Mesopotamian Journal of Artificial Intelligence in Healthcare
Vol.2024, **pp**. 118–117
DOI: https://doi.org/10.58496/MJAIH/2024/014; ISSN: 3005-365X
https://mesopotamian.press/journals/index.php/MJAIH

Research Article

An optimized Framework for kidney disease Detection and Prediction Using DL Techniques

El-Sayed M. El-Kenawy [1,2,3,*], Marwa M. Eid [4], Ahmed M. Elshewey [5], Ahmed M. Osman [6]

[1] *School of ICT, Faculty of Engineering, Design and Information & Communications Technology (EDICT), Bahrain Polytechnic, PO Box 33349, Isa Town, Bahrain*

[2] *Jadara University Research Center, Jadara University, Jordan*

[3] *Applied Science Research Center, Applied Science Private University, Amman, Jordan*

[4] *Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura 11152, Egypt*

[5] *Department of Computer Science, Faculty of Computers and Information, Suez University, P.O.Box:43221, Suez, Egypt*

[6] *Department of Information Systems, Faculty of Computers and Information, Suez University, P.O.Box:43221, Suez, Egypt*

ARTICLE INFO

Article History

Received 18 May 2024
Accepted 13 July 2024
Accepted 19 Aug 2024
Published 20 Sep 2024

Keywords

Health risk prediction
Deep learning
kidney disease Detection
CKD
CDK Ecosystem

ABSTRACT

In recent years, chronic kidney disease (CKD) has become a crucial public health concern and issue, resulting in serious mortality rates. Therefore, we introduce a powerful system for kidney disease diagnosis that develops up-to-date approaches and applications. Our powerful system is meant for early detection, individual patient monitoring, and optimal CKD patient outcomes. The study introduces a CKD detection ecosystem primarily intended for optimized early detection techniques to be able to manage significant weaknesses in current systems and also for proactive treatment. This ecosystem also provides medical practitioners with up-to-date diagnostic technology, incorporating advanced Convolutional Neural Networks (CNN) that guarantee accuracy and efficiency of the diagnosis phase. Furthermore, it facilitates patient empowerment via an individualized mobile application meant for passive surveillance and active engagement in the treatment cycle. Our study introduces this ecosystem with two components, including (1) a deep sequential model for CKD diagnosis and (2) a CNN-based diagnosis for kidney conditions. We have developed the system in order to manage the shortcomings of CKD treatment, deal with the challenges faced by CKD patients by empowering them, and support the global attempts intended to reduce the CKD mortality rates. In addition, we anticipate it to significantly transform the kidney disease treatment field all over the world, demonstrating an effective performance with an accuracy of 98.75%.

1. INTRODUCTION

CKD is considered a global health threat that results in millions of mortalities on an annual basis. Bangladesh and Egypt experience an extensive prevalence of kidney failure, with 40,000 patients per annum in Bangladesh along with thousands of premature mortalities in Egypt. The annual CKD mortality rate in Egypt has significantly reached 25%, surpassing the global average rate of 10%.

CKD is characterized by limited kidney function due to a failure of the kidneys to purify blood. The major role associated with the kidneys is to carefully remove excess waste as well as water from the bloodstream and eliminate them through the mechanism of urine formation. When an individual has CKD, there is a pile of waste in their body, leading to symptoms that include problems with their back, constipation, nose bleeding, a rash, a high temperature, vomiting, and abdominal cramps [1]. Due to the slow accumulation of damage over time, it will have an impact on every part of the human body and give rise to numerous diseases. As the condition progresses and reaches its last phases, it might lead to mortality [2]. The implementation of modern computer-based detection techniques in medical studies has resulted in improved mortality rates by enabling early detection of diseases and aiding doctors in providing prompt treatments.

*Corresponding author. Email: skenawy@ieee.org

The CKD is regarded as a global threat, which our Kidney Disease Diagnosis System could handle. It causes millions of mortalities each year, globally, reflecting a significant impact on mortality rates among patients. The CKD situation becomes highly severe in Egypt, reflecting high mortality rates among kidney failure patients [3]. The healthcare system in Egypt stresses the requirement of advanced CKD treatment technologies. The system is developed for early detection, individual patient monitoring, and optimized outcomes, supporting the global attempts intended to reduce CKD mortality rates.
This paper specifically implements a kidney disease diagnosis system for early detection and prevention of CKD, as well as for health care monitoring of CKD patients. Ecosystem aims to improve early detection methods for proactive intervention, develop a sophisticated diagnostic tool using convolutional neural networks for CKD diagnosis, and empower patients through a mobile application for active participation in their health management journey. The system introduced in our paper comprises two applications. The first one is desktop software developed for kidney disease diagnosis and individual patient monitoring in medical facilities, while the second is a smartphone app designed for tracking the patient's biometric parameters and alerting health abnormalities or threats. Our ecosystem targets early CKD prediction by examining the application of various deep learning models, which represents a significant system of intelligent technologies for early disease prediction and features substantial life-saving advantages.
Beside this section is the remaining body of the paper. Section 2 presents the relevant research, which includes a range of investigations carried out in the area of interest. The proposed methodology is simply explained in Section 3. The simulations performed in Section 4 are intended to evaluate the performance of the proposed technique. At the end of the paper, the conclusion and prospective areas for future investigation are offered in Section 5.

