A machine learning methodology for diagnosing chronic kidney disease

Shraddha Parve[1], Dr. Tushar T. Phadtare[2]

[1]*Student, Department of Computer Engineering, Ajeenkya DY Patil School of Engineering, Lohegoan, Savitribai Phule Pune University Pune, Maharashtra*

[2]*Assistant Professor, Department of Computer Engineering, Ajeenkya DY Patil School of Engineering, Lohegoan, Savitribai Phule Pune University Pune, Maharashtra*

Mail Id: shraddhaparve.8@gmail.com[1], phadtaretulshidas@dypic.in[2]

Abstract—**chronic kidney disease (CKD) presents a significant global health burden, often progressing silently until reaching severe stages. Accurate and early diagnosis is critical to effective intervention and treatment. This paper outlines a machine learning-based diagnostic framework that leverages multiple algorithms to improve prediction accuracy, reduce false diagnoses, and assist healthcare professionals. The study evaluates various machine learning models methods, benchmarking their performance using a publicly available CKD dataset. The findings highlight the optimal approach for integrating machine learning into clinical practice, promising better patient outcomes**

Keywords— *kidney disease Detection, Machine Learning, Classification*

I. INTRODUCTION

Chronic kidney disease is a long-term condition marked by the gradual loss of kidney function. The World Health Organization (WHO) has highlighted CKD as a leading cause of morbidity, emphasizing the need for early diagnosis. Despite advancements in medical technology, traditional diagnostic methods can be time-consuming and often depend heavily on expert analysis. Machine learning (ML) offers a powerful alternative, using data-driven techniques to predict outcomes with higher accuracy and efficiency.

The potential of ML in healthcare is well-documented, and its application to CKD diagnosis can revolutionize early detection strategies. This research bridges the gap between clinical knowledge and computational intelligence to propose a robust diagnostic model for CKD

II. LITERATURE SURVEY

Mehdi Hassan Jony, et al. [1] describe Lung cancer is such an ailment that can be portrayed as one of the trading off ones and has become incredibly testing these days. All around, the cells that pass on threat cover with each other and it's hard to recognize in the hidden stage. However, the reality of the situation is, early area brings an answer; it can reduce the destruction unquestionably. All through the assessment, we use a technique (GLCM); The Grey Level Co-occurrence to isolate the lung photos of the affected. The acknowledgment of the sporadic Lung image can be perceived by the Support Vector Machine (SVM).

Deep Prakash Kaucha, et al. [2] describe Image getting ready techniques are by and by regularly used in the clinical field in the early era of contaminations. Using a combination of image planning tools and data mining, this evaluation seeks to increase the accuracy, affectability, and identification of the early region of lung cancer. The Region of Interest (ROI) is isolated, retained, and compressed to use a DWT (Discrete Waveform Transform) technique on the computed tomography (CT) scan image of the lungs. The resulting ROI picture is divided into bunches LL, HL, LH, and HH and rotated into four sub frequencies. Once again, the ROI-based picture is processed using a 2-level DWT after the LL sub repetition has been spun into four sub-gatherings. Additionally, characteristics including entropy, co-association, energy, change, and homogeneity are separated from the 2-level DWT pictures using a GLCM (Gray Level Co-occurring Matrix), with data collection controlled by an SVM (Support Vector Machine). Request determines whether a CT scan is safe or unsafe. For planning and testing, the Lung Image Database Consortium dataset (LIDC) has been utilised. Aims of this investigation. The structure is separated from the show via a twist on receiver operating characteristics (ROC). The system has an accuracy of 95.16%, an affectability of 98.21%, and an expressivity of 78.69% overall.

Özge Günaydin, et al. [3] describe Lung cancer is a dangerous and challenging condition to diagnose. It often results in passing in the two people, therefore quick, accurate evaluation of handles is much more important for therapy. Different methods have been used to detect a detrimental development in its early phases. In this study, AI techniques were taken into account while identifying lung cancer in lung handles. To detect anomalies, we used Artificial Neural Networks (ANNs), Decision Trees, Naive Bayes, Support Vector Machines, Principal Component Analysis, K-Nearest Neighbors, and Support Vector Machines. We looked at every process, including those that followed pre-handling and those that didn't. The test results demonstrate that Decision Tree provides the best result with 93,24% accuracy without picture taking care, while Artificial Neural Networks provide the best result with 82,43% precision after image planning.

