Forecasting gait freezing event in Parkinson's patients utilizing machine learning approach from accelerometer signals

Bal Gopal Mishra[1], Animesh Sarangi[2], Satyabhama Dash[3]
[1]*Post Graduate Scholar, EIE Department, Odisha University of Technology and Research, Odisha, India.*
[2]*Post Graduate Scholar, EIE Department, Odisha University of Technology and Research, Odisha, India.*
**Corresponding author: mishrabalgopal1996@gmail.com*
[3] *Assistant Professor, EIE Department, Odisha University of Technology and Research, Odisha, India.*

Abstract

Neurological illnesses are one of the most common medical conditions affecting human races and societies worldwide. Parkinson's disease is a neurological ailment caused by absence of dopamine in the human brain and has an impact on the afflicted person's daily routine. Gait freezing event is the most concerning symptom of Parkinson's disease, and it affects around half of people with severe Parkinson's. Machine learning methods are used in this study to detect and forecast gait freezing events. Two hundred thirty seven gait freezing instances from eight patients were collected from tri-axial accelerometer data set and used to train four machine learning classification models. After comparing different performance measures of the four classification models it was found that the Random forest classification model was the most suitable one for predicting gait freezing events in Parkinson disease as it had the best accuracy, sensitivity ,selectivity and least error among the four models.

Keywords: *Parkinson's disease , Gait freezing , machine learning , tri-axial accelerometer , Random forest classification model.*

1. Introduction

Parkinson's disease is a worsening neurological disorder marked by uncontrollable shaking , stiffness, hypotonia, dyskinesia and gait freezing.[1]. Other mental symptoms like melancholy, stress, and anxiety may also impact Parkinson patients [2]. It often first manifests in people between the ages of 50 and 60 [3]. Additionally, it affects males more often than women, while the causes are unclear [3].

One of the most crippling motor symptoms of Parkinson's disease patients is the gait freezing event, which is defined as "an abrupt, brief, and transitory incapacity to move the feet forward despite the desire to walk"[4]. It is mostly seen in patients in the advanced stage of Parkinson.[5].The gait freezing episode significantly hampers the routine lifestyle of the patients. It can lead to falls, fractures and frequent hospital admissions.[6].Consequently, this gives rise to social isolation and dependence on caretakers. This is described in flowchart given below:

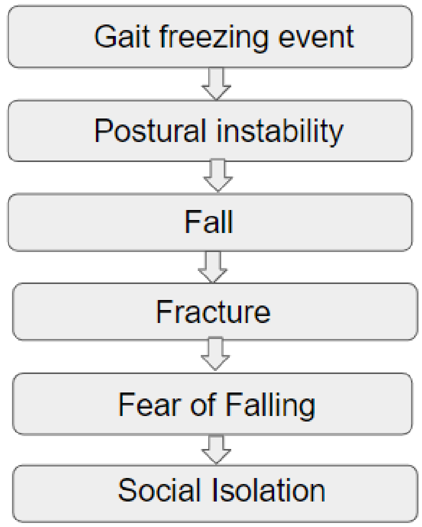

Figure .1 clinical impact of gait freezing event in Parkinson's patients

According to Hausdorff et al. gait freezing event does not completely stop the motion of feet, there is still shaking of legs, but the patient is unable to advance on walking.[7], Gait freezing episode is distinguished by a number of sub-categories, including start hesitation, turn hesitation, hesitation in confined spaces, destination hesitation, and hesitation in an open area [8].As gait problem is difficult to treat through drugs, in order to encourage the patient to continue walking, visual [9] and auditory cues can significantly help in reducing it. Predicting the gait freezing event and sending out audible cues can assist the patient to completely avoid it, which is a step in the right direction. There are, however, few studies that use machine learning algorithms and a tri-axial accelerometer to examine and forecast the change from walking to gait freezing. [10].

Our study focuses on forecasting gait freezing events in Parkinson disease patients by using four machine learning classification models namely K-Nearest Neighbor (KNN),Support Vector Machine(SVM),Decision tree and Random forest algorithm through tri-axial accelerometer data set of Parkinson patients. We then compare the performance measures like accuracy,sensitivity,specificity of the above classification models so that we can find the most suitable model to anticipate gait freezing event before it occurs in Parkinson patients. This will be of great aid to develop various innovative wearable devices and intelligent real-time health assistants in future which can provide different types of auditory and visual cues that will alleviate the gait freezing event and provide relief to Parkinson patients. Furthermore, this model can also be used for telemonitoring and diagnosis of patients suffering from Parkinson.

