International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Neural Nephroinformatics: Ensemble Strategies in Deep Learning for CKD Detection

Jeena Jose [*1], S. Sheeja[2]

Submitted: 25/01/2024 **Revised**: 03/03/2024 **Accepted**: 11/03/2024

Abstract: A burden on global health, chronic kidney disease (CKD) affects about 10% of adult population worldwide. It is acknowledged as one among the top 20 global causes of death. Although there exists treatment for chronic kidney disease, early detection of the illness can help mitigate the damage and slow down the disease's progression. Consequently, to improve the accuracy and effectiveness of the conventional chronic kidney disease diagnosis system, contemporary computer-aided approaches must be used. In the suggested ensemble model, Support Vector Machine (SVM) used as the stacking ensemble model, which mixes two hybrid deep learning models. The dataset of CSV files used to detect CKD was obtained from the Kaggle repository in order to validate our model. Convolutional neural network- Gated Recurrent Unit (CNN-GRU) and Convolutional neural network-Long short-term memory (CNN-LSTM) are the two hybrid models employed in the model. With a high accuracy of 98.95%, the model produced generally accurate predictions. Recall and precision ratings of 98.56 and 100 respectively, show how accurate the classifications model can be. The suggested stacking ensemble model was contrasted with both our own implementation and alternative methods for detecting CKD. The proposed approach outperforms other existing techniques in terms of performance, while utilizing the model to prevent overfitting.

Keywords: *Creatinine, Convolutional neural network, chronic kidney disease, long short-term memory, Cardiovascular disease, Gated Recurrent Unit.*

1. Introduction

"Health is the greatest of all possessions; a pale cobbler is better than a sick king" is a proverb stated by Mr. Isaac Bickerstaff. Health is wealth; it is the most valuable thing. Various illnesses can have an impact on human health. Long-lasting illnesses known as chronic diseases are regarded as the primary global cause of disability and mortality. Globally, the prevalence of these illnesses is increasing, spreading to every area and impacting all socioeconomic groups. According to the World Health Report, chronic illnesses account for more than 79% of fatalities in poor nations. There is a strong correlation between cardiovascular disease [1] (CVD), diabetes mellitus [2], and CKD. Consequently, these three illnesses are increasingly prevalent risk factors among people in the modern period. The results of the study indicate that at least 50% of people with heart disease will eventually experience CKD [3]. The prognosis of these conditions will be negatively impacted by the diagnosis of CKD, which will also increase the morbidity of CVD. It is also evident that a patient has a very significant chance of getting CKD if their CDM remains undetected [4].

Figure 1 provides a visual representation of the internal structure of the kidney. The kidneys are two bean-shaped organs that are located behind the belly and beneath the ribs on either side of the spine [5]. They have a role in controlling blood pressure, maintaining electrolyte balance, and helping the body produce red blood cells. The two renal arteries supply blood to the kidney, which filters it before it leaves the kidney through the two renal veins. Urine excretion is transported to the bladder by the urethra, a tube. Because the liver is larger on the right side of the body, the left kidney is situated slightly higher than the right kidney [6].

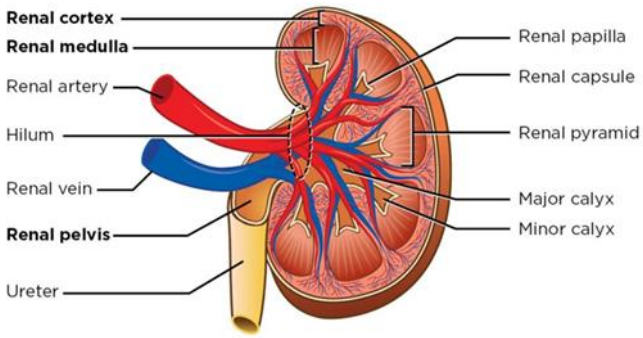

Fig. 1. Internal Structure of Kidney

Because of the insulin resistance in diabetic patient's bodies, they are more likely to develop high levels of bad cholesterol. Diabetes patients are prone to elevated levels of bad cholesterol due to insulin resistance in their body. As a result, there is a rise in bad cholesterol and a fall in good cholesterol, increasing the risk of cardiovascular disease [7]. If the CDM is not stopped right away, the blood's increased glucose concentration damages the kidney's blood vessels, ultimately leading to renal failure.

[1] *Department of Computer Science and Engineering, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu*
[2] *Department of Computer Science and Engineering, Karpagam Academy of Higher Education Coimbatore, Tamil Nadu*
** jeena4762@gmail.com*

A substantial correlation has been seen between CVD and CKD [8]. Kidney failure is the end result of blockage and a reduced flow of oxygenated blood to the kidney [9]. Kidney may become more vulnerable to heart disease as a result of the heart having to work harder to pump blood to the damaged kidney. Figure 2 illustrates the increasing link that exists between CDM, CKD, and CVD.

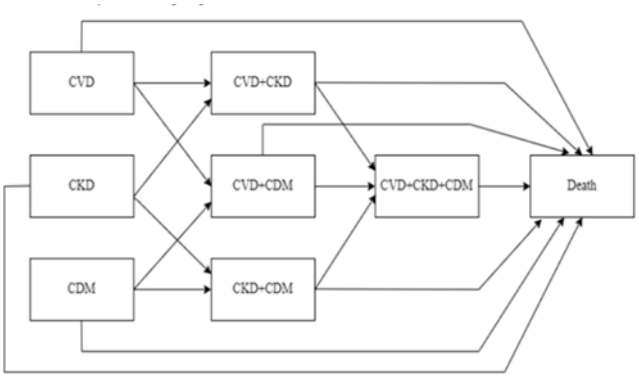

Fig. 2. Relation between Chronic Diseases

Chronic kidney disease (CKD) is an asymptotic chronic illness in which the kidneys' capacity to operate steadily declines over time. For kidney to function properly, the blood must include a healthy balance of salts and minerals, including potassium, phosphorus, and sodium [10]. Large volumes of healthcare data are generated from healthcare applications through variable technologies like computers, smart health devices, and embedded systems in this digital age because of the quick advancements in science and technology. By 2025, the World Health Organization (WHO) predicts that 73% of deaths would be related to chronic illnesses. Thus, the conditions are referred to as the Global Burden of Disease [11].

One of the biggest issues facing the health sector is identifying chronic diseases and determining their severity. Even with the tremendous advancements in medical science, it is still difficult to identify illnesses early and treat them promptly. The complexity of diagnostic techniques, their growing expense, and the lack of necessary resources in certain places, especially isolated areas, make it imperative to create faster and more effective approaches. Creating a diagnosis model could help doctors learn more about the risk factors associated with chronic illnesses and group patients according to their own health patterns for more personalized care. When working with high-dimensional medical data, many methods are inefficient due to several issues. Because of this, only a small number of input features that have an effect on the algorithms' predictive outcome should be kept; the rest should be eliminated. Sufficient data pre-processing techniques are needed to address the heterogeneity problem. The following is a brief summary of the paper's main contributions:

- Implemented an ensemble model that combines the strengths of CNN [12] with LSTM [13] and Gated Recurrent Unit [14] architectures to effectively capture both spatial and temporal features in medical data related to CKD.
- To enhance the performance, SVM [15] with stacking framework is integrated in the model.
- Assessed the performance of the model with existing CKD detection approaches.

