

Original Article

An Optimized Cluster Convolution based Hybrid Hierarchical Deep Learner for Chronic Kidney Disease (CKD)

P. Usha¹, N. Kavitha²

^{1,2}Department of Computer Science, Nehru Arts and Science College, Coimbatore, India

¹Corresponding Author : blessed.usha@gmail.com

Received: 25 August 2022

Revised: 02 November 2022

Accepted: 16 November 2022

Published: 26 November 2022

Abstract - Chronic Kidney Disease (CKD) remains a universal health issue. Learning the attributes relevant to CKD patients might help improve the early prediction of CKD. However, the outliers in the CKD database may affect the prediction accuracy. To solve this issue, a Moth Flame (MF)-based DBSCAN with Pearson Correlation (MFDBSCAN-PC) algorithm was suggested that adjusts the DBSCAN variables during clustering of the CKD-related attributes and creates the cleansed database without outliers. Also, various machine learning classifiers were performed for CKD prediction. But, the complex and implicit temporal relationships between local and global attributes were not learned, influencing the learning of the time-series CKD database. Therefore, this article proposes a Deep Clustering-based outlier removal and Convolutional Neural network with a Hierarchical Bi-directional Long Short-Term Memory (DC-CNN-HBLSTM) model to process the time-series database for CKD prediction. Primarily, the CKD information-rich dataset is pre-processed using data imputation to fill up any blanks in the rows. Then, MF optimization performs to select the optimal DBSCAN variables and attributes simultaneously. Based on the best variables, the DBSCAN is employed as a new clustering layer in the CNN structure to cluster the data points and remove the outliers from the database, resulting in a newly cleaned database. After that, the HBLSTM classifier is trained by learning the temporal correlation between the local and global attributes to create a trained model. Further, the trained classification model is used to classify the test instances into healthy and CKD patients. Finally, the experimental results realize the DC-CNN-HBLSTM model achieves 96.21% accuracy compared to the SVM, recursive ANN, and ELM classifiers, which achieved 88.69%, 90.57%, and 92.9% accuracy for the CKD database.

Keywords - Attribute selection, Chronic kidney disease, Clustering, CNN, DBSCAN, Hierarchical bidirectional LSTM, Moth flame optimization, Outliers removal.

1. Introduction

Chronic kidney disease (CKD) is a medical disorder in which a patient gradually loses kidney function. As a result, their standard of living as a whole declined. One out of every ten persons globally suffers from CKD. The prevalence of CKD is increasing, and by 2040, it is expected to rank as the fifth leading cause of death worldwide [1]. It is a major contributor to excessive medical expenses. The expense of transplantation and dialysis amounts to 2%-3% of the yearly medical expenditure in high-income countries [2]. Dialysis and kidney transplants are the only treatment options for people with renal failure in low- and middle-income countries [3]. Unpredictably, the number of people with renal failure is likely to rise in developing countries like China and India [4].

Due to CKD, it becomes more difficult for the kidneys to filter out extra fluid from the body. High levels of fluid, electrolytes, and wastes in the body are potentially lethal consequences of severe chronic renal impairment [5]. Risks

like excessive blood hypertension, anaemia, weak muscles, and sciatica are possible. CKD prediction is poor, mainly since the disease is frequently undetectable until it is severe [6]. Conversely, delayed CKD prognosis leads to less opportunity to avoid adverse effects. Physicians must be attentive to CKD to apply evidence-based therapy that can delay the course of renal impairment, avoid diabetic problems and improve cardiac mortality [7]. CKD is generally incurable, and therapy focuses on preventing the growth of kidney and heart disorders [8]. Despite existing therapies, there is still a concern about the potential effects and CKD development. As a result, early identification, diagnosis, and monitoring of CKD are essential. It is also important to predict the evolution of CKD with fair precision due to its adaptive and hidden character in the early phases and patient heterogeneity. The severity of CKD is frequently defined in terms of stages. The stage, whether a patient is advancing, and the rate of advancement all impact clinical decisions. Further, characterizing the illness stage is critical since it provides various indicators that help determine the necessary intervention and therapies.



So, data mining may play a significant role in extracting confidential data from massive patient medical and clinical datasets that physicians regularly gather from patients to get insights regarding diagnostic information and execute accurate treatment plans [9-11]. Data mining is the technique of retrieving hidden data from massive datasets. Numerous situations and fields make use of data mining techniques. Researchers can try to predict, classify, filter, and cluster data using data mining techniques. From this perspective, many scientists developed and emerged various data mining methods involving pre-processing, attribute extraction, clustering, and classification algorithms for predicting CKD patients and providing precise diagnoses [12]. However, due to the vast number of patients, it is impractical to examine each individual, and persons at a higher risk of having CKD will be advised to undertake comprehensive testing. Maintaining clinical databases is currently a complex task in the healthcare business.

1.1. Research Gap

Most of the classifier models in the literature need more attributes to predict CKD and non-CKD patients appropriately. Also, they have separate pre-processing, clustering, and classification algorithms, resulting in high computation time for medium-sized and large-sized databases. Most studies involve machine learning classifiers, which were unsuitable for large-scale databases due to slow training. The temporal correlation between the attributes was essential to predict CKD patients in an earlier stage.