Ning Guo, et al. [4] a unique evaluation of intra-tumor heterogeneity that shows promise for identifying and assembling lung cancer. We demonstrated the viability of SVM-based damage assessment utilising distinguishing PET/CT picture characteristics. Tumor putting together in oncological practise may be enhanced by SVM examination and request using a mix of tumour heterogeneity and other useful aspects.

Nidhi S., et al. [5] describe Possibly the most prevalent and actual contaminant that consistently accounts for a huge number of passings is dangerous development. Lung cancer is the disease with the highest mortality rate among all other types of life-threatening diseases. Tomography channels that have been registered are utilised to identify lung cancer because they provide a rapid, unrefined view of the tumour within the body and follow its course. Although CT is favoured over the other imaging modalities, visually understanding of these CT scan pictures may be a difficult undertaking that delays the identification of lung cancer. To begin identifying the stage of lung cancers, image planning approaches are widely employed in the clinical areas. This study describes a robotized method for detecting lung cancer in CT scan pictures. The partition of the lung district of premium utilising mathematical morphological workouts is one method suggested for the proposed lung cancer area calculation. Another method is centre filtering for picture pre-preparing. Numerical characteristics from the premium district are utilised to categorise CT scan pictures into the normal and abnormal using a help vector machine.

Moffy Vas, et al. [6] describe Malignant development is the primary cause of a huge number of deaths worldwide, with lung cancer accounting for the majority of these deaths. Radiologists use computerised tomography examination to identify illness in the body and monitor its progression. When malignant development is discovered at a later stage thanks to a visual interpretation of the data set, treatment may begin much later, which only serves to increase overall malignancy passing rates. Devices for image preparation may be used in this manner to find diseases early. The lung region of interest is divided using numerical morphological tasks for the proposed lung cancer identification computation, from which Haralick highlights are extracted and used for the order of illness using fake neural organisations.

Snehal Dabade, et al. [7] describe Controlled cell growth in lung tissues describes lung cancer, a dangerous lung tumour. The condition that has been studied the most globally is lung cancer. Lung cancer is the cause of more fatalities than any other illness combined. Early detection and therapy are very beneficial and effective for the survival of persistent cancer. Different image preparation and careful figuring techniques may be used to identify the malignant development cells from clinical imaging. High aim, improved clarity, minimal disturbance, and low twisting are some characteristics of CT images. As a result, CT-images are the most often used kind of image management. The best image-based method for identifying tiny knobs is this one. The likelihood of a patient surviving lung cancer is higher due to early detection. As a result, lung cancer CAD frameworks have been proposed. Essentially, there are three steps in the CAD structure. Pre-handling, lung division, and knob applicant order are those mean places. In this article, we provide a method for dividing extricated lung sections from human chest CT scans. The classifier model of an artificial neural network is that method. The edge recognition of lung area flaps has been improved using a combination of touch planes for each pixel.

R. Sathishkumar, et al. [8] in this PC time, we are thoroughly going with the computerization of everything, similarly, the clinical business is additionally mechanized with the assistance of image handling and information investigation. The most ideal approach to control the passing brought about by malignancy is early location. The clinical image or a CT filter image is pre-handled. The difference of the image is expanded with the CLAHE Equalization strategy. At that point, it is divided with the assistance of the arbitrary walk division technique. In division, the three cycles will happen the ROI of the image is portioned, and afterward then the line remedy is finished. As the third part, the consistent pixel change is portioned. The arrangement is the significant part where the malignant and non-carcinogenic is related to the pre-prepared model. All the techniques utilized above arrangements with the conventional method of image preparing and information investigation. In the Future this exactness will be supported with the cutting edge XGboost calculation where less information is utilized to get high precision.

Kyamelia Roy, et al. [9] describe Lung cancer has been identified and determined using unique information investigation and arranging procedures. Early detection of lung tumours is the primary method for treating lung carcinoma since preventing it becomes impossible because the causes of the disease are unending. Thus, the interplay between biomedical image preparation and AI is employed to classify the existence of lung cancer. The goal of this study is to use a mix of biological image processing techniques and knowledge discovery in data to establish precision and choose a clear incentive for early lung cancer detection. The region of interest (ROI), Random Forest is used to divide the picture of the lungs that is obtained from the CT (Computerized Tomography) filter images.

