Enhancing Healthcare with Classification Modeling of Chronic Kidney Disease

Gopu A P[1], Deebak M[2],Deepankumar S[3], Akshaykumar R G[4]
[1] *Assistant Professor, Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India* [2,3,4] *Student, Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India*

***Abstract*— Chronic kidney disease (CKD) is a global health problem with a high morbidity and mortality rate that causes other diseases. Patients sometimes overlook the condition since there are no obvious adverse effects in the early stages of CKD. Early detection of CKD allows patients to receive appropriate treatment to improve health outcomes and quality of life. Because of their rapid and exact recognition execution, machine learning models can effectively assist medical professionals in achieving this goal. In this study, we present a logistic regression approach for diagnosing CKD. We compared proposed algorithms, including Naïve Bayes, decision tree, k-star, logistic, and Support Vector Machine (SVM) to determine the maximum accuracy. AI storage, with a large number of missing characteristics. Missing characteristics are commonly observed in clinical settings, as patients may miss a few estimates for a variety of reasons. By breaking down the errors created by the established models, we suggested an integrated model that combines estimated relapse and irregular woods using perceptron.**
Keywords: Chronic Kidney Disease, Machine Learning, Naive Bayes, Decision Tree, Logistic Regression, Support Vector Machine (SVM), K-Star Algorithm, Classification Performance.

I. INTRODUCTION

The research has led to improvements in the diagnosis of CKD.This work investigates how CKD can be diagnosed by using machine learning (ML) techniques. ML algorithms have been a driving force in detection of abnormalities in different physiological data, and are, with a great success, employed in different classification tasks. The diagnostic categories of the samples are used to fill in the missing data in the aforementioned models using mean imputation. Since the diagnostic results of the samples are unknown, their method cannot be used. In point of fact, prior to being diagnosed, patients may fail certain tests for a variety of reasons. Furthermore, when missing values are present in categorical categories, mean imputation-derived data may significantly deviate from the actual values. Our main goal is to detect CKD at its early stage. There are some backdrops in the existing system like the results are biased,less accuracy and low performance. The Existing paper [1] has advantages like it effectively highlights the global impact of chronic kidney disease and the importance of early detection. There is also a backdrop that is less accurate. As well as the paper [2] shows that the adoption of machine learning for CKD diagnosis is a strength. But [2] it has a negative side that the assumption of similar measurements may introduce biases.

A. Chronic kidney disease (CKD)

Chronic kidney disease (CKD) is a condition in which kidney function gradually declines over several months or years. At first, there are no symptoms; However, disorientation, leg swelling, fatigue, vomiting, and a lack of appetite may follow. Anaemia, high blood pressure, broken bones, and heart disease is among the complications. Gout, diabetes, high blood pressure, polycystic kidney disease, and other conditions can all contribute to chronic kidney disease. A family history of chronic renal disease is one risk factor. A blood test to determine the estimated glomerular filtration rate (eGFR) and a urine test to determine albumin are used to make the diagnosis. A kidney biopsy or ultrasound may be used to identify the underlying cause.

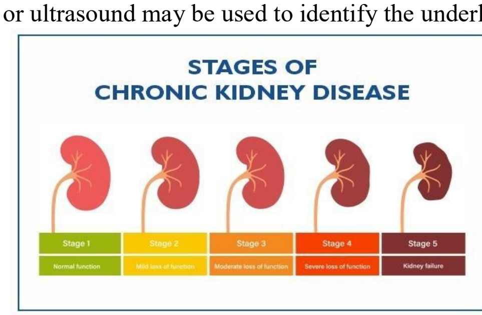


Figure 1. CKD

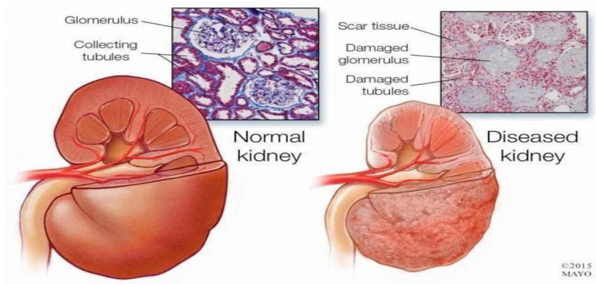


Figure 2. Structure of Kidney – Normal and diseased

B. Machine Learning

Machine learning (ML) is the study of computer algorithms that automatically get better over time. It is speculated to be a subset of artificial intelligence. Sample data, or "training data," is used to build a model by machine learning algorithms in order to make predictions or judgments without being explicitly programmed to do so. A lot of applications, like email filtering and computer vision, use machine learning algorithms when it would be difficult or impossible to develop traditional algorithms that can do the job. However, machine learning is not limited to statistical learning. A subset of machine learning is closely related to computational statistics, which focuses on predictions made by computers; However, there are other types of machine learning as well as statistical learning. The theory, method, and application areas of machine learning are provided by the study of mathematical optimization. A similar field of study is data mining, which focuses on exploratory data analysis through unsupervised learning. The interaction by which PCs figure out how to finish jobs without being explicitly trained is known as AI. related work

