



PREDICTIVE MODELING IN MENTAL HEALTH: USING DATA SCIENCE TO DETECT EARLY SIGNS OF CLINICAL DEPRESSION AND ANXIETY

Dr Anju Chauhan

Research Scholar

mabhagwatikripa@gmail.com

DECLARATION: I AS AN AUTHOR OF THIS PAPER /ARTICLE, HERE BY DECLARE THAT THE PAPER SUBMITTED BY ME FOR PUBLICATION IN THE JOURNAL IS COMPLETELY MY OWN GENUINE PAPER. IF ANY ISSUE REGARDING COPYRIGHT/PATENT/OTHER REAL AUTHOR ARISES, THE PUBLISHER WILL NOT BE LEGALLY RESPONSIBLE. IF ANY OF SUCH MATTERS OCCUR PUBLISHER MAY REMOVE MY CONTENT FROM THE JOURNAL WEBSITE. FOR THE REASON OF CONTENT AMENDMENT /OR ANY TECHNICAL ISSUE WITH NO VISIBILITY ON WEBSITE /UPDATES, I HAVE RESUBMITTED THIS PAPER FOR THE PUBLICATION.FOR ANY PUBLICATION MATTERS OR ANY INFORMATION INTENTIONALLY HIDDEN BY ME OR OTHERWISE, I SHALL BE LEGALLY RESPONSIBLE. (COMPLETE DECLARATION OF THE AUTHOR AT THE LAST PAGE OF THIS PAPER/ARTICLE

ABSTRACT

Mental illnesses such as depression and anxiety are becoming common at an alarming rate in today's world. The current research explored the use of predictive models and data science methods in early diagnosis of mental illnesses using a sample size of 90 people. Psychological test questions along with behavioral factors such as stress, sleep disorders, mood swings, and anxiety symptoms were used to gather data. Various types of machine learning models like Logistic Regression, Decision Trees, Random Forest, and Support Vector Machine (SVM) were used for making predictions. According to results, high levels of stress, anxiety symptoms, and sleep problems were prevalent amongst the subjects. Amongst all the prediction models tested, Random Forest produced the most accurate results with a prediction accuracy of 93%.

Keywords: *Predictive Modelling, Mental Health, Depression, Anxiety, Data Science.*

1. INTRODUCTION

Mental health constitutes a vital aspect of well-being for all humans and helps contribute greatly to their psychological and social stability. Mental health issues like depression and anxiety have recently become a significant global health concern that has affected many people irrespective of their age and socioeconomic status. Lifestyle changes, increased academic/occupational pressure, isolation, financial pressure, and excessive digital interaction have been among the factors that have been contributing immensely to the increase in psychological issues among many people. Depression, for example, usually manifests itself through feelings of sadness, low energy/motivation levels, exhaustion, and a general disinterest in most activities, while anxiety disorders manifest themselves through fear, nervousness,



stress, and emotional instability. Mental health disorders continue to be a difficult problem to address due to the lack of adequate awareness, social stigma, and mental healthcare access.

The development of new methods and technology in AI, machine learning, and data science has paved ways to make predictions and diagnoses of diseases using new techniques. Predictive models employ the use of statistics and various machine learning techniques to analyze big data sets in order to uncover any behavioral or psychological patterns that can help diagnose various mental disorders. Such predictors include symptoms of stress, lack of sleep, change in moods, social withdrawal, and behavioral problems that can help predict any signs of depression and anxiety.

Different machine learning techniques such as Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine have performed impressively in predicting mental disorders. This study seeks to investigate the potential of employing predictive modeling and data science approaches to detect early signs of depression and anxiety among the participants. This study also aims to evaluate the efficiency of various machine learning techniques used in predicting mental disorders, and their importance in the field of preventive mental healthcare systems.

1.1 Background of the Study

The prevalence of mental disorders has emerged as one of the major causes of impairment and psychological problems across the globe. Depression and anxiety are some of the commonly reported mental health disorders that affect millions of people and influence their emotions, interactions, and employment. The conventional techniques for assessing mental well-being heavily rely on questionnaires, interviews, and psychological evaluation, and this approach cannot always recognize early warning signs of psychological issues. Early diagnosis could help to address the problem at its nascent stage.

