e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

6381 907 438 6381 907 438 ijmrset@gmail.com www.ijmrset.com

International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705100 |

CLOUD CLEAN: Cloud Storage Space Sustainability Using Data Life Cycle Guard Scheme

Dr.J Stanly Jayaprakash, Amirthavarshini C, Monika M, Monisha A

Professor, Department of Computer Science and Engineering, Mahendra Institute of Technology, Namakkal, Tamilnadu, India

Department of Computer Science and Engineering, Mahendra Institute of Technology, Namakkal, Tamilnadu, India

ABSTRACT: The prevalence of chronic kidney disease (CKD) is currently rising globally. If not treated promptly, CKD can result in early death, increasing up healthcare expenses. Early detection of chronic kidney disease (CKD) may be possible with the use of artificial intelligence (AI) and machine learning (ML), which may prevent additional kidney damage. But if the models' logic is unclear, medical professionals can be eager to use AI. This study describes the establishment and inspection of an explainable CKD forecasting model that offers insights on how numerous clinical characteristics determine early CKD diagnosis. eXplainable AI (XAI) addresses the need for physicians recognize the output of AI models. A computational strategy that seeks a compromise between explainability and classification accuracy was used in the development of the model. The most important aspect of the paper is its explicable, data-driven methodology, which provides statistical insights into the role played by specific clinical variables in the early detection of chronic kidney disease (CKD). Consequently, the best explainable prediction model uses three features (specific gravity, hypertension, and hemoglobin) to implement an extreme gradient boosting classifier. It achieves accuracy of 99.3% (standard deviation 0.8) and 97.4% with new, unseen data and a 5-fold cross-validation, respectively. In addition, an explainability study reveals that hemoglobin, specific gravity, and hypertension are the features that have the most impact on the prediction. The inexpensive nature of an early CKD diagnosis due to the minimal number of features chosen suggests that treatment is practical for developing nations.

I. INTRODUCTION

Chronic kidney disease (CKD) has become a worldwide public health problem with increasing incidence (more than 800 million individuals) and prevalence (13.4% globally) which can lead to premature mortality for many patients (1.2 million people died from CKD).One of the few non-communicable diseases that has seen an increase in related deaths over the last two decades is chronic kidney disease (CKD). This is placing a significant burden on healthcare systems, particularly in middle-income and low-income countries where a high death rate is caused by inadequate transplantation of kidneys. Cardiovascular disorders are the primary cause of early morbidity and death experienced by CKD patients. Chronic kidney disease (CKD) is a non-communicable chronic disease with associated comorbidities that is typically brought on by diabetes and hypertension.

