Parkinson’s Disease Action Tremor Detection with Supervised-Learning Models

Minglong Sun,
Computer Science Department, William & Mary, Williamsburg, United States

Woosub Jung,
Computer Science Department, William & Mary, Williamsburg, United States

Kenneth Koltermann,
Computer Science Department, William & Mary, Williamsburg, United States

Gang Zhou,
Computer Science Department, William & Mary, Williamsburg, United States

Amanda Watson,
The PRECISE Center, University of Pennsylvania, Philadelphia, United States

Ginamari Blackwell,
School of Nursing, Virginia Commonwealth University, Richmond, United States

Noah Helm,
School of Nursing, Virginia Commonwealth University, Richmond, United States

Leslie Cloud,
Department of Neurology, Virginia Commonwealth University, Richmond, United States

Ingrid Pretzer-Aboff
School of Nursing, Virginia Commonwealth University, Richmond, United States

Abstract

People with Parkinson’s Disease (PD) have multiple symptoms, such as freezing of gait (FoG), hand tremors, speech difficulties, and balance issues, in different stages of the disease. Among these symptoms, hand tremors are present across all stages of the disease. PD hand tremors have critical consequences and negatively impact the quality of PD patients’ everyday lives. Researchers have proposed a variety of wearable devices to mitigate PD tremors. However, these devices require accurate tremor detection technology to work effectively while the tremor occurs.
This paper introduces a PD action tremor detection method to recognize PD tremors from regular activities. We used a dataset from 30 PD patients wearing accelerometers and gyroscope sensors on their wrists. We selected time-domain and frequency-domain hand-crafted features. Also, we compared our hand-crafted features with existing CNN data-driven features, and our features have more specific boundaries in 2-D feature visualization using the t-SNE tool. We fed our features into multiple supervised machine learning models, including Logistic Regression (LR), K-Nearest Neighbours (KNNs), Support Vector Machines (SVMs), and Convolutional Neural Networks (CNNs), for detecting PD action tremors. These models were evaluated with 30 PD patients’ data. The performance of all models using our features has more than 90% of F1 scores in five-fold cross-validations and 88% F1 scores in the leave-one-out evaluation. Specifically, Support Vector Machines (SVMs) perform the best in five-fold cross-validation with over 92% F1 scores. SVMs also show the best performance in the leave-one-out evaluation with over 90% F1 scores.

Keywords
Parkinson’s Disease; Action Tremor Detection; Wearable Device; Supervised-learning

1 INTRODUCTION

Parkinson’s disease (PD) is a chronic and continuous central nervous degeneration disease with more than ten million patients suffering worldwide[13]. PD has these main motor symptoms, including bradykinesia, postural instability, rigidity, freezing of gait, and tremors[29]. Typical PD tremors are hand and finger tremors, which can cause inconvenience and potentially severe consequences in patients’ daily lives[2].

Currently, PD and PD tremors cannot be cured entirely. However, multiple treatments and mitigation methods are performed to decrease tremor severity. Traditional clinical approaches have been utilized to mitigate the symptoms, including deep brain stimulation surgery[3] and Levodopa replacement therapy [9]. However, these are not always successful and have side effects that can be more disruptive than the symptom itself. In recent years, less intrusive wearable devices like vibration devices have been utilized to mitigate tremors in the home environment[11, 38].

In order to design a system to mitigate or quantify PD tremors, we need to detect the PD tremors accurately. Typically, PD tremors have three types: rest, postural, and action tremors[17]. The patients’ activities help distinguish the tremor types. Rest tremors are seen when patients sit or lie down without doing any activities. Postural tremors occur when patients maintain a stationary pose. Action tremor appears as patients are involved in activities such as writing and drawing. TremorSense[36], a system to detect PD tremors, has used a multi-layer Convolutional Neural Network (CNN) method. The CNN method detects rest tremors and postural tremors with above 95% accuracy. However, the accuracy and F1-score for action tremor detection was only 70%–80% in our evaluation. This problem motivates us to investigate more reliable features and robust models to classify PD action tremors.
To accurately detect PD action tremors, we need to solve the two following research questions:

RQ1: How to design features to accurately detect action tremors?