Data science and artificial intelligence have revolutionized the field of healthcare in recent years, providing new ways of making predictions and intelligent decisions. Predictive modeling has recently attracted great interest in mental healthcare because of its ability to analyze various behavioral, emotional, and lifestyle-related aspects of a patient's state to detect

potential risks. Various machine learning algorithms can handle high volumes of data to discover patterns related to depression and anxiety. The use of predictive modeling and machine learning techniques opens up possibilities for creating prevention-oriented and personalized mental healthcare services.

Due to the growing digitalization and the popularity of mobile devices, a vast amount of valuable behavioral data can be obtained to predict an individual's mental condition. There is currently increasing interest in using predictive modeling to improve various aspects of mental healthcare. To assess the efficacy of these solutions in this sphere, the following research was designed to examine how predictive modeling techniques can be applied in mental healthcare systems.

1.2 Objectives of the Study

1. To analyze the role of predictive modeling in mental health assessment.
2. To identify behavioral and psychological indicators associated with depression and anxiety.
3. To evaluate the effectiveness of machine learning algorithms in predicting mental health disorders.
4. To compare the prediction accuracy of different data science models.
5. To examine the applicability of predictive analytics in early mental health intervention.

1.3 Hypotheses of the Study

H1: Predictive modeling techniques significantly improve early detection of depression and anxiety.

H2: Machine learning algorithms can accurately classify individuals at risk of mental health disorders.

H3: Behavioral and emotional indicators significantly influence predictive outcomes in mental health prediction.

2. REVIEW OF LITERATURE

Nemesure et al. (2021) The research concentrated on predictive modeling of depression and anxiety based on electronic health records through artificial intelligence machine learning algorithms. In this case, the study aimed at the analysis of health data, behavioral indicators, and other clinical information to determine the pattern related to mental disorders. Advanced machine learning models were used to predict depression and anxiety symptoms in patients. Results showed that artificial intelligence models enabled identification of mental disorder patterns and prediction accuracy. This study shows that predictive analytics can play a vital role in early diagnosis of psychological illnesses among patients.

Norouzi and Machado (2024) investigated the use of machine learning methods in the prediction of depression, anxiety, and stress disorders. In order to assess the effectiveness of these predictions, predictive analysis was employed by the authors to examine emotional and behavioral data related to mental disorders. The research emphasized the increasing role of data science solutions in the field of mental healthcare services. Moreover, it showed that the use of machine learning solutions enhanced the process of detecting psychological distress considerably. Predictive analysis could also help to develop prevention measures and personalized psychological treatments.

Karimian (2025) Explored the use of machine learning in diagnosing and forecasting mental disorders. In the review, different types of artificial intelligence and forecasting models used in mental health care research have been evaluated. According to the review, it was found that the use of machine learning has been very helpful for analyzing behavioral, emotional, and psychological information in diagnosing mental disorders. It was also seen that predictive modeling is useful for recognizing hidden patterns of data regarding depression, anxiety, and other mental health disorders. In addition to these, it can be said that using an artificial intelligence system would help in efficient decision making.

Al Amin et al. (2024) explored advanced machine learning approaches in forecasting and tracking cases of anxiety and depression. The main aim of the study was to examine how effective various predictive models were in analyzing behavioral and emotional factors linked to mental health conditions. As per the findings, the use of machine learning algorithms was



very effective in grouping patients based on their risk level of anxiety and depression with a high degree of predictability. What is more, the study revealed that factors such as level of stress, emotional instability, sleep disturbance, and social isolation had a significant influence on the accuracy of the results.

3. RESEARCH METHODOLOGY

The current study made use of the quantitative analytical research approach in order to examine the efficacy of predictive modeling and data science applications in detecting early symptoms of clinical depression and anxiety. The approach used ensured that behavioral and psychological data were collected, analyzed, and interpreted systematically. The study sought to evaluate the accuracy of different machine learning models in detecting the risks of depression and anxiety among the participants.

3.1 Research Design

Quantitative Analytical Research Design Approach was employed in this study. This approach was selected due to its suitability for making behavioral and emotional measures of mental disorder quantifiable and subject to statistical analysis. With this analytical approach, the researcher could explore the relationships that exist between the behavioral measures and the resultant mental outcomes, and determine the efficacy of prediction using Machine Learning Models. Predictive Analytics approaches were employed to classify individuals on the basis of depression and anxiety risk levels.

