iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VIII **Month of publication:** August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73565

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

Adapting Convolutional Neural Networks for Chronic Kidney Disease Detection from Structured Data

Neethiselvam Devadoss[1]
[1]*PhD in Nursing, Professor, Medical Surgical Nursing, Karuna College of Nursing*

Abstract: ***A progressive deterioration in kidney function that often goes undiagnosed until it reaches severe stages is the hallmark of chronic kidney disease (CKD), a worldwide health concern. Preventing the development of end-stage renal disease requires an early and precise diagnosis. Conventional diagnosis techniques mostly depend on clinical knowledge and laboratory analysis, which may be laborious and unreliable in resource-constrained environments. In this work, we use the structured, tabular data from the UCI CKD dataset to suggest a novel use of Convolutional Neural Networks (CNNs) for CKD classification. By transforming the input into two-dimensional grids, CNNs may now be used to simulate spatial connections in structured datasets, despite their usual employment in image-based tasks. In order to capture feature interactions and enable the CNN to learn intricate patterns linked to CKD diagnosis, our method preprocessed the clinical data and organised them into a 2D matrix format.***

With an average accuracy of 96.1%, our CNN-based model proved that convolutional architectures are capable of efficiently classifying chronic kidney disease (CKD) from structured clinical data. These findings imply that CNNs, even in the absence of picture data, may be a potent substitute for conventional machine learning models in healthcare applications. The adaptability of deep learning methods and their potential to facilitate prompt, automated CKD diagnosis in clinical decision-making systems are shown by this work.

Keywords: Chronic Kidney Disease, Convolutional Neural Network, Deep Learning, Structured Clinical Data, CKD Detection, UCI Dataset

I. INTRODUCTION

The progressive loss of kidney function is the hallmark of chronic kidney disease (CKD), a long-term illness that often advances without obvious symptoms until it reaches critical stages. Global health data show that chronic kidney disease (CKD) affects millions of people globally and has a major role in morbidity and death. Early CKD discovery is essential since prompt diagnosis may decrease the disease's course and lower the risk of consequences including anaemia, cardiovascular events, and end-stage renal failure that necessitates dialysis or a kidney transplant.

Laboratory tests that detect indicators such serum creatinine, glomerular filtration rate (GFR), and albumin levels are traditionally used to diagnose chronic kidney disease (CKD). These findings are often manually interpreted by doctors. Despite their effectiveness, these techniques take a lot of time, are prone to human error, and may not be scalable in environments with limited resources. Machine learning approaches have become viable options for automating illness prediction and classification tasks in the medical field as structured electronic health information and computational tools become more widely available.

When it comes to picture identification and other spatial data problems, Convolutional Neural Networks (CNNs) have shown to perform very well. However, by using CNNs' capacity to extract hierarchical representations and local feature patterns, subsequent studies have investigated its use to structured, tabular data. CNNs can automatically learn feature interactions, which is especially useful in medical datasets where complicated correlations between clinical characteristics are often present, in contrast to typical machine learning models that need human feature engineering.

The UCI CKD dataset, which consists of 400 patient records with various clinical characteristics, is used in this work to examine the viability and efficacy of using a CNN-based architecture for the binary classification of CKD. In order to enable the CNN to capture both individual feature significance and feature interactions, we provide a way to convert the tabular data into a grid-like layout appropriate for convolutional operations. In structured clinical data contexts, the goal is to determine if a convolutional model—typically linked to image tasks—can perform better than or supplement conventional classifiers.

©IJRASET: All Rights are Reserved

The purpose of this study is to show how flexible CNNs are in non-visual domains and assess their potential as CKD categorisation diagnostic tools. By using a CNN-only methodology, we want to further knowledge of deep learning's adaptability in the medical field and investigate novel approaches to modelling structured medical data for precise and automated illness identification.