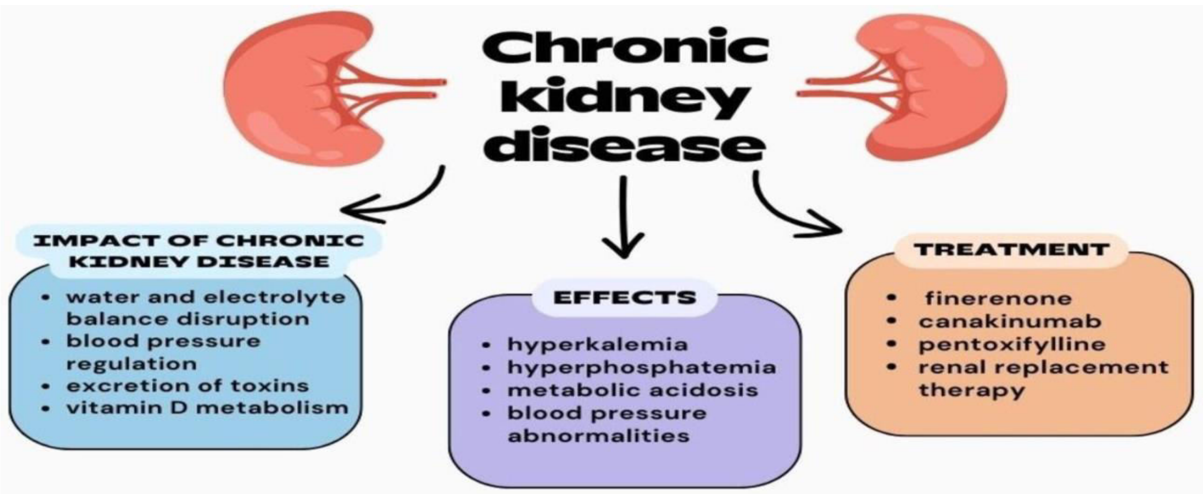

Fig 1: Novel Therapeutic Approaches in the Management of Chronic Kidney

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705100 |

In the event chronic kidney disease (CKD) is identified by laboratory testing that measures the estimated glomerular filtration rate (eGFR), the kidney has often lost 25% of its capacity and is experiencing severe damage that is progressing toward end-stage kidney disease. At this point, symptoms including leg water retention, acute exhaustion, weakness throughout the body, breathing problems, loss of appetite, or confusion can occur. In order to prevent an exponential increase in the patient's risk of death, hemodialysis or even kidney transplantation become essential if the irreversible progression cannot be prevented by managing underlying risk factors (hypertension, obesity, heart disease, age). As a result, early detection of chronic kidney disease (CKD) based on risk factors and its surveillance enable the start of therapeutic and preventive treatments that delay the development of kidney damage and extend patients' lives.

II. LITERATURE SURVEY

R. W. Major et al. developed a method for machine learning techniques were used to explore a model for prediction in ESKD on people with chronic kidney disease. Based on easy to access patient data that is practical for medical translation, the majority of classifiers showed satisfactory performance. Overall, this study's three machine learning models—logistic regression, naïve Bayes, and random forest—performed comparably to the KFRE. These results suggest that machine learning (ML) is a viable method for estimating the course of CKD, which may help doctors create individualised treatment plans for patients with the condition earlier on. Practically speaking these ML models with greater sensitivity scores might be preferred over the KFRE for patient screening[1].

A technique based on machine learning was developed by AnusornCharleonnan et al. to identify chronic kidney disease. The study applied four machine learning classifiers, namely Decision Tree, K-Nearest Neighbours (KNN), Logistic Regression (LR), and Support Vector Machine (SVM). To determine the best classifier, the outcomes from these were evaluated. With an accuracy of 99%, SVM was shown to be the most efficient classifier among the others [2].

In their research, PramilaArulanthu and EswaranPerumal employed feature selection techniques to decrease the amount of features needed to identify kidney diseases.Aclassification method for chronic kidney disease using machine learning techniques. The study included four machine learning classifiers, namely Decision Tree, K-Nearest Neighbors (KNN), Logistic Regression (LR), and Support Vector Machine (SVM). Jrip, SMO, IBK, and Naïve Bayes were the several algorithms that were employed. The outcome from the original dataset and the outcome obtained from the reduced features were compared [3].

Wasle et al. developed a method for statistical analysis.Theresearchers utilised different machine learning techniques to analyse the Chronic Kidney Disease dataset. The Random Forest algorithm outperforms the other approaches in terms of accuracy in classification when they were applied along with Nave Bayes, Decision Trees, and Random Forest to enhance prediction [4].

Bemando et al. implemented classifier methods such Gaussian NB, Bernoulli NB, and Random Forest in order to investigate the association between blood-related disorders and their characteristics. These three algorithms use a range of techniques to predict and provide statistical findings. We found that the Nave Bayes predicted accuracy in this experiment was higher than the other the techniques [5].

SwathiPrabhu et al. has observed the early diagnosis and treatment of carcinoma can decrease mortality rates and increase survival rates. Medical/Clinical examination (CT), laboratory testing (MRI, PET), imaging tests (X-ray) and biopsy are involves in the diagnosis of tumor. Examining slides using a low magnifications microscope (5X, 9X), as well as at high magnifications (30X, 50X, 90X), to assess the histomorphology at the cellular level. The outcome depends on tissue-level annotations unlike the contemporary methods that depend on pixel/ROI annotation [6].