1.2. Problem Description

To provide high-quality service, the patient's data contains many aspects, and illness diagnosis must be delivered with exceptional care. Because the data saved in the hospital database may contain missing or unneeded information, mining the patient data becomes difficult. Thus, improved data processing and data reduction methodologies are required before applying data mining algorithms. For this reason, an uncertainty handling attribute selection was performed to solve the issue of extensive database handling, choose the significant attributes and predict the CKD patients.

First, the CKD database was grouped by classical the Density-Based Spatial Clustering of Applications with Noise (DBSCAN), which uses randomly assigned epsilon and mid-point variables [13]. However, the clustering was not satisfactory when such variables were not appropriately assigned, i.e., with no prior data. So, a Moth Flame-based DBSCAN with Pearson Correlation (MFDBSCAN-PC) algorithm has been designed to optimize the variables based on the flight route characteristics of the moth flame [14]. The fitness value of each moth flame was determined according to their location, and the attribute's linear relationship among the class variable was measured by Pearson's linear correlation, which helps to group similar attributes. Also, the attributes which do not belong to any group were termed outliers and removed from the database.

Further, the processed database was passed to the different machine learning classifiers for training. Later, the trained classifiers were utilized to predict CKD from testing instances. On the other hand, the isolated processes of clustering and classification were time-consuming for medium-sized and large-sized databases. Also, the complex and implicit correlations between local and global attributes were not considered.

1.3. Research Objective

This research focuses on designing a unified deep learning-based clustering and classification model, which supports clinicians in efficiently predicting CKD patients from the large-scale clinical database.

1.4. Research Contribution

The DC-CNN-HBLSTM framework is developed in this research to handle the time-series database for CKD prediction. The CKD database is initially gathered and pre-processed via data imputation to manage missing data. The moth flame optimization is then done to find the best DBSCAN parameters and the most appropriate attributes. According to the best parameters, the DBSCAN clustering is implemented as an additional layer in the CNN model to generate the clusters and remove the outlier data from the CKD database. It provides a new cleansed CKD database, passed into the HBLSTM classifier to learn the hierarchically temporal correlation between the characteristics and class labels, i.e., CKD data in each interval. As a result, a trained classification model is created and used to categorize the test examples as healthy or CKD patients. Thus, this DC-CNN-HBLSTM model enhances the accuracy of predicting CKD from time-series data.

1.5. Organization of the Paper

Section 2 presents the research performed on predicting and classifying CKD. The presented paradigm is explained in Section 3, and its effectiveness is depicted in Section 4. Conclusions and suggestions for further research are presented in Section 5.

2. Literature Survey

Pasadana et al. [15] presented CKD prediction depending on various decision tree schemes, including Random Forest (RF). But, it needs pre-processing methods to solve the missing values in the dataset. Ogunleye and Wang [16] discussed many machine-learning algorithms for CKD diagnosis and selected Extreme Gradient Boosting (XGBoost) as a base system, which was optimized and trained on each attribute related to CKD. However, feature selection schemes must choose the most significant attributes and reduce the attribute dimensionality.

Qin et al. [17] presented an integrated model which combines logistic regression and RF by perceptron to

diagnose CKD patients. First, the CKD database was acquired and given to the K-Nearest Neighbour (KNN) imputation scheme to handle the missing values. But, the number of instances was limited, so the generalization efficiency of this framework was not satisfactory.

Khamparia et al. [18] designed a new deep learner for CKD categorization based on the stacked autoencoder structure, which applies multimedia information with a SoftMax classifier. First, the CKD database was pre-processed to handle the outliers using the data imputation schemes. Then, the stacked autoencoder was used to capture the significant traits, and those were passed to the SoftMax categorizer to determine the probabilities of each class, i.e. predict the healthy and CKD patients. However, the hyperparameters were not optimized, impacting the efficiency of large-scale datasets.

Ma et al. [26] developed a Heterogeneous Modified Artificial Neural Network (HMANN) for the early identification, partition, and prognosis of CKD. First, the ultrasound kidney scans were collected and pre-processed to enhance the image contrast using Gabor filtering. After that, the images were partitioned to differentiate essential and non-essential pixels. Then, the Haralick, histogram, and adaptive characteristics were retrieved and classified by the HMANN, which involves the SVM, Multi-Layer Perceptron (MLP), and backpropagation algorithms. But the efficiency was less for a limited dataset and needed to analyze the efficiency of other CKD datasets.

Elkholy et al. [20] presented a smart categorization and prediction framework, which employs a modified Deep Belief Network (DBN) to predict kidney-related disorders. In this framework, SoftMax was utilized as an activation factor, and the categorical cross-entropy was utilized as a loss factor. But, its computation cost was high during initialization. Ilyas et al. [21] presented decision tree algorithms for CKD prognosis. First, the database was created using the CKD patients' clinical files. Then, pre-processing was applied to manage missing values. Moreover, J48 and RF classifiers were performed to predict the CKD severity levels. However, the random forest classifier was challenging to apply in real-time prediction due to the enormous amount of trees and the slow training process.

Peng et al. [22] designed a 2-level neural network to predict CKD using ultra-dimensional training databases. The pre-processing feature selection was conducted at the initial level using different algorithms. The second level used the single-layer neural network to learn the database and predict CKD patients. Nevertheless, the entire framework was not effective while using several redundant attributes.

