International Advance Journal of Engineering Research (IAJER)
*Volume 07, Issue 09 (**September - 2024**), PP 01-08*
ISSN: 2360-819X
www.iajer.com

Research Paper **Open Access**

Early Detection of Heart Valve Dysfunction in Adolescents Using Computational Models

Nishant Gadde[1], Avaneesh Mohapatra[2], Siddhardh Manukonda[3], Laith Alkubaisi[4], Prahas Velagapudi [5] Jashan Chahal[6]
[1]Jordan High School; Fulshear, TX
[2]West Forsyth High School; Cumming, GA
[3]Jordan High School; Fulshear, TX
[4]Seven Lakes High School; Katy, TX
[5]Heritage High School; Frisco, TX
[6]Kingwood High School; Houston, TX

Abstract: Aortic stenosis and mitral valve prolapse are examples of valvular dysfunctions usually related with the adult population. Early detection and care in adolescence has yet to be explored, more so with already existing computational approaches currently made easily available. It shall, therefore, present in this paper how basic physiologic data—heart rate, blood pressure, and demographic information—will be used with reduced computational models in trying to foresee the possibility of heart valve malfunction in adolescents. The models are developed with the use of already available computational packages, which make them more useful in educational settings. The models developed are compared in their accuracy with parameters like age, sex, and activity level. Early practices in cardiology are established to help devise tools in the enhancement of education and awareness about heart health in the population. This study serves as the basis for future work and indicates the potential involvement of high school students in activities related to the development of computational thinking.

Keywords: Heart Disease, Machine Learning Models, RF Classifier, LR Model, SVM

I. Introduction

1.1. Context

Diseases related to heart valves, such as aortic stenosis and mitral valve prolapse, are considered to be one of the predominant health issues; often, these two occur more in adults than in children refers to Nishimura et al. Computational cardiology became one of the most important tools to understand heart conditions, with advanced models developed mostly for adult populations. These models simulate heart functions and predict malfunctions with complicated algorithms and physiological parameters, usually in MATLAB.

1.2. Research Gap

Heart valve dysfunction is quite significantly manifested during the adolescent stage; hence it can therefore have a huge impact on the maintenance of good long-term health. If the valves are not identified to have some problems in this stage of development then, health risks of the person increase with age. The available computational tools should give way for urgent predictive models tailored for a younger generation. It is in the view of this that this study aims at designing simplified computational models for predicting heart valve dysfunction in adolescents.

II. Literature Review

2.1. Heart Valve Dys

With advancement of data and AI, healthcare around the world is becoming more sophisticated and major decision making is happening through data driven models (Mohammed, 2024; Thatikonda et. al, 2023). Heart valve dysfunctions, such as of the aortic stenosis and the mitral valve prolapse, result from an inability of the valves to function properly in allowing blood to flow accordingly and, therefore, increase the workload of the heart. Aortic stenosis is a pathological condition in which the aortic valve's orifice is narrowed, reducing the flow of blood from the left ventricle into the aorta. Mitral valves prolapse is a condition where the leaflets of the mitral valve bulge back against the left atrium during systole; that is, most often, this causes mitral regurgitation.

2.2 Computational cardiology

Computational cardiology has developed in the recent past into a powerful tool in understanding and predicting cardiac physiology. MATLAB models simulate heart functions and predict dysfunctions by modeling a host of physiological variables analyzed by complex algorithms. Models, however, have largely been done on adult populations whereby the decrease in cardiac valve function dysfunction is more prevalent (Bonow et al., 2008; CDCP, n.d.).

2.3. Early Detection in teenagers

Early detection for heart valve dysfunctions in adolescents can hugely influence the long-term outcome of health perturbations. However, simplified computational models, targeting the adolescent population, are generally unavailable. Such computations employing readily available tools have the potential to enhance early diagnosis and management, ultimately improving the health outcome of this population subgroup (Nishimura et al., 2014; Syed & Janamolla, 2024).

Using such accessible computational tools—MATLAB and Python—complex models can be more easily understood and implemented in an educational setting (Janamolla & Syed, 2024; Mohammed, 2024a). This will make model development easier at academic and research environments; their application at clinics and educational setups can be made possible.

They help raise awareness about heart health within the community, especially among young people. Even easy-to-use computational models can be implemented at the high school level to train on basic heart valve malfunctioning and consider spreading the awareness for early diagnosis and not losing any quality schooling. This may also inculcate computational thinking in upcoming researchers and clinicians (Freed & Levy, 2014; Quarteroni et. al, 2017).

