Asia-Pacific Journal of Neural Networks and Its Applications
Vol.4, No.1 (2020), pp.29-36
http://dx.doi.org/10.21742/ajnnia.2020.4.1.04

Machine Learning Algorithms for Parkinson's Disease Detection

Ch Sekhar[1], M S Rao[2] and D Bhattacharyya[3]

Dept of Comp Sci. and Engineering, Vignan's IIT(A), India
[1]sekhar1203@gmail.com, [2]srinivasamarada@gmail.com

Abstract

Machine learning now days plays a crucial role in real-time problem analysis and providing solutions with its popular algorithms. Nowadays, in the health care sector, machine learning algorithms are involved in detecting the health issues of patients. This paper elaborated detailed information about how the ML algorithms are detecting Parkinson's disease. Parkinson's sickness is caused by the interruption of the brain cells that generate an essence to permit synapses to speak with one another, called dopamine. The cells with the purpose of produce dopamine in the cerebrum are answerable for the control, adjustment and familiarity of developments. At the point when 60-80%of, these cells are missing, at that point adequate dopamine isn't delivered, and Parkinson's engine indications show up Here we used random forest and XGBoost algorithms to detect the disease XGBoost giving the best performance than the Random forest.

Keywords: *Parkinson's disease, Random forest, XGBoost, Machine learning, Neural networks*

1. Introduction

Parkinson's illness is brought about by the interruption of the synapses that bring into being a substance to permit synapses to speak through one another, named as dopamine. The cells with the purpose of produce dopamine in the cerebrum are liable for the control, adjustment and familiarity of developments. At the point when 60-80%of these cells are lost, at that point, enough dopamine isn't created, and Parkinson's engine side effects show up. It is felt that the malady starts numerous years before the engine (development related) indications and in this manner. Scientists are searching for approaches to perceive the non-engine side effects that show up from the get-go in the ailment as right on time as could reasonably be expected, subsequently stopping the movement of the ailment. Right now, learning-based conclusion of Parkinson's infection is displayed. The proposed finding technique comprises of highlight determination and characterization forms. Highlight significance and Recursive Feature eradication strategies were considered for include choice undertaking regression analysis and classifying techniques [1].

Parkinson's illness is a dynamic neurodegenerative issue that causes both engine and non-engine side effects. The engine indications of PD, which incorporate tremor, unbending nature, postural unsteadiness, and extreme slowness, are an immediate consequence of inadequate dopamine motioning because of the particular degeneration of dopamine creating neurons in the substantia nigra district of the midbrain. What's more, the more differing and

Article history:
Received (May 19, 2020), Review Result (June 24, 2020), Accepted (July 28, 2020)

Print ISSN: 2207-8738, eISSN: 2207-8746 AJNNIA

varied non-engine indications, which incorporate autonomic brokenness, hyposmia, psychological decrease, sadness, and rest brokenness, appear likewise to be connected to the obsessive collection of the α-synuclein protein. The specific reasons for PD stay misty however are thought to include a mind-boggling transaction between hereditary qualities, science, and the earth, which adds to the fundamental heterogeneity of PD manifestations, yet also their paces of movement over time. This gives vulnerability to patients concerning ultimate personal satisfaction and exhibits difficulties for helpful preliminaries as far as characterizing capable endpoints [2].

Parkinson's illness (PD) persistent consideration is restricted by deficient, sporadic indication checking, rare access to the mind, and inadequate experiences with medicinal services experts prompting weak restorative dynamic and imperfect patient wellbeing related results. Progressively visit persistent checking and treatment changes can prompt better symptomatic administration and decrease in treatment-related intricacies, for example, engine vacillations [3].

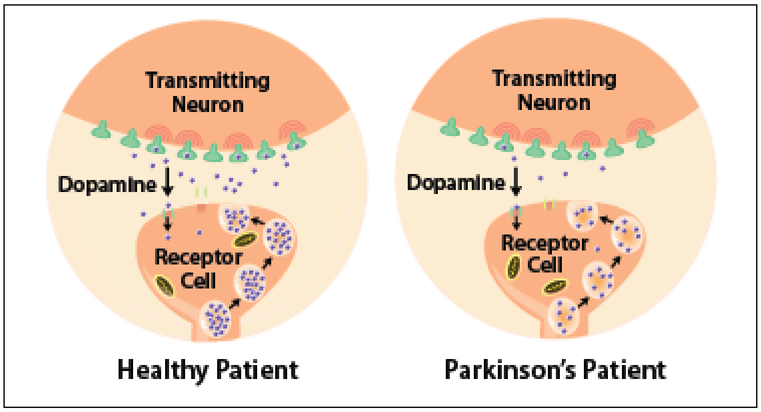


Figure 1. Healthy patient vs Parkinson's patient

The symptoms commonly caused by Parkinson disease patients as mentioned below [2]