2. Methodology

Our approach for developing the various classification models is shown in the flowchart in fig 2 which consists of following steps : data collection , data pre-processing , feature extraction, dividing the data set into training and testing; then we use various machine learning algorithms to model the classifier and finally analyze the classifier using various performance measures like accuracy, sensitivity, selectivity, etc.

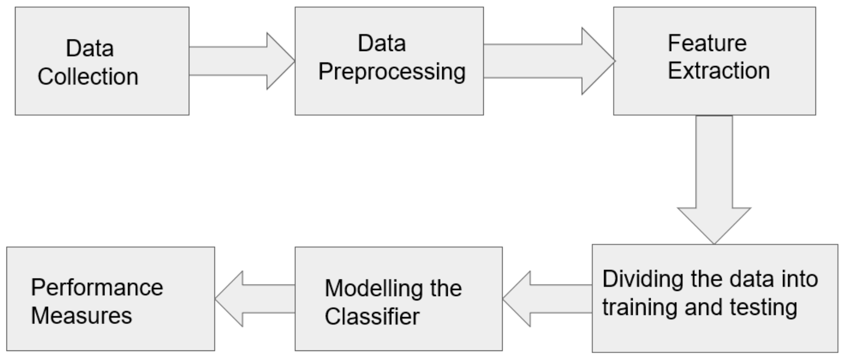

Figure .2 General methodology used in our study

2.1. Data Collection

The Daphnet Freezing of Gait data set, which was obtained from the UCI Machine Learning Repository, is used for this work [11]. Data of ten Parkinson patients are included in the data set; however, two individuals were not considered since they had no prior exposure to gait freezing episodes. Three wearable sensors that collect tri- axial accelerometer signals (sampled at 64 Hz) from the ankle, knee and hip were used to collect the data. Using the video records that have been examined by physiotherapists, the walking and gait freezing episode were identified.The dataset was captured in the lab with a focus on producing plenty of freeze occurrences. Users completed a variety of activities, such as walking in a straight line, making many turns while walking, and ultimately, a more realistic activity of routine tasks that included entering various rooms and doing routine chores including getting a cup of coffee, opening doors, etc. Eight hours and twenty minutes of data recording was collected in total where average duration of gait freezing episode was found to be around 8 seconds.

2.2 Data pre-processing

The data comprises of following indexes denoted by:

0 - not a part of our experiment.

1 - experiment when freezing does not occur

2 - freezing occurs

First the data with index “0” is removed as it is not part of our experiment. Due to the lack of knowledge on the typical length of time between walking and gait freezing events, it is crucial to use a variety of observation periods to pinpoint the most productive window of time. The main dataset now includes a five second window between walking and gait freezing episode. Each designated walking event's last five seconds were substituted by the window. The channel description across X,Y and Z axes from ankle, knee and hip sensors is plotted in the fig 3 below:

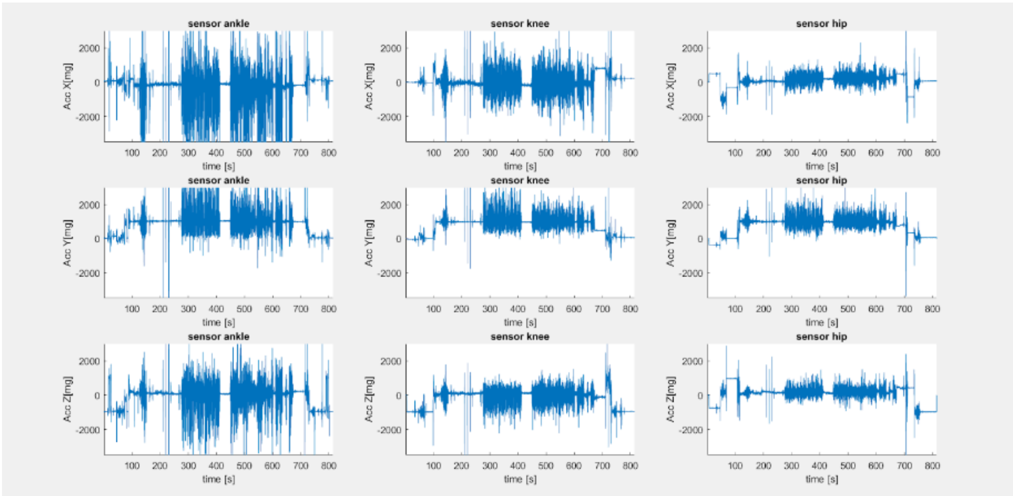

Figure .3 Channel description

2.3 Feature Extraction

Data from the channels were sequenced in accordance with the set of extracted features in order to give adequate pattern information for the various classification models taken into consideration in this research. Our methodology's concept is in line with [12]. Subsequently, the following seven time-domain and non-linear features were extracted from each channel employing five second windows:

2.3.1 Mean of the signal: It is usually the signal's average value denoted by equation (1) given below:

$$\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t \tag{1}$$

Here it is obtained by considering the X, Y-axis acceleration of the ankle along with Z-axis acceleration of the hip

2.3.2 Proportion Above the Mean: It is the number of observations within the window whose values are more than the average of the window denoted by equation(2) given below:

$$\text{RMS(velocity)} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (y_t)^2} \tag{2}$$

Here, the Y-axis acceleration of ankle was used to obtain this feature

2.3.3 Proportion Below the Mean: It is the number of observations within the window whose values are less than the average of the window . It is denoted by equation(3) given below:

$$P_{AM} = \frac{1}{n}\left|\left\{y_t > \bar{y}\right\}\right|, \forall t = 1, 2, \ldots\ldots n \tag{3}$$

Here,it is extracted taking into account the Y-axis acceleration of ankle

2.3.4. RMS Velocity of the signal: It is calculated as the square root of the mean of the square of the velocity signal ‘y’ in the time domain which is denoted by equation(4) given below:

$$P_{BM} = \frac{1}{n}\left|\left\{y_t < \overline{y}\right\}\right|, \forall t = 1, 2, \ldots\ldots n \tag{4}$$

Here,it is extracted taking into account the Y-axis acceleration of ankle

2.3.5 Sum of variations: It is calculated as the sum total of the successive differences of the window which is denoted by equation (5) given below:

$$\text{SoC} = \sum_{t=1}^{n} dif(y_t)^2 \tag{5}$$

Here, it is extracted taking into account the X-axis acceleration of the thigh.

2.3.6 Madogram dimension of the signal: It is a spatial function which is used to quantify the roughness (or smoothness) of time series data taking one as the power index. It is denoted by equation [6] given below:

$$\beta_p(t) = \frac{1}{2}E\left|y_h - y_{t+h}\right|^P, P = 1 \tag{6}$$

Here it is obtained by considering Y-axis acceleration of thigh and hip along with Z-axis acceleration of hip.

2.3.7 Variogram dimension of the signal: It is a spatial function which is used to quantify the roughness (or smoothness) of time series data taking two as the power index. It is denoted by equation [7] given below:

$$\beta_p(t) = \frac{1}{2}E\left|y_h - y_{t+h}\right|^P, P = 2 \tag{7}$$

Here, it is extracted taking into account the Y-axis acceleration of the hip.

2.4 Dividing the data into training and testing

In our study we varied the training data set from 20 % to 90% and testing data set from 80% to 10% so that we can get a comprehensive and holistic idea of the performance measures of each classification model. Furthermore, it would also help in better comparison of the models.

2.5 Modeling the classifier

In our study we employed various algorithms like K-Nearest Neighbor (KNN), Support Vector Machine(SVM),Decision Tree(DT),Random Forest(RF) to generate 4 classifiers that can predict gait freezing episodes.

KNN is a similarity based classifier which first stores the data and then classifies it based on similarity features without making any assumptions. [13] It’s learning is slow but it is quite simple and easy to implement.

SVM is a binary classifier that divides the data points into their respective classes by tracing the best suitable hyperplane . It is done by maximizing the margin(distance between hyperplane and support vectors) [14]

In the DT classifier , binary choices are organized into branches that resemble a tree. A feature is then compared against a threshold at each decision node to select the next node. The sequence ends and the sample is given a class label when there are no decision nodes left. [15]

RF[16] is based on predictions from several decision trees in place of just one, forecasting the ultimate class label depending on the majority of the outcomes derived from each decision tree.

2.6 Performance measures for evaluating and comparing each model

In our work we evaluate different performance measures like accuracy, sensitivity and specificity by generating a confusion matrix of each model as shown in fig 4-7.