Rashed-AI-Mahfuz et al. [23] presented machine learning frameworks using selective major pathological types to detect medical analysis characteristics, which supports the precise

early prognosis of CKD. First, the database was processed and converted into the desired form with missing value imputations. Then, the shapely additive explanations scheme was applied to choose the significant attributes. Further, those attributes were fed to the RF, gradient boosting, XGBoost, logistic regression, and SVM algorithms for CKD classification. But these classifiers were not suitable for processing the time-series attributes, which may need further increase the classification accuracy.

Antony et al. [24] designed an intelligent system to categorize patients into CKD or non-CKD, which supports physicians in coping with several patients and rapidly offers a prognosis. First, data collection and regularization schemes were performed on the database to remove the outliers. Then, different attribute selection schemes like filtering, wrapper-based, embedded-based and unsupervised schemes were conducted to select the most significant attributes and decrease the number of input parameters. Moreover, those attributes were fed to the K-means clustering, Isolation forest (I-forest), DBSCAN, and autoencoder for classification. However, it needs to learn the temporal attributes to enhance the classification efficiency.

Singh et al. [25] designed a Deep Neural Network (DNN) to identify and predict CKD earlier. First, the mean of the related attributes utilizes to substitute each missing value in the dataset. Then, the best parameters of the neural network were set by initializing the parameters and performing many tests. The primary relevant attributes were chosen by Recursive Feature Elimination (RFE) and fed to the DNN for categorizing healthy and CKD patients. Nevertheless, the dataset was limited and needed additional attributes to improve the prediction efficiency.

3. Proposed Methodology

This section describes the DC-CNN-HBLSTM model briefly. A block diagram of the presented study illustrates in Figure 1. The significant contributions of this study are:

- First, the CKD database is obtained from the UCI repository and pre-processed using the data imputation technique to fill in the missing values in the database.
- Then, the moth flame optimization is performed to choose the most optimal attributes and DBSCAN's parameters simultaneously.
- After that, unsupervised CNN clustering is performed, which involves the DBSCAN algorithm as an additional layer in the standard CNN model to cluster the database and identify the outliers. The identified outliers are discarded from the database, and the new cleansed database is generated for further processing.

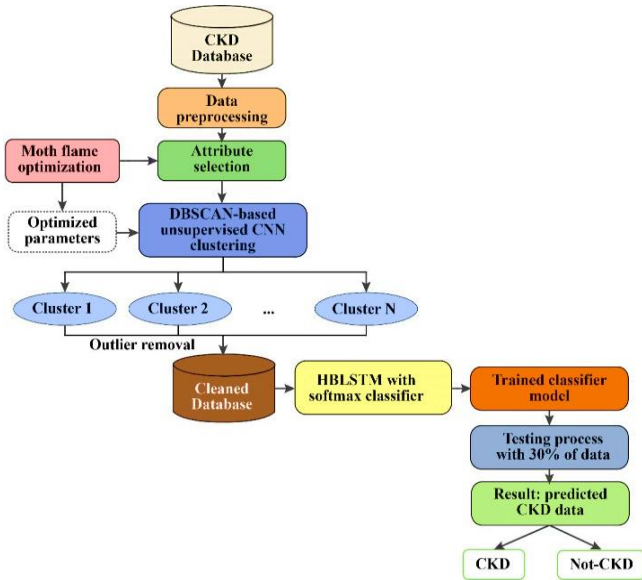


Fig. 1 Block diagram of the presented study for CKD prediction

- Moreover, the HBLSTM classifier is trained by the cleansed database, which learns the temporal correlation between the local and global attributes for the effective prediction of CKD. The trained classification model is later used to classify the test data into CKD or not-CKD.

3.1. Database Description

The CKD database is retrieved from this study's UCI machine learning repository. This database is acquired from the hospitals closely in 2 monthly intervals. In this database, 24 variety of attributes include, namely age, blood pressure, specific gravity, albumin, sugar, red blood cells, pus cell, pus cell clumps, bacteria, blood glucose random, blood urea, serum creatinine, sodium, potassium, hemoglobin, packed cell volume, white blood cell count, hypertension, diabetes mellitus, coronary artery disease, appetite, pedal edema, and anaemia. Also, it has a single class attribute that represents CKD and not-CKD.

3.2. Pre-processing and Attribute Selection

After collecting the CKD database, a moth flame optimization algorithm performs to choose the most optimal attributes and DBSCAN's parameters for further processing. Figure 2 portrays the moth flame optimization algorithm [14] for selecting both CKD-related attributes and DBSCAN's parameters simultaneously. This optimization relies on the genetic nature of the moth-flighting flames. First, the moth population is created, where DBSCAN's parameters and CKD attributes represent each moth. The fitness value (prediction accuracy) of each moth is determined. If the iteration initiates, the flames and the moths are combined. The best combination is called flames; otherwise, the moths are ordered according

to their fitness values and assigned as flames. Then, the position of each moth is updated. This process is continued until the termination criteria are achieved, i.e., until the maximum number of iterations. Finally, the best solution obtained, such as the most relevant CKD attributes and optimal DBSCAN parameters, i.e., fine-tuned DBSCAN. Once the optimal DBSCAN parameters are obtained, the pre-processed CKD database with the most relevant attributes is given to the unsupervised CNN clustering for outlier removal.

