ISSN: 0974-5823 Vol. 6 No. 3 October-December, 2021

International Journal of Mechanical Engineering

A Non-Invasive Early Diagnosis of Osteoarthritis Using Random Forest and ANN

Anisha.C.D
Research Scholar, Dept of CSE, PSG College of Technology, Coimbatore, India.
ani.c.dass@gmail.com

Arulanand.N
Professor, Dept of CSE, PSG college of Technology, Coimbatore, India.
naa.cse@psgtech.ac.in

Abstract

Osteoarthritis or degenerative arthritis is one of the joint disorders which affects the cartilage. This type of joint damage, if left untreated in the early stage will lead to degeneration of joints. There is a need for a non-invasive early diagnosis of osteoarthritis which is made possible through the usage of Machine Learning Techniques. Surface Electromyography signal (sEMG) is a non-invasive manner of collecting muscle data from the surface of the skin, which is cheaper and easier to capture from the individuals. The proposed framework uses the dataset retrieved from the UCI repository. It consists of sEMG data of three movements namely Gait, Flexion and Extension procured from the lower limb of the normal and abnormal subjects. The time series data can't give meaningful information for classification, so the data is sampled and the sampled sEMG signal data is pre-processed using Butterworth Band Pass Filter which is then fed to feature extraction stage wherein Time Domain (TD) features are procured. The extracted features are fed to the Random Forest classifier and Artificial Neural Network (ANN) classifier. The evaluation metrics used to analyse the prediction of the classifier are Confusion Matrix, Accuracy, Precision, Recall and F1 Score.

Keywords— Osteoarthritis, Butterworth band pass, Time Domain, Random Forest, ANN

I INTRODUCTION

Osteoarthritis or degenerative arthritis is a progressive disorder wherein it affects the cartilage of the joints.[1][3] This disorder affects the lower limb functionality.[5]. The prevailing Osteoarthritis diagnosis system is invasive and costly. Therefore, it is necessary to use surface Electromyography Signals (sEMG) which is a non-invasive method for collecting data from individuals. sEMG represents the electrical potentials recorded from the individual while performing movements.[15]. The movements considered includes Gait which refers to walking, extension of the leg from the sitting position and flexion of leg up from the standing position.

Signal Processing is the vital step in sEMG data classification. The filtering of signals removes the noise in the sEMG signal Machine Learning Algorithms are used for Classification in more efficient manner.

The organization of the paper is as follows: Section II presents the related works which provides the techniques used in the paper, Section III presents the methodologies used in the proposed system and Section IV provides the results analysis and Section V presents the Conclusion and the future work.

II RELATED WORKS

In [1] Jean de Dieu Uwisengeyimana proposes a Knee pathology diagnosis using Support Vector Machine (SVM) and deep learning Neural Network and the sEMG is processed using Overlapped windowing technique. Least-Squares Kernel, Linear Discriminant Analysis (LDA) are used for classification in [2] for the data which involves control subjects and Rheumatoid arthritis and hip osteoarthritis. Different Machine Learning and Deep learning algorithms were compared to classify the osteoarthritis individual data from healthy subject data in [3] and a discussion on quantum perspective is also provided. Xin chen et al in [5] presents Entropy based Measures are used to distinguish Knee Osteoarthritis (KOA) individuals from control subject. The three types of Entropy measures considered are Approximate entropy, Sample Entropy and Fuzzy Entropy.

Jihye Lim et al, in [16] has presented the deep learning method to predict the occurrence of the Osteoarthritis using the statistical information of the patients. Principal Component Analysis (PCA) is used for generation of features.

Multi-Layer Perceptron, a type of Artificial Neural Network (ANN) has been used for the assessment of the Knee Injuries from the surface Electromyography signals and goniometric signals wherein these signals sent to the processing

stage of time-frequency space obtained from the spectrogram and wavelet transform.[17].

Nima Befrui et al in [18] has presented a Vibro-arthrography based diagnosis for Knee Osteoarthritis. The normalized features are extracted from vibratory signals which is processed using segmentation stage, normalizing stage using Hann Window. The Linear Support Vector Machine (SVM) has been used for classification with input as normalized feature vectors obtained from the raw vibratory signals.