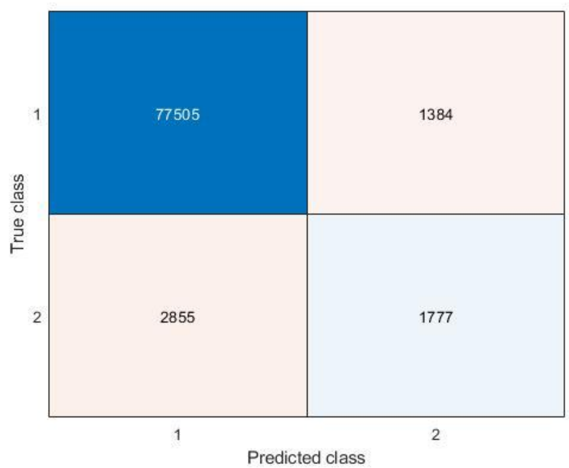

Figure .4 Confusion chart of KNN model taking 70% training and 30% testing data from 83521 samples

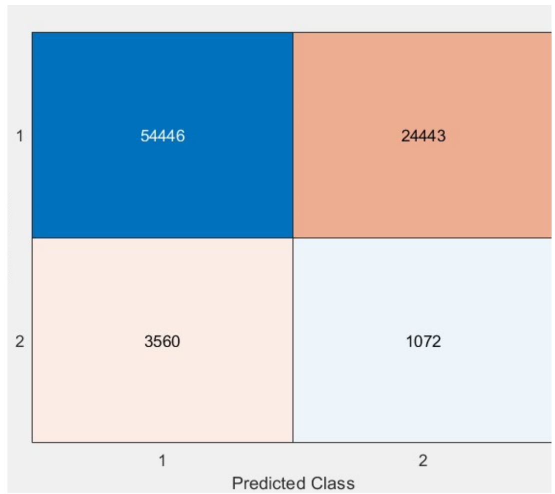

Figure .5 Confusion chart of SVM model taking 70% training and 30% testing data from 83521 samples

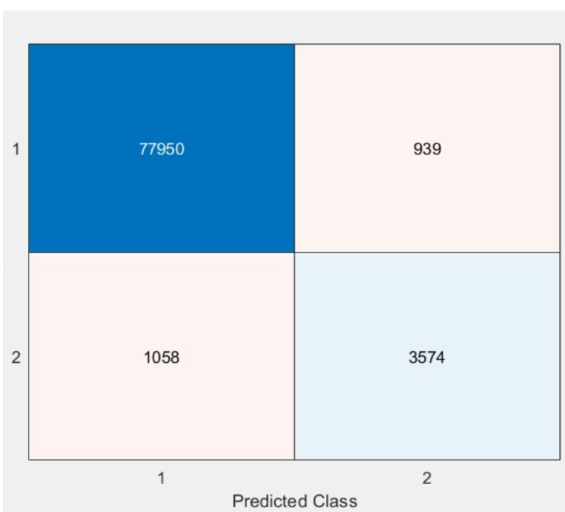

Figure .6 Confusion chart of DT model taking 70% training and 30% testing data from 83521 samples

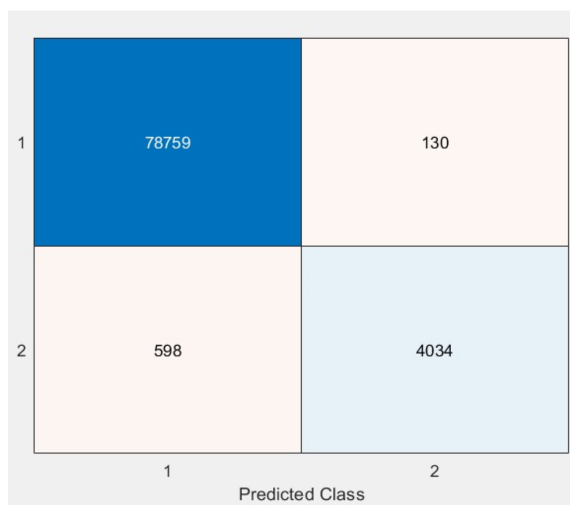

Figure .7 Confusion chart of RF model taking 70% training and 30% testing data from 83521 samples

3. Results

The performance measures namely accuracy, sensitivity and specificity of KNN, SVM,DT,RF classification models when training data is varied from 20% to 90% is shown in the figures below:

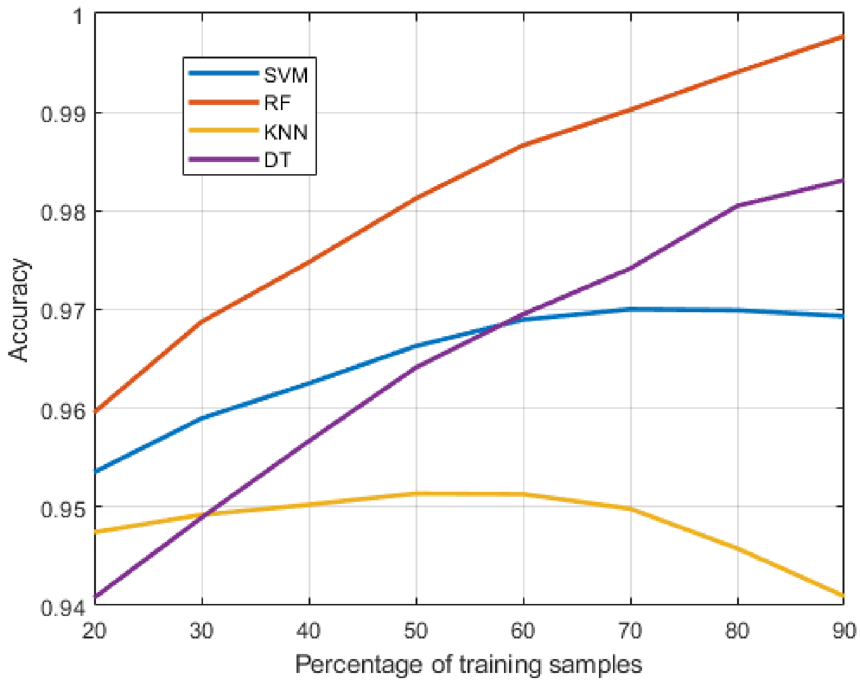

Figure .8 Accuracy chart of 4 models obtained by varying training data from 20% to 90%

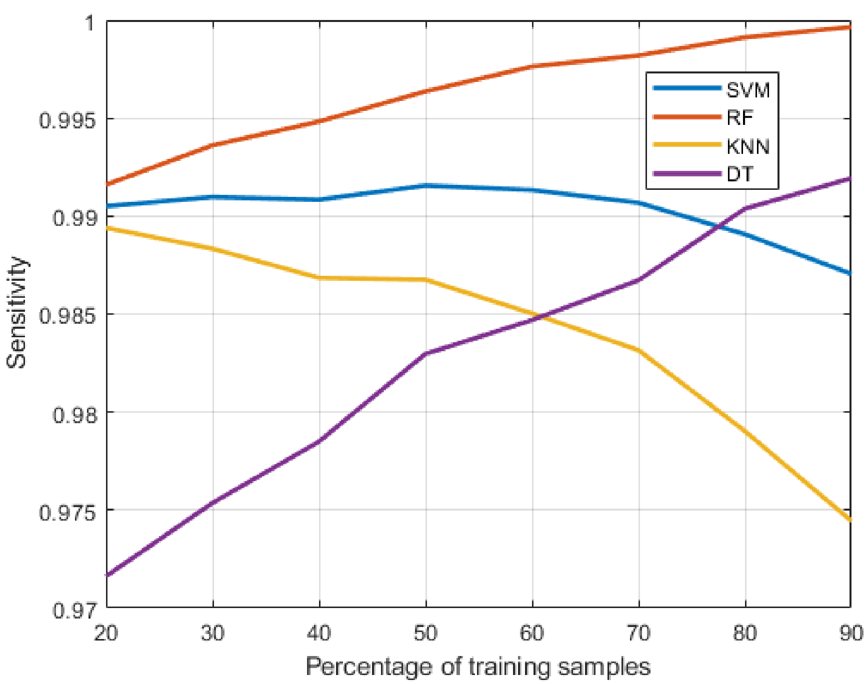

Figure .9 Sensitivity chart of 4 models obtained by varying training data from 20% to 90%