2. RELATED WORK

In medical research, machine learning commonly employs supervised learning techniques. Every supervised learning experience reflects an input object, along with a comparator that returns the intended output value, representing a supervised signal [4]. Commonly employed techniques for supervised learning encompass decision trees (DTs), naive Bayes (NB) classification, least squares regression, logistic regression (LR), and support vector machine (SVM) algorithms [5]. Recent studies highlighted the efficacy of deep neural networks (DNN) performing in the classification phases of both natural and biomedical images [6]. In merit of their performance, along with their ability to facilitate hypothesis development, they stress the vital role of deep learning (DL) in supporting medical studies [7]. DL's adaptability to diverse dataset processing and the availability of open-source DL tools further enhance its role in facilitating medical research.
Several researchers have proposed algorithms to forecast health risks associated with various diseases, with the goal of decreasing mortality rates. Because of the high level of risk that kidney illness poses to human health, scientists have made efforts to identify it at an early stage or predict its emergence earlier. Disease detection refers to the identification of an existing disease in a patient, whereas disease prediction refers to the anticipation of its occurrence in the future. Consequently, research has been categorized into two distinct types: detection and prediction.
Almansour et al. [1] employed SVM and artificial neural networks (ANN) to identify CKD in its initial phase. The dataset underwent preprocessing, followed by the replacement of missing values. Subsequently, ten-fold cross-validation was employed. The study determined that ANN achieved higher accuracy compared to SVM, with accuracy reaching up to 99.75%. The study has been limited by a small number of samples, resulting in the issue of dimensionality. The SVM algorithm was used to tackle this problem. This paper proposes the utilization of a deep learning technique for the detection of CKD.
Elhoseny et al. [8] proposed a method called Density based Feature Selection (DFS) with Ant Colony based Optimization (D-ACO) to address CKD by eliminating unnecessary characteristics in medical data. By implementing this technique, the ability for different systems to work together, the speed of calculations, and the problem of overfitting were all enhanced. The study employs precisely 14 out of the 24 features provided, revealing a high accuracy of 95%.
A deep neural network model has been developed by Kriplani et al. for early CKD detection [9]. This model has applied cross-validation to overfitting risk management and demonstrating optimized performance, which could surpass other developed models. Compared to Naïve Bayes, Logistic Regression, Random Forests (RF), Adaboost, and SVM, it could reach a high accuracy of 97%.
For the best accuracy outcomes, employing the previous dataset, Jongbo et al. have applied an ensemble approach that combines random subspace and bagging, which could attain an accuracy of 100% [10]. This approach is perfectly developed for effective CKD diagnosis. The data undergoes preprocessing, followed by handling of missing values, and ultimately normalization. The algorithm utilized majority voting among three base-learners: KNN, Naïve Bayes, and Decision Tree. This study found that combining the fundamental classifiers improved classification efficiency. According to the experimental data, the suggested model outperformed the separate classifiers in the performance matrices. The random subspace algorithm generally exhibited superior performance compared to the bagging approach.

Ekanayake et al. [11] introduced a highly effective approach for identifying CKD using medical data. The method starts with preprocessing the data and then uses the K Nearest Neighbors-imputer technique to fill in missing values. This makes the detection model more accurate. Ultimately, the categorization method was employed. Their attention was on the pragmatic elements of gathering data, with an emphasis on the significance of integrating domain expertise in the identification of chronic kidney disease through machine learning. The authors conducted tests on 11 classifiers and found that the extra tree and RF classifiers were superior in detecting CKD compared to other classifiers such as logistic regression, KNN, SVC with a linear kernel, SVC with an RBF kernel, Gaussian NB, decision tree classifier, XGB classifier, Adaboost classifier, and a classical neural network. The K-Nearest Neighbors- algorithm is suggested for this study to address missing values in different diseases. Moreover, it is crucial to incorporate additional elements into the analysis, including dietary categories, water intake, and genetic expertise.
Gudeti et al. [12] evaluated the effectiveness of several machine learning methods in 2020 by assessing their precision in analyzing CKD and differentiating between CKD and non-CKD patients. The authors employed Logistic Regression, SVM, and K-Nearest Neighbors (KNN) models to identify CKD. The SVM model demonstrated superior performance compared to the other approaches, obtaining an accuracy rate of 99.2%. The primary advantage of this research is the expedited identification process, which enables prompt initiation of patient treatment and efficient categorization of the patient population within a shorter timeframe. Nonetheless, they utilized a limited dataset consisting of 400 cases.
Chittora et al. [13] employed a range of strategies to identify CKD, such as utilizing all available features, employing correlation-based feature selection, implementing the wrapper methodology, utilizing LASSO, and employing synthetic minority oversampling and synthetic minority oversampling methods. Additional classifiers encompassed C5.0, CHAID, ANN, LSVM, logistic regression, random tree, and KNN. The LSVM attained the maximum level of accuracy, reaching 98.86%, when utilizing all available features.
Senan et al. [14] implemented a machine learning approach for early detection of CKD, employing mean and mode to substitute missing values and recursive feature elimination to identify crucial features. They partitioned the dataset, allocating 75% for training and 25% for testing and validation. Four machine learning methods employed in this study include SVM, RF, K-nearest neighbors (KNN), and DT. The RF algorithm achieved a 100% perfect accuracy rate.
Table 1 provides a brief review of the most recent studies on health risk detection and prediction for CKD.

