Nat. Volatiles & Essent. Oils, 2021; 8(5): 12320-12340

NVEO Natural Volatiles & Essential Oils

Detection and Predictive analytics of Parkinson's disease employing Tremor analysis and Deep Learning Algorithm

Krishna Prasad S J [1], Mohana Kumar S [2], Puttabore Gowda G [3],Swati Swami [4]

[1]Dept Electronics and Telecommunication, Associate Professor, M S Ramaiah Institute of Technology,Bangalore, India,

[2]Dept of Computer Science and Engineering, Associate Professor, M S,Ramaiah Institute of Technology, Bangalore, India.

[3]Dept of Mechanical Engineering, Professor, S Ramaiah Institute of Technology Bangalore, India,

[4]Dept of Electronics and Telecommunication, Student, M S Ramaiah Institute of Technology, Bangalore, India,

Abstract

Among medical diseases affecting human races and communities globally neurological diseases are dominant one. Latest investigations have revealed statistics that more than seven million globally and one million in United States which is a developed country are suffering from Parkinson's disease. In fact studies reveal that predicted statistics is so large that medical communities are underprepared to treat these patients affected in near future .Parkinson's disease (PD) is a disease of nervous system. This disease is caused due to paucity of dopamine in human brain and effect being on the daily routine activity profile of affected person. Symptoms abstractly remain same among communities but may vary slightly among genders and age groups. Onset of Parkinson's disease detection and predictive analytics is an innovation in medical research. Clinical diagnosis of the disease as practiced by clinicians globally is by examination of tremor signals. Tremor quantification methods encompass clinical analysis of drawn figures, computerized signal analysis in time/frequency domains and using tremor rating scales. Objectively to complement clinical research, varieties of Tremor rating scales are standardized for assessment of intensity of disease among neurological patients. Primitive scale employed being Fahn-Tolosa-Marin Tremor Rating Scale (TRS). It is a scale rated from 4 (serious tremor) to 0 (no tremor) based on amplitude of tremor which happens to be a 5-point scale. This study scope is development of a hardware tremor detection module using Raspberry Pi processor interfaced with flex and accelerometric sensors attached on wrist of subjects from which dataset of tremor is built. To complement same machine learning algorithms namely Random Forest and Convolutional neural network (CNN) algorithms are developed in Python for Predictive analytics of onset of disease. Comparison of algorithms indicates that prediction accuracy is 96.97% of CNN algorithm.

Key words: Parkinson's disease, Raspberry Pi, Accelerometer Sensors, Tremor rating scale, Convolutional Neural Network, Random Forest algorithm

1. Introduction:

Neurological disorders are associated with the human body nervous systems. These symptomatic effects are paralysis, muscle weakness, poor physical and mental coordination, sensation reduction, seizures, tremors and altered cognitive effects. A timely and accurate diagnosis is a focus in diagnostic research. Neuron loss in

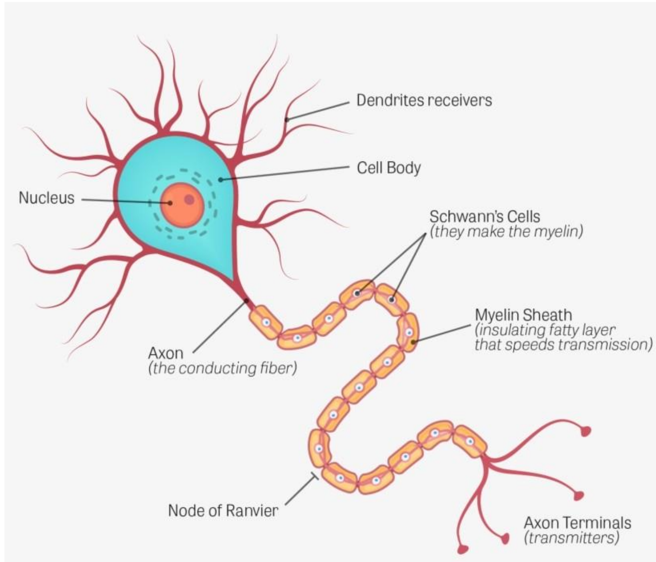

Figure 1. Neuron Image[22]

increasing trend and nervous system rupture in different areas are causes for onset and progression of neurological disorders. Neurons are the functional units of brain and all along body neurons are present. A neuron is termed as biological cell consisting of cell body, dendrites extensions/receivers, axons to connect other neurons and a nucleus. Neurons in billions are spread throughout human body. The figure1 shows a schematic of neuron. On the onset of sickness in neurons, it loses ability of impulse transmission to surrounding neurons and hence its metabolism decreases [1][2]. In the last stage a neuron loses its axon and cyton reducing to vacuoles [1]. Irrespective of neurological disorders its nature is to cause deterioration in coordination of nervous system. The amount of deterioration is felt by reduced nerves communication, hearing and cognition levels among subjects. Hundred billion neurons (nerve cell) are present in the human body. The structural and functional units of nervous system are neurons. Cyton of the neuron (nerve cell) contains DNA and RNA. The cell cycles of neurons are similar to the life cycle of biological cell namely its creation, reproduction and eventual death. Cells will divide by process of mitosis. Once cells are created, it will be in the first growth phase G1, in which major of proteins for duplication of DNA are synthesized. Next, cells enter Synthesis phase in which the DNA replication takes place. Further, cell will enter Second growth phase G2 where proteins essential for mitosis are formed. Unlike other body cells however, neurons are generally considered not capable of reproducing once they grow as they lack centrosomes, which are the initiators of cell division as visualized in nervous system of adults. Nevertheless, it can be possible that neurons can take reentry to the cell cycle under special conditions. Sympathetic and cortical neurons, which tries to reactivate cell growth cycle when exposed to severe insults namely oxidative stress, DNA damage and excitotoxicity. This phenomenon is called as "abortive cell cycle re-entry". Normally cells die in G1/S phase prior to DNA replication. When neurons gets impaired and damaged, it loses its ability of communication with other similar entities, its metabolism becomes low. In the last stage a neuron loses its axon and cyton reducing to vacuoles [2]. Parkinson's disease is prominent among movement disorders caused due to loss of dopaminergic neurons in Substantia niagra of brain cells that produce such symptoms. Hypothesis says that the misfolding of proteins and dysfunction in certain neurological pathways are the genesis of the Parkinson's disease. James Parkinson in his famous monograph " Essay of shaking palasy" in 1817 has Illustrated details of disease popularly named as paralysis agates. After later times it was further popularized by medical community as Parkinson's disease. Parkinson's disease obviously affect Substantia niagra brain neurons whose presence is a needed for easy overall body movements. It is an exclusive disorder of the nervous system, caused and alleviated due to nonreversible loss of brain cells which mainly produces dopamine [3] [4]. Parkinson's disease (PD) is evolved in both genders. Nevertheless, occurrence probability in men is large. This disease statistically will be around 50% higher in men. PD onset is high among aged people. Statistics reveal that subjects whom disease affects more than 60 years are large. Subjects for which evolvement by PD between age of 50 and 60 are 5 to 10%. In Small sets of subjects below 50 years with whom specific gene mutations are evolved there will be onset of Parkinsonism [5]. It is disease of nervous system eventual guidelines of the management and diagnosis of PD patients have been published from National health and clinical excellence in UK [3][5]. Predominant features in subjects with Parkinson's disease are as follows:

- Shaking in arms, hand and body parts named as tremors.
- Balance loss in posture and poor coordination
- Stiffness or rigidity in joints and muscular region
- Bradykinesia which in common man language called as slowness in movement [5].