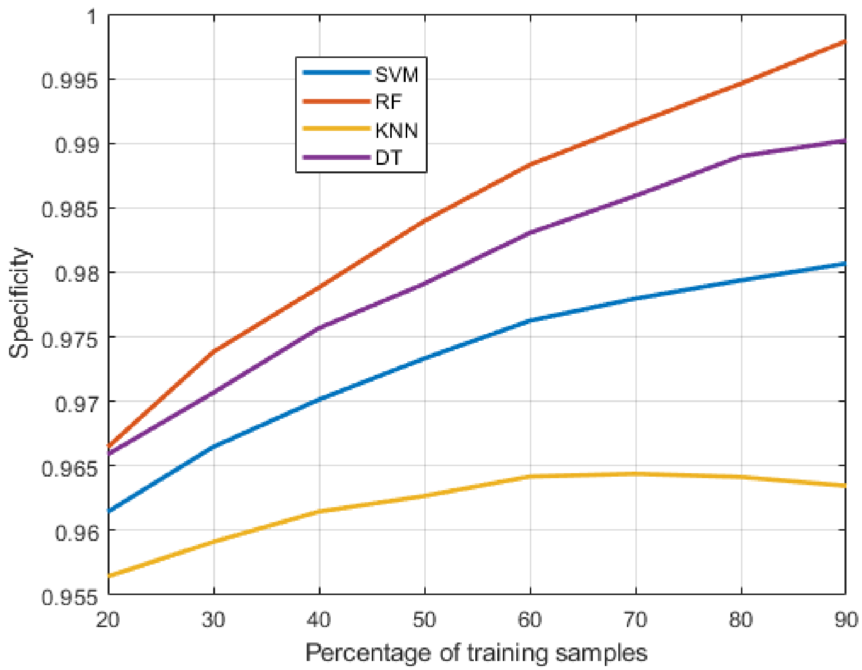

Figure .10 Specificity chart of 4 models obtained by varying training data from 20% to 90%

The bar graphs showing error in accuracy, sensitivity and specificity of KNN, SVM,DT,RF classification models when training data is varied from 20% to 90% is shown in the figures below:

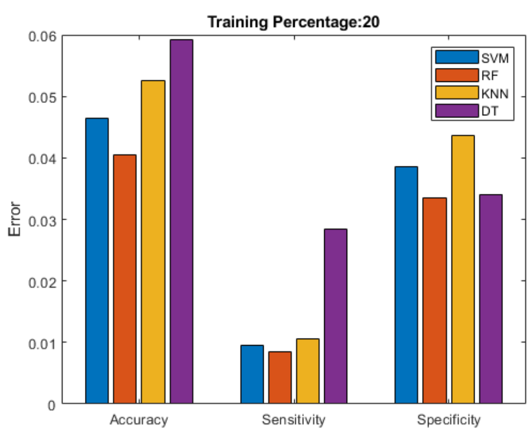

Figure. 11 Errors in accuracy,sensitivity and specificity by taking 20 % training data

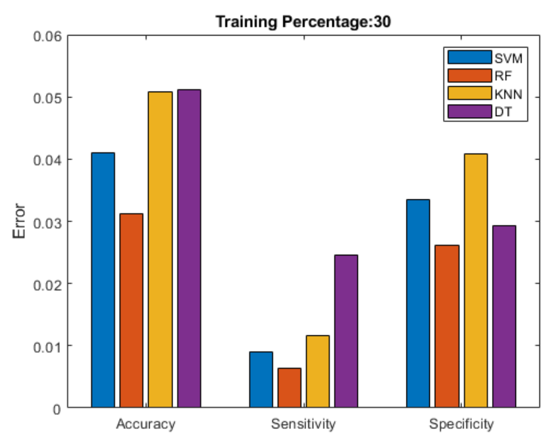

Figure. 12 Errors in accuracy,sensitivity and specificity by taking 30 % training data

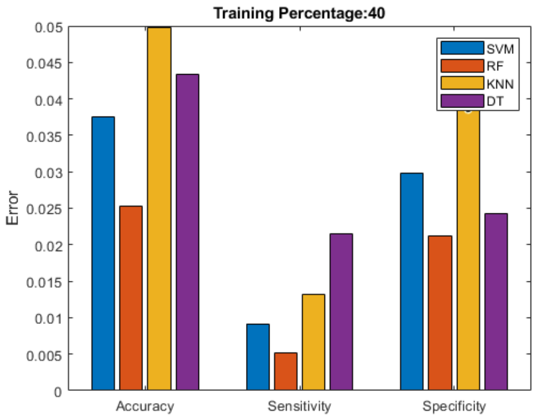

Figure. 13 Errors in accuracy,sensitivity and specificity by taking 40 % training data