TABLE 1. SUMMARY OF THE MOST RECENT STUDIES ON HEALTH RISK DETECTION AND PREDICTION FOR KIDNEY DISEASE.

Used techniques	Best performance	Dis adv.	Adv.	Ref. No
SVM/ANN	The accuracy of ANN is 99.75%.	The sample size is restricted, and the goal is to identify or recognize.	A thorough comparison was conducted between SVM and ANN, yielding encouraging outcomes.	[1]
DFS / (D-ACO)	95%	The system is highly complex. Its main purpose is to detect.	less features	[8]
DNN, SVM, NB, Adaboost and LR	96.6%	Limited dataset and main purpose to detect.	A new methodology with high accuracy has been introduced, making its implementation accessible and reproducible.	[9]
The ensemble consisting of KNN, NB, and DT.	100%	lacking comparison to the current leading methods, aimed at detection.	Ensemble strategy surpasses the performance of current methods.	[10]
Utilizing a total of 11 distinct classifiers.	Extra Tree Classifier and Random Forest with a 100% accuracy rate.	Limited dataset with the aim of detection.	Assesses 11 machine learning algorithms with notable accuracy.	[11]
SVM, LR, and KNN.	SVM achieved an accuracy of 99.2%.	Limited dataset with no examination of the influence for various FS techniques on performance, focusing on detection as the goal.	A limited number of features, ensuring the reproducibility of the proposed approach.	[12]
C5.0, CHAID, ANN, LSVM, LR, RT, and KNN.	LSVM achieved an accuracy rate of 98.86%.	Limited dataset and detection were the main target.	Target seven algorithms and identifies significant features.	[13]

SVM, RF, KNN, DT.	Random Forest achieving a perfect accuracy rate of 100%.	Limited dataset and no comparison with our ecosystem.	Achieving exceptional accuracy by employing Recursive Feature Elimination (RFE) to identify crucial features and optimizing classifier parameters.	[14]

3. MATERIALS AND METHODS

3.1. Dataset

The research utilized a shared dataset from Kaggle repository available at https://www.kaggle.com/datasets/mansoordaku/ckdisease/data. This dataset consists of several patient variables which seeks to forecast the probability of kidney disease by analyzing missing values, outliers, and inconsistencies. Table 2 displays the statistical analysis of Kidney dataset numerical data.

TABLE II. THE STATISTICAL ANALYSIS OF KIDNEY DATASET NUMERICAL DATA.

Attribute	**Mean**	**Std**	**Min**	**Max**
age	51.48	16.97	2.00	90.00
bp	76.47	13.48	50.00	180.00
sg	1.017	0.005	1.005	1.025
al	1.01	1.27	0.00	5.00
su	0.45	1.03	0.00	5.00
bgr	148.04	74.78	22.00	490.00
htn	0.37	0.48	0.00	1.00
dm	0.34	0.47	0.00	1.00
cad	0.08	0.27	0.00	1.00
appet	0.79	0.40	0.00	1.00
pe	0.19	0.39	0.00	1.00
ane	0.15	0.36	0.00	1.00
classification	0.62	0.48	0.00	1.00

Figure 1 displays the heatmap for numerical data of the shared kidney dataset.

