ARTICLE

# ESTIMATION OF PARKINSON'S DISEASE RISK BY STATISTICAL MODEL

**S. Anita [1*], P. Aruna Priya [2]**

[1]*Department of Electronics and Communication Engineering, SRM University, Chennai, INDIA*

[2]*Faculty of Electronics and Communication Engineering, SRM University, Chennai, INDIA*

ISSN 0976-3104

## ABSTRACT

***Background:*** *The constant progress and complexity of clinical and non- clinical treatment of Parkinson's disease (PD) become very tough to diagnose the disease. This may lead to delayed diagnosis, misdiagnosis and excessive medical cost. Rapid advances in diagnostic techniques have offered an effective way for tracking the different stages of the disease. This paper focuses the early PD diagnosis and its progression is estimated by Gama Amino Butyric Acid (GABA) concentration level with the help of Striatal Binding Ratio (SBR) values. SBR values of Caudate and Putamen (left & right) are calculated from Single Photon Emission Computed Tomography (SPECT) images as treated as input variables and the response variable is GABA concentration level. The mathematical model for GABA concentration level is developed by using the input variables like SBR values of Caudate and Putamen (left & right). The performance of the model is analyzed using ANOVA (Analysis of variance), normal probability curve and residual plot. The coefficient of determination ($R^2$) gives 99.3% of fitness rate with the regression line. The progression rate of PD is measured for three consecutive years and it is compared with the threshold value of GABA to find the severity of the Parkinson's disease.*

**KEY WORDS**
*Parkinson's disease, SBR, level of GABA concentration, Mathematical model, progression rate*

Received: 17 March 2017
Accepted: 14 April 2017
Published: 10 June 2017

***Corresponding Author**
Email: sranitaa@gmail.com
Tel.: +91- 9787931123

## INTRODUCTION

Parkinson's disease is a progressive movement disorder which constantly affects the mid brain neurons of human called substantia nigra. It is clinically defined by the major symptoms of resting tremor, rigidity, postural instability, bradykinesia, cognitive and psychiatric disturbances.The diagnosis of PD is easy when symptoms are full blown. But an accurate diagnosis is quite tough when the disorder is mild, which demands the formulation of an early detection technique for PD [1-3].The inhibitory neurotransmitter dopamine regulates and controls movements, motivation and cognition. Degeneration of these dopamergic nerve cells along the nigrostraital pathway affects the gait system of human, which in turn leads to Parkinson's disease [4].

SPECT images take the crucial role to discriminate PD patients from the healthy group by calculating dopamine deficit in Caudate and Putamen of the midbrain, even in the premature stage of PD. Thus, calculating dopamine deficiency (ie.SBR) in Caudate and Putamen of the human brain from the SPECT image is a valuable diagnostic tool for discriminating Parkinson's disease from the healthy control [5, 6]. Still the cases have high SBR value misdiagnosed as PD(supposed to be low) and the cases have low SBR value misdiagnosed as healthy control (supposed to be high). Hence the misdiagnosed rate is significant [3].

Gamma-Amino Butyric Acid (GABA) is also a most essential inhibitory transmitter in the central nervous system (CNS) and spinal card. GABA mediates pre-synaptic inhibition of primary blood vessels in the motor neuron system. The disturbances of the GABA concentration level in CNS influence the motor system [7]. Several neurological disorders including Parkinson's disease, anxiety, depression, insomnia, and epilepsy are negatively related with the level of GABA concentration in the human brain [8]. Hence the neurotransmitters dopamine and GABA are found to be a novel diagnostic tool for detecting PD.

The model that the log of the odds states how the variables are related with prediction of the Parkinson's disease. Hence formulating the mathematical model is a new approach to diagnosis PD. In statistics regression analysis focuses on the relationship between a dependent variable and one or more independent variables. Primarily it supports to understand how the dependent variable changes when any one of the independent variable is varied while the remaining independent variables are held fixed. When the regression analysis has considerable overlap with the field of machine learning, it can predict the various disorders. It is also used to realize how the independent variables are related to the dependent variable, and to investigate the forms of these relationships [9].

