RESEARCH PAPER

A Robust Deep Hybrid Framework for Early DKD Detection Using Dynamic Routing-Based R-MOE with TabTransformer and BiLSTM

Konne Madhavi[1]∗ and Harwant Singh Arri[2]

[1]∗School of Computer Science and Engineering, LPU, India-144411
[2]School of Computer Science and Engineering, LPU, India-144411
**Corresponding Author: konnemadhavi@gmail.com*

Received: 28th Feb, 2026; Revised: 6th March 2026; Accepted: 7th April, 2026; Available Online: 20th April, 2026

ABSTRACT
One of the most serious microvascular consequences of diabetes mellitus, diabetic kidney disease (DKD) or diabetic nephropathy is a major global cause of end-stage renal disease (ESRD) and chronic kidney disease (CKD). Nearly 40% of diabetes individuals are at risk of developing chronic kidney disease (CKD), which is characterized by a progressive decrease of renal function brought on by long-term damage to the renal vasculature. The coexistence of diabetes and hypertension further accelerates this decline, creating a significant global health and economic burden. Despite the clinical importance of early intervention current diagnostic approaches such as microalbuminuria have proven unreliable, as they fail to consistently capture the complex and nonlinear relationships among biochemical and therapeutic variables, leading to delayed or inaccurate diagnoses. To address these limitations, we propose a novel hybrid deep learning framework for early DKD detection that combines advanced preprocessing, feature selection and hybrid classification. The input dataset undergoes preprocessing with an improved Gaussian filter enhanced by a Laplacian operator to improve data quality, followed by feature selection using the Boruta algorithm to identify the most informative attributes. The selected features are then passed into an enhanced TabTransformer integrated with a BiLSTM network for effective classification. In the proposed architecture, the conventional feed-forward network of the TabTransformer is replaced with a Routed Mixture-of-Experts (R-MoE) module, where a dynamic router replaces the standard gating mechanism to extract richer and more meaningful representations. These representations are subsequently processed by the BiLSTM, which performs classification to predict DKD or non-DKD thereby enabling accurate, reliable and early diagnosis to support improved clinical management. Our model was successfully implemented in Python, achieving a high accuracy of 99.65

Keywords: *DKD, Tab Transformer, BiLSTM, R-MoE, dynamic router, Boruta algorithm and hybrid classification model*
How to cite this article: Madhavi K and Arri HS, A Robust Deep Hybrid Framework for Early DKD Detection Using Dynamic Routing-Based R-MOE with TabTransformer and BiLSTM. Int J Drug Deliv Technol. 2026;16(6): 530-545. DOI: 10.25258/ijddt.16.6.90
Source of support: Nil.
Conflict of interest: None

I. INTRODUCTION

The Kidney is one of the most vital organs of a Human body that helps a lot by filtering blood and throws the waste toxic contents through excretory organs. But the food habits and other bad habits can affect the functions of Kidney and even kidney fails. Most vital Kidney diseases can include Chronic Kidney Disease (CKD), Diabetic Nephropathy at the 2nd stage of Diabetes and still many more. If the kidneys fail to work, it can severely impact the immune system and lead to acute renal failure [1]. DKD is significantly more common in several demographic groups, such as women, elderly persons, racial underrepresented groups and individuals with comorbidities such as diabetes mellitus or hypertension [2].

The KDIGO 2022 guidelines propose using the terms "diabetes and kidney disease" and "diabetic kidney disease" to provide a more inclusive nomenclature. This approach ensures comprehensive treatment across the full spectrum of patients, differentiated only by eGFR range and albuminuria. [3]. Several machine learning methods, including K-Nearest Neighbour (KNN), Naive Bayesian Classifier (NB), Supreme Boosting Classifier (SB) and Decision Tree (DT) are used for early diagnosis and successful therapy for timely detection and prediction of DKD was considered [4]. To improve the availability of diabetic kidney disease (DKD) screening, various artificial intelligence (AI), deep learning and machine learning systems have been incorporated to address the issues and challenges associated with DKD [5].

Traditional methods include a hybrid deep learning approach [7] that combines CNNs and SVMs for detection, an AI-driven predictive framework [6] for early kidney disease diagnosis, and predictive models [8] such as

**Author for Correspondence: konnemadhavi@gmail.com*

CURA that use AI/ML for dataset analysis, prediction, and monitoring of diabetic kidney patients. The SCI-XAI framework [9] enables early identification of high-risk kidney patients, while an explainable MLP model [10] uses LIME for transparent CKD detection. Another interpretable ML approach [11] applies SHAP and LIME to improve model interpretability and sup-port clinicians in understanding predictions.

An important step forward in this direction is the intro-duction of an explainable robust Deep Hybrid Framework. Our proposed method does more than merely forecast DKD by combining an improved tab transformer with a BiLSTM model. It also describes how and why this prediction was realized. Because of this transparency, the cause of the disease can be better understood, leading to more individualized and knowledgeable clinical advice. The primary issues with the present and ambiguous AI methods for diagnosing DKD are being considered to be addressed by the proposed explainable

Deep learning model. The model explores the factors that affect its predictions in great detail making it a vital tool for nephrologists. This invention improves the accuracy of DKD detection and provides insight into the disease's progression. Patients may receive more individualized, efficient care as a result.

Tu, Z et al [13] analysed 2,048 hospitalized individuals with type 2 diabetes in order to evaluate the correlation between the risk of diabetic peripheral neuropathy (DPN) and diabetic kidney disease (DKD) and the Triglyceride Glucose (TyG) index, a marker of insulin resistance. From the results they observed that higher TyG values increased risk of DKD and the combination of DKD and DPN. However, there was no significant association between TyG and DPN in multivariable analyses. In general, these results suggest that the TyG index may be useful for detecting DKD early in individuals using type 2 diabetes. However, this study was confined to a cross-sectional design of hospitalized patients, which may restrict its generalizability. Other limitations include insufficient evalua-tion of potential confounding variables such as medication, lifestyle factors, and the duration of diabetes; additionally, the current study lacked longitudinal follow-up to determine whether TyG is predictive of DKD and DPN behaviours over time.

Shi et al.[14] discovered the potential of employing artificial intelligence (AI) to categorize diabetic kidney disease (DKD) by integrating clinical and retinal vascular characteristics in people with type 2 diabetes. The researchers utilized retinal imaging data in conjunction with demographic information and various longitudinal clinical variables, such as blood pressure, blood glucose, and lipids, to develop predictive AI models. Changes in the retinal blood vessels, such as the size of the vessels and the way they branch out in angiopathy, were found to be strongly linked to kidney damage. This means they can be used as a useful non-invasive biomarker. The retinal parameters, referred with clinical data, yield a more accurate prediction of DKD, suggesting the AI model could classify DKD, based on retinal features from a non-invasive retinal assessment. This methodology could provide an inexpensive and viable screening technique for DKD to promote advocacy for further assessment of kidney health in patients at risk, especially in low-resource settings where kidney function testing is limited. The current classification of DKD facilitates better patient management or timely intervention in at-risk patients with type 2 diabetes. However this model needs additional types of testing and validation of the AI model in diverse and larger populations to assess generalizability and clinical application.

Nayak, S et al.[15] developed an interpretable machine learning approach for analyzing the stages and progression of Diabetic Kidney Disease (DKD) through retrospective analysis of clinical data from 227 patients treated at one South Indian tertiary care hospital. This approach was based on Bayesian optimization to tune models using XGBoost classifier that pinpointed serum uric acids, urea, phosphorus, red blood cell count, calcium, and absolute eosinophil count to be key predictors. A gradient boost regression model performed well in identifying progression of DKD. Although this method performs well in all aspects, it lacks explainability and interpretability in the context of recognizing methods for individualized patient care.