RQ2: How to design models to accurately detect action tremors?

In this paper, we introduced an approach to classify PD action tremors from daily activities. We used the dataset collected from 30 PD patients with a wearable device consisting of accelerometer and gyroscope sensors on the patient’s wrists. We divide the data into instances with a time frame of 1.28 seconds. The sampling rate of the sensors was 100Hz. Each instance has 1.28*100*6 data points. We employed 4,000 action tremor instances and 4,000 activity instances. We extracted hand-crafted features based on domain knowledge and selected the features with the LASSO algorithm. Based on the t-SNE visualization tool, the high dimension distribution of the our hand-crafted features is more apparent than the data-driven TremorSense features. We evaluated our features using Linear Regression (LR), Support Vector Machines (SVMs), K-Nearest Neighbors (KNNs), and Convolutional Neural Network (CNN). The performance of all four models has more than 90% of F-1 scores and 92% accuracies.

Our contributions are summarized as follows:

- To answer research question 1, we extracted hand-crafted features based on domain knowledge and selected the features with the LASSO algorithm. From the t-SNE visualization results, our hand-crafted features have more apparent boundaries than the data-driven TremorSense features.

- To answer research question 2, we evaluated our features on four supervised learning models. The performance of all models has more than 90% of F-1 scores and 92% of accuracies, which indicates that different classification algorithms perform robustly and similarly with our features.

We organize our paper as follows. In Section 2, we first introduce the details of our dataset. Then, we present feature extraction and selection methods. Section 3 presents the machine-learning models we utilized to classify PD actions tremors. In Section 4, we introduce the classification performance of these machine-learning models. In Section 5, we discuss our limitations and propose potential works. In Section 6, we present the related works about tremor detection. In Section 7, we conclude our research results.

2 FEATURE EXTRACTION AND SELECTION

2.1 Dataset

We used the same dataset as TremorSense [36]. The dataset was collected in a clinical environment with the help of medical specialists. The dataset includes 30 patients performing seven UPDRS activities and three Fahn-Tolosa-Marin scale activities. Table 1 shows the specific activities, and Table 2 describes all the 30 patients’ demographic information.
The data for the TremorSense dataset was collected using two UG sensor bands containing a three-axis accelerometer and a three-axis gyroscope [43]. The UG sensor band is shown in Figure 1. The sensor data was transmitted via Bluetooth to the TremorSense App. During the tests with patients, medical professionals used a camera to record patients upper extremity movement to serve as the ground truth data. The ground truth videos were then used to label the action tremor output recorded by the UG sensor band.

The UG sensor sampling rate is 100Hz in the TremorSense dataset. The dataset was divided into windows with a 1.28 seconds time windows and a 0.64-second step size. Since the UG sensor has a three-axis accelerometer and a three-axis gyroscope, each window includes 128*6 data points. We selected 4,000 action tremor instances and 4,000 non-tremor activities from the dataset. The total data points are 6,144,000 (128*6*4000*2), enough to train the TremorSense CNN classifier [36].

2.2 Features Extraction and Selection

As we evaluated the TremoSense data-driven features in Section 4, the accuracy is only about 80%. This result motivated us to select more hand-crafted features. Therefore, we extracted 6 * 14 features from both the time and frequency domains. Table 3 displays the extracted features. The time-domain features include Min, Max, Mean, Standard Deviation, and Zero crossing values. These features are computed for each raw component of the accelerometer and gyroscope readings. Similarly, the frequency-domain features are Energy, Entropy, Spectral-centroid, and the first three dominant frequencies and their magnitudes between 2Hz-10Hz. After we extract these features, we use feature selection techniques to select the best features to predict action tremors.

Figure 2 shows the three axes directions of both sensors in the UG sensor. The white arrows indicate the direction of motion of PD action tremors. Intuitively, the PD action tremors have very little acceleration in the Y direction along the arms. Conversely, PD tremors have reasonable acceleration variation in both the X and the Z directions. This observation leads us to select more acceleration features in the X and Z direction. Feature selection is also beneficial for reducing high-dimension classification costs and helps avoid overfitting. The feature selection algorithm we have chosen in a Logistic Regression method is Least Absolute Shrinkage and Selection Operator (LASSO).

