International Journal of Computing and Artificial Intelligence

E-ISSN: 2707-658X
P-ISSN: 2707-6571
IJCAI 2023; 4(1): 71-79
www.computersciencejournals.com/ijcai
Received: 16-03-2023
Accepted: 21-04-2023

Jayakeshav Reddy Bhumireddy
University of Houston, Texas, United States

Rajiv Chalasani
Sacred Heart University, Connecticut, United States

Mukund Sai Vikram Tyagadurgam
University of Illinois at Springfield, Illinois, United States

Venkataswamy Naidu Gangineni
University of Madras, Chennai, Tamil Nadu, India

Sriram Pabbineedi
University of Central Missouri, Missouri, United States

Mitra Penmetsa
University of Illinois at Springfield, Illinois, United States

Corresponding Author:
Jayakeshav Reddy Bhumireddy
University of Houston, Texas, United States

Predictive models for early detection of chronic diseases in elderly populations: A machine learning perspective

Jayakeshav Reddy Bhumireddy, Rajiv Chalasani, Mukund Sai Vikram Tyagadurgam, Venkataswamy Naidu Gangineni, Sriram Pabbineedi and Mitra Penmetsa

DOI: https://www.doi.org/10.33545/27076571.2023.v4.i1a.169

Abstract
Many patients do not realize they have chronic kidney disease (CKD) until it is advanced. Early identification is crucial. CDK is a global health issue, especially for elderly people, requiring early and correct diagnosis to optimize treatment outcomes. This study proposes a VGG16 DL model for automated early detection of chronic kidney disease (CKD) using the CKD dataset. The model uses advanced data preparation techniques like feature selection, normalization, and missing value management to improve input quality. With a 98% F1-score, 98% accuracy, 97% precision, and 99% recall, the proposed VGG16 model exhibits remarkable performance. It is compared against baseline ML models, such as Random Tree, XG Boost, as well as Decision Tree, all of which perform worse. These findings demonstrate how well the VGG16 model captures intricate clinical patterns, offers a strong and trustworthy tool for CKD prediction, and facilitates better patient care by enabling prompt data-driven decision-making.

Keywords: Chronic Kidney Disease (CKD) dataset, chronic diseases, elderly populations, machine learning, disease classification.

1. Introduction
Patient-centred medical homes and accountable care organisations enhance clinician coordination and integration across venues [1]. However, they do not yet address the potential damage or treatment burden of caring for a patient with many disorders. Regarding value-based buying, the bulk of the 33-quality metrics that were initially selected to evaluate accountable care organizations (and decide a portion of the payment) had to do with treating specific ailments. Innovations in health care delivery must provide integration and coordination across conditions, as well as between doctors and chronic illness settings, in order to promote effective treatment for patients with multiple chronic ailments. The term "chronic kidney disease" encompasses many disorders that affect kidney structure and function [2]. Its death rate is really high. Diagnosis of chronic disease of the kidneys (CDK) requires the presence of albuminuria or reduced kidney function for a duration of three months or more. The most severe consequence of CDK is kidney failure, which is often brought on by complications from decreased kidney function [3]. Dialysis and transplantation are the sole options for treating severe symptoms; end-stage renal disease is the term used to describe kidney failure treated in this manner.
The risks of cardiovascular disease, kidney damage, infections, cognitive decline, and poor physical performance are all increased in patients with a reduced glomerular filtration rate (GFR). In order to provide more reliable and early-as-possible detection, ML is used for the early detection. Tasks involving the categorization and prediction of diseases have made extensive use of ML methods [4]. These days, neural networks and ML algorithms in general are being effectively used as aids for illness diagnosis as well as early detection. Large amounts of data, such as diagnostic pictures, lab results, and medical records, are used to train these algorithms. To forecast Colombian chronic renal disease prognosis, this research suggested using a neural network [5]. The human body's condition is already evaluated, disease-related factors are analyzed, and several ailments are diagnosed using ML.

For the purpose of developing models for the detection of cancer, diabetes, cardiovascular disease, retinopathy, and acute renal injury, ML algorithms were employed. encounter massive volumes of data, including patient histories, laboratory results, and diagnostic images [6].

A. Motivation and Contribution of the study: The increasing prevalence and severity of chronic kidney disease (CKD), particularly among the elderly, prompted this study since early diagnosis is key to effective treatment and better patient outcomes. It is necessary to create automated, precise, and scalable alternatives to traditional diagnostic methods because they are frequently expensive, time-consuming, and vulnerable to human mistake. As DL as well as AI continue to progress, particularly convolutional neural networks like VGG16, there is a compelling opportunity to enhance diagnostic precision using real-world clinical data. The purpose of this project is to use these technologies to help medical personnel make timely and accurate forecasts, which will eventually improve patient care and illness management. The contribution of the study are as follows:

- The study presents a new use of the VGG16 DL architecture for CKD early detection, allowing for effective and automated classification from CKD datasets.
- Using deep convolutional layers for automated and hierarchical feature extraction, this end-to-end approach does away with human feature engineering.
- Employs a robust pre-processing pipeline, including missing value handling, normalization using Min-Max scaling as well as optimized feature selection to ensure high-quality model input.
- The suggested model's accuracy, precision, F1-score, and recall standard performance metrics showcase its dependability and efficacy in clinical diagnostic prediction in contrast to more conventional ML models.

B. Justification and novelty of paper: This work is unusual because it uses real-world clinical data to integrate a DL-based method more precisely, the VGG16 architecture for the early identification of CKD. Unlike traditional ML techniques that often depend on weak learning ability and manual feature engineering, the proposed model leverages automatic feature extraction and deep hierarchical learning to identify intricate patterns in medical data. The use of a standardized and widely accepted healthcare dataset further strengthens the practical relevance of the study. This approach introduces a robust, scalable, and data-driven framework that can be generalized to similar medical classification tasks, offering a meaningful advancement in predictive analytics for early-stage chronic disease detection.