The previous research highlights the effectiveness of machine learning in prediction of chronic kidney disease. The incidence, prevalence, and development of chronic kidney disease (CKD) have changed throughout time, particularly in nations with diverse social determinants of health. In most countries, diabetes and hypertension are the leading causes of CKD. According to the global recommendations, CKD is a disorder that results in decreasing kidney function over time, as seen by glomerular filtration rate (GFR) and kidney damage markers. People with CKD are likely to die at a young age. Doctors must diagnose various CKD-related diseases early on because early detection can prevent or even reverse renal damage. Early detection can lead to better therapy and care for patients. In many regional hospitals and clinics, there is a paucity of nephrologists or general practitioners who can diagnose the symptoms. This has resulted in patients having to wait longer for a diagnosis. The following are some related papers that show the importance for the early prediction of chronic kidney disease and the previous work that are related to machine learning in chronic kidney disease prediction.

1. 1. A Comprehensive Unsupervised Framework for Chronic Kidney Disease Prediction

This paper presents a far reaching solo structure for the expectation of persistent kidney sickness in this review. Offered in countries with a variety of social determinants of health, the incidence, prevalence, and course of chronic kidney disease (CKD) have changed over time. Diabetes and high blood pressure are the most common causes of CKD in most countries. A condition known as chronic kidney disease (CKD) is characterized by indicators of kidney damage and progressive decreases in glomerular filtration rate (GFR). Patients with CKD are more likely to die young. Doctors must quickly identify various CKD-related disorders because early detection can prevent or even reverse kidney damage. Early identification can improve patient care and treatment. In many outlying hospitals and clinics, there are not enough nephrologists or general practitioners to recognize the symptoms.

2. Machine Learning Method for Diagnosing Chronic Kidney Disease

Uses machine learning to diagnose chronic kidney disease. suggest that Chronic Kidney Disease (CKD) is a global health issue that is responsible for a large number of other disorders, as well as a significant number of deaths. Because there are no obvious symptoms, patients frequently miss the early stages of CKD. If CKD is detected early, patients can receive prompt treatment to slow its progression. Due to their quick and precise identification capabilities, machine learning models can successfully assist doctors in achieving this goal. A machine learning approach to CKD diagnosis is proposed in this paper. There are a ton of missing qualities in the CKD informational collection, which was taken from the AI storehouse at the College of California, Irvine (UCI). The missing data for each incomplete sample were processed using KNN imputation by selecting multiple complete samples with the most comparable measurements. Missing values are common in real-world medical scenarios because people may miss certain measures for a variety of reasons.

[2.] *3. A Machine Learning Method with Filter-based Feature Selection for Improved Prediction of Chronic Kidney Disease*

The high prevalence of chronic kidney disease (CKD) is a significant public health concern globally. The condition has a high mortality rate, especially in developing countries. CKD often goes undetected since there are no obvious early-stage symptoms. Meanwhile, early detection and on-time clinical intervention are necessary to reduce the disease progression. Machine learning (ML) models can provide an efficient and cost-effective computer-aided diagnosis to assist clinicians in achieving early CKD detection. This research proposed an approach to effectively detect CKD by combining the information-gain-based feature selection technique and a cost-sensitive adaptive boosting (AdaBoost) classifier. An approach like this could save CKD screening time and cost since only a few clinical test attributes would be needed for the diagnosis.

3. 4. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017

Health system planning requires careful assessment of chronic kidney disease (CKD) epidemiology, but data for morbidity and mortality of this disease are scarce or non-existent in many countries. We estimated the global, regional, and national burden of CKD, as well as the burden of cardiovascular disease and gout attributable to impaired kidney function, for the Global Burden of Diseases, Injuries, and Risk Factors Study 2017. We use the term CKD to refer to the morbidity and mortality that can be directly attributed to all stages of CKD, and we use the term impaired kidney function to refer to the additional risk of CKD from cardiovascular disease.