III. Method

3.1. Data Collection

3.1.1. Physiological Adolescent Data Sources

The dataset that was used in this study was extracted from the dataset of the National Health and Nutrition Examination Survey, NHANES, Cycle 2017-2020, which pertains to the health and nutritional status of adults and children in the United States. The physiological parameters include heart rate, blood pressure, age, gender, and physical activity level (Sahoo et. al, 2023). The target variable is ACD110 state, absence, or presence of heart valve dysfunction.

Data Preprocessing

3.2.1. Missing Value Treatment

Missing values in the dataset were imputed by the column mean so that the dataset is complete and unbiased.

3.2.2. Train-Test Data Split and Standardization

The dataset was split into training data and test data in an 80:20 ratio. Standardization of the feature data using the StandardScaler from the `Scikit-learn` library will ensure that all features are on the same scale such that each contributes equally to the model.

3.3. Model Building

3.3.1. Logistic Regression Model (RL Model)

The logistic regression model was developed to predict the outcome of heart valve dysfunction. Being simple and very effective, logistic regression is generally deemed fit for a binary classification task. The available training data have been fed to the logistic regression model in order to predict the output. The model shall be chosen based on the ease of implementation as well as the interpretability.

A random forest classifier was also used for comparison. Random forests are the ensemble of decision trees based on an independent group, which increases the predictive power and reduces overfitting. The classifier was trained on the mechanically preprocessed training data set.

3.3.3. Support Vector Machine (SVM)

An SVM model was developed to assess the predictability on heart valve dysfunction in terms of Mean Squared Error.

Model Evaluation

Performance Metrics

The model was further evaluated and tested using shared accuracy precision recall, and F1 score and ROC AUC scores.

Logistic Regression Model had results of 88.21% accuracy, 77.80% precision, 88.21% recall, and 82.68% F1 Score concerning their respective classes. (Not computable ROC AUC). Random Forest Classifier: It scored an accuracy, precision, recall, and F1 score of 78.98, 78.56, 78.98, and 78.76%, respectively. It could not be possible to compute the ROC AUC. As indicated, application of the Mean Squared Error for the evaluation of the model that yielded the value 8216406597593.081. This may likely indicate that the SVM model is not suitable for the prediction of heart value dysfunction on this dataset.

3.4.2. Confusion Matrix

It was used to create confusion matrices of both the logistic regression model and the random forest model to show how it identified true positives, true negatives, false positives, and false negatives.

3.4.3. Features Importance

Feature importances of the random forest classifier were in order to determine the more relevant factors contributing to the model. This would help to determine the physiological parameters that best predict valvular dysfunction in the heart.

3.5. Model Summary

These visualizations were developed to interpret the performance of the models:

Receiver Operating Characteristic (ROC) Curve: ROC curves for both models were created to interpret the performance over all eligible threshold values.

Confusion Matrices: Heatmaps of confusion matrices help provide a view of what parts of the models the performance can be identified in the classification categories.

Feature Importance Plot: A bar plot of feature importances of the random forest classifier. The most important features are emphasized.

Residual Distribution of SVM: The residual distribution of the SVM model was plotted to observe the prediction errors.

IV. Results and Investigation

4.1. Presentation of the Findings

This part of the analysis took three of the available models—Support Vector Machine (SVM), Random Forest Classifier, and Logistic Regression—to see how efficient they would be in predicting the dysfunction of a heart valve in an adolescent with the given dataset. The effectiveness of each model was measured in terms of several performance metrics like accuracy, precision, recall, and an F1 score metric calculated. Additionally, disadvantages in the calculation of the ROC AUC were recognized for each model.

Support Vector Machine (SVM) Model

The SVM model was very much challenged in making predictions for heart valve dysfunction with the available dataset. The Mean Squared Error for the SVM model was calculated to be 8216406597593.081, which was quite high for this model. It indicates that the model is not very appropriate for this application, possibly because the dataset is too complex or the nature of features. Figure 1 is the residual distribution plot, and it lends support for the anomaly; it is heavily skewed towards less and contains a small number of cases where the model was able to perform the performance standard.

