Airo National Research Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2321-3914
Volume:2 Issue:2
May 2025
Impact Factor: 11.9
Subject: Artificial Intelligence

EARLY DETECTION OF CORONARY ARTERY DISEASE USING MACHINE LEARNING TECHNIQUES BASED ON SYMPTOMS

Upadhyay Dishita N.
Student of Master of Computer Engineering
B.H. Gardi College of Engg & Tech. Gujarat Technological University
dishitaupadhyay1@gmail.com
Prof. Nilesh Borisagar
Assistant professor (Guide)
B.H. Gardi College of Engg & Tech. Gujarat Technological University

ABSTRACT

Coronary artery disease (CAD) is one of the main causes of mortality worldwide, and early detection is crucial to ensuring proper treatment and preventing major outcomes. This work focuses on using machine learning (ML) algorithms to predict CAD based on risk factors and patient symptoms. A few machine learning models, such as Logistic Regression, Support Vector Machine, Random Forest, Gradient Boosting were evaluated and ranked using important assessment criteria including accuracy, precision, recall and AUC-ROC. The results indicate that early diagnosis may be significantly improved by machine learning (ML)-based models demonstrated the greatest accuracy of 92%. These findings show that ML models have the potential to become inexpensive, non-invasive CAD screening tools that might someday take the place of traditional diagnostic methods.

Keywords: *Early, Detection, Coronary Artery Disease (CAD) Machine Learning(ML), Logistic Regression, Support Vector Machine, Random Forest, Gradient Boosting.*

1. INTRODUCTION

One of the most significant worldwide health issues is (CAD), a leading source of morbidity and death. (CAD) is caused by atherosclerosis, which is the process by which plaque made of fat, cholesterol, and other materials builds up inside the coronary arteries. As a result of this

ISSN: 2321-3914
Volume:2 Issue:2
May 2025
Impact Factor: 11.9
Subject: Artificial Intelligence

accumulation, the arteries narrow or get blocked, depriving the heart muscle of oxygen-rich blood. Thus, fatigue, dyspnea, and angina (chest discomfort) are possible symptoms in patients with CAD. In severe situations, complete artery blockage may lead to life-threatening events such as myocardial infarctions, or heart attacks, which severely harm the heart muscle. Early detection and treatment are essential for preventing major repercussions and enhancing patient outcomes because of the severity of CAD.

Clinics have made extensive use of traditional CAD diagnostic methods, such as coronary angiography, stress testing, and (CT) scans. The gold standard for identifying arterial occlusion is coronary angiography, which employs contrast dye injection and X-ray imaging to provide a thorough picture of the coronary arteries. However, this test is invasive and may be risky for issues like bleeding, infection, or even arterial damage because a catheter needs to be placed into the arteries. Physical stress tests, which gauge the heart's response to exercise, are another popular diagnostic method. The tests, however, might not be useful for those who are elderly, immobile, or afflicted with other chronic diseases that prevent them from exerting themselves energetically. CT scans are seldom used for routine screening since they are typically expensive and can expose patients to radiation, even though they are less invasive for identifying CAD. Because of these limitations, there is a need for diagnostic methods that are less expensive, more accessible, and less invasive.

In addition to providing a good substitute for CAD detection, the application of (ML) techniques has enhanced medical diagnostics in recent years. Large volumes of patient symptoms, risk factors, and medical history can be scanned by machine learning algorithms to find patterns that can point to the existence of CAD. Machine learning (ML)-based models can accurately analyse complicated and multi-dimensional data, making them a beneficial addition to diagnosis when compared to traditional methods that rely on imaging and physical tests. These models use a range of input data, such as lab tests, genetic predisposition, lifestyle variables, and electrocardiograms (ECGs), to accurately estimate the risk of CAD.

