DOI: https://doi.org/10.33103/uot.ijccce.22.4.11

Machine Learning Approach Based on Automated Classification for Leg Rehabilitation

Ayat Naji Hussain[1], Sahar Adil Abboud[2] , Basim Abdul baki Jumaa[3], Mohammed Najm Abdullah[4]

[1,2,3,4] Computer Engineering Department, University of Technology, Baghdad, Iraq

[1]ce.19.14@grad.uotechnology.edu.iq, [2]sahar.a.abboud@uotechnology.edu.iq, [3]120003@uotechnology.edu.iq, [4]mohammed.n.abdullah@uotechnology.edu.iq

Abstract *—Human gait data follows distinct and identifiable patterns that are critical for movement analysis and evaluation Like other biological signals. The success of a rehabilitation program is dependent on the execution of proper progress monitoring. To ensure success, diagnosis of gait anomalies, as well as the implementation of therapy to address them, must be validated in a constant and timely manner in developing youngsters. In this paper, machine learning techniques were utilized to classify foot diseases and the purpose is to increase the accuracy of disease detection and diagnosis because intelligent systems can contribute significantly in the medical field and have proven their worth in diagnosing many diseases. The results show high accuracy of the used machine learning algorithms, where the accuracy of the classifiers reached 100% for Random Forest (RF), Decision Tree (DT), and k-nearest neighbors (KNN), while it reached 98% for Logistic Regression.*

Index Terms*—Biometrics, Machine Learning (ML), Drop foot (DF), Leg Rehabilitation, and Human gait.*

I. INTRODUCTION

Many factors are used to describe persons and their actions, regardless of their look, form, or hair and skin colors. Gait, facial characteristics, fingerprints, voice, iris, and other unique identifiers such as skin spots are all examples of biometrics. Height, stride length, and the lengths of the silhouette bounding box are all fixed variables in gait [1]. Hundreds of joints and muscles can be employed to perform a variety of tasks that describe each individual’s distinctive walking style, according to biomechanical and clinical studies [2]. Walking is one of the most common human physical activities, and it may be done in a variety of environments. Human gait patterns can reveal a lot about a person’s physical and neurological functioning, and they can also help with the identification of human motor abnormalities in pathological situations. For these goals, human gait patterns must be recognized and classified according to the situation or clinical condition of the researched locomotor function [3]. ML approaches are beneficial when algorithmic solutions are unavailable, formal models are lacking, or information about the application area is inadequately specified. ML has been increasingly studied in the context of medical sciences throughout the last few decades [4]. Machine learning is being utilized in this context to examine the usefulness of clinical parameters in combination with prognosis, such as sickness development prediction, and to extract medical information for outcomes research, treatment planning and support, and overall patient care [5]. A sample (for example, a patient) is represented in machine learning (ML) by a set of features that may include the patient's attributes, risk factors, shape/texture characteristics in medical images, and clinical history data. These properties are frequently concatenated to produce a multidimensional feature vector to aid in the learning process. As

Received 17/December/2021; Accepted 08/March/2022

DOI: https://doi.org/10.33103/uot.ijccce.22.4.11

indicated in *Fig. 1*, the two stages of ML systems are the learning phase (training) and the testing phase (testing) [6].

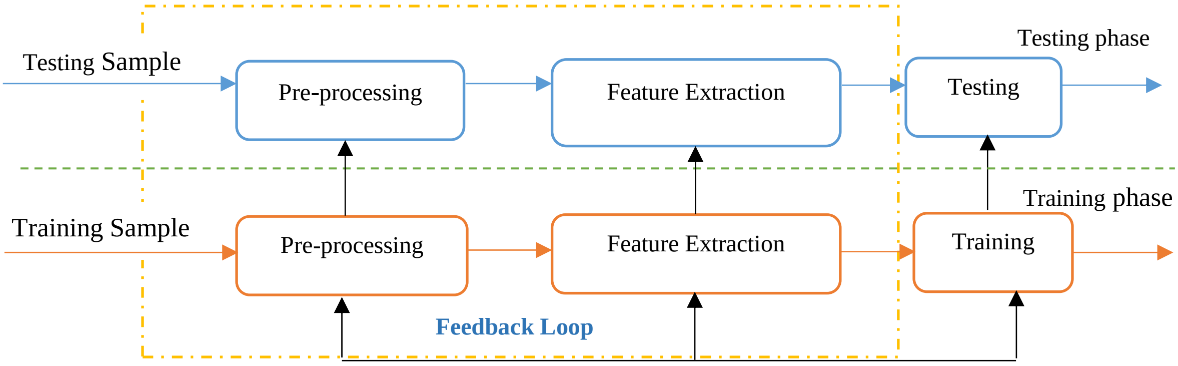


FIG. 1. EXAMPLE OF A MACHINE LEARNING SYSTEM [6].

II. RELATED WORKS

Drop Foot (FD) refers to the inability to elevate one's foot due to dorsiflexion muscular paralysis induced by a stroke. Recognizing the patient's FD and offering therapy as needed is a crucial part of rehabilitation. Some of the prior work relevant to this subject will be covered in this section:

In 2016, the researchers investigated the Extreme Learning Machine (ELM) method's capacity to distinguish between diseased and healthy muscles in the leg using EMG, which supplies the FD. The outcomes are compared to the Support Vector Machine (SVM) and the Neural Network (NN). ELM outperforms SVM and NN in classification performance, achieving up to 97 percent classification accuracy when two channels are used on each side of the leg [7]. In 2017, the researcher presented a variety of classification methods for detecting Parkinson's Disease using EMG readings. In a comparison examination, three distinct categorization systems were applied. They are, in that sequence, Neural Network (NN), Naïve Bayes, and Logistic Regression. To determine the performance of these classifiers, many assessment criteria were used. The classification accuracy of the Naïve Bayes Classifier was determined to be the best, with an accuracy of 99%, while the accuracy of the NN and Logistic Regression was 97% and 95%, respectively [8]. In 2019, the researchers used supervised machine learning methods to classify sagittal gait patterns in children with cerebral palsy and spastic diplegia. Data from 200 children with spastic diplegia CP was utilized to develop gait characteristics that reflected the most essential kinematic elements of each child's stride. On the same gait data, the following supervised machine learning approaches were investigated: ANN, Discriminant Analysis, Naïve Bayes (NB), Decision Trees (DT), k-nearest neighbors (KNN), support vector machine (SVM), and random forest are all examples of artificial neural networks. According to the data, the ANN approach provides the most accurate predictions (93.5%), followed by the SVM and random forest techniques, all of which have strong prediction accuracy (> 77.9%). In terms of classification, discriminant analysis, Naïve Bayes, and KNN all perform badly [9]. In 2019, The authors developed and validated a computerized technology that classified people into three groups based on lower-body motion data and pattern recognition algorithms: healthy seniors, geriatrics, and Parkinson's Disease patients. Using just accelerometer data, an ideal collection of gait characteristics was produced using a suggested feature selection approach based on maximum information gain and minimum correlation (MIGMC) among the features. This research looks at a new set of features developed in partnership with machine learning techniques including Support Vector Machine, Random Forest, AdaBoost, Bagging, and Naïve Bayes across a variety of feature sets. The efficacy of the best set of gait characteristics derived using our proposed feature selection approach was validated using a Similarity

Received 17/December/2021; Accepted 08/March/2022
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.22.4.11

Network Model. The results show that ensemble approaches, notably boosting classifiers, outperformed alternative classifiers [10]. In 2020, The researchers using spatiotemporal gait data, identified lower limb amputees (transtibial and trans femoral) from those who had their limbs removed (control). To identify the data, machine learning, KNN (K-nearest neighbors), and RF approaches were utilized (random forest). The researchers looked into three treadmill walking options: horizontal (0°), uphill (+8%), and downhill (–8%). These factors were important in deciding which circumstances the data is more discriminating. The accuracy of data categorization was 75.8% for KNN and 77.7% for RF when all situations were considered [11].

III. THE PROPOSED SYSTEM ARCHITECTURE

This research presents a Computational Intelligence for automatic classification of Foot Drop Rehabilitation with machine learning approaches. To achieve this goal, four machine learning techniques were used: Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), and k-nearest neighbors are all examples of decision trees (KNN). As a consequence, classifiers based on machine learning techniques will be utilized to assess input signals in order to identify the best and most significant one for Foot Drop detection. The diagram of the proposed system is shown in *Fig. 2* and Algorithm (1) below.

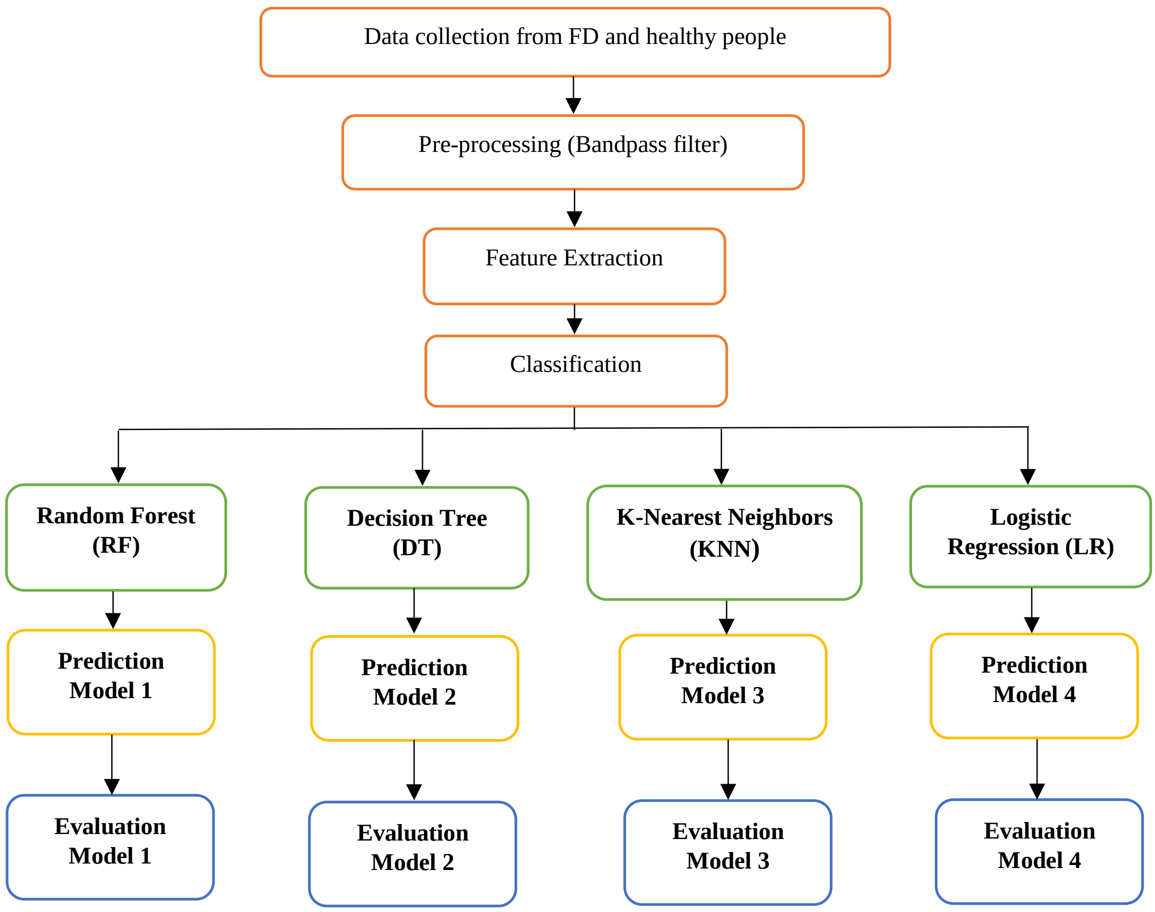


FIG. 2. THE PROPOSED SYSTEM ARCHITECTURE BASED ON MACHINE LEARNING.