A Multi Model based Knee Osteoarthritis has been presented in [19], The X ray images has been processed using Convolutional Neural Network (CNN), clinical information and the baseline characteristics are fused using Gradient Boosting Machine (GBM).

From the related works, it is evident that the osteoarthritis prediction was mainly performed using the sEMG time series data and X ray Images data. The prominent classifiers used were Artificial Neural Network (ANN), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) for sEMG data and Convolutional Neural Network (CNN) for X-ray Images.

III. METHODOLOGIES

Figure 1 depicts the block diagram of the proposed framework with the flow of the process. The proposed system uses Butterworth Bandpass filter to the input signal wherein it provides an averaged smooth signal and performs better than the filters used in the existing system [1]. Another improvisation incorporated to the existing system is the construction of the hyperparameter space for the Random Forest Classifier, identifying and implementing the exact preprocessing technique essential for each classifier in the analysis which impact in the accurate prediction.

A. Dataset Description

The sEMG dataset is retrieved from the publicly available UCI repository. The dataset consists of sEMG recordings of 11 Healthy subjects and 11 Abnormal subjects who exhibit osteoarthritis. The five muscle positions considered for the collection are Recto Femoral, Biceps Femoral, Vasto Medial, EMG Semitendinoso, Flexo-Extension of the lower limb. The movements performed by the subjects are gait which refers to the walking movement, leg extension from a sitting position, and flexion of the leg up from the standing position. Figure 2,3,4 represents the above specified movements. The sources of the figures are specified in [21][22][23].

B. Data Processing- Splitting of Data

Each sEMG data of an individual is unique. The dataset considered for the process involves one Normal subject sEMG data and one Abnormal subject sEMG data. Table 1 represents the input given to this splitting process and the output obtained from this process. The data is split into 100 samples for the range of instances from 5000 and 200 samples for the range greater than 10000 instances.

C. Signal Processing – Filtering and Feature Extraction

The sEMG data obtained from splitting process is filtered using Butterworth band pass filter as Butterworth provides a thin frequency signal and band pass filter provides the average of low pass and high pass filter.[13] The filtered signal is then sent to the feature Extraction stage wherein 17 Time Domain Features are extracted.[11][12]

D. SPLITTING OF TRAIN AND TEST SET

The extracted features are split into 80% training set and 20% testing set. Table 2 represents the Training and Testing set split.

