Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Early diagnosis of end-stage renal disease risk in type 2 diabetes mellitus using advanced analysis of clinical laboratory data

Raafat M. Munshi [1, 2, *], Othman Y. Alyahyawy [2], Lammar R. Munshi [3], Shashi Kant Gupta [1, 4]

[1]*Department of Computer Science and Engineering, Lincoln University College, Petaling Jaya, Malaysia*
[2]*Department of Medical Laboratory Technology (MLT), Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia*
[3]*Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia*
[4]*Center for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab 140401, India*

ARTICLE INFO

Article history:
Received 1 July 2025
Received in revised form
4 November 2025
Accepted 9 December 2025

Keywords:
End-stage renal disease
Type 2 diabetes
Risk prediction
Machine learning
Decision tree

ABSTRACT

End-stage renal disease (ESRD) is a serious complication of Type 2 Diabetes Mellitus (T2DM) and has a significant negative effect on patient health. Early and accurate detection is essential but difficult to achieve in clinical settings. This study introduces an Optimized Grey Wolf Convolutional Decision Tree (OGW-ConvDT) classifier to predict the risk of ESRD by combining advanced machine learning techniques with clinical laboratory data. The model uses Z-score standardization for data normalization, Principal Component Analysis (PCA) to reduce data dimensions, and the SelectKBest method for selecting the most important features. A Convolutional Neural Network (CNN) is used to extract spatial features, and a Decision Tree (DT), optimized using the Grey Wolf (GW) algorithm, performs the final classification. The proposed method was tested on a publicly available dataset from Kaggle and achieved strong performance: precision (0.996), F1-score (0.996), recall (0.997), accuracy (0.997), AUC (0.999), specificity (0.959), log loss (0.009), and AUC-PRC (0.824). These results show that the OGW-ConvDT model performs better than traditional methods and provides an effective and reliable tool for early ESRD risk detection in T2DM patients.

1. Introduction

Comprehensive testing is essential in the rapidly emerging healthcare field to ensure that innovations meet demanding standards for patient security, organizational reliability, and compliance with regulations (Tupsakhare and Kulkarni, 2025). The principles of traditional medicine are contributing to the developments of the digital age in an exciting moment that presents both opportunities and difficulties. Medical laboratories are often observed as the foundation of precise healthcare diagnosis. Artificial Intelligence (AI) is extensively used in the biomedical industry for surgeries that are minimally invasive, endoscopic navigation, forecasting disease, medical diagnosis, prognosis, and the identification of novel medicinal indicators (Xie et al., 2024). Diabetes mellitus is a metabolic disease resulting from a number of reasons, such as inadequate insulin or the body's incapacity to use insulin. Several patients initially could not realize that they had diabetes mellitus due to poor routine variables, such as eating excessive food that is heavy in sugar and salt, not exercising, and not being aware of the disease. AI's influence on healthcare is distinguished by its capacity to examine enormous datasets, spot intricate patterns, and generate predictions with formerly unattainable precision. AI improves the rapidity and precision of the evaluation of images in the fields of pathology and radiography. Furthermore, AI is crucial in helping patients through telemedicine, virtual help, and online surveillance, radically changing the usual model of patient-doctor relationships.

Labs have emerged as vital centers for development, as combining big data with technology, like Machine Learning (ML), and AI has improved the diagnostic capacity (Almutairi et al., 2025). Laboratory medicine has undergone an important shift as a result of the combination of automation and ML. Such a shift marks the transition from

* Corresponding Author.
Email Address: rmonshi@kau.edu.sa (R. M. Munshi)
https://doi.org/10.21833/ijaas.2026.01.002
Corresponding author's ORCID profile:
https://orcid.org/0000-0001-7696-0452

traditional physical and partially computerized processes to a digital age marked by greater efficacy, accuracy, and reliability (Ain et al., 2024). AI algorithms have also proven highly successful in identifying some cancers in initial phases, including hepatitis or intestinal tumors, enhancing patient quality of life, and also eliminating the financial problem (Aamir et al., 2024). Through improved therapy that is less costly, faster, and more secure, improvements in outcomes were achieved. AI has increasingly facilitated personal treatment through the evaluation of information and prediction of implications (Adeoye and Adams, 2024). End-Stage Renal Disease (ESRD) in Type 2 Diabetes Mellitus is difficult to diagnose early because of the Silent, Early symptoms of the disease, variable patient profile, and complicated evolution of the disease. The conventional diagnostic techniques are not able to identify the premature renal deterioration. The analysis of advanced clinical laboratory information should be integrated to improve the accuracy of prognosis with the help of timely intervention and better patient outcomes. The proposed research aims at establishing a sophisticated machine learning model, OGW-ConvDT, that outperforms previous efforts in accurately estimating the risk of a person developing end-stage renal disorder (ESRD), in the presence of type 2 diabetes, based on their available clinical laboratory data. The experiment, therefore, seeks to allow early identification, effective management of patients, and timely medical intervention through a combination of data preprocessing, optimal feature selection, and classification. Key contributions of this research can be summarized as follows:

- The research aims to address the ESRD hazard in people with T2DM, which has the possibility to improve clinical decision-making and patient outcomes.
- Employing intricate algorithms along with clinical laboratory data to increase the effectiveness of ESRD risk estimation.
- Z-score standardization is used to normalize the data to ensure reliable feature scaling. Principal Component Analysis (PCA) is utilized to reduce the highly complex clinical features into a simpler space. SelectKBest is applied to choose the appropriate features for diagnosing ESRD risk.
- A novel model is introduced, combining Convolutional Neural Network (CNN) for spatial extraction of features, a Decision Tree (DT) for classification, and Grey Wolf Optimization (GWO) for efficient model training.

The relevant works of existing models are discussed in Section 2, along with how the current research bridges the gaps. Section 3 explains the methodology context. Section 4 delivers illustrations of the results and findings. The discussion portion is shown in Section 5, while Section 6 describes the conclusion along with the limitations and its future scope.

2. Literature survey

Better accessibility in healthcare was provided by the MERN-stack-based Healthcare Management Platform (Reddy et al., 2025), which had features, including specific role monitoring, symptom evaluation, and appointment booking. Performance metrics showed increased patient involvement, improved efficiency, and high safety with potential applications in telemedicine, blockchain, and AI diagnostics. The limitations were around the scalability and accuracy in complex medical cases.

Developed an ML model to forecast CKD/HF risk in early-stage T2DM was the aim of the research (Kanda et al., 2022). Extreme gradient boosting was externally evaluated and achieved the highest accuracy using Japanese claims data (217,054 patients) (AUC 0.718/0.837). Patients at higher risk fared worst. Limitations include the use of claims data, a retrospective approach, and potential unmeasured confounders.

Developing an ML model for T2DKD diagnosis was the goal of the experiment (Liu et al., 2023). CatBoost performed best (AUC 0.86) using 3,624 T2DM patient data (multi-center, 2019); a reduced version with 12 characteristics was still effective (AUC 0.84). The interpretability was enhanced via SHAP. Limitations included the requirement for prospective clinical validation, potential selection bias, and the retrospective design.

Using baseline data from 748 individuals, the research (Bai et al., 2022) assessed machine learning's ability to predict 5-year ESKD risk in CKD patients. Random forest, naïve Bayes, and logistic regression all exhibited better sensitivity and Kidney Failure Risk Equation (KFRE) in accuracy, which helped with early screening. Limitations included restricted predictor variables, a single cohort, a small sample size, and no external validation.