Received 17/December/2021; Accepted 08/March/2022

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.22.4.11

Algorithm (1). The proposed System Based on Machine Learning
Input: Dataset
Output: Accuracy of (RF, DT, KNN, and LR), Best classifier accuracy
<u>Begin</u> **1:** Load dataset **// Input//** **2:** Pre-Processing phase • Applying Bandpass filter **3:** Feature Extraction Phase • (min, max, average, mood, variance, standard division) **4:** Apply Holdout splitting // **Training Phase //** **5:** Test the remainder of the data that have been entered for training //**Testing Phase** // **6:** Classify Samples **// Classification Phase //** • Results (1) = RF (Features Set, Targets) • Results (2) = DT (Features Set, Targets) • Results (3) = KNN (Features Set, Targets) • Results (4) = LR (Features Set, Targets) **7:** Best Classifier = High (Accuracy) **// Output//** **<u>End</u>**

A-Preprocessing Phase

The first phase is pre-processing; in this phase, the bandpass filter achieves this purpose.

1-Apply Bandpass Filter

Bandpass filters are named for the center or peak wavelength they transmit and will block longer and shorter wavelengths, resulting in increased contrast and better control over variations in ambient illumination conditions that may occur over time. Most machine vision applications require “wide” Bandpass Filters.

B-Feature Extraction Phase

The second phase is a feature extraction phase (min, max, average, mood, variance, standard division) which is applied in such sequence:

1-Maximum and Minimum Features Extraction

By preserving the largest feature variance and the lowest reconstruction error, the dominant features preserve the majority of the information.

2- Average Features Extraction

This means the average value of the data set which calculate in Eq. (1) as follows:

$$\bar{X} = \frac{x_1+x_2+ \ldots +x_n}{n} \tag{1}$$

3-Mode Features Extraction

The most often occurring number in the data set, the mode is always the number from the data set.

For example: the mode for the data set: 19, 19, 34, 3, 10, 22, 10, 15, 25, 10, 6. The number that occurs the most is number 10, mode = 10.

4-Variance Features Extraction

Received 17/December/2021; Accepted 08/March/2022

ISSN (Print) 1811-9212 *ISSN (Online) 2617-3352*

DOI: https://doi.org/10.33103/uot.ijccce.22.4.11

Measures how far the data set's values deviate from the mean on average. The population variance is the average of the squared deviations. Eq. (2) explain the calculation of variance as follows:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} \tag{2}$$

5-Standard Division Features Extraction
It is a square root of the variance as shown in Eq. (3).

$$\sigma^2 = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}} \tag{3}$$

C-Classification Phase
The third and most important stage in the work is the stage of data classification using machine learning. The classification will be into four classes, as follows:

- Healthy person
- Patient in the early stages of the disease
- A patient in the middle of the disease
- The patient's condition is severe

This section explains the four classifiers that were used to classify a patient's state in this work as follows:

1-Random Forest Classifier (RF)
Random Forest is a flexible learning model that can solve both sorts of issues, whether they are regression or classification problems. It works by building several "decision trees" throughout the training phase and producing average forecasts of all the decision trees. In regression, the goal variable is continuous, but in classification-related issues, it is categorical. Random Forest is a data analysis algorithm that delivers a high accuracy score [12]. The goal is to discover an f(X) function that can predict Y. The prediction function, L (Y, f (X)), in Eq. (4) is determined to minimize the expected loss value [13]:

$$E_{xy} = \mathrm{L}\left(\mathrm{Y}, \mathrm{f(X)}\right) \tag{4}$$

2-Logistic Regression Classifier (LR)

The dependent variable is a two-category categorical variable, such as normal/abnormal. When one or more of the following properties are present in a category variable, it is used:

- There are two distinct types.
- The value range [14].

It is a technique in which learning functions are represented as f: A→B or P(B|A) for discrete-valued B, where A = (A1...An) is any vector with discrete or continuous values, and B = (A1...An) is any vector with discrete or continuous values. It looks into a parametric version of the P(B|A) distribution in which the parameters are derived directly from the training data. When B is Boolean, the parametric model is as follows [15]:

$$P(B=1|A) = \frac{1}{1+\exp(w_0+\sum_{i=1}^{n} w_i A_i)} \tag{5}$$

$$P(B=0|A) = \frac{\exp(w_0+\sum_{i=1}^{n} w_i A_i)}{1+\exp(w_0+\sum_{i=1}^{n} w_i A_i)} \tag{6}$$

Received 17/December/2021; Accepted 08/March/2022

DOI: https://doi.org/10.33103/uot.ijccce.22.4.11

The total of the two probabilities derived from Eq. (5) and Eq. (6) must equal one. The form P(B|A) is very useful since it leads to a straightforward linear equation-based classification strategy. To categorize an A, assign it a y k-value that optimizes P (B = y_k|A). If the following condition is true, the label B = 0 is produced.

$$1 < \frac{P(B = 0|A)}{P(B = 1|A)} \tag{7}$$

When Eq. (5) and Eq. (6) are substituted it becomes:

$$1 < \exp(w_0 + \sum_{i=1}^{n} w_i B_i) \tag{8}$$

Using the natural logarithm on both sides of Eq. (8), a linear classification method can be created that gives label B = 0 if A meets the criteria.

$$0 < w_0 + \sum_{i=1}^{n} w_i A_i \tag{9}$$

Otherwise, B = 1 is assigned [16].

3-Decision Tree Classifier (DT)

Decision trees are a strong tool that may be utilized in a variety of domains, including machine learning, image processing, and pattern recognition [17]. DT is a sequential model that links a set of fundamental tests in which a numerical property is effectively and uniformly compared to a threshold value in each test. The numerical weights in the neural network of connections between nodes are substantially more difficult to create than the conceptual principles [18]. Decision trees offer a wide range of applications due to their simple analysis and accuracy across many data types [19].

