



Diagnosis of Chronic Kidney Disease Based on CNN and LSTM

Elif Nur Yildiz¹, Emine Cengil², Muhammed Yildirim³, Harun Bingol^{1*}

¹ Department of Software Engineering, Faculty of Engineering and Natural Sciences, Malatya Turgut Ozal University, 44200 Malatya, Turkey

² Department of Computer Engineering, Faculty of Engineering and Architecture, Bitlis Eren University, 13000 Bitlis, Turkey

³ Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Malatya Turgut Ozal University, 44200 Malatya, Turkey

* Correspondence: Harun Bingol (harun.bingol@ozal.edu.tr)

Received: 04-09-2023

Revised: 05-02-2023

Accepted: 05-20-2023

Citation: E. N. Yildiz, E. Cengil, M. Yildirim, and H. Bingol, "Diagnosis of chronic kidney disease based on CNN and LSTM," *Acadlore Trans. Mach. Learn.*, vol. 2, no. 2, pp. 66–74, 2023. <https://doi.org/10.56578/ataiml020202>.



© 2023 by the authors. Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: Kidney plays an extremely important role in human health, and one of its important tasks is to purify the blood from toxic substances. Chronic Kidney Disease (CKD) means that kidney begins to lose its function gradually and show some symptoms, such as fatigue, weakness, nausea, vomiting, and frequent urination. Early diagnosis and treatment increase the likelihood of recovery from the disease. Due to high classification performance, artificial intelligence techniques have been widely used to classify disease data in the last ten years. In this study, a hybrid model based on Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) was proposed using a two-class data set, which automatically classified CKD. This dataset consisted of thirteen features and one output. If the features showed, CKD was diagnosed. Compared with many well-known machine learning methods, the proposed CNN-LSTM based model obtained a classification accuracy of 99.17%.

Keywords: Artificial intelligence; Convolutional Neural Network (CNN); Machine learning; Kidney diseases; Long Short-Term Memory (LSTM)

1 Introduction

Kidney provides internal balance of the human body by producing urine. Blood flows to the kidney, glomerular filtration, tubular reabsorption, and tubular secretion to make the kidney perform its functions [1]. Kidney provides the salt and water ratio in the body, helps remove waste materials constantly, keeps blood pressure at a normal level, increases blood production, plays a role in bone structure, and excretes the substances taken during the day with urine. Although kidney has so many functions, it suffers damage resulting from wrong reactions in the human body, which leads to various kidney diseases and affects other organs, because the human body is considered as a whole. People know that the kidney is not healthy in accordance with several symptoms, such as fatigue, vomiting, swelling in the feet and hands, difficulty in breathing, frequent urination, and burning sensation in the urine. CKD diagnosis result is a structural or functional abnormality in the kidney. CKD is caused by congenital or acquired diseases progressively destroying the kidney parenchyma, such as inflammatory (chronic glomerulonephritis), infectious (chronic pyelonephritis), or degenerative (amyloidosis) diseases. Sometimes, congenital diseases (ren polycystic) lead to CKD and renal insufficiency. In affluent nations, diabetic nephropathy is the most typical cause of chronic renal failure, which is followed by primary (hypertensive) nephrosclerosis caused by hypertension, chronic glomerulonephritis, and other diseases [2]. Early disease diagnosis is important, including CKD. Necessary treatments can be applied under the control of a doctor to prevent morbidity or mortality, thus obtaining positive results. With the technological advance, it is increasingly common to detect diseases with artificial intelligence techniques today, and one of their most important features is the highly accurate classification.

Sharma et al. [3] examined the effectiveness of various machine learning algorithms in predicting medical diagnosis and focused on CKD identification. The dataset was composed of 400 samples and 24 characteristics, and twelve categorization approaches were applied to examine CKD. The predictions made using the potential approaches were compared with the subject's actual medical results in order to determine efficacy. Accuracy,

precision, sensitivity, and specificity metrics were used for performance evaluation, which achieved 98.6% accuracy in the decision tree classifier. Eyupoglu [4] developed a new clinical decision support system based on PCA and RF techniques for the early diagnosis of CKD. Several metrics were used to test the performance of the proposed system, such as accuracy, precision, sensitivity, F-measure, MCC, and AUC. Test results were compared with classical machine learning algorithms and previous studies in the literature, which showed that the proposed system was effective and could be used as an auxiliary tool in the early diagnosis with an accuracy of 99.75%. According to the study of Chittora et al. [5], a professional quickly identified chronic renal illness by utilizing classifier algorithms in machine learning. The CKD dataset was taken from the UCI repository. Seven classifier techniques were used in the study, including a random tree, artificial neural network, C5.0, Chi-square automatic interaction detector, logistic regression, and linear support vector machine with L1 and L2 penalties. A technique for selecting significant features was also used on the data set. The experimental results of each classifier were computed using the following methods: (i) full features; (ii) correlation-based feature selection; (iii) Wrapper method feature selection; (iv) least absolute shrinkage and selection operator regression; and (v) synthetic minority oversampling technique. Avci and Doğantekin [6] used the genetic algorithm - wavelet kernel - edge machine learning method for CKD diagnosis, with 24 different features of 400 people as the data. The developed method was used to classify CKD data, with the 400*24 feature vector as the input. The developed model was evaluated in terms of classification accuracy, sensitivity, and specificity metrics, which showed 98.42% classification success rate. Amirgaliyev et al. [7] used a support vector machine (SVM) algorithm to examine the effects of classifying patients with CKD using clinical features. Clinical history, physical exams, and laboratory tests were considered as the foundation for the CKD dataset. Based on three performance parameters, experimental results showed a success rate of over 93% in diagnosing individuals with kidney illness. Ghosh et al. [8] used machine learning classifiers in SVM, AdaBoost, linear discriminant analysis, and gradient boosting experiments. The highest prediction accuracy with approximately 99.80% was obtained from gradient boosting classifiers. Sobrinho et al. [9] investigated the early detection of CKD in underdeveloped nations using machine learning methods, and obtained 93.33%, 88.33%, 76.66%, 75.00%, and 71.67% accuracy, respectively, by using random forest, naive Bayes, SVM, multilayer perceptron and KNN classifiers in the experiments.