Zheng, X et al.[16] investigated the usefulness of non-invasive lens advanced glycation end product (AGE) measure-ment for individuals with type 2 diabetes to screen for and evaluate the severity of diabetic kidney damage (DKD). 134 patients in all were taken in for the examination and randomly allocated to low, medium and high risks group for progression of DKD risk as well as DKD and non DKD groups based on Chinese clinical instructions. The analysis demonstrated lens AGE levels are much greater in the high-risk and DKD groups. Regression analysis affirmed value in DKD diagnosis and in assessing severity, while Cox regression found positive correlation relative to duration. The severity evaluation was conducted only as a single-center, cross-sectional study with a limited sample size that limits the conclusions drawn from the analysis and highlights the need to examine lens AGE measurement in longitudinal, multi-centric studies for DKD screening.

Brondani, L. D. A et al.[17] used LC-MS/MS peptidomic profiling in combination with a Random Forest algorithm to characterize people suffering from type 2 diabetes (T2DM). The investigators obtained urine samples from 60 T2DM patients and identified more than 1,080 naturally-occurring peptides, which represented approximately 100 proteins from all three stages. In the predictions, there were identifiable differences in the urinary peptidomic pattern, relating to the severe DKD group and many of the peptides that were predictive in the model included fragments of collagen or an alpha-1 antitrypsin, proteins that had also been previously linked with DKD Predictions done using Proteasix tool shows 48

specific proteases, such as metallopeptidases, cathepsins and calpains, etc to make up the most important peptide markers. The results support urinary peptidomics as a way to determine molecular signatures of DKD, and identifying possible underlying pathogenic mechanisms for discovering new therapeutic targets. The urine parameters analysed for small sample size and cross-sectional design could limit the ability to generalize and causal interpretation of the results.

Guo, X et al.[18] examined lipid alterations of the development of diabetic kidney disease (DKD) and determined to evaluate new serum biomarkers which distinguish DKD in patients with diabetes mellitus (DM). The lipidomic anal-ysis using serum samples extracted from individuals and heathy controls assigned in three groups - DM without DKD (n=55), early DKD (n=21), advanced DKD (n=32), and healthy controls(n=22) and found that certain levels of lipid classes, such as lysophosphatidylethanolamines (LPEs), phosphotidylethanolamines (PEs), ceramides (Cers) and dia-cylglycerols (DAGs) increased with KD, and LPCs, monoa-cylglycerols (MAGs), and triacylglycerols (TAGs) increased noticeably in early rate to advanced stage of disease. Positive associations were also observed with higher levels of LPCs, LPEs, PEs and DAGs, which showed significant relationships with DKD risk in logistic regression, and strong positive correlations were found between LPEs and clinical indices such as UACR and eGFR. A machine learning biomarker panel was developed to distinguish DKD from DM. However, its findings are limited by the observational design and small sample size.

II. LITERATURE SURVEY:

Several researchers have introduced effective methods for detecting Diabetic Kidney Disease and their approaches can be summarized as follows: Li et al. [12] conducted a cohort that was retrospective study spanning 10 years, examining diabetic patients' medical records from a major hospital. Multiple machine learning models were applied to this historical data to predict the likelihood or onset of diabetic kidney disease (DKD), with the goal of identifying patterns and risk factors that support early detection and prevention in Type 2 Diabetes patients including Diabetes with peripheral circulatory complications particularly diabetic gangrene, diabetic peripheral an-giopathy, and diabetic ulcerative DKD emerged as significant risk factors in many studies. Because they share risk factors with diabetic retinopathy, these microvascular consequences are more strongly linked to DKD than to other diabetes-related diseases. urine protein is still an important early indicator for DKD diagnosis, and urine albumin levels are a critical predictor of the illness's development as well as a measure of the disease's severity in a number of predictive models for end-stage renal disease in DKD. however this model cannot adopt to complex machine learning models.

Xia et al. [19] designed a cross-sectional study across six provinces in China to assess the screening prevalence of diabetic kidney disease (DKD) in peoples with type 2 diabetes. This study encompasses a substantial cohort of patients from diverse healthcare settings across different geographical regions to evaluate the regional status of DKD screening. In general, the rates of screening for DKD were low, which means that many people with type 2 diabetes are not being checked for problems with their kidneys. Older age, longer diabetes duration, comorbidities, and better access to health-care services were all linked to higher rates of DKD screening. The analysis reveals substantial variations in DKD screening contingent upon geographical location and the category of healthcare facility utilized by patients. However, the study's cross-sectional design limited its ability to establish causal relationships and to capture changes in screening over time.

*A. **Motivation and Research Gap***

Early Diabetic Kidney Disease (DKD) detection has several advantages, such as lowering the risk of kidney failure and related complications through prompt intervention that may decrease or completely stop the disease's development. Reliable prediction models help doctors make well-informed treatment choices, which enhance patient outcomes and quality of life. Healthcare systems can maximize resources, reduce treatment costs and reduce the need for dialysis or transplan-tation by early detection of high-risk patients. Additionally, incorporating predictive technologies improves clinical judg-ment, guarantees individualized treatment, and increases self-assurance in the management of long-term diabetes-related disorders. All of these benefits emphasize how crucial it is to create accurate and comprehensible models for early DKD detection.

More reliable and accurate early detection techniques are desperately needed, especially considering the severity and progressive nature of Diabetic Kidney Disease (DKD), which frequently results in end-stage renal disease. Because some patients with protein uria may return to normal levels, tradi-tional diagnostic markers like microalbuminuria have demon-strated limitations. This uncertainty affects risk assessment and the timing of interventions. This variability highlights the challenge of identifying DKD in its early stages and casts doubt on the validity of traditional methods. In order to create sophisticated, trustworthy and interpretable diagnostic models that can precisely capture intricate clinical patterns, facilitate prompt intervention and enhance patient outcomes in DKD management, there is a significant research gap.

*B. **Problem Statement and Novelty***

The increased care needs, treatment complications and expenses associated with type 2 diabetes and its comorbidities put a significant strain on the healthcare system. In particular, diabetic kidney disease (DKD) presents a significant challenge because its pathophysiology is still poorly understood and current

treatment results are frequently insufficient. Early prevention and diagnosis are therefore critical to slowing progression. While machine learning (ML) and artificial in-telligence (AI) shown remarkable success in healthcare by improving diagnosis, prediction and personalized treatment through analysis of complex datasets, their application in DKD detection and management remains limited. This gap highlights the need for advanced ML-based approaches to enhance early diagnosis and improve clinical outcomes in DKD.

To overcome these challenges, we propose a hybrid deep learning model that offers the following key advantages:

- Initially, data is sourced from the DKD dataset [20].
- The preprocessing phase is processed using an improved Gaussian filter, further enhanced by incorporating the Laplacian operator to improve data quality
- Next, feature selection is performed using the Boruta algorithm and the selected features are passed to a deep hybrid model to diagnose whether the condition is DKD or non-DKD.
- A hybrid deep learning model, combining TabTrans-former and BiLSTM is utilized for effective classification.
- A Routed Mixture-of-Experts (R-MoE) module is used in place of the conventional feed-forward network (FFN) layer in an enhanced TabTransformer. In this config-uration, a dynamic router which serves as the central component of the R-MoE and is in charge of extracting significant features from the input data replaces the conventional gating mechanism in the MoE layer.
- Following the extraction of these features, the BiLSTM layer makes the final classification to figure out whether the input is DKD or not. The proposed method's exper-imental assessments include specificity, F1-score, accu-racy and precision.

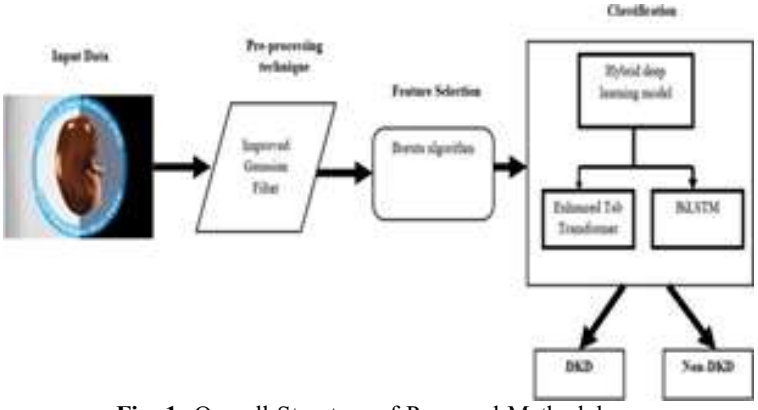

Fig. 1: Overall Structure of Proposed Methodology

- The Python platform is used to build and test the entire model, proving its usefulness and practicality.