PD subjects have difficulties in walking, talking and completion of any specific day to day actions. Statistics reveal that subjects whom disease affects who are more than 60 years are large. In reality, initial or early problems discussed above are experienced by subjects, nevertheless symptoms intensity increases in slower rate. Time of onset of disease severity varies among varieties of human races. Hence progression of disease is specific to a subject [5]. Severity can be at early or later time depending on the patient's environmental conditions. Execution of day to day activities of subjects becomes hard to realize as severity increases. To add to complexity features of Parkinson's disease are connected to brain encompassing mental depression, cogwill rigidity, Postural reflex impairments, anxiety disorders in daily activities, vocal cordal paresis, speech difficulties. There has been a study which was conducted on these subjects suffering from symptoms highlighted and epidemiologic studies have been made on subjects. This study and development was enacted with consent from Institutional Review Board of the Kaiser Foundation Research Institute and study revealed that the incidence of Parkinson's disease varies with gender and ethnicity [6]. In another study conducted researchers have observed that there will be variations of speech signal features for PD patients. Researchers in that work have used linear regression models derived from speech signals of PD patients speaking vowels [7].

This study has a focus on development of innovative machine learning algorithms for identification of onset and eventual predictive analytics of disorder of neurology associated with Parkinson's disease which is alarmingly increasing trend among genders and ethnicity. Article has been organized as Section II which focuses on previous works done on estimation and detection of disease. Section III highlights broad methodology/implementation of the system proposed. Section IV, highlights results realized from work done and also from predictive CNN and Random forest algorithms. Conclusions and future scope are illustrated in in section V.

2. Previous Works

In this section a brief review of the previous works are discussed. Also associated literatures are reviewed in the domain of predictive analytics of Parkinson's disease employing various sensors.

Sachin et.al., in their works [8] state that Specific gait characteristics are analyzed which would classify Parkinson's disease from other neurological disease like Amytropic lateral sclerosis. From time series data charectaristics,featrue vectors are analyzed and best features are extracted from which classification are undertaken. Support vector machines were used for classification and results reported that accuracy obtained is 83.33% and detection rate of Parkinson's disease has been 75%.

Ennas Abdulhai et.al., in their works [9] have investigated gait and tremor signals using machine learning algorithms for diagnosis of Parkinson's disease. It is done by making use of Chebyshev order II high pass filter with 0.8 Hz cutoff frequency. Filtered data is used for extracting gait features using peak detection algorithm. Here various kinetic feature of pressure of force is used .Work documents accuracy of 92% using machine learning and tremor analysis.

Shu et.al.,[10] in their works have developed a novel wearable, electronic wireless module for textile sensors. It renders huge scale of measurement, minimal errors, wireless communication, robustness in isolation for power, optimal cost and also works as real event quantification. It can be employed in health monitoring with help of wearable sensors in indoor or outside environments .It has less than 1% error .It is used in foot pressure measurements.

Christopher L.P et.al., in their works [11] have assessed the fluctuations of tremor, bradykinesia and dyskinesia with respect to levodopa administration. Here in this work wearable sensors have been used instead of diaries for the comfort of patients. Here motion data acquired from sensor from ankle and joints have been used. They have used validated algorithm to extract spectral power in tremor frequencies and deploy them in multiple regression models based on expert clinician ratings of subjects executing daily tasks and scheduled activities. Here accuracy obtained has been documented as 90%.

V.Dietz et.al., in their works [12] have explored three main characteristics in Parkinson's subjects based tremor these are rhythmic discharge during rest, during voluntary contraction abnormal low firing rates and consistent firing rate difference between large and small motor units. Inference from work is tremor is strongly correlated by voluntary contraction related force. At rest or during weak muscular effort it is strongest and with increasing value of force it becomes continuously of smaller amplitude and higher frequency.

K.Takei et.al., in their works[13]introduce flexible wearable devices for health monitoring interacting with a person by detecting and providing stimulations to body. This is new class of electronics which not only monitor the health of subjects but also to give remedial measures to problems subjects who are suffering from disorders.In this study latest advancements and innovations toward health-monitoring devices with flexibility are discussed, inclusive of conceptual designs of devices interactive in humans.

Maitin et.al., in their works [14] have employed machine learning (ML) to detect PD through state of rest or motor activation through electroencephalography (EEG) tests. PD classification accuracy was found to be 62–99.62%. There is absence of cleaning protocol for the EEG and a lot of heterogeneity in the features that were acquired from EEG.

Boon G.Lee et.al., in their works have [15] have developed a smart wearable sensor identifying sign language of deaf persons. This wearable system utilizes five flex-sensors, two pressure sensors, and a three-axis inertial motion sensor to classify characters in the American sign language alphabet. Authors claim that without pressure sensors accuracy recorded being 65.7%.Whereas, second generation device has drastically increased the accuracy of detection of signed language.

In one of the literature [2] varieties of disorders of neurology its information and genesis, are debated in detail. Chief factors for the causes, symptoms of varieties of disorders of neurology have been exhaustively explained in this literature.

Objectives and scope of this work realized are being motivated from several of these works mentioned. Primitive objective of work is not only to design and analyze a module employing latest Raspberry Pi microcontroller and accelerometer sensor but also to design and implement benchmarking machine/deep learning algorithms from constructed data set.

DESIGN AND IMPLEMENTATION OF THE EXPERIMENTAL SYSTEM

The broad methodology is highlighted in this part of discussion. It employs Raspberry Pi processor interfaced with accelerometer and flex sensors. Design specifications and realization techniques of modules implemented for acquisition of signals of tremors employing Raspberry Pi processor tagged with accelerometeric sensor are also highlighted. Raspberry Pi is microcomputer is a handy module used popularly for its ability to connect to wireless systems like IOT systems. It provides a framework for beginner to unfold his learning using programming constructs like Python and Pycharm associated with huge data processing. Raspberry Pi has ability to interact with outside environment with the help of sensors and has been deployed in various applications ranging from agriculture, engineering and medicine [16].