Classification of kidney diseases was discussed in Section 2. The dataset, CNN, LSTM, machine learning methods, and the proposed hybrid model were given in Section 3. Experimental results were presented in Section 4. Finally, the investigation results were discussed in Section 5.

2 Methodology

CNN and LSTM were utilized in the model constructed in this study, which was compared with other machine learning techniques, namely, the K Nearest Neighbor (KNN), SVM, logistic regression, naive Bayes, random forest, and AdaBoost.

2.1 Dataset

The dataset used for CKD detection in the experiments was obtained from the Kaggle platform, and was publicly available from www.kaggle.com/datasets/abhia1999/chronic-kidney-disease [10]. The data set consisted of 400 samples and 13 features. This two-class data set was labeled by experts according to whether a patient had CKD or not. Five randomly selected samples from the data set are presented in Figure 1.

Bp	Sg	Al	Su	Rbc	Bu	Sc	Sod	Pot	Hemo	Wbcc	Rbcc	Htn	Class
90.0	1.025	1.0	0.0	1.0	18.0	1.2	140.0	4.2	12.53	8406.0	4.71	0.0	1
70.0	1.010	1.0	3.0	0.0	55.0	1.7	138.0	4.5	12.60	10200.0	4.10	1.0	1
90.0	1.010	4.0	1.0	1.0	309.0	13.3	124.0	6.5	3.10	5400.0	2.10	1.0	1
80.0	1.025	0.0	0.0	1.0	10.0	1.2	135.0	5.0	15.00	10400.0	4.50	0.0	0
80.0	1.025	0.0	0.0	1.0	36.0	1.0	150.0	4.6	17.00	9800.0	5.00	0.0	0

Figure 1. Samples from the CKD data set

Data normalization in the data preprocessing stage was completed before classifying the data in the data set. The standardization procedure changed the mean and standard deviation of each column to zero and one, respectively, which was crucial for many machine learning algorithms, because it obtained more precise benchmarks when the data had different scales or units. As a result, the preprocessed data was employed in machine learning classifiers of the constructed model in this study.

2.2 CNN, LSTM and Classifiers

In this study, a CNN-LSTM based model was developed to detect CKD. CNNs transmitted to the artificial neural network according to the parameters determined on the image with the feature extraction layer, which obtained the features distinguished from each other, thus helping train the neural network. The feature extraction layer generally

consisted of more than one layer and obtained results [11]. LSTM, on the other hand, was a type of architecture that controlled the recurrent neural network (RNN) style and the information stage of cells. It changed during the process with different door structures. The filtered part of the cell in this process was known as the final output [12].

In order to compare the performance metrics of the proposed model, different classifiers mentioned in the literature were used. Due to simple implementation and outstanding performance, KNN was widely used as a classifier in data mining and machine learning applications, which observed the proximity of the data to be classified to k of the previous data. The testing and training data were compared in the classification phase of the KNN algorithm, with Euclidean connection as the preferred neighborhood distance measurement method [13]. SVM, on the other hand, covered many principles and solved classification problems. It was basically based on two ideas. According to the first idea, it was used to specify the accuracy of models trained in different classes and converted them to the feature space. The other idea was to introduce the concept of the best margin to find the best hyperplane [14]. The logistic regression model was used to analyze the effects of the results obtained. It was useful in diagnosing disease, and determining whether the disease was in the human body. When done more than once, the model was called a multivariate logistic model. The logistic regression model was known as the statistical model, which was frequently used in the field of medicine [15]. The naive Bayes model had good classification performance. It was used in complex methods and provided correct output. The model estimated the probability and conveyed the correct part to the class even if it was wrong [16]. Random forest model was known as a classification or regression tree. When making predictions, the result was obtained with the majority or average vote [17]. AdaBoost was the boosting algorithm, and was an efficient and useful model. The theory also had a strong effect and obtained successful results in practice, which contributed to algorithm design method and idea [18].