- Shaking of figures in hands or Tremors, legs, arms, jaw, and face
- Firmness or stiffness of limbs and chest
- Movement is slowdown
- Unable to balance, verbal communication, and harmonization

Similarly, like motor symptoms, nonmotor also having symptoms, they can identify before motor based symptoms, they mentioned below

- Unfortunate sense of odour
- Constipation
- Hopelessness
- Cognitive impairment
- Low energy

2. Machine learning algorithms

There are various ways an algorithm can show an issue dependent on its collaboration with the knowledge or condition or anything we desire to call the data. It is well known in AI and man-made consciousness course readings to initially believe the learning styles that a calculation can embrace. There is just a couple of standard learning styles or learning models that an algorithm can have, and we'll experience them here with a couple of instances of algorithm and issue types that they suit [4].

a) **Super Learning:** These Learning Algorithms data is marked with naming while training the data. When considering the results of students each record of students can be labelled as pass class or failed class in the results status. A method was establish with the well prepared process with pre defined labels for each class. New data point when received, we can easily place it in well know the class label.

b) **Unsupervised Learning:** Learning Algorithms when new data point received that can be placed randomly as in this, there is no class labels are not available. It was considered as learning without any labelled data marked or without teacher learning concept. It might be through a mathematical procedure to efficiently decrease repetition, or it might be to arrange data by likeness.

c) **Reinforcement Method:** Uses this algorithm, the machine is prepared to settle on explicit choices. It works along these lines: the machine is presented to a situation where it trains itself consistently utilizing experimentation. This machine gains from past understanding and attempts to catch the ideal information to settle on exact business choices.

The given below are the set of most frequent used ML algorithms. These algorithms can be applied to almost any data problem.

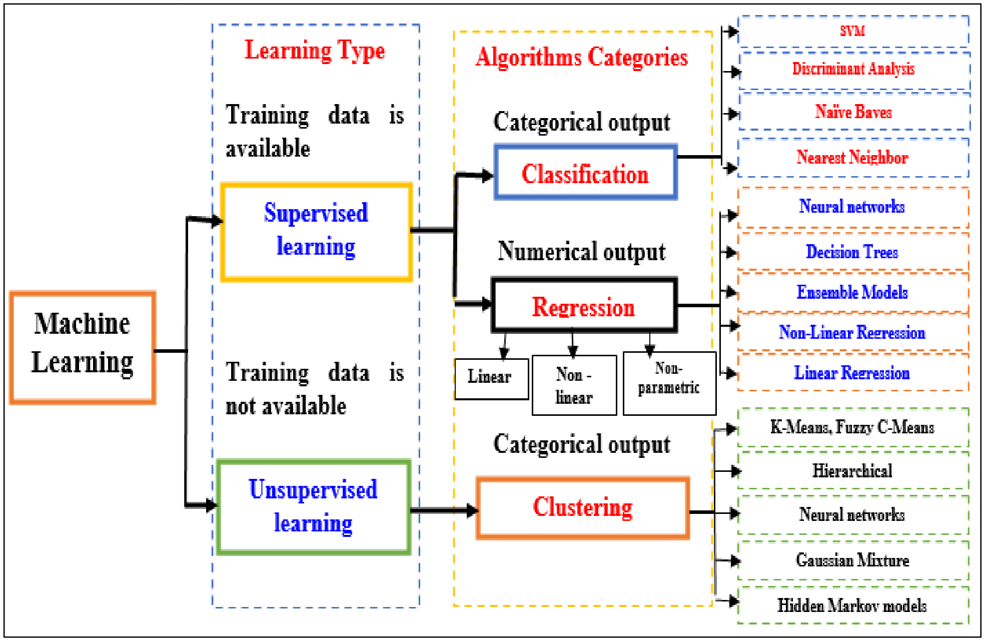


Figure 2. Various types of machine learning algorithms

Now we are using Random forest and XGBoost algorithms to analyze and forecast the Parkinson's disease using the available data set from UCI ML Parkinson's dataset. The dataset has 195 samples or instance with 24 fields.