An ML-based risk score for predicting DKD in type 2 diabetes was created and validated in the present research (Hosseini Sarkhosh et al., 2023). Six important characteristics were chosen from 1,907 patients using Recursive Feature Elimination with Cross-Validation, with development and validation AUCs of 79% and 75.8%, respectively. The model was built as an easily navigable web-based application, employed regular screening data, and demonstrated satisfactory calibration. The single-center data was a limitation.

ML and Electronic Medical Records (EMR) were used to create a 3-year risk forecasting model for Diabetic Kidney Disease (DKD) in patients with T2DM (Dong et al., 2022). The Light GBM model highlighted important risk factors, such as age, histamine, and glucose control. Since data from one hospital was used, it was unable to fully represent the diversity of patient populations.

12,190 T2DM patients' 3-year DKD risk was predicted by the experiment (Zou et al., 2025) using machine learning-based models. Five important predictors were found, and LightGBM was the most accurate of the seven methods. Early intervention

might benefit from the concept. A limitation that might restrict generalizability was the retrospective single-dataset design without external validation.

The used HL7 FHIR to integrate an ML model into an EMR in order to predict the 5-year ESRD risk in 19,159 T2DM patients was objective of the research (Wang et al., 2022). AUCRC 0.79 and AUROC 0.95 indicated that XGBoost performed best. Albumin, eGFR, and creatinine were important predictors. The design was retrospective, and there was no external validation that was prospective or multi-center.

An ML model was created in the research (Watanabe et al., 2025) to forecast fast renal deterioration in DKD patients with intact estimated glomerular filtration rate (eGFR) (>60). Utilizing time-series data from 2,533 patients, the model's highest AUC was 0.81, with urine protein variability and eGFR serving as important predictors. The absence of external validation and the small number of endpoint cases were limitations.

Evaluated the predicted fast-progressing renal impairment and the necessity for a referral to nephrology in T2DM patients with eGFR≥60, the research (Hsu et al., 2023) developed a 6-month machine learning model. XGBoost, RF, LR, and an ensemble voting classifier were evaluated using EMR data; the ensemble performed the best overall. A retrospective single-cohort design without external validation is a limitation.

2.1. Research gaps

The limited utilization of AI for exactly synthesizing extensive clinical data and identifying numerous medical diseases concurrently represents a research gap. Under early disease identification, the conventional approach shows less efficiency, which contributes to later interventions. As an example, the MERN-stack developed Healthcare Management Platform was designed to enhance the system of access by monitoring the role, evaluating the symptoms, and scheduling the appointment, but had a problem of flexibility in complex instances and precision (Reddy et al., 2025). Correspondingly, the LightGBM-based 3-year Diabetic Kidney Disease (DKD) risk prediction model found the important prognostic factors such as age and glucose control, but had shortcomings due to the selection of a single hospital's data, which limited the possibility to scale in different populations (Dong et al., 2022). Both point to the accessibility requirement for more scalable, precise, and generally applicable AI healthcare applications. The proposed research introduces OGW-ConvDT, which can address this research gap through the use of temporal and spatial trends within clinical lab data, parameter optimization of the Decision Tree via a metaheuristic Grey Wolf, and strict data preprocessing for interpretability.

This not only outlines the predictive capabilities of the already existing models but also gives higher priority to clinically meaningful variables like recall and specificity, which directly lead to earlier and more targeted interventions with regard to ESRD in T2DM patients.

3. Methodology

The methodology describes the usage of clinical lab data to create an OGW-ConvDT Classifier to diagnose ESRD in people with T2DM. The method relates spatial patterns and optimal decision-making to develop patient care and diagnostic precision. The methodological process for predicting ESRD risk in T2DM using AI-based classification is shown in Fig. 1.

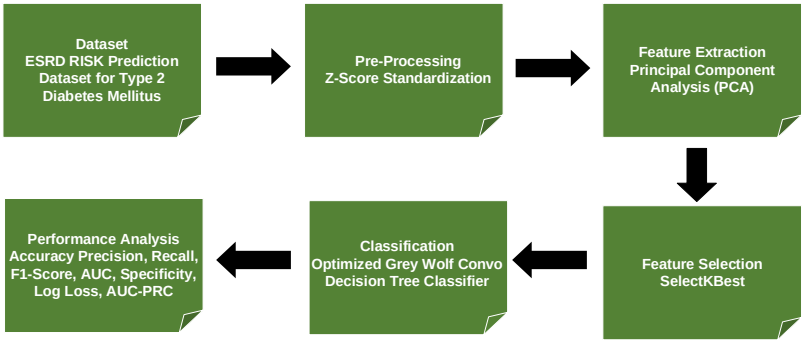


Fig. 1: Methodological workflow

3.1. Dataset description

The dataset called “ESRD Risk Prediction Dataset for Type 2 Diabetes Mellitus” is gathered from Kaggle (https://www.kaggle.com/datasets/ziya07/esrd-risk-prediction-dataset/data). The purpose of the dataset is to support the formation and evaluation of ML algorithms for evaluating the danger of ESRD in individuals with T2DM. The dataset includes demographic data, clinical features, medications, and comorbidities of the T2DM patients. The Kaggle dataset "ESRD Risk Prediction Dataset for Type 2 Diabetes Mellitus" includes 53,477 patient records with 16 demographic and clinical characteristics. Patients at risk of End-Stage Renal Disease are represented by the binary target variable ESRD, where 0 denotes no risk for ESRD. The dataset is extremely unbalanced, with 348 (0.65%) ESRD patients and 53,129 (99.35%) non-ESRD cases in the class distribution. To maintain the initial class ratio, an 80:20 stratified train–test split was used, yielding 42,781 samples for training and

10,696 samples for testing. The demographic, clinical, comorbidity, medication, and goal characteristics of the ESRD Risk Prediction Dataset for patients with Type 2 Diabetes Mellitus are described in Table 1.

Table 1: Feature description for the ESRD risk prediction dataset in type 2 diabetes mellitus patients

Category	Feature	Description
Demographic data	Age	Patient's age (20–90 years)
	Gender	Binary indicator (0 = Male, 1 = Female)
Clinical features	Baseline creatinine	Baseline serum creatinine levels (mg/dL)
	Mean creatinine	Mean serum creatinine over one year (mg/dL)
	HS CRP	High-sensitivity C-reactive protein levels (mg/dL)
	UPCR	Spot urine protein-to-creatinine ratio (g/g)
	HbA1c	Glycated hemoglobin levels (%)
	Hemoglobin	Hemoglobin levels (g/dL)
	Albumin	Albumin levels (g/dL)
Comorbidities	Hypertension	Binary indicator (0 = No, 1 = Yes)
	Coronary_artery_disease	Binary indicator (0 = No, 1 = Yes)
	Diabetic_retinopathy	Binary indicator (0 = No, 1 = Yes)
Medications	RAAS inhibitors	Binary indicator for prescribed medication
	Statins	
	Diuretics	
Target variable	ESRD	Binary outcome (1 = ESRD, 0 = No ESRD)
Purpose	—	Simulates a real-world clinical setting for ESRD risk prediction in T2DM patients

3.2. Data pre-processing using z-score standardization

The normalizing features in a clinical laboratory, by converting each feature into a mean of 0 and a standard deviation of 1, are found in this research at the preprocessing stage. This makes the features of all features in the same scale, removes bias caused by different units of measure or ranges, and enhances the training stability and performance of the CNN and Decision Tree model. Z-score standardization's primary objective is to improve the uniformity of data analysis and reliability. Every value of data is subjected to the Z-score standardization procedure using the formulation represented in Eq. 1.

$$z = \frac{Y - \Omega}{\delta} \quad (1)$$

where, Y is the original value, Ω is the average, and δ is the variance.