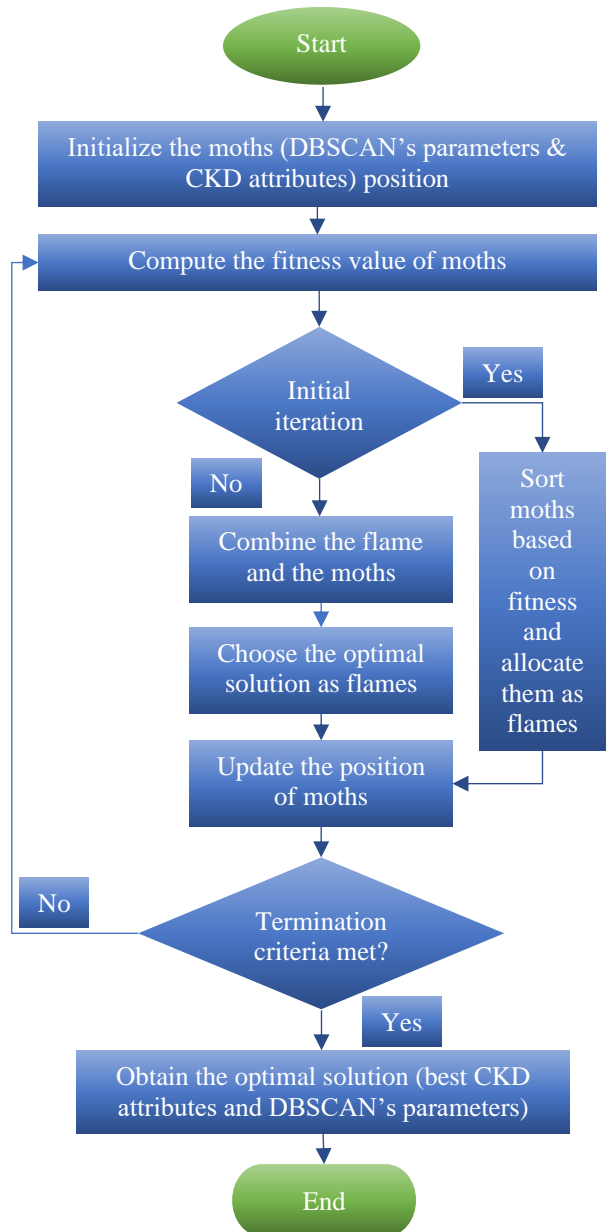


Fig. 2 Flow diagram of moth-flame optimization for CKD attributes and DBSCAN's parameter selection