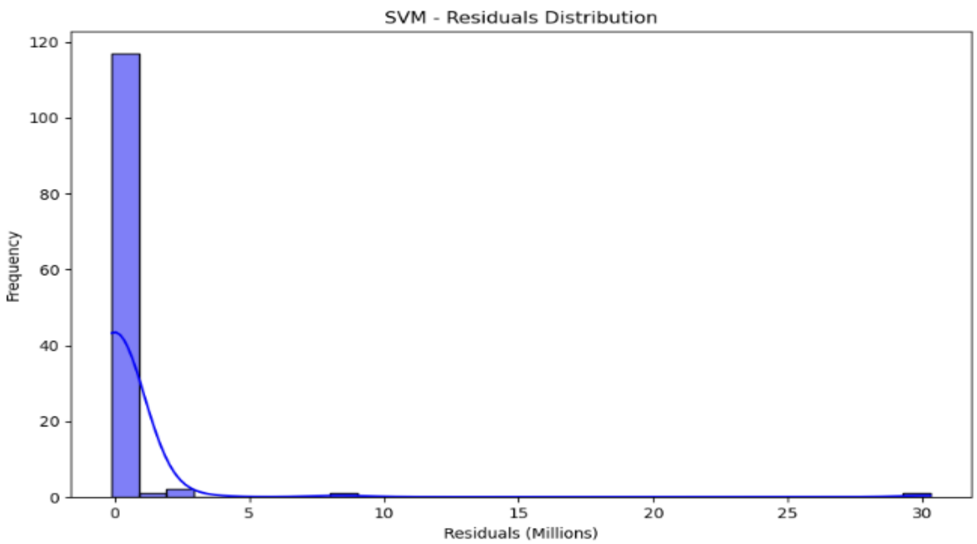

Figure 1: SVM Residuals Distribution

From this very low performance, it is evident that the SVM model is not suitable to accomplish this task as it did not perform well enough to benefit the situation of the problem when compared to the other models.

Random Forest Classifier:

The Random Forest Classifier was slightly more encouraging. The identified model achieved 78.98% accuracy, with a precision of 78.56%, a recall of 78.98%, and an F1 score of 78.76%. These measures suggested that the Random Forest model was quite good, except it still holds some weaknesses that are common to all rare classes. The ROC AUC score was not available for this model, hence evaluation across several threshold values was limited to get a complete picture. The confusion matrix of the Random Forest model in Figure 2 shows that the model misclassified almost all classes, with a higher rate of misclassification in classes with lower representation.

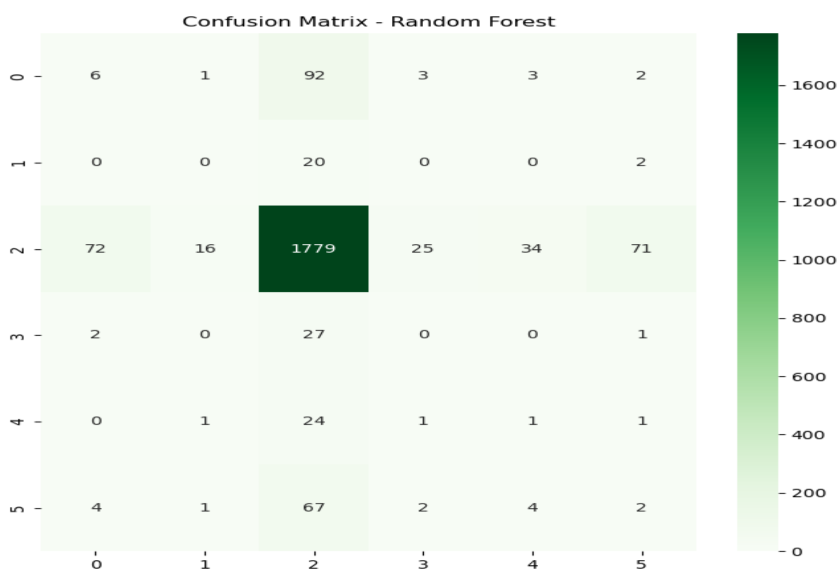

Figure 2. Random Forest Confusion Matrix

The feature importance plot (Figure 3) for the Random Forest model will project who the top contributors are in making predictions. The topmost physiological parameters indicating heart valve dysfunction point out that this analysis will be important for driving future clinical practice and research activities.