LASSO is a popular regularization and feature selection algorithm for machine learning classifiers. The LASSO model has a constraint on the sum of the absolute values of the model parameters, and the sum has to be less than an upper bound. The method applies a shrinking process, also known as the regularization process, where it penalizes the weights of the regression features and forces some of them to zero. If the features still have a non-zero weight after the shrinking process, they are selected to be a part of the model.

The LASSO starts with a linear regression model and can be expressed as:

\[ Y_i = \beta_0 + X_1\beta_1 + \ldots + X_n\beta_n \]  

(1)
$X_i$ is the input features, $Y$ is the prediction matrix, and the parameters $\beta_0, \beta_1, \cdots, \beta_p$ are the regression weights. The cost function of LASSO regression is as follows:

$$\sum_{i=1}^{n} \left( y_i - \sum_{j=1}^{p} x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$  \hspace{1cm} (2)

$\lambda \sum_{j=1}^{p} \beta_j^2$ is the LASSO penalty addition to the L1 optimization cost function. If $\lambda$ is zero, then it will be a common L1 optimization problem. However, as $\lambda$ becomes larger, LASSO will shrink the weight of less significant features to zero, which works perfectly for feature selection. During our feature selection implementation, we select the $\lambda$ that perfectly drops the Y-axis of acceleration features and maintains other useful axes features.

Table 4 shows the selected features. As expected, the Y-axis features of the accelerometer are dropped. Also, the maximum and minimum value time-domain features are dropped for all axes. Those features are dropped because the maximum and minimum values are typically not predictable when the patients perform activities, and they are not related to the action tremors. In total, the selected features are 5 * 3 time-domain features and 5 * 9 frequency-domain features. Table 5 shows the top 20 selected features and their importance scores. The importance scores are calculated as follows:

$$\text{Importance score} = \frac{\text{Certain Feature Weight}}{\sum |\text{Weight}|}$$  \hspace{1cm} (3)

The Top 20 features total weights are 54.96% of the total weights, representing the most important features. Next, we feed all the selected features into the supervised-learning models.

3 MODELING

We implemented multiple supervised learning classifiers with the selected features to detect PD action tremors. We employed classic learning algorithms, such as Linear Classification Models, K-Nearest-Neighbors (KNNs), and Support-Vector Machines (SVMs). We also applied Convolutional Neural Networks (CNNs) to perform the classification. In the following section, we present each classifier we designed to differentiate PD action tremors from regular activities.

3.1 Linear Classification Model

The first model we employ is a Linear Classification Model. The model $Y$ can be presented as follows:

$$Y = \beta_0 + \sum_{j=0}^{p} \beta_j X_j$$  \hspace{1cm} (4)

where $X$ is the feature matrix, and $\beta$ is the weight for each predictor.
\[
\hat{\beta} = \arg\min_{\beta} L(\beta) 
\]  
(5)

where \( \hat{\beta} \) is the weight vector that can minimize the loss function \( L(\beta) \). The loss function can be presented as follow:

\[
L(\beta) = \| y - \beta X \|^2 
\]  
(6)

The final hyperplane that separates the two classes is:

\[
\hat{Y} = X\hat{\beta} = X(X^TX)^{-1}X^TY 
\]  
(7)

\( \hat{Y} \) is the projection of \( Y \) onto columns of \( X \). The hyperplane separates the two classes of space. If a data point lies in a certain space, then the instance is classified as that category.

### 3.2 KNN Classification Model

The k-nearest-neighbors classification model is another frequently used supervised-learning model. A majority decision of the nearby data points can classify an input instance. According to a Euclidean distance function, an instance is categorized into the class with the highest prevalence among its \( K \) closest neighbors:

\[
D = \sqrt{\sum_{i=1}^{k} (x_i - y)^2} 
\]  
(8)

Intuitively, if a particular instance is near most of the tremor class points, it is classified as a tremor instance. The KNN model has a principal parameter \( K \); as \( K \) increases, the model becomes more inflexible. We implement five-fold cross-validation to choose \( K \) with the highest accuracy to pick the \( K \).