In the healthcare industry, Kumar and Polepaka developed a method for predicting sickness. They used CNN and Random Forest in addition to other machine learning techniques. These methods perform better for F1-score, precision, recall, and classification of sickness datasets. Random Forest performed better statistically and accurately than the other algorithms in this trial [7].

Ahmed J. Alijaaf et al. developed a model that uses the UCI repository's dataset to diagnose chronic kidney disease (CKD). Two urine and blood tests yielded twenty-four attributes for the dataset. The model was built using only thirty percent of the attributes. At a sensitivity of 98.97%, specificity of 100%, and AUC of 99.5%, four supervised machine learning classifiers were employed to predict the illness [8].

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705100 |

kidney disease using four algorithms: Multilayer Perceptron (MLP), Probabilistic Neural Networks (PNN), Radial Basis Function (RBF), and Support Vector Machine (SVM) techniques. The model was created using the various stages of chronic kidney disease (CKD), which were categorised based on the measured Glomerular Filtration Rate (GFR) values. When compared to other classification algorithms, PNN achieved the most impressive results, with a general precision of 96.7%. The MLP took a few seconds to execute, while the PNN took 12 seconds [9].

A method for predicting datasets of diseases related to medicine was created by Sing et al. The several Machine Learning ensemble algorithms are used to prepare the model using the training set. The hyperparameters of every ensemble classifier are adjusted to yield the optimal parameters that yield the best model for forecasting the patient's chronic renal disease.They employed a support vector machine classifier for better prediction. Over the course of time, the author increased accuracy to 91 percent, with a range of 73 to 91 percent [10].

III.METHODS

In ROC curves, the graphical comparison of two or more analytical tests can be performed at the same time in one graph, which is an advantage over individual values of precision and recall [53]. Furthermore, the classifier which provides a curve closer to the left upper corner shows better performance [37]. Figure 4 shows that the curves provided by the classifiers used in this study are almost on the left upper corner, providing evidence of the high performance of the trained models for detecting and diagnosing CKD.
The aforementioned tables and figures show that the models trained based on the CKD data are significantly reliable in terms of model accuracies, model performance, model sensitivities, F-measures, and the significantly reliable curves provided by the classifiers. This study has trained several models described above with an outcome of higher performance for all; therefore, they can be used as predictive models to help healthcare practitioners in detecting and diagnosing chronic kidney diseases and can also be an integral part of the CKD intervention decision-making process.