C. Structure of Paper: This paper is organized as follows: Section II reviews CKD prediction literature. Section III outlines the approach, including how to use the model as well as preprocess the data. Section IV analyses findings and model comparisons. Section V provides conclusions and suggests ways to improve the accuracy of CKD predictions in the future.

2. Literature Review

This study focuses on predicting health outcomes in elderly populations with chronic diseases using ML. It investigates the factors that influence human resource quality of life the probability of coronary heart disease (CHD), Parkinson's disease (PD), the classification of kidney illnesses, and death.

Krishnani *et al*. (2019) Suggest a comprehensive pre-processing method for CHD prediction. The process begins with empty columns being replaced, followed by resampling, standardisation, normalisation, classification, and lastly, prediction. The primary goal is to use ML techniques like K-Nearest Neighbours, Decision Trees, and Random Forest to forecast the likelihood of coronary heart disease. Also, the algorithms are compared according to how accurate their predictions are. In addition, K-fold Cross Validation is used to randomise the data. The "Framingham Heart Study" dataset, which has 4,240 records, is used to test these approaches. In the experimental research, Random Forest achieved 96.8% accuracy, K-Nearest Neighbours 92.7%, and Decision Tree 92.89% [7].

Amirgaliyev, Shamiluulu and Serek (2018) this research looks at how well the support vector machine method and clinical factors work for chronic renal illness classification. Laboratory, clinical, and physical exams underpin the CDK dataset. Based on accuracy, sensitivity, and specificity, the trial classified over 93% of renal illness patients. CDK are hard to identify, making the process tedious, expensive, and hazardous [8].

Prashanth and Dutta Roy (2018) SVM, logistic regression, random forests, and boosted trees are prominent biomedical ML approaches. They are utilised in the Patient Questionnaire (PQ) part of the widely used Movement Disorder Society-Unified Parkinson's Disease Rating Scale. Prediction models that distinguish early PD from healthy normal are desired. Subject-wise and record-wise validation assessed ML algorithms. They find that both methods can identify early Parkinson's disease and healthy normal individuals (>95%) using accuracy and area under the ROC curve measures [9].

Gjoreski *et al*. (2017) provide a machine-learning technique for identifying heart sounds that indicate chronic heart failure. Filtering, segmentation, feature extraction, and ML make up the approach. A leave-one-subject-out assessment strategy was used to evaluate the procedure using data collected from 122 participants in the research. With an accuracy rate of 96%, the method was fifteen percentage points better than a majority classifier. To be more specific, it can identify 87% of patients with chronic heart failure (87% recall rate). The research demonstrated that the identification of chronic heart failure may be accomplished by using sophisticated ML to real-world sounds captured using an inconspicuous digital stethoscope [10].

Propose new approaches to using chest computed tomography (CT) for the prediction of 5-year mortality in the elderly. This prediction is made by a classifier using data taken from the CT image; the approaches also include segmentation maps of various anatomical components. One method uses deep learning to optimise the classifier and features at the same time; another uses a multi-stage framework to learn the classifier and features in stages, beginning with the design, selection, and extraction of features derived from radiomics data. While radiomics achieves an average accuracy ranging from 56% to 66% (dependent on the classifier and feature selection/extraction technique), the DL algorithm achieves an average accuracy

of 68.5% when using a dataset of 48 annotated chest CT scans for 5-year mortality prediction [11].

The Early Detection of Chronic Diseases literature review overview, including dataset, methodology, conclusions, limitations, and future work, is shown in Table I.

Table 1: Summary of comparative analysis based on Cronic Diseases using Machine Learning

Authors	Dataset	Methods	Findings	Limitations	Future Work
Krishnani *et al.* (2019) [7]	Framingham Heart Study dataset (4240 records)	Data preprocessing (null value replacement, resampling, standardization, normalization), Random Forest, Decision Trees, K-NN, K-Fold Cross Validation	Achieved accuracy: RF-96.8%, DT-92.7%, K-NN-92.89% in predicting CHD risk	Focuses only on structured data; may lack generalizability to other datasets	Extend to DL models; evaluate on more diverse datasets
Amirgaliyev, Shamiluulu & Serek (2018) [8]	CDK dataset (clinical history, lab tests)	Support Vector Machines (SVM)	Over 93% accuracy, sensitivity, and specificity in classifying kidney disease based on clinical features	Dataset details (size, diversity) not fully disclosed; reliance on invasive clinical data	Explore non-invasive features; test other ML models; expand dataset diversity
Prashanth & Dutta Roy (2018)	MDS-UPDRS PQ section (early Parkinson's disease)	SVM, Boosted Trees, Random Forests, and Logistic Regression; validation by subject and record	Classification accuracy and AUC > 95% in differentiating between healthy people and those with early PD	May be limited to questionnaire-based assessment; external validation not detailed	Extend to multimodal data (e.g., imaging, wearable sensors); longitudinal studies
Gjoreski *et al.* (2017) [10]	Heart sounds from 122 subjects	Filtering, Segmentation, Feature Extraction, ML; Leave-one-subject-out evaluation	96% accuracy in detecting chronic heart failure; Recall and precision: 87% each	Small dataset; real-world variability in heart sounds may affect generalization	Larger datasets; integration with wearable tech; real-time monitoring
Carneiro *et al.* (2016)	48 annotated chest CT scans	DL (unified framework), Radiomics (hand-crafted features + classifiers)	DL accuracy: 68.5%; Radiomics accuracy: 56%–66% depending on feature selection method	Very small dataset; limited feature interpretability in DL model	Expand dataset size; improve model interpretability; test integration with clinical decision systems