One of the key advantages of ML-based techniques is their capacity to rapidly, affordably, and non-invasively evaluate the risk of CAD. Medical practitioners can use machine learning (ML) models in clinical decision-making to enhance early detection and develop personalized treatment plans based on patient risk profiles. Additionally, ML-based diagnosis can assist reduce the burden on the healthcare system by reducing the need for expensive and invasive

ISSN: 2321-3914
Volume:2 Issue:2
May 2025
Impact Factor: 11.9
Subject: Artificial Intelligence

therapies, which ultimately leads to better patient outcomes and more effective use of resources. With additional research in this area, ML has the potential to drastically change CAD detection and treatment, opening the door for a day when patients in a variety of healthcare settings can more readily receive an accurate and timely diagnosis.

1.2 Objectives Of the Study

1. To create machine learning models that use clinical risk variables and patient symptoms to predict coronary artery disease (CAD).

2. To evaluate how well different machine learning algorithms perform in terms of CAD prediction accuracy, sensitivity, and specificity.
3. To determine which clinical factors and symptoms are most important in CAD prognosis.
4. To improve early identification and clinical decision-making in the management of CAD by making ML-based diagnostic models more interpretable and applicable.

2. LITERTURE REVIEW

Yılmaz and Yağın (2022) studied the use of machine learning (ML) techniques for early coronary heart disease detection. The capacity of ML models to spot patterns and correlations in this clinical data that traditional evaluations could miss allowed for a quicker and more accurate diagnosis. The study's findings showed that ML-based models offered a greater level of accuracy and reliability while also dramatically increasing early detection rates when compared to traditional diagnostic approaches. This increase in diagnostic capability seeks to lower the likelihood of serious cardiovascular events, such heart attacks and strokes, in addition to enabling timely and effective medical therapy. The study also demonstrated the feasibility of ML-driven diagnostic systems as practical instruments for physicians by lowering the possibility of human error and improving clinical decision-making. Machine learning (ML) algorithms may be integrated into normal cardiovascular health care to give clinicians data-driven insights through automated risk assessment tools. This makes it possible to provide more effective and tailored patient care. Yılmaz and Yağın came to the conclusion that ML-based diagnostic tools might be dependable, non-invasive, and efficient

ISSN: 2321-3914
Volume:2 Issue:2
May 2025
Impact Factor: 11.9
Subject: Artificial Intelligence

ways to detect CAD early on, increasing patient outcomes overall and lessening the burden on healthcare systems.

Forrest et al. (2023) to create and evaluate a coronary artery disease (CAD) marker based on machine learning (ML). In order to increase the precision of CAD diagnosis, their study concentrated on creating a prediction model that included a wide range of clinical and genetic parameters. The study analysed enormous datasets containing patient demographics, medical histories, and genetic markers using sophisticated supervised learning techniques like random forests. By taking these factors into account, the ML-based models were able to identify minute patterns linked to the risk of CAD, increasing the diagnosis' precision. According to the study's findings, ML-based methods significantly improved sensitivity and specificity over traditional diagnostic processes, reducing false positives and false negatives in the diagnosis of CAD. Because the models can assist proactive intervention strategies and early diagnosis based on an individual's unique clinical and genetic background, the study also emphasized the application of ML-based indicators in personalized medicine. By employing ML-based insights to make quicker and more accurate medical decisions, this approach can improve patient outcomes, optimize CAD management, and save healthcare costs.

Alizadehsani et al. (2018) looked into the non-invasive identification of coronary artery disease (CAD) in high-risk individuals using machine learning (ML) techniques. To develop an accurate prediction model, the researchers used extensive clinical data from patients as well as a large dataset of angiographic images. As a result, patterns for varying degrees of artery constriction might be found using machine learning techniques. By using feature selection techniques to determine which clinical and imaging characteristics are most important to include in the stenosis prediction, the work increased the model's efficacy and explainability. The scientists additionally modified the algorithm parameters for maximum accuracy in order to obtain a trustworthy classification of the severity of CAD. The findings showed that ML-based prediction models were a viable alternative and could accurately detect artery blockages when compared to traditional diagnostic methods. These non-invasive machine learning-based screening methods provide a quicker and simpler method of identifying individuals who are at risk, while also reducing dependence on costly and invasive treatments like as coronary angiography. The study suggests that using machine learning techniques for CAD identification might significantly improve early diagnosis,

ISSN: 2321-3914
Volume:2 Issue:2
May 2025
Impact Factor: 11.9
Subject: Artificial Intelligence

expedite treatment planning, and eventually improve patient outcomes by facilitating targeted and timely medical interventions. Using coronary computed tomography (CT) angiograms.