III. PROPOSED METHODOLOGY

We present a hybrid deep learning approach in this paper for the early detection of DKD using data collected from the DKD dataset, where the preprocessing phase uses an improved Gaussian filter and a Laplacian operator to improve the quality of the data, followed by feature selection using the Boruta al-gorithm to keep only the most relevant attributes, and then the selected features are sent to a hybrid classification model that combines a BiLSTM and an enhanced TabTransformer, where the TabTransformer replaces the traditional feed-forward net-work with an R-MoE module, where a dynamic router extracts meaningful representations. The BiLSTM layer then processes the features, performing the final classification to differentiate between DKD and non-DKD. Figure 1 shows the proposed model's overall architecture.

A. Pre-processing using Improved Gaussian Filter

The DKD dataset provides the input images, which are then preprocessed with an improved Gaussian filter. In this ap-proach, the standard Gaussian filter is enhanced by integrating the Laplacian operator which helps in improving the overall data quality. A detailed explanation of this process is provided below.

Pre-processing is performed as the initial step, where Gaussian-based filtering plays a crucial role. Filtering is con-sidered an essential stage in data pre-processing. In this paper, the Laplacian of Gaussian (LoG) is applied to detect image edges effectively. This filtering process helps in identifying sudden variations within the images. The Gaussian filter model is mathematically represented in Eq. (1), where σ denotes the standard deviation of the filter and k represents the Gaussian index.

$$G(n) = \frac{1}{\sqrt{2\pi\sigma^2}} \quad \frac{k^2}{2\sigma^2} \qquad (1)$$

The second derivative of the Gaussian filter is calculated as given in Eq. (2). Subsequently, the normalized Laplacian of Gaussian (LoG) is formulated as shown in Eq. (3).

$$G(n) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{k^2}{\sigma^2} - 1 \; e^{-\frac{k^2}{2\sigma^2}} \quad (2)$$

$$LoGfilter_k = \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \frac{k^2}{\sigma^2} - 1 \; e^{-\frac{k^2}{2\sigma^2}}}{\sum_k e^{-\frac{k^2}{2\sigma^2}}} \quad (3)$$

When, the sum of the filter coefficients approaches zero the filter functions as a High-Pass Filter (HPF). Based on the proposed Gaussian filter concept, the Improved Laplacian of Gaussian (ILoG) is expressed in Eq. (4). Here, C denotes the number of taps (i.e., filter coefficients), σ is fixed at 25 and the order of ILoG is set to 10.

$$ILoGfilter_k = \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \frac{k^2}{\sigma^2} - 1 \; e^{-\frac{k^2}{2\sigma^2}}}{\sum_k e^{-\frac{k^2}{2\sigma^2}}} - \frac{1}{C} \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \frac{k^2}{\sigma^2} - 1 \; e^{-\frac{k^2}{2\sigma^2}}}{\sum_k e^{-\frac{k^2}{2\sigma^2}}} \quad (4)$$

Once the pre-processing stage is completed, the process proceeds to feature selection.

B. Feature Selection using Boruta algorithm

After pre-processing, the Boruta algorithm is used to select features, which is a wrapper-based method built around the Random Forest (RF) classifier. This algorithm distinguishes between important and non-important attributes within the dataset. To achieve this, Boruta randomizes the input features (A1, A2, A3, . . . , An) by creating duplicate copies, referred to as shadow attributes (S1, S2, S3, . . . , Sn).

An RF classifier is then trained using the expanded dataset to ascertain the maximum Z-score among the shadow attributes (MZSA). The remaining features are designated as question-able, features with Z-scores below the MZSA are rejected and features with better scores are verified as relevant. Finally, the selected features are utilized in the classification process.

Therefore, the Boruta algorithm was applied to select the most relevant variables for the model. The top 15 features were determined by analyzing feature importance scores, as shown in Fig. 2. Hematocrit, globulin, glycated hemoglobin (HbA1c), red blood cell (RBC) count, absolute eosinophil count (AEC), phosphorous, serum uric acid, calcium, thyroid stimulating hormone (TSH), erythrocyte sedimentation rate (ESR), fasting blood sugar (FBS), urea, random blood sugar (RBS), and absolute neutrophil count (ANC).

C. Classification Using a Hybrid Deep Learning Model

The selected features are then used in the classification stage, where a hybrid deep learning model is employed. This model integrates an enhanced TabTransformer with a BiLSTM for effective classification.

In the enhanced TabTransformer, the conventional feed-forward network (FFN) layer is replaced with a Routed

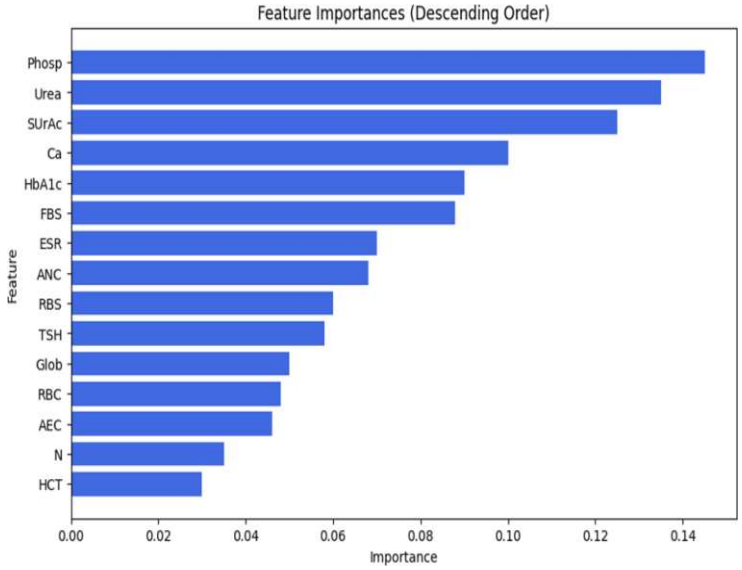

Fig. 2: Selected Features using Boruta algorithm

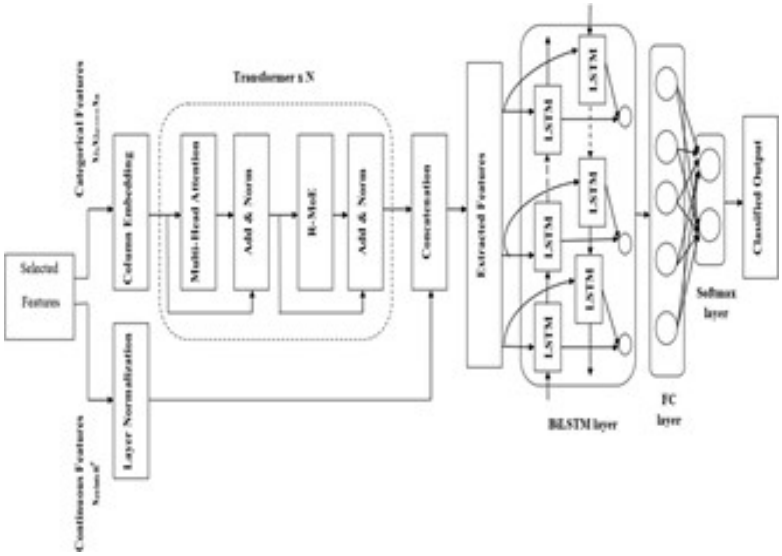

Fig. 3: Architecture of the Hybrid Deep Learning Model

Mixture-of-Experts (R-MoE) module. The R-MoE uses a dynamic router as its central component, in contrast to the conventional gating mechanism used in traditional MoE layers. This allows for the extraction of more significant features from the given input data. The BiLSTM layer then receives these improved features and uses them to make the last classification, identifying whether the input is DKD or not. Below is a thorough explanation of this procedure.

