Detection of Parkinson's Disease using Machine Learning Algorithm

Shikha Singh
RajarshiShahu College of Engineering, Tathawade, Pune

Priti Sarote
RajarshiShahu College of Engineering, Tathawade, Pune

Nikita Shingade
RajarshiShahu College of Engineering, Tathawade, Pune

Deepti Yelale
RajarshiShahu College of Engineering, Tathawade, Pune

Nihar Ranjan
RajarshiShahu College of Engineering, Tathawade, Pune

ABSTRACT

Parkinson's disease is a condition in which dopamine-producing cells in the brain die. Parkinson's disease symptoms appear as the amount of dopamine in the brain diminishes. Parkinson's disease is a slow-progressing condition with symptoms such as tremors in the hands, arms, legs, chin, and face that get worse with time. People may have trouble walking and speaking as the condition advances. Although there is no cure for Parkinson's disease, the symptoms of the disease can be alleviated with the use of some medications. There are a number of common symptoms that may or may not suggest that the patient has Parkinson's disease. In this study, a new rating system was developed to aid in determining the severity of Parkinson's disease. However, a person with identical symptoms does not necessarily have Parkinson's disease. Because Parkinson's disease is an unsolved problem, the study focuses

On relevant aspects, medicines, and common approaches used to identify or assess the disease. Patients with Parkinson's disease often experience voice difficulties in the early stages of the condition. As a result, recent investigations for the identification of Parkinson's disease have focused on diagnosis systems based on voice disturbances.

Keywords

Parkinson Disease, Dopamine, Ensemble Learning, Boosting

1. INTRODUCTION

How does machine learning function, and what does it entail? Machine learning (ML) is an artificial intelligence (AI) technique that enables software to improve prediction accuracy without being particularly designed to do so.. Machine learning algorithms use historical data as input to forecast newly introduced output values. Parkinson's disease (PD) can be difficult to diagnose, especially in its early stages, because the symptoms of other neurologic conditions might be confusing. Motor signs such as bradykinesia (slowed movement and loss of spontaneous movement), muscle rigidity, a resting tremor, and postural instability are used to make the current diagnosis (balance issues).After Alzheimer's disease, Parkinson's disease (PD) is the second most common neurological disease .In general, there are two types of PD symptoms: motor and non-motor symptoms. Tremor, bradykinesia, stiffness (rigidity), and impaired balance are the main motor symptoms of Parkinson's disease (postural instability). Mood problems, cognitive dysfunction, pain, sensory dysfunction, and dysautonomia are the most common non-motor symptoms .Patients with Parkinson's disease frequently experience motor speech problems. More than half of the patients have speech problems, such as very quiet and rushed speaking .Speech signal analysis is a popular non-invasive way for diagnosing Parkinson's disease. Clinicians and neuroscientists are interested in noninvasive PD detection and prediction technology. Furthermore, detecting speech changes in Parkinson's patients would allow for early detection and intervention before the onset of disabling physical symptoms, which would have a significant impact on both patient healthcare system and patient life span as well as quality of life.

2. LITERATURE SURVEY

[1] In this paper, authors have introduced a system which is useful for detection of Parkinson's disease. Parkinson's disease is a central nervous system condition that affects the body's motor processes. It's a long-term illness with symptoms that worsen over time. It usually affects the elderly, whose symptoms steadily worsen until they reach a peak. Hearing, walking, speech, and other basic bodily functions can all be affected by the condition. Generic machine learning methods that provide varied degrees of accuracy can be used to analyse this disease. As a result, the best one is picked, as it will provide the maximum level of accuracy in predicting whether or not the disease is present in the patient.

[2]In this paper, authors have introduced a system which is useful for detection of Parkinson's disease. Different approaches to optimizing the creation of machine learning classification models for the early identification of Parkinson disease were investigated in this study. The goal was to use feature selection techniques to sort the medical measures and select the most relevant characteristics in order to develop a faster and more accurate model. Reducing the number of characteristics used to develop a model may result in a more efficient machine learning method, allowing doctors to focus on the most relevant measurements to consider. We compared the Filter and Wrapper approaches for feature selection. Then, by calculating the crossover scores for each strategy, we chose a solid machine learning algorithm to predict which technique could aid us.

[3] In this paper, authors have introduced a system which is useful for detection of Parkinson's disease. In this paper, an unique detection approach is proposed. Considering the qualities gathered from the speech of Parkinson's disease signals. Early detection and diagnosis of Parkinson's disease are critical. In terms of illness progression and treatment, this is critical. The dataset for Parkinson's disease used in this investigation was taken from the machine learning repository at UC Irvine. The hybrid.