The machine learning techniques such as Multivariate Logistic Regression (MLR), Artificial Neural Network (ANN), and Support Vector Machine (SVM) are effectively used to formulate a prediction model for diagnosing neural disorders. Machine learning techniques consent with individual level characterization rather than group level characterization. Hence high level of clinical translation is obtained. MLR aims at the determination of probability based on SBR values, which classify the subjects into different risk categories. SVM finds the hyper plane in order to classify the subjects and high accuracy was achieved [10-12].

IIOAB JOURNAL
ISSN: 0976-3104

In the present study, the level of GABA concentration of Parkinson's disease is calculated using SBR values. The developed mathematical model of GABA is to predict the early diagnosis of PD and its progression rate by using regression analysis. The developed model has potential to assist the clinician in the diagnostic process.

## COMPUTATIONAL METHODOLOGY

### *PPMI Database*

All PD subjects taken from the international PPMI database are in three stagesof Hoehn and Yahr (HY). The corresponding SBR values are taken for the analysis [13]. The SBR values are calculated as follows:
Iterative reconstruction was performed on SPECT raw projection data using hybrid ordered subset expectation maximization (HOSEM) algorithm. Iterative reconstruction was done without any filtering to ensure consistency of the reconstructions. The HOSEM reconstructed files were processed for Attenuation correction, which was filtered and normalized to get the same anatomical alignment. Striatal uptake count densities of the region of interest (ROI) were extracted and used to calculate striatal binding ratios (SBRs) for each region of the four striatal regions.SBR is calculated by PPMI as follows and compared with Occipital cortex region below the Putamen as reference region [14].

SBR= (target region/reference region) – 1 (1)

Where,
Target region: left caudate, right caudate, left putamen, right putamen.
Reference region: occipital cortex.

### *Statistical significance of SBR*

All statistical analysis was carried out using Minitab software with 5% significant level (95% of confidence level). Histogram plots were drawn for three consecutive years and it is presented in Fig.1. The plots illustrate the different stages of SBR values distribution for Caudate and putamen (both left and right) of PD Patients. It also shows the amount of overlap between the three consecutive years. The overlap is higher in Caudate (left and right) than Putamen. Higher the overlap, difficult to discriminate the progress rate of the disease [3]. Hence the classification tool plays a major role as they integrate all the characteristics of disease, train the model and categorize them accordingly.

### *The level of GABA concentration*

The gradual changes of neurochemical lead to PD in the human body. The GABA and dopamine in Caudate and Putamen is interrelated to measure progression of PD. From the literature [15] it is found that the threshold level of GABA for PD is 0.265µmol/g. If the measured GABA lies below 0.265µmol/g the cases are affected by PD. Similarly it lies above 0.265µmol/g it is a healthy control. The Level of GABA concentration is measured using radio receptor array.

### *Mathematical model*

The mathematical model for the level of GABA concentration is framed using regression analysis, for investigating the relationship between the variables called SBR and GABA. It demonstrates the relationship between GABA, Y (called as response or output or dependent variable) and four SBR values $X_1 \ldots\ldots X_p$ (called as predictor or input or independent or explanatory variables).

The linear regression equation for GABA is:

$$Y = a + bX \quad (2)$$

Where, Y= Level of GABA; X= SBR used to predict GABA; a = the intercept; b = the slope.

Here, regression is used to estimate the qualitative effects of the variables, namely Caudate (left, right) and Putamen (left, right) upon the variable GABA [16]. The standard error (SE) of regression indicates that the observations are closer to the fitted line, and the following equation calculates it.

$$\text{Standard error (SE) of regression} = \frac{\sqrt{\Sigma(yi - \hat{y}i)2}}{\Sigma(xi - x)2} \quad (3)$$

Where,

yi is the value of the GABA, ŷi is estimated value of GABA, xi is the observed value of the SBR, x is the mean of the SBR, and n is the number of observations.