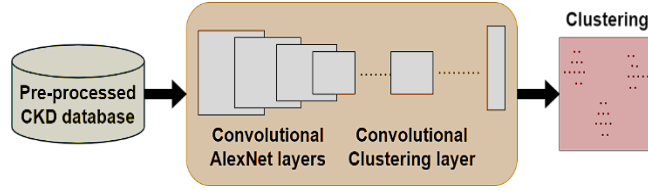


Fig. 3 Block diagram of the deep clustering model

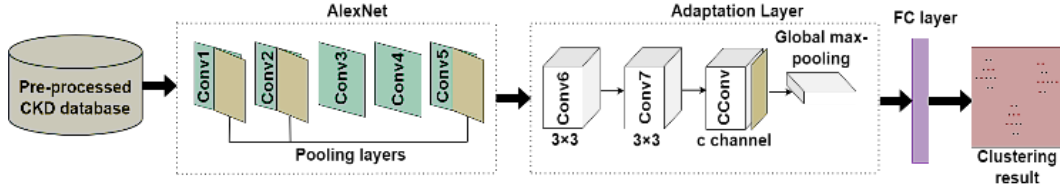


Fig. 4 Presented unsupervised CNN-based network structure for data clustering

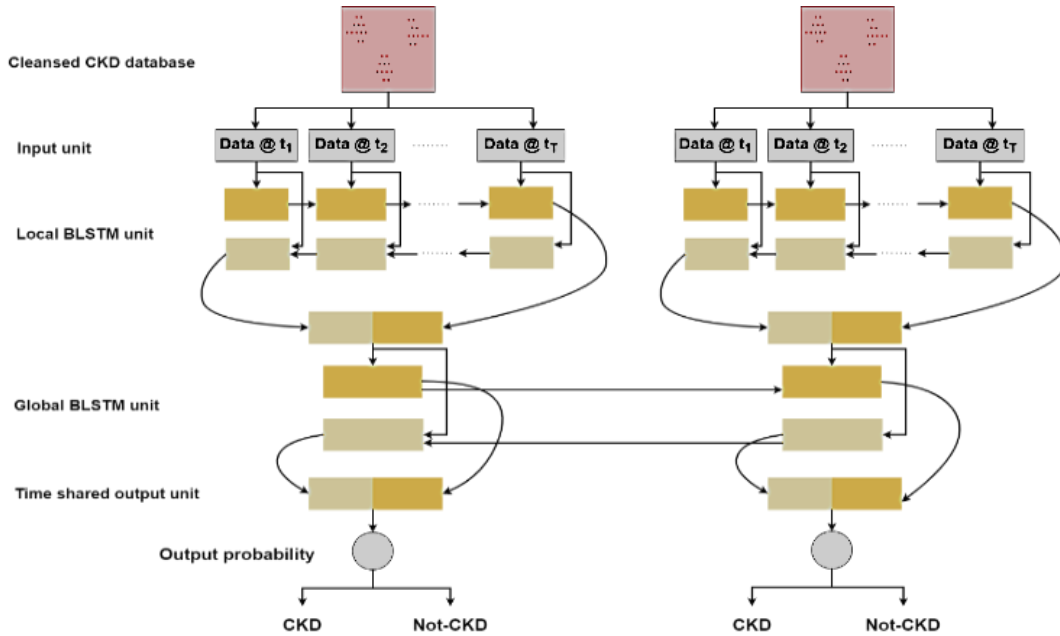


Fig. 5 Structure of HBLSTM

3.3. Unsupervised CNN Clustering based on Fine-tuned DBSCAN Algorithm

To improve the efficiency of the clustering process, unsupervised CNN integrates with the DBSCAN algorithm in this study, which is illustrated in Figure 3. In the learning phase, the clusters are updated by the fine-tuned DBSCAN algorithm. Figure 4 depicts the network structure of this unsupervised CNN, which has multi-convolution with a single clustering layer. In this study, the unsupervised CNN clusters comprise 5 convolutional layers obtained from the primary 5 convolutional layers (Conv1 – Conv5) of AlexNet, followed by 3 adaptation layers (Conv6, Conv7, and CConv) with channel numbers 6144, 2048 and c , correspondingly, that swap the Fully Connected (FC) layer in AlexNet. The adaptation layers engage 2 convolutional layers (Conv6 – Conv7) and single Clustering Convolutional layer (CConv)

with c clusters, each with 3×3 kernels followed by global max-pooling. The highest range for all channels of the CConv is the result of the max-pooling; thus, the dimension $1 \times c$. At last, the FC layer is added to obtain the number of clusters.

If the database comprises n_i instances $D = \{x_1, x_2, \dots, x_{n_i}\}$. The primary goal is to cluster n_i instances into c clusters $C = \{C_1, C_2, \dots, C_c\}$. Consider $H = \{h_1, h_2, \dots, h_{n_i}\}$ is the database without outliers from the FC layer of unsupervised CNN using filters $h_i = f(W_{FC}/D_n)$, where W_{FC} is the group of variables (weights) of the FC layer. The fine-tuned DBSCAN algorithm is utilized to obtain the clusters. During the training phase, the weight W_{FC} and other DBSCAN parameters are adjusted simultaneously by the moth flame optimization algorithm.

3.4. Hierarchical Bi-directional Long Short-Term memory Classification

After clustering the database, the data instances at a specific time $t \in T$ are fed to the HBLSTM, which involves both local and global BLSTM layers. As illustrated in Figure 5, global BLSTM has 4 units: input unit, local BLSTM, global BLSTM, and time-shared output unit. The input unit forwards all attributes among the CKD-fixed window into a 1×24 matrix and then passes it into the local BLSTM unit. The dimension of the CKD-fixed window for global BLSTM is assigned to 7. Local BLSTM aims to analyze the matrices for time-variant attributes and capture the temporal correlations.

As depicted in Figure 6, the local BLSTM has a many-to-single BLSTM model. The local BLSTM for each CKD-fixed window has BLSTM layers; all LSTM layers comprise t memory blocks related to instances at t in all CKD-fixed windows. In this model, local BLSTM only provides the hidden layer state of the final time step in all directions. After that, forward and backward hidden states merge, and the local BLSTM can produce a single hidden state to define all CKD-fixed windows.

The state created by the local BLSTM unit can be passed to the global BLSTM for further analysis. The global BLSTM unit is arranged above the local BLSTM unit (see Figure 7), which has a many-to-many BLSTM model. Each time step incorporates the hidden states generated by the local BLSTM unit and uses the arrays of such hidden states to enhance the training process.

At last, the time-shared output unit provides results for each temporal node in the global BLSTM unit. Every node in the output layer can produce a binary vector to define the probabilities of the given instance being CKD or not-CKD. The output with a higher probability can be chosen as the final prediction, which depends on the hidden state of the corresponding global BLSTM time step. Thus, this HBLSTM can predict CKD and healthy patients according to the 2 probability values.

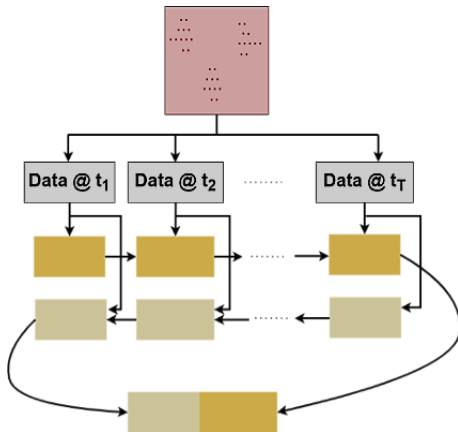


Fig. 6 Structure of local-BLSTM unit

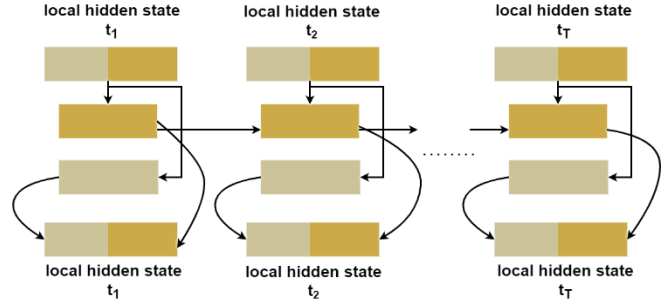


Fig. 7 Structure of global-BLSTM unit

4. Results and Discussion

In this section, the efficacy of the DC-CNN-HBLSTM was assessed by implementing it in Python using the CKD database (discussed in Section 3.1).