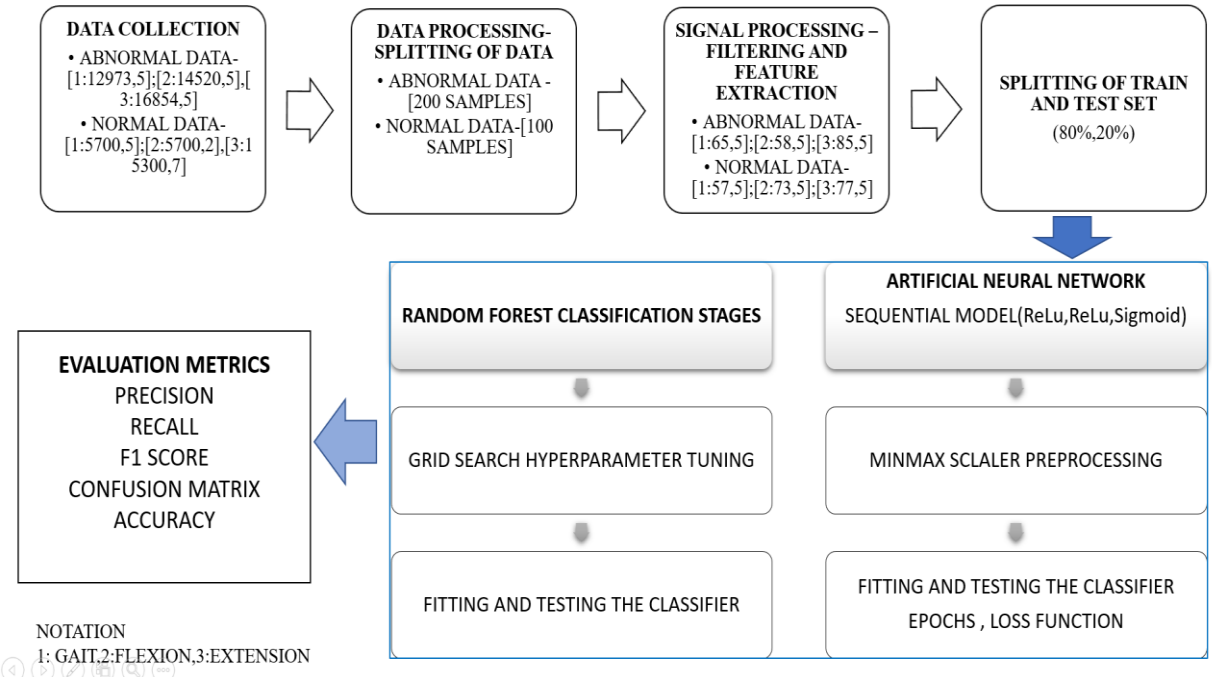

Figure 1 – Block Diagram of the Proposed Framework

Table 1 - INPUT TO SPLITTING STAGE

MOVEMENT	ACTUAL INSTANCES IN DATSET	SAMPLES SPLIT	AFTER SPLIT OUTPUT INSTANCES
GAIT	{Normal :(5700,5); Abnormal:(12973,5)}	{Normal: 100 Abnormal :200}	Normal:57 Abnormal:65 {122,5}
FLEXION OF LEG UP	{Normal :(14520,5); Abnormal:(12973,5)}	{Normal: 200 Abnormal :200}	Normal:73 Abnormal:58 {131,5}
SITTING-EXTENSION	{Normal :(11403,5); Abnormal:(16854,5)}	{Normal: 200 Abnormal :200}	Normal:77 Abnormal:85 {162,5}

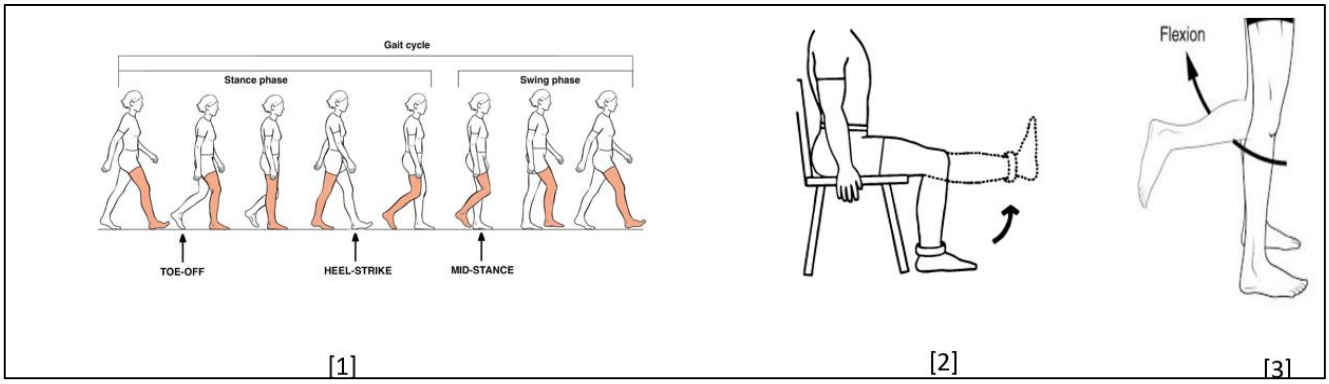

Figure 2 – Movements in the dataset [1] Gait(walking) [2] Extension from sitting position [3] Flexion of leg up from standing position.

Table 2- TRAINING AND TESTING SET SPLIT

MOVEMENT	TRAINING SET INSTANCES	TESTING SET INSTANCES
GAIT	97	25
FLEXION OF LEG UP	104	27
SITTING-EXTENSION	129	33

E. Classification

The classification algorithm considered are Random Forest and Artificial Neural Network (ANN).

1. Random Forest Classifier

i. Working Principle of Random Forest Classifier

Random Forest is an ensemble algorithm wherein it uses majority voting technique. [8]. The parameter bootstrap in the Random Forest Classifier specifies the sampling of data, if the bootstrap is set to True then the dataset is split into samples and send to each decision tree, if it set to false then the whole dataset is sent to each decision tree. The number of decision tree is specified using number of estimators. Each decision tree in the forest is trained using bootstrap approach and Each decision tree provides prediction which is then combined using majority voting approach.

ii. Hyperparameter Tuning – Grid Search Method

The hyperparameters are the prominent parameters for the classifier which directly influences the performance. They are tuned using Grid Search Method to increase its performance. It forms all possible combinations from the hyperparameter space and finds the best combination.[6][7].