4-K-Nearest Neighbors (KNN)

In a statistical pattern recognition classification approach, the k-nearest neighbor decision rule (k-NN) is used. Each class contains a set of example prototypes that serve as a training set for the class's pattern vectors. The nearest k neighbors of an unknown vector are chosen from all prototype vectors, and the class name is selected by a majority rule. The value of k should be odd to avoid ties on class overlap zones [20]. For continuous variables, the distance metric was Euclidean Distance, which was determined using Eq. (10) [21]:

$$d(A, B) = \sqrt{\sum_{i=1}^{n} (fa_i - fb_i)^2} \tag{10}$$

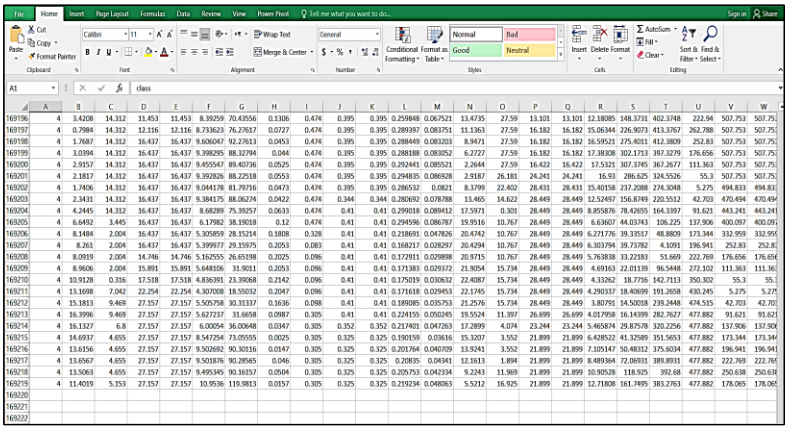
In this formula, A and B are observations and fa_i and fb_i are the i^{th} feature of observation A and observation B respectively.

IV. DATASET DESCRIPTIONS

The categorization was based on the sample data that was used to make it. The data was gathered under the supervision of Metro-Rehab Hospital's Stroke Coordinator. Data was gathered in three steps for each knee joint location (flexion or extension). An example of this data is shown in *Fig. 3*, it is included information about 169219 patients.

Received 17/December/2021; Accepted 08/March/2022
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.22.4.11



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
169196	4	3.4208	14.312	11.453	11.453	8.39259	70.43556	0.1306	0.474	0.395	0.395	0.259848	0.067521	13.4735	27.59	13.101	13.101	12.18085	148.3731	402.3748	222.94	507.753	507.75
169197	4	0.7984	14.312	12.116	12.116	8.733623	76.27617	0.0727	0.474	0.395	0.395	0.289397	0.083751	11.1363	27.59	16.182	16.182	15.06344	226.9073	413.3767	262.788	507.753	507.75
169198	4	1.7687	14.312	16.437	16.437	9.606047	92.27613	0.0453	0.474	0.395	0.395	0.288449	0.083203	8.9471	27.59	16.182	16.182	16.59521	275.4011	412.3809	252.83	507.753	507.75
169199	4	3.0394	14.312	16.437	16.437	9.398295	88.32794	0.044	0.474	0.395	0.395	0.288188	0.083052	6.2727	27.59	16.182	16.182	17.38308	302.1713	397.3279	176.656	507.753	507.75
169200	4	2.9157	14.312	16.437	16.437	9.455547	89.40736	0.0525	0.474	0.395	0.395	0.292441	0.085521	2.2644	27.59	16.422	16.422	17.5321	307.3745	367.2677	111.363	507.753	507.75
169201	4	2.1817	14.312	16.437	16.437	9.392826	88.22518	0.0553	0.474	0.395	0.395	0.294835	0.086928	2.9187	26.181	24.241	24.241	16.93	286.625	324.5526	55.3	507.753	507.75
169202	4	1.7406	14.312	16.437	16.437	9.044178	81.79716	0.0473	0.474	0.395	0.395	0.286532	0.0821	8.3799	22.402	28.431	28.431	15.40158	237.2088	274.3048	5.275	494.833	494.83
169203	4	2.3431	14.312	16.437	16.437	9.384175	88.06274	0.0422	0.474	0.344	0.344	0.280692	0.078788	13.465	14.622	28.449	28.449	12.52497	156.8749	220.5512	42.703	470.494	470.49
169204	4	4.2445	14.312	16.437	16.437	8.68289	75.39257	0.0633	0.474	0.41	0.41	0.299018	0.089412	17.5971	0.301	28.449	28.449	8.855876	78.42655	164.3397	91.621	443.241	443.241
169205	4	6.6492	3.445	16.437	16.437	6.17982	38.19018	0.12	0.474	0.41	0.41	0.294596	0.086787	19.9516	10.767	28.449	28.449	6.63607	44.03743	106.225	137.906	400.097	400.09
169206	4	8.1484	2.004	16.437	16.437	5.305859	28.15214	0.1808	0.328	0.41	0.41	0.218691	0.047826	20.4742	10.767	28.449	28.449	6.271776	39.33517	48.8809	173.344	332.959	332.95
169207	4	8.261	2.004	16.437	16.437	5.399977	29.15975	0.2053	0.083	0.41	0.41	0.168217	0.028297	20.4294	10.767	28.449	28.449	6.303794	39.73782	4.1091	196.941	252.83	252.8
169208	4	8.0919	2.004	14.746	14.746	5.162555	26.65198	0.2025	0.096	0.41	0.41	0.172911	0.029898	20.9715	10.767	28.449	28.449	5.763838	33.22183	51.669	222.769	176.656	176.65
169209	4	8.9606	2.004	15.891	15.891	5.648106	31.9011	0.2053	0.096	0.41	0.41	0.171383	0.029372	21.9054	15.734	28.449	28.449	4.69163	22.01139	96.5448	272.102	111.363	111.36
169210	4	10.9128	0.316	17.518	17.518	4.836391	23.39068	0.2142	0.096	0.41	0.41	0.175019	0.030632	22.4087	15.734	28.449	28.449	4.33262	18.7716	142.7113	350.302	55.3	55.
169211	4	13.1698	7.042	22.254	22.254	4.307008	18.55032	0.2047	0.096	0.41	0.41	0.171618	0.029453	22.1745	15.734	28.449	28.449	4.290337	18.40699	191.2658	430.245	5.275	5.27
169212	4	15.1813	9.469	27.157	27.157	5.505758	30.31337	0.1636	0.098	0.41	0.41	0.189085	0.035753	21.2576	15.734	28.449	28.449	3.80791	14.50018	239.2448	474.515	42.703	42.70
169213	4	16.3996	9.469	27.157	27.157	5.627237	31.6658	0.0987	0.305	0.41	0.41	0.224155	0.050245	19.5524	11.397	26.699	26.699	4.017958	16.14399	282.7627	477.882	91.621	91.621
169214	4	16.1327	6.8	27.157	27.157	6.00054	36.00648	0.0347	0.305	0.352	0.352	0.217401	0.047263	17.2899	4.074	23.244	23.244	5.465874	29.87578	320.2256	477.882	137.906	137.90
169215	4	14.6937	4.655	27.157	27.157	8.547254	73.05555	0.0025	0.305	0.325	0.325	0.190159	0.03616	15.3207	3.552	21.899	21.899	6.428522	41.32589	351.5653	477.882	173.344	173.34
169216	4	13.6156	4.655	27.157	27.157	9.502692	90.30116	0.0147	0.305	0.325	0.325	0.201764	0.040709	13.9241	3.552	21.899	21.899	7.105147	50.48312	375.6034	477.882	196.941	196.941
169217	4	13.6567	4.655	27.157	27.157	9.501876	90.28565	0.046	0.305	0.325	0.325	0.20835	0.04341	12.1613	1.894	21.899	21.899	8.489364	72.06931	389.8931	477.882	222.769	222.76
169218	4	13.5063	4.655	27.157	27.157	9.495345	90.16157	0.0504	0.305	0.325	0.325	0.205753	0.042334	9.2243	11.969	21.899	21.899	10.90528	118.925	392.68	477.882	250.638	250.63
169219	4	11.4019	5.153	27.157	27.157	10.9536	119.9813	0.0157	0.305	0.325	0.325	0.219234	0.048063	5.5212	16.925	21.899	21.899	12.71808	161.7495	383.2763	477.882	178.065	178.06
169220																							
169221																							
169222																							