3.2 Sample Size

A total of 90 subjects was used in the research. The selection process utilized purposive sampling method depending on the readiness and availability of respondents. The selected respondents cut across various age and socio-economic backgrounds to ensure that the study was able to obtain different reactions from the participants concerning mental disorders.

3.3 Data Collection

The primary data was gathered by the use of structured questionnaires and assessment tests that had been specially developed to measure behavioural and emotional characteristics indicative

of depression and anxiety. Among the areas covered by the questionnaires were the levels of stress, quality of sleep, mood variations, social interactions, emotional stability, physical activity, and screen time.

4. RESULT AND DISCUSSION

Table 1 shows the socio-demographic profile of the individuals who participated in the study. This table includes data regarding gender and age-group composition among the 90 subjects involved in the research.

Table 1: Demographic Profile of Participants

Variable	Category	Frequency	Percentage
Gender	Male	41	45.6%
	Female	49	54.4%
Age Group	18–25 Years	38	42.2%
	26–35 Years	33	36.7%
	36–45 Years	19	21.1%

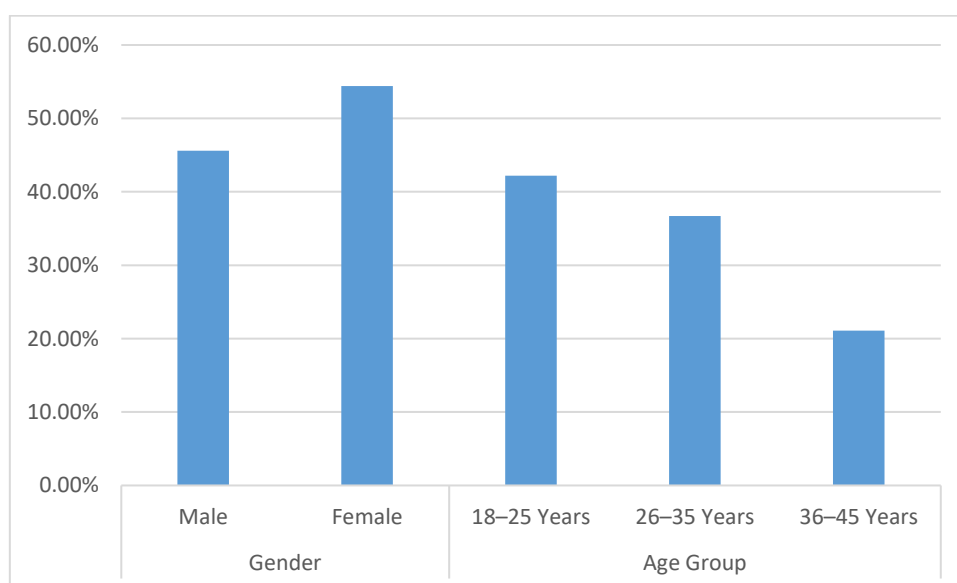


Figure 1: Graphical Representation of the Percentage of Demographic Profile of Participants

As can be seen from the results, out of the 90 respondents, there were 41 respondents who were males, accounting for 45.6%, while there were 49 respondents who were females, making up 54.4%. Regarding the age distribution of the respondents, most were within the age bracket of 18-25 years, with 38 respondents (42.2%). Secondly, there were 33 respondents who were within the age range of 26-35 years, making up 36.7%. There were 19 respondents within the age range of 36-45 years, representing 21.1% of the total.

Table 2: Frequency and Percent Distribution of Major Psychological and Behavioural Indicators Related to Depression and Anxiety.

Table 2: Psychological and Behavioural Indicators

Indicator	Frequency	Percentage
Sleep Disturbance	63	70.0%
High Stress Level	71	78.9%
Mood Swings	57	63.3%
Social Withdrawal	48	53.3%
Anxiety Symptoms	65	72.2%