The remaining sections of the document are organized as follows: In section 2 the literature review and research gaps are provided. The methodology is highlighted in section 3. The detailed results of the proposed method and the comparative study with existing research work are presented in section 4. Finally, the conclusion and future scopes are discussed in section 5.

2. Background of the Study

Machine learning (ML) techniques in the healthcare industry has advanced significantly in recent years, with an emphasis on early detection. This improvement has also been shown in the field of CKD, where a number of notable studies have advanced our understanding of the disease. By carefully going over the pertinent articles, we offer a thorough overview of the state of CKD research as it is in this literature review. We have thoroughly examined the methods used, the results, and the limitations found in each study as part of our analysis. By doing this, we hope to provide a thorough and objective knowledge of the advancements and difficulties in CKD research.

Integration of feature selection techniques and ML classification algorithms were suggested by Abdel-Fattah et al. [16] based on big data platforms. The condition known as chronic kidney disease (CKD) is becoming quite common among adults. Medical scientists can greatly benefit from the use of ML algorithms in precisely diagnosing diseases at the early stages of their development. Healthcare is enhanced by the integration of algorithms with big data platforms. Relief-F and chi-squared are used as the feature selection method along with six ML classification techniques. A ML model for CKD prediction was presented by Arif et al. [17]. Authors utilized sequential data scaling, z-standardization, min-max scaling, robust scaling, along with iterative imputation for missing values to overcome difficulties with medical datasets. The Boruta method is used for feature selection, while ML algorithms are used to create the model. Study used UCI CKD dataset and the primary drawback of the study was its dependence on a single dataset, which has a significant number of missing values. The missing data was estimated by the authors using iterative imputation; nonetheless, it is important to recognize that imputation

techniques introduce uncertainty, which may affect the prediction power of the model.

Venkatesan et al.'s [18] ML techniques let experts in research of CKD preventive methods through earlier detection. With a publicly accessible dataset, authors use ML techniques to predict and categorize CKD. CKD dataset with 400 cases were collected from the Irvine ML Repository, a publicly accessible dataset. Extreme Gradient Boosting (XGBoost) has been used to compare the performance of six ML techniques as base learners. Every ML algorithm is assessed based on various performance metrics.

Srivastava et al. [19] concentrated on accurate disease prediction by utilizing ML and data mining to give decisionmakers vital data-driven insights. The fundamental properties of the dataset determine the classification systems' accuracy, which is why feature selection strategies are used to simplify algorithm models and maximize classification precision. There are five different classification schemes used to identify chronic renal illness. The goal of Silveira et al.'s work [20] is to aid in the early prediction of CKD by addressing issues with imbalanced and small-scale datasets, using three ML methods. The following characteristics were gathered by the authors from Brazilians' medical records, regardless of whether they were diagnosed with CKD: age, gender, creatinine, urea, albuminuria, and diabetes mellitus. The technique of oversampling is based on both automated and manual augmentation.

Singh et al. [21 build a deep neural network and assess how well it performs in comparison to other modern ML methods. In order to address this, they put up a deep learning model for CKD early detection and prediction. During testing, all missing values in the database were replaced with the average of the related features. After setting the parameters and completing several trials, the neural network's ideal parameters were determined. Recursive Feature Elimination was used to choose the most crucial features (RFE). The limitation of the model was that it had only been evaluated on limited data sets, which limited the generalizability of the method.

The main goal of Prakash et al. [22] is to present and advance predictive algorithms for CKD prediction. The authors proposed for the efficient administration of huge dataset samples. Random forests and ensemble nonlinear SVM are used to formulate a binary classification issue. Thus, in prototype examples, the work generates nonlinear combinations of kernel activations instead of employing the standard linear combination of activations. The study's main drawback is that because it depends so heavily on a single dataset, it cannot be applied in real time.

AdaBoost classifier and an information-gain-based feature selection technique are used in Ebiaredoh-Mienye et al.'s [23] suggested approach to efficiently identify CKD. The University of California, Irvine (UCI) ML repository has made the dataset freely accessible. The study's main flaw is that it depends too heavily on the scant data. For the purpose of diagnosing CKD, Rama et al. [24] suggested a ML methodology. The study used CKD dataset from UCI, which contains missing values. Six algorithms are typically employed to create the models. There are just 400 data samples total that were used in the study, making them relatively tiny. As a result, the model's capacity for generalization may be restricted. The hybrid approach for diagnosing chronic renal disease was proposed by Khalid et al. [25]. The study makes use of the CKD dataset from the UCI repository.

While previous research has shown that integrating different neural network architectures can be effective, more thorough studies are required to examine the possible advantages of integrating not only different deep learning models but also taking non-neural network techniques into account. Furthermore, there is little investigation of interpretability and explainability features in the context of CKD detection, and there is disagreement on the best combination tactics and ensemble sizes. By filling in these research gaps, ensemble models for chronic kidney disease diagnosis could become more reliable and accurate, and they could also be easier to understand.

3. Materials and methods

Deep stacking ensemble modelling is used in the suggested CKD detection model. Two hybrid deep learning models are combined in the suggested model, and the SVM serves as the meta learner. The illustration in Figure 3 visually depicts the representation of the proposed model. The dataset for CKD is gathered via Kaggle. Preprocessing and data augmentation techniques are applied to the dataset. CNN-LSTM, CNN-GRU are the hybrid model utilized in the approach. The convolutional layer, fully connected layer, flatten layer, LSTM layer, max-pooling layer, and an output layer make up the CNN-LSTM model. The convolutional layer, GRU layer, fully connected layer, max-pooling layer, flatten layer, and an output layer make up the CNN-GRU architecture.