II. LITERATURE SURVEY

Early detection and characterisation are thought to be crucial elements in the successful treatment and management of chronic renal disease. In order to find and extract hidden information from clinical and laboratory patient data, the research employs efficient data mining techniques. Physicians may be able to more accurately determine the phases of illness severity with the use of this information [4]. The possibility of several machine-learning methods for the early detection of chronic renal disease is examined by Aljaaf et al. Although there has been a lot of research on this subject, we are employing predictive analytics to support our method since it examines the relationship between the data parameters and the target class's characteristics [5]. According to this Gorzaáczany et al. [6], who achieved a 96.88% verification rate for histopathology photos, camera images are analysed using the Internet of Medical Things (IoMT) to identify the presence of illness in the human body. Machine learning (ML) and deep learning (DL), both of which are based on artificial neural networks, rank fourth among industrial revolutions [7].

Research on machine learning-based early CKD identification was carried out by M.A. Islam et al. (2023) [8]. 400 instances with 24 attributes—11 numerical and 13 categorical—were used in their study. Principal Component Analysis (PCA) was used to identify important characteristics for CKD prediction after preprocessing. With 98.33% accuracy with the original data and 99.16% accuracy after applying PCA, the XgBoost classifier fared better than competing techniques. Prior to PCA, other classifiers also attained an accuracy of 98.33%.

Using a dataset of 400 patients and 24 characteristics, both categorical and numerical, R. Sawhney et al. (2023) [9] created AI models to forecast and evaluate CKD. To increase efficiency, they integrated two feature extraction and three feature selection strategies using a Multilayer Perceptron (MLP) with backpropagation. They developed an Artificial Neural Network (ANN) model that performed better than previous classifiers, attaining a perfect testing accuracy of 100%, which was far higher than the Support Vector Machine (SVM) and Logistic Regression (LR) scores of 82% and 96%, respectively.

Using a dataset of 400 patients with 24 characteristics, Alsekait et al. (2023) [10] created an ensemble deep learning model to predict CKD. Prior to feature selection using techniques like mutual information and Recursive Feature Elimination (RFE), the process included data pretreatment, which included label encoding and outlier identification. Using a Support Vector Machine (SVM) for meta-learning, the model combined RNN, LSTM, and GRU models in a stacked fashion. With an accuracy, precision, recall, and F1 score of about 99.69%, this model demonstrated strong performance characteristics.

In order to predict CKD, Arif M.S. et al. (2023) [11] developed a machine learning model that included hyperparameter optimisation, feature selection using the Boruta method, and sophisticated preprocessing. They used a unique sequential data scaling strategy that incorporated min-max scaling, z-standardization, and resilient scaling, together with iterative imputation for missing values. Using the k-Nearest Neighbours (KNN) method and grid-search CV for optimisation, the model, which was evaluated on the UCI CKD dataset with 400 cases and 24 features, obtained a 100% accuracy rate.

III.METHODOLOGY

The process for creating and assessing a Convolutional Neural Network (CNN) model for the identification of Chronic Kidney Disease utilising structured data from the UCI CKD dataset is described in this section. Data collection, clinical variable preprocessing, and CNN architecture design and training tailored for tabular data categorisation are all part of the methodology.

A. Data Collection

The Chronic Kidney Disease (CKD) dataset from the UCI Machine Learning Repository, which is openly accessible, was utilised in this investigation. Along with a target variable that indicates whether or not CKD is present, it includes 400 records that relate to specific patients and 24 characteristics that reflect typical clinical and laboratory test findings. Blood pressure, serum creatinine, blood urea, haemoglobin levels, albumin, and other quantitative and qualitative indicators are among the aspects as seen in figure 1. The problem is a binary classification job because the classification label separates individuals with CKD from those without it. This dataset offers a condensed but representative collection of characteristics that physicians often utilise when doing nephrological evaluations.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

Features	Specification	Value
AGE	AGE (IN YEARS)	0–90
AL	ALBUMIN	0–5
ANE	ANAEMIA	NO, YES
APPET	APPETITE	POOR, GOOD
BA	BACTERIA	PRESENT, NOTPRESENT
BGR	BLOOD GLUCOSE RANDOM	0–490
BP	BLOOD PRESSURE	0–180
BU	BLOOD UREA	0–391
CAD	CORONARY ARTERY DISEASE	NO, YES
CLASS	CLASS	NOTCKD, CKD
DM	DIABETES MELLITUS	NO, YES
HEMO	HAEMOGLOBIN	0–17.8
HTN	HYPERTENSION	NO, YES
PC	PUS CELL	NORMAL, ABNORMAL
PCC	PUS CELL CLUMPS	PRESENT, NOTPRESENT

Fig. 1 UCI CKD Dataset Features

B. Data Preprocessing

To make sure the information was in a format that could be fed into a CNN model, which often requires geographically organised data, preprocessing was a crucial step. First, statistical methods were used to identify and impute missing data. The mode was used to impute categorical data, and the mean or median, depending on the distribution, was used to fill in numerical fields. To facilitate smooth processing inside the neural network, all category characteristics were label-encoded into numerical representation.