FIG. 3. DATASET CONTENT.

Here, we review images of different positions of patients' feet during the examination as follows:

- **The first test** is to bend and straighten his or her lower leg at the knee joint from a resting posture while seated in a chair as shown in *Fig. 4 and 5.*

FIG. 4. FLEXION OF THE KNEES WITH FLEXION DORSIFLEXION A.

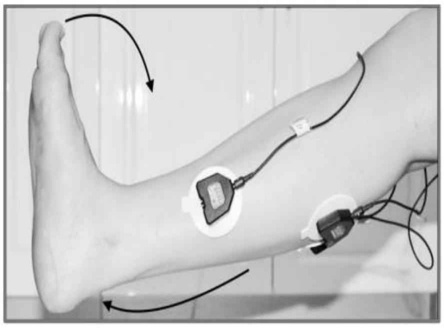


FIG. 5. FLEXION DORSIFLEXION WITH KNEE EXTENSION B.

- **The second test** He/she sat in the chair, as shown in *Fig. 6 and 7*, as from the position of rest, he/she stretched his/her foot as much of it as allowed up or down.

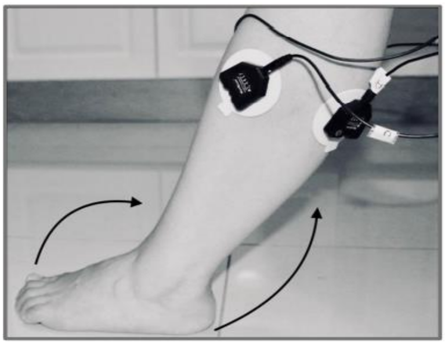


FIG.. 6. FLEXION OF THE KNEES WITH FLEXION DORSIFLEXION C.

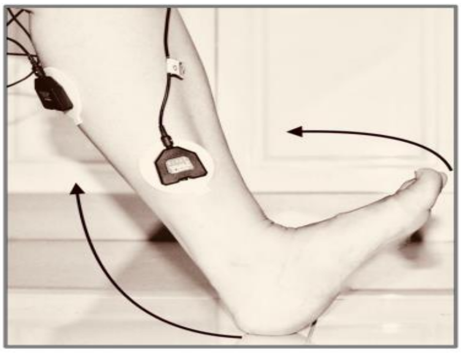


FIG. 7. FLEXION DORSIFLEXION WITH KNEE EXTENSION D.

Received 17/December/2021; Accepted 08/March/2022

DOI: https://doi.org/10.33103/uot.ijccce.22.4.11

- **Third test** While seated in a chair, flex or extension (bend and straighten) his or her lower limb (foot and leg) at the knee joint using Extension Plantarflexion and Flexion Dorsiflexion, as shown in *Fig. 8 and 9*.

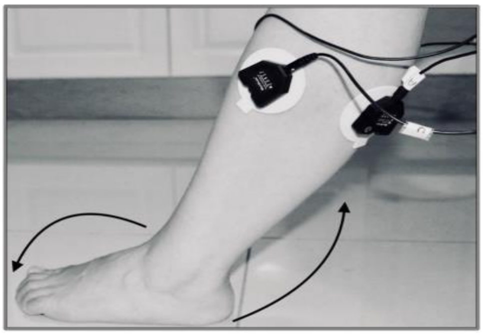


FIG. 8. FLEXION OF THE KNEES WITH FLEXION DORSIFLEXION E.

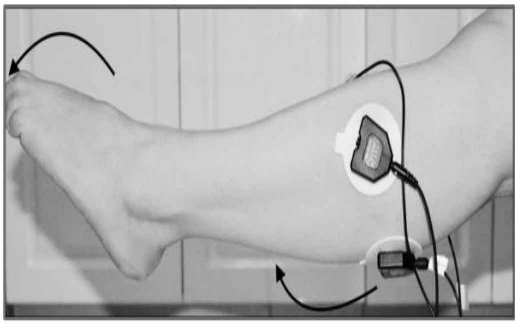


FIG. 9. FLEXION DORSIFLEXION WITH KNEE EXTENSION F.