Barath Narayanan, et al. [10] describe Aspiratory knob placement often reveals the essence of lung cancer. Lung cancer screening would be greatly aided by PC Aided Detection (CAD) of such knobs in CT sweeps. A competitor finder and just a classifier based on components are included in the common CAD framework. In this investigation, we look at and explore the Support Vector Machine (SVM) display in light of a significant number of highlights. We look at the SVM presentation as a part of the number of highlights. Our findings demonstrate that SVM is more robust and computationally faster than traditional classifiers, with a wide variety of features and a lower propensity for over-training. Additionally, we provide a computationally efficient strategy for selecting highlights for SVM. Results for a publicly available Lung Nodule Analysis 2016 dataset are presented. According to our results based on 10-crease approval, the fisher straight discriminant classifier is defeated by the SVM-based arrangement approach by 14.8%

III. RESEARCH METHODOLOGY

In general study the kidney disease diagnosis involves many different technologies. The CKD systems are developed for identification of kidney disease in initial stage. A general Computer Aided Diagnosis set-up consists of several steps in identification of the kidney disease. The following techniques are 1) Data Collection and Extraction 2) Pre-processing 3) Feature Extraction 4) Classification. Figure 1.1 shows the structure of a basic CKD set-up in identification of kidney disease
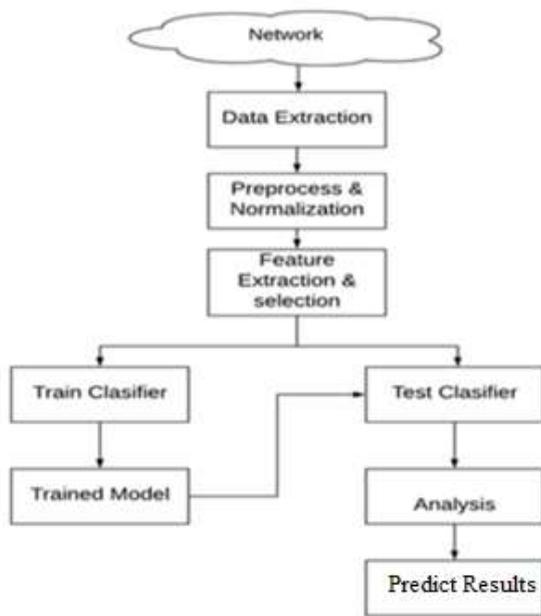
Figure 1.1 Research Methodology

1. Data Collection: The research uses a publicly available CKD dataset comprising patient data, including age, blood pressure, specific lab values and comorbidities.

2. Data Preprocessing: Missing values are imputed using mean or median strategies, and data normalization ensures consistency. Feature scaling is applied to standardize the dataset.

3. Feature Selection: Techniques such as Recursive Feature Elimination (RFE) and correlation matrices identify the most relevant features contributing to CKD diagnosis.

4. Model Training: A range of ML algorithms, including ANN, SVM, J48, RF and NB, are implemented.

5. Validation: The models are evaluated using k-fold cross-validation and performance metrics like precision, recall, F1 score, and accuracy.

Algorithm

Input: Test Dataset which contains various test instances TestDB-Lits [], Train dataset which is built by training phase TrainDB-Lits [] , Threshold Th**.**
Output: HashMap < class label, Similarity Weight > all instances which weight violates the threshold score.
Step 1: For each testing records as given below equation

$$testFeature(k) = \sum_{m=1}^{n} (. featureSet[A[i] ……..A[n] \leftarrow TestDBlits)$$

Step 2: Create feature vector from $testFeature(m)$ using below function.

$$Extracted\_FeatureSetx [t.....n] = \sum_{x=1}^{n}(t) \leftarrow testFeature (k)$$

Extracted_FeatureSetx[t] holds the extracted feature of each instance for testing dataset.
Step 3: For each train instances as using below function

$$trainFeature(l) = \sum_{m=1}^{n} (. featureSet[A[i] ……..A[n] \leftarrow TrainDBList)$$

Step 4: Generate new feature vector from $trainFeature(m)$ using below function

$$Extracted\_FeatureSet\_Y[t……n] = \sum_{x=1}^{n}(t) \leftarrow TrainFeature(l)$$