1) ***Enhanced TabTransformer:*** The improved TabTransformer is used for feature extraction in the hybrid deep learning model using the self-attention based Transformer architecture; this framework consists of a stack of N Trans-former layers after a column embedding layer. The model can effectively capture complex feature interactions because each Transformer layer is made up of an R-MoE layer and a multi-head self-attention mechanism.

The TabTransformer works with feature target pairs (x,y) that include continuous features (xcont) and categorical fea-tures (xcat). A parametric embedding method known as col-umn embedding is used to process the categorical features, mapping each feature into a dense representation. After that,

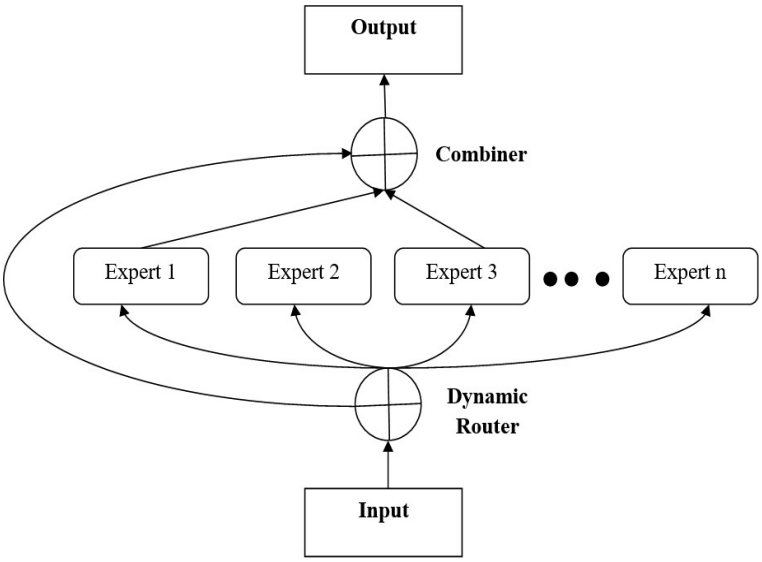

Fig. 4: Framework of the R-MoE Layer

this embedding is run through several Transformer layers, each of which consists of an R-MoE layer and a multi-head self-attention mechanism. Contextual embeddings are made possible by the self-attention mechanism, which permits each feature embedding to focus on every other one in order to capture inter-feature dependencies. Below is a presentation of the R-MoE layer's comprehensive workflow.

- **R-MoE layer (Routed Mixture-of-Experts)**

The Mixture-of-Experts (MoE) is a combinatory method of learning that uses several expert models, every specializing in different sub-spaces of the input domain, MoE divides the input space into sub-spaces, trains expert models within each sub-space and then integrates their knowledge through a router, also referred to as a gating function. The fundamental concept of MoE is to select the most appropriate experts for a given input using this router. The input is first passed to the router, which determines the most suitable experts to process it. In this example, Expert 1 and Expert 3 are chosen. Their outputs are then combined by the combiner to produce the final result.

In general, Following three fundamental components make up the MoE framework:

- **Experts:** These are individual models or neural networks designed to specialize in different regions of the input space or in distinct tasks. The experts can exist in lots of ways, such as linear models, decision trees, or deep neural networks
- **Router (Gating Function):** The router is responsible for deciding which expert, or combination of experts, should process a given input. Typically implemented as a neural network, it is trained jointly with the experts. During inference, only a selected subset of experts is activated, chosen by the router based on the input.
- **Combiner:** The combiner aggregates the outputs of the selected experts, weighting them according to the proba-bilities or scores provided by the router, and generates the final prediction. In Figure 4, the R-MoE layer's framework is shown.

In the R-MoE layer, the standard gating function (router) is replaced with a dynamic router, thereby transforming the conventional MoE into an improved and more effective format. The R-MoE layer utilizes a dynamic routing strategy guided by model confidence. In this approach, the model evaluates whether the initially selected experts are sufficient for handling a given input. If the confidence is low, additional experts are gradually incorporated.

Here, P in Equation 5 represents the confidence distribution of input x across different experts, where Pi indicates the confidence that the i-th expert can effectively handle x. If the highest probability in P is sufficiently large, only the corresponding expert is used. However, if the highest prob-ability is too small, more experts are gradually added to improve reliability. Experts are included in descending order of their probabilities until the cumulative confidence exceeds a predefined threshold p. At this point, the model is considered confident that the selected experts can adequately process the input x, while minimizing the number of experts activated.

Formally, the elements in P are first arranged in descending order, producing a sorted index list I. Next, the smallest subset of experts is identified such that their cumulative probability surpasses the threshold p. The number of selected experts, denoted as t is then determined by:

$$t = \arg \min_{k \in \{1,\ldots,N\}} \sum_{j \le k} P_{i,j} \ge p \qquad (5)$$

Here, p serves as the threshold that determines the confidence level required for the model to stop adding additional experts. It is a hyperparameter with a range from 0 to 1. A greater p value leads to the activation of more experts.

Within the dynamic routing mechanism, the computation of g∗X) is given as:

$$g(X) = \begin{cases} P_{i,e_i}, & e_i \in S \\ 0, & e_i \in S \end{cases} \qquad (6)$$

Here, S denotes the set of experts selected, determined by t in Equation (5)

$$S = \{e_{i_1}, e_{i_2}, \ldots, e_{i_t}\} \qquad (7)$$

The router then assigns probability weights PIj to these experts based on the confidence distribution P. Finally, the outputs are aggregated in a weighted manner to form the final representation of the R-MoE layer, expressed as:

$$y = \sum_{j=1}^{t} P_{\hat{s}}\, e_{ij}(x) \qquad (8)$$

The continuous features are concatenated with the contextual embeddings derived from the categorical features. The key aspect of column embedding is the unique identifier approach, specifically designed for tabular data. In this method, each cat-egorical feature is associated with an embedding lookup table containing vectors for each class, along with an additional embedding for missing values. The unique identifier helps distinguish classes within a column, while the embedding itself combines a category-specific component with a feature-value-specific component.

Figure 3 depicts the systematic process where categorical features are first transformed through column embedding and processed by a Transformer block, while numerical features undergo layer normalization. The outputs of both are then combined to generate the extracted feature representations, which are subsequently used for the classification stage.

Let (x,y) denote a feature target pair, where x is composed of categorical features (xcat) and continuous features (xcont). The continuous features (xcont) belong to Rc with c representing the total number of continuous variables. The categorical features (xcat) are expressed as x1,x2,. . . ,xm, where every xi corresponds to a categorical variable

for i=1,. . . ,m.

All the categorical feature x_i is mapped into a parametric vector of dimension d using column embedding. The embed-ding of x_i is represented as $e_{\phi i}(x_i) \in R^d$, whereas, ϕ_i denotes the embedding parameters associated with x_i. The complete set of categorical embeddings can thus be expressed as:

$$E_\phi(x_{cat}) = \{e_{\phi_1}(x_1), \ldots, e_{\phi_m}(x_m)\} \quad (9)$$

The embeddings $E_\phi(x_{cat})$ are passed into the first Trans- former layer.

Each layer processes these input embeddings and gener- ates contextual embeddings by aggregating information from the surrounding embeddings across successive layers. This transformation is represented by a function $f\theta$, which takes the parametric embeddings $e_{\phi 1}(x_1), \ldots, e_{\phi m}(x_m)$ and pro- duces the contextual embeddings $h_1, \ldots, h_m$, whereas every h_i R^d for $i = 1, \ldots, m$.