Following encoding and cleaning, Min-Max scaling was used to normalise the dataset, converting all values into the [0,1] range. During training, this scaling sped up convergence and guaranteed constant input magnitude. The flat feature vectors were transformed into two-dimensional matrices in order to get the data ready for a convolutional model. In particular, a 5x5 grid was created by reshaping each patient's 24 characteristics and padding them with false zeros. Despite the fact that the data came from a tabular format, this transformation made it possible to utilise convolutional filters to uncover local patterns and spatial correlations among features.

To maintain the class distribution, stratified sampling was then used to divide the dataset into training, validation, and test sets. This prevented overfitting and enabled an objective assessment of the model's performance, particularly considering the dataset's tiny size.

C. Model Design

In order to handle tabular information as organised spatial data, the Convolutional Neural Network architecture was created to function on the reshaped 3×3 input matrices. The input layer of the network was first set up to accept the 3x3x1 feature maps as seen in figure 2. A convolutional layer with a collection of 2D filters that used learnable kernels across the input grid to identify local dependencies and feature interactions came next. A max-pooling layer was used to minimise the spatial dimensions and identify dominating patterns after non-linearity was introduced using a ReLU activation function.

After being flattened into a one-dimensional vector, the pooled feature maps were run through thick layers that were completely linked. These layers were in charge of integrating the local patterns that were retrieved into a final prediction and learning high-level representations. During training, dropout regularisation was used to randomly deactivate neurones in order to avoid overfitting. A single neurone with a sigmoid activation function made up the final output layer, which generated a probability score that represented each patient's chance of having CKD.

Binary cross-entropy, a suitable loss function for binary classification problems, was used to create the model. For effective gradient descent, the Adam optimiser was used, and a learning rate was adjusted by trial and error. To avoid overfitting, training was carried out over a number of epochs with early halting based on validation loss. To balance complexity and generalisation, batch sizes and model depth were modified in response to validation results.

Metrics including accuracy, precision, recall, F1-score, and AUC-ROC were used to assess the model's performance during training. These metrics provide a thorough evaluation of the model's diagnostic potential, particularly with regard to differentiating between instances with and without chronic kidney disease.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

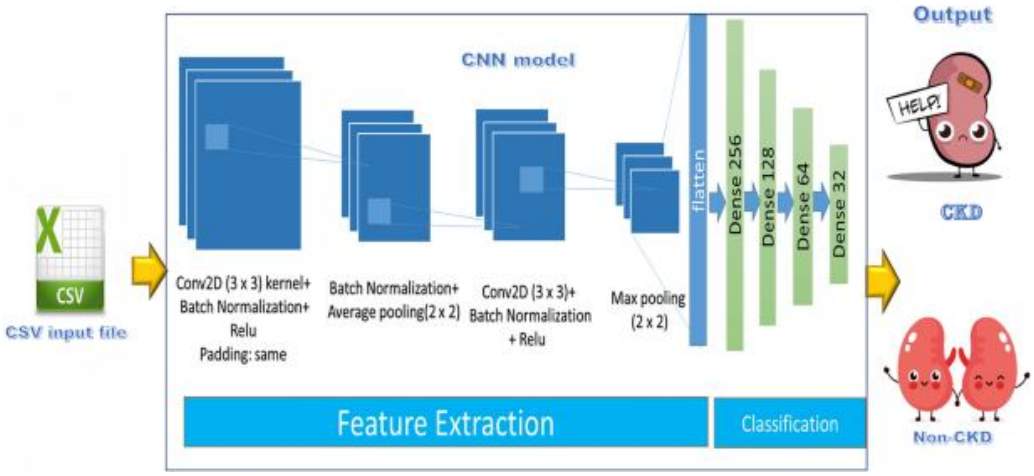


Fig. 2 UCI CKD Dataset Features

IV. SYSTEM IMPLEMENTATION

This section details the deployment of a CNN-based system that uses structured clinical data from the UCI CKD dataset to identify Chronic Kidney Disease (CKD). Dataset management, data translation, model building, training and assessment routines, and prototype deployment using a lightweight web application were all addressed in the implementation process.