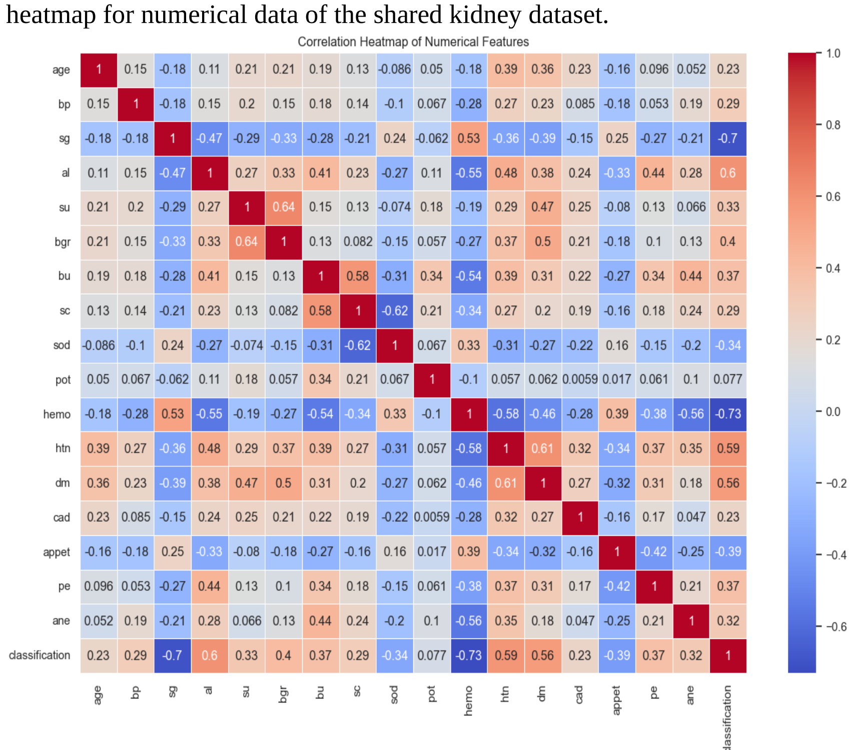

Fig. 1. heatmap of Kaggle kidney dataset.

Figure 2 displays the boxplot of the Kaggle kidney dataset attributes.

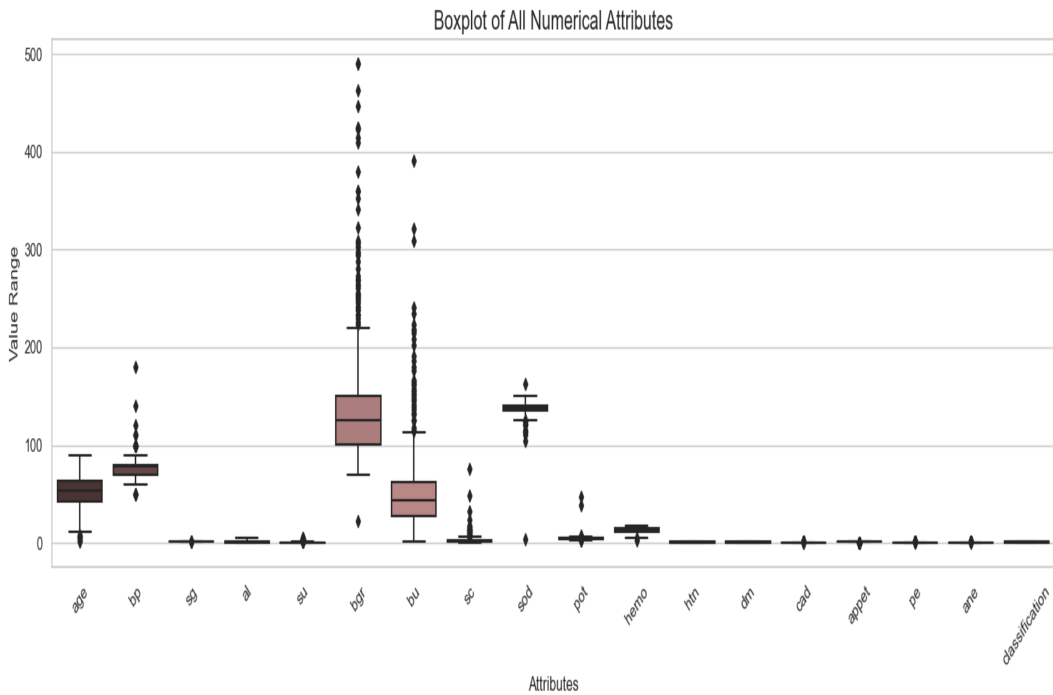

Fig. 2. boxplot of the Kaggle kidney dataset attributes.

Figure 3 displays the histogram of dataset attributes.