Methodology

The methodology for early detection of CKD follows a systematic and structured method illustrated in figure 1It starts with gathering actual clinical data, which is thereafter thoroughly preprocessed. The data scientist must perform data cleaning to fix issues with inconsistency or missing values and also reduce the data by removing redundant or unhelpful parts. The data goes through Min-Max normalization after preprocessing so that all the feature scales are similar and after that, the features are carefully selected to make the model more efficient and correct. When data is all set, it is divided to form testing and training sets used for model development. Training the models is done with Decision Tree, XGBoost, Random Tree and VGG16, using the prepared street detection dataset. Standard metrics such as precision, accuracy, F1-score and recall are used to check how well each model performs. This way of proceeding guarantees accurate predictions for spotting CKD early which allows for speedy diagnosis and better health care results. Figure 1 shows the process clearly by representing each step and the steps for each technique.

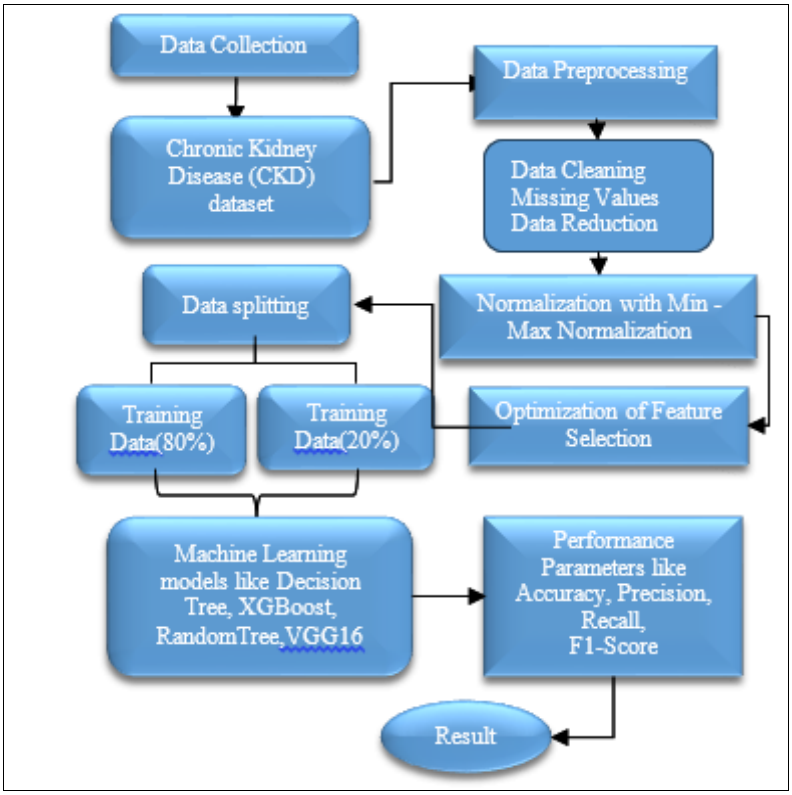

Fig 1: Proposed flowchart for early detection of chronic diseases

The following is a thorough explanation and walkthrough of every step suggested in the flowchart for early identification of chronic diseases in older adults.

A. Data Collection

In this study, the CKD dataset was employed to find out if patients had the disease or not. The public can acquire this two-month dataset through the UCI ML Repository; it was collected at Apollo Hospital in Tamil Nadu, India. The database contains details on four hundred individuals who have renal illness, each person summarized by 25 major medical features. The dataset contains numbers, decimals and categorical values that stand for various clinical readings and patient information required for correct disease prediction [12].

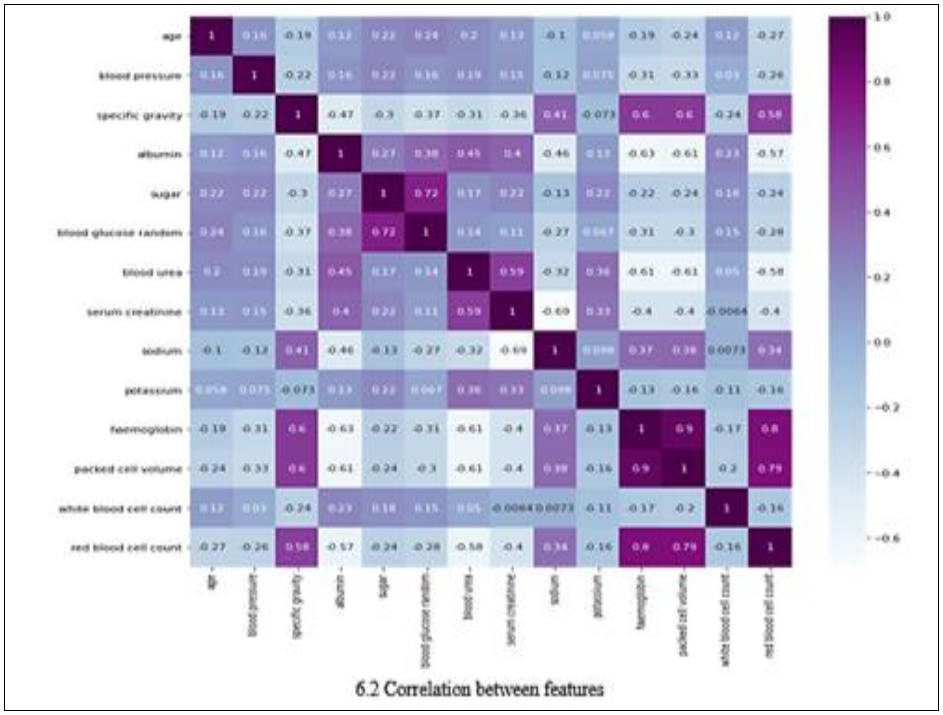

Fig 2: Correlation Matrix of Feature

Figure 2 contains the correlation table for different clinical factors related to chronic kidney disease. A heatmap indicates whether there is a strong positive or negative relationship between age, blood pressure, albumin, hemoglobin and serum creatinine. Stronger correlations are shown with darker shades. The analysis enables one to understand how key features relate which aids in picking the best features and boosts the CKD prediction model performance.