2.3 Proposed Model for CKD Detection

A hybrid model was developed for CKD classification using CNN and LSTM. In the developed model, two Convolution, two LSTM, two Dropout, one Maximum Pooling, two Dense, and one Flatten were used. Leaky Relu and Softmax were used as activation functions, and Adam was chosen as the optimizer. The proposed model was run for 200 epochs. An architecture of the proposed model is given in Figure 2. In brief, the transform layer was the convolution layer, which was the foundation of CNN. The most popular rectifier unit for the outputs of CNN neurons was the Relu layer, which came after the convolution layer. Following the Relu layer was the maximum pooling layer, which mainly reduced the input size of the following convolutional layer. The dropout layer prevented the network from memorizing when the training process was performed. The Softmax layer, on the other hand, performed probabilistic calculation of the probabilistic value produced in the layer within the deep learning network, and revealed the probability value for each class [19, 20].

CNN and LSTM were combined to create a hybrid model for CKD diagnosis. The automatic feature selection in the architecture was used to distinguish deep learning architecture from classical machine learning method. In manual methods, feature extraction was finished by an expert. Therefore, this was one of the advantages of the proposed model. In addition, the model had quite low number of layers and parameters.

3 Results

Results were obtained by comparing different machine learning classifiers with the proposed model based on CNN and LSTM. 70% of the data in the data set was used for training and 30% for testing. Confusion matrix was created, which was the result of the training phase, and the values of the matrix were used to obtain performance evaluation metrics. KNN, SVM, logistic regression, naive Bayes, random forest, and AdaBoost classifiers were used in this study. In addition, results were obtained on the same data set as the proposed hybrid model.

3.1 Results of Classifiers

In this study, whether the patient was diagnosed with CKD was determined in accordance with the application. A comparison was made between the classifiers and the proposed model, which showed that the proposed model obtained the best diagnosis results. This study obtained results from six classifiers using the default parameters.

As shown in Figure 3, a 400-sample data set is used for the KNN model classifier, with 70% samples for training and 30% for testing. Thus, 120 samples were allocated for the test. By observing the two-class confusion matrix, it was found that the first class had 36 correct samples and 8 incorrect ones, and the second class had 53 correct ones and 23 incorrect ones. The accuracy rate of the KNN model classifier was 74.16%.

As shown in Figure 4, a 400-sample data set is used for the SVM model classifier, with 70% samples for training and 30% for testing. Thus, 120 samples were allocated for the test. By observing the 2-class confusion matrix, it was found that the first class had 23 correct samples and 21 incorrect ones, and the second class had 67 correct ones and 9 incorrect ones. The accuracy rate of the SVM model classifier was 75%.

As shown in Figure 5, a 400-sample data set is used for the logistic regression model classifier, with 70% samples for training and 30% for testing. Thus, 120 samples were allocated for the test. By observing the 2-class confusion