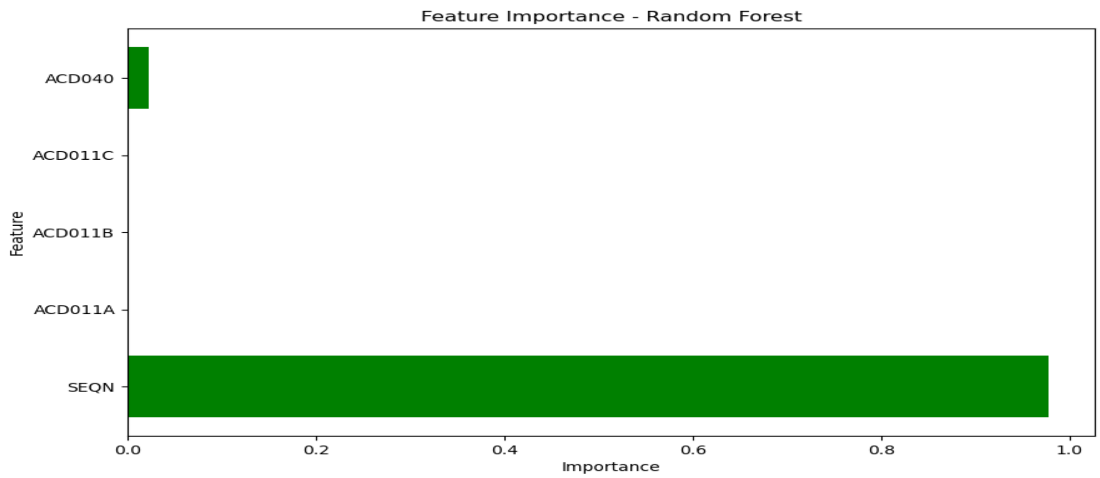

Figure 3: Feature Importance - Random Forest

Logistic Regression Model

The performance of the Logistic Regression model was better compared to the Random Forest Classifier under almost all metrics analysed. It attained an accuracy of 88.21%, precision of 77.80%, recall of 88.21%, and F1 score of 82.68%. These results may suggest that the Logistic Regression model, in particular, excels at identifying the true positive class of heart valve dysfunction. The confusion matrix of the model clearly specifies that the model did quite well in classification, making lesser errors as compared to the model developed using the Random Forest algorithm (Figure 4). Just like the Random Forest, it would not calculate but that should be noted as another limitation of the ROC AUC score for this model.

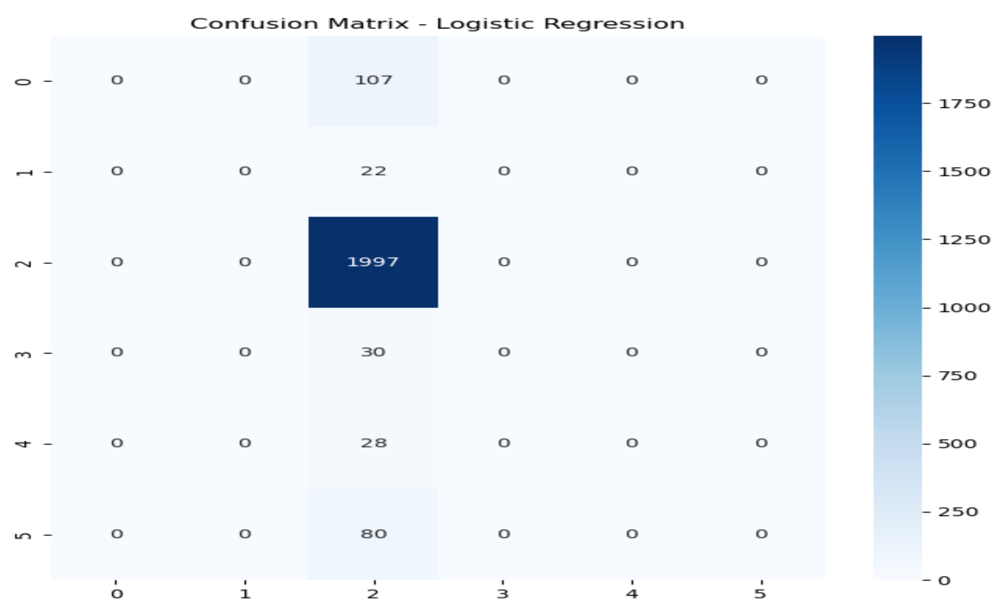

Figure 4: Confusion Matrix - Logistic Regression

Finally, the Receiver Operating Characteristic (ROC) curve for both models is in Figure 5. Although the ROC AUC scores are not obtained, the ROC curves continue to show how the trade-offs between the true positive rates and false positive rates of both models differ. They show both models' performance by how much it varies at different threshold values and also provide a visual representation of the performance of each model.