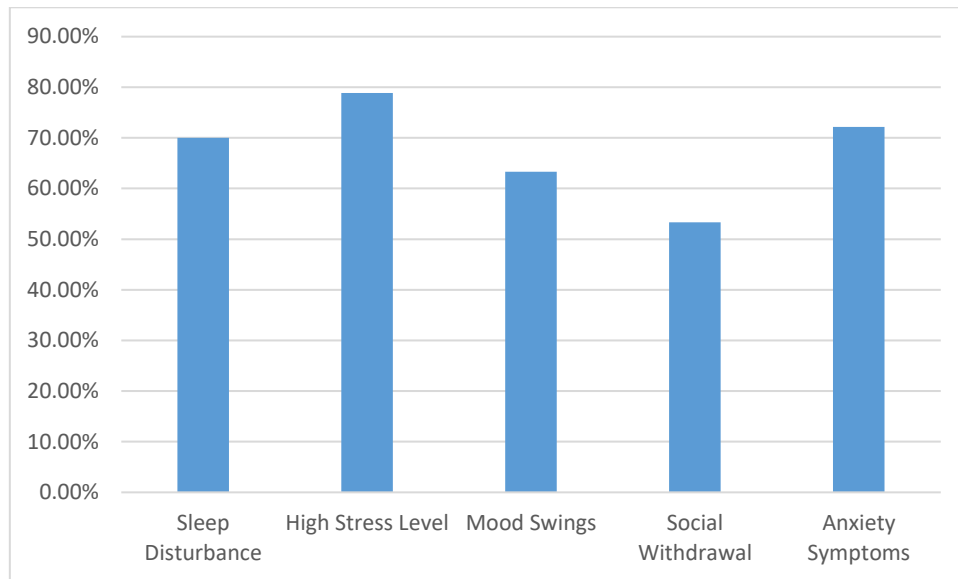


Figure 2: Graphical Representation of the Percentage of Psychological and Behavioural Indicators

The results show that high-stress levels were the most prevalent symptom, found in 71 students (78.9%). Anxiety was found in 65 students (72.2%) whereas insomnia was noted in 63 students (70.0%). Mood swings were found in 57 students (63.3%) while social isolation was found in 48 students (53.3%). From the results, it can be seen that high-stress levels, anxiety symptoms, and insomnia were quite prevalent in the students and could be significant risk factors for poor mental health.

Table 3 shows the prediction accuracy of various algorithms used for identifying early symptoms of depression and anxiety.

Table 3: Accuracy of Predictive Models

Predictive Algorithm	Prediction Accuracy
Logistic Regression	81%
Decision Tree	85%
Random Forest	93%
Support Vector Machine	89%

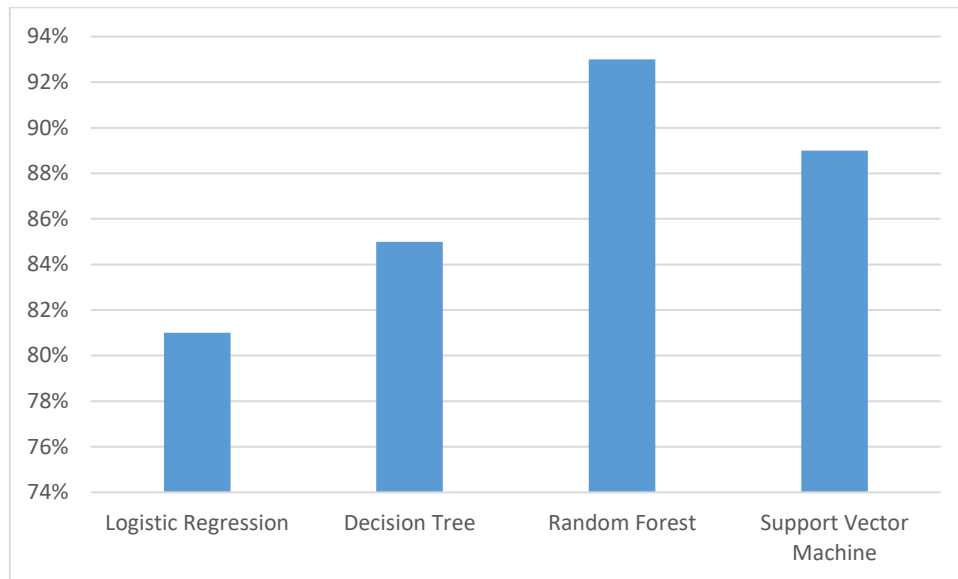


Figure 3: Graphical Representation of the percentage of Accuracy of Predictive Models

Random Forest algorithm gave the highest accuracy prediction of 93%, indicating its superiority in predicting the mental well-being of subjects. SVM gave an accuracy score of 89%, while Decision Tree had an accuracy rate of 85%. On the other hand, Logistic Regression gave the least accuracy of 81%. It is clear from the above results that the Random Forest is the most efficient prediction model among the machine learning models discussed here.