The contextual embeddings are then combined with the continuous features x_{cont}, resulting in a combined feature vector of dimension (d $m + c$). Whereas, d represent the dimension of each categorical feature embedding, m denotes the number of categorical features and c represent the number of continuous features.

Therefore, the extracted features are passed to the BiLSTM model for classification within the hybrid deep learning frame-work, as illustrated below.

D. ***BiLSTM model***

A complex type of recurrent neural network design called BiLSTM, or Bidirectional Long Short-Term Memory, com-bines two LSTM networks. The input sequence is processed forward by one and backward by the other. The model is especially useful for sequence modelling tasks because of its bidirectional structure, which allows it to extract contextual information from both past and future inputs. BiLSTM has been widely used in fields like classification, named entity recognition and sentiment analysis.

The LSTM model, which was first created to solve the vanishing gradient issue with conventional RNNs, serves as the basis for this architecture. Building on this, researchers subsequently showed that BiLSTMs capture richer contextual dependencies than unidirectional models, making them par-ticularly effective for tasks like phoneme classification and handwriting recognition.

The BiLSTM layer receives the output from the tabtrans- former layer in the proposed hybrid framework. The BiLSTM can efficiently model sequential information in both directions while maintaining long-range dependencies. By combining its own sequence modelling capabilities with the contextual representations learned from the TabTransformer, the BiLSTM functions as a feature extractor, improving the predictive accu-racy of the model. The following collection of formulas can be used to mathematically characterize the BiLSTM architecture:

1) Input Gate (i_t):

$$i_t = \sigma\ W^f_{ix} x_t + W^f_{ih} h^f_{t-1} + W^f_{ic} c^f_{t-1} + b^f_i \odot \sigma\ W^b_{ix} x_t + W^b_{ih} h^b_{t+1} + W^b_{ic} c^b_{t+1} + b^b_i \quad (10)$$

2) Forget Gate (f_t):

$$f_t = \sigma\ W^f_{fx} x_t + W^f_{fh} h^f_{t-1} + W^f_{fc} c^f_{t-1} + b^f_f \odot \sigma\ W^b_{fx} x_t + W^b_{fh} h^b_{t+1} + W^b_{fc} c^b_{t+1} + b^b_f \quad (11)$$

3) Cell State Update (C_t):

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh\ W^f_{cx} x_t + W^f_{ch} h^f_{t-1} + b^f_c + i_t \odot \tanh\ W^b_{cx} x_t + W^b_{ch} h^b_{t+1} + b^b_c \quad (12)$$

4) Output Gate (o_t):

$$o_t = \sigma\ W^f_{ox} x_t + W^f_{oh} h^f_{t-1} + W^f_{oc} c^f_t + b^f_o \odot \sigma\ W^b_{ox} x_t + W^b_{oh} h^b_{t+1} + W^b_{oc} c^b_t + b^b_o \quad (13)$$

5) Hidden State (h_t):

$$h_t = o_t \odot \tanh(C_t) \quad (14)$$

A. Experimental Results Based on DKD Datasets

The following section presents the early diagnosis of DKD using an evaluation framework that incorporates key performance metrics, including Accuracy, Precision, Sensitivity, Specificity, F1-Score and Kappa coefficient. Each metrics is explained in detail to guarantee a thorough comprehension of the model's overall functionality and efficacy.

Figure 5 presents aq confusion matrix that illustrates the models effectiveness with early detection of DKD by compar-ing a detected outcomes with an actual ground truth labels. The vertical axis denotes the actual classes, where class 0 represents non-DKD cases and class 1 corresponds to DKD cases, while the horizontal axis represents the detected classes. The results, expressed in percentage distribution, highlight the model's strong diagnostic capability. With very high true positives 155 and true negatives 143 along with extremely low false positives 1 and false negatives 1 the model demonstrates

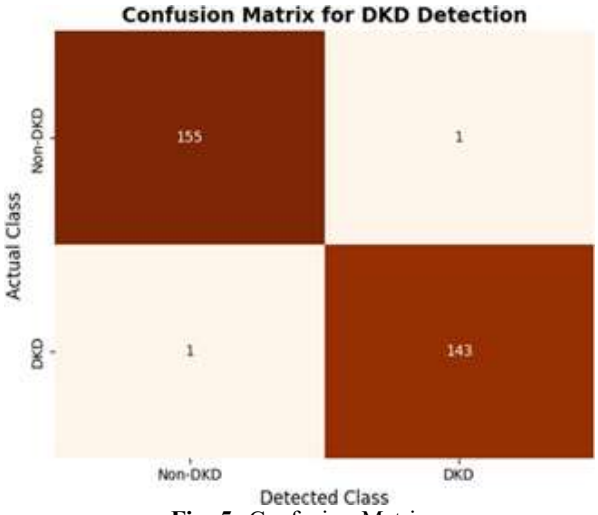

Fig. 5: Confusion Matrix

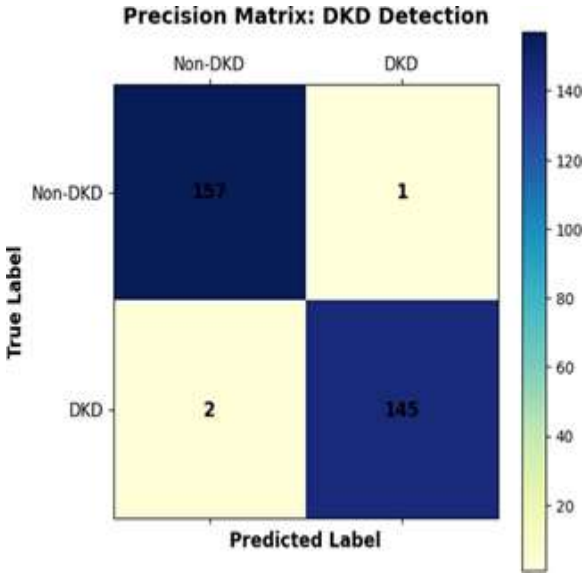

Fig. 6: Precision Matrix

remarkable accuracy and significantly reduces misclassifica-tion in the early detection of DKD.

The precision matrix in Figure 6 provides a summary of the models performance in detecting DKD. It records 157 true positives, 1 false positive, 0 false negatives and 145 true negatives, indicating an exceptionally low rate of misclassification. This reflects the models strong ability to determine the difference between situations with and without DKD. The model continuously produces accurate predictions while minimizing errors, achieving a precision and recall of 98.7%. These results validate its stability and dependability in aiding DKD detection.

The recall matrix used to evaluate the models' performance in DKD detection is shown in Figure 7. The model's sensitivity in accurately identifying DKD cases is demonstrated by the results, which show 156 true positives, 2 false positives, 1 false

TABLE I: DKD Dataset Samples

Id	157	109	17	347	24
Age	62	54	47	43	42
Bp	70	70	80	60	100
Sg	1.025	–	–	1.025	1.015
Al	3	–	–	0	4
Su	0	–	–	0	0
Rbc	normal	–	–	normal	normal
Pc	abnormal	–	–	normal	abnormal
Pcc	Not present	Not present	Not present	Not present	Not present
Ba	Not present	Not present	Not present	Not present	present
Bgr	122	233	114	108	–
Bu	42	50.1	87	25	50
Sc	1.7	1.9	5.2	1	1.4
Sod	136	–	139	144	129
Pot	4.7	–	3.7	5	4
Hemo	12.6	11.7	12.1	17.8	11.1
Pcv	39	–	–	43	39
Wc	7900	–	–	7200	8300
Rc	3.9	–	–	5.5	4.6
Htn	yes	no	yes	no	yes
Dm	yes	yes	no	no	no
Cad	no	no	no	no	no
Appet	good	good	poor	good	poor
Pe	no	no	no	no	no
Ane	no	no	no	no	no
Classification	dkd	dkd	dkd	Not dkd	dkd