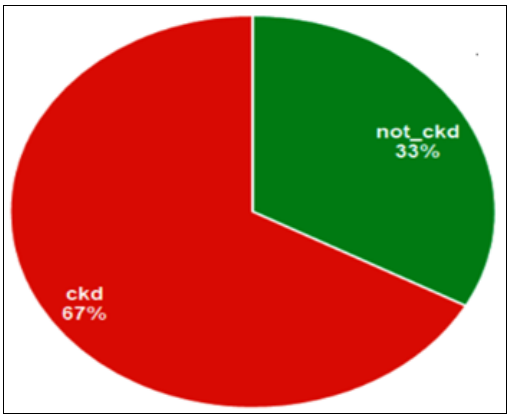

Fig 3: Class distribution pie chart of CDK

Figure 3 shows the CKD dataset class distribution. About two-thirds of all the records are from patients diagnosed with CKD and the remaining third are patients who don't have CKD. The data in the chart suggests there is a clear class distinction and this must be considered when relying on ML for correct analysis and predictions.

B. Dataset Preprocessing

ML models rely on well-prepared data and so Data pre-processing is needed before using the data for chronic disease analysis. The process consists of dealing with missing fields, reducing the data, standardizing it, choosing the needed features and making sure the data is consistent. As a result of these processes, the data becomes better, less noisy and the models work more accurately. The pre-processing steps are mentioned in the next paragraph:

- **Data Cleaning:** A large part of the data is hard to predict, may be incomplete and can be quite noisy. So, in the process of data cleaning, the dataset has its missing values removed, noise is polished and any distorted data is corrected. It focuses on checking for and handling any spots where the data is missing, is not complete or is inconsistent. For example, missing data may be eliminated completely or imputed using methods like mean or median imputation [6].
- **The missing values:** Missing values are clearly seen in medical data. Discarding such observations is the appropriate course of action when the percentage of missing values is low. However, in many cases, the missing ratio [13], Applying imputation techniques is preferable than removing them.
- **Data Reduction:** In data reduction, redundant data are removed in addition to the data's dimensions being

decreased. Data reduction in CDK research involves minimizing the dataset's dimensionality by removing irrelevant, redundant, or noisy features. This enhances ML model efficiency, reduces computational costs, and improves prediction accuracy while preserving essential clinical information for reliable diagnosis [14].

C. Data normalization with min-max normalization

This section describes the data normalization method employed for rescaling feature values. Raw data often has a range of values with varying scales. In this situation, objective functions in some ML algorithms won't operate well without data normalization. Numerous models and classifiers employ Euclidean distance to measure distance between locations. The simplest method is min-max scaling, often known as min-max normalisation. This includes scaling feature values to [0, 1] or [−1, 1]. The target range is chosen based on the kind of data. Equation (1) provides the generic formula for a minmax of [0, 1]

$$x' = \frac{x - \min(x)}{\max(x) - min(x)'} \quad (1)$$

Where the original value is denoted by x and the normalized value by *x'*. In Equation (2), the formula for rescaling a range between an arbitrary set of values [a, b] is as follows:

$$x' = a + \frac{(x - \min(x))(b - a)}{\max(x) - min(x)'} \quad (2)$$

The minimum and maximum values are denoted by a and b.

D. Optimization of Feature Selection

Feature selection is optimization technique that is used remove the irrelevant feature subset from original feature space and improves the classification accuracy using relevant or important feature subset. This research work used four ranking based FSTs to rank the features. Ranking-based feature selection determines the significance of features according to their ranking [15].

E. Data Splitting

The training and testing set of the dataset were separated in an 80-20% ratio. This random allocation ensured that 20% of the data was reserved for performance testing and 80% of the data was used to train the model.

F. Classification of VGG16 Models

Classifying images is important for foreseeing classes in images and checking how well DL models perform in image recognition. The explanation of VGG16-based models as well as their improved versions is provided below:

The VGG16 design, developed by the Visual Graphics Group at Oxford, gained fame for its ability to combine efficiency with simplicity. Its structure involves 16 layers and 13 of them are convolutional while the rest are fully connected. Convolutional layers have 3x3 filters and there are carpooling layers between each convolutional layer to reduce the input's spatial dimensions and notice patterns at different levels. Visual Geometry Group at the University of Oxford developed VGG16 as a deep network for use in computer vision. It is formed with 16 layers: 13 layers for convolution and 3 layers for fully connecting. The model relies on filters that are 3×3 and apply one pixel shift (1 stride) and max-pooling layers that are 2×2. In the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC), VGG16 achieved remarkable results thanks to its consistent design and user-friendliness. People commonly rely on it for image classification, getting relevant features and transferring skills from other datasets. Formula (3) is derived from using shortcut connections from VGG16's residual learning unit [16].

$$y = F(x, \{Wi\}) + x \quad (3)$$

$F(x, \{Wi\})$ means a function that does convolution and activation and (x) is the input feature map, all shaped by the weights $\{Wi\}$. Adding the input to the shape of the output is what $+x$ does and y becomes the final output for each input.