The hyperparameters of the Random Forest are n_estimators and Bootstrap.

a) n_estimators: corresponds to the number of decision tree in the forest, the value for n_estimators is provided based on the number of features in the dataset.
b) Bootstrap: It is Boolean value, True indicates that the dataset sample is split and send to each decision tree otherwise the whole dataset is sent to each decision tree.

The hyperparameters Space of the Random Forest is presented in the table.

TABLE 3 – HYPERPARAMETER SPACE – FOR ALL THREE MOVEMENT CLASSES

MOVEMENT	HYPERPARAMETER INPUT SPACE	HYPERPARAMETERS OUTPUT
GAIT	n-estimators= {1,33}, Bootstrap= {True, False}	n_estimators=31, Bootstrap=True

FLEXION OF LEG UP	n-estimators= {1,33}, Bootstrap= {True, False}	n_estimators=1, Bootstrap=True
SITTING-EXTENSION	n-estimators= {1,33}, Bootstrap= {True, False}	n_estimators=3, Bootstrap=True

iii. Fitting and Testing the Classifier.

The classifier is fitted to the training set and the trained classifier is tested using the testing set.

2. Artificial Neural Network (ANN)

i. Working Principle of ANN

Artificial Neural Network consists of an input layer, the hidden layers and the output layer. The activation function used for the hidden layers are ReLu, for two layers and Sigmoid for one layer.[9] The number of epochs parameter specifies the number of iterations required to train the classifier. The number of epochs is set to 2000 by trial and error method.

ii. Min Max Scaler Preprocessing

The dataset is preprocessed using min max scaler for providing better results from ANN. The dataset is preprocessed by standardizing the values in the dataset to the range of [0,1]. [10]

iii. Fitting and Testing the Classifier

The classifier is fitted to the training set by execution of the specified number of epochs and tested using the testing set wherein the prediction is done for the unknown data.

F. Evaluation Metrics

1. Confusion Matrix

The confusion matrix is 2 X 2 Matrix as it is a binary classification. The 0 indicates the presence of Osteoarthritis and 1 indicates the healthy status. The confusion Matrix is constructed based on True Positive (TN), True Negative (TN), False Positive (FP) and False Negative (FN).[14]

2. Accuracy:

Accuracy is computed using the formula:

Accuracy = Total number of correctly classified instances / Total number of instances.

3. Precision:

It is the number of positive instances correctly classified divided by the total number of positive instances in the test set [20].

4. Recall:

It is the number of positive instances correctly classified divided by the total number of predicted instances (True Positive + False Negative)

5. F1 Score:

It is weighted average of Precision and Recall.