Table 4 presents the frequency and percentage distributions for participants based on depression and anxiety risks.

Table 4: Depression and Anxiety Risk Levels Among Participants

Risk Level	Frequency	Percentage
Low Risk	22	24.4%
Moderate Risk	38	42.2%
High Risk	30	33.4%
Total	90	100%

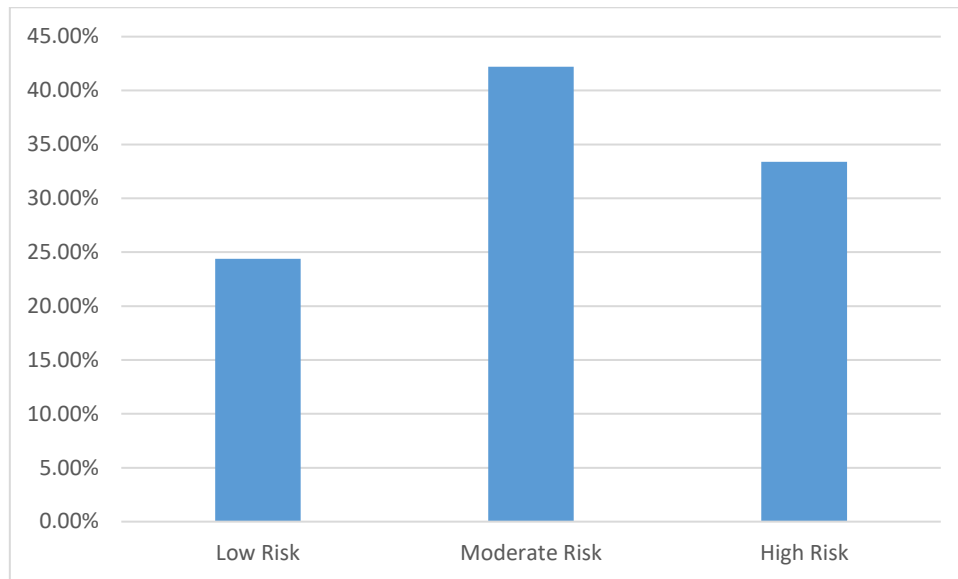


Figure 4: Graphical Representation of the Percentage of Depression and Anxiety Risk Levels Among Participants

Based on the research findings, it is clear that there are 38 participants (42.2%) in the moderate risk group, the highest proportion in terms of participants. There are 30 participants (33.4%) in the high-risk category, and those in the low-risk category are 22 participants (24.4%). From the above findings, it is evident that many participants are facing moderate to high levels of risk.

5. CONCLUSION

This study investigated the efficiency of predictive modeling and data science methods in the early identification of clinical depression and anxiety among the participants. It was established that behavioral and psychological factors like stress level, sleep problems, anxiety levels, mood variations, and social isolation highly affect the predictions of the mental state of a person, and a large number of participants belong to moderate- and high-risk categories. As far as the machine learning algorithms used in the research go, Random Forest model provided the best prediction accuracy rate, which reached 93%, followed by the Support Vector Machine algorithm with an accuracy rate of 89%. This study clearly illustrates the efficiency of using predictive analytics in the early identification of people who may suffer from mental health conditions.