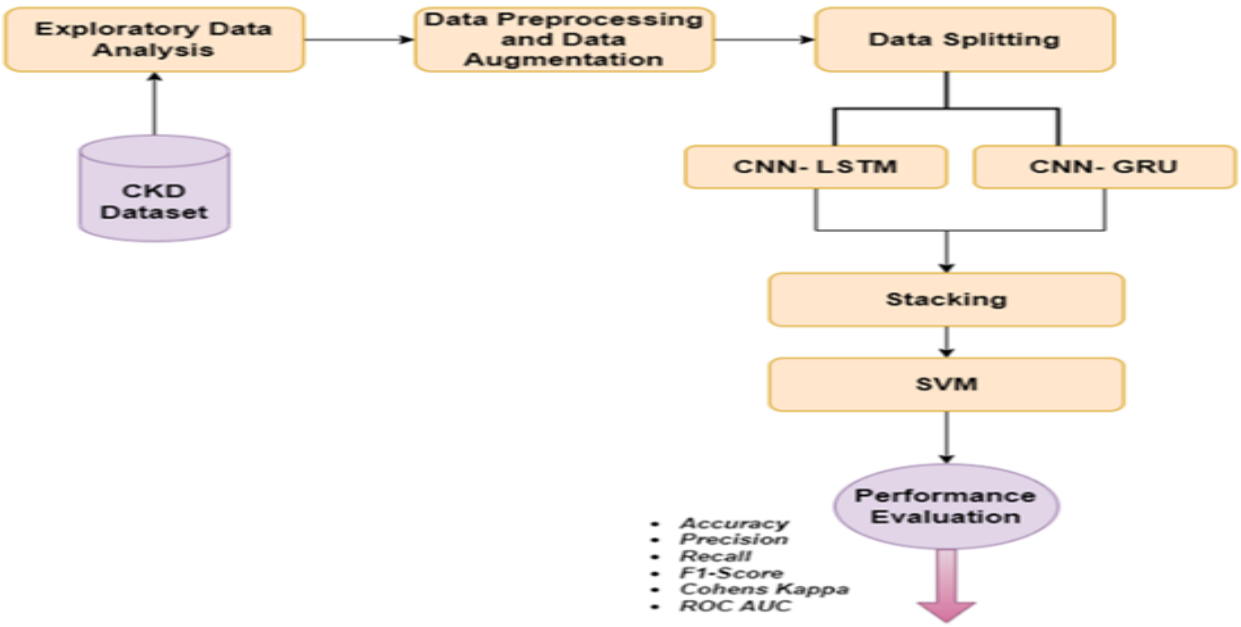

Fig. 3. Block diagram of the proposed model

3.1. Dataset Description

The CSV file dataset used for CKD detection was obtained from the Kaggle repository to validate the model. The depiction of sample data is presented in Figure 4. The dataset is a tabular data structure in two dimensions that supports row and column data storage. There are 400 samples in the collection, and each sample has 37 distinct properties. Essential clinical characteristics, test results, and demographic data on patients with CKD are included in the dataset to help identify patterns and temporal dependencies in the data.

3.2. Data Preprocessing and Augmentation

Data preprocessing is a crucial step in preparing datasets for Before the dataset is executed by the algorithm, it is pre-processed to look for missing values and other irregularities. Preprocessing data can also help to cut down on the amount of time needed to train the model. We can minimize the quantity of data that the algorithm must process by eliminating duplicated or unnecessary data, which can significantly cut down on the time and resources needed to train the model. Overfitting can potentially be avoided by preprocessing the data. Model overfitting is occurred when it performs well on training data but poorly on, untrained data. The interpretability of the model can also be enhanced by preprocessing the data. Understanding the links between various variables and how they affect the model's predictions will be made simpler by cleaning and formatting the data.

The term statistical analysis refers to the procedure of collecting and analysing data. It is a method for applying numerical analysis to remove bias from the evaluation of data. Massive data sets should be gathered and analysed in order to turn common patterns and trends into insightful knowledge. In data analysis, measurements like mean, standard deviation, and others are commonly employed.

Exploratory data analysis can make use of visual data analysis techniques. Data visualization refers to data analysis done with graphs or maps. Figure 4 provides visualization of dataset. A score of 1 denotes a patient with CKD, while a value of 0 denotes no CKD.

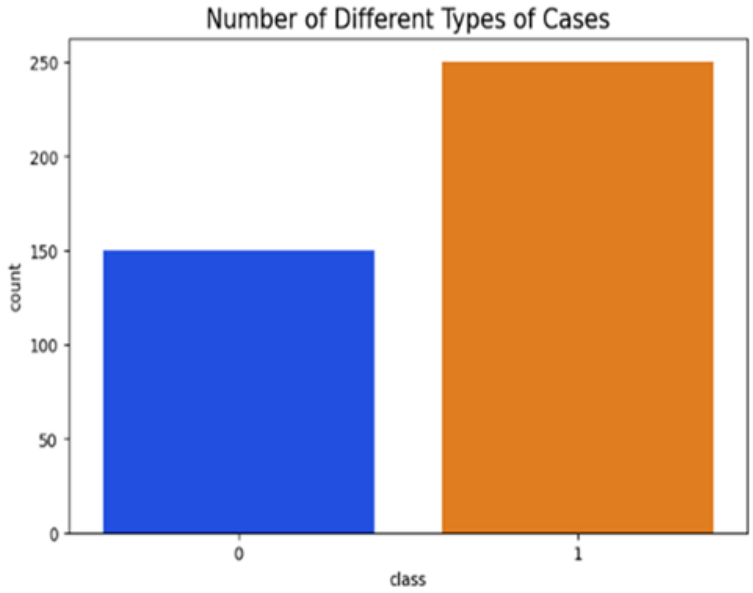

Fig. 4. Data Visualization

High correlation features have approximately the same influence on the dependent variable since they are more linearly dependent. Figure 5 illustrates the correlation of features with CKD. Thus, we can exclude one of the two traits when the other has a significant correlation. Feature selection attempts to eliminate uninformative characteristics in order to simplify models since highly correlated features provide redundant information. It can determine redundant features and choose a small number of crucial attributes that most accurately describe the target variable by looking at correlations.

.

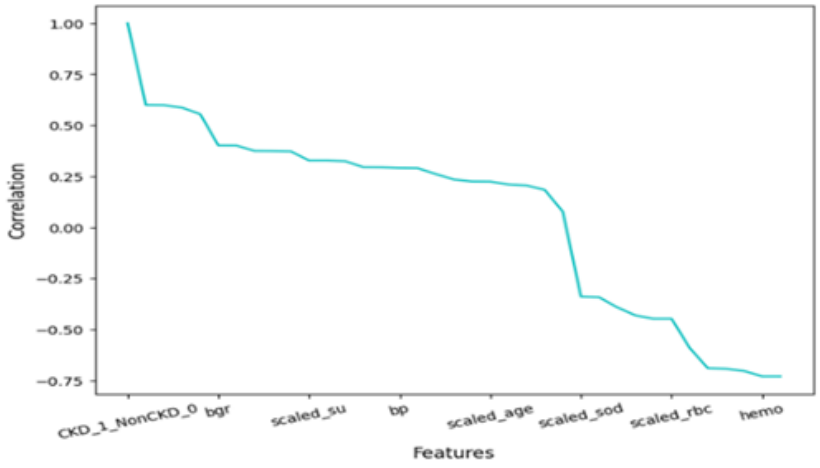

Fig. 5. Correlation of Features with CKD

Figure 6 's bar plot of the dataset illustrates how each attribute correlates with the goal variable, "CKD_1_NonCKD_0." The relationship between every feature and the goal variable will be shown on the plot.

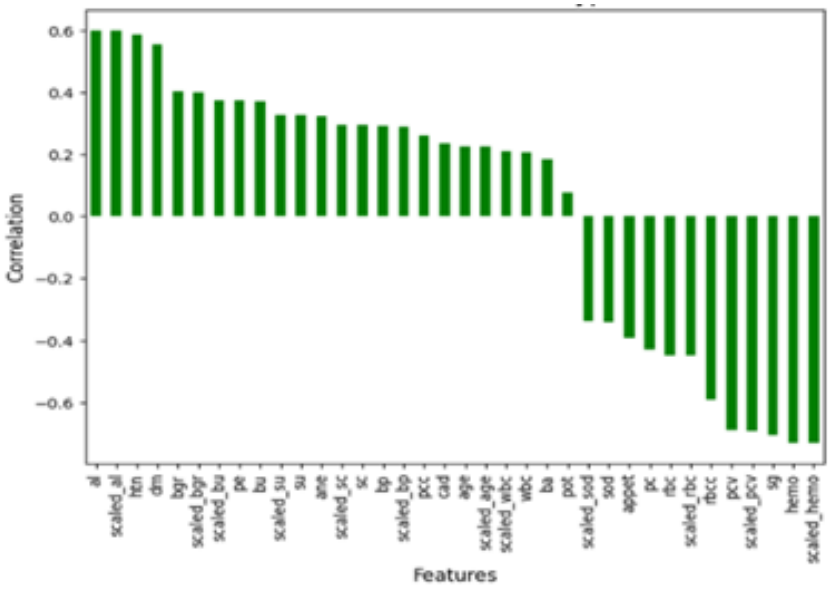

Fig. 6. Correlation with Case type

The pattern of the medical variables produced by categorization was shown using a heatmap in Figure 7. The heatmap was a graphical display of data that showed clustering on both rows and columns along with colour grids containing individual values in a matrix. A collection of patients was grouped using clustering based on the results of their health examinations. Using hierarchical clustering technique, participants were separated into many clusters, and the biomarker patterns for each cluster were displayed in the heatmap's centre as coloured patterns.