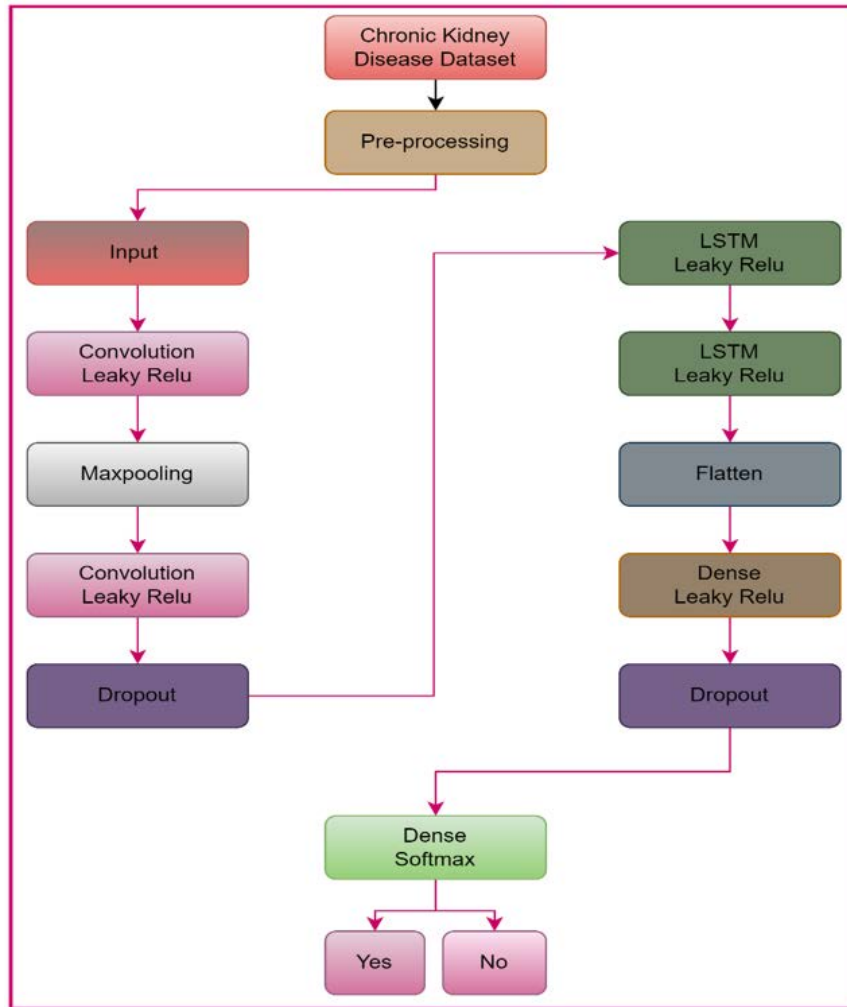


Figure 2. Architecture of the proposed CNN-LSTM model

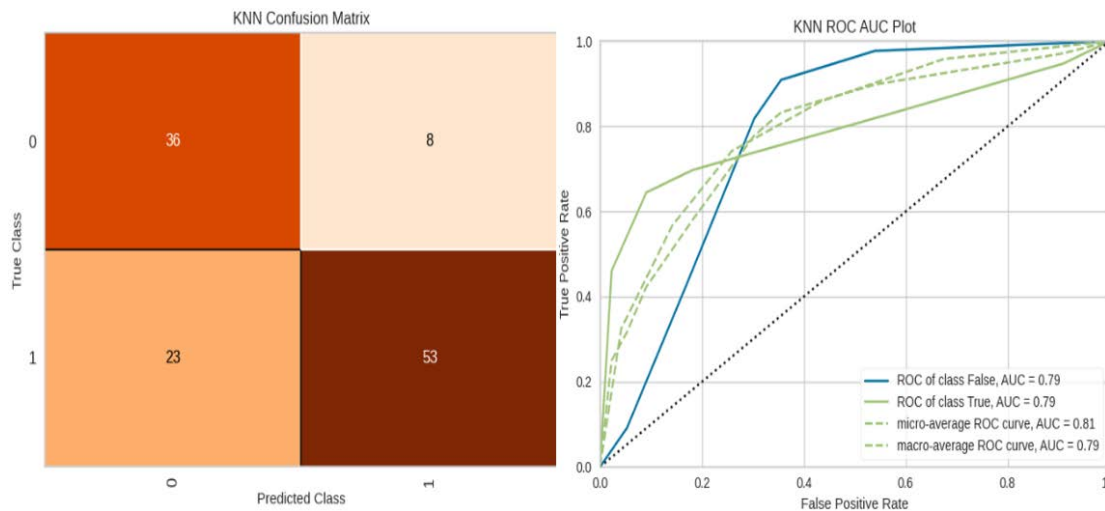


Figure 3. KNN model confusion matrix and ROC/AUC curves

matrix, it was found that the first class had 43 correct samples and 1 incorrect sample, and the second class had 74 correct ones and 2 incorrect ones. The accuracy rate of the logistic regression model classifier was 97.5%.

As shown in Figure 6, a 400-sample data set is used for the naive Bayes model classifier, with 70% samples for training and 30% for testing. Thus, 120 samples were allocated for the test. By observing the two-class confusion