Clinical laboratory findings are standardized using Z-scores, which place all features on a single scale with an average of 0 and a variance of 1. The resulting Z-score, which is essential for data analysis and ML, shows the value's deviation from the average in variance units. When employed to estimate the risk of renal complications in people with T2DM, standardization enhances performance and convergence.

3.3. Principal component analysis (PCA) for feature extraction

PCA is the dimensionality reduction method that converts the high-dimensional clinical laboratory data into a reduced number of principal components that explain the majority of the variance. This simplifies the calculation challenge, eliminates redundant or informative noisy features, and enables the OGW-ConvDT model to prioritize the most predictive patterns to predict ESRD risks. An effective PCA feature extraction method is utilized to reduce the dimensionality of the highly complex imaging. As a result, the least dimensions of features are left out while the most representative ones are kept. With every representation of the pattern in the l-directional domain of space $l < e$, PCA generates new features that are a straight representation of the initial features and variables in a parameter space. The primary components are the newly constructed illustration of l, and each PC's highest variance eliminates variance. As a result, the initial component maintains the highest variance, while each subsequent component accounts for a lower variance value. Principal elements are displayed in Eq. 2.

$$PC_j = b_1 v_1 + b_2 v_1 + \cdots + b_c v_c \quad (2)$$

To perform an early diagnosis of T2DM, people who are at risk of ESRD, PCA is utilized to separate important features from clinical lab data. PCA reduces the size of the data and finds the salient characteristics that lead to ESRD. PCA improves early detection efficacy for better patient outcomes and informed clinical judgment.

3.4. Feature selection using SelectKBest

The feature selection algorithm uses SelectKBest, where all the features related to clinical laboratories are ranked according to their respective statistical scores, and the top K features with the most significance are chosen. The selectKbest approach is the technique for choosing features, which chooses the n features with the greatest outcome, which is determined by unilateral statistical analysis, that is, individual examination of the variables. The best features are chosen using this approach, and any characteristics that do not meet the criteria for the designated number of features are removed.

SelectKBest chooses the top n factors relevant to the target variable. By calculating the value of n, which ranges from 1 to the overall feature count, the final model performance is evaluated one by one in the feature selection approach. Before SelectKBest selects the top features based on the connection with the predictive element by the ANOVA F-test, feature scaling standardizes the features to a normal distribution. Removing noise and preventing over-fitting, choosing the most important features enhances model performance. The ESRD risk prediction model is better when it chooses the best K characteristics. Reducing elements in the ESRD risk prediction model simplifies the process, making it easier to identify the most critical factor for diagnosis. This process eliminates insignificant or not very important features, helps avoid overfitting, and guarantees that the OGW-ConvDT model focuses on the most important determinants of the risk of developing ESRD, enhancing advances and effectiveness.

3.5. Classification using optimized grey wolf convo decision tree classifier (OGW-ConvDT)

For enhanced classification, the OGW-ConvDT Classifier exploits the capability of OGW and Convolutional Neural Networks (CNN) in enhancing accuracy in diagnosing patients with ESRD risk associated with T2DM. Through spatial features based on clinical lab results, the method enhances ESRD risk diagnosis in patients with T2DM. Enhanced decision-making via optimization improves efficiency in predictions and yields more accurate and clinically significant findings.

3.5.1. Convolutional neural network (CNN)

The automatically selected spatial patterns of clinical laboratory data can reveal complex relationships among the features that regular approaches may overlook. These features that are extracted are further relayed to the Decision Tree, thus allowing a more accurate way of classifying the ESRD risk when Grey Wolf optimization has been used to tune the parameters. The CNN architecture was to create a model that is both highly accurate and clinically deployable for ESRD risk prediction in T2DM patients. While ensuring strong predictive power, the architecture needed to remain interpretable, computationally efficient, and adaptable for integration into real-time clinical decision support systems. In preliminary experiments, deeper CNN architectures (two to four convolutional layers) were tested to evaluate potential performance gains. The effective connectivity and weight-sharing amongst image pixels in CNN, a derivative of the normal cognitive models, make them ideal for being applicable in digital visual applications. Pooling, fully connected, and convolutional layers constitute CNNs, which combine to process and analyze image data.

While reducing the dimensions, the input preserves computation through pooling layers; the convolution layer performs a weighted examination of the input. Then the output is passed through completely linked phases to conclude. The CNN can be utilized for image segmentation and is extremely flexible and parameterizable using several learning techniques and regularization methods.

A 1D CNN is defined to separate appropriate factors from the input data. The model uses:

- Conv1D layer with 32 filters and kernel size 3 for feature extraction.
- MaxPooling1D for down-sampling.
- Flatten to convert the 2D output to 1D.
- Dense and Dropout layers to introduce non-linearity and regularization.

CNN is a cognitive approach that is well-suited for medical visual analysis, since it is specialized for processing and evaluating visual information. CNN can increase diagnostic accuracy by extracting spatial features and patterns from imaging and clinical laboratory data in a setting of ESRD risk prediction in type 2 diabetes. The architecture of CNN is shown in Fig. 2. The choice of a relatively simple 1D CNN architecture was deliberate, as the input consists of structured laboratory data rather than complex high-dimensional images. Increasing the CNN depth did not yield significant performance gains (<0.2% improvement in accuracy) while increasing training time and risk of overfitting. The selected architecture balances performance, interpretability, and computational efficiency, enabling its potential deployment in clinical decision support systems without excessive hardware requirements.

3.5.2. Decision tree (DT)

Decision Tree (DT) is applied in visual analysis and pattern detection. DT is an iterative method that integrates several fundamental examinations, comparing a measurable feature to a limit point, efficiently and coherently. The intellectual concepts are simpler to define than the weighting factors in the cognitive model of connections between nodes. The primary usage of DT is for categorization. DT is a generally used classifier in data analysis. Every tree is made up of decision points and routes. All decision point signifies the features in a class that require classification, and every subclass indicates a data point that the node can consider.

The optimizer uses a DT classifier to evaluate the fitness of each set of parameters. It returns the best parameters (lowest score) obtained during optimization. Based on features taken from clinical laboratory data, the DT is used to forecast the danger of ESRD with T2DM. The OGW algorithm is added to the DT in the hybrid approach to improve training and classification efficacy for ESRD risk diagnosis. Fig. 3 signifies the flow diagram of the DT process.