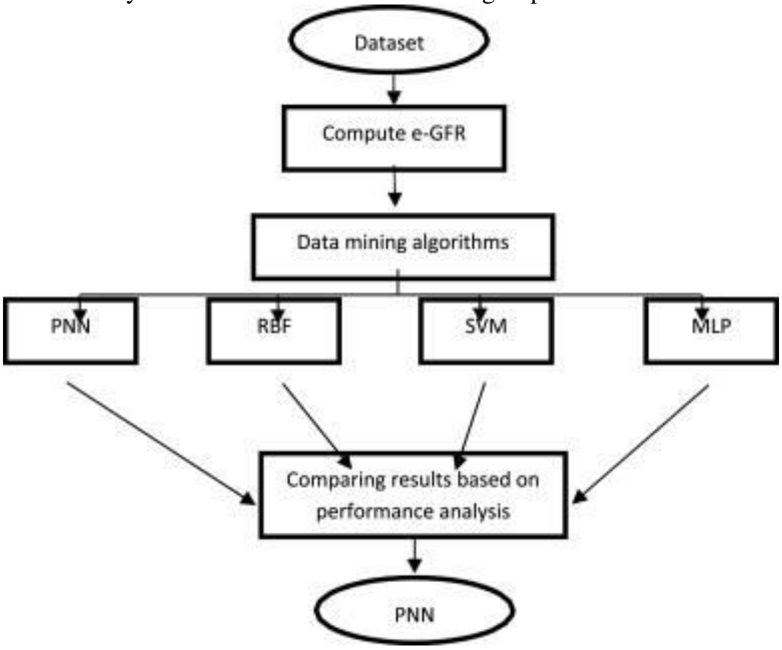

Fig 2: Work Flow

Moreover, due to the higher performance of the proposed models, they can be used as a decision support system for quick medical decisions in order to diagnose the CKD patients early based on the predominant features discussed in this study. Similarly, the feature selection process was applied in order to select the most relevant features for detecting and diagnosing CKD. Therefore, the soaring costs can be controlled by conducting fewer clinical tests and avoiding other identical tests, which may aid Third World survival.
The study employed different evaluation methods to examine the models, which increases the reliability of diagnosing the cases. In addition, the simplicity of the proposed method makes the implementation and deployment of such a system achievable.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705100 |

IV. RESULT ANALYSIS

The use of artificial intelligence, in general, and machine learning, in particular, has made it possible to organize and structure the unorganized and unstructured data in such manner to have an essential part of a business decision support system. The extraction of meaningful insights from raw data and the subsequent construction of prediction models based on those data are advantages of ML methods which are broadly used in the healthcare industry for predictive analytics and decision support systems that help medical practitioners in diagnosing several diseases, among other clinical practices. There are numerous studies available in the literature that utilized ML techniques for predicting CKD. The commonly used methods in the literature are DT, KNN, RF, SVM, and NB. In this study, the ML methods used for detecting and diagnosing CKD are discussed in the following sections.