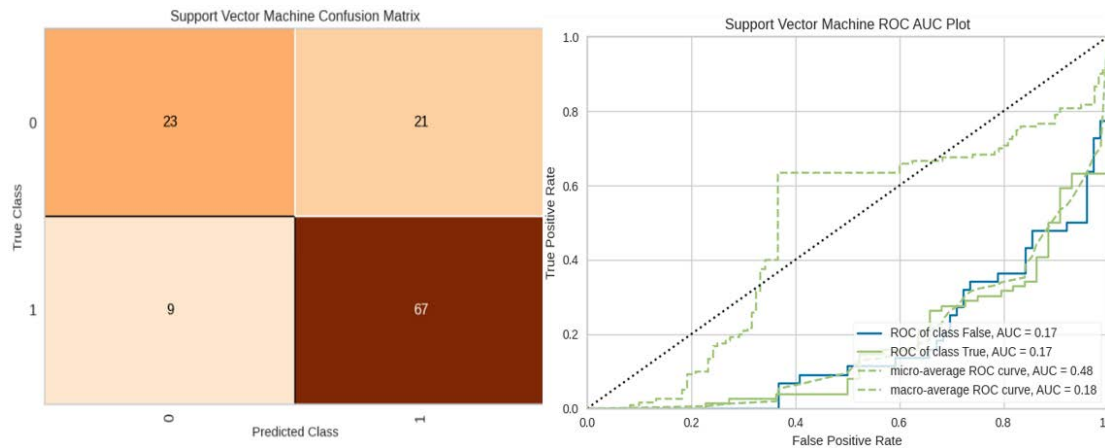


Figure 4. SVM model confusion matrix and ROC/AUC curves

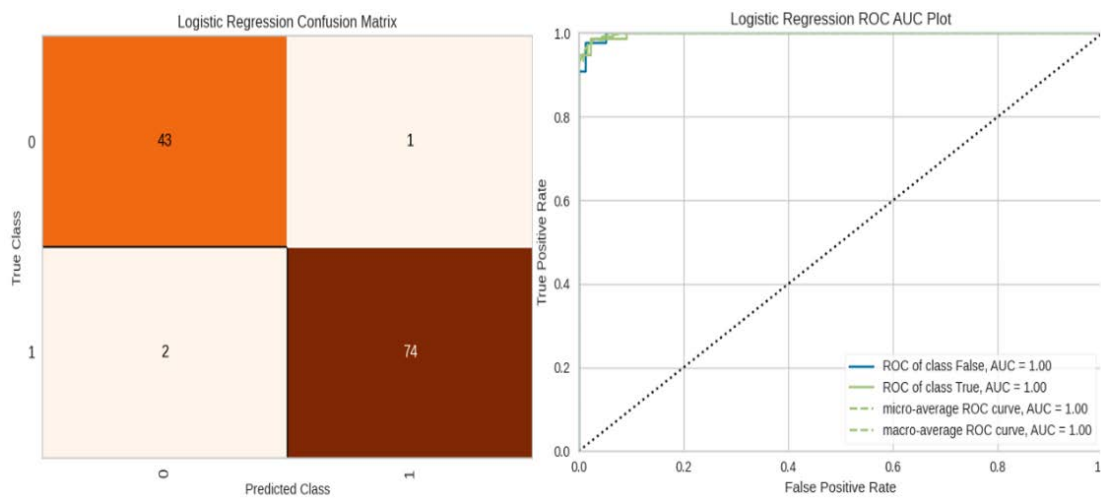


Figure 5. Logistic regression model confusion matrix and ROC/AUC curves

matrix, it was found that the first class had 44 correct samples, and the second class had 67 correct ones and 9 incorrect ones. The accuracy rate of the naive Bayes model classifier was 92.5%.

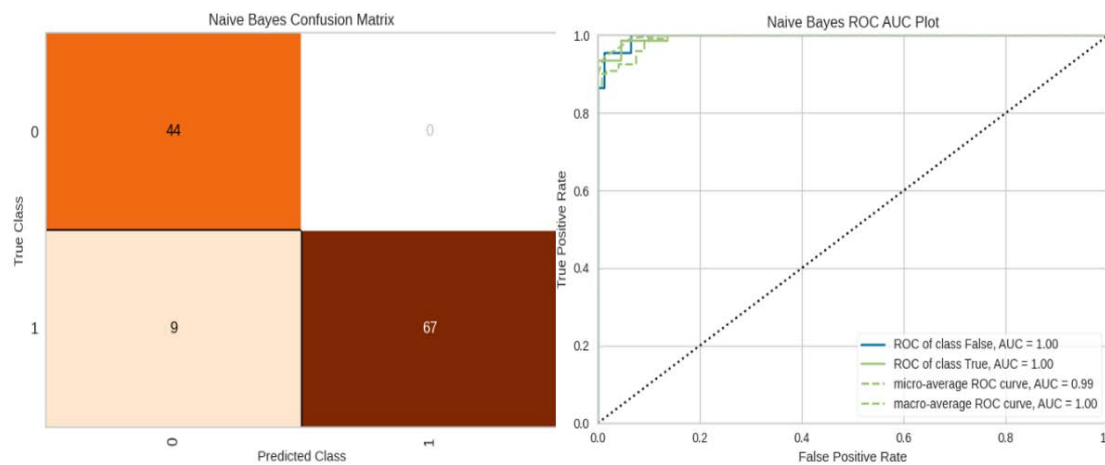


Figure 6. Naive Bayes model confusion matrix and ROC/AUC curves

As shown in Figure 7, a 400-sample data set is used for the random forest model classifier, with 70% samples for training and 30% for testing. Thus, 120 samples were allocated for the test. By observing the two-class confusion matrix, it was found that the first class had 44 correct samples, and the second class had 75 correct ones and 1

incorrect sample. The accuracy rate of the random forest model classifier was 99.16%.