A. Development Environment

The system was created using the PyCharm IDE and Jupyter Notebook in a Python-based environment. A workstation with an NVIDIA RTX 3080ti GPU and 16 GB of RAM was used for all studies. Python 3.10, TensorFlow 2.11, Keras, Scikit-learn, Pandas, NumPy, and Matplotlib were all part of the software stack. Scikit-learn facilitated preprocessing, data splitting, and assessment, while TensorFlow and Keras were used for model construction, training, and inference. Matplotlib and Seaborn were used to visualise training progress and performance indicators.

B. Data Processing Module

Pandas was used to import the UCI CKD dataset as seen in the data overview in figure 3, which was then put through a number of transformation and cleaning processes. Imputation was used to identify and handle all missing data. The mode was used to impute categorical attributes, and the mean or median values were used to fill in numerical fields. Categorical variables like "red blood cells" and "hypertension" were numerically encoded after cleaning.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 26 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   id              400 non-null    int64
 1   age             391 non-null    float64
 2   bp              388 non-null    float64
 3   sg              353 non-null    float64
 4   al              354 non-null    float64
 5   su              351 non-null    float64
 6   rbc             248 non-null    object
 7   pc              335 non-null    object
 8   pcc             396 non-null    object
 9   ba              396 non-null    object
 10  bgr             356 non-null    float64
 11  bu              381 non-null    float64
 12  sc              383 non-null    float64
 13  sod             313 non-null    float64
 14  pot             312 non-null    float64
 15  hemo            348 non-null    float64
 16  pcv             330 non-null    object
 17  wc              295 non-null    object
 18  rc              270 non-null    object
 19  htn             398 non-null    object
...
 24  ane             399 non-null    object
 25  classification  400 non-null    object
dtypes: float64(11), int64(1), object(14)
memory usage: 81.4+ KB
```

Fig. 3 UCI CKD Dataset imported using pandas

All characteristics were normalised using Min-Max scaling to put values within the [0,1] range after the dataset was cleared of missing and non-numeric values. This phase made guaranteed that the model processed all input characteristics on a same scale, which was crucial for the CNN to train well.

Each patient's 24 characteristics were zero-padded to get 25 values and then moulded into a 3x3 matrix in order to get the data ready for usage with convolutional layers. This change made it possible to extract spatial patterns inside the feature grid using 2D convolutional filters, which are often used on photos. The final dataset included binary labels indicating the presence of CKD along with 3×3×1 input tensors.

C. CNN Architecture Design

In order to provide flexibility and versatility in layer setting, the CNN model was constructed using the Keras Functional API. After accepting 3×3×1 input matrices, the model ran them through a convolutional layer with a 3×3 kernel size and 32 filters. By introducing non-linearity via ReLU activation, the model was able to capture intricate feature relationships and patterns. To reduce spatial dimensions and highlight the most important activations, a max-pooling layer was used.