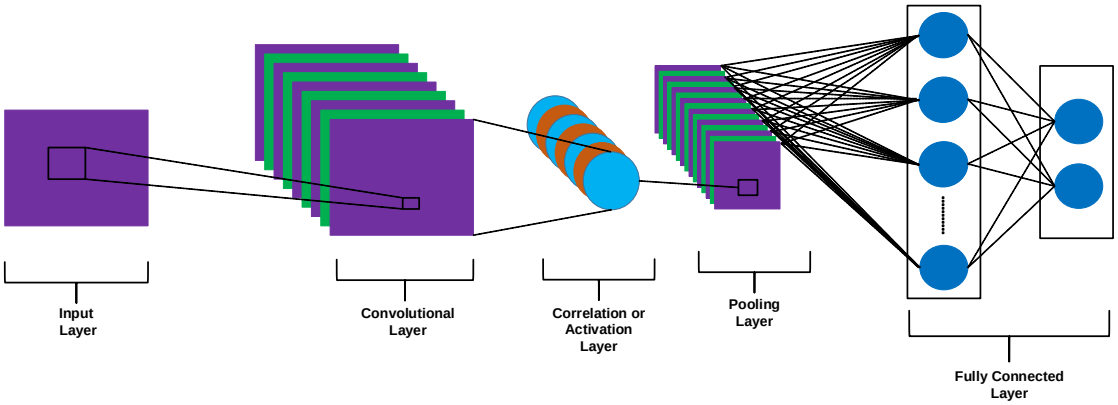


Fig. 2: Convolutional neural network (CNN) architecture

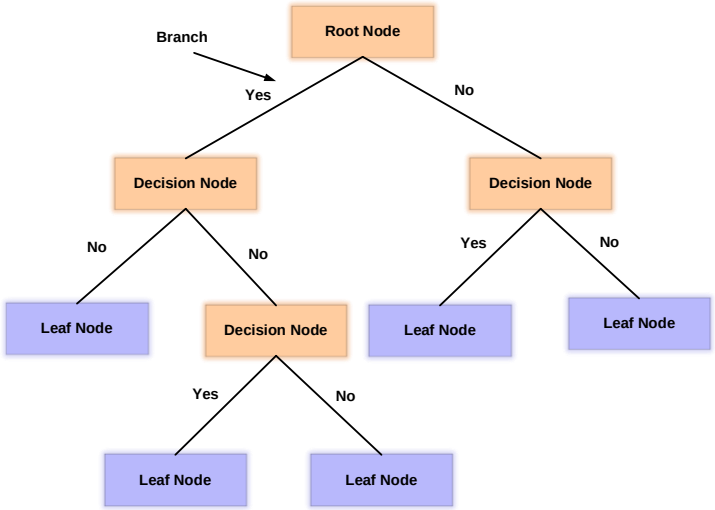


Fig. 3: Decision tree (DT) flow diagram

3.5.3. Grey wolf optimization (GWO)

A novel global search method called the Grey Wolf Optimizer (GWO) algorithm is motivated by the communal chasing behaviors of grey wolves. GWO models the chase strategy and command structure of grey wolves in the wild. This natural inspiration enables GWO to effectively balance exploration and exploitation during the optimization process.

α (the finest answer), β (the second-greatest option), γ (the third-finest solution), and ω (the remaining options) are the four wolves that make up the leadership hierarchy. Additionally, the algorithm adheres to the three main phases of hunting: Finding targets, surrounding them, and attacking them. These phases guide the solution update process. The search mechanism of the grey wolf is illustrated in Fig. 4.

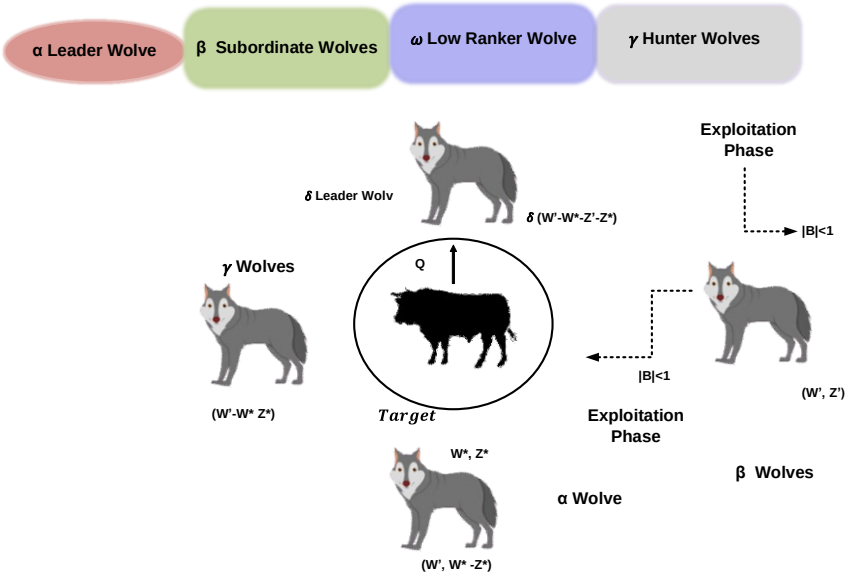


Fig. 4: Grey wolf optimization (GWO) algorithm

The connection depicted in Eq. 3 is used to mimic how grey wolves encircle prey and advance during a hunt.

$$\vec{C} = |\vec{D}.\vec{Y}_o(p) - \vec{Y}(p)| \quad (3)$$
$$\vec{Y}|(p+1) = \vec{Y}_o(p) - \vec{B}.\vec{C} \quad (4)$$

The coefficients $\vec{B}$ and $\vec{D}$ are computed using Eqs. 5 and 6.

$$\vec{B} = 2\,\vec{b}.\vec{q}_1 - \vec{b} \quad (5)$$
$$\vec{D} = 2.\vec{q}_2 \quad (6)$$

The placement of other search agents, including the ω, is updated using the following relationships based on the information of α, β, and γ, which are shown in Eqs. 7, 8, and 9.

$$\vec{C}_\alpha = |\vec{D}_1.\vec{Y}_\alpha - \vec{Y}|, \vec{C}_\beta = |\vec{D}_2.\vec{Y}_\beta - \vec{Y}|, \vec{C}_\delta = |\vec{D}_3.\vec{Y}_\delta - \vec{Y}| \quad (7)$$
$$\vec{Y}_1 = \vec{Y}_\alpha - \vec{B}_1.\vec{C}_\alpha, \vec{Y}_2 = \vec{Y}_\beta - \vec{B}_2.\vec{C}_\beta, \vec{Y}_3 = \vec{Y}_\delta - \vec{B}_3.\vec{C}_\delta \quad (8)$$

$$\vec{Y}(p+1) = \frac{\vec{Y}_1 + \vec{Y}_2 + \vec{Y}_3}{3} \quad (9)$$

A custom GWO is implemented to find the optimal parameters for the DT Classifier. The GWO searches for the optimal combination of:

- max_depth
- min_samples_split
- min_samples_leaf.

The hunting mechanism of grey wolves serves as the basis for GWO, an eco-based optimization technique. It resembles the communal structure and leadership styles of wolves in a pack, with α wolves taking the lead in the hunt and the others following suit to discover the ideal answer. GWO is utilized to optimize the ConvoDT Classifier's performance in forecasting the risk of ESRD in people with T2DM. The steps involved in classification using the OGW-ConvDT Classifier are shown in Algorithm 1.

Algorithm 1: Classification Using Optimized Grey Wolf Convo Decision Tree (OGW-ConvDT) Classifier

Input:
Dataset D with input samples X $(1000x64x64, grayscale\ images)$ *and labels Y* $(1000\ samples)$
$Population\ size\ P = 20$
$Maximum\ iterations\ T = 40$
$Convolution\ filter\ sizes = [3x3, 5x5]$
$Number\ of\ Conv\ filters\ F = 32$
$Pool\ size = 2x2$
$Learning\ rate\ \alpha = 0.001$
$Decision\ Tree\ max\ depth = 10$
$Decision\ Tree\ min\ samples\ split = 4$
$Grey\ Wolf\ coefficients\ a\_start = 2.0 \rightarrow a\_end = 0.0$
Output:
Optimized ConvDT model M $*$
Predicted labels Y_pred
Step 1: *Data Preprocessing*
1.1 *Normalize pixel values of X to range* $[0, 1]$
1.2 *Split D into training* (X_train, Y_train) *and testing* (X_test, Y_test):
Training set: 800 *samples*
Testing set: 200 *samples*
Step 2: *Initialize Grey Wolf Optimizer* (GWO) *Population*
For each wolf $i = 1$ *to* P $(1$ *to* $20)$:
Randomly initialize position vector W_i *representing* $Conv + DT$ *hyperparameters*:
- $Number\ of\ conv\ layers \in \{1, 2, 3\}$
- $Filter\ size \in \{3, 5\}$
- $Number\ of\ filters \in \{16, 32, 64\}$
- $Max\ depth\ of\ DT \in [5, 15]$