0020REFERENCES

1. Al Amin, M., Liza, I. A., Hossain, S. F., Hasan, E., Haque, M. M., & Bortty, J. C. (2024). *Predicting and monitoring anxiety and depression: Advanced machine learning techniques for mental health analysis. British Journal of Nursing Studies, 4(2), 66-75.*
2. Alqaidoom, M. M., Ateeq, A., & Aljazrawi, Z. (2025). *AI-driven predictive analytics for early detection of high-functioning anxiety symptoms. In Sustainable Data Management: Navigating Big Data, Communication Technology, and Business Digital Leadership. Volume 2 (pp. 729-738). Cham: Springer Nature Switzerland.*
3. Hahn, T., Nierenberg, A. A., & Whitfield-Gabrieli, S. (2017). *Predictive analytics in mental health: applications, guidelines, challenges and perspectives. Molecular psychiatry, 22(1), 37-43.*
4. Karimian, M. (2025). *A short review on diagnosing and predicting mental disorders with machine learning. International Journal of Applied Data Science in Engineering and Health, 1(1), 20-27.*
5. Kumar, S., & Chong, I. (2018). *Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states. International journal of environmental research and public health, 15(12), 2907.*
6. Moshe, I., Terhorst, Y., Opoku Asare, K., Sander, L. B., Ferreira, D., Baumeister, H., ... & Pulkki-Råback, L. (2021). *Predicting symptoms of depression and anxiety using smartphone and wearable data. Frontiers in psychiatry, 12, 625247.*
7. Nemesure, M. D., Heinz, M. V., Huang, R., & Jacobson, N. C. (2021). *Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. Scientific reports, 11(1), 1980.*
8. Norouzi, F., & Machado, B. L. M. S. (2024). *Predicting Mental Health Outcomes: A Machine Learning Approach to Depression, Anxiety, and Stress. International Journal of Applied Data Science in Engineering and Health, 1(2), 98-104.*
9. Pandit, M., Azwaan, M., Wani, S., Ibrahim, A. A., Abdulghafor, R. A. A., & Gulzar, Y. (2023). *Examining factors for anxiety and depression prediction. International Journal on Perceptive and Cognitive Computing, 9(1), 70-79.*



10. Place, S., Blanch-Hartigan, D., Rubin, C., Gorrostieta, C., Mead, C., Kane, J., ... & Azarbajani, A. (2017). *Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. Journal of medical Internet research, 19(3), e75.*
11. Priya, A., Garg, S., & Tigga, N. P. (2020). *Predicting anxiety, depression and stress in modern life using machine learning algorithms. Procedia Computer Science, 167, 1258-1267.*
12. Salazar de Pablo, G., Studerus, E., Vaquerizo-Serrano, J., Irving, J., Catalan, A., Oliver, D., ... & Fusar-Poli, P. (2021). *Implementing precision psychiatry: a systematic review of individualized prediction models for clinical practice. Schizophrenia bulletin, 47(2), 284-297.*
13. Silva, C., Saraee, M., & Saraee, M. (2019, November). *Predictive modelling in mental health: a data science approach. In 2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET) (pp. 6-11). IEEE.*
14. Srividya, M., Mohanavalli, S., & Bhalaji, N. (2018). *Behavioral modeling for mental health using machine learning algorithms. Journal of medical systems, 42(5), 88.*
15. Zhai, Y., Zhang, Y., Chu, Z., Geng, B., Almaawali, M., Fulmer, R., ... & Du, X. (2025). *Machine learning predictive models to guide prevention and intervention allocation for anxiety and depressive disorders among college students. Journal of Counseling & Development, 103(1), 110-125.*



Author's Declaration

I as an author of the above research paper/article, here by, declare that the content of this paper is prepared by me and if any person having copyright issue or patent or anything otherwise related to the content, I shall always be legally responsible for any issue. For the reason of invisibility of my research paper on the website /amendments /updates, I have resubmitted my paper for publication on the same date. If any data or information given by me is not correct, I shall always be legally responsible. With my whole responsibility legally and formally have intimated the publisher (Publisher) that my paper has been checked by my guide (if any) or expert to make it sure that paper is technically right and there is no unaccepted plagiarism and hentriacontane is genuinely mine. If any issue arises related to Plagiarism/ Guide Name/ Educational Qualification /Designation /Address of my university/ college/institution/ Structure or Formatting/ Resubmission /Submission /Copyright /Patent /Submission for any higher degree or Job/Primary Data/Secondary Data Issues. I will be solely/entirely responsible for any legal issues. I have been informed that the most of the data from the website is invisible or shuffled or vanished from the database due to some technical fault or hacking and therefore the process of resubmission is there for the scholars/students who finds trouble in getting their paper on the website. At the time of resubmission of my paper I take all the legal and formal responsibilities, If I hide or do not submit the copy of my original documents (Andhra/Driving License/Any Identity Proof and Photo) in spite of demand from the publisher then my paper maybe rejected or removed from the website anytime and may not be consider for verification. I accept the fact that as the content of this paper and the resubmission legal responsibilities and reasons are only mine then the Publisher (Airo International Journal/Airo National Research Journal) is never responsible. I also declare that if publisher finds Any complication or error or anything hidden or implemented otherwise, my paper maybe removed from the website or the watermark of remark/actuality maybe mentioned on my paper. Even if anything is found illegal publisher may also take legal action against me.

Dr Anju Chauhan