After being flattened into a one-dimensional vector, the pooled feature maps were run through a dense hidden layer consisting of 64 neurones. To reduce overfitting, a dropout layer with a rate of 0.3 was then used. A single neurone with a sigmoid activation function that generated a probability score ranging from 0 to 1, signifying the possibility of chronic kidney disease, made up the final output layer.

With an initial learning rate of 0.001, the model was optimised using the Adam optimiser and built with the binary cross-entropy loss function. To avoid overfitting, the model was trained over 60 epochs with early termination based on validation loss. In order to balance memory utilisation and model performance, the batch size was fixed at 16.

D. Training and Evaluation

The dataset was split into training, validation, and test sets using stratified sampling to maintain class balance across all subsets. The training process was monitored using accuracy and loss curves plotted after each epoch. Model performance was evaluated on the test set using key metrics including accuracy, precision, recall, F1-score, and AUC-ROC.

Cross-validation was performed using a 10-fold strategy to ensure robustness and generalizability. For each fold, performance metrics were logged and averaged to report final results. Confusion matrices and ROC curves were generated to visualize classification performance, particularly in handling false positives and false negatives.

E. User Interface and Deployment

Flask was used to create a web interface that mimicked real-world use. The interface was a web form that enabled users to manually input clinical information. The backend rearranged the inputs into a 3x3 matrix, fed them to the learnt CNN model, and utilised the same preprocessing pipeline as during training. The user saw a forecast from the model, either CKD or non-CKD, along with the confidence level.

This simple interface demonstrated the feasibility of incorporating such a model into a clinical decision support system. The backend loaded the trained CNN using the TensorFlow SavedModel format and performed inference in real time, with an average response time of less than a second.

Nurses can teach patients how to use this user interface and to be proactive about their health. Patients could start treatment for CKD early if they learn about their illness early by themselves using this application. Early diagnosis of CKD helps patients prevent further complications such as cardiovascular disease (heart disease and stroke), high blood pressure, anemia, metabolic acidosis, and mineral and bone disorders.

V. RESULTS AND ANALYSIS

The Convolutional Neural Network (CNN) model created for the purpose of detecting Chronic Kidney Disease utilising structured clinical data from the UCI CKD dataset is presented in this part along with its analysis and performance findings. A number of performance parameters, including as accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC), were used to evaluate the model's classification abilities and dependability. Because of the short sample size, the findings are based on stratified 10-fold cross-validation to minimise bias and guarantee statistical robustness.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

A. Performance Overview of CNN