ADXL345 is branded as accelerometer sensor which provides user Micro-electro mechanical system (MEMS) capacitive digital sensor with three axes. It has a resolution upto 13 bits, choosable range0 up to +/-16g, It has maximum data rate of 3200Hz and sensitivity 3.9

mg/LSB. It renders digital output that communicates data with the help of SPI and I2C protocols of 14-lead package. Interesting feature of ADXL345 is that it renders acceleration generated from movements making use of a constant of acceleration due to gravity. It can also be employed as tilt sensor, sense object falls and linear accelerations along three axes. In order to provide inherent sensing functions of fall detection it consists of dual interrupt pins [16].

To enact on Predictive analytics tremor data base is developed. Random Forest and Convolutional neural network machine /deep learning algorithms are developed. It is enacted only after suitable dataset is collected and developed from module realized. Further features extracted are acting as inputs to machine /deep learning algorithms modules implemented in various proportions of training and test data and classification results with its assoicated accuracy parameters are discussed in the result section.

Raspberry Pi

Raspberry Pi is connected with accelerometer sensor for the purpose of acquisition of tremor signals. Connection diagram is illustrated as given in the Fig 2. Raspberry Pi GPIO pins are used for that purposes. Accelerometer sensor can transfer data both over Inter-Integrated Circuit (I^2C) and Serial Peripheral (SPI) Protocols. In our study however I^2C protocol is employed to enable communications with ADXL345.The I^2C pins of ADXL345 will be connected directly to Raspberry Pi.I^2C pins are connected without any need of additional circuits. The digital output is16 bits two's complement. These generated values are transformed to acceleration (gravity units) and will be calculated according theoretical equations multiplying with 4mg raw value as highlighted in the equation [1].

$G = 0.004 * raw\ value \quad (1)$

This is mathematical technique of transforming raw value to acceleration values all along x, y and z axes represented in units of gravity. In order have further transformation from acceleration in units of gravity to a value of acceleration represented in m/s^2 then following equation [2] is employed

$Accel(m/s2) = 9.8 * Accel(In\ units\ of\ gravity)(2)$

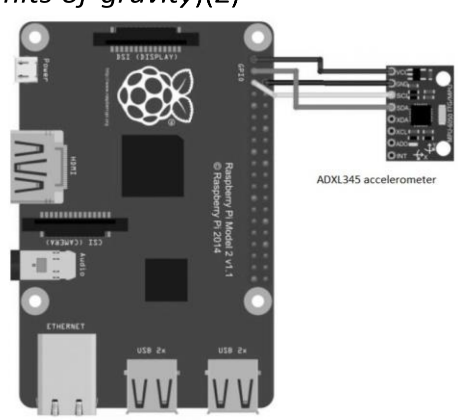

Figure 2. Interfacing of RasphberryPi with accelerometer sensor

3.2. Flex Sensor Interfacing

A type of sensors namely flex sensor may also be employed along with Raspberry Pi for acquisition of signals of tremor. This can be for the purpose of benchmarking results obtained from accelerometer sensors. ADC0804 which is an ADC enacts at 5V and renders output in values of +5V logic can be used. ADC0804 is equipped with 8 outputs indicating 8 bits in which each output provides user +5V output representing logic level 1. But Raspberry Pi can work mandatorily with 3.3V.Consequently If the output logic of ADC0804 +5V is directly connected to Raspberry Pi it may malfunction due to changed voltage levels. Solution for same is to have voltage divider circuit where two registers are employed for division of +5V into two equal half's of 2*2.5 logical voltages. One of those voltage +2.5V is supplied to Raspberry Pi. Whenever ADC0804 generates logic 1, then +2.5V is reached at GPIO pins of Raspberry Pi. The Interfacing circuit connections of flex sensors to Raspberry Pi is as depicted i.

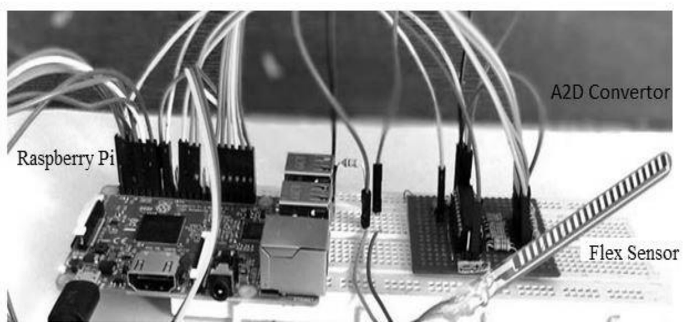

Figure 3. Circuit diagram of flex sensor interfacing

For acquisition of tremor signals, accelerometer sensors are suitable because accelerometer sensor consumes less power with only 3.3V logical requirements. Flex sensor renders output in analog values, for further conversions to digital values, an ADC0804 IC is employed requiring a logical level of 5V. Raspberry Pi requires only 3.3V logical 1 level and output of ADC0804 is 5V logic, eventually mismatch and malfunctioning may be onset in module. A voltage divider circuit has to be designed to divide 5V into two parts before giving input to the Raspberry Pi. Flex sensor power consumptions is more than accelerometer sensors. Accelerometer sensor exhibits ability of acquisition of output in all three axes that is x, y and z whereas, flex sensor can only detect in single axis and provides voltages with help of voltage divider. A little sensation in the axes of the accelerometer sensor detects small amplitude changes in tremor signals. Flex sensor has to use ADC whereas accelerometer sensor does not need same. Hence, accelerometer sensor is compatible for tremor signal acquisition. So after deliberations it was decided to construct dataset employing accelerometer sensor in this study.

Description of data set

To enact our study, development of module for tremors acquisition employing Raspberry Pi and accelerometer sensor is a requirement. Further, sensor is affixed on wrist of the subject of focus. Target person under test is made to be in stable sitting posture and readings of x, y and z axes of accelerometer sensor are tabulated. Samples collection of healthy subjects as well as the PD subjects is realized. Eventually data associated with labels of definite age groups have been developed. The dataset has been recorded based upon tremor signals variations that are in all three axes of sensor. Total 700nos data samples are collected for the further experimentation. After dataset are synthesized, next action is to eliminate unnecessary data and convert it into the desirable format. Data cleansing is process of removal of not only the null values of three axes but also unlabeled data. At primitive step 700 samples have been documented. It happens to be combination of normal and parkinsonian subject's tremor samples. After due cleansing of data, 219 samples which is combination of normal and parkinsonian samples are employed for predictive analytics of disease. In cleansing step initial transient data are rejected to obtain suitable classification results. Total of 219 samples were employed for predictive analytics.