2.1. Radom forest algorithm

RF (Random Fore) method is a teacher based learning method which is utilized for both classification and regression. It can work as a decision tree, Random Forest method makes choice trees on information tests and afterward gets the expectation from every one of them lastly chooses the best arrangement by methods for casting voting. Irregular backwoods or arbitrary choice woodlands are machine learning techniques for making choice trees to perform AI errands. Classification and relapse as well as different undertakings that work by developing an enormous quantity of option trees at preparing time and yielding the class that is the method of the classes (classification) or mean prediction (regression)of the individual trees [5]. Arbitrary choice backwoods right for choice tress propensity for over fitting to their preparation set. The principal method for RF (RANDOM forest) was prepared by Tin Kam Ho utilizing the unbalanced subspace technique, which, in Ho's plan, is a move toward to execute the "stochastic segregation" way to deal with characterization projected by Eugene Kleinberg. The development joins Breiman's "stowing" thought and irregular determination of highlights, offered first by Ho and later on autonomously by Amit and Geman so as to develop an assortment of choice tress with forbidden variance [6][7].

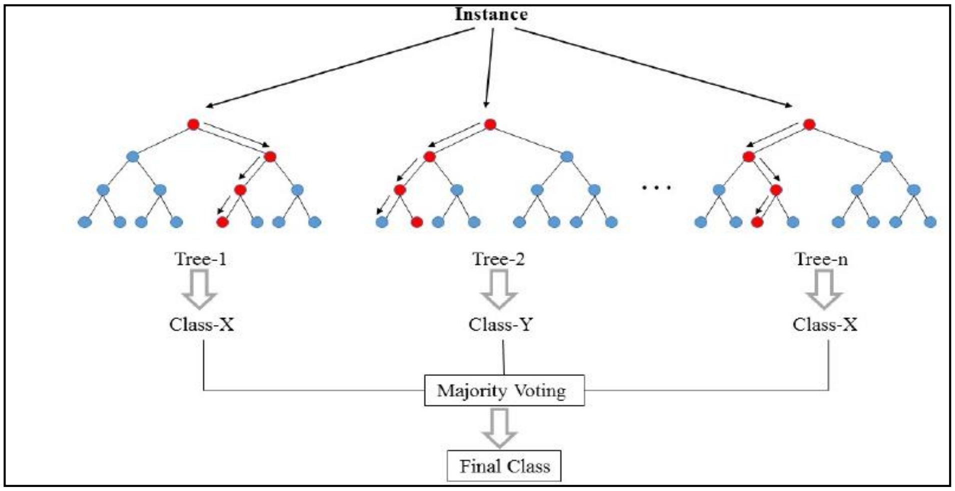


Figure 3. Illustration algorithm of random forest

2.2. XGBoost

Extreme Gradient Boost (XGBoost) is a performance of gradient boosted decision trees planned for rapidity and performance. Extreme boost algorithm is belongs to a children unit of boosting algorithms that exchange slow learners into healthy learners [8][9][10]. A slow learner is one which is slightly better than random guessing. Extreme gradient Boosting in the short form XGBoost. Boosting is a chronological procedure; i.e., trees are developed using the data from a formerly grown tree one after the further. This development gradually learns from data and tries to advance its forecast in successive repetitive.

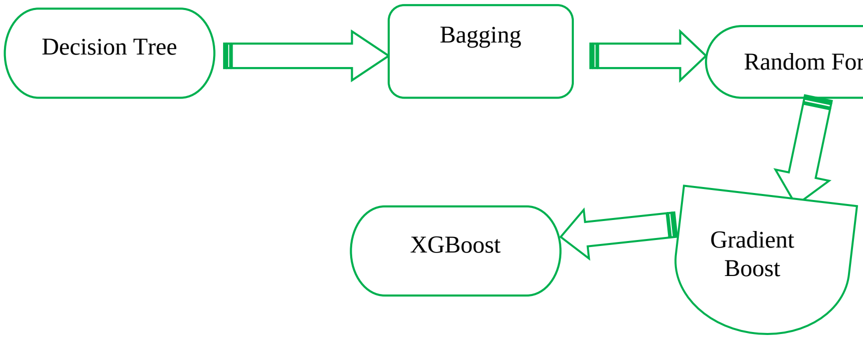


Figure 4. Transformation of decision tree to XGBoost

- A graphical representation of a possible solution to a decision based on certain conditions
- Bootstrapping aggregating or bagging is an ensemble meta algorithm combining predictions from multiple decision trees through a majority voting mechanism
- Bagging based algorithm where only a subset of features are selected at random to build a forest or collection of decision tree
- Models are built successively by minimizing the errors from prior models while growing or boosting the influence of high performing models