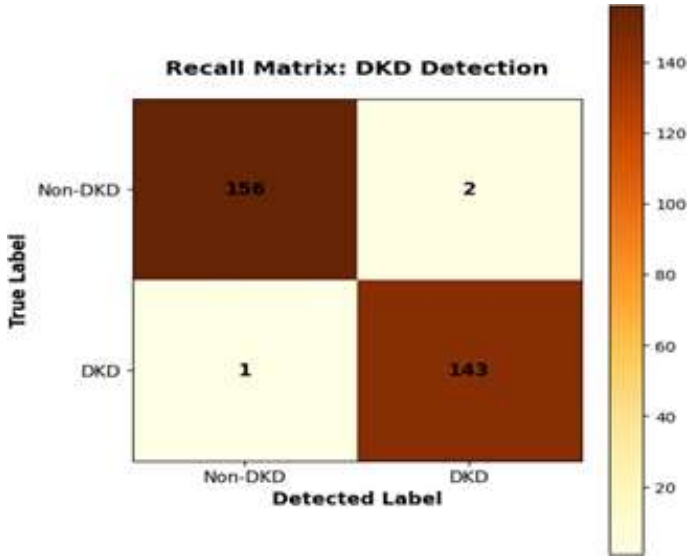

Fig. 7: Recall Matrix

negative and 143 true negatives. Reliable detection is ensured by the extremely low false negative count, which shows that DKD cases are rarely mislabeled as normal. The model is well-suited for early diagnosis and efficient monitoring because it exhibits robust capability in differentiating between healthy and DKD-affected cases, with strong recall and high overall accuracy.

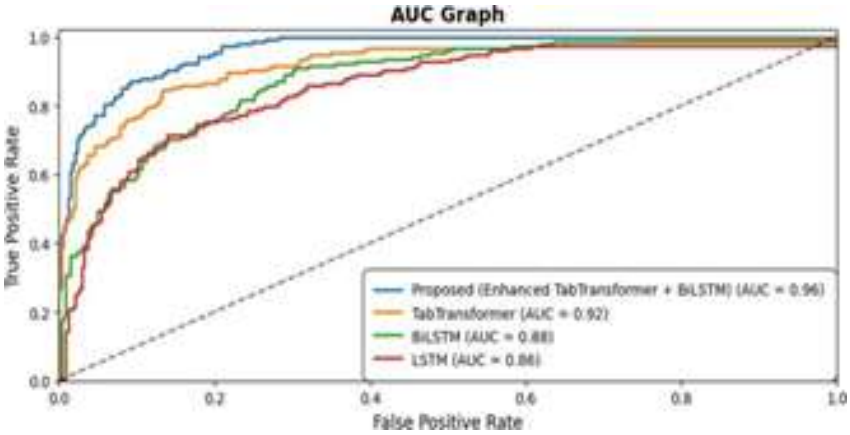

Fig. 8: AUC graph for the proposed method

Figure 8 compares the performance of the proposed hybrid model with three baseline models Tab Transformer, BiLSTM and LSTM in the early diagnosis of diabetic kidney disease (DKD) using the Area under the Curve (AUC) graph. With the greatest AUC value of 0.96, the enhanced Tab Transformer in conjunction with BiLSTM exhibits the strongest discrimina-tive ability, showing superior ability to differentiate between patients at risk of DKD and those who are not. On the other hand, the Tab Transformer achieves an AUC of 0.92, whereas LSTM and BiLSTM achieve 0.88 and 0.86, respectively. These findings demonstrate that the hybrid approach offers more

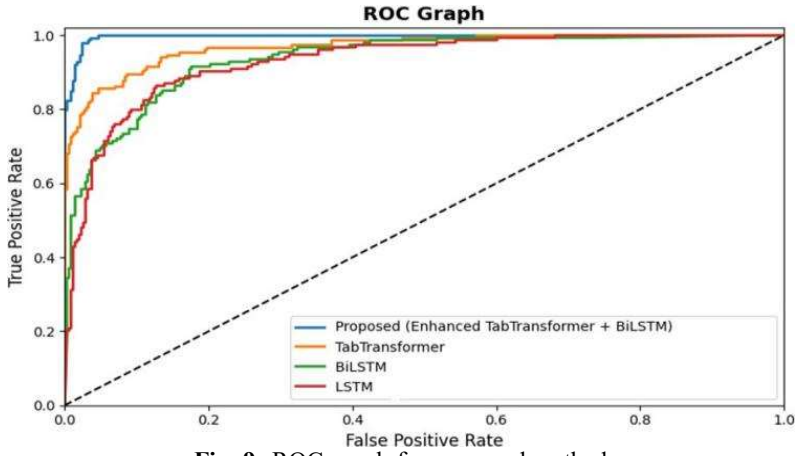

Fig. 9: ROC graph for proposed method

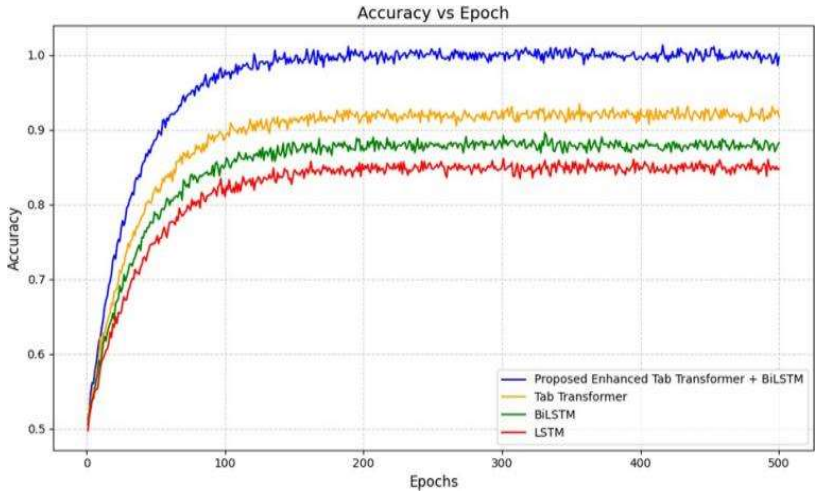

Fig. 10: Epochs Vs Accuracy for proposed method

accurate and dependable detection, highlighting its potential to enhance clinical decision-making and early diagnosis in DKD. The proposed improved Tab Transformer + BiLSTM is compared with Tab Transformer, BiLSTM and LSTM for early DKD diagnosis in Figure 9 is ROC curve. With a curve that is closest to the top-left corner and shows greater sensitivity and fewer false positives, the proposed model performs the best. While LSTM performs the worst, with results that are almost random prediction, Tab Transformer and BiLSTM exhibit moderate accuracy. This demonstrates the hybrid model's superiority for accurate DKD early detection.

The accuracy and loss curves of the proposed Enhanced Tab Transformer + BiLSTM model over 500 training epochs are shown in Figures 10 and 11, in comparison to Tab Transformer, BiLSTM and LSTM. While the baseline models Tab Transformer, BiLSTM and LSTM achieve relatively lower