The surface electromyography (EMG) signals for the Metro-Hospital dataset (TA) were collected from 13 patients' Gastrocnemius, Rectus Femurs, Tibias Anterior, and Soule. Surface electromyography (EMG) signals from ten persons are included in the OpenSim dataset for CG. Signals from the Biceps Femurs Long Head, Gastrocnemius, Medial Hamstrings, Tibias Anterior, and Rectus Femurs are recorded using EMG(TA).

V. PROPOSED SYSTEM IMPLEMENTATION

This work is implemented by a computer with Intel (R) Core i7- 6600U CPU @2.60 GHz, 16 GB Random Access Memory (RAM), a Hard disk with a capacity of 500GB, Intel(R) HD Graphics 520 MB. The following are the two phases of the system's implementation:

A. Training

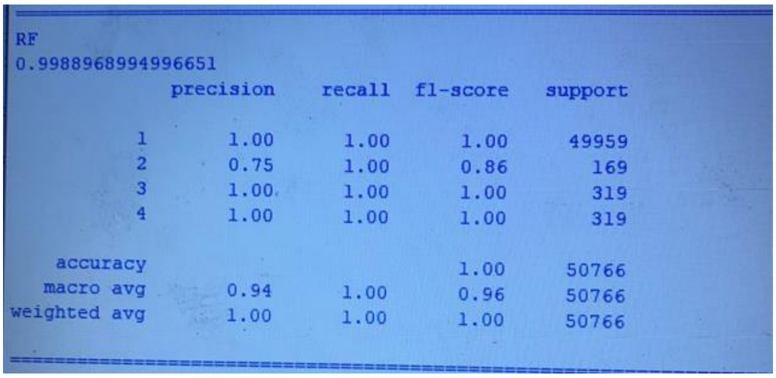
The suggested method's initial stage is to train the dataset using Holdout splitting, in which the bulk of data (70%) goes into the training phase and the remaining data (30%) goes into the testing phase.

B. Testing

The suggested system's testing phase is the second stage. The remaining data (30%) will be handled in the same way as the training data, as previously stated.

Fig. 10, 11, 12, and 13 illustrate the implementation of the system using four classifiers as follows:

Random Forest (RF)



RF
0.9988968994996651

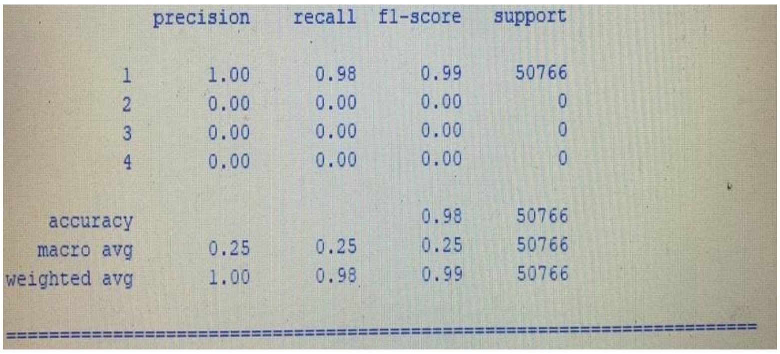
	precision	recall	f1-score	support
1	1.00	1.00	1.00	49959
2	0.75	1.00	0.86	169
3	1.00	1.00	1.00	319
4	1.00	1.00	1.00	319
accuracy			1.00	50766
macro avg	0.94	1.00	0.96	50766
weighted avg	1.00	1.00	1.00	50766

FIG. 10. EVALUATION METRICS OF RANDOM FOREST CLASSIFIER.

Received 17/December/2021; Accepted 08/March/2022

DOI: https://doi.org/10.33103/uot.ijccce.22.4.11

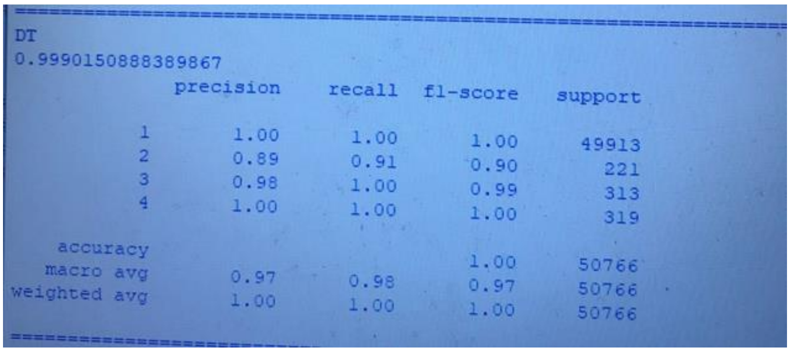
Logistic Regression (LR)



	precision	recall	f1-score	support
1	1.00	0.98	0.99	50766
2	0.00	0.00	0.00	0
3	0.00	0.00	0.00	0
4	0.00	0.00	0.00	0
accuracy			0.98	50766
macro avg	0.25	0.25	0.25	50766
weighted avg	1.00	0.98	0.99	50766

FIG. 11. EVALUATION METRICS OF LOGISTIC REGRESSION CLASSIFIER.

Decision Tree (DT)

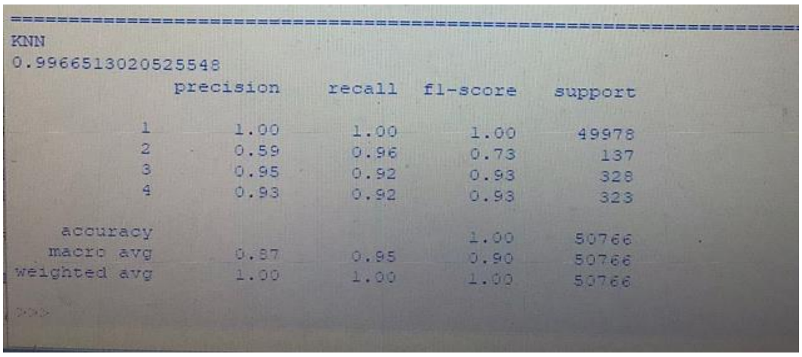


DT
0.9990150888389867

	precision	recall	f1-score	support
1	1.00	1.00	1.00	49913
2	0.89	0.91	0.90	221
3	0.98	1.00	0.99	313
4	1.00	1.00	1.00	319
accuracy			1.00	50766
macro avg	0.97	0.98	0.97	50766
weighted avg	1.00	1.00	1.00	50766

FIG. 12. EVALUATION METRICS OF DECISION TREE CLASSIFIER.