Classification Algorithms

Predictive analytics of Parkinson's disease is realized by employing data acquired and documented from module. Machine learning algorithms which have been employed are Random Forest (RF) and Convolutional neural network algorithms. Machine Learning is a part of artificial intelligence. Machine learning aids information and technology systems to acknowledge the given patterns based on its existing datasets and algorithms which are developed for a required solution. In machine learning data and algorithms must be given to a system in advance, so that desired analysis and pattern recognition of the data has to be determined accurately.

Random Forest: Random forest classification algorithm is one which generates a sets of a decision tree from chosen subset in random of training data. It assimilates voting function developed from decision trees and then decision of final class of the test sample is made. Random Forest algorithm helps one which increases estimation accuracy of the algorithm. Random forest happens to be simplest and popularly employed algorithm used not only for classification and but also for regression analysis. Random forest classifier is defined by tree set of classifiers. It is advancement over bagging classifiers such that randomness is added to the same. Algorithm splits each node employing best among subset of predictors chosen in random. New training set is thus developed. Tree grown is employing random feature selection. RF is fast and robust to

over fitting. It is possible to make any number of trees user wants. To initialize RF algorithm, one must deploy two parametric data. These parametric data are N and m, where N= number of trees to grow and m= number of variables used for node splitting respectively. At each node split is performed using GINI index [17]. GINI index measures class homogeneity and can be written as the equation [3].

$$\sum_{i \neq j}\sum (f(C_i, t)/|T|)(f(C_i, t)/|T|) \quad (3)$$

Where T is training set, C_i is the class that a random pixel belongs to, and $(f(C_i, t)/|T|)$ is the

probability that the selection sample has class C affinity.
The working of Random Forest algorithm is highlighted as following steps:
1 The first step is to select the random samples from a given dataset.
2 In the second step, a decision tree is constructed for every sample. The prediction result is documented from every decision tree.
3 Based on the prediction result, voting will be performed.
4 In the final step, the prediction result with the most voted is considered as the final.

Working steps of Random Forest algorithm is Illustrated in figure 4 [21] shown:

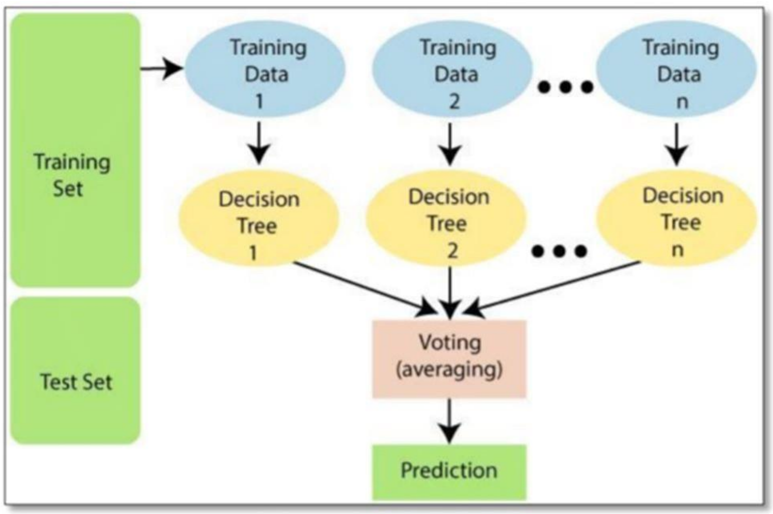

Figure 4. Technique of Random Forest Algorithm[17]

CNN uses the minimal pre-processing steps in the image classification algorithms. It also called the Convnet's as they belong to the class of Deep Neural Networks. These neural networks are mostly used to analyze the visual imagery. These CNN's are the multilayer perceptron. The perceptron are nothing but the mathematical model of a biological neuron. These multilayer perceptron are fully connected networks in case of a CNN. This can be explained as, each neuron in one layer in every layer is connected to all other neurons in the corresponding layer. These convnet's were inspired by the biological processes. These CNN's are very eminent in image recognition. The CNN has different layers namely input layer, convolution layer, max pooling layer, RELU layer, fully connected layer and the output layer [18].

INPUT layer: The input layer comprises of pixel values in the three axes of accelerometer sensor.

CONV layer: The conv layer computes the dot product of the weights between output neuron and the local regions of the input.
RELU layer: To keep the size of the volumes unchanged, the Relu layer applies an activation function element wise.

POOL layer: The spatial dimensions are down sampled by the pooling layer.

FULLY-CONNCECTED layer: The class score shall be computed by this layer. This layer as the name suggests every single neuron is connected to ever other neuron in the previous layer.

The fig5 shows a multi-layer convolution neural network.

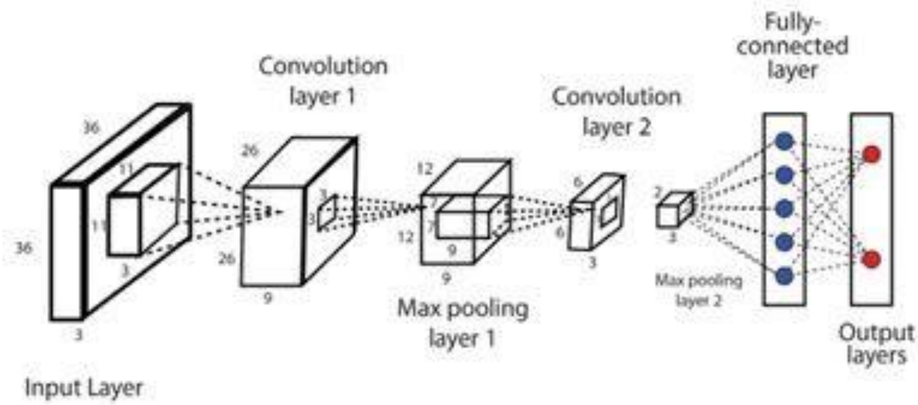

Figure 5. A multi layer convolutional neural network [20]

Fig. 6 depict complete working technique of system realized. The tremors generated in the subjects are acquired by attaching accelerometer sensor on wrist area on subjects of study. Hand is oriented straightway in a stable condition and the tremor signals of healthy subjects are documented. The output accelerometer data will be on all three axes. Here in the accelerometric data features of the tremor are extracted and then the same features are fed to the classification algorithms developed. Classification algorithms will get trained on the data set of both healthy persons and of that subject's suffering from Parkinson's disease. Various proportions of the training and test sets are considered for noting the classification performance of the algorithms developed. These classification results are then compared against benchmarked labeled data (These data are labeled on the priori knowledge of the subject). The classes identified are the Parkinsonian and normal healthy class. Accuracy of the both the algorithms are compared and presented in the results section.