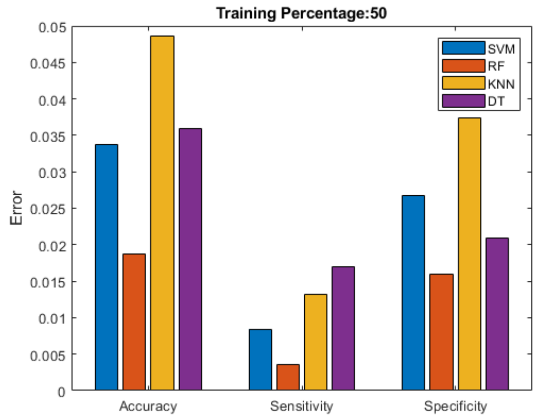

Figure. 14 Errors in accuracy,sensitivity and specificity by taking 50 % training data

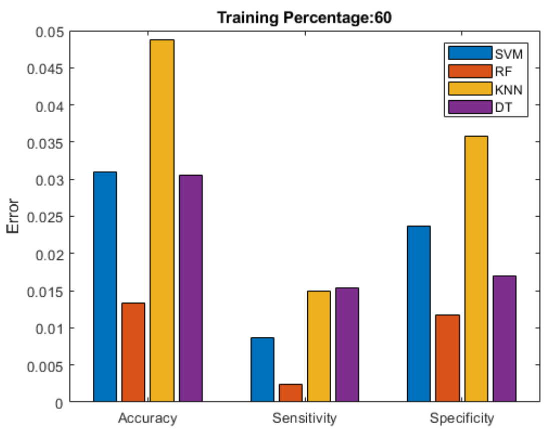

Figure. 15 Errors in accuracy,sensitivity and specificity by taking 60 % training data

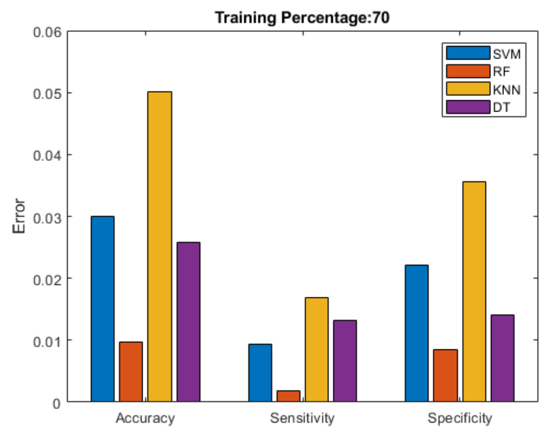

Figure. 16 Errors in accuracy,sensitivity and specificity by taking 70 % training data

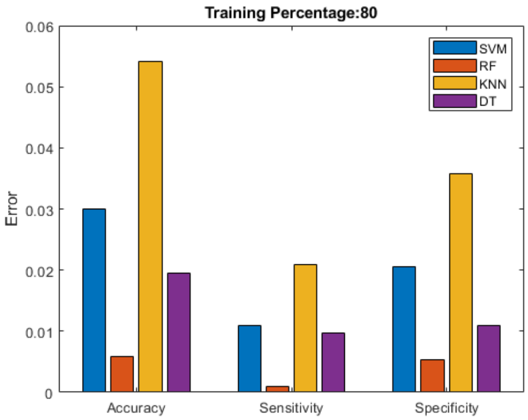

Figure. 17 Errors in accuracy,sensitivity and specificity by taking 80 % training data

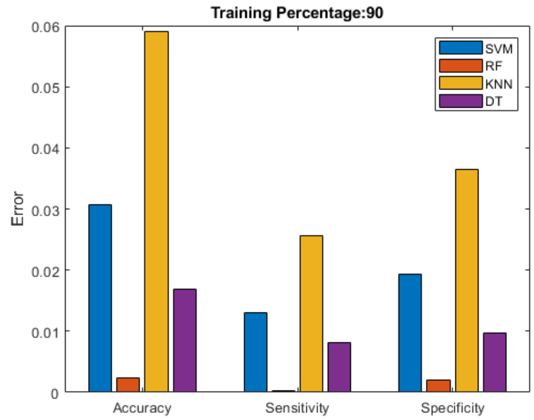

Figure. 18 Errors in accuracy,sensitivity and specificity by taking 90 % training data