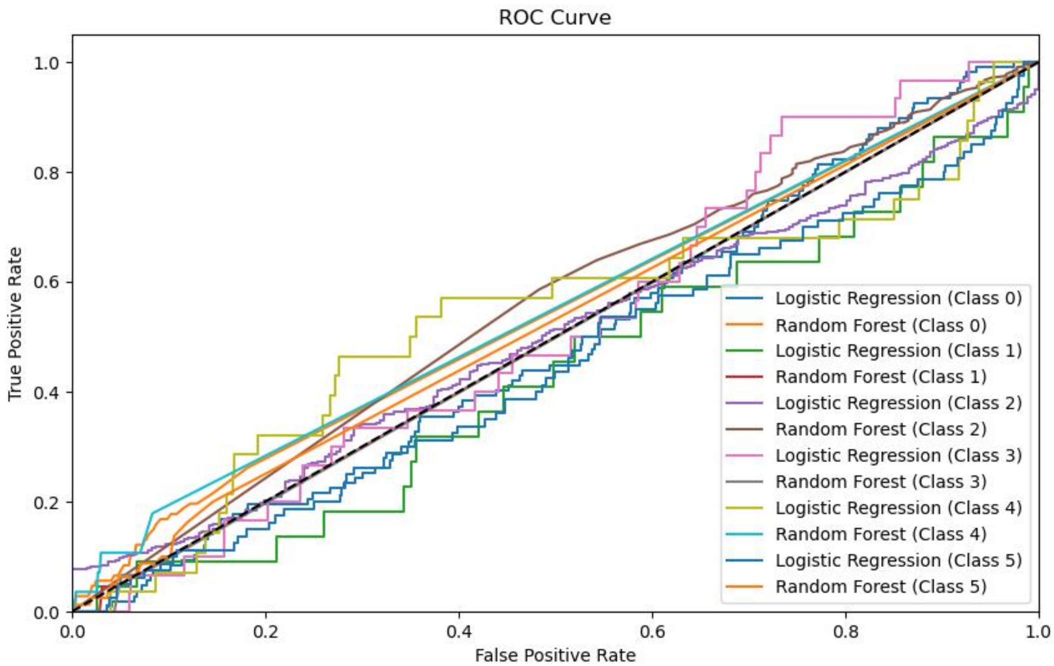

Figure 5: ROC Curve

4.2. Importance of Findings

Suitability of this kind of model evaluation could highlight the utility of the Logistic Regression model, to the point of being a possible tool in the prediction of heart valve dysfunction among adolescents. A logistic regression model scored higher accuracy, recall, and F1 value, which reflects the model's performance, strength, and area to easily identify heart valve dysfunction true positive cases. It is simple and easily interpretable, making it an ideal candidate in beneficial education, increasing awareness of heart health among the young, and encouraging early diagnosis.

The Random Forest Classifier is slightly less accurate than the Logistic Regression model, but it is just as informative with respect to the ability to manage intricate interactions of variables. Feature importance knowledge will provide you with the insight of which physiological parameters are most important for prognosis of heart valve dysfunction; hence, this may come in handy for further fine-tuning of the model in question or their clinical application in general.

Specifically, the high MSE of the SVM model indicates that its performance would be rather poor for this application. That said, in light of this result, it is also really important to say that model selection and the process of tuning computational tools to the particular properties of the dataset and the problem are vital.

4.3. Limitations

There are, however, several enhancements to the study in order to make generalization. The dataset mentioned has been observed to be of help only within its comprehensiveness to define the general viewpoint of the entire population of adolescents; otherwise, the dataset basically was exploring the health and nutritional status of either adult or child populations of the U.S. This limitation may thus affect any generalization to other adolescent populations, needing further validation studies on more specific datasets.

Another major limitation is that the ROC AUC score could not be computed into either the Logistic Regression or the Random Forest model. The ROC AUC score is quite instrumental as it allows understanding how the model performs across different threshold values, and without it, the depth of understanding is limited. This increased interaction between features represents a further improvement that could likely enhance performance for the same reason it should take into account more factors from the general physiological state that could be involved in heart valve disease.

More important, the developed models in this study are based on one data set and, therefore, are likely not general—probably because of other populations or other data sets. This will thus stand out as evidence for further validation in other diverse populations for the robustness of the developed models.