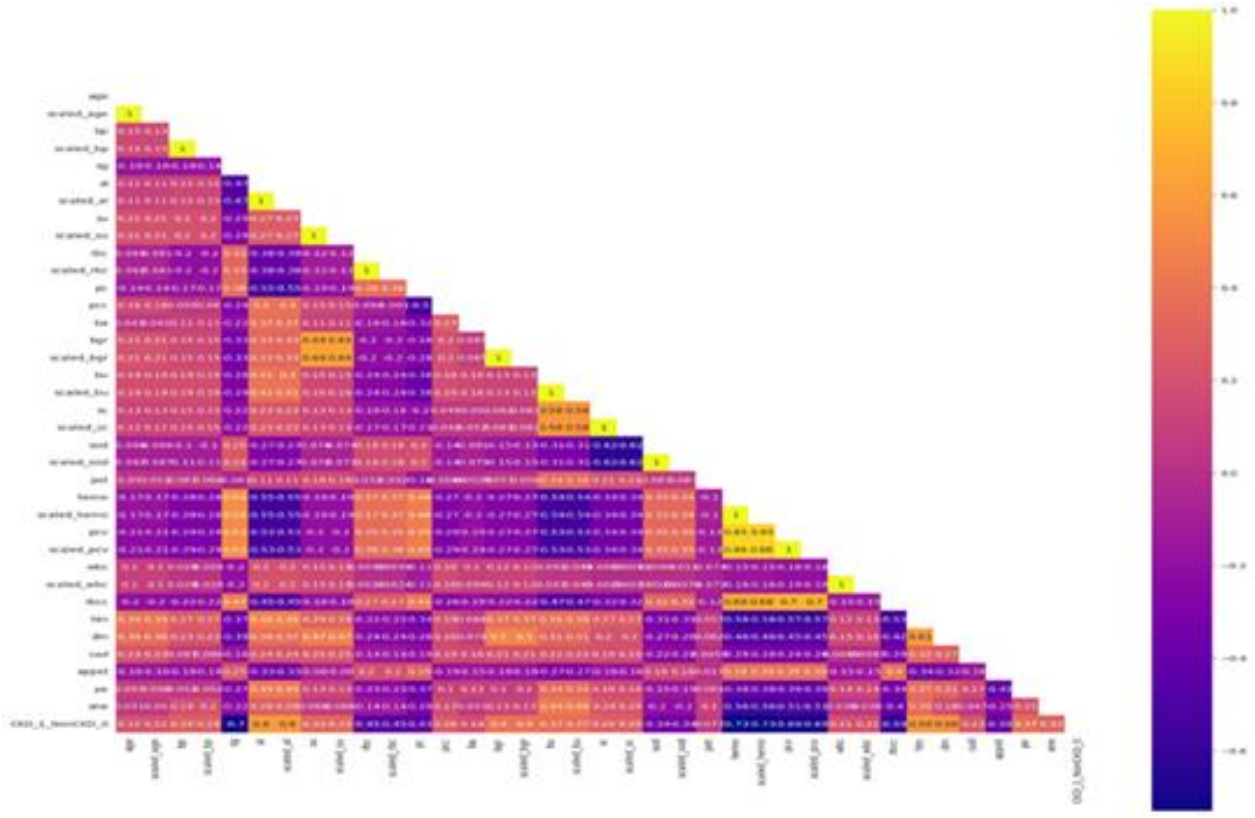

Fig. 7. Heatmap Visualization of Data

3.3. CNN- LSTM Model

CNN is frequently employed in feature engineering because of its propensity to focus on the most noticeable elements in the field of view. For time series analysis LSTM is the best option, due to its ability to expand based on time sequence. A CNN-LSTM based model is built for CKD detection based on the features of CNN and LSTM. Figure 8 illustrates the architecture of the basic LSTM model.

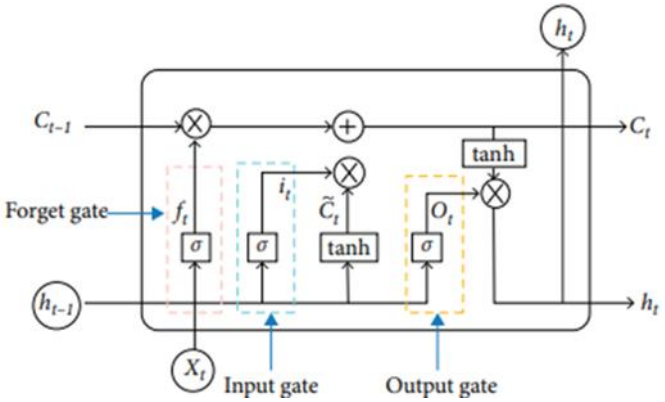

Fig. 8. Basic CNN Architecture

CNN consist of layers of interconnected neurons with learnable parameters, including convolutional layers that apply convolution operations to input data, pooling layers that down sample spatial dimensions, and fully connected layers for high-level feature integration. CNNs excel at capturing hierarchical and spatial dependencies in data, making them well-suited for tasks like image recognition and classification. The convolutional layers enable the network to automatically learn spatial hierarchies of features, allowing it to recognize patterns and representations at various scales. Overall, CNNs have become instrumental in computer vision applications, demonstrating significant success in tasks ranging from image classification to object detection and segmentation. Eq (1) displays the computation formula for each convolution layer, which has several convolution kernels.

$$LT_t = \tanh(XT_t * KT_t + BT_t) \quad (1)$$

where tanh is the activation function, $\boldsymbol{XT_t}$ is the input vector, $\boldsymbol{KT_t}$ is the convolution kernel's weight, and $\boldsymbol{BT_t}$ is the convolution kernel's bias. The output value after convolution is represented by $\boldsymbol{LT_t}$. Four of the LSTM modules function in a unique interactive way and resemble the typical RNN modules. An LSTM memory cell is composed of input gate, forget gate, and output gate. The forget gate's output value can be calculated using (2)

$$FT_t = \sigma(w_f.[HT_{t-1}, XT_t] + BT_f \quad (2)$$

For the current time XT_t is the input value, HT_{t-1} denotes the output value at the most recent moment, w_f is the weight of forget gate, and BT_f is its bias. The value range of FT_t is therefore (0,1).