G. Performance Metrics

The confusion matrix, F1-score, recall, accuracy, and precision are often used metrics to assess machine learning models for identifying chronic illnesses in the elderly. A summary of how much each model prediction matches or differs from the observed data is provided by the confusion matrix. Using these criteria is very important for comparing models, especially to recognize situations where the model thinks there is an infection when there really isn't one. Correct understanding of these key characteristics guarantees that prediction models for early detection in ageing people are useful and reliable.

- **False Positives (FP):** The positive classes that were incorrectly predicted are FP.
- **True Positives (TP):** Number of positive instances predicted correctly (TP).
- **False Negatives (FN):** FN denotes the negative classes that were incorrectly anticipated.
- **True Negatives (TN):** The negative classes that were accurately predicted are TN.

Accuracy: In the field of classification models, accuracy is a crucial parameter that provides a gauge of overall performance. This is mathematically represented by the following Equation (4).

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (4)$$

Precision: Precision, or positive predictive value, measures how well our model predicts positive outcomes. The Equation below illustrates this mathematically (5):

$$Precision = \frac{TP}{(TP + FN)} \quad (5)$$

Recall: Recall, also known as sensitivity or true positive rate, compares the model's positive detection to the dataset's real positives. This relationship is defined using the Equation shown below (6):

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

F1 Score: The F1-score fairly evaluates both recall and accuracy since it is the harmonic mean of the two. The formula below is used to calculate the F1-score mathematically Equation (7):

$$F1 - Score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (7)$$

These four criteria are utilised to determine the precision, accuracy, recall, and F1 score. Each of these criteria has a number that ranges from 0 to 1.

Results and Discussion

The results of using ML to identify chronic illnesses in older persons at an early stage are described in this section. An Intel dual-core i6 CPU running at 3.3 GHz, with 1 TB of RAM and Windows 11 Pro installed, was used for the tests. Each model was evaluated by taking into account its F1-score, recall, accuracy, and precision. It summarizes the results of the VGG16 model predicting chronic disease results using clinical information. The proposed model based on VGG16 excels at medical image classification which is clear from Table II. It was able to correctly predict 98% of the information which is a very high rate. The high precision is due to the model detecting few false positives and since recall is 99%, it typically detects almost all actual positive cases. A high F1-Score of 98% demonstrates that the VGG16-based method has excellent balance between getting accurate predictions and remembering all images, making the model a dependable option for this task.

Table 2: proposed vgg16 model Performance on Chronic Kidney Disease (CKD) Dataset

Metric	VGG16
Accuracy	98
Precision	97
Recall	99
F1 Score	98

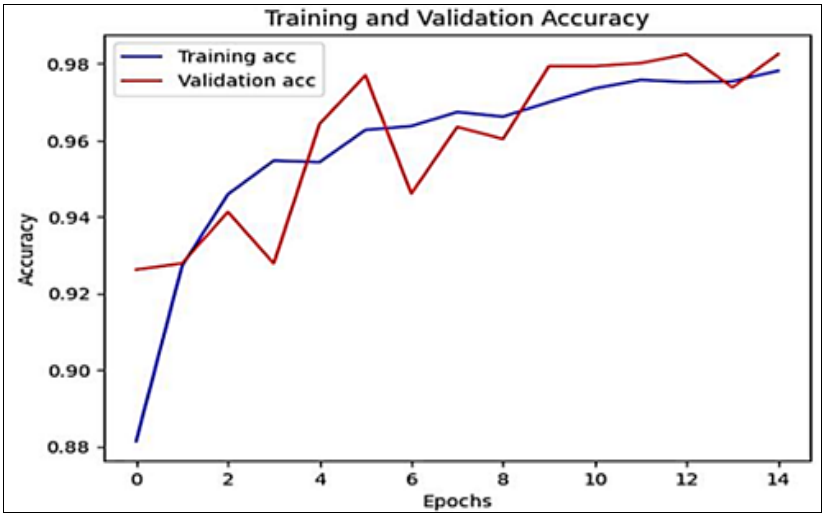

Fig 4: Accuracy of VGG16 model training and validation

Figure 4 graphs present the VGG16 model's performance on training and validation sets after 14 training rounds. At the beginning, the accuracy jumps from 88% to over 94% in only a few episodes. The accuracy in verifying is consistent with some short-term differences. When training finishes, both accuracy scores are nearly identical, both being around 98% by the last episode. Aligning the accuracy of validation with that of training means the model learns and generalizes well and there is not much overfitting

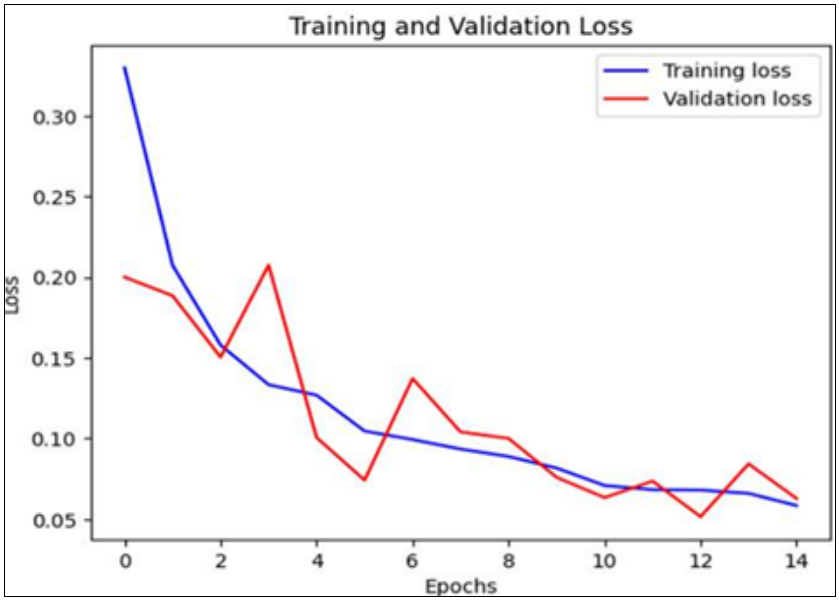

Fig 5: VGG16 Model Training and Validation Loss graph.