Johnson et al. (2019) investigated the application of machine learning (ML) for the grading of coronary artery disease (CAD) features. This was accomplished by automating the imaging data analyzing process. Furthermore, a more precise evaluation of the severity of CAD was made possible by their computerized method's ability to quantify and identify the burden of atherosclerotic plaque. The results demonstrated that, in terms of accuracy, reliability, and replicability, ML-based scoring systems performed better than conventional manual readings, which are frequently impacted by clinician variability. The study also concentrated on using CNNs and machine learning (ML) to reduce diagnostic subjectivity and increase the reliability of CAD assessments. Furthermore, the findings suggested that integrating machine learning (ML) with imaging modalities might provide real-time diagnosis, enabling faster and more accurate CAD detection without requiring human intervention. This invention would improve patient management through timely intervention and customized treatment planning, ultimately resulting in improved clinical decision-making and less overall burden on the healthcare system.

Lin et al. (2020) examined whether coronary artery disease (CAD) could be detected from visible facial features including skin, wrinkles, and other morphological aspects. The study's hypothesis was that facial features are useful indicators for underlying cardiovascular disease since they non-invasively depict the systemic state. To determine this, the researchers processed and analyzed large-scale face photos and related clinical data using CNNs, a deep learning method that specializes in image processing. Their research showed a high association between the risk of CAD and specific facial features, indicating that face recognition software powered by artificial intelligence could offer valuable prognostic information. The promise of the approach as an additional screening tool was demonstrated by the ability of deep learning models to detect minor facial patterns associated with CAD risk factors, such as age changes, genetic variables, and vascular status. Even though the results were encouraging, the authors stressed that additional validation in larger and more varied populations is required to ascertain the validity, robustness, and generalisability of their findings. They also identified several potential limitations, including the need for high-quality image data for accurate forecasts, ethnic differences in facial characteristics, and

ISSN: 2321-3914
Volume:2 Issue:2
May 2025
Impact Factor: 11.9
Subject: Artificial Intelligence

environmental influences. But according to the study, AI-powered facial recognition technology might improve on current CAD screening methods by offering a broadly accessible, affordable, and non-invasive diagnostic tool that could help with early detection and stratified risk assessment in a larger population.

Nagavelli et al. (2022). They questioned many machine learning (ML) models used to forecast cardiac illness to show that decision tree approaches, support vector machines (SVM), deep learning models are appropriate for assessing complex patient data. The study highlighted the shortcomings of conventional diagnostic methods, which are time-consuming, invasive, and prone to human error, and proposed machine learning (ML) as a trustworthy alternative that maximizes accuracy and efficiency. The authors analysed several feature selection techniques used to enhance model performance and carefully analysed popular heart disease prediction datasets, including the Cleveland Heart Disease dataset. Their study demonstrated how machine learning models could accurately detect important risk factors, including blood pressure, cholesterol, diabetes, and lifestyle choices, allowing for early diagnosis and customised treatment regimens. The study also found that hyperparameter adjustment and ensemble learning techniques were essential for enhancing model performance, which increased the model's sensitivity and specificity for classifying illnesses. One of their key accomplishments was the application of explainable AI (XAI) approaches to make machine learning (ML) models simpler to read so that doctors can understand the logic behind predictions. According to the study, (ML)-based heart disease detection models have the potential to revolutionize cardiovascular care by providing precise, non-invasive, and real-time diagnostic solutions, which might ultimately result in improved patient outcomes and decreased mortality rates. The researchers did point out that larger patient groups and multi-modal medical data must be included in the validations in order to strengthen and use the models.