Cell state and the output value are as follows

$$IT_t = \sigma(w_i.[HT_{t-1}, XT_t] + BT_i \quad (3)$$

$$\widetilde{ct_t} = \tanh(wt_c.[HT_{t-1}, XT_t] + BT_c \quad (4)$$

where the input gate's weight w_i, bias BT_i and input gate's bias BT_c are all expressed in terms of the (0,1) value range of IT_t.

Current state of the cell is as follows

$$c_t = FT_t{*}c_{t-1}, I_{Tt} * \widetilde{c_t} \quad (5)$$

The value of c_t is between (0,1), OT_t from the output gate is obtained as follows: At time t, HT_{t-1} and X_{Tt} are received as input values of the output gate.

$$OT_t = \sigma(w_0.[HT_{t-1}, XT_t] + BT_0 \quad (6)$$

where w_0 denotes the weight output gate and BT_0 its bias, OT_t is between (0,1).

$$HT_t = OT_t * \tanh(c_t) \quad (7)$$

Equation (7) is employed for computing the LSTM output, providing a mathematical expression to determine the resulting value based on the specified inputs and parameters.

3.4. CNN- GRU Model

GRU networks have the ability to selectively retain and forget past inputs, which allows them to handle long-term dependencies in sequential data. Compared to the LSTM, the GRU is simpler because it only contains two gates: reset and updated gates. This gate is used to decide whether or not the data is useful. The following formula determines the hidden layer output: Output of forget gate can be calculated using (8).

$$H_t^i = (1 - Z_t) * H_{t-1}^i + Z_t * H_t^{-i} \quad (8)$$

where Z_t represents an update gate, and H_{t-1}^i indicates the memory data. Z_t controls how many data of the preceding and present memory would be added.

The data value H_t^{-i} is estimated as

$$H_t^{-i} = \tanh(w.[R_t * H_{t-1}^i, H_{t-1}^i]) \quad (9)$$

in which R_t determined is the GRU reset gate that effectively reset the data in the memory.

$$R_t = \text{sigmoid}\,(w_r.[H_{t-1}^i, H_t^{i-1} \quad (10)$$

In order to transform the classification process into a sequential task and improve the accuracy of CNN detection, the CNN-FC layer is replaced by the GRU network. The classification results of each feature map are then included in the classification of the next feature map in the same hidden layer. The CNN-GRU architecture does not modify the parameters of the convolution layers with sizes, pooling, or the original CNN input in order to do feature extraction. The pth convolution layer's output feature maps are all computed using

$$Con\_Output_j^i = sigmoid\left(\sum_{p=1}^{m} con\left(A_p^{i-1} * k_{pq}^i\right) + \boldsymbol{B}_q^i\right) \quad (11)$$

Subsequently, every feature across all feature maps in the final CNN pooling layers is linked to an appropriate GRU technique. This suggests that there are twelve GRU networks in total. Each GRU network consists of three layers: ten layers for the output neuron, twenty-five layers for the input neuron, and fifty layers for the hidden neuron. Finally, the softmax function found in the equation to classify the CKD is used to activate the GRU output. Each GRU output has been selected during the testing phase in order to determine the ultimate detection result.

$$Y_r = \frac{ex(W_r H)}{\sum_r ex(W_r.H)} \quad (12)$$

The model architecture is depicted in Figure 9.CNN-LSTM, CNN-GRU are the hybrid models employed in this ensemble approach. Ultimately, stacking is used to train the SVM, and its performance is assessed.

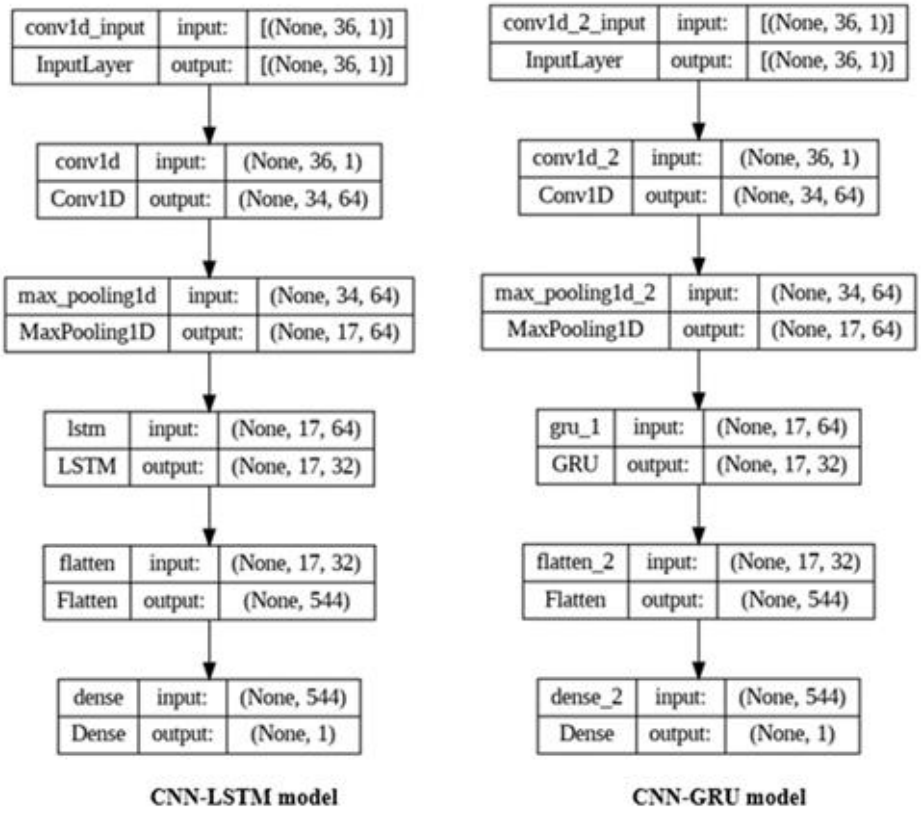

Fig. 9. Proposed Hybrid Model Architectures

3.5. Proposed Ensemble Approach

A stacking ensemble model serves as the foundation for the suggested CKD detection model. SVM is used as the meta-learner and combines two hybrid deep learning models. Stacking is an effective framework that combines the predictions of several base learning algorithms in the most practical way using a meta-learning algorithm. There are two layers in the stacking framework. Base models are found in layer 0 and a meta-learning process is found in layer 1. The original data set is first divided into datasets

for testing and training. Each of the base models goes through this process once more. The SVM model is trained using the last new feature training dataset, and predictions are made using the trained model.

The selection of the base learning models affects stacking performance. CNN-LSTM and CNN-GRU were selected as the basis learning models for this investigation. SVM is used to classify the incorrectly categorized samples, creating fresh feature training and test data sets. It is possible to forecast the new feature test data when SVM trained on fresh feature training data discovers new feature patterns. Stated otherwise, the combined strength of the basic learners equals the resultant accuracy.

3.6. Model Optimization and Hyperparameters

TensorFlow and Python were used to implement the suggested models following the preparation of the dataset. Deep neural network hyperparameters are empirically set and have a significant impact on learning. As a result, a large range of values are tested to identify the most optimized model in terms of the offered classification performance.