3. Implementation mechanism

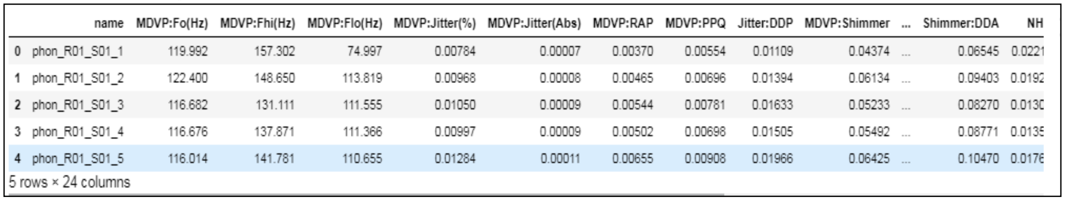
Here we are going to put into operation the machine learning algorithms to detect the disease of Parkinson's. During the implementation process, we are following the various stages as follows:

- Loading the dataset of Parkinson's disease and do the preprocessing operation, to eliminate the irrelevant samples or incorrect sample.
- After preprocessing, we need to choose the samples having a status of disease Zero or one (i.e Yes or NO)
- Make the samples normalize by applying the normalization method Min max method, to convert the values with the range 1 to -1
- Split the data set into two sets Train & Test sets
- Apply the Random forest algorithm using respective libraries from Python
- Apply the XGBoost algorithm using respective libraries from Python

Based on the above procedure, real time implementation was done stepwise on Kaggle platform.

Step 1: Fetching the dataset

```
# Data Set Reading
df=pd.read_csv ('/kaggle/input/parkinsons.data')
df.head ()
```

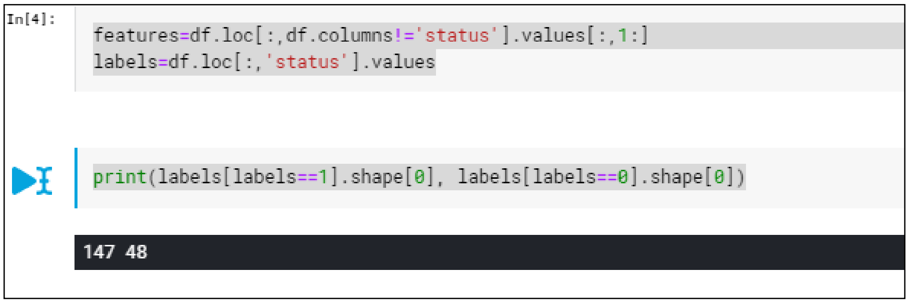


	name	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shimmer	...	Shimmer:DDA	NH
0	phon_R01_S01_1	119.992	157.302	74.997	0.00784	0.00007	0.00370	0.00554	0.01109	0.04374	...	0.06545	0.0221
1	phon_R01_S01_2	122.400	148.650	113.819	0.00968	0.00008	0.00465	0.00696	0.01394	0.06134	...	0.09403	0.0192
2	phon_R01_S01_3	116.682	131.111	111.555	0.01050	0.00009	0.00544	0.00781	0.01633	0.05233	...	0.08270	0.0130
3	phon_R01_S01_4	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505	0.05492	...	0.08771	0.0135
4	phon_R01_S01_5	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	0.06425	...	0.10470	0.0176

5 rows × 24 columns

Figure 5. UCI ML Parkinson's dataset with 195 samples and 24 fields

Step 2: Feting the et the features and labels with Status Zeros(0) and Ones(1):



```
In[4]:
features=df.loc[:,df.columns!='status'].values[:,1:]
labels=df.loc[:,'status'].values

print(labels[labels==1].shape[0], labels[labels==0].shape[0])

147 48
```

Figure 6. Samples with 147 ones and 48 zeros as status fields

Step 3: Normalization with Min_max method to scale the feature between 1 to -1 range