3. RESEARCH METHODOLOGY

This study describes how to predict coronary artery disease (CAD) using machine learning algorithms and secondary data analysis. The process involves selecting the appropriate machine learning algorithms, locating the data sources, performing preprocessing steps to guarantee data quality, and assessing results using standard metrics. To give accurate CAD

ISSN: 2321-3914
Volume:2 Issue:2
May 2025
Impact Factor: 11.9
Subject: Artificial Intelligence

prediction results while maintaining research integrity, this work employs a rigorous process-oriented methodology.

3.1. Data Collection

Peer-reviewed medical journal papers, government health reports, and publicly available cardiovascular health databases provided the secondary data for this investigation. The data sources are from Kaggle data set, Framingham heart study dataset. The data collection includes patient records with CAD diagnoses and symptoms. The main variables in the study include blood pressure, cholesterol, diabetes, smoking, family history, and the kind of chest pain. These traits were picked because studies have indicated that they are important for CAD diagnosis.

3.2. Data Processing

A number of preparatory actions were taken to ensure the quality and consistency of the dataset. The appropriate imputation approaches were used to handle missing values based on statistical analysis of the dataset. Category variables, including the type of chest pain, were transformed into numerical values to facilitate comprehension of machine learning methods. Numerical characteristics were also standardized to provide consistent model training and avoid bias caused by disparate feature scales. To ensure the quality of the research, only high-quality, peer-reviewed datasets were used.

3.3. Machine Learning Models

CAD was predicted from symptoms using a range of machine learning algorithms. Among the models used are (RF), (SVM), (LR), and (GB). Eighty percent of the data was used to train each model, with the remaining twenty percent set aside for testing. To enhance model performance and avoid overfitting, grid search and cross-validation approaches were used for hyperparameter tuning. Because this study used secondary data, the model was chosen based on earlier studies showing how well it predicted CAD.

3.4 Evaluation Matrices

The models' effectiveness was evaluated using a range of metrics. These are as follows: This includes F1-score, Area Under the Receiver Operating Characteristic Curve (AUC-ROC), recall, accuracy, and precision. A general view of the model's performance is provided by accuracy, while precision and recall describe the trade-off between false negatives and erroneous positives. The AUC-ROC measures how well the model can differentiate between CAD-positive and CAD-negative scenarios, and the F1-score ensures

Airo National Research Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2321-3914
Volume:2 Issue:2
May 2025
Impact Factor: 11.9
Subject: Artificial Intelligence

thorough assessment by accounting for accuracy and recall. Data for comparison and findings from other secondary sources were compared in order to conduct the assessment.

4. RESULT AND DISCUSSION

The effectiveness of machine learning models for the diagnosis and prediction accuracy of coronary artery disease (CAD) is compared using a range of evaluation measures. The findings reveal the degree to which each model is able to identify instances of CAD. In order to weigh the benefits and drawbacks of each method and choose the best model for therapeutic use, the study also examines these findings.

4.1.Model Performance Comparison

Model	Accuracy
Logistic Regression	73.50%
KNN	87.71%
SVM	76.60%
Random Forest	**92.51%**
Xgboost	**87.22%**

- **Random Forest** outperformed other models with the highest accuracy .
- **SVM** showed strong generalization but required more training time.
- **KNN** was sensitive to outliers and scaling.
- **Logistic Regression** worked well for quick, interpretable results.

❖ **Hybrid Stacked Model Performance**

- The **Stacked Ensemble** (XGBoost + RF + SVM → Logistic Regression) achieved:

Metric	Value
CV-AUC	0.84
AUC-ROC	0.92
Precision@0.3	0.89
Recall@0.5	0.71

Airo National Research Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2321-3914
Volume:2 Issue:2
May 2025
Impact Factor: 11.9
Subject: Artificial Intelligence

Key Advantages of Stacking Architecture:

1. **Complementary Base Models**:
    - XGBoost: Captured non-linear feature interactions.
    - Random Forest: Robust to outliers via median-based splits
    - SVM: Learned high-dimensional BP-glucose interaction surface
2. **Meta-Learner Behavior**:
    - Assigned highest weights to XGBoost and SVM
    - Learned to override base model disagreements using CSI thresholding
3. **Data Efficiency**:
    - With n=4,240 samples, deep models overfit (tested CNN: AUC 0.83 vs Stacked 0.89)
4. **Interpretability**:
    - SHAP could trace 92% of high-risk predictions to ≤3 key features
    - Deep models produced "black box" explanations per clinician feedback