Flowchart of representive algorithms employed and also work flow of the study undertaken is shown in fig7. The diagrammatic depiction of any algorithm is termed as flowchart. The flowchart of realized and working

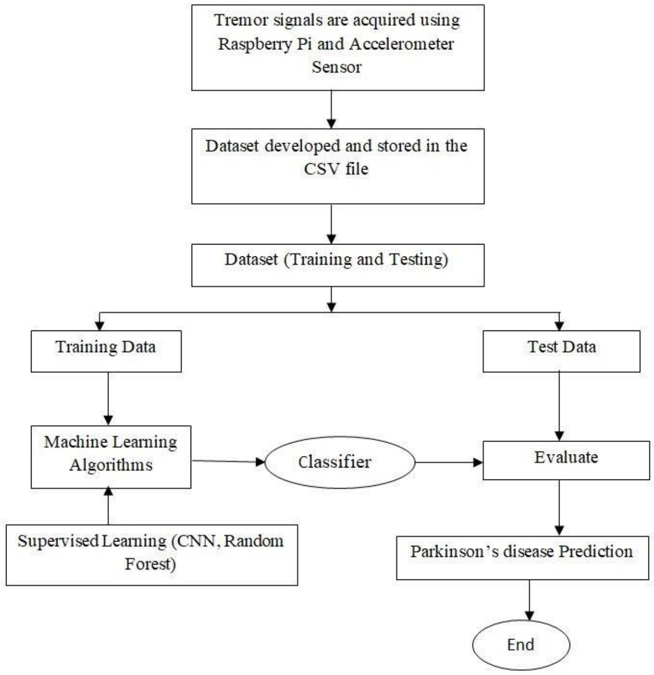

Figure 6. Block Diagram of the Complete Proposed System

model, its methods and material employed are also depicted in the fig 7. Explaining stepwise, initial step in work is to acquire tremor signals by employing Raspberry Pi with accelerometer sensor affixed to wrist of subject under study.

Subjects are being held in stability and tremor signal are acquired. Here the features are further extracted in the tremor signals. From the tremor features the thresholds are set to classify as belonging to either of two classes already defined.

The features extracted are spectrum entropy indicating high energy assimilation at specific bands of interest. Entropy of Spectrum H is calculated by formula indicated in equation [4].

$$H = \sum_{f} p(f) \log p(f) \quad (4)$$

Following features are also defined in the study. One is the low frequency energy E_L(*Low frequency*) and another is high frequency energy E_H(*High frequency*) defined as follows and are given in equations [5] and [6].

$$E_L = \sum_{i\epsilon\omega} ((S^L{}_x)^2 + (S^L{}_y)^2) + (S^L{}_z)^2)) \quad (5)$$

$$E_H = \sum_{i\epsilon\omega} ((s_x^H)^2 + (s_y^H)^2) + (s_z^H)^2))(6)$$

There is also Ratio high to total energy in the tremor signals; Energy is present in both E_L and E_H
bands of signals, there is energy mainly in HF bands .The ratio of high to total energy bands are
modeled in following formula and represented in equation [7].

$$R = E_H /(E_H + E_L \ (7)$$

Where ϵ is a minimal positive constant avoiding division by zero. This value is high in action tremor and low in rest tremor [19].

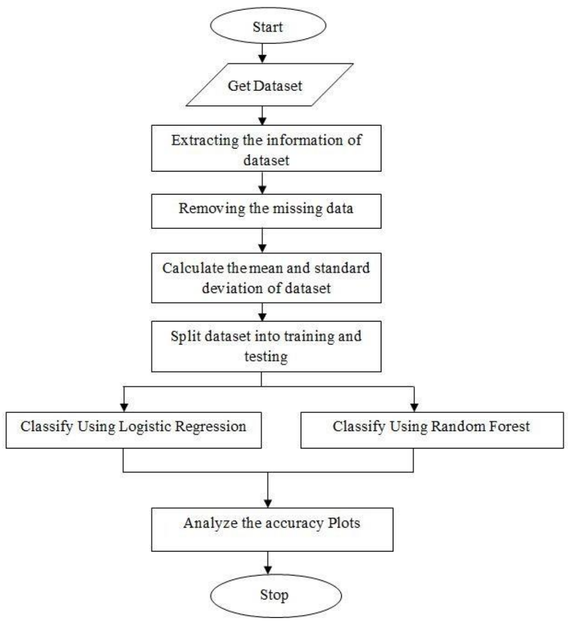

Figure 7. Flowchart of the Proposed System

Here the Ratio of High frequency to energy in total is considered for the classification purposes as the feature fed to the classification algorithm and so also High frequency energy and Low frequency energy. The threshold data after several iterations are decision threshold for x, y and z axes highlighted as follows in the equation [8],[9] and [10].

Experimentation are done on the subjects within the age groups of 20-40 years. Irrespective of the age help of the M S Ramaiah Medical College is considered to have subjects with Parkinsonian disease. Some of the data for statistical conformance are synthesized.

$$-20 < X < 30 \ or \ -0.08G < X < 0.12G \ or$$

$$-0.7848m/s^2 < X < 1.176m/s^2 \ (8)$$

$30 < Y < 90$ *or* $0.12G < Y < 0.36G$ *or*

$1.178 m/s^2 < Y < 3.528 m/s^2$

$Z > 220$ *or* $Z > 0.88G$ *or* $Z > 8.624 m/s^2$

(10)

If x, y and z are in the range of these thresholds, then the subject is classified as healthy class otherwise the subject is Parkinsonian class are inferences drawn from algorithms developed. Basis data is chosen with
averaging calculated from dataset developed. Datasets are synthesized from experimentation which has 700 samples and 11 columns. First three Columns data are x,y and z axes data, time stamp is 4th column and labels drawn from 11th column. Dataset are saved as a CSV file. The numerical value of x y and z are those considered after consolidating all features highlighted in the equations [5][6] and [7].
The dataset is documented in the similar folder where Python code for machine learning algorithms is realized. The code employed to realize machine learning algorithms are in Python. Jupyter note book is employed to debug stepwise execution. Steps are highlighted as follows:

Initial steps realization is to import Operating System LINUX and to employ all requisite Python libraries. Libraries needed are matplotlib,seaborn. numpy and pandas
Then next step is to load the data-set and this data set is developed by extracting features highlighted. Further actions are to eliminate null , unlabeled and transient data employing library of dropna() available in Python.
Data set is reduced to 219 samples are left after cleansing. Calculation of mean and standard deviation of data-set for both the classes. The labels Parkinson's or healthy are coded as 1 or 0 respectively employing python library Encoder (). The threshold as indicated in equations[8] and [9] are calculated from average values of individual classes and basis of classifications are calculated to be fed to machine learning algorithms.