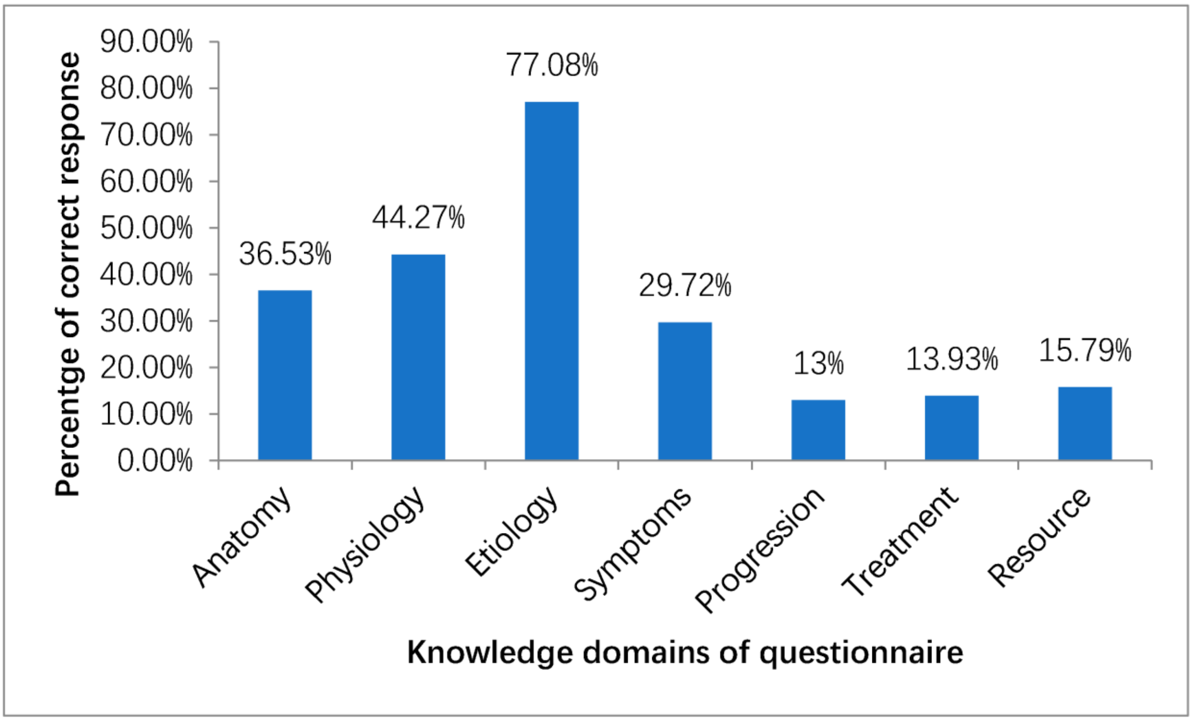

Fig 3: Result analysis

This method utilizes the entered data and creates multitudes of DTs at the time of training and delivers a mean prediction of each tree. In RF, the classification is conducted through nominating different randomized DTs on the final score where each DT is randomized based on a bootstrap resampling method with arbitrary feature selection. This practice is repeated throughout the forest for all trees based on various bootstrap data, and the new samples are labeled to the class having the majority of votes

To advance in the line of trustworthiness and transparency of our model, we propose as future works to perform an external validation with other datasets that contain the same group of features to evaluate the generalization capability of the model in the early CKD diagnosis. Additionally, this external validation could be deployed in a clinical setting with the aim at also gathering insights from clinicians about the explainability results and discussing how it could affect the CKD treatment plans. Therefore, we could confirm that the explainability approach presented in this paper would provide clinicians with an easier understanding and interpretability of how CKD is diagnosed early with a reduced group of indicators. This information would allow them to also focus on tackling relevant features and their values to avoid the CKD onset or even to revert its progress.

V. CONCLUSION

The development and evaluation of an explainable prediction model for CKD early diagnosis. The main goal is to show how XAI contributes to improving prediction models used in the medical field. This research also pursues to exemplifying how to address the existing trade-off between accuracy and explainability when dealing with black box AI models. Therefore, using an automated optimization framework, the best combination of the ensemble tree

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705100 |

algorithm and the number of features are selected to provide the best balanced model according to the classification and explainability metrics. The optimal balanced explainable model detected by the framework was an XGBoost classifier that used three features for the CKD prediction: hemoglobin (hemo), specific gravity (sg), and hypertension (htn). After conducting an explainability analysis with different post-hoc techniques, the features' relevance in descending order of importance was found to be hemo, sg, and htn. The prediction model developed in this work achieved the classification performance of the best CKD prediction models identified in the literature with the least number of features selected compared to the other works.

