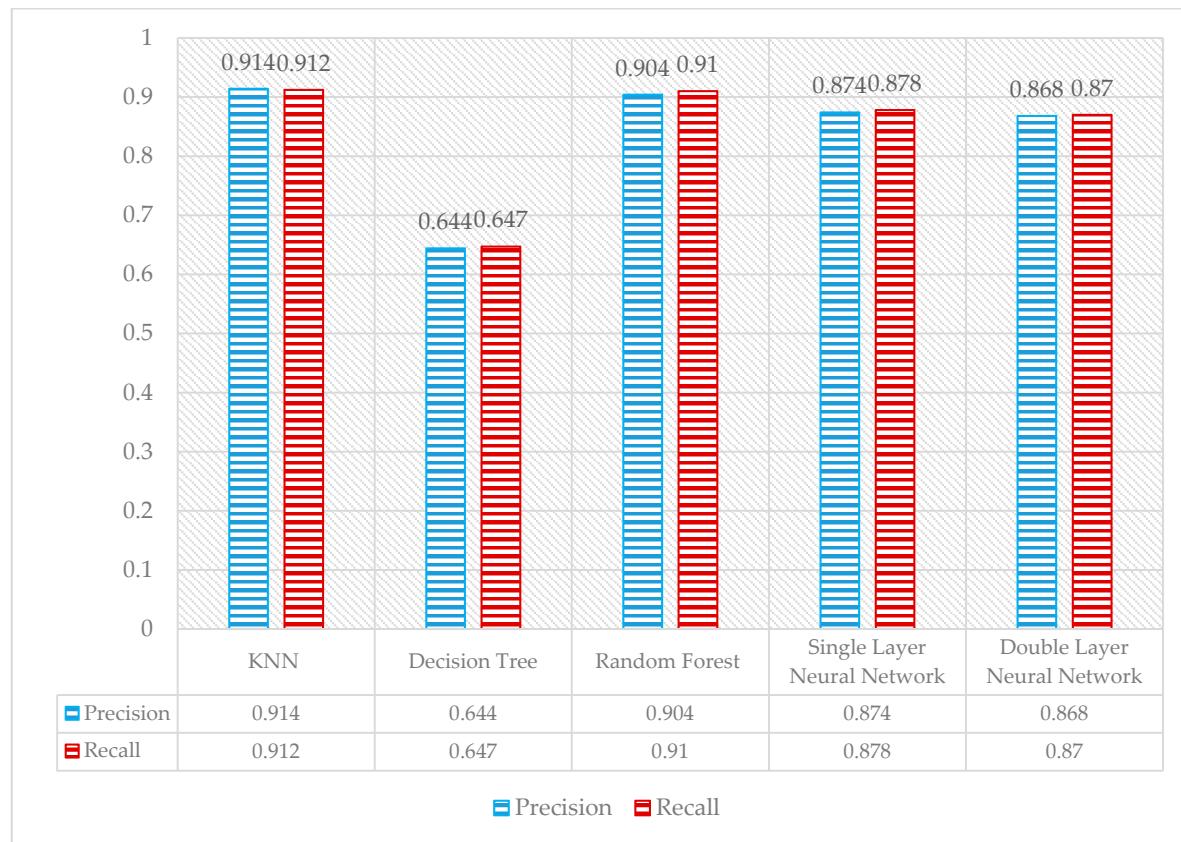


Figure 9. Average comparison of precision and recall values of all classifiers.

4.4. Performance Comparison of the Classifiers Using Leave-One-Person-out Cross-Validation

In the leave-one-person-out cross-validation (LOOCV), we used the nine subjects for training and one subject for testing. The subject used for testing performed each gesture seven times. We present the results of RF and NN using backpropagation. Figure 10 shows the performance comparison of the LOOCV for RF, which shows that RF using information gain had a recognition accuracy greater than 91%, which was achieved by the 10-fold cross-validation.