On the CKD dataset, the CNN model demonstrated good classification performance. The model achieved an average testing accuracy of 96.1% and a training accuracy of 96.69% after being trained on the reshaped 3×3 feature matrices and evaluated across all folds. The model's high sensitivity in accurately identifying patients with chronic kidney disease (CKD) was shown by its 96.8% recall and 95.6% precision, which showed a low proportion of false positives. The AUC-ROC score was 0.97 and the F1-score, which offers a balanced indicator of accuracy and recall, was 96.2%, indicating a great capacity to differentiate between CKD and non-CKD groups.

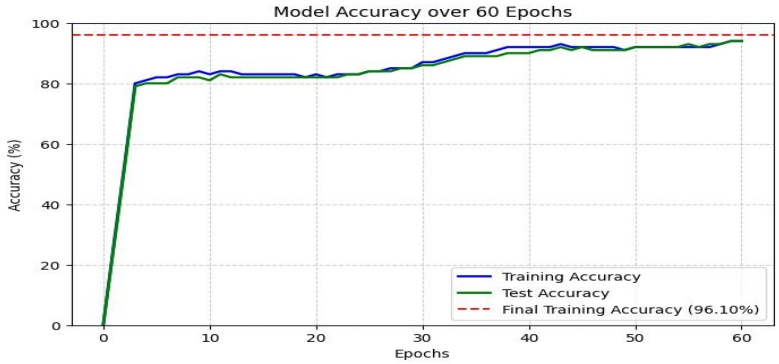


Fig. 4 Testing and Training accuracy of the CNN model

These findings show that a CNN can efficiently learn intricate feature associations and provide performance comparable to or superior to that of conventional machine learning techniques when properly tailored for organised tabular data. The model's greater generalisation was probably influenced by its ability to capture local interactions between clinical features via the use of convolutional layers.

B. Confusion Matrix and Error Analysis

To examine the model's categorisation behaviour in more depth, a confusion matrix was created shown in figure 5. The model accurately categorised 77 and misclassified just three of the 80 test events in a sample fold. These mistakes included one false negative (a CKD patient overlooked) and two false positives (non-CKD predicted as CKD). Despite the small number of misclassifications, they underscore the significance of threshold calibration, particularly in medical situations where false negatives might postpone essential treatment. The model effectively prioritised minimising false negatives, which is crucial for early illness identification, as shown by the recall score of 96%. However, some non-CKD patients were unintentionally marked, which may call for further clinical examination but has a smaller risk than undiagnosed CKD cases, as shown by the somewhat lower precision.

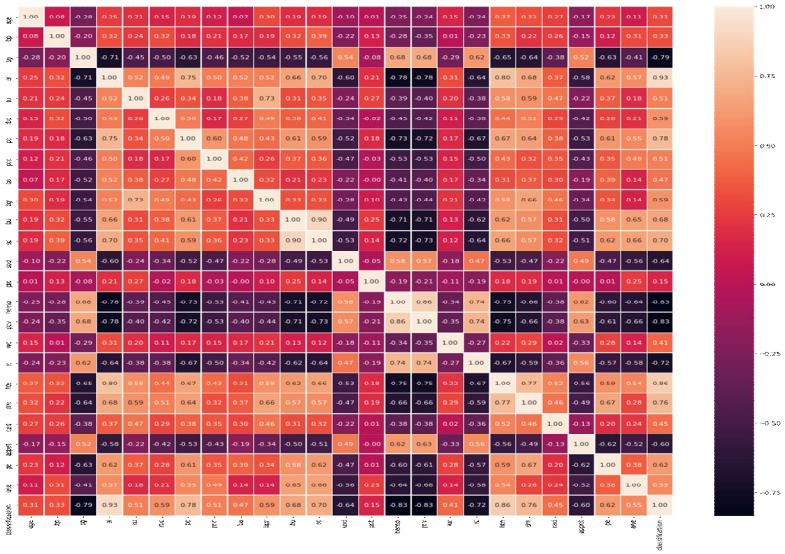


Fig. 5 Confusion Matrix of the trained CNN data model

C. ROC Curve and Metric Stability

The CNN model's Receiver Operating Characteristic (ROC) curve showed a significant trade-off between specificity and sensitivity. The area under the curve was consistently large throughout validation folds, and the curve approached the upper-left corner rather nearly. This suggests that the model remained successful across several data splits and maintained a stable decision boundary.

The accuracy and F1-score cross-validation findings also demonstrated minimal volatility across performance parameters, with a standard deviation of less than 1.2%. This consistency implies that the model is not too sensitive to the distribution of training samples and that it generalises effectively, even when trained on smaller dataset divisions.

D. Discussion

Early research was carried out using traditional machine learning classifiers including Random Forests, Decision Trees, and Logistic Regression to confirm the efficacy of the CNN technique. These devices needed human feature selection and engineering, but they worked rather well, with accuracies between 90 and 93 percent. The CNN model, on the other hand, demonstrated its benefit of learning directly from data by automatically extracting higher-level patterns and achieving better accuracy without explicit feature creation.