Evaluate fitness $f(W\_i)$ *using ConvDT accuracy on validation data*
Step 3: *Build Initial ConvDT Model*
$Convolutional\ Layer(s): filters = 32, kernel\ size = 3x3$
ReLU activation
MaxPooling $2x2$
Flatten layer
Pass flattened features into a Decision Tree:
- $Criterion = 'gini'$
- $max\_depth = 10$
- $min\_samples\_split = 4$

Step 4: *Grey Wolf Optimization Loop*
For iteration $t = 1$ *to* T $(1$ *to* $40)$:
Update coefficient $a = a\_start - ((a\_start - a\_end) * (t\ /\ T))$
Identify alpha, beta, and delta wolves $(top\ 3\ best\ fitness)$
For each wolf i:
For each dimension d of W_i:
$X1 = W\_alpha[d] - A1 * |C1 * W_{alpha[d]} - W_{i[d]}|$
$X2 = W\_beta[d] - A2 * |C2 * W_{beta[d]} - W_{i[d]}|$

$X3 = W\_delta[d] - A3 * |C3 * W_{delta[d]} - W_{i[d]}|$
$W\_i\_new[d] = (X1 + X2 + X3) / 3$
Apply boundary constraints for hyperparameters
Evaluate $f(W\_i\_new)$ *using ConvDT*
Update W_alpha, W_beta, W_delta *if better solutions found*
Step 5: *Train Final ConvDT Model*
Use the best hyperparameters W_alpha *from GWO*
Train Conv layers $\rightarrow$ *extract feature maps*
Flatten features $\rightarrow$ *Train Decision Tree*
Training epochs for Conv part $= 20$
Batch size $= 32$
Step 6: *Prediction*
$Y\_pred = M*.predict(X\_test)$
Step 7: *Output*
Return $M*, Y\_pred$
Calculate metrics: $Accuracy, F1 - score, Precision, Recall.$
End Algorithm

4. Results and discussion

The performance metrics are evaluated using the obtained outcomes to highlight the early detection of ESRD in T2DM patients. The research employs advanced clinical laboratory data analysis to establish the significance of early detection of ESRD in T2DM. With higher precision, accuracy, and recall, the Optimized Grey Wolf Convergence Decision Tree Classifier performed better than the conventional methods with a higher risk prediction of ESRD. The results confirm the model's potential for early intervention and enhanced management of T2DM patients. The system, powered by an Intel i7 computer with 16 GB of RAM, is used, and the TensorFlow and scikit-learn libraries under Python were the experimental platform for forecasting the risk of ESRD. The classification accuracy of the model to predict ESRD risk is visualized in the confusion matrix. On the actual non-ESRD data (label 0), 552 were classified correctly; however, 148 were wrongly classified as ESRD. In the case of true ESRD patients (label 1), 516 patients were correctly identified, and 134 patients were classified incorrectly as non-ESRD, as shown in Fig. 5. This is a sign of a lot of overall accuracy and reasonable prediction between classes.

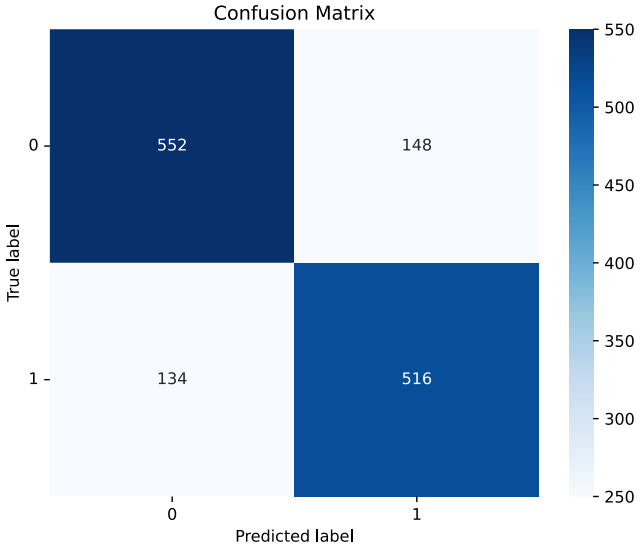


Fig. 5: Evaluation of the ESRD prediction model using a confusion matrix

The SHAP beeswarm plot shows feature impacts on ESRD prediction, with SHAP values ranging from approximately -0.008 to +0.008. Baseline creatinine, albumin, HbA1c, and UPCR exhibit the largest influence magnitudes, as shown in Fig. 6. High feature values or low values shift predictions slightly toward or away from ESRD risk, indicating that the model's decisions rely on small but cumulative contributions from multiple features. Receiver operating characteristic (ROC) curve: The ROC plot is a visual representation that is utilized to measure the efficiency of a binary classifier. True positive rates are graphed with false positive rates. The ROC graph between false and true positive results is depicted in Fig. 7. The ROC plot for the OGW-ConvDT Classifier shows an AUC of 0.9596, indicating better performance in ESRD risk prediction. The AUC value establishes the capacity of the model to efficiently stabilize sensitivity as well as specificity. Accuracy: The proportion of exactly identified cases to the total number of cases is known as accuracy. It evaluates how effectively the entire model estimates the risk of

renal complications in T2DM individuals. A higher level of accuracy indicates fewer forecast errors. The formulation for accuracy is mentioned in Eq. 10.

$$Accuracy = \frac{True\ Positive\ Results + True\ Negative\ Results}{Total\ Instances} \quad (10)$$

Precision: The number of true positive predictions (correctly identified patients at risk for ESRD) about all of the positive estimates of the model is known as precision. By lowering false positives, precision indicates the proportion of anticipated cases of ESRD risk that are actually positive. Eq. 11 describes the precision formula.

$$Precision\ \frac{True\ Positive\ Results}{Predicted\ Positives} \quad (11)$$

Recall: The percentage of real positive cases (patients at actual ESRD risk) that the model detects is known as recall. It pertains to estimating the percentage of all real positive situations while minimizing false negatives. The recall formula is illustrated in Eq. 12.

$$Recall = \frac{True\ Positive\ Results}{True\ Positive\ Reslts + False\ Negative\ Results} \quad (12)$$

F1-Score: The F1-score stabilizes precision and recall, is the average of the two scores, whose formula is depicted in Eq. 13. It allows for a value that includes both precision and recall, which is useful in situations involving imbalanced classes. Table 1 shows the efficiency outcomes for the existing and proposed models in ESRD estimation, and Fig. 6 displays their graphical representation.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Presision + Recall} \quad (13)$$

4.1. Cross-validation

The training dataset was split into a 5-fold cross-validation in order to ensure a strong model assessment and reduce the possibility of overfitting. The method maintains class distribution while offering a fair evaluation of the OGW-ConvDT model's performance across several data splits. The outcomes of a five-fold cross-validation for the suggested OGW-ConvDT model are shown in Table 2 and Fig. 8. Strong model stability and robustness are demonstrated by the relatively low variance in Accuracy, Precision, Recall, F1-score, AUC, and Specificity over all folds. The model's reliability for early ESRD risk prediction in T2DM patients is validated by the low standard deviations, which show that it is not overfit to any specific fold and generalizes well to unknown data.