Table 1. Hyperparameters

Parameters	CNN- LSTM Model	CNN- GRU Model
Optimizer	Adam	Adam
Loss function	Binary Crossentropy	Binary Crossentropy
Batch size	16	16
Number of Epochs	10	10

3.7. Simulation Setup

Utilizing an Intel Core i7-6850K 3.60 GHz 12-core processor and a NVIDIA GeForce GTX 1080 Ti GPU with 11GB VRAM, the system operates on the Google Collaboratory and Microsoft Windows 10 environment. Keras, an accessible open-source library, is employed for constructing efficient neural network models that leverage TensorFlow, specifically tailored for swift computations in deep learning, serving as an ideal framework for chronic kidney disease (CKD) detection.

4. Result and Discussion

4.1. Performance Evaluation

Table 2 displays the metrics employed in this study to assess the effectiveness and performance of the proposed architecture. These performance parameters provide different perspectives on the model's effectiveness in detecting CKD.

Table 2. Performance parameters

Parameters	Equation
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$
Precision	$(TP)/(TP+FP)$
Recall	$(TP)/(TP+FN)$
F1- Score	$2*\frac{(Precision \times Recall)}{(Precision+Recall)}$
TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative	

- Accuracy: It measures how often a model correctly classify CKD and non-CKD.
- Precision: In predictive analytics, precision is the degree to which the model's predictions agree with the observed data. The data points match the predictions more closely when the model is more accurate.
- Recall: The recall is a metric that indicates how well our model finds True Positives. Recall thus indicates the number of individuals that we accurately identified as having CKD out of all those who do.
- F1 Score: The precision and recall's harmonic mean is known as the F1-score.
- Cohen's Kappa: It assesses the level of agreement between the model's predictions and the actual presence or absence of CKD, while accounting for the agreement that could occur by chance.
- ROC AUC: It indicates how well the model can differentiate between CKD and non-CKD.

The proposed model was trained and evaluated on the dataset of CSV file for the detection of CKD to reveal their efficiencies in terms of the classification of CKD. According to the experimental result, the proposed model obtained the best accuracy of 98.95%.

Table 3. Classification Report

Parameters	CNN-LSTM	CNN-GRU	Ensemble Model
Accuracy	0.9875	0.9886	0.9895
Precision	1.0000	1.0000	1.0000
Recall	0.9787	0.9832	0.9856
F1- Score	0.9892	0.9885	0.9934
Cohens Kappa	0.9743	0.9756	0.9812
ROC AUC	0,9893	0.9893	0.9893

The accuracy and loss plot depicted in Figure 10; Figure 11 give an overview of how the model trained over a series of epochs. The graph shows that accuracy increases proportionately to the number of epochs. Furthermore, as the number of epochs reduced, the loss also fell until it reached its lowest point, suggesting that the model was sufficiently trained and that the CKD illness categorization could be completed.

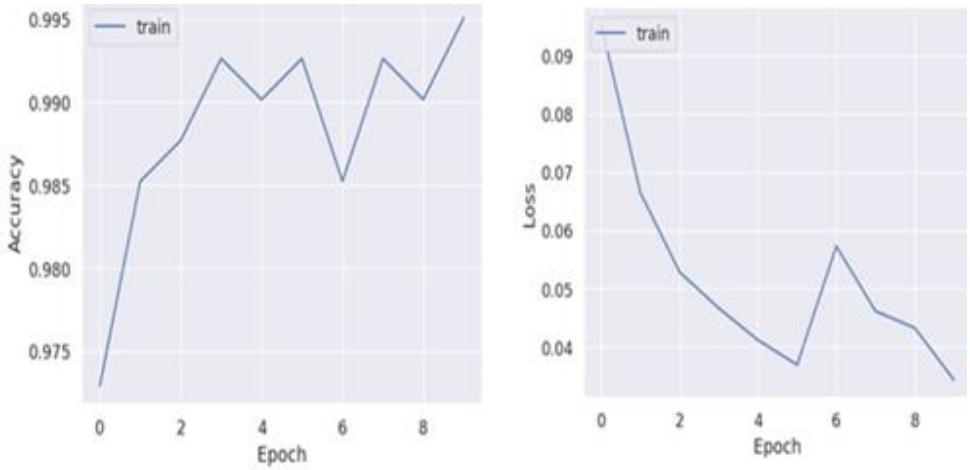

Fig. 10 Visual Representation of Accuracy and Loss for the CNN-LSTM Model

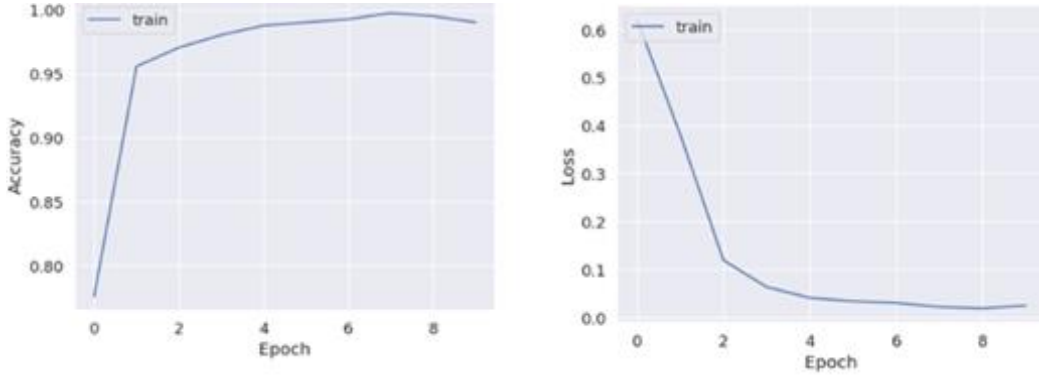

Fig. 11. Visual Representation of Accuracy and Loss for the CNN-GRU Model

In Figure 12, the confusion matrix [26] illustrates the performance evaluation of the suggested architecture, providing a comprehensive overview of its classification outcomes. The confusion matrix shows how many images the model properly and erroneously detected. The output dimension of the classification issue determines the matrix's size. It has a 2 × 2 size for binary categorization. The classification model's target and projected outputs are compared in the matrix.