Figure 5 shows the VGG16 model's fourteen-epoch training and validation loss curves. Both losses initially decrease quickly, suggesting successful learning in the early phases. The training loss consistently decreases with minor fluctuations, while the validation loss shows slight variations but follows a similar downward trend. As training progresses, both losses gradually converge around 0.05, reflecting good model performance with reduced overfitting. The model seems to generalize effectively to unknown data based on the near alignment of training as well as validation losses.

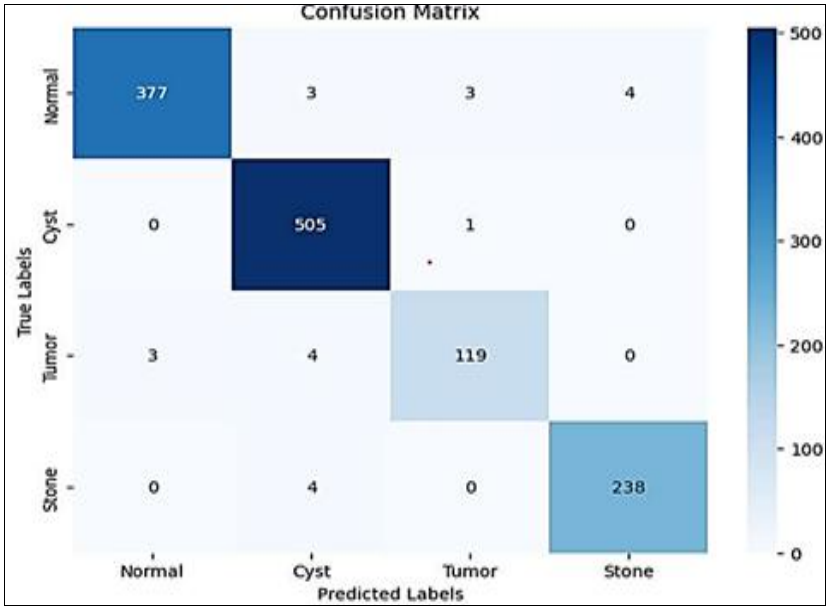

Fig 6: VGG16 model confusion matrices

Figure 6 shows the confusion matrix of the VGG16 model, which demonstrates how well it classified four types of data: normal, cyst, tumour, and stone. The matrix shows the model's correct as well as incorrect predictions. High values along the diagonal represent accurate classifications, with the highest count in the 'Cyst' category (505), followed by 'Normal' (377), 'Stone' (238), and 'Tumor' (119). Very few misclassifications are observed, indicating the model's strong and reliable performance. The overall distribution suggests that the VGG16 model effectively distinguishes between different classes with minimal errors, reflecting its high accuracy and generalization capability. Show in Table III.

Table 3: Comparative Analysis of proposed and existing models using Chronic Kidney Disease (CKD) Dataset

Models	Accuracy	Precision	Recall	F1 Score
Decision Tree [17]	91.75%	85.02	94.66	89.58
XG Boost [18]	83%	77.0	83.0	80.0
Random Tree [19]	95.5%	95.6	95.5	95.5
VGG16	98%	97	99	98

A comparison of the proposed VGG16 model with the baseline models, Random Tree, XG Boost, and Decision Tree, revealed that the DL approach performed better. Among all evaluation metrics, the VGG16 model stands head and shoulders above the competition with its 98% accuracy, 97% precision, 99% recall, and 98% F1 Score. In contrast, the Decision Tree model, while showing strong recall 94.66% and respectable accuracy 91.75%, lags in precision 85.02% and F1 Score 89.58%. With an accuracy of 83%, precision of 77% and F1 Score of 80%, XG Boost seems to be the least effective at the task. Random Tree does much better than Decision Tree and XG Boost by consistently getting scores around 95.5%, although it cannot match the results of VGG16. The investigations confirm that the proposed VGG16 does better than others, especially with high demands for precision and recall, so it is an ideal choice for medical classifications.

VGG16 is proposed as a useful model because it excels in medical image classification. With its deep convolutional architecture, it can automatically find important features, so feature engineering is unnecessary. Therefore, the accuracy and generalization on challenging amounts of data improve. The model's precision and recall reduce false positives and negatives, which is crucial for healthcare. Another aspect of VGG16, transfer learning model, is that it easily transfers past learning and performs well even with minimal data for new tasks which makes it both effective and efficient.

5. Conclusion & Future Work

Today most people try to focus on their health, but it is usually when they feel sick that they think about it. The study described an approach using the VGG16 model for early identification of CKD from the CKD dataset. The accuracy rate for the proposed model was 98% and it did better than Decision Tree, XGBoost and Random Tree. The visualizations show that the VGG16 model can detect intricate clinical details, so it is dependable and useful in spotting CKD risks. Following this approach, healthcare staff may help patients get diagnosed and treated faster. A problem with the data is that it covers a small group and a small duration which raises concerns about how widely its results apply. The model has not been assessed on a wide variety of real or active clinical settings yet.

Future studies may incorporate additional DL architectures such as Reset or Efficient Net, to boost how precise and general the predictions are. Besides, using bigger and more varied data such as long-term patient records, could boost the model's stability. If medical staff could trust AI, it

would be more acceptable to them and build more confidence among them. A clinical decision support system created using this approach would be easier to put into use at hospitals or clinics.