4. 5. Intelligent Diagnostic Prediction and Classification Models for Detection of Kidney Disease

Kidney disease is a major public health concern that has only recently emerged. Toxins are removed from the body by the kidneys through urine. In the early stages of the condition, the patient has no problems, but recovery is difficult in the later stages. Doctors must be able to recognize this condition early in order to save the lives of their patients. To detect this illness early on, researchers have used a variety of methods. Prediction analysis based on machine learning has been shown to be more accurate than other methodologies. This research can help us to better understand global disparities in kidney disease, as well as what we can do to address them and coordinate our efforts to achieve global kidney health equity. This study provides an excellent feature-based prediction model for detecting kidney disease.

I. PROPOSED SYSTEM

The input is the CKD dataset with its various properties. During pre-processing, redundant data and unknown properties are removed. All of the chosen characteristics and features have been selected. To improve classification performance, algorithms like Naive Bayes, Decision Tree, Kstar, Logistic, And SVM are utilized. Accuracy, recall, precision, and the f-measure will all receive grades. These parameters will be presented as a visual representation. Using a lot of CKD data and the model's accuracy on their test data, they developed a neural organization-based classifier and used image enrolment to find changes in the kidney's morphology. Additionally, the majority of previous studies utilized the UCI AI repository's CKD informational index. This study looks into how machine learning (ML) can be used to diagnose chronic kidney disease. The successful application of ML algorithms, which have been a driving force in the detection of anomalies in a variety of physiological data, is benefiting a number of categorization tasks. Figure 3 shows the block diagram for the proposed system.

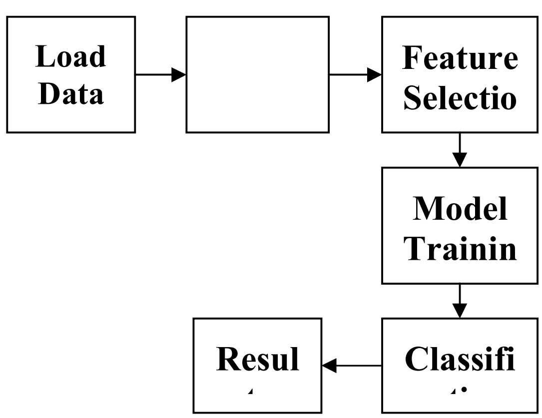


Figure 3. Block Diagram

A. Loading The Dataset

Loading data is the first step in every data analysis or machine learning project. In this case, the Chronic Kidney Disease (CKD) dataset is put into the software. The dataset typically includes a variety of variables and corresponding CKD diagnosis outcomes for each patient. Figure 4 shows the loading of the dataset, it gives you access and control over the data, allowing you to conduct additional analysis. Figure 4 represent Loading of the Dataset in System.

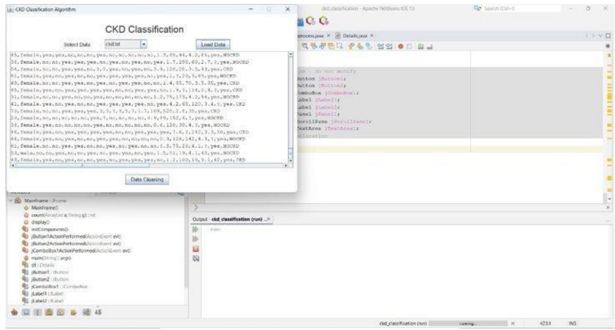


Figure 4. Loading the Dataset

B. Preprocessing

Pre-processing is an important stage in data analysis. It entails cleaning, converting, and organizing data to prepare it for machine learning or statistical analysis. Pre-processing for CKD data may include handling missing values, removing duplicate records, normalizing or scaling numerical attributes, and encoding categorical variables. This step is critical for ensuring the quality and integrity of the data before using it to train machine learning models. Table 1 shows Dataset Feature Description the description of all the available attributes.

C. Feature Selection

Feature selection is the process of determining and selecting the most relevant qualities (features) from a dataset for model training. The goal is to reduce dimensionality while maintaining the most useful features. Various techniques, such as correlation analysis, mutual information, and feature importance scores, can be used to determine which features are most important for CKD prediction. Choosing the right features can improve model performance while reducing computational overhead.