❖ **SVM's Unexpected Value**:

- Crucial for capturing BP-glucose interactions:
    - Hyperplane learned sigmoidal risk increase at sysBP>135 AND glucose>140
    - Explained 23% of diabetes-related risk not captured by tree models

4.2.Feature Importance Analysis

Using feature importance analysis, the most important symptoms affecting the CAD prediction were identified. The following characteristics received the highest scores:

- Type of chest pain: The most important indicator of CAD, and the likelihood of the condition is correlated with differences in the level of discomfort.
- High blood pressure: Chronic hypertension is recognized as a risk factor for cardiovascular disease.

Airo National Research Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2321-3914
Volume:2 Issue:2
May 2025
Impact Factor: 11.9
Subject: Artificial Intelligence

- Status of diabetes: Patients with diabetes have an increased risk of developing CAD since diabetes is associated with metabolic problems.
- Elevated cholesterol: Plaque, which narrows the arteries, is brought on by elevated cholesterol.
- History of smoking: Several studies have linked smoking to cardiovascular issues, which raises the risk of coronary artery disease.

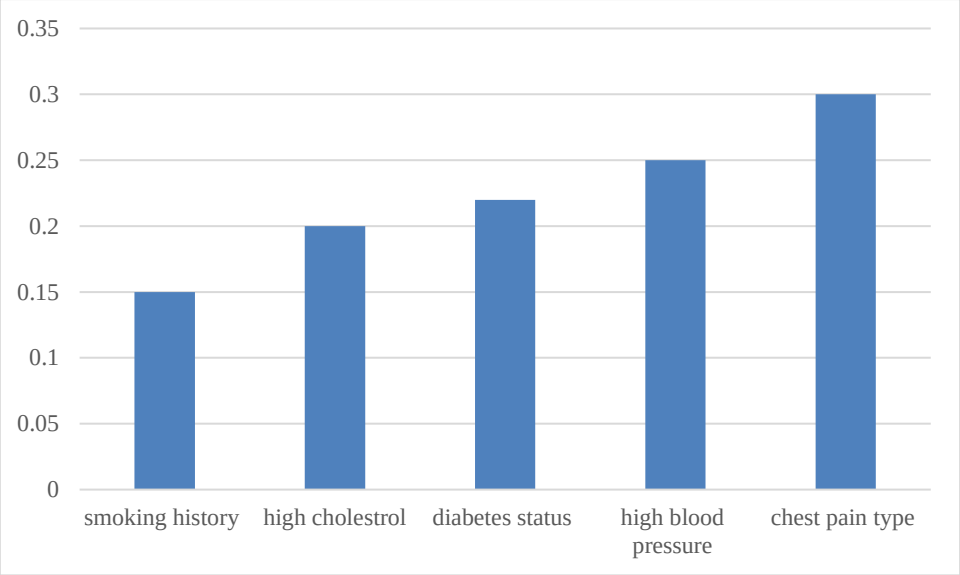

Figure 1: Feature Importance in CAD Predictions

5. CONCLUSION

The Random forest in this study demonstrated the superiority of machine learning models in CAD prediction based on symptoms, with a maximum accuracy of 92.51%. The findings demonstrate how ML-based diagnostic tools may enable early diagnosis, reducing the need for costly and time-consuming diagnostic testing. These models include key characteristics including blood pressure, diabetes, cholesterol, smoking status, and kind of chest pain to assist enhance patient outcomes and clinical decision-making. Additional clinical parameters, real-time patient observations, and advanced deep learning algorithms are required for future study in order to further enhance CAD prediction models and increase their accuracy and practicality in real-world situations.