6. References

1. Tinetti ME, Fried TR, Boyd CM. Designing health care for the most common chronic condition-multimorbidity. JAMA. 2012 Jun 12;307(23):2493-2494. DOI: 10.1001/jama.2012.5265.
2. Sharma S, Sharma V, Sharma A. Performance based evaluation of various machine learning classification techniques for chronic kidney disease diagnosis. Int J Mod Comput Sci. 2016;4(3):11-16. DOI: 10.48550/arXiv.1606.09581.
3. Kolluri V. Machine learning in managing healthcare supply chains: how machine learning optimizes supply chains, ensuring the timely availability of medical supplies. J Emerg Technol Innov Res. 2016;3(6).
4. Morales VGR, Monterrubio MSM, Ger MP, Garcia RJA. Explainable prediction of chronic renal disease in the Colombian population using neural networks and case-based reasoning. IEEE Access. 2019;7:152900-152910. DOI: 10.1109/ACCESS.2019.2948430.
5. Elhoseny M, Shankar K, Uthayakumar J. Intelligent diagnostic prediction and classification system for chronic kidney disease. Sci Rep. 2019;9(1):9583. DOI: 10.1038/s41598-019-46074-2.
6. Carneiro G, Rayner OL, Bradley AP, Nascimento J, Palmer L. Automated 5-year mortality prediction using deep learning and radiomics features from chest computed tomography. In: Proceedings, International Symposium on Biomedical Imaging, 2017. DOI: https://doi.org/10.48550/arXiv.1607.00267.
7. Krishnani D, Kumari A, Dewangan A, Singh A, Naik NS. Prediction of coronary heart disease using supervised machine learning algorithms. In: TENCON 2019-2019 IEEE Region 10 Conference (TENCON), 2019, p. 367-372. DOI: 10.1109/TENCON.2019.8929434.
8. Amirgaliyev Y, Shamiluulu S, Serek A. Analysis of chronic kidney disease dataset by applying machine learning methods. In: 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT), 2018, p. 1-4. DOI: 10.1109/ICAICT.2018.8747140.
9. Prashanth R, Dutta Roy S. Early detection of Parkinson's disease through patient questionnaire and predictive modelling. Int J Med Inf. 2018;119:75-87. DOI: 10.1016/j.ijmedinf.2018.09.008.
10. Gjoreski M, Gradišek A, Gams M, Simjanoska M, Peterlin A, Poglajen G. Chronic heart failure detection from heart sounds using a stack of machine-learning classifiers. In: Proceedings-2017 13th International Conference on Intelligent Environments, 2017. DOI: 10.1109/IE.2017.19.
11. Carneiro G, Rayner OL, Bradley AP, Nascimento J, Palmer L. Automated 5-year mortality prediction using deep learning and radiomics features from chest computed tomography. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), 2017, p. 130-134. DOI: 10.1109/ISBI.2017.7950485.
12. Drall S, Drall GS, Singh S, Naib BB. Chronic kidney disease prediction using machine learning: A new approach. Int J Manag Technol Eng. 2018;8(V):278-287.
13. Mahboob T, Ijaz A, Shahzad A, Kalsoom M. Handling missing values in chronic kidney disease datasets using KNN, K-means and K-medoids algorithms. In: ICOSST 2018-2018 International Conference on Open Source Syst. Technol. Proc. 2018:76-81. DOI: 10.1109/ICOSST.2018.8632179.
14. Besra B, Majhi B. An analysis on chronic kidney disease prediction system: cleaning, preprocessing, and effective classification of data. 2019;473-480.
15. Mugisha JO, *et al*. Chronic disease, risk factors and disability in adults aged 50 and above living with and without HIV: findings from the Wellbeing of Older People Study in Uganda. Glob Health Action. 2016;9:31098. DOI: 10.3402/gha.v9.31098.
16. Qassim H, Verma A, Feinzimer D. Compressed residual-VGG16 CNN model for big data places image recognition. In: 2018 IEEE 8th Annual Computing and Communication Work. Conf. CCWC 2018; 2018:169-175. DOI: 10.1109/CCWC.2018.8301729.
17. Tekale S, Shingavi P, Wandhekar S. Prediction of chronic kidney disease using machine learning algorithm. IJARCCE. 2018;7:92-96. DOI: 10.17148/IJARCCE.2018.71021.
18. Xiao J, *et al*. Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. J Transl Med. 2019;17(1):1-13. doi: 10.1186/s12967-019-1860-0.
19. Pasadana IA, *et al*. Chronic kidney disease prediction by using different decision tree techniques. J Phys Conf Ser. 2019;1255(1):012024. DOI: 10.1088/1742-6596/1255/1/012024.
20. Polu AR, Buddula DVK R, Narra B, Gupta A, Vattikonda N, Patchipulusu H. Evolution of AI in software development and cybersecurity: unifying automation, innovation, and protection in the digital age. SSRN. 2021;5266517. Available at: https://ssrn.com/abstract=5266517.
21. Polu AR, Vattikonda N, Buddula DVK R, Narra B, Patchipulusu H, Gupta A. Integrating AI-based sentiment analysis with social media data for enhanced marketing insights. SSRN. 2021;5266555. Available at: https://ssrn.com/abstract=5266555.
22. Narra B, Buddula DVK R, Patchipulusu HH S, Polu AR, Vattikonda N, Gupta A K. Advanced edge computing frameworks for optimizing data processing and latency in IoT networks.
23. Buddula D V K R, Patchipulusu HH S, Vattikonda N, Polu AR, Narra B, Gupta AK. Predictive analytics in e-commerce: effective business analysis through machine learning.
24. Kalla D. AI-powered driver behavior analysis and accident prevention systems for advanced driver assistance. Int J Sci Res Mod Technol. 2022;1.
25. Kuraku D S, Kalla D, Samaah F. Navigating the link between internet user attitudes and cybersecurity awareness in the era of phishing challenges. Int Adv Res J Sci Eng Technol. 2022;9(12).
26. Kalla D, Smith N, Samaah F, Polimetla K. Enhancing early diagnosis: machine learning applications in diabetes prediction. J Artif Intell Cloud Comput. 2022;1(1):2-7. doi: 10.47363/JAICC/2022(1):191.