REFERENCES

1. Major, R. W. *et al.* The Kidney Failure Risk Equation for prediction of end stage renal disease in UK primary care: An externalvalidation and clinical impact projection cohort study. *PLoS Med.* **16**, e1002955. https:// doi. org/ 10. 1371/ journ al. pmed. 1002955(2019).
2. A. Charleonnan, T. Fufaung, T. Niyomwong, W. Chokchueypattanakit, S. Suwannawach, and N. Ninchawee, "Predictive analytics for chronic kidney disease using machine learning techniques," in *2016 Management and Innovation Technology International Conference (MITicon)*, Oct. 2016, p. MIT-80-MIT-83, doi: 10.1109/MITICON.2016.8025242.
3. P. Arulanthu and E. Perumal, "Predicting the Chronic Kidney Disease using Various Classifiers," in *2019 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT)*, Dec. 2019, pp. 70–75, doi: 10.1109/ICEECCOT46775.2019.9114653.
4. R.S. Walse, G.D. Kurundkar, S.D. Khamitkar, A.A. Muley, P.U.Bhalchandra, S.N. Lokhande, Effective use of naive bayes, decisiontree, and random forest techniques for analysis of chronickidney disease, in *International Conference on Information andCommunication Technology for Intelligent Systems*. ed. by T. Senjyu, P.N. Mahalle, T. Perumal, A. Joshi (Springer, Singpore, 2020).
5. C. Bemando, E. Miranda, M. Aryuni, "Machine-Learning-BasedPrediction Models of Coronary Heart Disease Using Naive Bayesand Random Forest Algorithms," in *2021 International Conferenceon Software Engineering & Computer Systems and 4th InternationalConference on Computational Science and InformationManagement (ICSECS-ICOCSIM)*, (IEEE, 2021), pp. 232237.
6.Prabhu, S., Prasad, K., Robels-Kelly, A. and Lu, X., 2022. AI-based carcinoma detection and classification using histopathological images: A systematic review. *Computers in Biology andMedicine*,p.105209.7.R.P.Ram Kumar, SanjeevaPolepaka, Performance comparison of random forest classifier andconvolution neural network in predicting heart diseases, in *Proceedings of the Third InternationalConference on Computational Intelligence and Informatics*. ed. by K. SrujanRaju, A. Govardhan, B. PadmajaRani, R. Sridevi, M.RamakrishnaMurty (Springer, Singapore,2020).8. Ahmed J. Aljaaf*et al.*, "Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics," in *2018 IEEE Congress on Evolutionary Computation(CEC)*, Jul. 2018, pp. 1–9, doi:10.1109/CEC.2018.8477876. .9.E.-H. A.Rady and A. S. Anwar, "Prediction of kidney disease stages using data miningalgorithms," *Inform. Med. Unlocked*, vol.15, p.100178, Jan. 2019, doi: 10.1016/j.imu.2019.100178. 10. H. Singh, N. V. Navaneeth, G. N. Pillai, "Multisurface proximal SVM based decision trees for heart disease classification," in *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, (IEEE 2019), pp. 13–18.11.Wickramasinghe M, Perera D, Kahandawaarachchi K. Dietary prediction for patientswith chronic kidneydisease (ckd) by considering blood potassium level using machine learning algorithms. 2017 IEEE Life Sciences Conference (LSC). IEEE; 2017. p. 300–303.12.G. Chen *et al.*, "Prediction of Chronic Kidney Disease Using Adaptive Hybridized DeepConvolutional Neural Network on the Internet of Medical Things Platform," *IEEE Access*, vol. 8, pp. 100497–100508, 2020, doi: 10.1109/ACCESS.2020.2995310.13.S.D. Desai, S. Giraddi, P. Narayankar, N.R. Pudakalakatti, S. Sulegaon, *Backpropagation neural network versus logistic regressionin heart disease classification in advanced computing and communicationtechnologies* (Springer, Singapore, 2019). 14. D.D. Patil, R.P. Singh, V.M. Thakare, A.K. Gulve, Analysis of ecg arrhythmia for heart disease detection using svm and cuckoosearch optimized neural network. Int. J. Eng. Technol. **7**(217),27–33(2018). (2018).15. N. Liu, Z. Lin, J. Cao, Z. Koh, T. Zhang, G.-B. Huang, W. Ser, M.E.H. Ong, An intelligent scoring system and its application tocardiac arrest prediction. IEEE Trans. Inf Technol. Biomed.**16**(6),1324–1331(2012). 16.U. Rajendra Acharya, Oh. ShuLih, Y. Hagiwara, J.H. Tan, M. Adam, A. Gertych, R.S. Tan, A deep convolutional neural network model to classify heartbeats. Comput. Biol. Med. **89**, 389–396(2017). 17.Nusinovici, S. et al. Logistic regression was as good as machine learning for predicting major chronic J. Clin. Epidemiol. 122, 56–69. https://doi.org/10.1016/j.jclinepi.2020.03.002(2020).
18. Tangri, N., Stevens, L.A., Griffith, J., Tighiouart, H., Djurdjev, O., Naimark, D., Levin, A., Levey, A.S., 2011. A predictive model for progression of chronic kidney disease to kidney failure. Jama 305, 1553–1559.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705100 |

19. Xiao, J. *et al.* Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. *J.Transl. Med.* **17**, 119. https:// doi. org/ 10. 1186/ s12967
019- 1860-0 (2019).
20. Sterne, J. A. *et al.* Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ (Clin. Res. Ed.)* **338**, b2393. https:// doi. org/ 10. 1136/ bmj. b2393 (2009).

INNO SPACE
SJIF Scientific Journal Impact Factor

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com