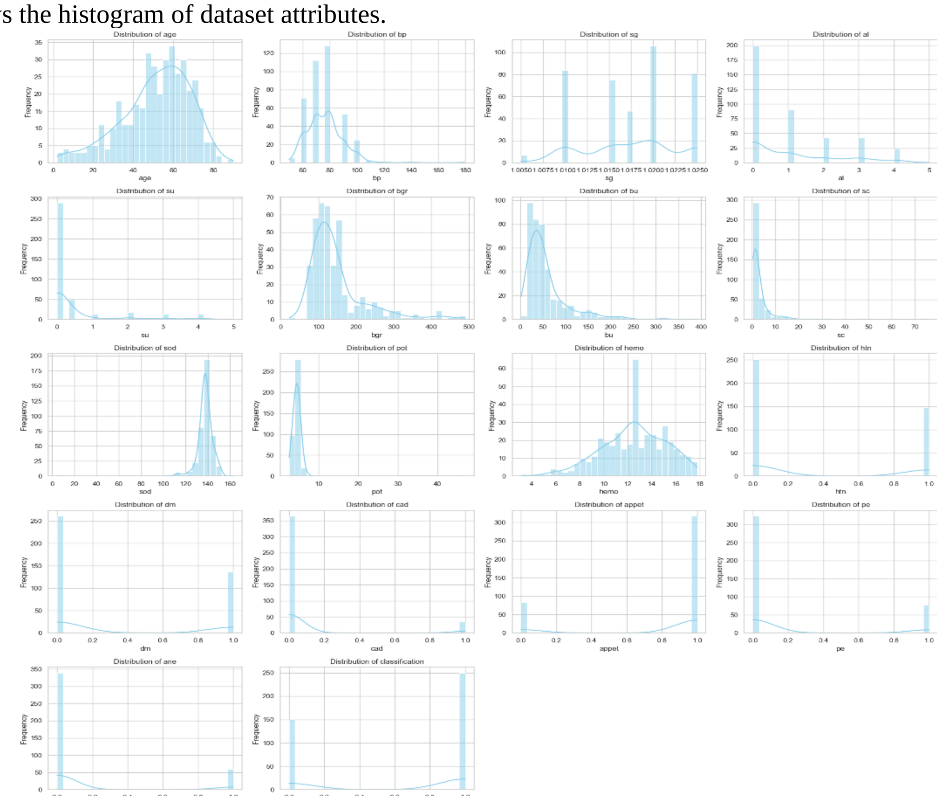

Fig. 3. histogram of the Kaggle kidney dataset attributes.

Figure 4 summarizes preprocessing tasks.

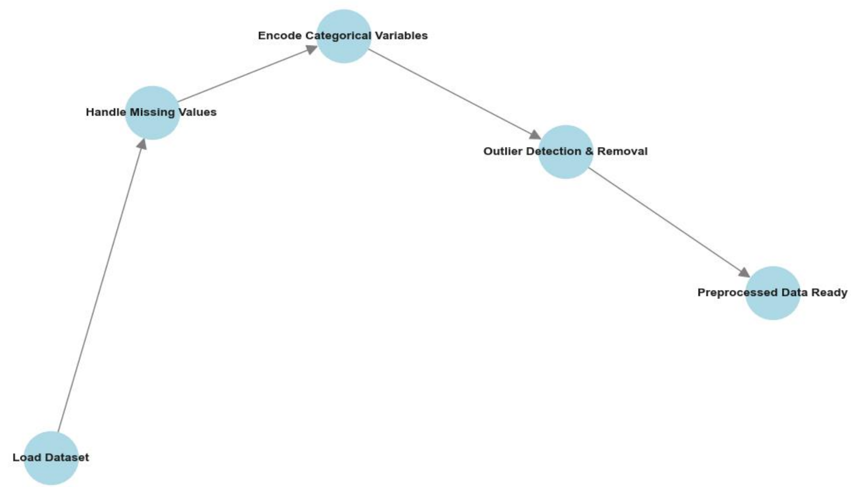

Fig. 4. preprocessing summary flowchart.

3.2 Proposed Framework

Deep learning and LSTM and CNN networks are used to classify renal illness in the proposed model. Standardized numerical data describing patient characteristics is fed into the model architecture and moulded into 2D for CNNs or 3D for LSTMs. The feature extraction layer identifies local patterns, succeeded by max-pooling layers that decrease dimensionality and collect predominant characteristics. The LSTM model learns dependencies and correlations among characteristics by processing input as a series. The output from the feature extraction layer is flattened and sent to dense layers, while dropout layers reduce overfitting by randomly excluding a subset of neurons during training. The terminal output layer is a dense layer utilizing a sigmoid activation function, generating a probability score ranging from 0 to 1, signifying the chance of kidney disease occurrence.

The model workflow encompasses many stages: data preparation, model training, hyperparameter tuning, assessment, and deployment and prediction. The model is evaluated through many metrices like accuracy, precision, recall, F1-score, and ROC AUC. A confusion matrix is created to evaluate true positives, false negatives, false positives, and true negatives. The resulting model is preserved and implemented for real-time prediction and may be included into clinical decision support systems to aid healthcare practitioners in the efficient diagnosis of kidney illness.

The integration of CNN and LSTM models provides an effective method for feature extraction and sequence learning. Figure 5 displays the model framework.