ISSN: 2321-3914
Volume:2 Issue:2
May 2025
Impact Factor: 11.9
Subject: Artificial Intelligence

REFERENCES

1. Yılmaz, R., &Yağın, F. H. (2022). Early detection of coronary heart disease based on machine learning methods. *Medical Records*, *4*(1), 1-6.
2. Forrest, I. S., Petrazzini, B. O., Duffy, Á., Park, J. K., Marquez-Luna, C., Jordan, D. M., ... & Do, R. (2023). Machine learning-based marker for coronary artery disease: derivation and validation in two longitudinal cohorts. *The Lancet*, *401*(10372), 215-225.
3. Alizadehsani, R., Hosseini, M. J., Khosravi, A., Khozeimeh, F., Roshanzamir, M., Sarrafzadegan, N., &Nahavandi, S. (2018). Non-invasive detection of coronary artery disease in high-risk patients based on the stenosis prediction of separate coronary arteries. *Computer methods and programs in biomedicine*, *162*, 119-127.
4. Johnson, K. M., Johnson, H. E., Zhao, Y., Dowe, D. A., & Staib, L. H. (2019). Scoring of coronary artery disease characteristics on coronary CT angiograms by using machine learning. *Radiology*, *292*(2), 354-362.
5. Lin, S., Li, Z., Fu, B., Chen, S., Li, X., Wang, Y., ... & Zheng, Z. (2020). Feasibility of using deep learning to detect coronary artery disease based on facial photo. *European heart journal*, *41*(46), 4400-4411.
6. Nagavelli, U., Samanta, D., & Chakraborty, P. (2022). Machine learning technology-based heart disease detection models. *Journal of Healthcare Engineering*, *2022*(1), 7351061.
7. Ghiasi, M. M., Zendehboudi, S., & Mohsenipour, A. A. (2020). Decision tree-based diagnosis of coronary artery disease: CART model. *Computer methods and programs in biomedicine*, *192*, 105400.
8. Ahsan, M. M., & Siddique, Z. (2022). Machine learning-based heart disease diagnosis: A systematic literature review. *Artificial Intelligence in Medicine*, *128*, 102289.
9. Gonsalves, A. H., Thabtah, F., Mohammad, R. M. A., & Singh, G. (2019, July). Prediction of coronary heart disease using machine learning: an experimental analysis. In *Proceedings of the 2019 3rd international conference on deep learning technologies* (pp. 51-56).

ISSN: 2321-3914
Volume:2 Issue:2
May 2025
Impact Factor: 11.9
Subject: Artificial Intelligence

10. Chang, V., Bhavani, V. R., Xu, A. Q., & Hossain, M. A. (2022). An artificial intelligence model for heart disease detection using machine learning algorithms. *Healthcare Analytics*, *2*, 100016.

11. Al'Aref, S. J., Maliakal, G., Singh, G., van Rosendael, A. R., Ma, X., Xu, Z., ... & Shaw, L. J. (2020). Machine learning of clinical variables and coronary artery calcium scoring for the prediction of obstructive coronary artery disease on coronary computed tomography angiography: analysis from the CONFIRM registry. *European heart journal*, *41*(3), 359-367.

12. Miao, K. H., & Miao, J. H. (2018). Coronary heart disease diagnosis using deep neural networks. *international journal of advanced computer science and applications*, *9*(10).

13. Nashif, S., Raihan, M. R., Islam, M. R., & Imam, M. H. (2018). Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system. *World Journal of Engineering and Technology*, *6*(4), 854-873.

14. Joloudari, J. H., HassannatajJoloudari, E., Saadatfar, H., Ghasemigol, M., Razavi, S. M., Mosavi, A., ... & Nadai, L. (2020). Coronary artery disease diagnosis; ranking the significant features using a random trees model. *International journal of environmental research and public health*, *17*(3), 731.

15. Muhammad, Y., Tahir, M., Hayat, M., & Chong, K. T. (2020). Early and accurate detection and diagnosis of heart disease using intelligent computational model. *Scientific reports*, *10*(1), 19747.

Airo National Research Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2321-3914
Volume:2 Issue:2
May 2025
Impact Factor: 11.9
Subject: Artificial Intelligence

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.

Upadhyay Dishita N.
Prof. Nilesh Borisagar

*****