27. Kalla D, Kuraku D S, Samaah F. Enhancing cybersecurity by predicting malware using supervised machine learning models. Int J Comput Artif Intell. 2021;2(2):55-62.
28. Katari A, Kalla D. Cost optimization in cloud-based financial data lakes: techniques and case studies. ESP J Eng Technol Adv. 2021;1(1):150-157.
29. Kalla D, Smith N, Samaah F, Polimetla K. Facial emotion and sentiment detection using convolutional neural network. Indian J Artif Intell Res. 2021;1(1):1-13.
30. Boppudi V, Bodepudi V, Boppana S, Katnapally N, Babu S, Boppana MS. Exploring the role of neural networks in big data-driven ERP systems for proactive cybersecurity management. Kurdish Stud. 2022.
31. Routhu K, Bodepudi V, Jha KM, Chinta PCR. A deep learning architectures for enhancing cybersecurity protocols in big data integrated ERP systems. SSRN. 2020;5102662. Available at: https://ssrn.com/abstract=5102662.
32. Chinta PCR, Katnapally N. Neural network-based risk assessment for cybersecurity in big data-oriented ERP infrastructures. Neural Netw Appl. 2021.
33. Katnapally N, Chinta PCR, Routhu K, Velaga V, Bodepudi V, Karaka LM. Leveraging big data analytics and machine learning techniques for sentiment analysis of Amazon product reviews in business insights. Am J Comput Eng. 2021;4(2):35-51.
34. Chinta PCR. Enhancing supply chain efficiency and performance through ERP optimisation strategies. J Artif Intell Cloud Comput. 2022;1(4):10-47363.
35. Sadaram G, Sakuru M, Karaka LM, Reddy MS, Bodepudi V, Boppana SB, Maka SR. Internet of Things (IoT) cybersecurity enhancement through artificial intelligence: a study on intrusion detection systems. Universal Libr Eng Technol. 2022.
36. Karaka LM. Optimising product enhancements: strategic approaches to managing complexity. SSRN. 2021;5147875. Available at: https://ssrn.com/abstract=5147875.
37. Jha KM, Bodepudi V, Boppana SB, Katnapally N, Maka SR, Sakuru M. Deep learning-enabled big data analytics for cybersecurity threat detection in ERP ecosystems.
38. Chandrasekaran A, Kalla D. Heart disease prediction using chi-square test and linear regression. Comput Sci Inf Technol. 2023;13:135-146.
39. Kalla D, Kuraku D. Phishing website URLs detection using NLP and machine learning techniques. J Artif Intell. 2023;5:145.
40. Kuraku DS, Kalla D. Impact of phishing on users with different online browsing hours and spending habits. Int J Adv Res Comput Commun Eng. 2023;12(10).
41. Kuraku S, Kalla D, Samaah F, Smith N. Cultivating proactive cybersecurity culture among IT professionals to combat evolving threats. Int J Electr Electron Comput. 2023;8(6).
42. Kuraku DS, Kalla D, Smith N, Samaah F. Exploring how user behavior shapes cybersecurity awareness in the face of phishing attacks. Int J Comput Trends Technol. 2023.
43. Chinta PCR. Leveraging machine learning techniques for predictive analysis in merger and acquisition (M&A). J Artif Intell Big Data. 2023;3(1):10-31586.
44. Kuraku DS, Kalla D, Smith N, Samaah F. Safeguarding FinTech: elevating employee cybersecurity awareness in the financial sector. Int J Appl Inf Syst. 2023;12(42).
45. Moore C. AI-powered big data and ERP systems for autonomous detection of cybersecurity vulnerabilities. Nanotechnol Perspect. 2023;19:46-64.
46. Chinta PCR. The art of business analysis in information management projects: best practices and insights.
47. Madhav JM, B Varun, Niharika K, Rao MS, Murthy KL. Optimising sales forecasts in ERP systems using machine learning and predictive analytics. J Contemp Edu Theo Artific Intell. 2023;104.
48. Maka SR. Understanding the fundamentals of digital transformation in financial services: drivers and strategic insights. SSRN. 2023;5116707. Available at: https://ssrn.com/abstract=5116707.
49. Routhu KK, Katnapally N, Sakuru M. Machine Learning for Cyber Defense: A Comparative Analysis of Supervised and Unsupervised Learning Approaches. Journal for ReAttach Therapy and Developmental Diversities, 2023, p. 6. DOI: 10.53555/jrtdd.v6i10s(2).3481.
50. Chinta PCR, Moore CS. Cloud-based AI and big data analytics for real-time business decision-making. J Artif Intell Big Data. 2023;36:96-123. DOI: 10.47363/JAICC/2023.
51. Madhav JK, Varun B, Niharika K, Rao MS, Murthy KL. Optimising sales forecasts in ERP systems using machine learning and predictive analytics. J Contemp Educ Theo Artif Intell, 2023, p. 104.
52. Bodepudi V. Understanding the fundamentals of digital transformation in financial services: drivers and strategic insights. J Artif Intell Big Data. 2023;3(1):10-31586.