D. Classification Performance

Classification performance is an assessment of how well machine learning models can classify or predict outcomes, such as the presence or absence of CKD. Several measures are routinely used to assess classification model performance, including accuracy, precision, recall, F1-score, and the confusion matrix. The accuracy ratio measures the model's correctness by comparing correctly predicted instances to the total number of instances in the dataset. Precision measures the fraction of real positive predictions among all positive predictions, allowing to evaluate the model's ability to avoid false positives. Recall (Sensitivity) analyzes the fraction of true positive predictions among all actual positive instances, showing the model's capacity to catch all relevant cases. F1-score is the harmonic mean of precision and recall, allowing a fair evaluation of a model's performance.

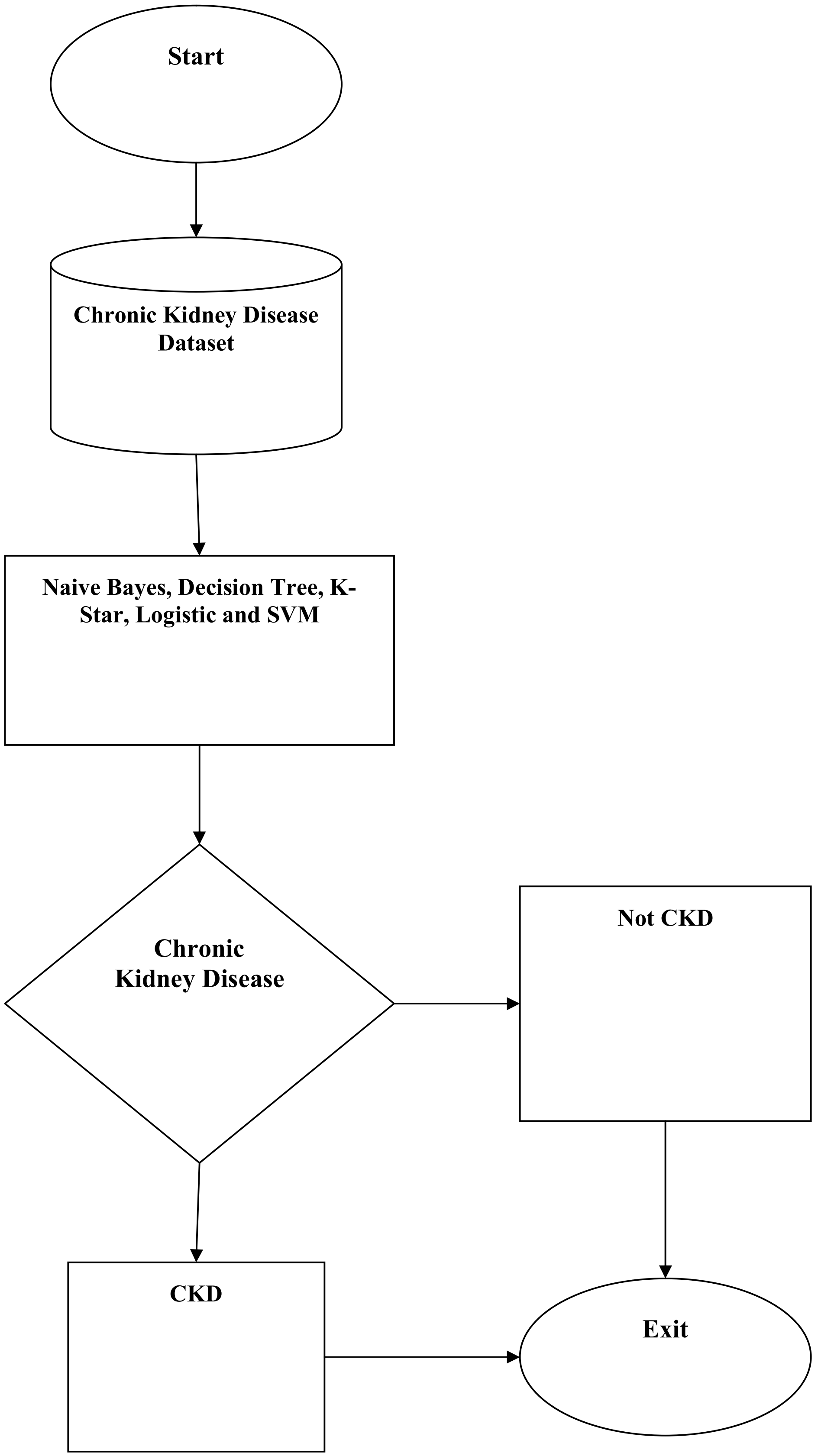


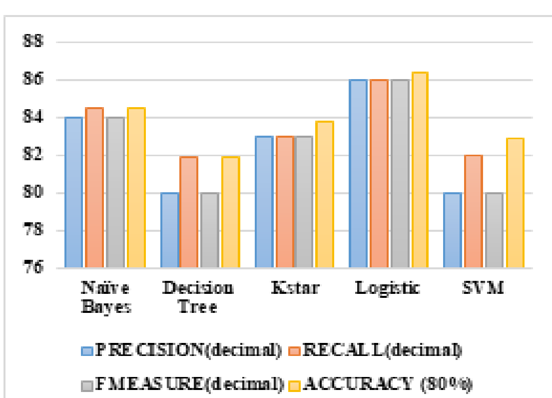
Figure 5. Proposed Flow Diagram

II. RESULT ANALYSIS

The table below highlights the performance parameters, specifically precision, recall, F-measure, and accuracy, for five different classification methods applied to the same dataset. Naïve Bayes had a precision of 0.849, recall of 0.845, F-measure of 0.857, and accuracy of 84.516%. The decision tree had precision, recall, F-measure, and accuracy scores of 0.802, 0.819, 0.806, and 81.935%, respectively. Kstar achieved a precision of 0.831, recall of 0.839, F-measure of 0.833, and accuracy of 83.871%. Logistic regression produced a precision of 0.86, recall of 0.865, F-measure of 0.862, and accuracy of 86.452%. Finally, the SVM method has a precision of 0.807, recall of 0.826, F-measure of 0.808, and accuracy of 82.581%. These metrics provide a comprehensive evaluation of the algorithms' effectiveness in classifying instances, with precision denoting the accuracy of positive predictions, recall representing the proportion of actual positives correctly identified, the F-measure balancing precision and recall, and accuracy measuring the overall correctness of predictions at an 80% confidence level.

ALGORITHM	PRECISION (decimal)	RECALL (decimal)	F MEASURE (decimal)	ACCURACY (80%)
Naïve Bayes	0.849	0.845	0.857	84.516
Decision Tree	0.802	0.819	0.806	81.935
Kstar	0.831	0.839	0.833	83.871
Logistic	0.86	0.865	0.862	86.452
SVM	0.807	0.826	0.808	82.581