K-Nearest Neighbors (KNN)



KNN
0.9966513020525548

	precision	recall	f1-score	support
1	1.00	1.00	1.00	49978
2	0.59	0.96	0.73	137
3	0.95	0.92	0.93	328
4	0.93	0.92	0.93	323
accuracy			1.00	50766
macro avg	0.87	0.95	0.90	50766
weighted avg	1.00	1.00	1.00	50766

>>>

FIG. 13. EVALUATION METRICS OF K-NEAREST NEIGHBORS (KNN) CLASSIFIER.

VI. PERFORMANCE METRICS

The performance of the recommended machine learning model may be measured using a variety of criteria. The following are the metrics [22]:

A. Accuracy

It is calculated from TP and TN and represents how well the model predicts the classes. It’s calculated like this: (TP + TN)/total samples:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{9}$$

Received 17/December/2021; Accepted 08/March/2022

DOI: https://doi.org/10.33103/uot.ijccce.22.4.11

B. Precision

The percentage of all samples expected to be from class I that is really from class I is calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (10)$$

When the amount of data by class is unbalanced, accuracy alone isn't always sufficient to assess the model's effectiveness. If the model predicts everything as class 0, and there are 99 cases of class 0 and 1 example of class 1, the accuracy is 99%. but when precision is taken into consideration, the model performs badly. The precision of class 0 will be zero in this case.

C. Recall

Sensitivity is another name for it. The proportion of all samples predicted to be class I that is really class I This is how it's defined:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (11)$$

As a result, the prior example's class 0 will also have zero recall. The goal of our model is to maximize both precision and recall.

D. F-score

It's a mix of recollection and accuracy. The harmonic mean is what it's called. This is how it is defined:

$$\text{F}_1 = 2 * \frac{\text{precision}*\text{recall}}{\text{precision+recall}} \quad (12)$$

Where,

TP = true positives: the number of positive examples anticipated that are actually positive.
FP = false positives: the number of examples predicted positively but turned out to be negative.
TN = true negatives: the number of expected negative examples that are truly negative.
FN = false negatives: the number of examples that were expected to be negative but turned out to be positive.

VII. EXPERIENTIAL RESULTS

In this section, the results obtained through the use of machine learning algorithms to classify the medical images that were collected from a Metro-Rehab hospital are presented. These images include leg images in different positions. The purpose of the work is to help diagnose the disease and then find solutions in a better and faster way. In order to rehabilitate the leg in many pathological cases, and achieve this, four classifiers were used, and the results as shown in Table I and *Fig. 14* the classifiers RF, DT, and KNN are the best with the highest accuracy, precision, and recall. While the LR is considered the worst one of these classifiers.

TABLE I. MACHINE LEARNING CLASSIFIER'S OUTCOMES

	RF	DT	LR	KNN
Accuracy	100%	100%	98%	100%
Precision	100%	100%	100%	100%
Recall	100%	100%	98%	100%
F-score	100%	100%	99%	100%

Received 17/December/2021; Accepted 08/March/2022

DOI: https://doi.org/10.33103/uot.ijccce.22.4.11

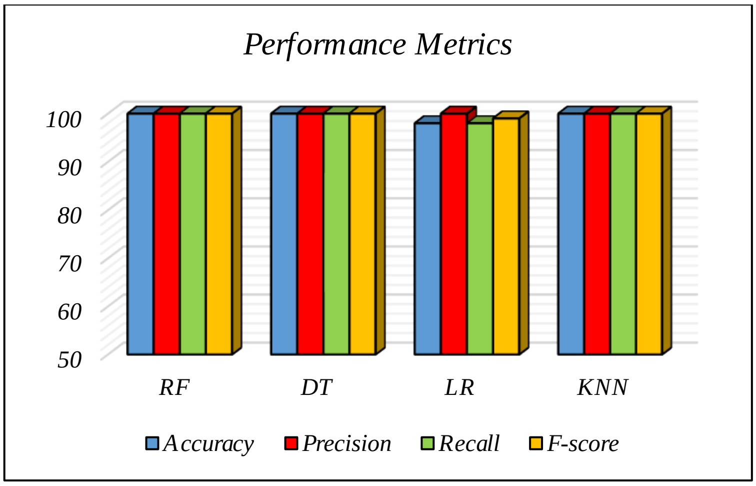


FIG.14. CHART OF PERFORMANCE METRICS OF MACHINE LEARNING CLASSIFIERS.

Now a comparison is made, as shown in Table II, between the research results and the results of previous studies that also relied on machine learning algorithms as classifiers.

TABLE II. EVALUATION METRICS OF STUDIES

	Year	Technique	Accuracy
[7]	2016	Extreme Learning Machine (ELM)	96%
		Support Vector Machine (SVM)	93%
		Neural Network (NN)	82%
[8]	2017	Naïve Bayes (NB)	99%
		Neural Network (NN)	97%
		Logistic Regression (LR)	95%
[9]	2019	Neural Network (NN)	93.5%
		Naïve Bayes (NB)	82.1%
		Decision tree (DT)	84.3%
		K- Nearest Neighbor (KNN)	77.9%
		Support Vector Machine (SVM)	85%
		Random Forests (RF)	83.6%
[10]	2019	Support Vector Machine (SVM)	83.6 %
		Random Forests (RF)	80.4%
		AdaBoost	83.6%
		Bagging	68.4%
		Naïve Bayes (NB)	67.7%
[11]	2020	K- Nearest Neighbor (KNN)	75.8%
		Random Forests (RF)	77.7%
This work	2021	Naïve Bayes (NB)	100%
		Random Forests (RF)	100%
		Decision tree (DT)	100%
		Logistic Regression (LR)	98%
		K- Nearest Neighbor (KNN)	100%