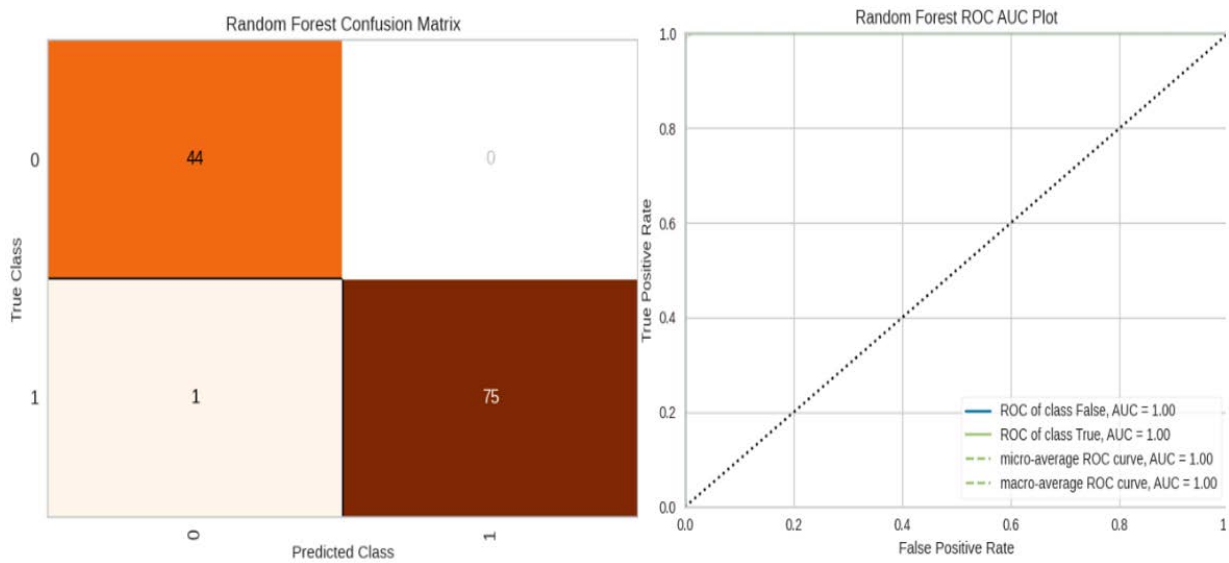


Figure 7. Random forest model confusion matrix and ROC/AUC curves

As shown in Figure 8, a 400-sample data set is used for the AdaBoost model classifier, with 70% samples for training and 30% for testing. Thus, 120 samples were allocated for the test. By observing the two-class confusion matrix, it was found that the first class had 42 correct samples and 2 incorrect ones, and the second class had 73 correct ones and 3 incorrect ones. The accuracy rate of the AdaBoost model classifier was 95.83%.

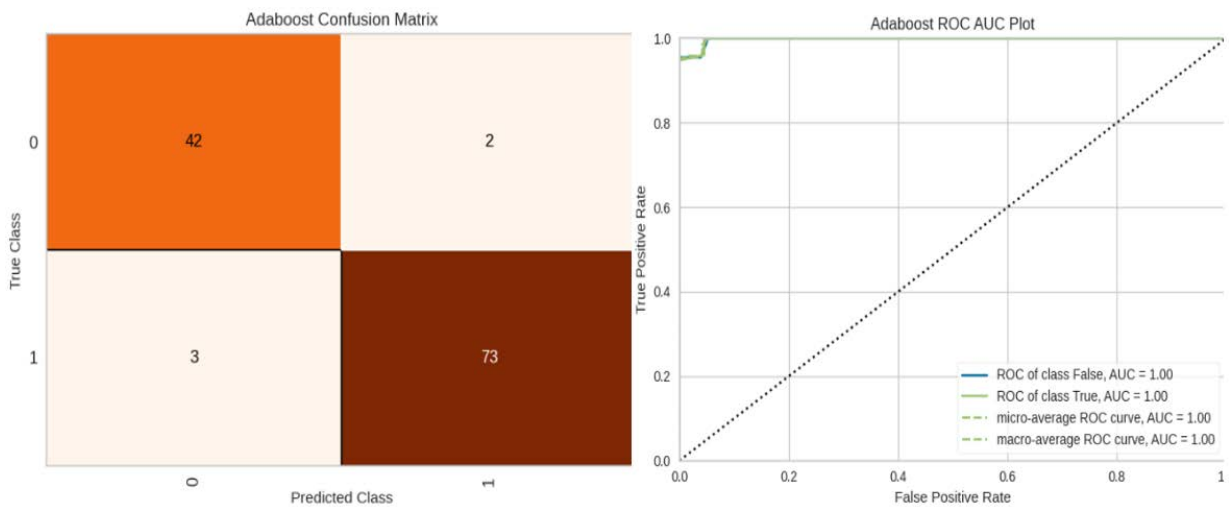


Figure 8. AdaBoost model confusion matrix and ROC/AUC curves

3.2 Results of the Proposed CNN-LSTM Based Model

The accuracy and loss curves obtained by running the CNN-LSTM based hybrid model for 200 epochs are shown in Figure 9 and Figure 10, respectively. As shown in Figure 9, an epoch value of 200 is sufficient, because the validation and training values become flatter as they approach the 200 epoch value.

Results from six classifiers mentioned in the literature were acquired to test the performance of the constructed model. The confusion matrix in Figure 10 reveals that the proposed model is competitive.

The confusion matrix of the proposed model is shown in Figure 10. It can be seen from the figure that only 1 of the 120 samples in the matrix is classified incorrectly by the proposed model.

3.3 Comparison of Results

It was observed that the model constructed in this study achieved successful results. Table 1 presents the performance evaluation metrics of the conventional classifiers and the proposed model for CKD detection.