The count number of Parkinson's and Healthy subject data are developed employing library value-counts () i.e. for eg., 188 of 1s and 31 of 0s.

Initialize and feed the data then further split data-set into requisite proportion (for eg 80% training and 20% for testing sets) by employing library of train -test.
The results are also documented in each algorithm by taking data of requisite proportions of training and test data (for eg 70% and 90% for training associated with 30% and 10% for testing sets respectively).

To repeat this steps for both Random Forest and Convolutional neural networks algorithms and benchmark accuracies by comparing classification accuracies and to tabulate classification accuracies of both algorithms.

Results and Discussion

The output window obtained on Raspbian OS as documented by accelerometer sensor can be viewed by employing VNC viewer on a Laptop workstation. Raspberry Pi is connected to laptop workstation using Wi Fi connections. The machine algorithms (Random forest and Convolutional neural networks) are running on Raspbian and employing requisite libraries of Python and also Jupyter Notebook (For testing conformance stepwise) on Laptop workstation.

Accelerometer Sensor Dataset

The Raspberry Pi is interfaced with accelerometric sensor for acquisition of signals of tremor by affixing same on wrists of the subject. Figure 8 shows the tremor dataset. 5th and 6th and 7th column represent three axes without G values. 8th 9th and 10th column represent x,y,z axes data multiplied with G (Acceleration due to gravity).

A	B	C	D	E	F	G	H	I	J	K
x	y	z	time	x in G unit	y in G unit	z in G unit	x in m/s²	y in m/s²	z in m/s²	label
-12	39	245	4.09.59	-0.048	0.156	0.98	-0.4704	1.5288	9.604	Parkinson
-12	59	238	4.10.01	-0.048	0.236	0.952	-0.4704	2.3128	9.3296	Parkinson
-3	54	233	4.10.03	-0.012	0.216	0.932	-0.1176	2.1168	9.1336	Normal
-23	52	238	4.10.05	-0.092	0.208	0.952	-0.9016	2.0384	9.3296	Parkinson
-17	74	234	4.10.07	-0.068	0.296	0.936	-0.6664	2.9008	9.1728	Parkinson
5	59	240	4.10.09	0.02	0.236	0.96	0.196	2.3128	9.408	Normal
6	43	239	4.10.11	0.024	0.172	0.956	0.2352	1.6856	9.3688	Normal
-3	55	251	4.10.13	-0.012	0.22	1.004	-0.1176	2.156	9.8392	Normal
2	35	246	4.10.15	0.008	0.14	0.984	0.0784	1.372	9.6432	Parkinson
-2	30	244	4.10.17	-0.008	0.12	0.976	-0.0784	1.176	9.5648	Parkinson
5	38	241	4.10.19	0.02	0.152	0.964	0.196	1.4896	9.4472	Parkinson
5	39	241	4.10.21	0.02	0.156	0.964	0.196	1.5288	9.4472	Parkinson
13	46	239	4.10.23	0.052	0.184	0.956	0.5096	1.8032	9.3688	Normal
20	38	240	4.10.25	0.08	0.152	0.96	0.784	1.4896	9.408	Parkinson
8	40	231	4.10.27	0.032	0.16	0.924	0.3136	1.568	9.0552	Parkinson
5	19	243	4.10.31	0.02	0.076	0.972	0.196	0.7448	9.5256	Parkinson
-93	26	219	4.10.33	-0.372	0.104	0.876	-3.6456	1.0192	8.5848	Parkinson
-85	14	220	4.10.35	-0.34	0.056	0.88	-3.332	0.5488	8.624	Parkinson
-21	140	219	4.10.37	-0.084	0.56	0.876	-0.8232	5.488	8.5848	Parkinson
11	7	237	4.10.39	0.044	0.028	0.948	0.4312	0.2744	9.2904	Parkinson
26	50	240	4.10.41	0.104	0.2	0.96	1.0192	1.96	9.408	Normal
21	38	241	4.10.43	0.084	0.152	0.964	0.8232	1.4896	9.4472	Parkinson
4	2	244	4.10.45	0.016	0.008	0.976	0.1568	0.0784	9.5648	Parkinson
4	1	244	4.10.47	0.016	0.004	0.976	0.1568	0.0392	9.5648	Parkinson
5	0	243	4.10.49	0.02	0	0.972	0.196	0	9.5256	Parkinson

Figure 8. Tremors Dataset

4.2. Random forest algorithm Results

By considering proportion 80% of total dataset for training and 20% for testing, matrix of confusion of classification is developed as depicted in Table [1]. This data set is of classification results of Random forest algorithm.

Table 1. Matrix of confusion of Random forest employing20% of testing and 80% of training proportion

Real Statistics	Predicted statistics Values	
	Healthy	Parkinsonian
Healthy	2	2
Parkinsonian	0	62

Irrespective of algorithms, 1 is termed as Parkinsonian (1) class and 0 is termed as Healthy (0). Notations used are TN ,TP,FN and FP which are True negative, True positive, False negative and False positive which are described as follows:

TP is notation indicates that Tremor signal statistic is Parkinsonian (1) and it is predicted as Parkinsonian (1).

TN is notation indicating that Tremor signal statistic is Healthy (0) and it is predicted as Healthy (0).

FP is notation indicating that the actual Tremor signal statistic is Healthy (0) but is predicted as Parkinsonian (1).

FN is notation indicating actual tremor signal statistic is Parkinsonian (1) but it is predicted as Healthy (0).

As indicated in Tables [1 and 2] we can term two classes as Healthy class and Parkinsonian class.

Further the performance parameters of classification are documented in the following equations [11] [12][13] and [14].

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (11)$$

$$Sensitivity = TP/(TP + FN) \quad (12)$$

$$Specificity = TN/(TN + FP) \quad (13)$$

$$Precision = TP/(TP + FP) \quad (14)$$

Derivations and reasoning of above equations [11] [12] [13] and [14] have been illustrated in various studies on machine learning [17][18]. Further testing of performance of Random forest algorithm is realized by data sample split in variety proportions. At First 70% of data statistic for training and 30% of data statistic for testing is employed. Among 219 samples, 66 statistics are employed for testing 153 statistics are employed for training. The matrix of confusion is documented for this test results is as depicted in Table [2] and Table [3] .Document values

Table 2. Matrix of confusion Random Forest employing 30% of testing and70%of training proportion

Real Statistics	Predicted statistics Values	
	Healthy	Parkinsonian
Healthy	2	2
Parkinsonian	2	60

are generated from Random forest algorithm by taking three sample proportions of data for training and testing respectively. First Proportion is to split value samples into 90% for training and 10% for testing. The next proportion is to split value samples into 80% for training and 20% for testing purpose. The third proportion is to split data into 70% for training and 30% for testing purpose. The comparison plots of above said data divisions are as depicted in Fig. 9.