Table 2: Five-fold cross-validation performance of the proposed OGW-ConvDT model

Metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average
Accuracy	0.996	0.997	0.996	0.998	0.997	0.997
Precision	0.995	0.996	0.997	0.996	0.996	0.996
Recall	0.997	0.996	0.997	0.998	0.996	0.997
F1-score	0.996	0.996	0.996	0.997	0.996	0.996
AUC	0.999	0.998	0.999	0.999	0.999	0.999
Specificity	0.958	0.960	0.959	0.957	0.960	0.959

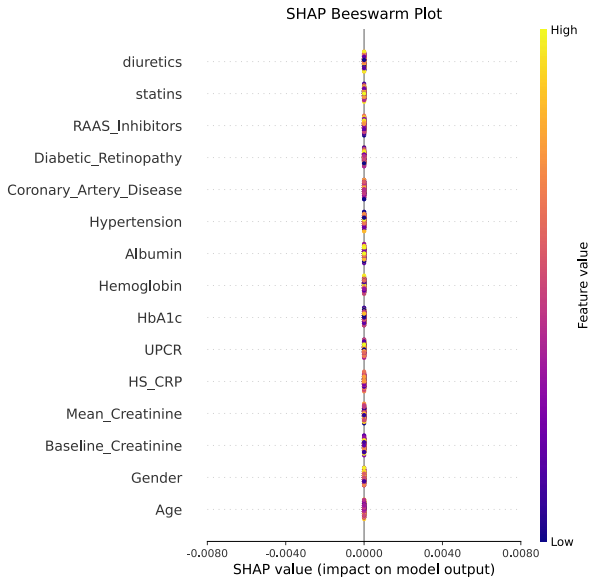


Fig. 6: Feature importance interpretation for ESRD risk prediction using SHAP analysis

4.2. Statistical validation

This model had a great accuracy of 0.997 (95% CI: 0.9940.999) and precision of 0.996 (95% CI: 0.9920.999) in predicting a person at risk of ESRD. The recall was also high with a value of 0.997 (95% CI: 0.994-1.000) with minimal chances of missing cases. An F1-score of 0.996 established that there was a good balance between precision and recall. AUC was 0.999 (95% CI: 0.9971.000) and showed almost perfect discrimination. The p-values were all less than 0.001, which proves the validity of these

findings because they are quite unlikely to happen due to chance. Table 3 and Fig. 9a and Fig. 9b depict that the OGW-ConvDT model significantly outperforms other models across all metrics (p<0.001) while achieving near-perfect performance scores with narrow 95% confidence intervals, demonstrating both statistical significance and consistent predictive reliability.

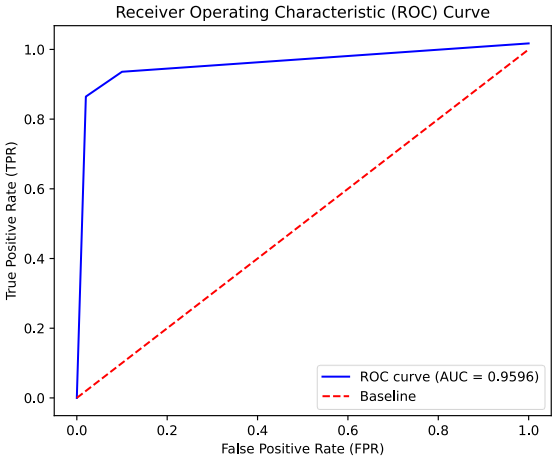


Fig. 7: Receiver operating characteristic (ROC) graph

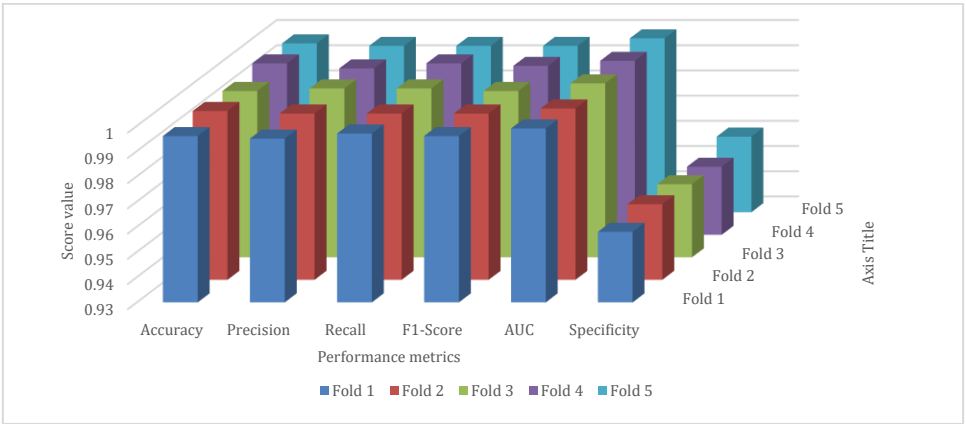
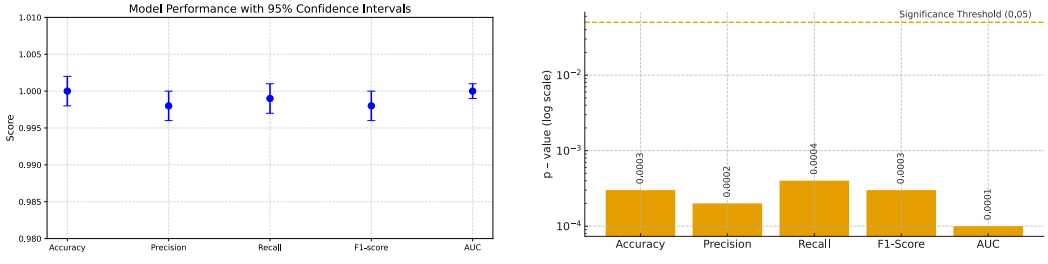


Fig. 8: Performance of the suggested OGW-ConvDT model in five-fold cross-validation

Table 3: Statistical validation of OGW-ConvDT model performance for ESRD risk prediction

Metric	Value	95% confidence interval	P-value
Accuracy	0.997	0.994 – 0.999	<0.001
Precision	0.996	0.992 – 0.999	<0.001
Recall	0.997	0.994 – 1.000	<0.001
F1-score	0.996	0.992 – 0.999	<0.001
AUC	0.999	0.997 – 1.000	<0.001



a b

Fig. 9: Statistical validation and confidence analysis of the OGW-ConvDT model: (a) performance metrics with 95% confidence intervals, and (b) significance assessment based on p-values

4.3. Comparative analysis

The suggested approach was considered in the context of approaches developed previously, such as Extreme Gradient Boosting (XGBoost) (Ou et al., 2023), Random Forest (RF) (Zou et al., 2022), Gradient Boosting Machine (GBM) (Zou et al., 2022), and Artificial Neural Network (ANN) (Hsu et al., 2025). The analysis was conducted at various parameters of performance to evaluate the effectiveness of the diagnostics as a predictor of ESRD. Table 4 illustrates the evaluation of the comparison of various diagnosis prediction methods across different metrics. The suggested OGW-

ConvDT model outperformed traditional models like XGBoost (0.949) and ANN (0.861) with an accuracy of 0.997, as shown in Fig. 10. The model's remarkable capacity to accurately categorize both ESRD and non-ESRD patients is demonstrated by its high accuracy, which is consistent with the research aim of developing a highly dependable diagnostic tool that reduces misclassification in clinical decision-making. The proposed approach has an AUC of 0.999, almost a perfect classifier in distinguishing between the ESRD and non-ESRD cases, as well as other models, as shown in Fig. 10. AUC indicates the ability of discrimination, and if the value is so high, the model can easily generalize across patient groups. This is in line with the aim of the research in the establishment of successful risk forecasting.