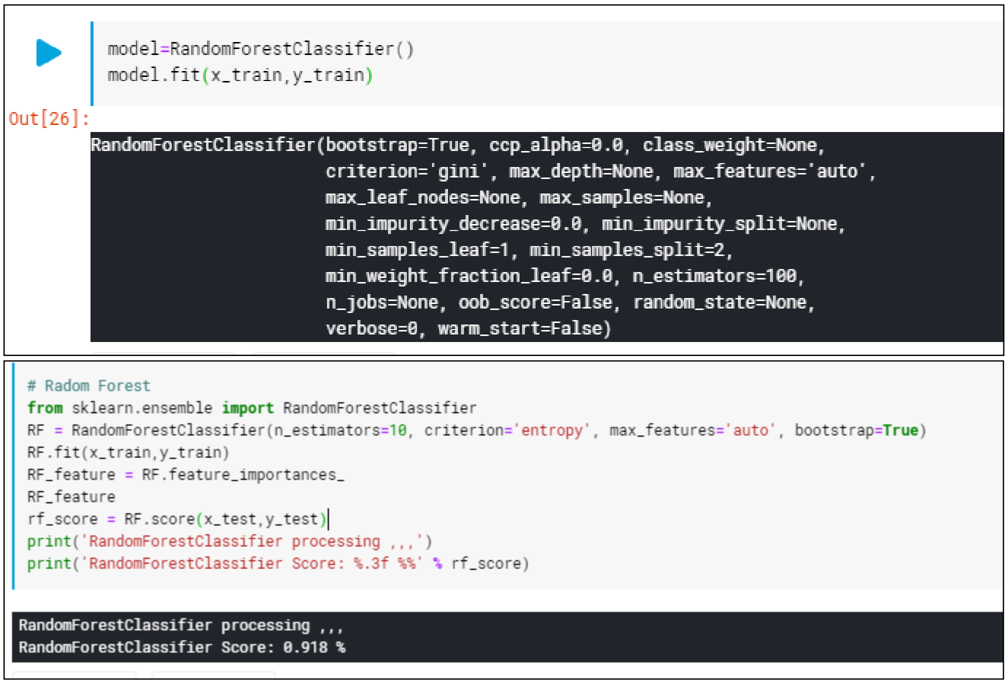
Norm=MinMaxScaler((-1,1))
x=scaler.fit_transform(features)
y=labels

Step 4: Make the dataset splits as train and test set with a ratio of 75 and 25%. The below code shows the data set split with specified portions.

x_train,x_test,y_train,y_test=train_test_split(x, y, test_size=0.25, random_state=7)

Step 5: Implementing Random Forest

Using random forest algorithms applying on training and test data, the performance was 91.8 %



```
model=RandomForestClassifier()
model.fit(x_train,y_train)

Out[26]:
RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=None,
                       criterion='gini', max_depth=None, max_features='auto',
                       max_leaf_nodes=None, max_samples=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, n_estimators=100,
                       n_jobs=None, oob_score=False, random_state=None,
                       verbose=0, warm_start=False)
```

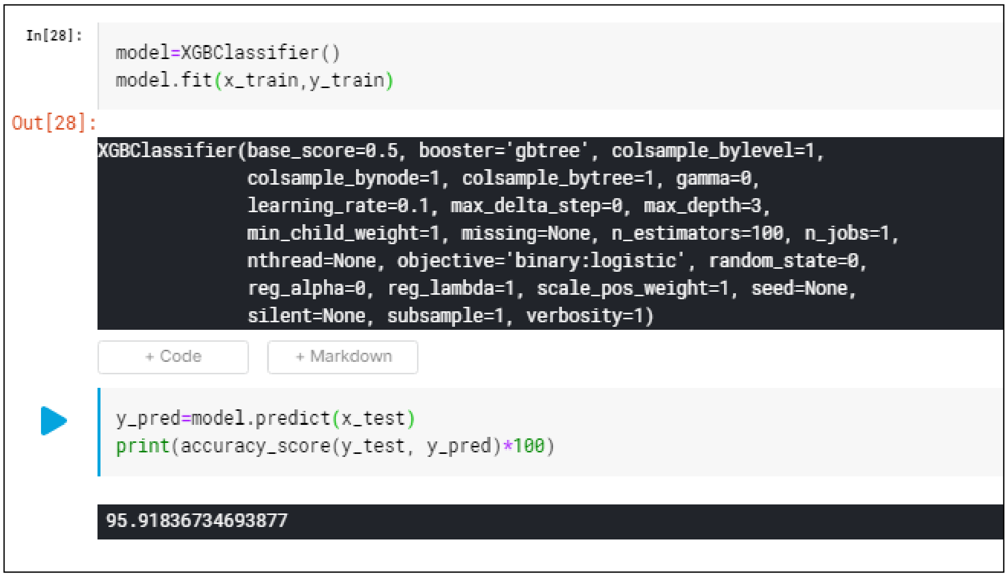
```
# Radom Forest
from sklearn.ensemble import RandomForestClassifier
RF = RandomForestClassifier(n_estimators=10, criterion='entropy', max_features='auto', bootstrap=True)
RF.fit(x_train,y_train)
RF_feature = RF.feature_importances_
RF_feature
rf_score = RF.score(x_test,y_test)
print('RandomForestClassifier processing ,,,')
print('RandomForestClassifier Score: %.3f %%' % rf_score)

RandomForestClassifier processing ,,,
RandomForestClassifier Score: 0.918 %
```