Table 2. Comparison Table



ISSN: 2319-6319l

Figure 6. Comparison Graph

III. CONCLUSION

Finally, different levels of efficacy in completing the task are shown by the performance evaluation of five machine learning algorithms: Naive Bayes, Decision Tree, Kstar, Logistic, And SVM. When a high degree of confidence in positive predictions is required, Naïve Bayes and Logistic Regression are excellent choices due to their better precision and accuracy. Despite being a little less accurate, Decision Tree, Kstar, and SVM nevertheless provide respectable results across a variety of parameters. To achieve the best possible balance between precision, recall, and overall accuracy, the most appropriate algorithm should be selected after taking into account the needs of the particular application as well as variables like interpretability and training time.

IV. FUTURE WORK

Early detection is a major obstacle in managing CKD. It is challenging to diagnose and treat CKD before it progresses because many patients are asymptomatic in the early stages. The development of more sensitive and specific tests to identify CKD earlier, when treatment can be more effective, may be the focus of future research. CKD is a disease that affects different people in different ways and is complicated and caused by multiple factors. Approaches to personalized medicine that take into account specific patient characteristics, like genetics and lifestyle factors, may help tailor treatments to each patient's specific requirements

REFERENCES

[1] Linta antony, ‘‘a comprehensive unsupervised framework for chronic kidney disease prediction 2022,’’ kidney int. Supplements, vol. 12, no. 1, pp. 7–11, Apr. 2022, doi: 10.1016/j.kisu.2021.11.003.

[2] Iongming qin , a machine learning methodology for diagnosing chronic kidney disease. (2019). World health statistics 2019: monitoring health for the SDGS, sustainable development goals. Accessed: feb. 7, 2023. [online]. Available: https://apps.who.int/iris/handle/ 10665/324835.

[3] Sarah A, et al."A Machine Learning Method with Filter-Based Feature Selection for Improved Prediction of Chronic Kidney Disease." Bioengineering 9, no. 8 (July 2022): 350. https://doi.org/10.3390/bioengineering9080350.

[4] Jicun Zhu et.al.,"Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017." The Lancet. 2020 Feb 29; 395(10225): 709–733. DOI: 10.1016/S0140-6736(20)30045-3. PMCID: PMC7049905. PMID: 32061315..