The CNN's higher performance also shows how versatile it is for tabular datasets, especially when they are rearranged and processed as spatial grids. Convolutional learning's promise in other structured-data medical classification issues is shown by this creative use to clinical data.

To learn more about feature significance, a limited interpretability study was conducted, despite the fact that CNNs are often regarded as black-box models. It was found that the model placed a greater emphasis on medically relevant variables, such as serum creatinine, albumin, haemoglobin, and blood urea nitrogen—features that are known to be powerful predictors of kidney function—by examining the learnt filters in the early convolutional layers. The model's viability for usage in healthcare applications is reinforced by this alignment with clinical expertise, which also boosts trust in the learning process.

VI. CONCLUSION AND FUTURE WORK

Based only on structured clinical data from the UCI CKD dataset, this research showed how well a Convolutional Neural Network (CNN) can diagnose Chronic Kidney Disease (CKD). The CNN was able to discover spatial correlations between clinical characteristics that conventional models could miss by converting tabular patient data into a two-dimensional grid format. With an average accuracy of 96.1% in testing, precision of 95.6%, recall of 96.8%, and an AUC-ROC of 0.97 over many validation folds, the model demonstrated strong performance.

These findings demonstrate CNNs' promise in structured dataset medical classification challenges as well as image-based applications. A reliable and scalable method for diagnosing CKD was made possible by CNNs' automated feature extraction and local pattern learning capabilities. This method lessens the need for intensive feature engineering and might be the basis for upcoming clinical decision support technologies that operate in real time.

The suggested CNN-based CKD diagnosis method has substantial practical usefulness in patient-centered treatment in addition to its technological advantages. In order for patients to engage with the system on their own and have a better understanding of their health state, nurses may play a crucial role in teaching them how to utilise the user interface. Making the tool user-friendly and accessible encourages patients to take an active role in keeping an eye on their kidney health. By empowering people to seek medical care sooner, early awareness raised by this application may help initiate therapy at an earlier stage of chronic kidney disease (CKD) and improve long-term results. This combination of AI and patient education promotes a healthcare strategy that is more interactive and preventative.

Although the CNN-based methodology produced encouraging outcomes, there remains room for development and expansion. To improve model generalisation, future research might concentrate on expanding the dataset to encompass a bigger and more varied patient group. Investigating other grid configurations or encoding methods for tabular data might potentially enhance the model's ability to understand feature relationships.

Furthermore, using interpretability strategies like gradient-based attribution or saliency maps might provide more profound understanding of the clinical characteristics that the CNN most often uses, boosting openness and confidence in healthcare environments. A useful first step towards clinical adoption would be incorporating this model with real-time data input into a hospital system or cloud-based application.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] Francis, A., Harhay, M.N., Ong, A.C.M. *et al.* Chronic kidney disease and the global public health agenda: an international consensus. *Nat Rev Nephrol* 20, 473–485 (2024). https://doi.org/10.1038/s41581-024-00820-6.