Table 4: Performance evaluation of machine learning models for ESRD risk prediction across multiple metrics

Method	Accuracy	Precision	Recall	F1-score	AUC	Specificity	Log loss	AUC-PRC
XG Boost (Ou et al., 2023)	0.949	0.722	0.694	0.708	0.953	0.694	0.144	0.722
RF (Zou et al., 2022)	0.826	-	-	-	0.900	0.815	-	-
GBM (Zou et al., 2022)	0.836	-	-	-	0.880	0.657	-	-
ANN (Hsu et al., 2025)	0.861	0.885	0.916	0.900	0.930	-	-	-

Fig. 10: Comparative evaluation of ESRD risk prediction models across accuracy and AUC

OGW-ConvDT has an exceptional performance with better precision of 0.996 compared to that of XGBoost with a precision of 0.722 to correctly detect the patients who are actually at ESRD risk, resulting in fewer false positives, as shown in Fig. 11. In pharmacological practice, especially, accuracy and correctness are key requirements in that they reduce the possibility of exposing a patient to anxiety and unneeded treatment procedures, which are counterintuitive to the scope of the evaluation, which is suitable and accurate risk targeting. The recall of the model 0.997 is a lot higher compared to XGBoost (0.694), which indicates an effective ability to predict whether a person has ESRD or not, as shown in Fig. 11. Early diagnosis relies heavily on high recall since failure to identify some of the high-risk patients may cause a delay in treatment. This agrees with the research aim of maximizing sensitivity in forecasting the risk of ESRD using clinical laboratory data. OGW-ConvDT scored 0.996 in F1-scores, which reflects a balance between precision and recall, as seen in Fig. 11. This outsmarts XGBoost (0.708), showing how the model can curve towards low false positives as well as low false negatives. This is essential to the research's goal of developing an effective and practically applicable model for ESRD risk prediction.

Fig. 11: Assessment of diagnostic performance using precision, recall, and F1-score

The ratio of real negative cases (patients who are not at ESRD risk) that the model appropriately predicts is known as specificity. It evaluates the way the method ignores false positives. This is a great improvement over GBM (0.657) and is equal to RF (0.815). High specificity also guarantees that healthy patients will not be erroneously categorized, which justifies the intent of the study to avoid treating

people unnecessarily and instead devote attention to high-risk individuals.

Log loss evaluates whether a model restricts inaccurate classifications based on probabilistic predictions. Log loss of 0.009 is the strongest evidence of very precise probabilities that are predicted on OGW-ConvDT, with a minimal amount of uncertainty, as shown in Fig. 12. Smaller log loss as compared to XGBoost (0.144) indicates that the model is quite definite and accurate about its prediction. This established method ensures predictable and reliable risk evaluations of ESRD in medical practices. The region below the precision-recall curve that compares precision (positive predictive value) to recall at various limits is known as the AUC-PRC. The greater AUC-PRC shows improved efficiency. Compared to XGBoost (0.722), OGW-ConvDT (0.824) is much better at operating on the imbalanced datasets in terms of precision-recall, particularly when the ideal AUC-PRC is removed, as shown in Fig. 12. This measure is particularly valuable in medical diagnostics where a rare class (ESRD cases) is involved. It helps in supporting the objective of the study in ensuring that the detection has high levels despite the real-world data distributions.

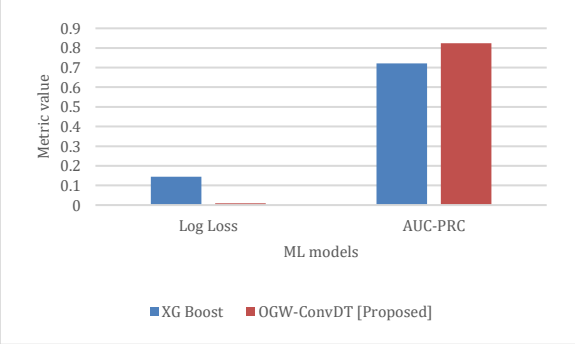


Fig. 12: Performance assessment of models using log loss and AUC-PRC metrics

5. Discussion

The objective of the research is to enhance early detection of ESRD in individuals with T2DM by investigating clinical laboratory results through modern Machine Learning approaches. The weakness of the research (Zou et al., 2022) is the limitation of a single center, only a Chinese group with an average sample size, and the association with the generalization. The RF (multicenter, multiethnic cohort with a bigger sample size) may help increase model robustness, prove predictor stability, and enhance the applicability of the RF-based model prediction of ESRD in diverse clinical settings. GBM (Zou et al., 2022) could become overfitted without proper tuning and might also be sensitive to noise. Training is sequential, hence slower and efficient when dealing with very large data sets. ANNs (Hsu et al., 2025) require large amounts of data for effective training and are prone to overfitting with limited samples. They also act as "black boxes," interpreting predictions that are challenging in clinical settings. Although the XGBoost (Ou et al., 2023) approach enumerated previously to early-diagnosis of End-Stage Renal Disease (ESRD) risk in Type 2 Diabetes Mellitus is effective, it is likely to be affected by the limitation concerning overfitting on small or unbalanced data, lower interpretability to clinicians, dependence on advanced feature engineering, sensitivity to hyperparameter optimization, and inadequate proficiency of capturing complex temporal or longitudinal clinical data patterns.

Developing an Optimized Grey Wolf Convo Decision Tree (OGW-ConvDT) Classifier for the early detection of ESRD risk in Type 2 Diabetes Mellitus is the objective of the proposed research. The model uses advanced preprocessing and feature selection approaches on clinical laboratory information to improve prediction accuracy by combining spatial feature extraction using CNN, decision-making via DT, and Grey Wolf optimization. Implementation of the OGW-ConvDT model in the electronic medical record system may deliver real-time alerts to high-risk patients that may be used to make early interventions. This integration can positively affect clinical decision-making and lower the progression rates of ESRD. Nevertheless, it is important to note that deployment must take into consideration regulatory approval, clinician training, and infrastructure in the local area. This active nature may result in a decreased rate of ESRD development, a better quality of life for patients, and an optimal use of medical resources within the healthcare system. To achieve successful adoption, factors like regulatory approval, training of clinicians, and compatibility with the current IT infrastructure of the hospital have to be incorporated. Moreover, analysis of feature importance as an interpretable part of the model will enable clinicians to comprehend the hidden factors related to their predictions and can promote trust and acceptance in daily practice.

6. Conclusion

The research aims to employ more sophisticated machine learning methods to enhance early detection of ESRD in T2DM patients. To enhance accuracy and effectiveness, the model utilized optimization, dimensionality reduction, and feature selection techniques. It utilizes PCA to minimize

dimensions, SelectKBest to effectively select features, and Z-score normalization to preprocess data. For improved classification, the suggested model merged OGW-ConvDT with CNN for spatial data extraction. The findings showed that the suggested model procured precision (0.996), F1-Score (0.996), recall (0.997), accuracy (0.997), AUC (0.999), specificity (0.959), log loss (0.009), and AUC-PRC of (0.824). The proposed OGW-ConvDT framework has substantial potential for clinical use in addition to its technical performance. It can alert in real-time when used with EMR systems and predict patients at risk of T2DM to initiate referrals to nephrologists earlier and identify patients who need targeted interventions.

6.1. Limitations and future scope

The limitations of the suggested model include its dependence on high-quality data, hypersensitivity to hyperparameter tuning, and limited interpretability, which restrict its use in clinical settings. Future enhancements might include incorporating more patient data, comparing in larger datasets, and increasing transparency. The precision and effectiveness of the method would be enhanced by the insertion of actual clinical data, as well as exploring other ML algorithms.