Table 3. The Random forest comparison results for different dataset proportions

Dataset Proportions for Random Forest	Accuracy	Precision	Sensitivity	Specificity
90% for training and 10% for testing	90.91%	0.95	0.952	0
80% for training and 20% for testing	93.18%	0.95	0.97	0.5
70% for training and 30% for testing	93.93%	0.96	0.96	0.5

Observation is that for 70% for training and 30% for testing proportion, Accuracy of classification is high. This is due to intense training with formidably large sample to increase the testing accuracy. It is also observed that specificity of algorithm is large indicating as further conformance the accuracy of classification.

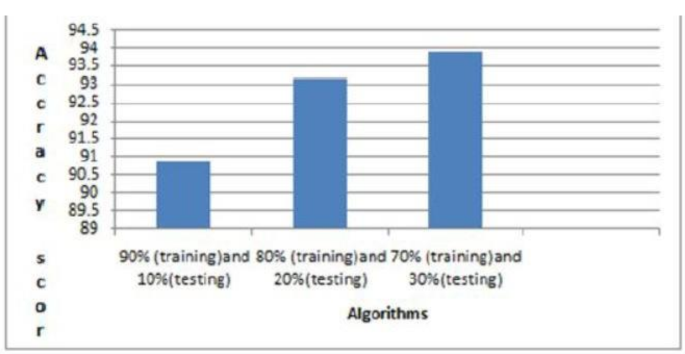

Figure 9. Dataset accuracy of Tremors for various proportions of statistics

4.3. Convolutional Neural Networks Results

Generated results of Convolutional neural network deep learning Algorithm is as realized by employing 80% of sample training sets and 20% of samples testing sets. Results of Matrix of confusion are as indicated in Table 4.

Table 4. Matrix of Confusion of CNN algorithm employing 20% of testing 80% of training Proportion

Real Statistics	Predicted Statistics Values	
	Healthy	Parkinsonian
Healthy	2	2
Parkinsonian	0	40

Now data analytics of CNN algorithm is further executed by employing 70% of data samples as training sets and 30% of samples as testing sets. In total of cleansed 219 statistics, 153 numbers are employed as training sets and 66 numbers are employed as testing sets. The Matrix of confusion resulting for this analytics is as documented in Table 5. In Table [6] comparison of results of CNN algorithm is done by taking three

Table 5. Matrix of confusion of CNN using 30% of testing sets and 70% of training sets of data

Real Values	Predicted Statistics Values	
	Healthy	Parkinsonian
Healthy	2	2
Parkinsonian	0	62

proportions. The first proportion data split is to divide into 90% for training sets and 10% as testing sets. The second proportion data split is to divide dataset into 80% for training sets and 20% for testing sets. The third proportion is to split the data into 70% for training sets and 30% for testing sets. The comparative plots are as depicted in Fig.10.

By analysis of all the results generated by CNN classifier, it has generated best classification results by employing 70% of sample as training sets and 30% of the samples testing sets. Table [7] documents the comparative results of both the algorithms. it can be observed that CNN is suitable for diagnosis and predictive analytics of Parkinson's disease as benchmarked to Random forest algorithm. Fig.11 depicts the comparison of

Table 6. Comparative results of CNN algorithm for various dataset proportions

Dataset proportions	Accuracy	Precision	Sensitivity	Specificity

for CNN				
90% for training and 10% for testing	94.0%	0.96	0.96	0.75
80% for training and 20% for testing	95.45%	0.95	1	0.5
70% for training and 30% for testing	96.97%	0.96	1	0.5

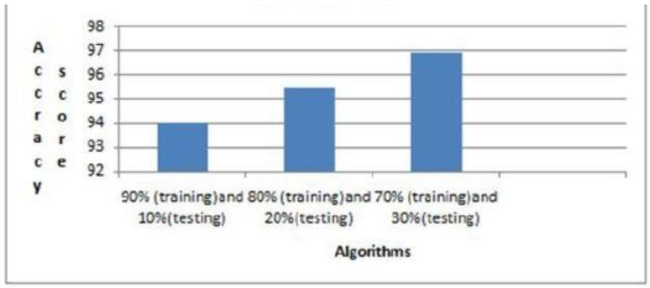

Figure 10. Comparison of CNN for different dataset distribution

accuracies of both Random forest and CNN Algorithms by making use of histogram plot. CNN has documented an accuracy of 96.97%.

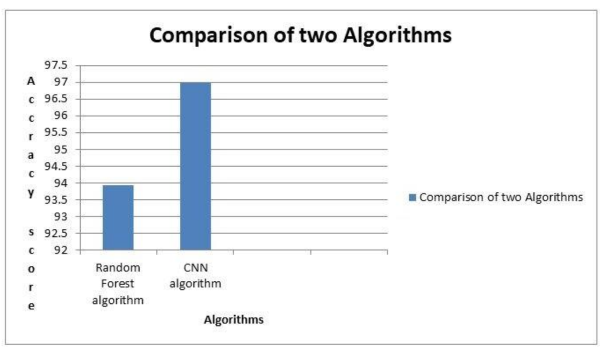

Figure 11. Comparison of Accuracies of Random Forest and CNN Algorithms using bar plot

Table 7. Accuracy comparison of Random Forest and CNN Algorithms

ML algorithms	Accuracy	Precision	Sensitivity	Specificity
Random Forest	93.93%	0.95	0.96	0.5
CNN Algorithm	96.97%	0.96	1.00	0.5

5. Conclusion and Future Scope

In this study, the signals of tremors acquisition and eventual Parkinson's disease analysis and predictive models are realized employing the data set collected from Raspberry Pi and Accelerometer sensors. Sensors are affixed on the wrist of a hand of subject and tremor signals are documented. Innovative and accurate classification algorithms have been realized for predictive analytics of onset of Parkinson's disease. Algorithms employed are Random Forest and CNN algorithms. Highlighted algorithms are employed to classify whether the subject is healthy or Parkinsonian subjects. These results are compared with priori labels associated with data. Further accuracy parameters of classification performances are drawn. Analytics exhibit that the accuracy of CNN is better than the Random forest. Accuracy of CNN algorithm is 96.9% and whereas the accuracy of Random forest is 93.93% . The overall experimentation confirms that these models can be deployed as classification models for predictive analytics of Parkinson's disease. CNN performance is best because of the constructs of deep learning layers like the Max-pooling layer, RELU layer and Soft Max layers in the algorithm development. Also it may be because of migration and convergence of large number of weights of deep learning layers.