[4]In this paper, authors have introduced a system which is useful for detection of Parkinson’s disease. So, in this research, we present a new approach for selecting feature sets by comparing performance metrics with other feature sets, such as original feature sets and Principal component Analysis-based feature reduction techniques. To compare the performance indicators, we employed a non-linear based categorization technique. Based on the data, this analysis will assist clinicians in distinguishing the PD group from the healthy group.

[5]In this paper, authors have introduced a system which is useful for detection of Parkinson’s disease. A neural technique based on employing Long-Short Term Memory (LSTM) neural networks is proposed in this paper to diagnose individuals with Parkinson's disease. The temporal patterns of the gait cycle are different for healthy people and sick, according to this study. As a result, the suggested method extracts temporal patterns to distinguish patients from healthy people using a recurrent structure like LSTM, which can assess the dynamic character of the gait cycle. The data used to determine the temporal characteristics of the gait cycle is based on shifting vertical Ground Reaction Force (vGRF), which is recorded by 16 sensors in each subject's shoe soles.

[6] In this paper, authors have introduced a system which is useful for detection of Parkinson’s disease. This necessitates the use of a machine learning model for the early diagnosis of Parkinson's disease. The examination of existing computational intelligence strategies in the field of research employed for PD detection is a precondition for discovering a full proof model. Many present models either focus on a single modality or analyse multiple modalities in a superficial manner. This prompted us to conduct a comparative literature review of four primary modalities

[7] In this paper, authors have introduced a system which is useful for detection of Parkinson’s disease.In this study, a new rating system was developed to aid in determining the severity of Parkinson's disease. However, a person with identical symptoms does not necessarily have Parkinson's disease. Because Parkinson's disease is an unsolved problem, the study focuses on relevant aspects, medicines, and common approaches used to identify or assess the disease. To address this issue, several methodologies will be employed to research and analyze the early diagnosis of Parkinson's disease. It can be examined with the use of a thorough knowledge of Parkinson's disease. However, the occurrence of several frequent symptoms has not yet been sufficiently characterized to assess the severity of Parkinson's disease.

[8] In this paper, authors have introduced a system which is useful for detection of Parkinson’s disease.In magnetic resonance imaging, an intensity-based texture segmentation technique for detecting regions with anomalous texture features is provided. Our algorithm was tested on many images from The Parkinson's Progression Markers Initiative (PPMI-database), and the findings show that it is acceptable for the successful identification and extraction of regions of interest whose attributes may be associated to signature traits of Parkinson disease.

[9] In this paper, authors have introduced a system which is useful for detection of Parkinson’s disease. Author used the MATLAB framework to analyze the frontal and temporal EEG of Parkinson's disease patients in this paper. Within a particular time frame, the Lyapunov exponent and inverse Lyapunov exponent for both PD and healthy participants were determined. The Lyapunov exponent for the temporal region of the brain is less than that for the frontal, and the inverse Lyapunov exponent is reverse order of the brain in PD patients. This means that in healthy people, the correlations between neurons are higher in the frontal region than in people with Parkinson's disease.

[10] In this paper, authors have introduced a system which is useful for detection of Parkinson’s disease. The difficulty of distinguishing between Essential and Parkinson's tremor is discussed in this article. A clinical study was conducted to achieve this goal, in which a group of volunteers, including Essential and Parkinson's tremor patients, underwent a series of pre-defined motion patterns, during which a wearable sensing setup was used to measure their lower arm tremor characteristics from multiple points .To provide a comparative study and evaluate the potential of using machine learning to accurately identify between different tremor types, extracted features from the acquired accelerometer signals were used to train classification algorithms.

3. EXISTING SYSTEMS

Clinicians usually make clinical decisions based on their intuition and experience, rather than using the database's knowledge-rich data. Unintentional biases, errors, and extravagant medical expenses are all consequences of this strategy, all of which have an impact on the quality of care provided to patients. As part of the Internet of Things, wearable technologies are being used. Handwriting was employed as a marker for Parkinson's disease diagnosis using a support vector machine. Using 3D visualization techniques to provide a simple tool for assessing the performance of Parkinson's disease patients Using data mining techniques, visually assisted tracking of PD patients' performance.