VIII. CONCLUSIONS

The goal of this work was to create and evaluate an automated gait analysis system that employed lower-body motion data and machine learning techniques to differentiate between healthy and sick patients. The detection and diagnosis of drop foot are extremely accurate because of the employment of five machine learning algorithms as classifiers. The results show that the machine learning approaches used are very accurate, with Random Forest (RF), Decision Tree (DT), and k-nearest

Received 17/December/2021; Accepted 08/March/2022

 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.22.4.11

neighbors (KNN) classifier accuracy exceeding 100%, and Logistic Regression classifier accuracy exceeding 98%. The explanation for the high accuracy of the system is the preprocessing method and the robustness of the method used to extract the features. As the data is divided into four classes, each class includes a certain number of attributes, and data that contains noise is ignored, leaving only useful data. Only this data will be input into the classification stage.

REFERENCES

[1] J. Delaram, P. James, L. Tracey, T. Larisa, M. Amirhosein and S. Saeid, “A Review on Accelerometry Based Gait Analysis and Emerging Clinical Applications”. IEEE Reviews in Biomedical Engineering, Vol. 11, pp. 177 - 194, 2018.

[2] S. Khandelwal and N. Wickstrm, “Evaluation of the performance of accelerometer-based gait event detection algorithms in different real-world scenarios using the MAREA gait database,” Gait Posture., vol. 51, pp. 84–90, 2017.

[3] F. Joana, S. Cristina, M. Juan, “Automatic recognition of gait patterns in human motor disorders using machine learning: A review”. Medical Engineering & Physics, 2018.

[4] P. Amorim, J. Paulo, P. Silva, P. Peixoto, M. Castelo-Branco, H. Martins H, “Machine Learning Applied to Low Back Pain Rehabilitation – A Systematic Review”. International Journal of Digital Health, Vol.1, Issue:1, pp. 1–14, 2021.

[5] C. Tack, “Artificial intelligence and machine learning | applications in musculoskeletal physiotherapy (in eng)”. Musculoskelet Sci Pract. Vol. 39, pp. 164–169, Feb 2019.

[6] C. Kokkotis, S. Moustakidis, E. Papageorgiou, G. Giakas and D. Tsaopoulos, “Machine learning in knee osteoarthritis: A review”. Osteoarthritis and Cartilage Open, Vol. 2, Issue: 3, 2020.

[7] A. Sahar, A. Tanvir, Adel. Al. Jumaily, “Extreme Learning Machine based sEMG for Drop-Foot after Stroke Detection”. IEEE 2016 Sixth International Conference on Information Science and Technology (ICIST) - Dalian, China, 6-8 May 2016.

[8] V. Jebakumari, D. Shanthi, S. Sridevi, P. Meha, "Performance evaluation of various classification algorithms for the diagnosis of Parkinson's disease". IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS) - India, 23- 25 March 2017.

[9] Y. Zhang, Y. Ma, “Application of supervised machine learning algorithms in the classification of sagittal gait patterns of cerebral palsy children with spastic diplegia”. © Elsevier, Vol. 106, pp. 33-39, 2019.

[10] E. Rastegari, S. Azizian, H. Ali, “Machine Learning and Similarity Network Approaches to Support Automatic Classification of Parkinson’s Diseases Using Accelerometer-based Gait Analysis”. Proceedings of the 52nd Hawaii International Conference on System Sciences, 2019.

[11] E. Mesquita, A. Rodrigues, A. Andrade, and M. Vieira, “Classification of Lower Limb Amputees Gait Using Machine Learning Algorithms”. © Springer Nature Switzerland, pp. 950–954, 2020.

[12] S. Sharma et al., "An Improved Random Forest Algorithm for Predicting the COVID-19 Pandemic Patient Health", Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, No. 1, pp: 67 - 75, 2021.

[13] J. Engelen and H. Hoos, "A survey on semi-supervised learning", Springer, Machine Learning, Vol. 109, pp: 373–440, 2020.

[14] E. Boateng and D. Abaye, “A Review of the Logistic Regression Model with Emphasis on Medical Research”. Journal of Data Analysis and Information Processing, Vol. 7, pp. 190-207, 2019.

[15] M. Mahdizah and E. Zamanzade “Efficient Body Fat Estimation Using Multistage Pair Ranked Set Sampling”. Statistical Methods in Medical Research, Vol. 28, pp. 223-234, 2019.

[16] E. Boateng, S. Bosson-Amedenu, E. Nortey, and D. Abaye, "Non-Medical Determinants of Caesarean Deliveries in Ghana: A Logistic Regression Approach". Open Journal of Applied Sciences, Vol. 9, pp. 492-505, 2019.

[17] B. Jijo, A. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning”. Journal of Applied Science and Technology Trends Vol. 02, No. 01, pp. 20 – 28, 2021.

[18] S. Patil and U. Kulkarni, “Accuracy Prediction for Distributed Decision Tree using Machine Learning approach,” International Conference on Trends in Electronics and Informatics (ICOEI), pp. 1365–1371, Apr. 2019.

[19] F. Yang, “An Extended Idea about Decision Trees,” International Conference on Computational Science and Computational Intelligence (CSCI), pp. 349–354, Dec. 2019.

[20] A. Zaoui, C. Denis and M. Hebiri, “Regression with reject option and application to KNN”. In 34th Conference on Neural Information Processing Systems (NeurIPS), Vancouver, Canada, 2020.

[21] M. Arowolo, M. Adebiyi, A. Adebiyi, and O. Okesola, “PCA Model for RNA-Seq Malaria Vector Data Classification Using KNN and Decision Tree Algorithm”. International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS), pp. 1–8, Mar 2020.

[22] F. Al-Turjman and I. Baali, “Machine learning for wearable IoT-based applications: A survey”. © John Wiley & Sons, Ltd, 17 April 2019.

Received 17/December/2021; Accepted 08/March/2022