IIOAB JOURNAL
ISSN: 0976-3104

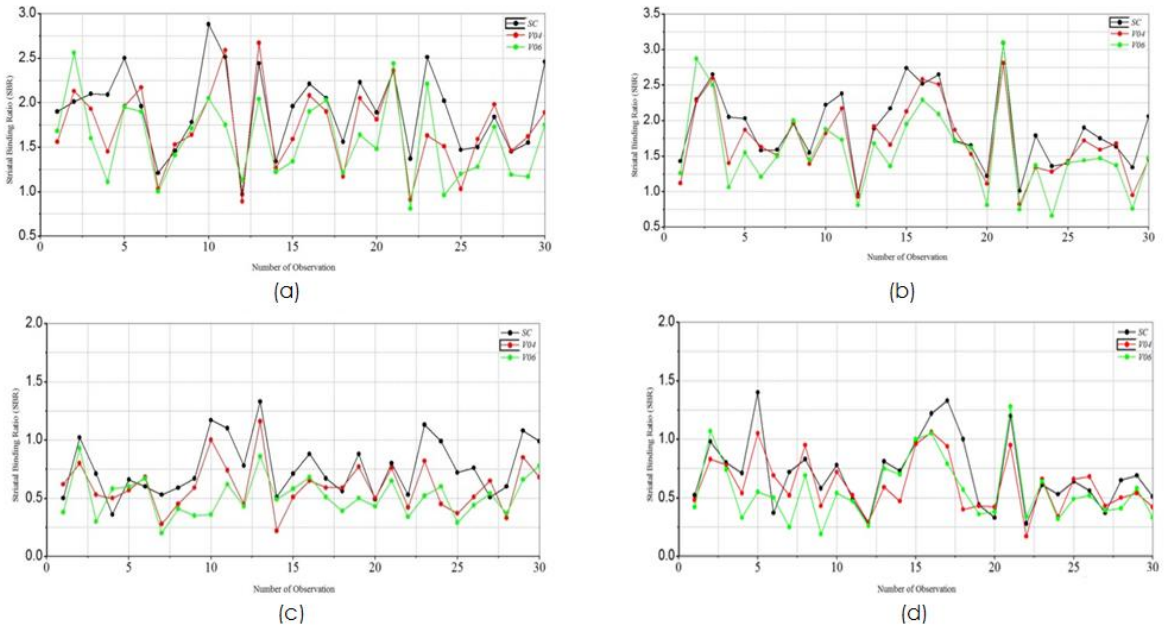


**Fig. 1:** Histogram plots of the striatal binding ratio (SBR) values for right caudate (a), left caudate (b), right putamen (c) and left putamen (d) for PD.

The p-value for each term tests the null hypothesis that the coefficient is equal to zero (no effect). A low p-value (< 0.05) indicates that the model is significant. And a high P-value (>0.05) indicates that the model is insignificant. The coefficient of determination (R-squared) is a statistical measure of how close the data are in the fitted regression line. R-squared is always between 0 and 100%. 0% indicates that the model has none of the GABA variables around its mean. 100% indicates that the model has all the GABA variables around its mean. In general, higher the R-squared, better the model that fits with the data.

## RESULTS AND DISCUSSION

### *Formulation of the mathematical model for the level of GABA concentration*

The coefficients of regression analysis of the level of GABA concentration for PD are shown in table 1 along with their P value of the parameters. It is observed that the P value of the level of the GABA concentration for Caudate (L) and Putamen (L) are most significant, whereas Caudate (R) and Putamen (R) are not so significant. It indicates the reliability of the model.

**Table 1:** Prediction time of parallel machines and prediction accuracy

| Predictor | B | SE | T | P value |
|---|---|---|---|---|
| Constant | 0.114468 | 0.005631 | 20.33 | 0.00 |
| Caudate(R) | 0.008960 | 0.004785 | 1.87 | 0.062 |
| Caudate(L) | 0.025654 | 0.004956 | 5.18 | 0.000 |
| Putamen(R) | -0.005756 | 0.006367 | -0.90 | 0.367 |
| Putamen(L) | -0.028090 | 0.007518 | -3.74 | 0.000 |

B is a regression coefficient for the predictors; SE is its standard error; T is test statistics; P value is the significance of the regression coefficient.

Where S is the estimated standard deviation about the regression line, R-squared is the coefficient of determination. Adjusted R-squared is an approximately unbiased estimate of the population R-squared.
The S value is the measurement of error.The model is better if it is smaller. The higher value of R-squared is better to determine the coefficients of a regression equation. The closeness of the adjusted R-squared with R-squared determines the fitness of the model [17]. In both cases, the adjusted R-squared value is closer to the R-squared value is shown in table 2.
The mathematical model given in equation 4 is framed from the coefficient (B) value of the regression table 1.

$$\text{GABA (PD)} = 0.114 + 0.00896 \times \text{Caudate(R)} + 0.0257 \times \text{Caudate (L)} - 0.00576 \times \text{Putamen (R)} - 0.0281 \times \text{Putamen (L)} \quad (4)$$

It is observed from the model that the GABA is positively related with SBR values of left and right caudate and negatively related with left and right putamen. Hence high value of putamen has less likely have PD.