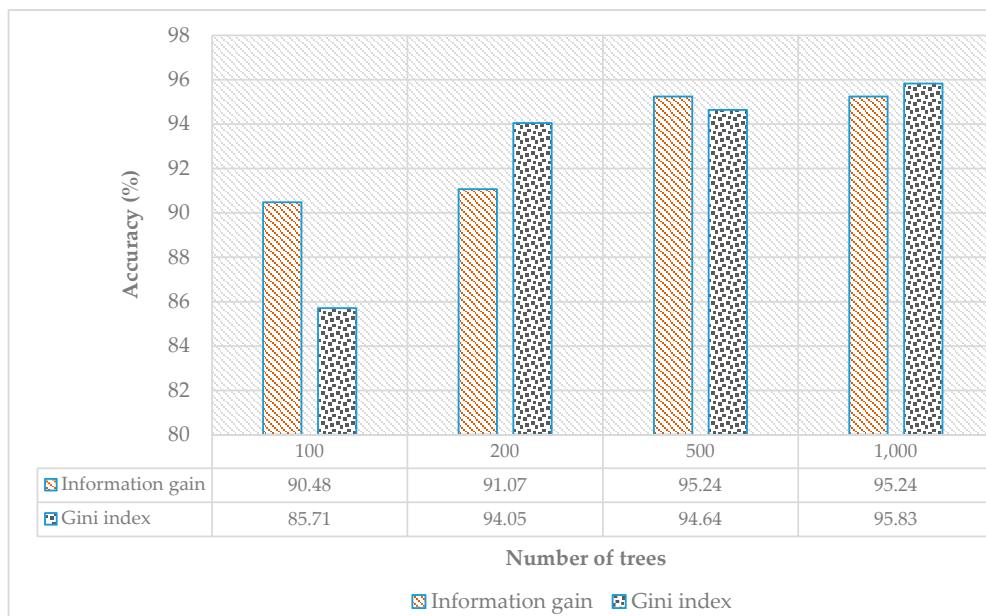


Figure 10. Performance of gestures recognition accuracy of random forest using LOOCV.

Figure 11 shows the confusion matrix of RF using LOOCV, which shows that each gesture was recognized with higher accuracy.

Actual \ Predicted	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	G20	G21	G22	G23	G24	Class Precision
G1	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G2	0	7	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	87.50%	
G3	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G4	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G5	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G6	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G7	0	0	0	0	0	0	6	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	75.00%	
G8	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G9	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G10	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G11	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G12	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G13	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	100.00%	
G14	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	100.00%	
G15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	100.00%	
G16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	100.00%	
G17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	100.00%	
G18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	100.00%	
G19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	100.00%	
G20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	100.00%	
G21	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	7	0	77.78%	
G22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	100.00%	
G23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	100.00%	
G24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	77.78%	
Class Recall	100	100	100	85.	100	100	85.	100	71.	100	100	100	100	100	85.	85.	100	85.	100	100	100	100	100.		
%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%		

Figure 11. Confusion matrix of random forest using LOOCV.

Figure 12 shows the accuracy of ASD children gestures recognition of NN back-propagation using LOOCV. The NN also performed better and each gesture was recognized with higher accuracy, as shown in Figure 13 for a learning rate of 0.02.