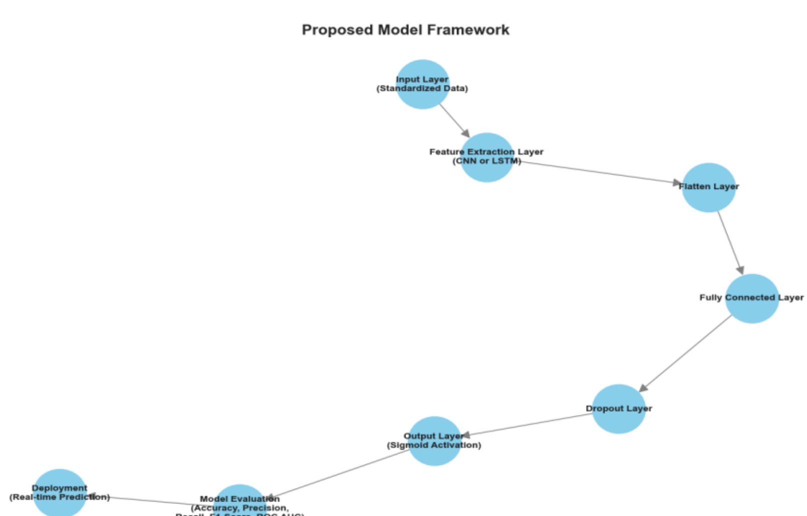

Fig. 5. The proposed framework.

4. RESULTS and DISCUSSION

The suggested CNN and LSTM deep learning models were applied to the renal disease classification problem utilizing a preprocessed numerical dataset. The findings demonstrate that the models are capable of effectively predicting the existence of renal disease based on the characteristics of clinical data. Within this section, the evaluation of metrics is presented, a comparison of the performance of the models is made, and a discussion of the implications of the findings take place.
Table 3 displays the performance of the suggested models.

TABLE III. THE PERFORMANCE EVALUATION OF THE SUGGESTED MODELS.

Metric	CNN Model	LSTM Model
Accuracy	98.75%	96.25%
Precision	100.00%	100.00%
Recall	98.08%	94.23%
F1-Score	99.03%	97.03%
AUC	99.04%	97.12%

This paper used evaluation metrics [15-17] such as accuracy, F1 score, sensitivity (recall), specificity, and AUC. Metrics calculated through the following equations

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{1}$$

$$Sensetivity = \frac{TP}{TP+FN} \tag{2}$$

$$Specificity = \frac{TN}{TN+FP} \tag{3}$$

$$F - score = \frac{2\times Recall \times Precision}{Recall+Precision} \tag{4}$$

$$AUC = 1/2\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \tag{5}$$

Figure 6 displays the confusion matrix of the CNN.

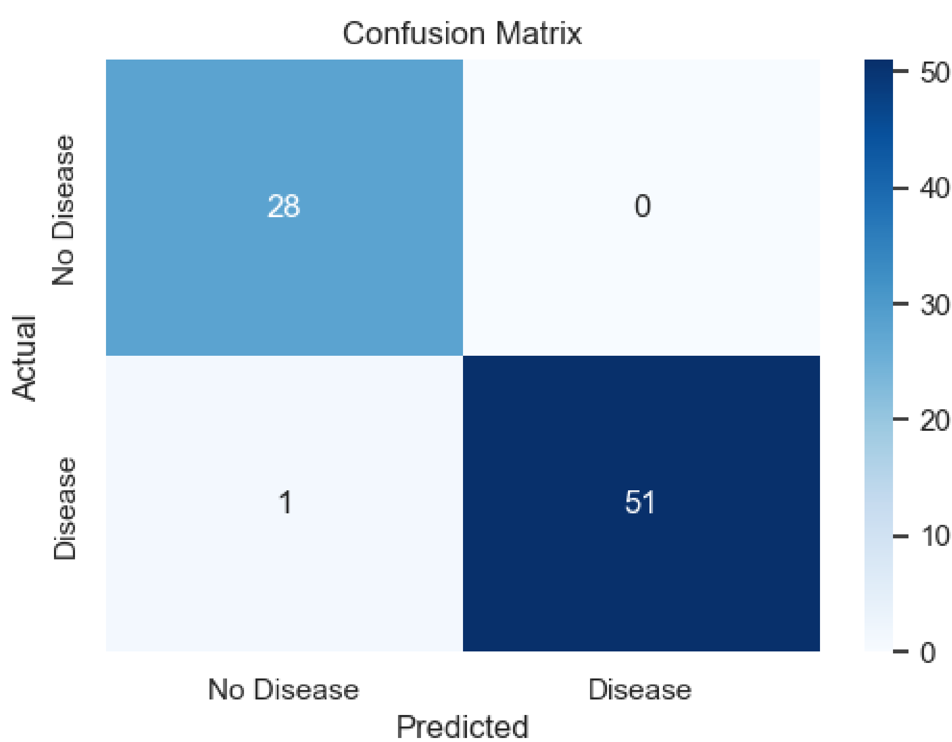

Fig. 6. CNN confusion matrix.

Figure 7 displays the confusion matrix of LSTM.