### 3.3 SVM Classification Model

Another well-known supervised-learning algorithm we employ is Support-Vector Machine. It is most suitable for binary classification, and we classify tremor and non-tremor instances in our model. The advantage of SVM compared with the previous algorithms is that SVM provides linear classification and non-linear classification. It can provide a more flexible model that is moderately helpful in classifying similar instances on the boundary between tremor and non-tremor instances.

The SVM maximizes the margin around the separating hyper-plane. A subset of training samples specifies the decision function called the support vectors. The equation that defines the decision surface separating the classes is a hyperplane of the following form:

\[
w^T + B = 0 
\]

\( W \) is a weight vector, \( X \) is the input vector, and \( B \) is the bias. In our case, the hyperplane separates tremor and non-tremor classes.
3.4 CNN Classification Model Design

In this subsection, we present our CNN design for action tremor classification. We classify each input action instance as a tremor event or a regular activity. Each instance includes time-domain and frequency-domain signal features. Our CNN model consists of eight layers. Next, we present each layer and its components of our CNN model in the following paragraphs.

3.4.1 Input Layer.—This input layer takes our 1-D (1 * 60) instance into the CNN model. It primarily organizes the input size for the subsequent layers, and it has any parameters or features to learn.

3.4.2 Convolutional Layer.—Our CNN model has one convolutional layer since our input size is small and simple to learn. The input size of our model is 1 * 60. We select a kernel size of 1 * 3 and the number of kernels is eight. The stride direction is only horizontal, and we use a stride size of 1 * 1. The convolutional layer generates the same amount of feature maps for each input instance as kernels. Therefore, the outputs of our convolutional layer are 1 * 58 * 8 feature maps.

3.4.3 Batch Normalization Layer.—To decrease the sensitivity to network initialization and prevent overfitting, we employ a batch normalization layer. Additionally, it can speed up CNN training and apply feature maps to the next ReLU layer. Eq. 10 illustrates the activation function. This function standardizes the input $s_i$ by a mini-batch for each input instance. Eq. 11 demonstrates the output of this layer.

$$\hat{s}_i = \frac{(s_i - \mu_B)}{\sqrt{\sigma_B^2 + \epsilon}}$$  \hspace{1cm} (10)

$\mu_B$ is the mean and $\sigma_B$ is the variance of the mini-batch.

$$y_i = k\hat{s}_i + \beta$$ \hspace{1cm} (11)

where $k$ is the scale factor, $\beta$ is the compensation, and $\hat{s}_i$ is the normalized output in Eq. 10.

3.4.4 Relu Layer.—The ReLU layer, also known as the rectified linear unit layer, is a function that uses a non-linear activation function to transform all the negative input values into zero. Eq. 12 shows the activation function:

$$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$ \hspace{1cm} (12)

The ReLU layer is more computationally efficient and helps avoid overfitting.

3.4.5 Max Pooling Layer.—The max pooling layer is primarily used to downsample the ReLU Layer’s output. The output of the max pooling layer is the max value of the inputs by a pooling filter. The max pooling layer utilized a pooling filter to generate the output by selecting the max value of the input matrix. In our model, the pool size is 1 * 2, and the stride size is 1 * 2. The output of our Max Pooling layer is 1 * 29 * 8.
3.4.6 **Fully-connected Layer.**—This fully-connected layer is the neural network to calculate the weights and bias based on the previous input features. Each neuron decides the weight that prioritizes the most relevant category and generates the probability for the classification decision. We use 32 neurons in our Fully-connected layer and have 7,456 parameters to train.

3.4.7 **Softmax Layer.**—This softmax layer uses the softmax function to generate the probability for each instance that belongs to ‘action tremor’ and ‘no action tremor’. In this softmax layer, we set our classification probability boundary as 50%, which means if the probability of an instance is more than 50% of the ‘action tremor’, then it will be classified as a tremor instance.

3.4.8 **Classification Output Layer.**—The output of our classification layer is a binary result of ‘action tremor’ and ‘no action tremor’. For each round of training and testing in the evaluation phase, the outcome is compared against the ground truth to create the confusion matrix.