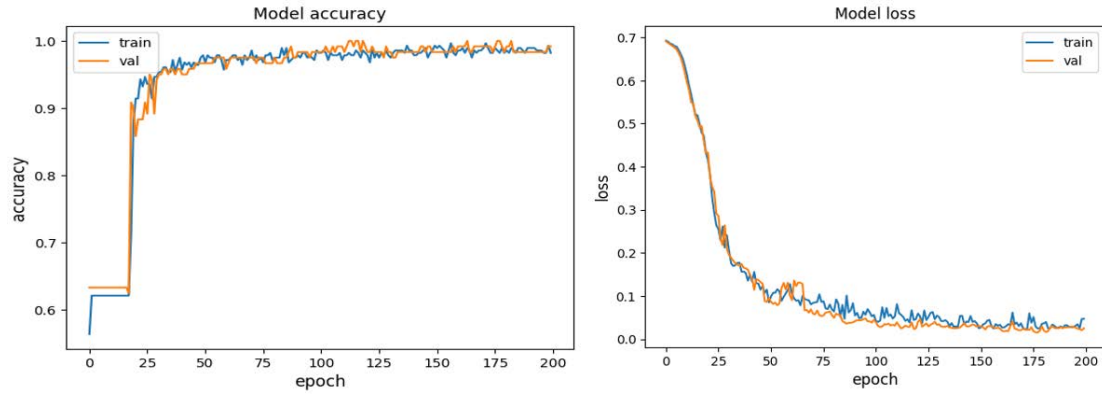


Figure 9. Accuracy and loss curves of the proposed model

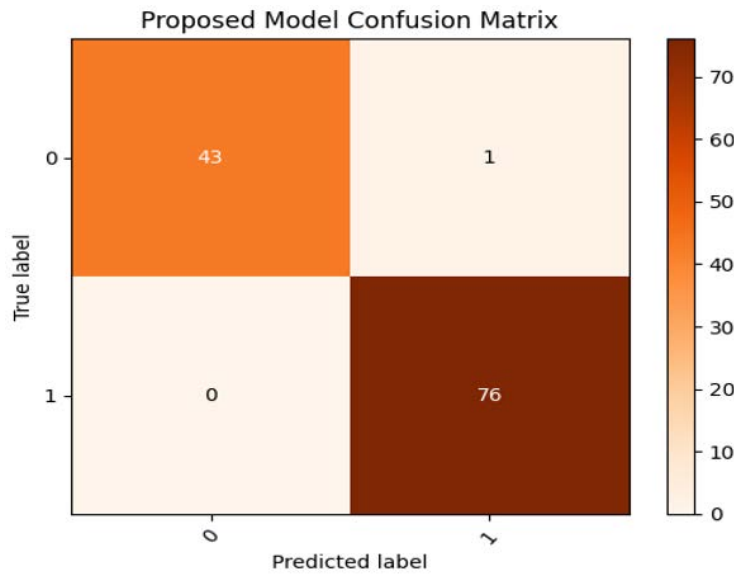


Figure 10. Confusion matrix of the CNN-LSTM based model

Table 1. Performance values of the classifiers and the proposed CNN-LSTM based model

	F1	FDR	FNR	FPR	Accuracy	Sensitivity	Specificity	MCC
KNN	69.90	18.18	38.98	13.11	74.17	61.02	86.89	49.70
SVM	60.53	47.73	28.13	23.86	75.00	71.88	76.14	60.53
Logistic regression	96.63	2.27	4.44	1.33	97.50	95.56	98.68	94.66
Naive Bayes	16.98	0	16.98	0	92.50	83.02	100	85.55
Random rorest	98.88	0	2.22	0	99.16	97.78	100	98.23
AdaBoost	94.38	4.55	6.67	2.67	95.83	93.33	97.33	91.09
Proposed model (CNN-LSTM)	98.85	2.27	0	1.30	99.17	100	98.70	98.21

As shown in Table 1, the CNN-LSTM based hybrid model has 99.17% accuracy, which is followed by random forest (99.16%), logistic regression (97.50%), Adaboost (95.83%), naive Bayes (92.50%), SVM (75%), and KNN (74.17%).

4 Conclusion

As one of the important organs of human body, kidney has various functions, and responds to the decrease of a healthy diet or large water consumption in the body in various ways. Therefore, early diagnosis, control and necessary treatment stop the progression of the disease. This study discussed the diagnosis of CKD in accordance with the response of human body. In this study, the accuracy of 74.16%, 75%, 97.5%, 92.5%, 99.16%, and 95.83% were obtained by the classifiers of KNN, SVM, logistic regression, naive Bayes, random forest, and AdaBoost,