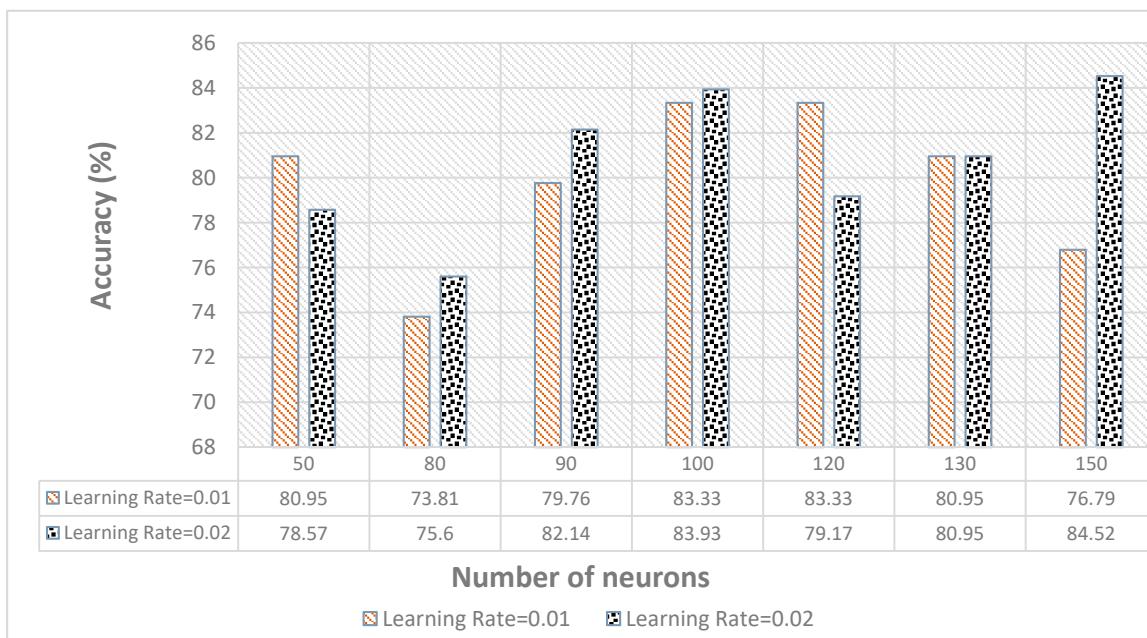


Figure 12. Performance of gestures recognition accuracy of the neural network using LOOCV using different learning rates.

Actual \ Predicted	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	G20	G21	G22	G23	G24	Class Precision	
G1	7	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	70.0%	
G2	0	5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	83.3%	
G3	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G4	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G5	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G6	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G7	0	0	0	0	0	0	6	0	1	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	60.00%	
G8	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G9	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G10	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G11	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G12	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	100.00%	
G13	0	1	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	1	0	77.78%	
G14	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	100.00%	
G15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00%	
G16	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	85.71%	
G17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	7	0	0	0	0	0	0	1	70.00%	
G18	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	5	0	0	0	0	0	0	71.43%	
G19	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	6	0	0	0	0	0	75.00%	
G20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	6	0	0	0	0	75.00%	
G21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	7	0	0	0	87.50%	
G22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	100.00%	
G23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	100.00%	
G24	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	2	1	0	0	0	0	6	54.55%	
Class Recall	100	71.	71.	100	100	100	85.	100	85.	71.	57.	85.	100	100	0.	85.	100	71.	85.	85.	100	100	85.7	85.7		
	.00	.43	.43	.00	.00	.00	.71	.00	.71	.43	.14	.71	.00	.00	.00	.71	.00	.43	.71	.71	.00	.00	.00	.1%	1%	
	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%			

Figure 13. Confusion matrix of the neural network using LOOCV.

5. Conclusions, Limitations, and Future Work

In this paper, we proposed a wearable-sensors-based platform for recognizing the gesture movements of children with autism spectrum disorder (ASD) using machine learning algorithms. This work focused on recognizing the daily gestures of ASD children to enable them to communicate with normal people without any hesitation. Modern Androids and smart watches are equipped with sensors such as gyroscopes, accelerometers, and GPS. Smartphones are the most widespread platform used for the recognition of human physical activities. However, their placement on the body creates some limitations, whereas wearable sensors can be placed easily on the human body. We used a wrist-worn sensors module consisting of an accelerometer and gyroscope for the x, y, and z axes and acquired the data at 50 Hz to obtain both the linear and angular motion for increased recognition accuracy of complex gestures. Twenty-four gestures were performed by ten subjects, and each gesture was repeated 7–12 times. A window of 3 s was used to extract various statistical measures (45 features) from the sensors data. The gesture was completed in about 3 s, so we considered a window of 3 s. We compared four classifiers: K-nearest neighbor (KNN), decision tree (DT), random forest (RF), and neural network (NN) using back-propagation using data-based cross-validation and leave-one-person-out cross-validation (LOOCV). Both the data-based 10-fold cross-validation and LOOCV produced accuracies greater than 91%. The RF, NN, and KNN showed about similar accuracy. Most of the individual gestures were recognized with accuracy greater than 90% by both data-partitioning-based cross-validation and LOOCV. This paper focused only the recognition of gestures of ASD children, but the proposed architecture can be utilized for remote health monitoring of ASD children.