4 **PERFORMANCE EVALUATION**

We implemented two evaluation experiments, cross-evaluation and leave-one-out evaluation, to illustrate the performance of our supervised learning classifiers and selected features. We then compared our models and features with the state-of-art model TremorSense. We utilized the 30 patients’ dataset mentioned in the previous section. The dataset included 4,000 tremor class instances and 4,000 non-tremor class instances. Since the action tremor always happens unpredictably, some patients had more action tremors than others. This issue had less influence on cross-evaluation. However, it will have an impact on the leave-one-out evaluation. Therefore, we picked the top 10 PD patients’ data that included more action tremor events, to test our model in the leave-one-out evaluation. Next, we present the evaluation results and the feature visualization in detail.

4.1 **Cross-Evaluation**

We trained and tested our classifiers using all 30 patients’ instances in cross-validation. We employed five-fold cross-validation methods to train and test for each round. We also evaluated the results with the TremorSense CNN classifier. Figure 4 shows the results of each classification confusion matrix results. Table 6 shows the overall accuracy, precision, recall, and F1-score for each classifier. The F1-scores for all four models are higher than 90%. SVM performs best with more than 92% of accuracy and F1-score. All four models fed with our hand-crafted features perform better than the TremorSense model fed with data-driven CNN features. The results demonstrate that our hand-crafted features are robust with different machine learning models.

4.2 **Leave-one-out Evaluation**

As we mentioned above, we selected ten patients for leave-one-out evaluation. In this experiment, we trained the classifiers with the nine patients’ data and tested the classifiers...
with the one remaining patient’s data. The total tremor instances are 2,000, and the total non-tremor instances are 2,000. Table 7 shows the leave-one-out results for each classifier. Even though the leave-one-out results are slightly less than those of cross-evaluation, the performance is still robust. We can also conclude that SVM has the highest performance with all confusion matrix parameters greater than 91%. Similar to the results of cross-evaluation, all four models perform better than the TremorSense model. The results demonstrate that our hand-crafted features and models are robust with new patients.

4.3 Features Visualization

We further demonstrate that our hand-crafted features perform better than the TremorSense data-driven features. We employed a tool, t-distributed stochastic neighbor embedding (t-SNE), to visualize the high-dimensional features. The t-SNE is a statistical technique to visualize high-dimensional data by assigning every data point a particular location in a 2-D or 3-D figure. Figure 6 shows the distribution of our hand-crafted features and TremorSense data-driven features in 2-D maps. The top two figures show the TremorSense features, and the bottom two figures show our hand-crafted features. We randomly selected 500 instances to plot in these figures. The red dots are the tremor instances, and the blue dots are the non-tremor instances. The X-axis comp-1 and The Y-axis comp-2 represent the two composite features from high dimensions. The different random states represent the random different projection angles to display the features in the 2-D dimension. We can conclude from the figures that our hand-crafted features have more specific boundaries than the TremorSense data-driven features. The results also demonstrate that our features perform better than TremorSense features with the supervised machine learning models.

5 RELATED WORKS

In recent years, many computer scientists have focused on solving PD-related problems with the assistance of a specialist in the field. Researchers utilize different sensors and algorithms to develop detection and quantification systems of various PD symptoms, including freezing of gait (FOG)[21, 37], hand tremors, speech difficulties[4, 24], etc. There have been increasing research works focusing on hand tremor detection and severity assessment in the past decade. They employ different sensors, such as accelerometers[16, 18–20, 22, 31, 32], gyroscopes[34, 35], electromyography (EMG)[5,23], WiFi-based sensors[6], and other motion sensors[26–28, 30, 40] to build detection devices. Based on various signals from these sensors, different models, such as threshold models[16, 18–20, 22, 31], machine-learning models[10, 41, 42], and deep learning models[6, 36, 39], are developed to detect hand tremor events.