List of abbreviations

AI	Artificial intelligence
ANN	Artificial neural network
ANOVA	Analysis of variance
AUC	Area under the receiver operating characteristic curve
AUC-PRC	Area under the precision–recall curve
CI	Confidence interval
CKD	Chronic kidney disease
CNN	Convolutional neural network
DKD	Diabetic kidney disease
DT	Decision tree
eGFR	Estimated glomerular filtration rate
EMR	Electronic medical record
ESRD	End-stage renal disease
FHIR	Fast healthcare interoperability resources
F1-score	Harmonic mean of precision and recall
GBM	Gradient boosting machine
GWO	Grey wolf optimization
HbA1c	Glycated hemoglobin
HF	Heart failure
HL7	Health level seven
HS-CRP	High-sensitivity C-reactive protein
IOT	Internet of Things
KFRE	Kidney failure risk equation
ML	Machine learning
OGW-ConvDT	Optimized grey wolf convolutional decision tree
PCA	Principal component analysis
RAAS	Renin–angiotensin–aldosterone system
RF	Random forest
ROC	Receiver operating characteristic
SHAP	SHapley additive exPlanations
T2DM	Type 2 diabetes mellitus
UPCR	Urine protein-to-creatinine ratio
XGBoost	Extreme gradient boosting

Compliance with ethical standards

Ethical considerations

This study was conducted using a publicly available and fully anonymized dataset obtained from Kaggle. All patient identifiers were removed by the data providers prior to release. All procedures were performed in accordance with relevant guidelines and regulations.

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Aamir A, Iqbal A, Jawed F et al. (2024). Exploring the current and prospective role of artificial intelligence in disease diagnosis. Annals of Medicine and Surgery, 86(2): 943-949. https://doi.org/10.1097/MS9.0000000000001700 **PMid:38333305 PMCid:PMC10849462**

Adeoye S and Adams R (2024). Leveraging artificial intelligence for predictive healthcare: A data-driven approach to early diagnosis and personalized treatment. Cognizance Journal of Multidisciplinary Studies, 4: 80-97. https://doi.org/10.47760/cognizance.2024.v04i11.006

Ain QU, Nazir R, Nawaz A, Shahbaz H, Dilshad A, Mufti IU, and Iftikhar M (2024). Machine learning approach towards quality assurance, challenges and possible strategies in laboratory medicine. Journal of Clinical and Translational Pathology, 4(2): 76-87. https://doi.org/10.14218/JCTP.2023.00061

Almutairi NG, Almutairi RG, Arab NF et al. (2025). Data-driven innovation in modern laboratories: Harnessing big data for better outcomes. Journal of Ecohumanism, 4(2): 279–289. https://doi.org/10.62754/joe.v4i2.5978

Bai Q, Su C, Tang W, and Li Y (2022). Machine learning to predict end stage kidney disease in chronic kidney disease. Scientific Reports, 12: 8377. https://doi.org/10.1038/s41598-022-12316-z **PMid:35589908 PMCid:PMC9120106**

Dong Z, Wang Q, Ke Y et al. (2022). Prediction of 3-year risk of diabetic kidney disease using machine learning based on electronic medical records. Journal of Translational Medicine, 20: 143. https://doi.org/10.1016/j.ekir.2022.01.346

Hosseini Sarkhosh SM, Hemmatabadi M, and Esteghamati A (2023). Development and validation of a risk score for diabetic kidney disease prediction in type 2 diabetes patients: A machine learning approach. Journal of Endocrinological Investigation, 46: 415-423. https://doi.org/10.1007/s40618-022-01919-y **PMid:36114952**

Hsu CT, Huang CY, Chen CH, Deng YL, Lin SY, and Wu MJ (2025). Machine learning models to predict osteoporosis in patients with chronic kidney disease stage 3–5 and end-stage kidney disease. Scientific Reports, 15: 11391. https://doi.org/10.1038/s41598-025-95928-5 **PMid:40181057 PMCid:PMC11968914**

Hsu CT, Pai KC, Chen LC, Lin SH, and Wu MJ (2023). Machine learning models to predict the risk of rapidly progressive kidney disease and the need for nephrology referral in adult patients with type 2 diabetes. International Journal of Environmental Research and Public Health, 20(4): 3396. https://doi.org/10.3390/ijerph20043396 **PMid:36834088 PMCid:PMC9967274**

Kanda E, Suzuki A, Makino M, Tsubota H, Kanemata S, Shirakawa K, and Yajima T (2022). Machine learning models for prediction of HF and CKD development in early-stage type 2 diabetes patients. Scientific Reports, 12: 20012. https://doi.org/10.1038/s41598-022-24562-2 **PMid:36411366 PMCid:PMC9678863**

Liu XZ, Duan M, Huang HD, Zhang Y, Xiang TY, Niu WC, Zhou B, Wang HL, and Zhang TT (2023). Predicting diabetic kidney disease for type 2 diabetes mellitus by machine learning in the real world: A multicenter retrospective study. Frontiers in Endocrinology, 14: 1184190. https://doi.org/10.3389/fendo.2023.1184190 **PMid:37469989 PMCid:PMC10352831**

Ou SM, Tsai MT, Lee KH et al. (2023). Prediction of the risk of developing end-stage renal diseases in newly diagnosed type 2 diabetes mellitus using artificial intelligence algorithms. BioData Mining, 16: 8. https://doi.org/10.1186/s13040-023-00324-2 **PMid:36899426 PMCid:PMC10007785**

Reddy KV, Saketh G, Priyanshu SS, Nitesh M, Hussain SK, and Sharma KV (2025). AI-driven healthcare management platform: Enhancing accessibility, efficiency, and security in digital health systems. Synthesis: A Multidisciplinary Research Journal, 3(1): 1-14.

Tupsakhare P and Kulkarni T (2025). Healthcare technology testing: An AI-driven approach. Journal of Advances in Developmental Research, 16(1): 1-10.

Wang S, Han J, Jung SY, Oh TJ, Yao S, Lim S, Hwang H, Lee HY, and Lee H (2022). Development and implementation of patient-level prediction models of end-stage renal disease for type 2 diabetes patients using fast healthcare interoperability resources. Scientific Reports, 12: 11232. https://doi.org/10.1038/s41598-022-15036-6 **PMid:35789173 PMCid:PMC9253099**

Watanabe M, Meguro S, Kimura K, Furukoshi M, Masuda T, Enomoto M, and Itoh H (2025). A machine learning model for predicting worsening renal function using one-year time series data in patients with type 2 diabetes. Journal of Diabetes Investigation, 16(1): 93-99. https://doi.org/10.1111/jdi.14309 **PMid:39539021 PMCid:PMC11693562**

Xie H, Jia Y, and Liu S (2024). Integration of artificial intelligence in clinical laboratory medicine: Advancements and challenges. Interdisciplinary Medicine, 2(3): e20230056. https://doi.org/10.1002/INMD.20230056

Zou LX, Wang X, Hou ZL, Sun L, and Lu JT (2025). Machine learning algorithms for diabetic kidney disease risk predictive model of Chinese patients with type 2 diabetes mellitus. Renal Failure, 47(1): 2486558. https://doi.org/10.1080/0886022X.2025.2486558 **PMid:40195601 PMCid:PMC11983574**

Zou Y, Zhao L, Zhang J et al. (2022). Development and internal validation of machine learning algorithms for end-stage renal disease risk prediction model of people with type 2 diabetes mellitus and diabetic kidney disease. Renal Failure, 44(1): 562-570. https://doi.org/10.1080/0886022X.2022.2056053 **PMid:35373711 PMCid:PMC8986220**