The data were collected in a constrained environment where the ASD children performed the gestures in standing positions. The system and data set have not been validated on sensors data from the ASD children in the sitting or any other body position. The Hexi-wear sensor has also limited processing and battery power. Its battery needs recharging after a certain time. The data were collected using a single sensor worn at the wrist position, so may not be capable of recognizing complex gestures.

In future work, we will work on a multiple- and heterogeneous-sensors-based platform for the gesture recognition of ASD children. We are also working with LSTM-based ML algorithms for validation and performance comparison with RF and NN back-propagation.

Author Contributions: Conceptualization, F.U.; data curation, S.P. and H.S.; formal analysis, F.U. and A.K.; funding acquisition, K.-S.K.; methodology, A.I.; software, U.A.S. and R.U.; supervision, K.-S.K.; writing—original draft, U.A.S., S.P. and H.S.; Writing—review & editing, F.U. and K.-S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a National Research Foundation of Korea Grant funded by the Korean Government (Ministry of Science and ICT)-NRF-2020R1A2B5B02002478.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available now and will be uploaded later to the public repository.

Conflicts of Interest: The authors declare no conflict of interest.

References

- National Institute of Mental Health. Autism Spectrum Disorder. 2018. Available online: <https://www.nimh.nih.gov/health-topics/autism-spectrum-disorders-asd/index.shtml> (accessed on 10 December 2020).
- Amiri, A.; Peltier, N.; Goldberg, C.; Sun, Y.; Nathan, A.; Hiremath, S.; Mankodiya, K. WearSense: Detecting Autism stereotypic behaviors through smartwatches. *Healthcare* **2017**, *5*, 11. [[CrossRef](#)] [[PubMed](#)]
- Alwakeel, S.S.; Alhalabi, B.; Aggoune, H.; Alwakeel, M. A machine learning based WSN system for autism activity recognition. In Proceedings of the 2015 IEEE 14th International Conference on Machine Learning ICMLA 2015, Miami, FL, USA, 9–11 December 2015; pp. 771–776. [[CrossRef](#)]
- Loomes, R.; Hull, L.; Mandy, W.P.L. What is the male-to-female ratio in Autism Spectrum Disorder? A systematic review and meta-analysis. *J. Am. Acad. Child Adolesc. Psychiatry* **2017**, *56*, 466–474. [[CrossRef](#)]
- Capirci, O.; Iverson, J.M.; Pizzuto, E.; Volterra, V. Gestures and words during the transition to two-word speech. *J. Child Lang.* **1996**, *23*, 645–673. [[CrossRef](#)]
- Goldin-Meadow, S.; Alibali, M.W. Gesture’s role in speaking, learning, and creating language. *Annu. Rev. Psychol.* **2013**, *64*, 257–283. [[CrossRef](#)]
- Perrault, A.; Chaby, L.; Bigouret, F.; Oppetit, A.; Cohen, D.; Plaza, M.; Xavier, J. Comprehension of conventional gestures in typical children, children with autism spectrum disorders and children with language disorders. *Neuropsychiatr. Enfance. Adolesc.* **2019**, *67*, 1–9. [[CrossRef](#)]
- Metcalf, D.; Milliard, S.T.J.; Gomez, M.; Schwartz, M. Wearables and the internet of things for health: Wearable, interconnected devices promise more efficient and comprehensive health care. *IEEE Pulse* **2016**, *7*, 35–39. [[CrossRef](#)] [[PubMed](#)]
- Clifton, L.; Clifton, D.A.; Pimentel, M.A.F.; Watkinson, P.J.; Tarassenko, L. Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors. *IEEE J. Biomed. Heal. Inform.* **2014**, *18*, 722–730. [[CrossRef](#)]
- Free, C.; Phillips, G.; Galli, L.; Watson, L.; Felix, L.; Edwards, P.; Patel, V.; Haines, A. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: A systematic review. *PLoS Med.* **2013**, *10*, e1001362. [[CrossRef](#)] [[PubMed](#)]