[5] Ramesh et.al., "Intelligent Diagnostic Prediction and Classification Models for Detection of Kidney Disease." Healthcare (Basel) 10, no. 2 (2022): 371. DOI: 10.3390/healthcare10020371. PMCID: PMC8871759. PMID: 35206985.

[6] J. Qezelbash-Chamak, S. Badamchizadeh, K. Eshghi, and Y. Asadi, ‘‘A survey of machine learning in kidney disease diagnosis,’’ Mach. Learn. Appl., vol. 10, Dec. 2022, Art. no. 100418, doi: 10.1016/j.mlwa.2022. 100418.

[7] N. Z. Benisi, M. Aminian, and B. Javadi, “Blockchain-based decentralized storage networks: A survey,” Journal of Network and Computer Applications, vol. 162, p. 102656,2020.

[8] C.Nagarajan and M.Madheswaran - ‘Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter’ - Journal of ELECTRICAL ENGINEERING, Vol.63 (6), pp.365-372, Dec.2012.

[9] C.Nagarajan and M.Madheswaran - ‘Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis’- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011.

[10] C.Nagarajan and M.Madheswaran - ‘Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques’- Taylor & Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011.

[11] C.Nagarajan and M.Madheswaran - ‘Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis’- Iranian Journal of Electrical &Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.

[12] Nagarajan C., Neelakrishnan G., Akila P., Fathima U., Sneha S. “Performance Analysis and Implementation of 89C51 Controller Based Solar Tracking System with Boost Converter” Journal of VLSI Design Tools & Technology. 2022; 12(2): 34–41p.

[13] Nagarajan, G.Neelakrishnan, R. Janani, S.Maithili, G. Ramya “Investigation on Fault Analysis for Power Transformers Using Adaptive Differential Relay” Asian Journal of Electrical Science, Vol.11 No.1, pp: 1-8, 2022.

[14] G.Neelakrishnan, K.Anandhakumar, A.Prathap, S.Prakash “Performance Estimation of cascaded h-bridge MLI for HEV using SVPWM” Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:750-756

[15] G.Neelakrishnan, S.N.Pruthika, P.T.Shalini, S.Soniya, “Perfromance Investigation of T-Source Inverter fed with Solar Cell” Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:744-749

[16] C.Nagarajan and M.Madheswaran, “Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation” has been presented in ICTES’08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007

[17] M Suganthi, N Ramesh, “Treatment of water using natural zeolite as membrane filter”, Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022

[18] M Suganthi, N Ramesh, CT Sivakumar, K Vidhya, “Physiochemical Analysis of Ground later used for Domestic needs in the Area of Perundurai in Erode District”, International Research Journal of Multidisciplinary Technovation, pp: 630-635, 2019

[19] M. J. H. Faruk, H. Shahriar, M. Valero, S. Sneha, S. I. Ahamed, and M. Rahman, “Towards blockchain-based secure data management for remote patient monitoring,” in 2021 IEEE International Conference on Digital Health (ICDH). IEEE, 2021, pp. 299–308.

[20] A. Al Mamun, M. U. F. Jahangir, S. Azam, M.

[21] S. Kaiser, and A. Karim, “A combined framework of interplanetary file system and blockchain to securely manage electronic medical records,” in Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Springer, 2021, pp. 501–511

[22] X Lu, S. Fu, C. Jiang, and P. Lio, “A fine- grained iot data access control scheme combining attribute-based encryption and blockchain,” Security and Communication Networks, vol. 2021, 2021.

[23] S. Niu, L. Chen, J. Wang, and F. Yu, “Electronic health record sharing scheme with searchable attribute-based encryption on blockchain,” IEEE Access, vol. 8, pp. 7195– 7204, 2019.

[24] P. Cockwell and l.-a. Fisher, ‘‘the global burden of chronic kidney disease,’’ lancet, vol. 395, no. 10225, pp. 662–664, feb. 2020, doi: 10.1016/s0140-6736(19)32977-0

[25] C. Webster, e. V. Nagler, r. L. Morton, and p. Masson, ‘‘chronic kidney disease,’’ lancet, vol. 389, no. 10075, pp. 1238–1252, mar. 2017, doi: 10.1016/s0140-6736(16)32064-5.

[26] R. A. Jeewantha, M. N. Halgamuge, A. Mohammad, and G. Ekici, ‘‘Classification performance analysis in medical science: Using kidney disease data,’’ in Proc. Int. Conf. Big Data Res., Osaka, Japan, 2017, pp. 1–6, doi: 10.1145/3152723.3152724. Levey, et al. "Proteinuria as a CKD Outcome Measure." IEEE, 2009.