Although Logistic Regression model holds good in the prediction of heart valve dysfunction among adolescents, the limitations of a study indicate the necessity for continued research as well as model development for precision and enhancement of relevance.

4.4 Future Work

Future work will remain in the development of more representative datasets for targeting potential heart valve dysfunctions amongst the adolescents, which the models are further very applicable to and of great accuracy. Future work on more advanced prediction models should involve deep learning techniques in dealing with larger and more diversified datasets so that more complex patterns in the data can be better captured. The models developed in this study need to be clinically validated to determine their practical usefulness. RH would certainly benefit from a collective intelligence model, in which these models could be further refined with the help of healthcare professionals to make them representative of real-world scenarios. Finally, educational tools developed to embed these models would increase the dissemination regarding heart health to young generations. Integration within the curricula of high schools would allow for students' hands-on experience dealing with real-world health problems, development of computational thinking, and preparation of the next generation of researchers and productivity in health professionals.

V. Conclusion

The findings of this research therefore suggest that ample opportunity is presented for prediction using the simplistic computational models to be performed for adolescents suffering from heart valve dysfunction. The Logistic Regression model depicted superior overall performance with optimized accuracy, making it a reliable early-detection tool. The increased recall and F1 score of the model further emphasize robustness for the identification of true positives. This study highlights the importance of early detection practices in cardiology across younger populations. Such models can be used in schools to sensitize students about cardiac health and early diagnosis. Especially for educational purposes, the interpretability and ease of use of Logistic Regression models render them quite suitable for teaching heart health and computational modeling to students.

This basically means that even with such simple computational models, the study can be able to predict the function of the heart valves reasonably well. Results basically lay the groundwork for further, more advanced studies, and an avenue for the application of computational thinking in educational and clinical settings. The future work will focus on model refinements, clinical validation, and generalization for a larger population.

GitHub Code: https://github.com/Nishant27-2006/HeartValveML

References

1. Bonow, R. O., Carabello, B. A., Chatterjee, K., de Leon, A. C., Faxon, D. P., Freed, M. D., ... & Smith, S. C. (2008). 2008 focused update incorporated into the ACC/AHA 2006 guidelines for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation, 118(15), e523-e661.
2. Centers for Disease Control and Prevention. (n.d.). National Health and Nutrition Examination Survey. Retrieved from https://wwwn.cdc.gov/Nchs/Nhanes/Search/variablelist.aspx?Component=Questionnaire&Cycle=2017-2020
3. Nishimura, R. A., Otto, C. M., Bonow, R. O., Carabello, B. A., Erwin, J. P., Guyton, R. A., ... & Thomas, J. D. (2014). 2014 AHA/ACC guideline for the management of patients with valvular heart disease: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation, 129(23), 2440-2492.
4. Freed, L. A., & Levy, D. (2014). Prevalence and clinical outcome of mitral-valve prolapse. New England Journal of Medicine, 341(1), 1-7.
5. Quarteroni, A., Lassila, T., Rossi, S., & Ruiz-Baier, R. (2017). Integrated heart–coupling multiscale and multiphysics models for the simulation of the cardiac function. Computer Methods in Applied Mechanics and Engineering, 314, 345-407. Bonow

6. Syed, W. K., & Janamolla, K. R. (2024). How AI-driven Robo-Advisors Impact Investment Decisionmaking and Portfolio Performance in the Financial Sector: A Comprehensive Analysis.
7. Mohammed, S. (2024). AI-Driven Drug Discovery: Innovations and Challenges.
8. Thatikonda, R., Vaddadi, S. A., Dash, B., Panigrahi, A., Pati, A., & Sahu, B. (2023, November). Novel AI-based Prediction Approach for Improving Patient Treatment in Healthcare. In 2023 2nd International Conference on Ambient Intelligence in Health Care (ICAIHC) (pp. 1-5). IEEE.
9. Janamolla, K. R., & Syed, W. K. (2024). Global Banking Exploring Artificial Intelligence Role in Intelligent Banking to Automate Trading Platform. International Journal of Multidisciplinary Research and Publications (IJMRAP), 6(12), 163–168.
10. Mohammed, S. (2024a). The Impact of AI on Clinical Trial Management.
11. Sahoo, G., Nayak, A. K., Tripathy, P. K., Dash, B., Pati, A., & Panigrahi, A. (2023). Enhanced breast Cancer Relapse Prediction Based on Ensemble Learning Approaches.