Researchers have explored new methods to detect PD hand tremors in the last three years. In 2021, TremorSense[36] proposed a CNN model to detect three types of tremors using the 3-D time-domain accelerometer and gyroscope data. The dataset they collected includes 11 UPDRS activities and three Fahn-Tolosa-Marin activities. In our paper, we employed the same dataset and used the action tremor activities and normal activities as a comparison group. The accuracy of detecting all types of tremors in the TremorSense was more than 94%. When we applied the TremorSense CNN model to the data of only action tremors with
everyday activities, the evaluation results were 70% to 80%. However, we beat TremorSense using our hand-crafted features, and our classification accuracy and F-1 score were more than 90%. In 2021, Tong et al. [39] used their designed IMU sensors to detect PD hand tremors. They used a CNN model to classify tremor events and got an F-1 score of 94.31%. Even though the F-1 score was high, they only collected data from five PD patients who walked in a circle back and forth. In comparison, our dataset is more diverse and results are more reliable since we collected data from 30 PD patients with 14 clinically relevant activities. In 2021[6], Chen et al. employed Wi-Fi sensing for hand rest tremor detection. In a small room, they simulated the frequency of hand tremors from 3Hz to 7Hz and collected channel state information (CSI) signals. They used pre-trained models VGG19 and Resnet152 from Keras and customized the last multiple layers for tremor classification. In 2021, Rini et al.[33] developed a glove with five accelerometers on each finger. They collected data from six patients and calculated the acceleration frequency of each finger. If the frequencies from all five fingers were between 1Hz to 10Hz, it would be a tremor event. This detection algorithm does not not consider when some fingers have 0Hz frequencies while the frequencies others are in the range of 1Hz to 10Hz. In 2020, Ibrahim et al. [12] fed the acceleration and angular velocity data into a hybrid convolutional-multilayer perceptron neural network to detect tremor events. The detection is limited to user-independent and task-independent activities, while our methods can work among different users and activities. In 2020 [1], Ahmed et al. utilized Photoplethysmograph (PPG) built-in smartwatch and EMG sensor to detect hand tremors. The highest accuracy of the tremor detection algorithm was 89%, and they collected data only from healthy subjects.

Researchers are using hand tremor signals to do more. They detect hand tremor signals to help assist wearable devices such as vibration devices to mitigate tremor symptoms in PD patients. They also use the signals to assess tremor severity[14, 15, 32], by applying similar sensors and machine learning algorithms to quantify tremor severity based on clinical standards such as UPDRS scores. In addition, some researchers employ hand tremor signals to detect PD[25] and quantify the severity of PD[8]. All these applications show the value and necessity of hand tremor signal detections.

6 DISCUSSION AND FUTURE WORK

Our paper used conventional motion sensors like accelerometers and gyroscopes to collect data. These sensors are practical and portable for PD patients to use in any environment. However, it is difficult for motion sensors to forecast tremor events. Motion sensors can detect tremor signals when a tremor has already happened. Future research can look into how to predict hand tremors and other PD symptoms using clinical sensors like EEG and EMG.

Our project collected the data in the clinical environment. We did the data processing and built tremor classification models offline. The future work for us is to develop an online system that can detect hand tremors in real-time with wearable devices and smartphones, and run a pilot study to test the robustness and accuracy of our system in different environments.
Our paper mainly selected supervised machine-learning models since we have labeled the action tremor events from tremor activities based on the videos. Future research can focus on unsupervised learning and deep learning models. Those models may potentially enhance the action tremor detection performance.

As we mentioned in the related work section, PD hand tremor signals may be used to evaluate the tremor severity and PD severity. Wearable technology can accurately detect tremors using our methods. PD patients may employ non-intrusive wearable technology to track the daily change in severity in a domestic environment setting. Additionally, accurate tremor detection helps wearable technology provide the appropriate care and mitigation. Our tremor detection techniques will be used in future studies to create these wearable devices.

7 CONCLUSION

Our paper focuses on detecting PD action tremors using the accelerometer and gyroscope sensor. We extracted hand-crafted features based on domain knowledge and selected the features with the LASSO algorithm. From the t-SNE visualization results, our hand-crafted features have more apparent boundaries than the state-of-art data-driven TremorSense features. We evaluated our features on four supervised learning models. The performance of all models has more than 90% of F-1 scores cross-validation results and more than 88% of F-1 scores leave-one-out results, which indicates that different classification algorithms perform robustly and similarly with our features. The performance of our models using hand-crafted features beat the TremorSense using data-driven features.