IV RESULT ANALYSIS

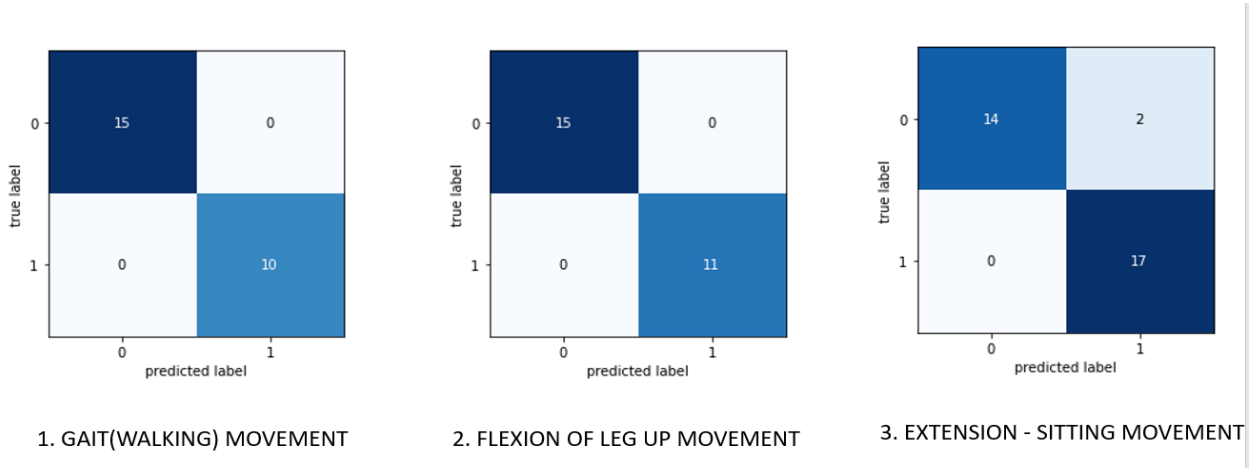

Figure 3 – Confusion Matrix of Random Forest Classifier

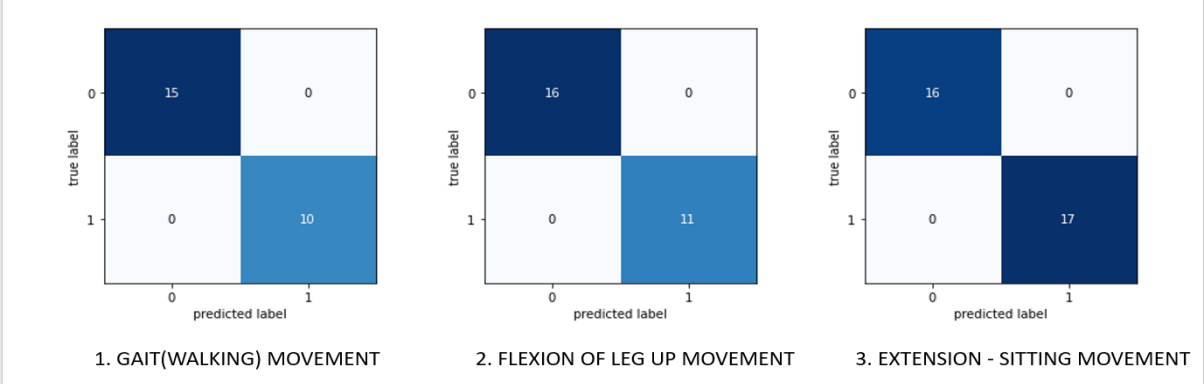

Figure 4 – Confusion Matrix of Artificial Neural Network (ANN)