Extracted_FeatureSet_Y[t] holds the extracted feature of each instance for training dataset.
Step 5: Now evaluate each test records with entire training dataset

$$weight = calcSim (FeatureSetx || \sum_{i=1}^{n} FeatureSety[y])$$

Step 6: Return Weight

IV. RESULTS

In order to determine the effectiveness of the proposed approach, a number of investigations covering a wide range of topics were carried out. A number of different machine learning algorithms, including NB, RF, AdaBoost, ANN, SVM and J48.
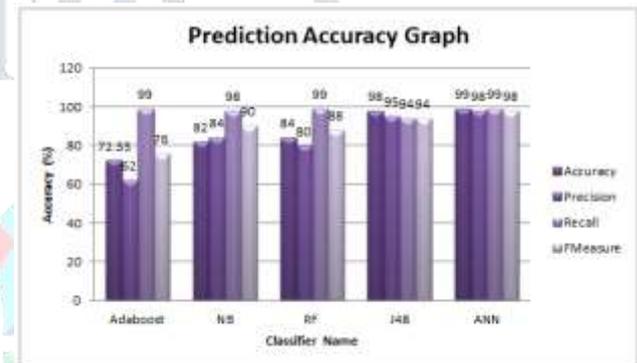

Figure 2: Comparative Analysis of Proposed ANN with Traditional ML

From figure 2, it is evident that the proposed hybrid machine learning algorithm used in this research is more accurate in predicting the disease than the machine learning (ML) Algorithms

V. CONCLUSION

Machine learning has significant potential to enhance CKD diagnosis by enabling early and accurate detection, thus facilitating timely treatment. The study's proposed model not only supports clinical decision-making but also paves the way for further integration of AI in nephrology. Future work should focus on larger, more diverse datasets and the practical deployment of these tools in healthcare settings to validate real-world applicability.

REFERENCES

[1] Mehdi Hassan Jony, Fatema Tuj Johora, Parvin Khatun, Humayan Kabir Rana, "Detection of Lung Cancer from CT Scan Images using GLCM and SVM", IEEE, 1st International Conference on Advances in Science, Engineering and Robotics Technology, 2019.
[2] Deep Prakash Kaucha, P. W. C. Prasad, Abeer Alsadoon, A. Elchouemi, Sasikumaran Sreedharan, "Early Detection of Lung Cancer using SVM Classifier in Biomedical Image Processing", IEEE International Conference on Power, Control, Signals and Instrumentation Engineering, 2017.

[3] Özge Günaydin, Melike Günay, Öznur Şengel, "Comparison of Lung Cancer Detection Algorithms", IEEE, 2019.

[4] Ning Guo, Ruoh-Fang Yen, Georges El Fakhri and Quanzheng Li, "SVM Based Lung Cancer Diagnosis Using Multiple Image Features in PET/CT", IEEE, 2015.

[5] Nidhi S. Nadkarni, Prof. Sangam Borkar, "Detection of Lung Cancer in CT Images using Image Processing", IEEE, Proceedings of the Third International Conference on Trends in Electronics and Informatics, 2019.

[6] Moffy Vas, Amita Dessai, "Lung cancer detection system using lung CT image processing", IEEE, 2017.

[7] Snehal Dabade, Shubhangi Chaudhari, Sneha Jadhav, Arjun Nichal, "A Review Paper on Computer Aided System for Lung Cancer Detection", 2017 International Conference on Big Data, IoT and Data Science (BID) Vishwakarma Institute of Technology, Pune, Dec 20-22, 2017.

[8] R.Sathishkumar, K.Kalaiarasan, A.Prabhakaran, M.Aravind, "Detection of Lung Cancer using SVM Classifier and KNN Algorithm", Proceeding of International Conference on Systems Computation Automation and Networking, 2019.

[9] Kyamelia Roy, Sheli Sinha Chaudhury, Madhurima Burman, Ahana Ganguly, Chandrima Dutta, Sayani Banik, Rayna Banik, "A Comparative Study of Lung Cancer Detection using Supervised Neural Network", IEEE, 2019.

[10] Barath Narayanan, Russell C. Hardie, Temesguen M. Kebede, "Performance Analysis of Feature Selection Techniques for Support Vector Machine and its Application for Lung Nodule Detection", IEEE, 2018