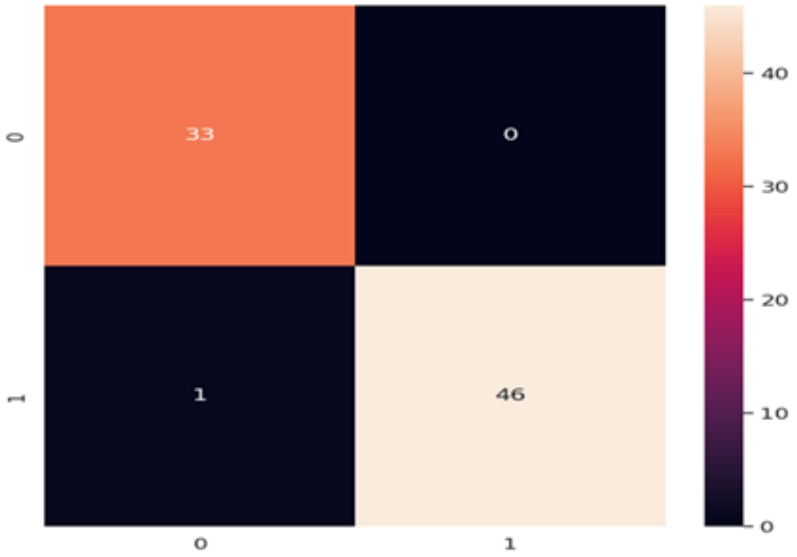

Fig. 12. Confusion Matrix of Proposed Model

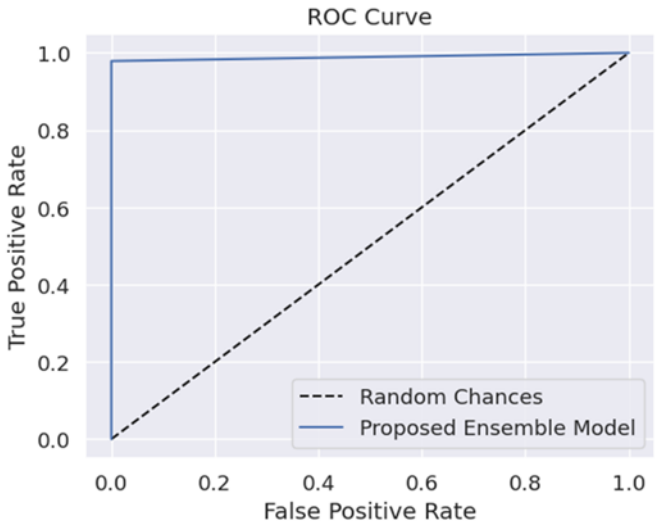

Fig. 13. ROC Curve of Proposed Model

To visualize the performance of CKD classification, the ROC curve is plotted is shown in Figure 13. The curve will showcase the model's performance at various classification thresholds, and calculating the corresponding TPR and FPR, indicating how well it distinguishes between CKD and non-CKD. The ROC curve serves as a visual depiction illustrating the classification model's performance. On the x-axis, it delineates the false positive rate, denoting the fraction of non-CKD instances inaccurately identified as CKD. Meanwhile, the y-axis portrays the true positive rate, signifying the proportion of CKD cases correctly identified as such. This graphical representation offers a comprehensive view of the model's ability to discriminate between CKD and non-CKD cases, highlighting the trade-off between sensitivity and specificity in the classification process. Random chances are important to consider when interpreting the ROC curve. The diagonal line in the ROC space represents the performance of a random classifier. Points above the diagonal line indicate that the classification model performs better than random chance, while points below the diagonal line suggest worse performance than random chance. The threshold is set at 1.0 at the lowest possible point, or at (0, 0). This implies that all patients are categorized by our model as without CKD.

- The threshold is set to 0.0 at the highest position, or (1, 1). This indicates that all patients are categorized with CKD.
- With FPR near to 0, we are achieving a TPR close to 1 at specific threshold values. At this point, the model will nearly accurately predict which patients will get CKD.
- AUC is the region bounded by the axes and a curve. This region is thought to be a sign of a high-quality model. A high AUC number is what we should strive for, as this metric ranges from 0

to 1. Models with good skills are those that have a high AUC.

The visual representation of the random sample provides insights into the accuracy of the predictions and any potential misclassifications. By plotting the images along with their predicted and true labels, we can assess how accurately the model is identifying CKD in the test dataset.

4.2. Performance Comparison with Existing Methods

Upon analyzing the table 4 comparing our system with existing methods, it becomes evident that our system outperforms other approaches in several key aspects. It achieves higher accuracy rates, demonstrating its superior ability to make correct predictions for CKD.

Table 4. Performance Comparison with Existing Methods

Author & Reference	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Cohens Kappa**	**ROC AUC**	**Sensitivity**
Ramat et al 2022 [28]	97.98	97.67	-	-	-	-	97.17
Prakash et al 2022 [29]	91.92	90.17	91.47	91.67	-		91.54
Venkatesan 2023 [27]	97	97.90	97.90	97.90	-	-	
Jain et al 2021 [30]	97.48	97	97.43	97.33	-	-	-
Alex et al 2022 [31]	96.67	96	96.32	-	-	-	-
Proposed	**98.95**	**100**	**98.56**	**99.34**	**98.95**	**98.93**	**-**

Moreover, our system demonstrates enhanced precision and recall metrics, underscoring its proficiency in reducing both false positives and false negatives, thereby enhancing the reliability of diagnostic outcomes. Furthermore, the system's F1-score surpasses that of current methodologies, underscoring a superior equilibrium between precision and recall in our approach. This balanced performance is crucial in accurately identifying high-risk individuals early and guiding appropriate personalized treatment plans. In Figure 14, a comprehensive comparison with existing approaches is presented, highlighting the relative performance and effectiveness of the proposed model in relation to other methods.

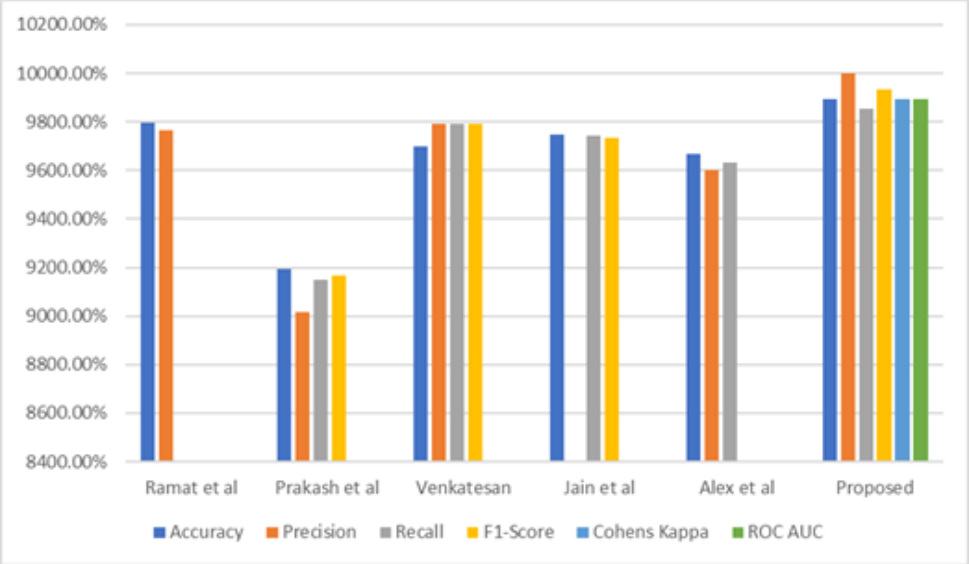

Fig. 14. Performance Comparison with Existing Approaches

5. Conclusion

The prognostic assessment of chronic disorders, particularly in cases like chronic kidney disease, is frequently impeded by substantial variability, introducing uncertainty into clinical decision-making and contributing to adverse outcomes. Traditional approaches for detecting chronic kidney disease (CKD) often suffer from accuracy limitations as they heavily depend on a restricted set of biological attributes for diagnosis. An ensemble model based on hybrid deep learning models for chronic kidney disease detection is designed and analysed in this work. CNN-GRU and CNN- LSTM were used as the hybrid model and SVM as the meta learner. The model demonstrated exceptional performance across various evaluation metrics, including accuracy, F1-score, recall, Cohen's Kappa, ROC AUC and precision. The model achieved a high accuracy of 98.95%, indicating overall correct predictions. Precision and recall scores of 1.0000 and 98.56, respectively, demonstrate the model's ability to make precise and comprehensive classifications. These results suggest that the model is robust and effective in its predictions for the given binary classification task.