Also, a comparison between the presented and existing prediction models is conducted based on the below metrics:

- Accuracy: It is the fraction of precise prediction over the total instances validated.

$$Accuracy = \frac{True\ Positive\ (TP) + True\ Negative\ (TN)}{TP + TN + False\ Positive\ (FP) + False\ Negative\ (FN)} \quad (1)$$

In Eq. (1), TP is the number of instances related to healthy people precisely predicted as themselves, whereas TN is the number of instances related to CKD patients precisely predicted as themselves.

Similarly, FP is the number of instances related to CKD patients imprecisely predicted as healthy, and FN is the number of instances related to healthy people imprecisely predicted as CKD patients.

- Precision: It is determined as:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

- Recall: It is determined as:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

- F-score: It is determined as:

$$F\text{-score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

In this experiment, the DC-CNN-HBLSTM model compares with various existing models such as RF [15], XGBoost [16], integrated model [17], HMANN [26], DBN [20], J48 [21], and DNN [25], which are implemented on the CKD database for CKD prediction.

The comparative analysis of proposed and existing CKD prediction models in terms of various performance metrics is provided in Table 1.

Table 1. Performance analysis of existing and proposed CKD prediction models

| Models | Precision | Recall | F-measure | Accuracy |
|----------------------|--------------|--------------|--------------|--------------|
| | (%) | | | |
| XGBoost | 86.14 | 86.37 | 86.26 | 86.49 |
| J48 | 88.35 | 88.52 | 88.44 | 88.69 |
| RF | 90.10 | 90.44 | 90.27 | 90.57 |
| Integrated model | 91.56 | 91.71 | 91.64 | 91.84 |
| HMANN | 92.45 | 92.63 | 92.54 | 92.90 |
| DNN | 93.50 | 93.78 | 93.64 | 94.00 |
| DBN | 93.80 | 94.22 | 94.01 | 94.40 |
| DC-CNN-HBLSTM | 95.86 | 96.18 | 96.02 | 96.21 |

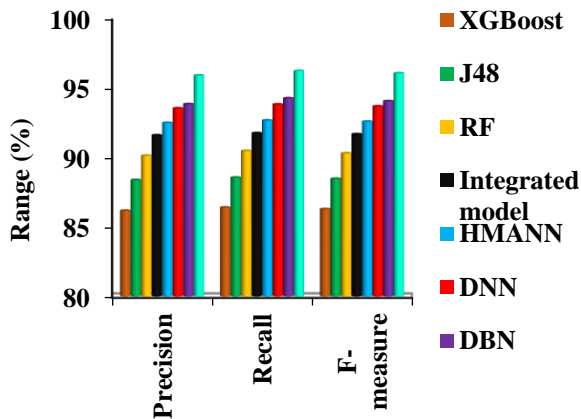


Fig. 8 Precision, recall, and f-measure for various CKD prediction models on the CKD database

Figure 8 illustrates the efficiency of various models developed for CKD prediction.

The precision of the DC-CNN-HBLSTM model is 11.28%, 8.5%, 6.39%, 4.7%, 3.69%, 2.52%, and 2.2% improved than the XGBoost, J48, RF, integrated model, HMANN, DNN, and DBN models, correspondingly.

The recall of the DC-CNN-HBLSTM model is 11.36%, 8.65%, 6.35%, 4.87%, 3.83%, 2.56%, and 2.08% greater than the XGBoost, J48, RF, integrated model, HMANN, DNN, and DBN models, respectively.

Likewise, the f-measure of the DC-CNN-HBLSTM is 11.32%, 8.58%, 6.37%, 4.79%, 3.76%, 2.54%, and 2.14% higher than the XGBoost, J48, RF, integrated model, HMANN, DNN, and DBN, correspondingly.

Figure 9 depicts the accuracy of various models applied to the CKD database.

It declares that the accuracy of the DC-CNN-HBLSTM is 11.24% larger than the XGBoost, 8.48% larger than the J48, 6.23% larger than the RF, 4.76% larger than the integrated model, 3.56% larger than the HMANN, 2.35% larger than the DNN and 1.92% larger than the DBN models.

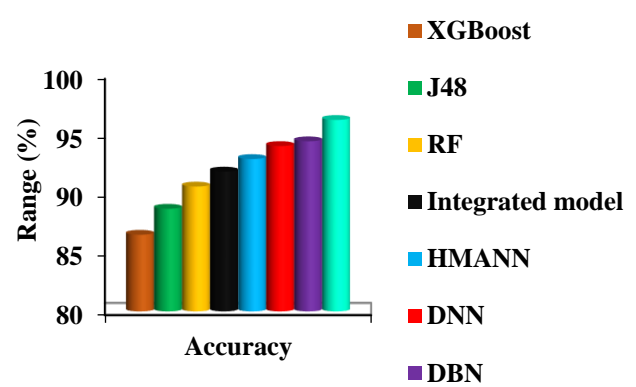


Fig. 9 Accuracy for various CKD prediction models on the CKD database

Thus, it realizes that the DC-CNN-HBLSTM increases the accuracy of predicting CKD patients compared to the other prediction models. It is because of selecting the most relevant attributes and removing the outliers from the database using the deep clustering model. Also, the HBLSTM considers the data attributes associated with each time step to learn the temporal correlation between them and predict the CKD patients efficiently.