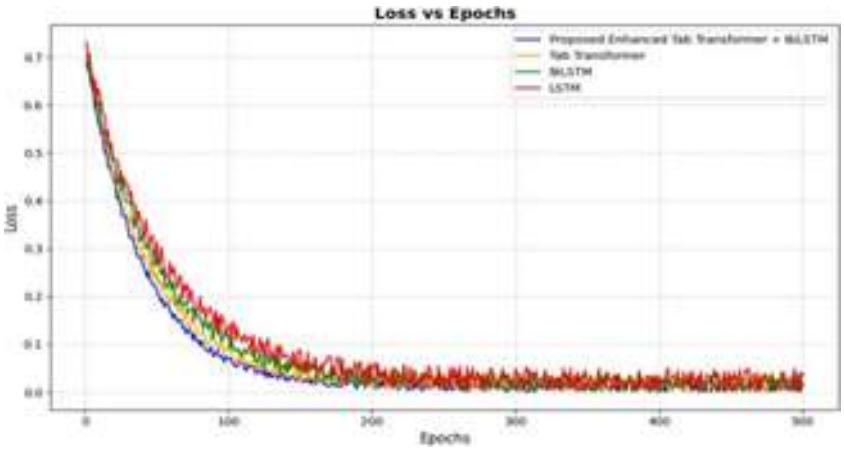

Fig. 11: Epochs Vs Loss for the proposed method

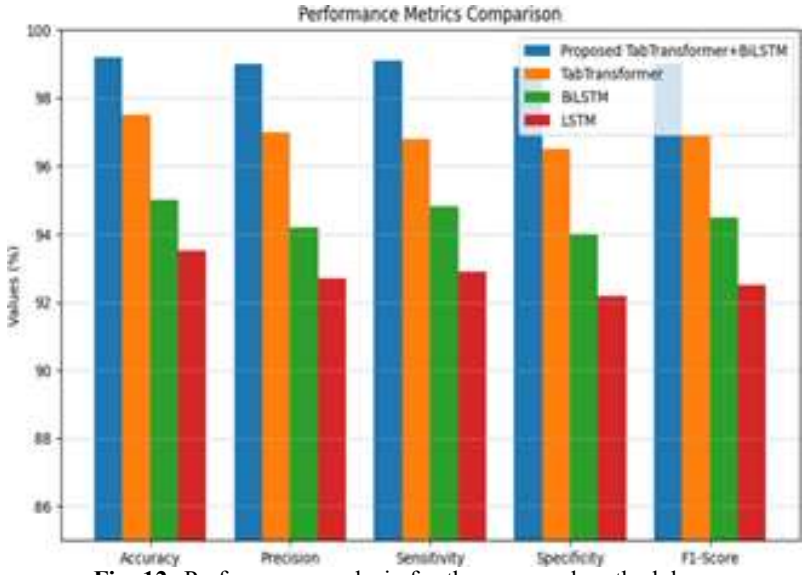

Fig. 12: Performance analysis for the proposed methodology

accuracy, indicating weaker learning capacity the proposed model achieves the highest and most stable accuracy in Figure 10, convergently approaching 1.0. Comparably, Figure 11 demonstrates the proposed method's excellent generalization and effective convergence by continuously maintaining the lowest loss throughout the training process. The proposed model performs significantly better than current techniques providing a more dependable and accurate framework for the early diagnosis of DKD, as further supported by the higher final loss values seen in BiLSTM and LSTM.

A comparative performance analysis for early DKD detection diagnosis based on important evaluation metrics, such as F1-score, Accuracy, Precision, Sensitivity and Specificity is shown in Figure 12. The x-axis represents the corresponding performance measures and the y-axis shows the metric values between 86 and 100. Each colored bar corresponds to a different model, providing a clear visual comparison of their effectiveness. The proposed enhanced Tab Transformer + BiL-STM model demonstrates superior performance, achieving the highest Accuracy of 99.65%, Precision of 99.60%, Sensitivity of 99.75%, Specificity of 98.9%, and F1-score of 99.6%. In comparison, the Tab Transformer attains slightly lower results with an Accuracy of 97.81%, Precision of 96.80%, Sensitivity of 98.5%, Specificity of 97.4%, and F1-score of 98.2%. The BiLSTM model follows with an Accuracy of 95.8%, Precision of 95.2%, Sensitivity of 95.71%, Specificity of 95.4%, and F1-score of 95.29%. The LSTM records the lowest performance, achieving an Accuracy of 94.91%, Precision of 94.52%, Sensitivity of 93.44%, Specificity of 93.81%, and F1-score of 92.11%. The above outcomes clearly emphasize the effectiveness and robustness of the proposed hybrid model in achieving highly accurate early diagnosis of DKD.

Figure 13 illustrates the comparative analysis of the Kappa coefficient for different detection approaches employed in the early diagnosis of DKD. When calculating the degree of agreement between the expected classifications and the actual clinical results, the Kappa coefficient takes chance agreement into consideration. Among the evaluated techniques, the pro-posed Enhanced TabTransformer + BiLSTM model obtained

TABLE II: Comparative analysis results

Reference	Technique	Dataset	Accuracy (%)	Precision (%)	Sensitivity (%)	F-Score (%)
[14]	EOAEDL-CKDD	CKD dataset	98.12	98.16	97.83	97.99
[15]	LASSO	Chronic Kidney Disease dataset	96.47	98.14	91.38	94.64
[13]	DSCNN	CKD dataset	99.18	Not measured	98.87	98.96
[12]	SGD	Not mentioned	94.39	84.98	94.50	Not measured
Proposed	Enhanced Tab Transformer + BiLSTM	DKD dataset	99.65	99.60	99.75	99.60

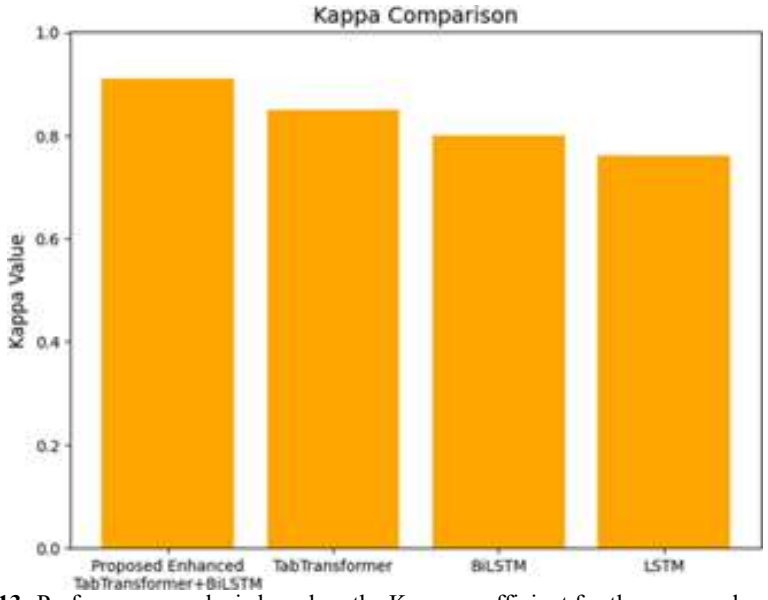

Fig. 13: Performance analysis based on the Kappa coefficient for the proposed method

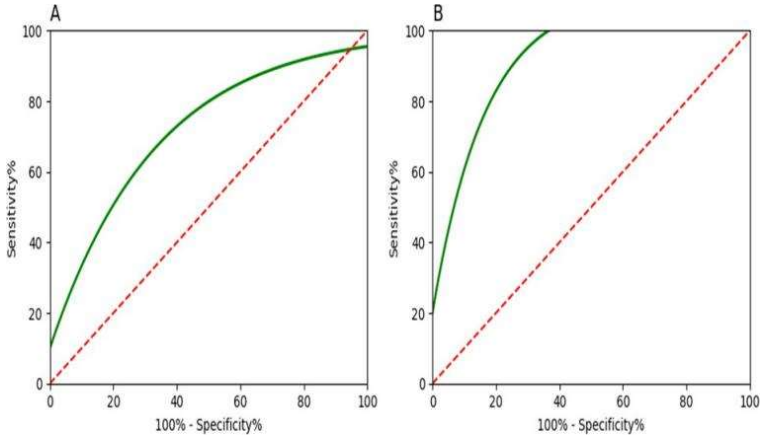

Fig. 14: ROC Analysis for Identifying Early-Stage Diabetic Kidney Disease

the best Kappa value of 0.91, indicating outstanding reliability and robustness. In comparison, TabTransformer obtained the lowest Kappa value of 0.85, followed by BiLSTM with 0.80 and LSTM with 0.76. These results highlight that the proposed hybrid model delivers the most consistent and accurate perfor-mance, offering superior capability in the reliable detection of DKD at an early stage.