References

[1] Fuchs, F. D., & Whelton, P. K. (2020). High blood pressure and cardiovascular disease. Hypertension, 75(2), 285-292.

[2] Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2021). Health and treatment of diabetes mellitus. International Journal of Health Sciences, 5(1)..

[3] Lamprea-Montealegre, J. A., Shlipak, M. G., & Estrella, M. M. (2021). Chronic kidney disease detection, staging and treatment in cardiovascular disease prevention. Heart.

[4] Lv, J. C., & Zhang, L. X. (2019). Prevalence and disease burden of chronic kidney disease. Renal fibrosis: mechanisms and therapies, 3-15.

[5] Venkatakrishna, S. S. B., Onyango, L. C., Serai, S. D., & Viteri, B. (2023). Kidney Anatomy and Physiology. In Advanced Clinical MRI of the Kidney: Methods and Protocols (pp. 3-12). Cham: Springer International Publishing.

[6] Merrill, A. E., & Chambliss, A. B. (2020). Water and electrolyte balance. In Contemporary Practice in Clinical Chemistry (pp. 651-663). Academic Press

[7] Gao, Z., Chen, Z., Sun, A., & Deng, X. (2019). Gender differences in cardiovascular disease. Medicine in Novel Technology and Devices, 4, 100025.

[8] Hannan, M., Ansari, S., Meza, N., Anderson, A. H., Srivastava, A., Waikar, S., ... & Chronic Renal Insufficiency Cohort (CRIC) Study Investigators. (2021). Risk factors for CKD progression: overview of findings from the CRIC study. Clinical Journal of the American Society of Nephrology, 16(4), 648-659.

[9] Srivastava, A., Kaze, A. D., McMullan, C. J., Isakova, T., & Waikar, S. S. (2018). Uric acid and the risks of kidney failure and death in individuals with CKD. American Journal of Kidney Diseases, 71(3), 362-370.

[10] Kalantar-Zadeh, K., Jafar, T. H., Nitsch, D., Neuen, B. L., & Perkovic, V. (2021). Chronic kidney disease. The lancet, 398(10302), 786-802.

[11] Mathers, C. D. (2020). History of global burden of disease assessment at the World Health

Organization. Archives of Public Health, 78(1), 1-13.

[12] Ketkar, N., Moolayil, J., Ketkar, N., & Moolayil, J. (2021). Convolutional neural networks. Deep Learning with Python: Learn Best Practices of Deep Learning Models with PyTorch, 197-242.

[13] Zhao, J., Huang, F., Lv, J., Duan, Y., Qin, Z., Li, G., & Tian, G. (2020, November). Do RNN and LSTM have long memory?. In International Conference on Machine Learning (pp. 11365-11375). PMLR.

[14] Dey, R., & Salem, F. M. (2017, August). Gate-variants of gated recurrent unit (GRU) neural networks. In 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS) (pp. 1597-1600). IEEE.

[15] Yang, Y., Li, J., & Yang, Y. (2015, December). The research of the fast SVM classifier method. In 2015 12th international computer conference on wavelet active media technology and information processing (ICCWAMTIP) (pp. 121-124). IEEE.

[16] Abdel-Fattah, M. A., Othman, N. A., & Goher, N. (2022). Predicting chronic kidney disease using hybrid machine learning based on apache spark. Computational Intelligence and Neuroscience, 2022.

[17] Arif, M. S., Mukheimer, A., & Asif, D. (2023). Enhancing the early detection of chronic kidney disease: a robust machine learning model. Big Data and Cognitive Computing, 7(3), 144.

[18] Venkatesan, V. K., Ramakrishna, M. T., Izonin, I., Tkachenko, R., & Havryliuk, M. (2023). Efficient Data Preprocessing with Ensemble Machine Learning Technique for the Early Detection of Chronic Kidney Disease. Applied Sciences, 13(5), 2885.

[19] Srivastava, S., Yadav, R. K., Narayan, V., & Mall, P. K. (2022). An Ensemble Learning Approach For Chronic Kidney Disease Classification. Journal of Pharmaceutical Negative Results, 2401-2409.

[20] Silveira, A. C. D., Sobrinho, Á., Silva, L. D. D., Costa, E. D. B., Pinheiro, M. E., & Perkusich, A. (2022). Exploring early prediction of chronic kidney disease using machine learning algorithms for small and imbalanced datasets. Applied Sciences, 12(7), 3673.

[21] Singh, V., Asari, V. K., & Rajasekaran, R. (2022). A deep neural network for early detection and prediction of chronic kidney disease. Diagnostics, 12(1), 116.

[22] Prakash, S., Raja, P. V., Baseera, A., Hussain, D. M., Balaji, V. R., & Venkatachalam, K. (2022). Ensemble Nonlinear Support Vector Machine Approach for Predicting Chronic Kidney Diseases. Computer Systems Science & Engineering, 42(3).

[23] Ebiaredoh-Mienye, S. A., Swart, T. G., Esenogho, E., & Mienye, I. D. (2022). A machine learning method with filter-based feature selection for improved prediction of chronic kidney disease. Bioengineering, 9(8), 350.

[24] Rama, M., Tejaswini, G., Niharika, T., & Murari, K. (2022). Detecting chronic kidney disease by using a voting classifier. South Asian Journal of Engineering and Technology, 12(3), 69-77.

[25] Khalid, H., Khan, A., Zahid Khan, M., Mehmood, G., & Shuaib Qureshi, M. (2023). Machine Learning Hybrid Model for the Prediction of Chronic Kidney Disease. Computational Intelligence and Neuroscience, 2023.

[26] Caelen, O. (2017). A Bayesian interpretation of the confusion matrix. Annals of Mathematics and Artificial Intelligence, 81(3-4), 429-450.

[27] Venkatesan, V. K., Ramakrishna, M. T., Izonin, I., Tkachenko, R., & Havryliuk, M. (2023). Efficient Data Preprocessing with Ensemble Machine Learning Technique for the Early Detection of Chronic Kidney Disease. Applied Sciences, 13(5), 2885.

[28] Rama, M., Tejaswini, G., Niharika, T., & Murari, K. (2022). Detecting chronic kidney disease by using a voting classifier. South Asian Journal of Engineering and Technology, 12(3), 69-77.

[29] Prakash, S., Raja, P. V., Baseera, A., Hussain, D. M., Balaji, V. R., & Venkatachalam, K. (2022). Ensemble Nonlinear Support Vector Machine Approach for Predicting Chronic Kidney Diseases. Computer Systems Science & Engineering, 42(3).

[30] Jain, D., & Singh, V. (2021). A two-phase hybrid approach using feature selection and adaptive SVM for chronic disease classification. International Journal of Computers and Applications, 43(6), 524-536.

[31] Alex, D. M., Chandy, D. A., Christinal, A. H., Singh, A., & Pushkaran, M. (2022). A Hybrid Random Forest Classifier for Chronic Kidney Disease Prediction from 2D Ultrasound Kidney Images. International Journal of Pattern Recognition and Artificial Intelligence, 36(07), 2256010.