5. Conclusion

This paper presents the DC-CNN-HBLSTM model for predicting CKD from time-series data. First, the CKD corpus was acquired and pre-processed. Then, the MF optimization was employed to concurrently choose the best DBSCAN variables and the traits related to CKD. Using the best variables, the DBSCAN clustering algorithm was implemented as an additional layer in the standard CNN model for clustering the data points and removing the outliers. Moreover, the cleansed database was fed to the HBLSTM classifier, which learns the temporal correlation between the local and global attributes for each time step to create a trained model. Later, the trained model was applied to classify the unknown data into healthy and CKD patients. At last, the investigational outcomes proved that the DC-CNN-HBLSTM model realizes an accuracy of 96.21% contrasted with the other prediction models applied to the CKD database.

References

- [1] K. J. Foreman, N. Marquez, A. Dolgert, K. Fukutaki, N. Fullman, M. McGaughey, and C. J. Murray, "Forecasting Life Expectancy, Years of Life Lost, and All-Cause and Cause-Specific Mortality for 250 Causes of Death: Reference and Alternative Scenarios for 2016–40 for 195 Countries And Territories," *The Lancet*, vol. 392, no. 10159, pp. 2052-2090, 2018. Crossref, [https://doi.org/10.1016/S0140-6736\(18\)31694-5](https://doi.org/10.1016/S0140-6736(18)31694-5).
- [2] R. Vanholder, L. Annemans, A. K. Bello, B. Bikbov, D. Gallego, R. T. Gansevoort, and F. Wieringa, "Fighting the Unbearable Lightness of Neglecting Kidney Health: The Decade of the Kidney," *Clinical Kidney Journal*, vol. 14, no. 7, pp. 1719-1730, 2021. Crossref, <https://doi.org/10.1093/ckj/sfab070>
- [3] C. S. D. Moraes Júnior, N. M. D. S. Fernandes, and F. A. B. Colugnati, "Multidisciplinary Treatment for Patients with Chronic Kidney Disease in Pre-Dialysis Minimizes Costs: A Four-Year Retrospective Cohort Analysis," *Brazilian Journal of Nephrology*, vol. 43, pp. 330-339, 2021. Crossref, <https://doi.org/10.1590/2175-8239-JBN-2020-0226>.
- [4] P. Suriyong, C. Ruengorn, C. Shayakul, P. Anantachoti, and P. Kanjanarat, "Prevalence of chronic kidney disease stages 3–5 in low-and middle-income countries in Asia: a systematic review and meta-analysis," *PloS One*, vol. 17, no. 2, pp. 1-18, 2022. Crossref, <https://doi.org/10.1371/journal.pone.0264393>
- [5] S. A. Amshi, A. B. Bababe, M. Saquib, and A. Adamu, A. "Chronic Kidney Disease Associated with Consumption of Vegetables Cultivated on Contaminated Soil in Gashua, Yobe State–Nigeria," *SSRG International Journal of Pharmacy and Biomedical Engineering*, vol. 7, no. 1, pp. 1-5, 2020. Crossref, <https://doi.org/10.14445/23942576/IJPBE-V7I1P101>
- [6] M. Evans, R. D. Lewis, A. R. Morgan, M. B. Whyte, W. Hanif, S. C. Bain, and W. D. Strain, "A Narrative Review of Chronic Kidney Disease in Clinical Practice: Current Challenges and Future Perspectives," *Advances in Therapy*, vol. 39, no. 1, pp. 33-43, 2022.
- [7] A. C. Simoes e Silva, A. S. Miranda, N. P. Rocha, and A. L. Teixeira, "Neuropsychiatric Disorders in Chronic Kidney Disease," *Frontiers in Pharmacol*, vol. 10, pp. 1-11, 2019. Crossref, <https://doi.org/10.3389/fphar.2019.00932>.
- [8] C. J. Sperati, S. Soman, V. Agrawal, Y. Liu, K. Abdel-Kader, C. J. Diamantidis, M.M. Estrella, K. Cavanaugh, L. Plantinga, J. Schell, J. Simon, J.A. Vassalotti, M.J. Choi, B.G. Jaar and R.C. Greer, "Primary Care Physicians' Perceptions of Barriers and Facilitators to Management of Chronic Kidney Disease: A Mixed Methods Study," *PloS One*, vol. 14, no. 8, pp. 1-19, 2019. Crossref, <https://doi.org/10.1371/journal.pone.0221325>
- [9] S. Shanmitha, N. Rathina Moorthy, and S. Suruthi, "Cancer Detection Technique in Data Mining," *SSRG International Journal of Computer Science and Engineering*, Special Issue ICETSST, pp. 40-42, 2018
- [10] E. H. A. Rady, and A. S. Anwar, "Prediction of Kidney Disease Stages Using Data Mining Algorithms," *Informatics in Medicine Unlocked*, vol. 15, pp. 1-7, 2019. Crossref, <https://doi.org/10.1016/j.imu.2019.100178>
- [11] H. S. Shad, Z. Jamal, S. M. Ahmed, S. Momen, and N. Mansoor, "Early Prediction of Chronic Kidney Disease using Data Mining Techniques," in *Proceedings of Computational Methods in Systems and Software*, Springer, Cham, pp. 947-957, 2021. Crossref, https://doi.org/10.1007/978-3-030-90321-3_79
- [12] G. M. Ifraz, M. H. Rashid, T. Tazin, S. Bourouis, and M. M. Khan, "Comparative Analysis for Prediction of Kidney Disease Using Intelligent Machine Learning Methods," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1-10, 2021. Crossref, <https://doi.org/10.1155/2021/6141470>
- [13] Y. Zhang, "DBSCAN Clustering Algorithm Based on Big Data is Applied in Network Information Security Detection," *Security and Communication Networks*, vol. 2022, pp. 1-8, 2022. Crossref, <https://doi.org/10.1155/2022/9951609>
- [14] P. Usha, and N. Kavitha, "Optimized Uncertainty Handling Feature Selection using Moth Flame Based DBSCAN Clustering for Diabetic Related Chronic Kidney," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 10, pp. 3328-3336, 2022.
- [15] I. A. Pasadana, D. Hartama, M. Zarlis, A. S. Sianipar, A. Munandar, S. Baeha, and A. R. M. Alam "Chronic Kidney Disease Prediction by using Different Decision Tree Techniques," in *Journal of Physics: Conference Series, IOP Publishing*, vol. 1255, no. 1. pp. 1-6, 2019. Crossref, <https://doi.org/10.1088/1742-6596/1255/1/012024>
- [16] A. Ogunleye, and Q. G. Wang, "XGBoost Model for Chronic Kidney Disease Diagnosis," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 6, pp. 2131-2140, 2019. Crossref, <https://doi.org/10.1109/TCBB.2019.2911071>
- [17] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, "A Machine Learning Methodology for Diagnosing Chronic Kidney Disease," *IEEE Access*, vol. 8, pp. 20991-21002, 2019. Crossref, <https://doi.org/10.1109/ACCESS.2019.2963053>
- [18] A. Khamparia, G. Saini, B. Pandey, S. Tiwari, D. Gupta, and A. Khanna, "KDSAE: Chronic Kidney Disease Classification with Multimedia Data Learning Using Deep Stacked Autoencoder Network," *Multimedia Tools Application*, vol. 79, no. 47, pp. 35425-35440, 2020. Crossref, <https://doi.org/10.1007/s11042-019-07839-z>
- [19] M. C. Shanker, M. Vadivel, "Hybrid Transfer Learning of Mammogram Images for Screening of Micro-Calcifications," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 9, no. 8, pp. 40-47, 2022. Crossref, <https://doi.org/10.14445/23488379/IJEEE-V9I8P105>
- [20] S. M. M. Elkholy, A. Rezk, and A. A. E. F. Saleh, "Early Prediction of Chronic Kidney Disease Using Deep Belief Network," *IEEE Access*, vol. 9, pp. 135542-135549, 2021. Crossref, <https://doi.org/10.1109/ACCESS.2021.3114306>