- Shoaib, M. Sitting Is the New Smoking: Online Complex Human Activity Recognition with Smartphones and Wearables. Available online: <https://research.utwente.nl/en/publications/sitting-is-the-new-smoking-online-complex-human-activity-recognit> (accessed on 10 April 2021).
- Lowry, B.L. “Tuning In” to Others: How Young Children Develop Theory of Mind. Available online: <http://www.hanen.org/Helpful-Info/Articles/Tuning-In-to-Others-How-Young-Children-Develop.aspx> (accessed on 10 May 2021).
- Levy, S.E.; Mandell, D.S.; Schultz, R.T. Autism. *Lancet* **2009**, *374*, 1627–1638. [[CrossRef](#)]
- Behaviour, A.; Cb, C. Stereotypes: A critical review. *Anim. Behav.* **1991**, *41*, 1015–1037.
- Camada, M.Y.O.; St, D.; Conceic, G.S. Recognition of affective state for austist from stereotyped gestures. In Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics, Lisbon, Portugal, 29–31 July 2016; Volume 1, pp. 197–204. [[CrossRef](#)]
- Guo, H.; Chen, L.; Chen, G.; Lv, M. Smartphone-based activity recognition independent of device orientation and placement. *Int. J. Commun. Syst.* **2015**, *29*, 2403–2415. [[CrossRef](#)]
- Mohammadian Rad, N.; Kia, S.M.; Zarbo, C.; van Laarhoven, T.; Jurman, G.; Venuti, P.; Marchiori, E.; Furlanello, C. Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders. *Signal Processing* **2018**, *144*, 180–191. [[CrossRef](#)]
- Muir, L.J.; Richardson, I.E.G. Perception of sign language and its application to visual communications for deaf people. *J. Deaf Stud. Deaf Educ.* **2005**, *10*, 390–401. [[CrossRef](#)] [[PubMed](#)]
- Piana, S.; Staglianò, A.; Camurri, A.; Odono, F. A Set of Full-Body Movement Features for Emotion Recognition to Help Children Affected by Autism Spectrum Condition. Available online: http://www.fdg2013.org/program/workshops/papers/IDGEI2013_idgei2013_4.pdf (accessed on 15 April 2021).
- Rad, N.M.; Bizzego, A.; Kia, S.M.; Jurman, G.; Venuti, P.; Furlanello, C. Convolutional Neural Network for Stereotypical Motor Movement Detection in Autism. *arXiv* **2015**, arXiv:1511.01865; pp. 15–19, 15–19.
- Faja, S.; Webb, S.J.; Jones, E.; Merkle, K.; Kamara, D.; Bavaro, J.; Aylward, E.; Dawson, G. The effects of face expertise training on the behavioral performance and brain activity of adults with high functioning autism spectrum disorders. *J. Autism Dev. Disord.* **2012**, *42*, 278–293. [[CrossRef](#)]
- Gonçalves, N.; Costa, S.; Rodrigues, J.; Soares, F. Detection of stereotyped hand flapping movements in Autistic children using the Kinect sensor: A case study. In Proceedings of the 2014 IEEE International Conference on Autonomous Robot Systems and Competitions ICARSC 2014, Espinho, Portugal, 14–15 May 2014; pp. 212–216. [[CrossRef](#)]
- Incertis, I.G.; García-Bermejo, J.G.; Casanova, E.Z. Hand gesture recognition for deaf people interfacing. In Proceedings of the International Conference on Pattern Recognition, Hong Kong, China, 20–24 August 2006; Volume 2, pp. 100–103. [[CrossRef](#)]
- Ángel-López, J.P.; de la Peña, N.A. *Proceedings of the VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, 26–28 October 2016*; Springer: Singapore, 2017; Volume 60, pp. 520–523. [[CrossRef](#)]
- Sombandith, V.; Walairacht, A.; Walairacht, S. Hand gesture recognition for Lao alphabet sign language using HOG and correlation. In Proceedings of the ECTI-CON 2017—2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Phuket, Thailand, 27–30 June 2017; pp. 649–651. [[CrossRef](#)]