ACKNOWLEDGMENT

This research was partially supported by the U.S. NIH NINDS under grant R01NS120560 and by the U.S. NSF under grant CNS-1841129 (CSR EAGER). The author would like to thank all co-authors who have contributed to collect the PD patients’ data.

REFERENCES

[7]. Cummings Jeffrey, Isaacsin Stuart, Mills Roger, Williams Hilde, Chi-Burris Kathy, Corbett Anne, Dhall Rohit, and Ballard Clive. 2014. Pimavanserin for patients with Parkinson’s disease


Figure 1:
UG Sensor band
Figure 2:
UG Sensor Axes
Figure 3:
Convolutional Neural Network Layers
Figure 4:
Cross-Validation Confusion Matrix
Figure 5:
Leave-one-out Confusion Matrix
Figure 6:
TremorSense and Our Features in t-SNE 2-D Maps
### Table 1:

Scales and Activities

<table>
<thead>
<tr>
<th>Number</th>
<th>Scale</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UPDRS</td>
<td>Finger Tapping</td>
</tr>
<tr>
<td>2</td>
<td>UPDRS</td>
<td>Hand Movements</td>
</tr>
<tr>
<td>3</td>
<td>UPDRS</td>
<td>Hands Pronation-Supination</td>
</tr>
<tr>
<td>4</td>
<td>UPDRS</td>
<td>Toe Tapping</td>
</tr>
<tr>
<td>5</td>
<td>UPDRS</td>
<td>Leg Agility</td>
</tr>
<tr>
<td>6</td>
<td>UPDRS</td>
<td>Gait</td>
</tr>
<tr>
<td>7</td>
<td>UPDRS</td>
<td>Kinetic Tremor of The Hands</td>
</tr>
<tr>
<td>8</td>
<td>Fahn-Tolosa-Marin</td>
<td>Handwriting</td>
</tr>
<tr>
<td>9</td>
<td>Fahn-Tolosa-Marin</td>
<td>Drawing</td>
</tr>
<tr>
<td>10</td>
<td>Fahn-Tolosa-Marin</td>
<td>Pouring Water</td>
</tr>
</tbody>
</table>
### Table 2: Patient Demographics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Number</td>
<td>30</td>
</tr>
<tr>
<td>Gender</td>
<td>18 Males / 12 Females</td>
</tr>
<tr>
<td>Patient Age</td>
<td>45 – 84 Years old</td>
</tr>
<tr>
<td>Patient Average Age</td>
<td>67.43 Years old</td>
</tr>
<tr>
<td>PD Symptom Onset Age</td>
<td>35 – 82 Years old</td>
</tr>
<tr>
<td>PD Diagnosed Age</td>
<td>38 – 82 Years old</td>
</tr>
<tr>
<td>Disease Duration Years</td>
<td>0–24 Years</td>
</tr>
<tr>
<td>Disease Duration Average Years</td>
<td>8.80 Years</td>
</tr>
</tbody>
</table>
### Table 3: Extracted Features

<table>
<thead>
<tr>
<th>Domain</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time-domain Features</td>
<td>Min, Max, Mean, Standard deviation, Zero crossing/ Mean value crossing</td>
</tr>
</tbody>
</table>
| Frequency-domain Features | Energy: Sum of FFT component magnitudes of the input sensor signal.  
                          | Entropy: It measures of the distribution of frequency components.   
                          | Spectral-centroid: Spectral power distribution balancing point.   
                          | First three dominant frequencies and their magnitudes between 2Hz-10Hz |
### Table 4: Selected Features

<table>
<thead>
<tr>
<th>Domain</th>
<th>Selected Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time-domain Features</td>
<td>All-axis Mean, Std, Zero crossing / Mean value crossing except ACC(Y)</td>
</tr>
<tr>
<td>Frequency-domain Features</td>
<td>All-axis frequency-domain features except ACC(Y)</td>
</tr>
</tbody>
</table>
Table 5:

Top 20 selected features

<table>
<thead>
<tr>
<th>Features*</th>
<th>Rank</th>
<th>Importance Score</th>
<th>Features</th>
<th>Rank</th>
<th>Importance Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gyro(Y) first Frequency</td>
<td>1</td>
<td>0.0483</td>
<td>Gyro(Y) Energy</td>
<td>11</td>
<td>0.0234</td>
</tr>
<tr>
<td>Gyro(Y) Std</td>
<td>2</td>
<td>0.0377</td>
<td>Gyro(X) first Frequency</td>
<td>12</td>
<td>0.0227</td>
</tr>
<tr>
<td>Gyro(Y) Mean</td>
<td>3</td>
<td>0.0325</td>
<td>Gyro(Y) first Magnitude</td>
<td>13</td>
<td>0.0197</td>
</tr>
<tr>
<td>Acc(X) first Frequency</td>
<td>4</td>
<td>0.0318</td>
<td>Acc(X) Entropy</td>
<td>14</td>
<td>0.0196</td>
</tr>
<tr>
<td>Acc(Z) first Frequency</td>
<td>5</td>
<td>0.0306</td>
<td>Gyro(X) Std</td>
<td>15</td>
<td>0.0195</td>
</tr>
<tr>
<td>Acc(X) Std</td>
<td>6</td>
<td>0.0297</td>
<td>Acc(X) Energy</td>
<td>16</td>
<td>0.0185</td>
</tr>
<tr>
<td>Gyro(Y) Entropy</td>
<td>7</td>
<td>0.0294</td>
<td>Acc(X) first Magnitude</td>
<td>17</td>
<td>0.0184</td>
</tr>
<tr>
<td>Acc(Z) Std</td>
<td>8</td>
<td>0.0292</td>
<td>Acc(Z) first Magnitude</td>
<td>18</td>
<td>0.0182</td>
</tr>
<tr>
<td>Acc(X) Mean</td>
<td>9</td>
<td>0.0288</td>
<td>Acc(Z) Entropy</td>
<td>19</td>
<td>0.0181</td>
</tr>
<tr>
<td>Acc(Z) Mean</td>
<td>10</td>
<td>0.0256</td>
<td>Gyro(Y) Zero crossing</td>
<td>20</td>
<td>0.0179</td>
</tr>
</tbody>
</table>

*Acc: Accelerometer; Gyro: Gyroscope. (X), (Y), (Z): represent X, Y, Z axes of the sensors.
## Table 6:

Cross-evaluation Confusion Matrix

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Model</td>
<td>91.37%</td>
<td>91.44%</td>
<td>91.05%</td>
<td>91.24%</td>
</tr>
<tr>
<td>KNN</td>
<td>90.21%</td>
<td>90.02%</td>
<td>90.45%</td>
<td>90.51%</td>
</tr>
<tr>
<td>CNN</td>
<td>92.24%</td>
<td>92.50%</td>
<td>91.98%</td>
<td>92.21%</td>
</tr>
<tr>
<td>SVM</td>
<td>92.28%</td>
<td>92.40%</td>
<td>92.13%</td>
<td>92.26%</td>
</tr>
<tr>
<td>TremorSense</td>
<td>83.60%</td>
<td>87.18%</td>
<td>79.27%</td>
<td>83.04%</td>
</tr>
</tbody>
</table>
**Table 7:**
Leave-one-out Evaluation Confusion Matrix Results

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Model</td>
<td>88.70%</td>
<td>88.86%</td>
<td>88.50%</td>
<td>88.68%</td>
</tr>
<tr>
<td>KNN</td>
<td>88.43%</td>
<td>88.64%</td>
<td>88.15%</td>
<td>88.39%</td>
</tr>
<tr>
<td>CNN</td>
<td>90.20%</td>
<td>90.28%</td>
<td>90.10%</td>
<td>90.19%</td>
</tr>
<tr>
<td>SVN</td>
<td>91.20%</td>
<td>91.08%</td>
<td>91.35%</td>
<td>91.21%</td>
</tr>
<tr>
<td>TremorSense</td>
<td>77.83%</td>
<td>77.13%</td>
<td>79.10%</td>
<td>78.10%</td>
</tr>
</tbody>
</table>