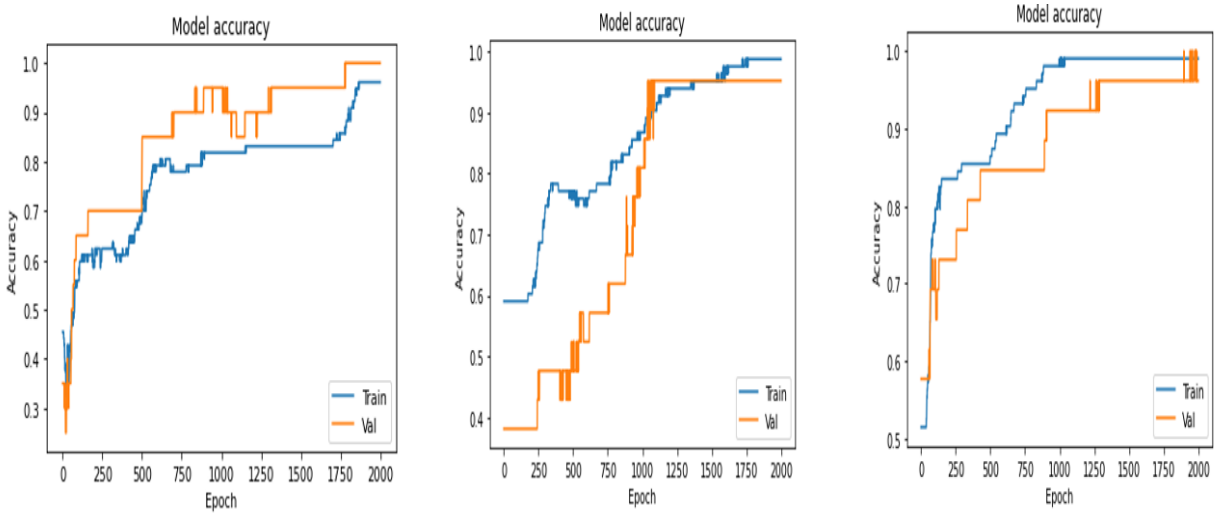

Figure 5 – Accuracy Graph of Epochs in ANN for three movements

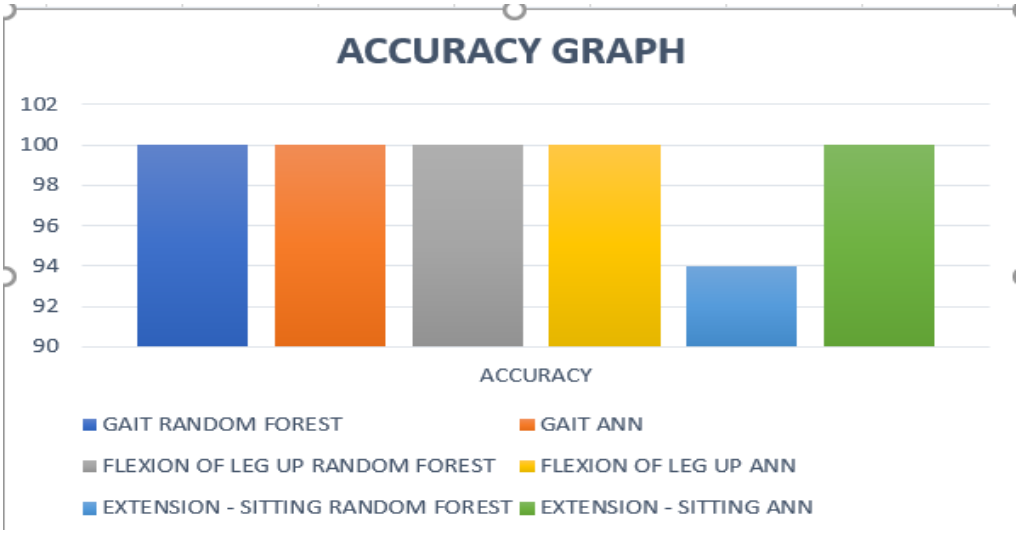

Figure 6- Accuracy Graph of Random Forest and ANN for all movements.

TABLE 5 – ACCURACY, PRECISION, RECALL AND F1 SCORE

Movement	Classifier	Accuracy	Precision	Recall	F1 Score
Gait	Random Forest	1.00	1.00	1.00	1.00
	ANN	1.00	1.00	1.00	1.00
FLEXION OF LEG UP	Random Forest	1.00	1.00	1.00	1.00
	ANN	1.00	1.00	1.00	1.00
SITTING-EXTENSION	Random Forest	0.94	0.95	0.94	0.94
	ANN	1.00	1.00	1.00	1.00

A. Inferences from the results

Figure 3,4 represents the Confusion Matrix of Random Forest and ANN. Figure 5 presents the accuracy obtained in each epoch for ANN. Figure 6 presents the accuracy obtained by Random Forest and ANN for all three movements. Table 5 presents the Evaluation results specified using Accuracy, Precision, Recall and F1 Score.

1. The Random Forest classifier correctly distinguishes Abnormal subjects from healthy subjects for the movement Gait and Flexion. It produces **100%** accuracy for Gait and Flexion movement and **94%** accuracy.
2. The ANN classifier correctly distinguishes Abnormal subjects from healthy subjects for all the three movements. It produces **100%** accuracy for three movements.
3. The Epochs is set to 2000 which is set by trial and error method.

i. Initially when Epochs was kept at 1000 the accuracy obtained was 87.5% but when Epochs increased to 2000 the accuracy obtained was 100%.
ii. Epochs which corresponds to the number of iterations plays a vital role in improving accuracy.

V CONCLUSION

The proposed Framework provided movement-based analysis for the discrimination of osteoarthritis patients from healthy subjects. The movement considered for the classification of Osteoarthritis are Gait, Flexion and Extension. The Random Forest and ANN both provided 100 % accuracy for the movement Gait and Flexion. ANN provided a higher accuracy of 100 % than the Random Forest for the movement Extension. The appropriate pre-processing stage are Hyperparameters Tuning of Random Forest and Min Max Scaling of Random Forest which are based on the nature of the classifier is one of the reasons for yielding good accurate results. The proposed Framework performs better than the existing system [1]. The results obtained prove that the classifiers are efficient in early diagnosis of the osteoarthritis. The future work is to test the algorithms on the real time data which is to be collected from the lower limb and to automate the complete process.