26. Pappas, I.P.I.; Popovic, M.R.; Keller, T.; Dietz, V.; Morari, M. A reliable gait phase detection system. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2001**, *9*, 113–125. [[CrossRef](#)]
27. Wu, F.J.; Solmaz, G. We hear your activities through Wi-Fi signals. In Proceedings of the 2016 IEEE 3rd World Forum Internet Things, WF-IoT 2016, Reston, VA, USA, 12–14 December 2016; pp. 251–256. [[CrossRef](#)]
28. Darwish, S.M.; Madbouly, M.M.; Khorsheed, M.B. Hand gesture recognition for sign language: A new higher order fuzzy HMM approach. *Int. J. Eng. Technol.* **2016**, *8*, 157–164. [[CrossRef](#)]
29. Bhattacharya, A.; Zope, V.; Kumbhar, K.; Borwankar, P.; Mendes, A. Classification of sign language gestures using machine learning. *Int. J. Adv. Res. Comput. Commun. Eng.* **2020**, *8*, 97–103. [[CrossRef](#)]
30. Žemgulys, J.; Raudonis, V.; Maskeliunas, R.; Damaševičius, R. Recognition of basketball referee signals from videos using histogram of oriented gradients (HOG) and support vector machine (SVM). *Procedia Comput. Sci.* **2018**, *130*, 953–960. [[CrossRef](#)]
31. Liu, S.; Gao, R.X.; John, D.; Staudenmayer, J.W.; Freedson, P.S. Multisensor data fusion for physical activity assessment. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 687–696. [[CrossRef](#)] [[PubMed](#)]
32. Gyllensten, I.C.; Bonomi, A.G. Identifying types of physical activity with a single accelerometer: Evaluating laboratory-trained algorithms in daily life. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 2656–2663. [[CrossRef](#)]
33. Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. Activity recognition using cell phone accelerometers. *ACM SigKDD Explor. Newsl.* **2011**, *12*, 74–82. [[CrossRef](#)]
34. Comput, J.P.D.; Uddin, Z. A wearable sensor-based activity prediction system to facilitate edge computing in smart healthcare system. *J. Parallel Distrib. Comput.* **2019**, *123*, 46–53. [[CrossRef](#)]
35. Zhou, S.; Liu, Y.; Li, K. Recognition of Multi-scale Multi-angle Gestures Based on HOG-LBP Feature. In Proceedings of the 2018 15th International Conference on Control, Automation, Robotics and Vision, ICARCV 2018, Singapore, 18–21 November 2018; pp. 407–412. [[CrossRef](#)]
36. Yao, R.; Lin, G.; Shi, Q.; Ranasinghe, D.C. Efficient dense labelling of human activity sequences from wearables using fully convolutional networks rui. *Pattern Recognit.* **2017**. [[CrossRef](#)]
37. Vaitkevičius, A.; Taroza, M.; Blažauskas, T.; Damaševičius, R.; Maskeliunas, R.; Woźniak, M. Recognition of American sign language gestures in a virtual reality using leap motion. *Appl. Sci.* **2019**, *9*, 445. [[CrossRef](#)]
38. Mukherjee, A.; Misra, S.; Mangrulkar, P.; Rajarajan, M.; Rahulamathavan, Y. SmartARM: A smartphone-based group activity recognition and monitoring scheme for military applications. In Proceedings of the 11th IEEE International Conference on Advanced Networks and Telecommunications Systems, ANTS 2017, Bhubaneswar, India, 17–20 December 2017; pp. 1–6. [[CrossRef](#)]