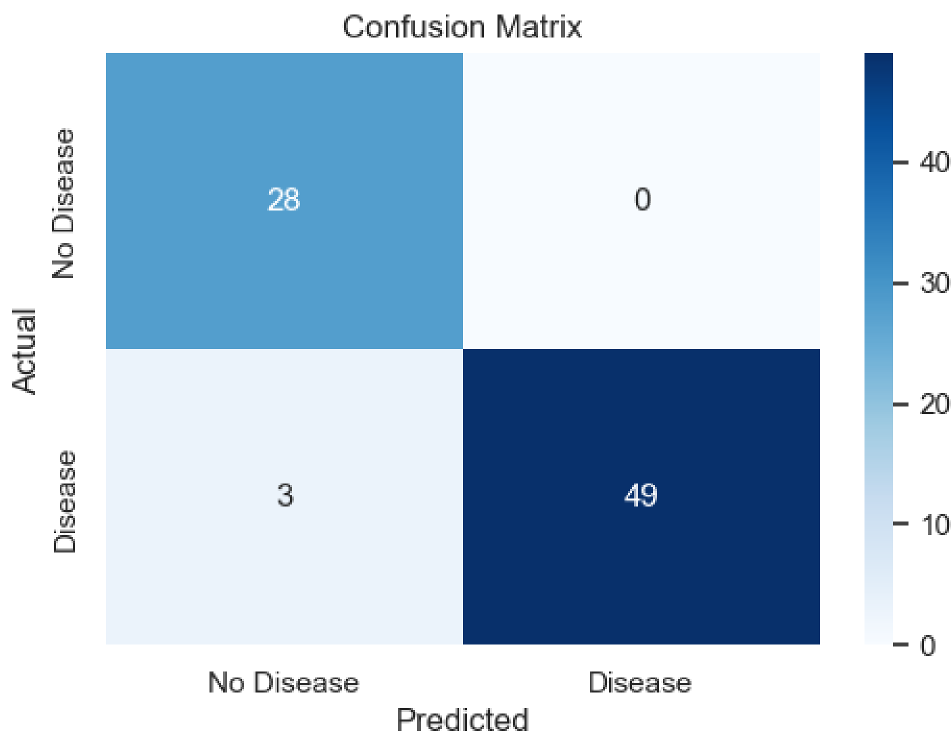

Fig. 7. LSTM confusion matrix.

Figure 8 displays training and validation accuracy of the CNN model.

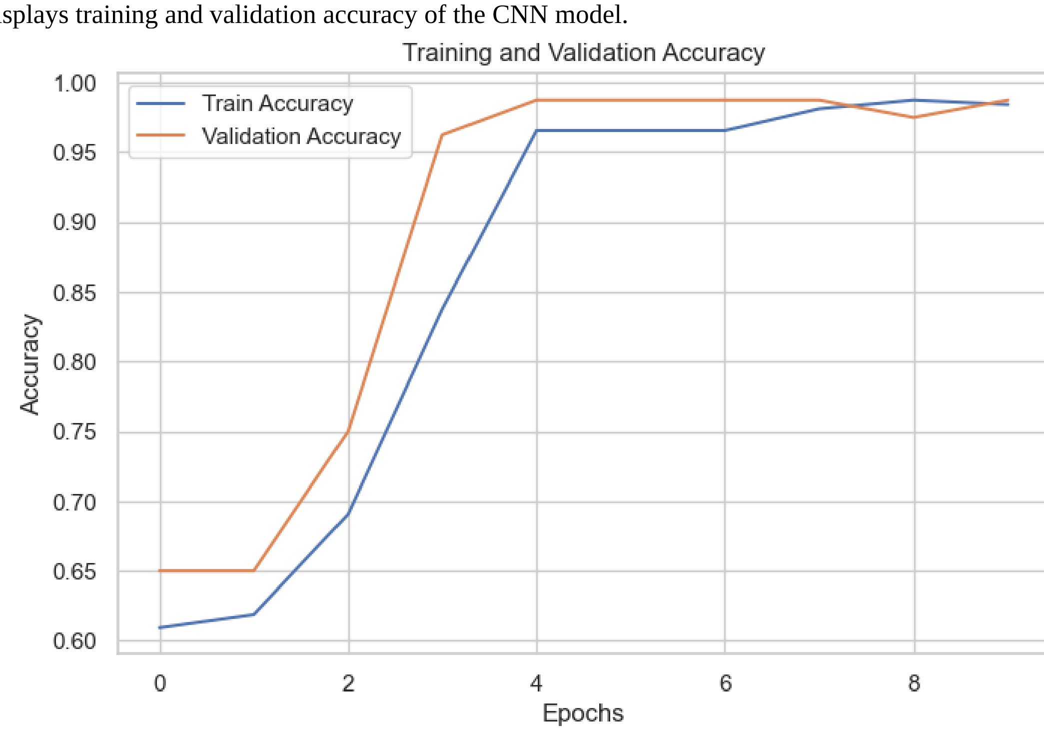

Fig. 8. training and validation accuracy of the CNN model.

.