REFERENCES

[1] Jean de Dieu Uwisengeyimana et al, "Diagnosing Knee Arthritis Using Artificial Neural Network and Deep Learning", Biomedical Statistics and Informatics, 29 March 2017.

[2] Sumitra S Nair et al, The Application of Machine Learning Algorithms to the Analysis of Electromyographic Patterns of Arthritis Patients, IEEE Transaction on Neural Systems and Rehabilitation Engineering, Vol 18, No2, April 2010.

[3] Serafeim Moustakidis et al, "Application of Machine Intelligence for Osteoarthritis classification: A Classical Implementation and Quantum Perspective, Quantum Machine Intelligence,29 September 2019.

[4] C. Kokkotis et al, "Machine Learning in Knee Osteoarthritis: A review", Osteoarthritis and Cartilage Open, 17 April 2020.

[5] Xin Chen et al, "Entropy Based Surface Electromyogram Feature Extraction for Knee Osteoarthritis Classification, IEEE Transactions and Journal, 2016.

[6] Pedregosa et al., "Scikit-learn: Machine Learning in Python", JMLR 12, pp. 2825-2830, 2011.

[7] Buitinck et al., "API design for machine learning software: experiences from the scikit-learn project", 2013.

[8] Gerard Biau et al, "Analysis of a Random Forests Model", Journal of Machine Learning Research 13 (2012).

[9] Vidushi Sharma et al, "A Comprehensive Study of Artificial Neural Networks", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10, October 2012

[10] C. Saranya et al, "A Study on Normalization Techniques for Privacy Preserving Data Mining.", International Journal of Engineering and Technology (IJET),2013

[11] J. Too, A. R. Abdullah, N. Mohd Saad, and W. Tee, "EMG Feature Selection and Classification Using a Pbest-Guide Binary Particle Swarm Optimization," Computation, vol. 7, no. 1, 2019.

[12] J. Too, A. R. Abdullah, and N. Mohd Saad, "Classification of Hand Movements based on Discrete Wavelet Transform and Enhanced Feature Extraction," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 6, 2019.

[13] Dipak C.Vaghela et al, "Design, Simulation and Development of Bandpass ", International Journal of Engineering Development and Research, Volume 3, Issue 2 1015

[14] Hossin, Mohammad & M.N, Sulaiman., "A Review on Evaluation Metrics for Data Classification Evaluations.", International Journal of Data Mining & Knowledge Management Process,2015.

[15] Jiajia Wu, "sEMG Signal Processing Methods: A Review", IOP Conf. Series: Journal of Physics: Conf. Series 1237 (2019)

[16] Jihye Lim et al, "A Deep Neural Network-Based Method for Early Detection of Osteoarthritis Using Statistical Data", International Journal of Environmental Research and Public Health — Open Access Journal,2019

[17] M. Herrera-González, G. Martínez-Hernández, J. Rodríguez-Sotelo and O. Avilés-Sánchez, "Knee functional state classification using surface electromyographic and goniometric signals by means artificial neural networks", Ing. Univ., vol. 19, no. 1, pp. 51-66, Ene., Jun., 2015.

[18] Nima Befrui et al, "Vibroarthrography for Early Detection of Knee Osteoarthritis Using Normalized Frequency Features", Medical & Biological Engineering & Computing, February 2018.

[19] Tiulpin, A., et al, "Multimodal Machine Learning-based Knee Osteoarthritis Progression Prediction from Plain Radiographs and Clinical Data.", Sci Rep 9, 20038 (2019). https://doi.org/10.1038/s41598-019-56527-3.

[20] Goutte C. et al, "A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation." In: Losada D.E., Fernández-Luna J.M. (eds) Advances in Information Retrieval. ECIR 2005. Lecture Notes in Computer Science, vol 3408. Springer, Berlin, Heidelberg.2005.

[21] Nazia Gillani, "Human Gait Phase Detection using Convolution Neural Network-Based Prediction Engine", innovate FPGA.

[22] "DVT: Guidelines for Activity and Exercise", North American Thrombosis Forum.

[23] Tonye Ogele CNX, "Anatomy and Physiology", OpenStax.