- [21] H. Ilyas, S. Ali, M. Ponum, O. Hasan, M. T. Mahmood, M. Iftikhar, and M. H. Malik, "Chronic Kidney Disease Diagnosis using Decision Tree Algorithms," *BMC Nephrology*, vol. 22, no. 1, pp. 1-11, 2021. Crossref, <https://doi.org/10.1186/s12882-021-02474-z>
- [22] H. Peng, H. Zhu, C. W. A. Jeong, T. Tao, T. Y. Tsai, and Z. Liu, "A Two-Stage Neural Network Prediction of Chronic Kidney Disease," *IET Systems Biology*, vol. 15, no. 5, pp. 163-171, 2021. Crossref, <https://doi.org/10.1049/syb2.12031>
- [23] M. Rashed-Al-Mahfuz, A. Haque, A. Azad, S. A. Alyami, J. M. Quinn, and M. A. Moni, "Clinically Applicable Machine Learning Approaches to Identify Attributes of Chronic Kidney Disease (CKD) for Use in Low-Cost Diagnostic Screening," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 9, pp. 1-11, 2021. Crossref, <https://doi.org/10.1109/JTEHM.2021.3073629>.
- [24] L. Antony, S. Azam, E. Ignatious, R. Quadir, A. R. Beeravolu, M. Jonkman, and F. De Boer, "A Comprehensive Unsupervised Framework for Chronic Kidney Disease Prediction," *IEEE Access*, vol. 9, pp. 126481-126501, 2021. Crossref, <https://doi.org/10.1109/ACCESS.2021.3109168>.
- [25] V. Singh, V. K. Asari, and R. Rajasekaran, "A Deep Neural Network for Early Detection and Prediction of Chronic Kidney Disease," *Diagnostic*, vol. 12, no.1, pp. 1-22, 2022. Crossref, <https://doi.org/10.3390/diagnostics12010116>
- [26] F. Ma, T. Sun, L. Liu, and H. Jing, "Detection and Diagnosis of Chronic Kidney Disease Using Deep Learning-Based Heterogeneous Modified Artificial Neural Network," *Future Generation Computer Systems*, vol. 111, pp. 17-26, 2020. Crossref, <https://doi.org/10.1016/j.future.2020.04.036>