5. CONCLUSIONS and FUTURE WORK

This ecosystem also provides medical practitioners with up-to-date diagnostic technology, incorporating advanced Convolutional Neural Networks (CNN) that guarantee accuracy and efficiency of the diagnosis phase. Furthermore, it facilitates patient empowerment via an individualized mobile application meant for passive surveillance and active engagement in the treatment cycle. Our study introduces this ecosystem with two components, including (1) a deep sequential model for CKD diagnosis and (2) a CNN-based diagnosis for kidney conditions. We have developed the system in order to manage the shortcomings of CKD treatment, deal with the challenges faced by CKD patients by empowering them, and support the global attempts intended to reduce the CKD mortality rates. In addition, we anticipate it to significantly transform the kidney disease treatment field all over the world, demonstrating an effective performance with an accuracy of 98.75%. Future research should focus on refining the models to further enhance their performance, robustness, and interpretability.

The use of the IoT to manage humidity in public spaces such as schools, offices and transportation hubs will improve IAQ while curbing the spread or airborne diseases ultimately improving overall public health [10][11]. And the deployment of AI and machine learning in conjunction with IoT systems allows for predictive maintenance and proactive interventions to keep environmental conditions stable at their optimum all times [12][13]. These advancements all go towards minimising energy, saving on operational costs while lowering carbon footprint [14]. In summary, IoT in humidity measurement and control has a significant positive impact on the quality of life in smart cities as it becomes integral to creating healthier, more comfortable and sustainable urban environments.

Funding

The author's paper asserts that the research was conducted on a voluntary basis and without any financial backing from institutions or sponsors.

Conflicts Of Interest

The absence of any competing relationships or biases that could affect the research is explicitly mentioned in the paper.

Acknowledgment

The author acknowledges the institution for their commitment to fostering a research-oriented culture and providing a platform for knowledge dissemination.

References

[1] N. A. Almansour, et al., "Neural network and support vector machine for the prediction of chronic kidney disease: a comparative study," Comput. Biol. Med., vol. 109, pp. 101–111, 2019.

[2] H. Kriplani, B. Patel, and S. Roy, "Prediction of chronic kidney diseases using deep artificial neural network technique," in Computer Aided Intervention and Diagnostics in Clinical and Medical Images, Springer, 2019, pp. 179–187.

[3] O. A. Jongbo, et al., "Development of an ensemble approach to chronic kidney disease diagnosis," Sci. Afr., vol. 8, p. e00456, 2020.

[4] W. Song, Y. Liu, L. Qiu, J. Qing, A. Li, Y. Zhao, et al., "Machine learning-based warning model for chronic kidney disease in individuals over 40 years old in underprivileged areas, Shanxi Province," Front. Med., vol. 9, p. 930541, 2023.

[5] K. Wang, J. Tian, C. Zheng, H. Yang, J. Ren, C. Li, et al., "Improving risk identification of adverse outcomes in chronic heart failure using SMOTE+ ENN and machine learning," Risk Manage. Healthc. Policy, vol. 14, pp. 2453–2463, 2021.

[6] R. Tapia-Conyer, H. Gallardo-Rincón, and M. Betancourt-Cravioto, "Chronic kidney disease in disadvantaged populations: online educational programs for NCD prevention and treatment," in Chronic Kidney Disease in Disadvantaged Populations, Academic Press, 2017, pp. 337–345.

[7] S. Krishnamurthy, K. Ks, E. Dovgan, M. Luštrek, P. B. Gradišek, K. Srinivasan, et al., "Machine learning prediction models for chronic kidney disease using national health insurance claim data in Taiwan," Healthcare, vol. 9, no. 5, pp. 1–13, 2021.

[8] M. Elhoseny, K. Shankar, and J. Uthayakumar, "Intelligent diagnostic prediction and classification system for chronic kidney disease," Sci. Rep., vol. 9, no. 1, pp. 1–14, 2019.

[9] H. Kriplani, B. Patel, and S. Roy, "Prediction of chronic kidney diseases using deep artificial neural network technique," in Computer Aided Intervention and Diagnostics in Clinical and Medical Images, Springer, 2019, pp. 179–187.

[10] O. A. Jongbo, et al., "Development of an ensemble approach to chronic kidney disease diagnosis," Sci. Afr., vol. 8, p. e00456, 2020.

[11] I. U. Ekanayake and D. Herath, "Chronic kidney disease prediction using machine learning methods," in 2020 Moratuwa Engineering Research Conference (MERCon), IEEE, 2020.
[12] B. Gudeti, et al., "A novel approach to predict chronic kidney disease using machine learning algorithms," in 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), IEEE, 2020.
[13] P. Chittora, et al., "Prediction of chronic kidney disease—a machine learning perspective," IEEE Access, vol. 9, pp. 17312–17334, 2021.
[14] E. M. Senan, et al., "Diagnosis of chronic kidney disease using effective classification algorithms and recursive feature elimination techniques," J. Healthc. Eng., vol. 2021, pp. 1–10, 2021.