4. Conclusion

In our study we utilized 4 machine learning classification models namely KNN, SVM, DT AND RF to forecast gait freezing episodes in Parkinson patients by taking into account the tri-axial accelerometer data set. We then extracted 7 features from the channels of acceleration in X,Y and Z-axes to train the above classifiers. Furthermore, the training data set was varied from 20% to 90% and testing data set from 80% to 10% to evaluate their performance measures. We found the RF classification model was the most suitable one for forecasting gait freezing events in Parkinson disease as it had the best accuracy, sensitivity,selectivity and the least error among the four models. The study's findings can help in developing intelligent real time health assistants and proper diagnosis of patients suffering from Parkinson's disease.

References

[1] J. A. Ross PhD and S. S. Singer PhD, "Parkinson's disease," Magill's Med. Guide Online Ed., 2016.

[2] P. J. Garcia-Ruiz, K. R. Chaudhuri, and P. MartinezMartin, "Non-motor symptoms of Parkinson's disease A review…from the past," J. Neurol. Sci., vol. 338, no. 1– 2, pp. 30–33, Mar. 2014.

[3] S. K. Van Den Eeden et al., "Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity," Am. J. Epidemiol., vol. 157, no. 11, pp. 1015–1022, Jun. 2003

[4] J. G. Nutt, B. R. Bloem, N. Giladi, M. Hallett, F. B. Horak, and A. Nieuwboer, "Freezing of gait: moving forward on a mysterious clinical phenomenon," Lancet Neurol., vol. 10, no. 8, pp. 734–744, Aug. 2011.

[5] N. Giladi, R. Kao, and S. Fahn, "Freezing phenomenon in patients with parkinsonian syndromes," Mov. Disord., vol. 12, no. 3, pp. 302–305, May 1997.

[6] D. Aarsland, J. P. Larsen, E. Tandberg, and K. Laake, "Predictors of Nursing Home Placement in Parkinson's Disease: A Population-Based, Prospective Study," J. Am. Geriatr. Soc., vol. 48, no. 8, pp. 938–942, Aug. 2000

[7] J. M. Hausdorff, Y. Balash, and N. Giladi, "Time series analysis of leg movements during freezing of gait in Parkinson's disease: akinesia, rhyme or reason?," Phys. Stat. Mech. Its Appl., vol. 321, no. 3–4, pp. 565–570, Apr. 2003.

[8] J. D. Schaafsma, Y. Balash, T. Gurevich, A. L. Bartels, J. M. Hausdorff, and N. Giladi, "Characterization of freezing of gait subtypes and the response of each to levodopa in Parkinson's disease," Eur. J. Neurol., vol. 10, no. 4, pp. 391–398, Jul. 2003

[9] R. Velik, U. Hoffmann, H. Zabaleta, J. F. Marti Masso, and T. Keller, "The effect of visual cues on the number and duration of freezing episodes in Parkinson's patients," Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf., vol. 2012, pp. 4656–4659, 2012.

[10] S. Mazilu, A. Calatroni, E. Gazit, D. Roggen, J. M. Hausdorff, and G. Tröster, "Feature Learning for

Detection and Prediction of Freezing of Gait in Parkinson's Disease," in Machine Learning and Data Mining in Pattern Recognition, P. Perner, Ed. Springer Berlin Heidelberg, 2013, pp. 144–158.

[11] M. Bachlin et al., "Wearable Assistant for Parkinson's Disease Patients With the Freezing of Gait Symptom," IEEE Trans. Inf. Technol. Biomed., vol. 14, no. 2, pp. 436–446, Mar. 2010.

[12] B. D. Fulcher and N. S. Jones, "Highly Comparative Feature-Based Time-Series Classification," IEEE Trans. Knowl. Data Eng., vol. 26, no. 12, pp. 3026–3037, Dec. 2014

[13] O. Kramer, "K-Nearest Neighbors," in Dimensionality Reduction with Unsupervised Nearest Neighbors, Springer, Berlin, Heidelberg, 2013, pp. 13–23.

[14] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," IEEE Intell. Syst. Their Appl., vol. 13, no. 4, pp. 18–28, Jul. 1998.

[15] C. D. Sutton, "Classification and Regression Trees, Bagging, and Boosting," Handb. Stat., vol. 24, pp. 303–329, Jan. 2005.

[16] L. Breiman, "Random Forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, Oct. 2001.