4. PROPOSED SYSTEM

Parkinson's disease (PD) is a neurological disease that has progressed to an advanced stage. In the early stages of Parkinson's disease, roughly 90% of people with the disease have speech problems. As a result, speech features were used to classify this condition in this study. Jitter, shimmer, basic frequency parameters, harmonicity parameters, Recurrence Period Density Entropy (RPDE), Detrended Fluctuation Analysis (DFA), and Pitch Period Entropy are some of the most well-known speech aspects employed in PD research (PPE). Those characteristics were dubbed baseline characteristics in this study. Refer Figure 1 for the proposed system model.

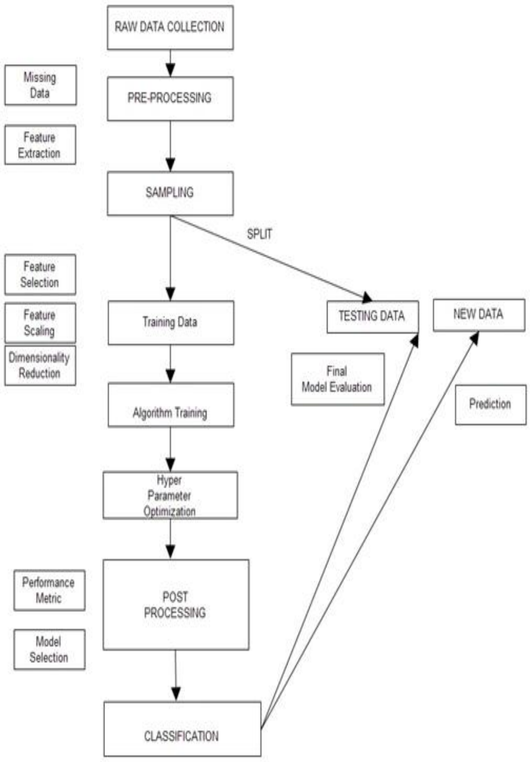

Figure 4. Proposed system

5. DATASET

The dataset used here combines 22 vocal features of 196 patients . This dataset is taken as input which consists of vocal features of PD infected patient as well as healthy person. The proposed system makes use of Machine learning ensemble method XGBOOST as it provides better accuracy irrespective of size of the dataset .Table 1 shows as the attributes of our dataset with their description.

Table 1. Dataset Description

ATTRIBUTE	DESCRIPTION
MDVP:Fo (Hz)	Average vocal fundamental frequency
MDVP:Fhi (Hz)	Maximum vocal fundamental frequency
MDVP:Flo (Hz)	Minimum vocal fundamental frequency
MDVP:Jitter(%) MDVP:Jitter(Abs) MDVP:RAP MDVP:PPQ Jitter:DDP	several measures of variation in fundamental frequency
MDVP:ShimmerMDVP:Shimmer(dB) Shimmer:APQ3 Shimmer:APQ5 MDVP:APQ Shimmer:DDA	Several measures of variation in amplitude
PDE,D2	Two nonlinear dynamical complexity measures
NHR, HNR	Two measures of ratio of noise to tonal components in the voice
DFA	Signal fractal scaling exponent
Spread1, spread2,PPE	Three nonlinear measures of fundamental frequency variation.
Status	Health status of the subject (one) - Parkinson's, (zero)-healthy

6. DATASET ANALYSIS

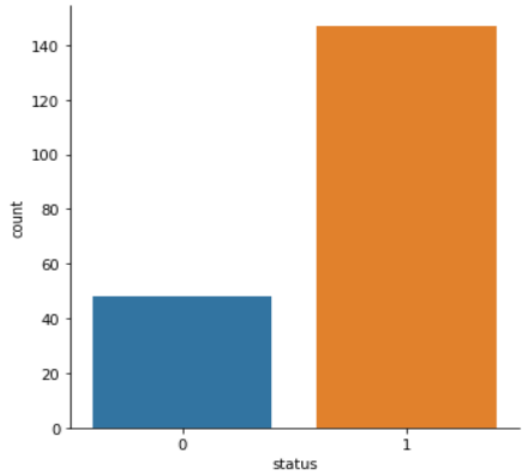

Figure 2. Status vs count plot

Through Figure2 we can see that around 25% of the people in dataset are healthy while 75% of people are PD infected.Here the various graphs which show the relation of vocal features to the status of person being healthy or not.