ISSN: 0976-3104

**Table 2:** Summary of regression analysis

| Responses | S value | $R^2$ (%) | Adjusted $R^2$ (%) |
|---|---|---|---|
| GABA for PD | 0.0248413 | 99.3 | 98.7 |

S is the standard deviation; $R^2$ is the coefficient of determination; Adjusted $R^2$modified version of $R^2$

A high value of the determination coefficient (R2) confirms model adequacy, the goodness of fit and high significance of the model. This indicates that the regression models for the output can be used for determining and estimating GABA for PD.

### *Overall model evaluation*

An analysis of variance (ANOVA) was performed for GABA to evaluate the model performance which is presented in table 3. The associated p-value for the model is lower than 0.05 (i.e. level of significance α=0.05, or 95% confidence), which indicates that the model can be considered statistically significant better than a null model.

**Table 3** Analysis of Variance for PD

| Source | DF | SS | MS | F | P value |
|---|---|---|---|---|---|
| Regression | 4 | 0.059270 | 0.014818 | 24.01 | 0.000 |
| Residual Error | 355 | 0.219067 | 0.000617 | | |
| Total | 359 | 0.278338 | | | |

DF is the degrees of freedom;SS is Sum of Squares; MS is Mean Squares;F calculated F value;P is a significance of regression coefficient

The normal probability plot for GABA is presented in Fig.2. It can be noticed that the residuals fall on a straight line, which means that the errors are normally distributed and the regression model is well fitted with the observed values. Fig. 3 shows the residual values with the fitted values for GABA. It indicates that the maximum variation of -0.075 to 0.050, which shows the high correlation that exists between fitted values and observed values.

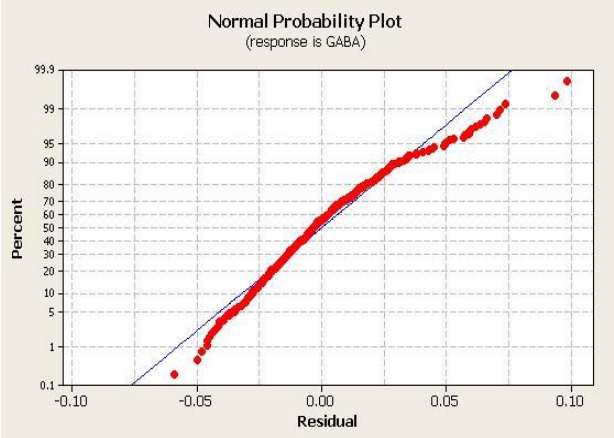


**Fig. 2:** Normal probability plot for GABA for PD

ISSN: 0976-3104

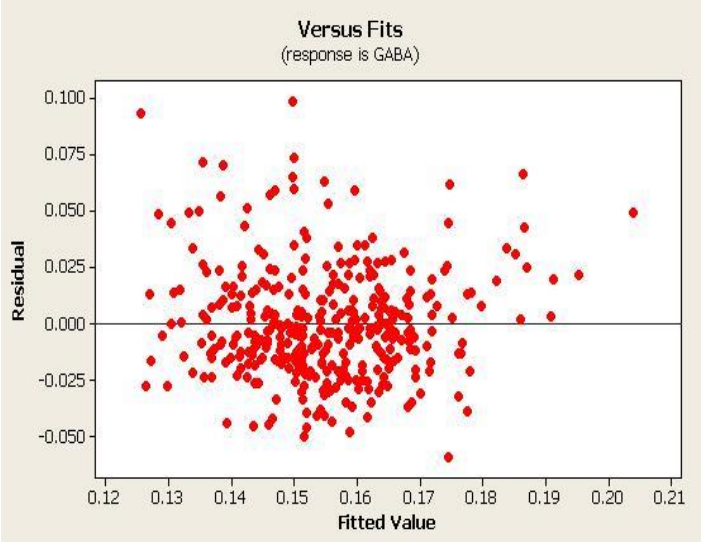


**Fig. 3:** Residual Vs fitted values for GABA for PD

### *Estimation of PD risk*

The progression rates of PD for the three consecutive years are compared and its risk is shown in Fig.4. It shows less overlap between the GABA values, which has high discriminative power between them. The notched box plot shown in Fig.5 gives the outliers of the PD progression rate. Since all values lie between the ranges of the level of GABA concentration, it is evident that the high performance is observed from the developed model.