Figure 6. Random forest algorithm implementation

Step 6: Implementing XGBoost Algorithm

Using XGBoost algorithms applying on training and test data, the performance was 95.9%



```
model=XGBClassifier()
model.fit(x_train,y_train)
```

```
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
              colsample_bynode=1, colsample_bytree=1, gamma=0,
              learning_rate=0.1, max_delta_step=0, max_depth=3,
              min_child_weight=1, missing=None, n_estimators=100, n_jobs=1,
              nthread=None, objective='binary:logistic', random_state=0,
              reg_alpha=0, reg_lambda=1, scale_pos_weight=1, seed=None,
              silent=None, subsample=1, verbosity=1)
```

```
y_pred=model.predict(x_test)
print(accuracy_score(y_test, y_pred)*100)
```

```
95.91836734693877
```

Figure 7. XGBoost algorithm implementation

4. Summary

In this paper, we had elaborated about the disease called Parkinson's, the symptoms and behavior of it. How the machine learning algorithms are useful towards the detection of disease. We have described about Random forest and XGBoost algorithms and performance analysis. Form this study, we understood to detect the cause of Parkinson's disease from the data set with individuals using various factors. We implemented both random forest and XGB Classifier. Random forest gave an accuracy of 91.8 %, and XGBoost gave an accuracy of 95.92%. From this analysis, we understood XGboost algorithm gives the best results to detect the disease.

References

[1] D. Ahmadi Rastegar, N. Ho, G.M. Halliday, and N. Dzamko, "Parkinson's progression prediction using machine learning and serum cytokines," NPJ Parkinsons Dis', **(2019)** DOI: 10.1038/s41531-019-0086-4

[2] W. Dong Yue and W. Taeg Keun, "An automatic diagnostic algorithm for parkinson's disease based on deep learning," Asia-Pacific Journal of Neural Networks and Its Applications, vol.2, no.2, pp.13-18, **(2018)** DOI: 10.21742/AJNNIA.2018.2.2.03

[3] P. Spyros, M. Georgia, and E. Albert, "Digital health revolution: Is it time for affordable remote monitoring for Parkinson's disease?" Frontiers in Neurology, vol.6, pp.34, **(2015)** DOI: 10.3389/fneur.2015.00034

[4] Types of classification algorithms in Machine Learning, https://medium.com/@Mandysidana/machine-learning-types-of-classification-9497bd4f2e14

[5] S. Zehra, "Early diagnosis of Parkinson's disease using machine learning algorithms," Medical Hypotheses, vol.138, **(2020)**

[6] L. Hyunsook Zin, "A case study of political analysis of chronic disease management system at an outpatient clinic in Korea," International Journal of Bio-Science and Bio-Technology, vol.7, no.4, pp.117-126, **(2015)** http://dx.doi.org/10.21742/IJBSBT.2015.7.4.12

[7] S. Priyamwada, "Prediction of heart disease using 2-tier SVM data mining algorithm," International Journal of Advanced Research in Big Data Management System, vol.1, no.1, pp.11-24, **(2017)** DOI: 10.21742/ijarbms.2017.1.1.02

[8] C. Young Sung, M. Song Chul, R. Kwang Sun, and R. Keun Ho, “A study on clinical and healthcare recommending service based on cardiovascular disease pattern analysis,” International Journal of Bio-Science and Bio-Technology, vol.8, no.2, pp.287-294, **(2016)** DOI: 10.21742/IJBSBT.2016.8.2.27

[9] P. S Kumar and B. Debnath, “A computational intelligence method for effective diagnosis of heart disease using genetic algorithm,” International Journal of Bio-Science and Bio-Technology, vol.8, no.2, pp.363-372, **(2016)** DOI: 10.21742/IJBSBT.2016.8.2.34

[10] Kumar N 2018, Algorithms Every Machine Learning Expert Should Know, https://www.marktechpost.com/2018/11/26/algorithms-every-machine-learning-expert-should-know/