39. Farooq, M.; Sazonov, E. Real time monitoring and recognition of eating and physical activity with a wearable device connected to the eyeglass. In Proceedings of the Eleventh International Conference on Sensing Technology, ICST, Sydney, Australia, 4–6 December 2017; pp. 1–6. [[CrossRef](#)]
40. Ramos, A.L.A.; Dalhag, G.D.M.; Daygon, M.L.D.; Omar, J.; La Cruz, K.D.; Macaranas, A.A.; Almodovar, K.L.J. Alphabet hand gesture recognition using histogram of oriented gradients, support vector machine and K-nearest neighbor algorithm. *Int. Res. J. Comput. Sci.* **2019**, *6*, 200–205.
41. Shukor, A.Z.; Miskon, M.F.; Jamaluddin, M.H.; Ibrahim, A.; Asyraf, M.F.; Bazli, M. A new data glove approach for Malaysian sign language detection. *Procedia Comput. Sci.* **2015**, *76*, 60–67. [[CrossRef](#)]
42. Nasri, N.; Orts-Escalano, S.; Cazorla, M. An semg-controlled 3d game for rehabilitation therapies: Real-time time hand gesture recognition using deep learning techniques. *Sensors* **2020**, *20*, 6451. [[CrossRef](#)]
43. Parcheta, Z.; Martínez-Hinarejos, C.D. Sign language gesture recognition using HMM. *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinform.* **2017**, *10255 LNCS*, 419–426. [[CrossRef](#)]
44. Mannini, A.; Rosenberger, M.; Haskell, W.L.; Sabatini, A.M.; Intille, S.S. Activity recognition in youth using single accelerometer placed at wrist or ankle. *Med. Sci. Sports Exerc.* **2017**. [[CrossRef](#)]
45. Krishnan, K.S.; Saha, A.; Ramachandran, S.; Kumar, S. Recognition of human arm gestures using Myo armband for the game of hand cricket. In Proceedings of the 2017 IEEE 5th International Symposium on Robotics and Intelligent Sensors, IRIS 2017, Ottawa, ON, Canada, 5–7 October 2017; pp. 389–394. [[CrossRef](#)]
46. Thabtah, F.; Kamalov, F.; Rajab, K. A new computational intelligence approach to detect autistic features for autism screening. *Int. J. Med. Inform.* **2018**. [[CrossRef](#)] [[PubMed](#)]
47. Raj, S.; Masood, S. Analysis and detection of autism spectrum disorder using machine learning techniques. *Procedia Comput. Sci.* **2020**, *167*, 994–1004. [[CrossRef](#)]
48. Iqbal, A.; Ullah, F.; Anwar, H.; Ur Rehman, A.; Shah, K.; Baig, A.; Ali, S.; Yoo, S.; Kwak, K.S. Wearable internet-of-things platform for human activity recognition and health care. *Int. J. Distrib. Sens. Netw.* **2020**, *16*. [[CrossRef](#)]
49. Guo, G.; Wang, H.; Bell, D.; Bi, Y.; Greer, K. KNN model-based approach in classification. In *On the Move to Meaningful Internet Systems*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 986–996.
50. Yoav, F.; Mason, L. The alternating decision tree learning algorithm. In Proceedings of the ICML, Bled, Slovenia, 27–30 June 1999; Volume 99, pp. 124–133.
51. Breiman, L. Random forests. In *Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2001; Volume 45, pp. 5–32.
52. Hecht-Nielsen, R. Theory of the backpropagation neural network. In *Neural Networks for Perception*; Academic Press: Cambridge, MA, USA, 1992; pp. 65–93.