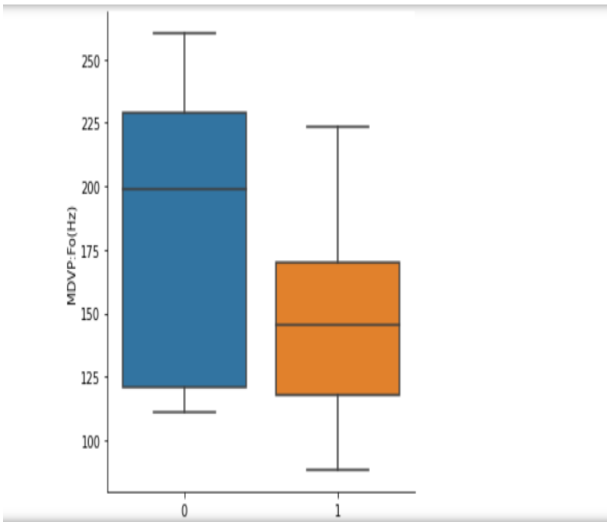

Figure 3. Status vs MDVP:FO(HZ) plot

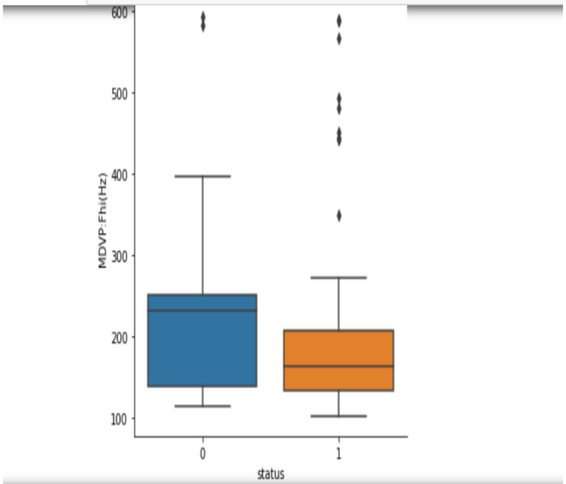

Figure 4. Status vs MDVP:FIH(HZ) plot

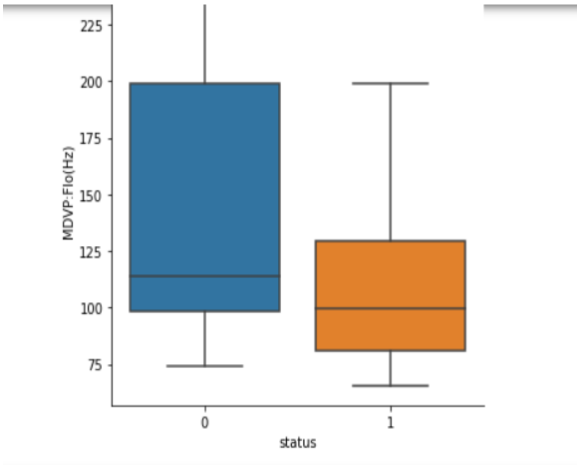

Figure 5. Status vs MDVP:FLO(HZ) plot

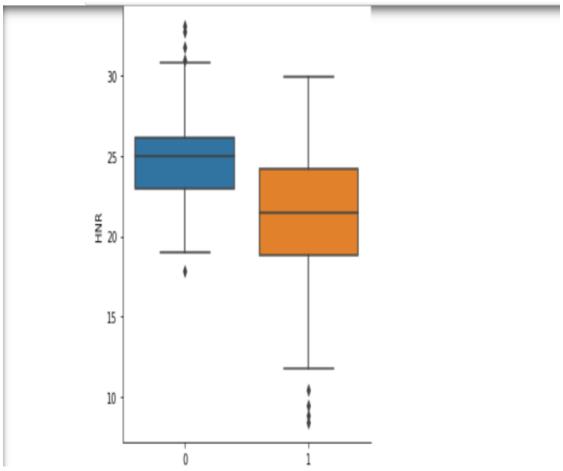

Figure 6. Status vs HNR plot

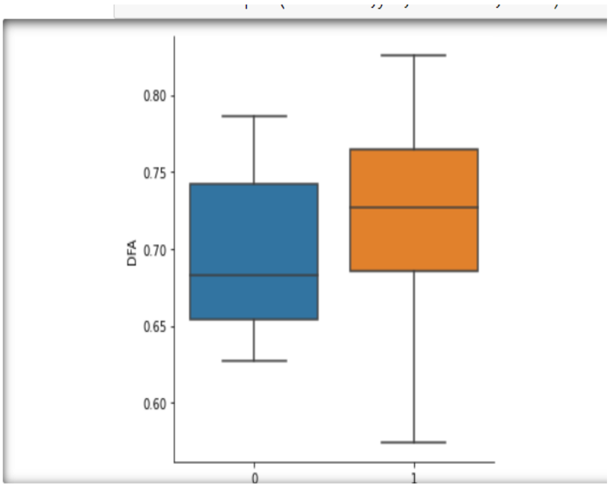

Figure 7. Status vs DFA plot

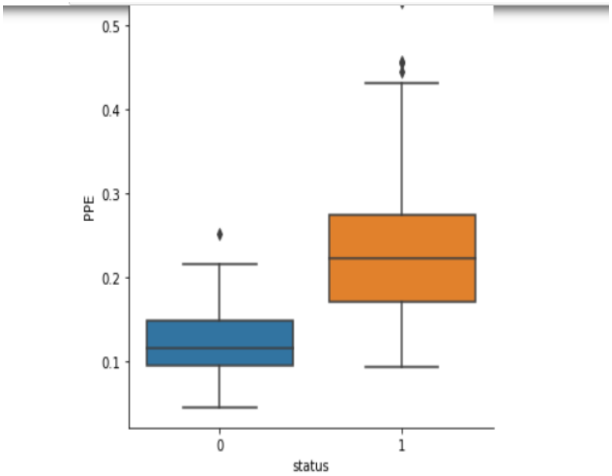

Figure 8. Status vs PPE plot

Refer Figure3, Figure4 and Figure5 to visualize how the fundamental vocal frequencies vary for a healthy person and a PD infected person.

Figure 6 shows the variation of ratio of tonal components in voice to noise for healthy and PD infected people.

Figure7 visualizes Signal fractal scaling exponent to the health status of subject whereas Figure8 shows PPE vs status.

7. ALGORITHMS

7.1 Ensemble learning

Ensemble learning is a machine learning process where various multiple models are combined to solve one computation. Using ensemble learning can improve the performance of a model .Commonly used ensemble learning methods are Boosting, Bagging, ADABoost etc.

7.2 XGBoost

As compared to other gradient boosting algorithms XGBoost is very much fast.XGBoost combines a convex loss function (based on the difference between the anticipated and target outputs) with a penalty term for model complexity to minimize a