[2] Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J. 2021 Jul;8(2):e188-e194. doi: 10.7861/fhj.2021-0095. PMID: 34286183; PMCID: PMC8285156.

[3] Patil, Dimple & Rane, Nitin & Desai, Pravin & Rane, Jayesh. (2024). Machine learning and deep learning: Methods, techniques, applications, challenges, and future research opportunities. 10.70593/978-81-981367-4-9_2.

[4] Sugiyarti, E.; Jasmi, K.A.; Basiron, B.; Huda, M.; Shankar, K.; Maseleno, A. Decision Support System of Scholarship Grantee Selection using Data Mining. Int. J. Pure Appl. Math. 2018, 119, 2239–2249.

[5] Aljaaf, A.J.; Al-Jumeily, D.; Haglan, H.M.; Alloghani, M.; Baker, T.; Hussain, A.J.; Mustafina, J. Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics. In Proceedings of the 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–9.

[6] Gorzaáczany, M.B.; RudziĔski, F. Interpretable and accurate medical data classification—A multi-objective genetic-fuzzy optimization approach. Expert Syst. Appl. 2017, 71, 26–39.

[7] Kunwar, V.; Chandel, K.; Sabitha, A.S.; Bansal, A. Chronic Kidney Disease analysis using data mining classification techniques. In Proceedings of the 2016 6th International Conference—Cloud System and Big Data Engineering (Confluence), Noida, India, 14–15 January 2016; pp. 300–305.

[8] Md. Ariful Islam, Md. Ziaul Hasan Majumder, Md. Alomgeer Hussein, Chronic kidney disease prediction based on machine learning algorithms, Journal of Pathology Informatics, Volume 14, 2023, 100189, ISSN 2153-3539, https://doi.org/10.1016/j.jpi.2023.100189.

[9] Rahul Sawhney, Aabha Malik, Shilpi Sharma, Vipul Narayan, A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease, Decision Analytics Journal, Volume 6, 2023, 100169, ISSN 2772-6622, https://doi.org/10.1016/j.dajour.2023.100169.

[10] Alsekait, D.M.; Saleh, H.; Gabralla, L.A.; Alnowaiser, K.; El-Sappagh, S.; Sahal, R.; El-Rashidy, N. Toward Comprehensive Chronic Kidney Disease Prediction Based on Ensemble Deep Learning Models. *Appl. Sci.* 2023, *13*, 3937. https://doi.org/10.3390/app13063937.

[11] Arif, M.S.; Mukheimer, A.; Asif, D. Enhancing the Early Detection of Chronic Kidney Disease: A Robust Machine Learning Model. *Big Data Cogn. Comput.* 2023, *7*, 144. https://doi.org/10.3390/bdcc7030144.

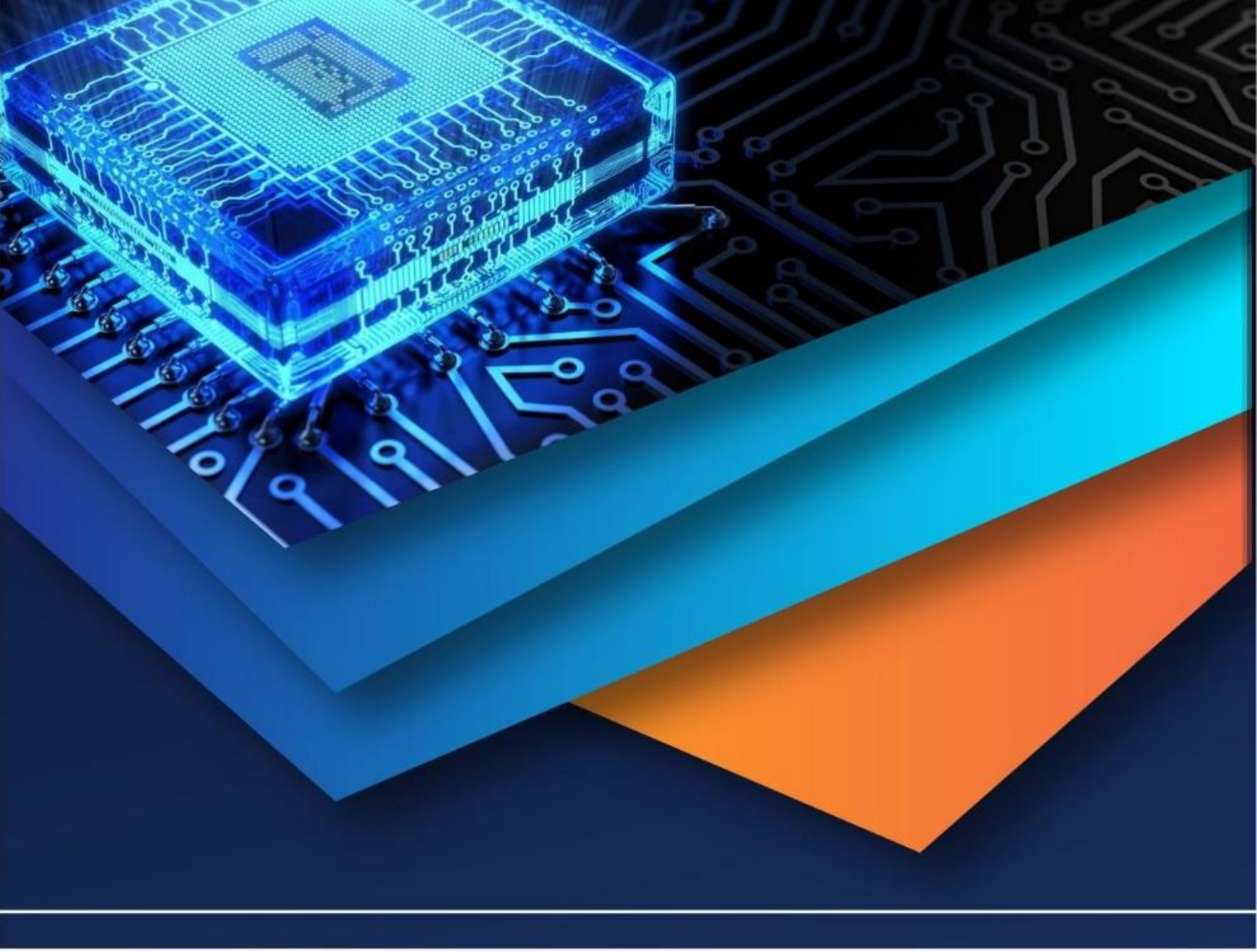
[12] Rubini, L., Soundarapandian, P., & Eswaran, P. (2015). Chronic Kidney Disease [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5G020.

[13] Singla R, Ringstrom C, Hu G, Lessoway V, Reid J, Nguan C, Rohling R. The open kidney ultrasound data set. arXiv preprint arXiv:2206.06657. 2022 Jun 14.

[14] Afkanpour M, Tehrany Dehkordy D, Momeni M, Tabesh H. Conceptual framework as a guide to choose appropriate imputation method for missing values in a clinical structured dataset. BMC Med Res Methodol. 2025 Feb 20;25(1):43. doi: 10.1186/s12874-025-02496-3. PMID: 39979819; PMCID: PMC11843774.

[15] Dina Elreedy, Amir F. Atiya, A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance, Information Sciences, Volume 505, 2019, Pages 32-64, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2019.07.070.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)