respectively. By comparing the results, the best classifier was selected for the model. The CNN-LSTM based model proposed by this study obtained the highest accuracy rate of 99.17%. The most important limitation of this study was the small number of data in the data set. In addition, the data was not collected from a multicenter data set. In future studies, new models can be developed to detect CKD on a multicenter data set with more data.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] C. Mangan, M. C. Stott, and R. Dhanda, "Renal physiology: blood flow, glomerular filtration and plasma clearance," *Anaesth. Intensive Care Med.*, vol. 19, no. 5, pp. 254–257, 2018. <https://doi.org/10.1016/j.mpaic.2018.02.013>
- [2] "Kronik böbrek hastalığı," 2020. <https://gesund.bund.de/tr/kronik-bobrek-hastaligi>
- [3] S. Sharma, V. Sharma, and A. Sharma, "Performance based evaluation of various machine learning classification techniques for chronic kidney disease diagnosis," *arXiv preprint arXiv:1606.09581*, 2016. <https://arxiv.org/abs/1606.09581>
- [4] C. Eyupoglu, "Kronik böbrek hastalığının erken tanısı için yeni bir klinik karar destek sistemi," *Avrupa Bilim Teknol. Dergisi*, no. 20, pp. 448–455, 2020. <https://doi.org/10.31590/ejosat.743652>
- [5] P. Chittora, S. Chaurasia, P. Chakrabarti, G. Kumawat, T. Chakrabarti, Z. Leonowicz, M. Jasinski, L. Jasin-ski, R. Gono, E. Jasinska, and V. Bolshev, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," *IEEE Access*, vol. 9, pp. 17 312–17 334, 2021. <https://doi.org/10.1109/access.2021.3053763>
- [6] D. Avci and A. Doğantekin, "Kronik Böbrek Hastalığının Teşhisi için Genetik Algoritma Dalgacık çekirdeği Uç öğrenme Makinesine Dayalı Uzman Bir Sistem," *Fırat Univ. Muh. Bilim. Dergisi*, vol. 34, no. 1, pp. 75–84, 2022. <https://doi.org/10.35234/fumbd.962824>
- [7] Y. Amirgaliyev, S. Shamiluulu, and A. Serek, "Analysis of chronic kidney disease dataset by applying machine learning methods," in *2018 IEEE 12th International Conference on Application of Information and Communi-cation Technologies*. IEEE, Almaty, Kazakhstan, 2018. <https://doi.org/10.1109/icaict.2018.8747140>
- [8] P. Ghosh, F. M. Javed Mehedi Shamrat, S. Shultana, S. Afrin, A. A. Anjum, and A. A. Khan, "Optimization of prediction method of chronic kidney disease using machine learning Algorithm," in *2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing*. IEEE, Bangkok, Thailand, 2020. <https://doi.org/10.1109/isai-nlp51646.2020.9376787>
- [9] A. Sobrinho, A. C. M. D. S. Queiroz, L. Dias Da Silva, E. De Barros Costa, M. Eliete Pinheiro, and A. Perkusich, "Computer-aided diagnosis of chronic kidney disease in developing countries: A comparative analysis of machine learning techniques," *IEEE Access*, vol. 8, pp. 25 407–25 419, 2020. <https://doi.org/10.1109/access.2020.2971208>
- [10] Kronik Kidney Disease. <https://www.kaggle.com/datasets/abhia1999/chronic-kidney-disease>
- [11] A. Emre and M. Turan, "Video duygu analizi," *Avrupa Bilim ve Teknoloji Dergisi*, no. 36, pp. 319–319, 2020.
- [12] M. J. Hamayel and A. Y. Owda, "A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms," *AI*, vol. 2, no. 4, pp. 477–496, 2021. <https://doi.org/10.3390/ai2040030>
- [13] S. Zhang, X. Li, M. Zong, X. Zhu, and D. Cheng, "Learning k for kNN Classification," *ACM T. Intel. Syst. Tec.*, vol. 8, no. 3, pp. 1–19, 2017. <https://doi.org/10.1145/2990508>
- [14] J. Chorowski, J. Wang, and J. M. Zurada, "Review and performance comparison of SVM- and ELM-based classifiers," *Neurocomputing*, vol. 128, pp. 507–516, 2014. <https://doi.org/10.1016/j.neucom.2013.08.009>
- [15] T. G. Nick and K. M. Campbell. *Topics in Biostatistics*, 2007.
- [16] E. Frank, L. Trigg, G. Holmes, and I. H. Witten, "Technical note: Naive bayes for regression," *Mach. Learn.*, no. 1, pp. 5–25, 2000. <https://doi.org/10.1023/a:1007670802811>
- [17] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, "Random Forest: A classification and regression tool for compound classification and QSAR modeling," *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 6, pp. 1947–1958, 2003. <https://doi.org/10.1021/ci034160g>
- [18] Y. Cao, Q. Miao, J. Liu, and L. Gao, "Advance and prospects of AdaBoost Algorithm," *Acta Automatica Sinica*, vol. 39, no. 6, pp. 745–758, 2013. [https://doi.org/10.1016/s1874-1029\(13\)60052-x](https://doi.org/10.1016/s1874-1029(13)60052-x)
- [19] M. Yildirim and A. Cinar, "A deep learning based hybrid approach for COVID-19 disease detections," *Trait. Signal*, vol. 37, no. 3, pp. 461–468, 2020. <https://doi.org/10.18280/ts.370313>

- [20] M. S. Bugday, M. Akcicek, H. Bingol, and M. Yildirim, "Automatic diagnosis of ureteral stone and degree of hydronephrosis with proposed convolutional neural network, relief, and gradientweighted class activation mapping based deep hybrid model," *Int. J. Imaging. Syst. Tech.*, vol. 33, no. 2, pp. 760–769, 2023. <https://doi.org/10.1002/ima.22847>