regularized (L1 and L2) objective function (in other words, the regression tree functions).The training process is repeated iteratively, with new trees being added that forecast the residuals or errors of previous trees, which are then integrated with previous trees to provide the final prediction. Gradient boosting gets its name from the fact that it uses a gradient descent approach to minimize loss when adding new models.

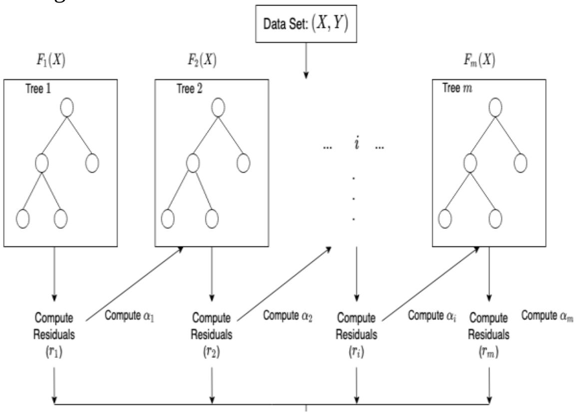

Figure 9. Diagrammatic representation of XGBoost Algorithm

8. PERFORMANCE PARAMETERS

Confusion matrix is a performance evaluation N X N matrix for classification model , where N is Number of target classes. For binary classification model we use 2 X 2 confusion matrix.

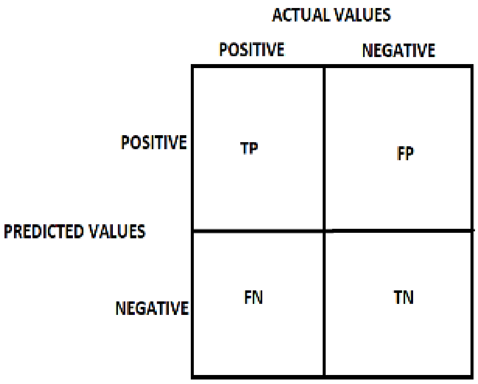

Figure 10. Confusion matrix

Columns represent actual values of target variable whereas rows represent predicted values of target variables. Target variables are either positive or negative

TP: True Positive predicted value matches actual value (actual value = True , predicted value = True)

FP: False Positive predicted value falsely predicted (actual value = False , predicted value = True)

TN: True Negative predicted value matches actual value (actual value = False , predicted value = False)

FN: False Negative predicted value falsely predicted (actual value = True , predicted value = False)