Figure 14 presents the ROC analysis for detecting early-stage diabetic kidney disease (DKD) in patients with T2DM. Fig A depicts the ROC curve for early DKD diagnosis, yielding an AUC of 0.78 (95% CI: 0.72–0.84, p ¡ 0.001) which demonstrates moderate discriminatory ability between affected and non-affected patients. Fig B illustrates the ROC curve for evaluating DKD severity, showing a higher AUC of 0.91 (95% CI: 0.87–0.95, p ¡ 0.001), indicative of excellent classification performance. Overall, these results highlight that the proposed method is effective for early DKD detection while exhibiting even stronger predictive capability for disease severity thereby supporting timely diagnosis and management.

*E. **Comparison with Published Work***

To confirm if the proposed strategy is successful for early diagnosis of DKD, a comparative performance evaluation was carried out against existing methods reported in the literature, specifically those presented in studies [12], [13], [14], and [15]. As shown in Table II, the proposed model obtained a significantly high accuracy of 99.65%, whereas the corre-sponding accuracies reported in the referenced studies were 94.39%, 99.18%, 96.47%, and 98.12%. These findings clearly indicate that the performance of our model is either on par with or superior to previously established methods. By employing a robust deep learning framework tailored for clinical data analysis, the proposed model not only improves diagnostic accuracy but also ensures efficient and reliable identification of DKD at its early stages.

IV. CONCLUSION

In this paper, a hybrid model combining an enhanced TabTransformer with a BiLSTM network was proposed for the early diagnosis of DKD diseases. The input data were obtained from the DKD dataset where preprocessing was carried out using an improved Gaussian filter enhanced by a Laplacian operator to refine data quality. Feature selection was then performed with the Boruta algorithm to identify the most informative attributes. These selected features were fed into the enhanced TabTransformer, integrated with a BiLSTM network for effective classification. The feed-forward network of the TabTransformer was swapped out for an R-MoE module in the proposed architecture. To extract richer and more significant representations, a dynamic router was used in place of the conventional gating mechanism. The BiLSTM then processed the resultant representations and performed the classification to determine whether they were DKD or not. A classification accuracy of 99.65% was attained by the proposed model. Future research could improve the system even more by adding sophisticated deep learning algorithms and bigger more varied healthcare datasets.

REFERENCES

[1] M. A. Mohammed, A. Lakhan, K. H. Abdulkareem, M. Deveci, A. K. Dutta, S. Memon, *et al.*, "Federated-reinforcement learning-assisted IoT consumers system for kidney disease images," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 4, pp. 7163–7173, 2024.

[2] V. Puri, I. Priyadarshini, A. Kataria, S. Rani, and H. Min, "Privacy-First ML for Chronic Kidney Disease Prediction: Exploring a Decentralized Approach Using Blockchain and IPFS," *IEEE Access*, 2025.

[3] N. Montero, L. Oliveras, A. Mart´ınez-Castelao, J. L. Gorriz, M. J. Soler, B. Fernandez-Fernandez, *et al.*, "Clinical Practice Guideline for detection and management of diabetic kidney disease: A consensus report by the Spanish Society of Nephrology," *Nefrolog´ıa (English Edition)*, vol. 45, pp. 1–26, 2025.

[4] M. Thakkar, D. Darji, and S. Verma, "Kidney Disease: Early-Stage Identification and Prevention Using Supervised Machine Learning Tech-niques," in *Proc. 2024 Int. Conf. Communication, Computer Sciences and Engineering (IC3SE)*, pp. 1868–1872, 2024.

[5] Z. Meng, Z. Guan, S. Yu, Y. Wu, Y. Zhao, J. Shen, *et al.*, "Non-invasive biopsy diagnosis of diabetic kidney disease via deep learning applied to retinal images: a population-based study," *The Lancet Digital Health*, vol. 7, no. 5, 2025.

[6] G. Eswar, C. Hariharan, S. Lavein, and V. Akilandeswari, "Predictive Analysis of Kidney Diseases with Artificial Intelligence and Data Sci-ence," in *Proc. 2025 Int. Conf. Visual Analytics and Data Visualization (ICVADV)*, pp. 779–784, 2025.

[7] R. Jain, V. Kukreja, S. Chattopadhyay, A. Verma, and R. Sharma, "Deep Learning Insights: Evaluating the Efficacy of CNN-SVM for Accurate Kidney Disease Identification," in *Proc. 2024 IEEE Int. Conf. Infor-mation Technology, Electronics and Intelligent Communication Systems (ICITEICS)*, pp. 1–5, 2024.

[8] M. Aljaafari, E. S. Hany, A. A. Hany, and S. E. Sorour, "Integrating Innovation in Healthcare: The Evolution of CURA's AI-Driven Virtual Wards for Enhanced Diabetes and Kidney Disease Monitoring," *IEEE Access*, 2024.

[9] P. A. Moreno-Sa´nchez, "Data-driven early diagnosis of chronic kidney disease: development and evaluation of an explainable AI model," *IEEE Access*, vol. 11, pp. 38359–38369, 2023.

[10] M. S. Arif, A. U. Rehman, and D. Asif, "Explainable machine learning model for chronic kidney disease prediction," *Algorithms*, vol. 17, no. 10, p. 443, 2024.

[11] S. K. Ghosh and A. H. Khandoker, "Investigation on explainable machine learning models to predict chronic kidney diseases," *Scientific Reports*, vol. 14, no. 1, p. 3687, 2024.

[12] G. Li, J. Li, F. Tian, J. Ren, Z. Guo, S. Pan, *et al.*, "A 10-year retrospective cohort of diabetic patients in a large medical institution: Utilizing multiple machine learning models for diabetic kidney disease prediction," *Digital Health*, vol. 10, p. 20552076241265220, 2024.

[13] Z. Tu, J. Du, X. Ge, W. Peng, L. Shen, L. Xia, *et al.*, "Triglyceride glucose index for the detection of diabetic kidney disease and diabetic peripheral neuropathy in hospitalized patients with type 2 diabetes," *Diabetes Therapy*, vol. 15, no. 8, pp. 1799–1810, 2024.

[14] S. Shi, L. Gao, J. Zhang, B. Zhang, J. Xiao, W. Xu, *et al.*, "The automatic detection of diabetic kidney disease from retinal vascular parameters combined with clinical variables using artificial intelligence in type-2 diabetes patients," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, p. 241, 2023.

[15] S. Nayak, A. Amin, S. R. Reghunath, G. Thunga, D. Acharya, K. N. Shivashankara, *et al.*, "Development of a machine learning-based model for the prediction and progression of diabetic kidney disease: A single centred retrospective study," *International Journal of Medical Informatics*, vol. 190, p. 105546, 2024.

[16] X. Zheng, Y. Gao, Y. Huang, R. Dong, M. Yang, X. Zhang, *et al.*, "Clinical value of noninvasive lens

advanced glycation end product detection in early screening and severity evaluation of patients with diabetic kidney disease," *BMC Nephrology*, vol. 24, no. 1, p. 379, 2023.

[17] L. D. A. Brondani, A. A. Soares, M. Recamonde-Mendoza, A. Dall'Agnol, J. L. Camargo, K. M. Monteiro, and S. P. Silveiro, "Urinary peptidomics and bioinformatics for the detection of diabetic kidney disease," *Scientific Reports*, vol. 10, no. 1, p. 1242, 2020.

[18] X. Guo, Z. Zhang, C. Li, X. Li, Y. Cao, Y. Wang, *et al.*, "Lipidomics reveals potential biomarkers and pathophysiological insights in the progression of diabetic kidney disease," *Metabolism Open*, vol. 25, p. 100354, 2025.

[19] H. Khalid, A. Khan, M. Z. Khan, G. Mehmood, and M. S. Qureshi, "Machine learning hybrid model for the prediction of chronic kidney disease," *Computational Intelligence and Neuroscience*, vol. 2023, no. 1, p. 9266889, 2023.

[20] S. K. David, M. Rafiullah, and K. Siddiqui, "Comparison of different machine learning techniques to predict diabetic kidney disease," *Journal of Healthcare Engineering*, vol. 2022, no. 1, p. 7378307, 2022.