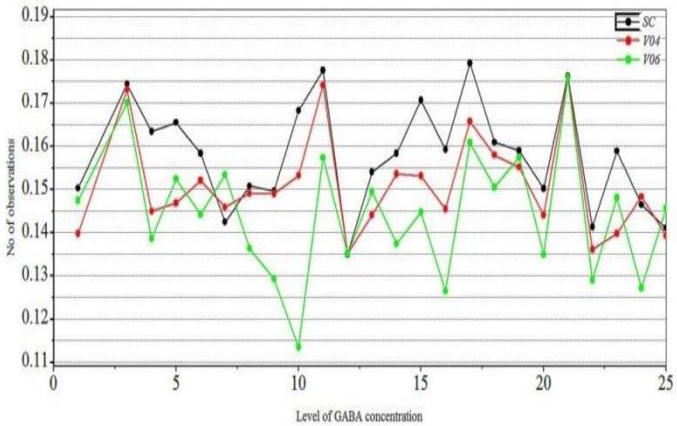


**Fig. 4:** Histogram plot for the level of GABA concentration

IIOAB JOURNAL
ISSN: 0976-3104

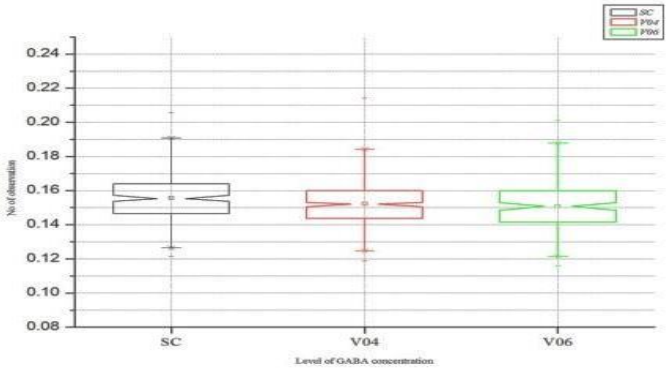


**Fig. 5:** Box plot for the level of GABA concentration

In each notched box plot, the central mark is the median (q2), the edges of the box are the 25th (q1) and 75th (q3) percentiles, the whiskers extend to the most extreme data points that are not considered outliers, and outliers are plotted individually. The extremes of the notches or the centers of the triangular markers correspond to q2 ± 1.57(q3 _ q1)/√n where n = 674 is the number of observations.

The SPECT images are again taken from the same patient for three consecutive years. The measured SBR values are applied to predict the early PD and its progress rate. The same procedure has been adopted for every three consecutive years to estimate the progression rate of PD. The Fig. 6 shows the progression rate of PD for three consecutive years. From the graph it is evident that the averaged GABA concentration level is increased for every year such as first year 0.1556, second year 0.15524 and the third year 0.1502.

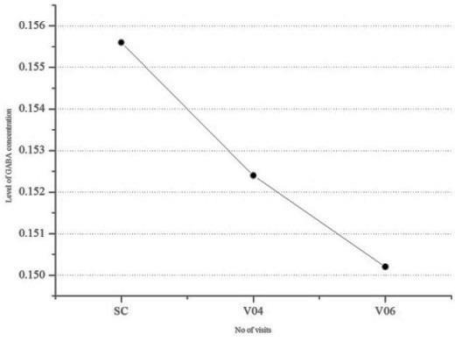


**Fig. 6:** Progression rate of PD

## CONCLUSION

The diagnosis of early PD and its progression rate is estimated for three consecutive years. The predictive mathematical models for the level of GABA concentration of PD patients are developed by using Regression analysis. These models have the potential to discriminate PD from healthy control. The model performance was tested using ANOVA, normal probability curve and residual plot. The residual plots of GABA for PD are generated and it is observed that regression model is well fitted and highly correlated with the observed values. The progression rate is measured for three years from the model; hence, the severity of the disease is estimated. The inference is that the prediction models for estimating GABA concentration level is a novel method to aid the clinicians for diagnosing PD. It overcomes the misdiagnoses of PD with high rate of accuracy compared with the related work.