1] Accuracy = (TP + TN) / (TP + FP + TN + FN)

2] Precision = TP / (TP + FP)

3] Recall= TP / (TP + FN)

4] F1 Score = TP / (TP + 0.5(FP + FN)

9. RESULTS AND DISCUSSIONS

We calculated various performance parameters for the proposed model

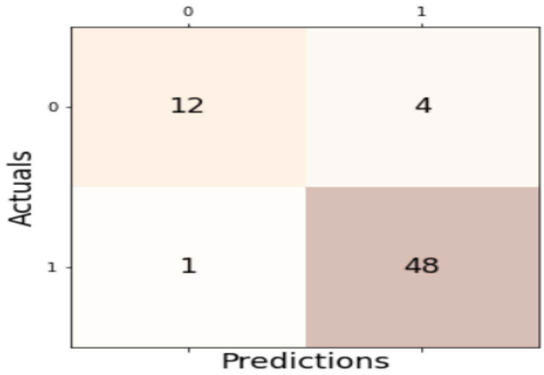

Figure 11. Confusion matrix for proposed model

Table 2. Obtained values of performance parameters for proposed model

Performance parameter	Accuracy	Precision	Recall	F1 score
Obtained value	92%	0.92	0.98	0.95

Table 3.Classification report

	Precision	**Recall**	**F1 Score**	**Support**
0	0.92	0.75	0.83	16
1	0.92	0.98	0.95	49
Accuracy			0.92	65
Macro avg.	0.92	0.86	0.89	65
Weighted avg,	0.92	0.92	0.92	65

10. CONCLUSION

Artificial intelligence and medical research have formed a partnership that aids in the treatment of ubiquitous disorders such as Parkinson's disease. For early detection of Parkinson's disease, symptoms such as Bradykinesia, Tremor at rest, Rigidity, and Voice Impairment can be noticed. There is no specific medical method or diagnosis for a person's parkinsonism, which also applies to bioinformatics. Strong techniques like Machine Learning, on the other hand, have sped up the process of detecting Parkinson's disease by making it more cost-effective and efficient. Machine learning can help doctors detect Parkinson's disease.

11. REFERENCES

[1] J. Jankovic, “Parkinson’s disease: Clinical features and diagnosis”, Journal of Neurology, Neurosurgery & Psychiatry, vol. 79, no. 4, pp. 368–376, 2008.

[2] M. Lawton, Y. Ben-Shlomo, M. T. May, and et al., “Developing and validating Parkinson’s disease subtypes and their motor and cognitive progression”, Journal of Neurology, Neurosurgery & Psychiatry, Dec.2018.

[3] R. Prashanth, S. Dutta Roy, P. Mandal, and S. Ghosh, “High-Accuracy Detection of Early Parkinson’s Disease through Multimodal Features and Machine Learning”, International Journal of Medical Informatics, vol. 90, Mar. 2016.

[4] NiharRanjan, Divyakumar et al. " Diagnosis of Parkinson Disease using Handwriting Analysis", International Journal of Computer Application, Volume 184, No. 1, pp13-16

[5] Z. Wang, A. R. Childress, J. Wang, and J. A. Detre, "Support vector machine learning-based fMRI data group analysis", NeuroImage, vol. 36, no. 4, pp. 1139–1151, 2007.

[6] Z Ghouse, NiharRanjan, N Hiwrale, "A multi-function robot for military application" Imperial Journal of Interdisciplinary Research, 2017

[7] G. Ma, L. He, C.-T. Lu, P. S. Yu, L. Shen, and A. B.Ragin, "Spatio- Temporal Tensor Analysis for Whole-Brain fMRI Classification", SDM, 2016.

[8] NiharRanjan, Rajesh S. Prasad " Author Identification in Text Mining for used in Forensics" ISSN 2321- 9637, Volume 1, Issue 5, December 2013, pp. 568-571

[9] D. R. Mehra, "Power Spectrum Estimation using Welch method for various Window techniques", International Journal of Scientific Research Engineering & Technology, vol. 2, pp. 389–392, Sep. 2013.

[10] A. Kakade, S. Jain, S. Das, N. M. Ranjan., "A neural network approach for text document classification and semantic text analytics", Journal of Data Mining and Management", Volume 2, Issue 3, 2017.

[11] SuyogGatkal, VinayakDhage, NiharRanjan et al., Healthware – A Smart GenerationHealthcare System" PENSEE Journal, Volume 51, Issue 7, pp 30-35, 2021.