In our studies machine learning and deep learning techniques are employed for Parkinson's disease predictive analytics. Though lots of studies are undertaken on machine learning few studies have been documented on deep learning algorithms. In near future, this study can be improvised by transforming and mapping many features to minimum features maintaining same classification accuracy. The dataset employed are tremor signals, but voice signals can be combined along with the tremor signals can be employed to increase confidence of classification. If Dataset is complex, in our study with complex dataset it would definitely documents better results. Future direction is to employ Django App and develop web page to communicate tremor data wirelessly and display results for doctors web portals. It will enable doctors with remote monitoring technologies to have remote supervision.

Acknowledgment

Authors of this study are thankful for support of management, authorities and department of electronics and telecommunication engineering of M.S.Ramaiah Institute of Technology, Bangalore. Facilities provided to conduct this study were helpful to increase the quality of work.

References:

1. World Health Organization *Neurological disorders: public health challenges*2006,World Health Organization
2. World Health Organization *Neurological disorders: public health challenges*,2006,Publisher :World Health Organization
3. Dauer, William and Przedborski, Serge *Parkinson's disease: mechanisms and models* 2003 Journal of Neuron,Vol396, pp 889-909 2003,Elsevier Publications
4. Jankovic, Joseph *Parkinson's disease: clinical features and diagnosis* 2008 Journal of Neurology

neurosurgery & psychiatry ,Vol794, pp 368-376 2008,Publisher BMJ Publishing Group Ltd.

5. Davie, Charles Anthony*A review of Parkinson's disease* 2008 British medical bulletin ,Vol861, pp 109-127 2008,Pub- lisher Citeseer
6. Van Den Eeden, Stephen K and Tanner, Caroline M and Bernstein, Allan L and Fross, Robin D and Leimpeter, Amethyst and Bloch, Daniel A and Nelson, Lorene M *Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity*, Proceedings of American journal of epidemiology, 2003, Vol157no11 pp1015–1022 Publisher Oxford University Press.

7. Sztah´o, D´avid and Kiss, Gabor and Vicsi, Kl´ara*Estimating the severity of Parkinson's disease from speech using linear regression and database partitioning*, Sixteenth Annual Conference of the International Speech Communication Association,2015
8. Shetty, Sachin and Rao, YS*SVM based machine learning approach to identify Parkinson's disease using gait analysis*
9. 2016 International Conference on Inventive Computation Technologies (ICICT) vol2,pp1-5,2016,IEEE
10. Abdulhay, Enas and Arunkumar, N and Narasimhan, Kumaravelu and Vellaiappan, Elamaran and Venkatraman, V *Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease* Proceedings of Journal of Future Generation Computer Systems,Vol83PP366-373,2018,Elseiver Publications
11. Shu, Lin and Tao, Xiao Ming and Feng, David Dagan*A wearable, wireless electronic interface for textile sensors lin shu* Proceedings of 2010 IEEE International Symposium on Circuits and Systems,PP3104–3107,2010,IEEE
12. Pulliam, Christopher L and Heldman, Dustin A and Brokaw, Elizabeth B and Mera, Thomas O and Mari, Zoltan K and Burack, Michelle A *Continuous assessment of levodopa response in Parkinson's disease using wearable motion sensors* Proceedings of the 2017 IEEE Transactions in Biomedical engineering Vol65,No1PP159-164,2017.IEEE Publications
13. Dietz, V. and Hillesheimer, W. and Freund, H.J -*Correlation between tremor, voluntary contraction, and firing pattern of motor units in Parkinson's disease*Journal of Neurology, Neurosurgery & Psychiatry,Vol37,8,PP927- 937,DOI:10.1136/jnnp.37.8.927 1974,BMJ Publishing Group Ltd
14. Takei, Kuniharu and Honda, Wataru and Harada, Shingo and Arie, Takayuki and Akita, Seiji *Toward flexible and wearable human-interactive health-monitoring devices* Journal of advanced health care materials ,Vol(4),4,PP487– 500,2015,Wiley Online Library
15. Mait´ın, Ana Mar´ıa and Garc´ıa-Tejedor, Alvaro Jos´e and Muñoz, Juan Pablo Romero*Machine Learning Ap- proaches for Detecting Parkinson's Disease from EEG Analysis: A Systematic Review* Journal of Applied Sci- ences,Vol10,23,PP8662,2020,Multidisciplinary Digital Publishing Institute
16. Lee, Boon Giin and Lee, Su Min*Smart wearable hand device for sign language interpretation system with sensors fusion*IEEE sensors Journal.Vol18No3PP1224-1232,2017,IEEE
17. Iwaniec, Marek and Holovatyy, Andriy and Teslyuk, Vasyl and Lobur, Mykhaylo and Kolesnyk, Kostyantyn and Mashevska, Marta *Development of vibration spectrum analyzer using the Raspberry Pi microcomputer and 3-axis digital MEMS accelerometer ADXL345*,In: 2017 XIIIth International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH),PP25–29,2017,IEEE Publisher

18. Enes Celik, Sevinc Ilhan Omurca, “Improving Parkinson’s Disease Diagnosis with Machine Learning Methods”, 2019 IEEE.
19. Onishi, Yuya and Teramoto, Atsushi and Tsujimoto, Masakazu and Tsukamoto, Tetsuya and Saito, Kuniaki and Toyama, Hiroshi and Imaizumi, Kazuyoshi and Fujita,Hiroshi,*Automated pulmonary nodule classification in computed tomography images using a deep convolutional neural network trained by generative adversarial networks*,
20. Journal of BioMed research international,Vol=(2019),2019,Hindawi Publisher
21. Rigas, George and Tzallas, Alexandros T and Tsipouras, Markos G and Bougia, Panagiota and Tripoliti, Evanthia E and Baga, Dina and Fotiadis, Dimitrios I and Tsouli, Sofia G and Konitsiotis, Spyridon*Assessment of tremor activity in the Parkinson’s disease using a set of wearable sensors*, Journal of IEEE Transactions on Information Technology in Biomedicine,Vol=(16),No(3)pp(478-487)2012 IEEE Publisher
22. Krishna Prasad SJ , Johnson A, Mohana Kumar S. Lung Cancer Detection with Prediction Employing Machine Learning Algorithms: A Recent Study. New Approaches in Engineering Research Vol. 14. 2021 Aug 25:99-109.
23. Swami S, SJ Krishna Prasad, Mohana Kumar S. Detection and Prediction of Neurological Disorder Using Tremor Signals and Machine Learning Algorithms. In2021 International Conference on Recent Trends on Electronics, Information, Communication and Technology (RTEICT) 2021 Aug 27 (pp. 43-50). IEEE.
24. http://www.clipart-library.com