## CONFLICT OF INTEREST

There is no conflict of interest.

ISSN: 0976-3104

ACKNOWLEDGEMENTS
None

FINANCIAL DISCLOSURE
None

## REFERENCES

[1] de Lau, L. M and Breteler, M. (2006) 'Epidemiology of Parkinson's disease', The Lancet Neurology, vol. 5, pp.525-535.
[2] Moore DJ, West AB, Dawson VL, Dawson TM. "Molecular pathophysiology of Parkinson's disease." Annual Review Neuroscience 28 (2005): 57-87.
[3] Prashanth , R., Sumantra Dutta Roy Pravat and Mandal Shantanu Ghosh, K. (2014)'Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging',Exepert system with Applications vol. 41, pp.3333-3342
[4] T.c. Booth.M, Nathan.A.D.waldman.A.Mquigley, A.H.Schapuria and J. Buscombe "The role of functional dopamine transporter AJNR AMJ neuroradial 36:229-35, Feb 2015
[5] Bairactaris, C., Demakopoulos, N., Tripsianis, G., Sioka, C., Farmakiotis, D., Vadikolias. K., Heliopoulos, I., Georgoulias, P., Tsougos, I., Papanastasiou, I., & Piperidou, C. (2009). 'Impact of dopamine transporter single photon emission computed tomography imaging using I-123 ioflupane on diagnoses of patients with parkinsonian syndromes', Journal of Clinical Neuroscience, vol.16,pp. 246-252..
[6] Booij, J., Tissingh, G., Boer, G. J., Speelman, J. D., Stoof, J. C., Janssen, A. G., Wolters, E. C., & van Royen, E. A. (1997) ' [123I]FP-CIT SPECT shows a pronounced decline of striatal dopamine transporter labeling in early and advanced Parkinson's disease', Journal of Neurology, Neurosurgery & Psychiatry, vol. 62, pp.133-140.
[7] (2007)'American college for advancement in medicine', Monograph, Alternative Medicine Review, Vol. 12, pp-274-279.
[8] Acta Med Okayama and Kuroda H. (1983) Gamma-aminobutyric acid (GABA) in cerebrospinal fluid. vol.37 (3), pp. 67-77.
[9] Douglas C. Montogomery, Elizabeth A. Peck and G.Geoffrey Vining (2001) 'Introduction to Linear Regression Analysis' Arizona State University, fifth edition,576 pages.
[10] Orru, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neuroscience & Biobehavioral Reviews, 36, 1140–1152.
[11] Shubhambind, arvind Kumar Tiwari, Anil Kumar Sahani, (2015). A Survey of Machine learning Based Approaches for Parkinson Disease prediction,vol6 (2),1648-1655
[12] Winogrodzka, A., Bergmans, P., Booij, J., van Royen, E. A., Janssen, A. G., & Wolters, E.C. (2001). [123I]FP-CIT SPECT is a useful method to monitor the rate of dopaminergic degeneration in early-stage Parkinson's disease. Journal of Neural Transmission, 108, 1011–1019.
[13] Francisco P M Oliveira, and Miguel Castelo-Branco, (2015) 'Computer-aided diagnosis of Parkinson's disease based on [123I]FP-CIT SPECT binding potential images, using the voxels-as-features approach and support vector', J. Neural Eng. Vol. 12 (10pp) doi:10.1088/1741-2560/12/2/026008.
[14] Seibyl, J. Jennings, D.Grachev, I. Coffey, C and Marek, K , (2013) '123-I Ioflupane SPECT Measures of Parkinson Disease Progression in the Parkinson Progression Marker Initiative (PPMI) Trial', in Society of Nuclear Medicine Annual Meeting Abstracts, pp. 190.
[15] RJ. Abbott, IF. Pye, SR. Nahorski, (1982) 'CSF and plasma GABA levels in Parkinson's disease', Journal of Neurology, Neurosurgery, and Psychiatry, vol. 45, pp. 253-256.
[16] Julian J. Faraway, (2002) 'Practical Regression and ANOVA using R' first edition,210 pages.
[17] Palanikumar.K (2007) 'modeling and analysis for surface roughness in machining glass fibre reinforced plastics using response surface methodology', Mater Des,Vol. 28, pp.2611–2618