ASEAN
Journal of Scientific and Technological Reports
Online ISSN: 2773-8752

Research article
Research article

Artificial Intelligence-Based Approach for Chronic Kidney Disease Detection

Gogulamudi Pradeep Reddy[1], Duppala Rohan[2], Kasaraneni Purna Prakash[3], Yellapragada Venkata Pavan Kumar[4*], and Maddikera Kalyan Chakravarthi[5]

[1] Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, Karnataka, India
[2] School of Computer Science and Engineering, VIT-AP University, Amaravati 522241, Andhra Pradesh, India
[3] Department of Computer Science and Engineering, Siddhartha Academy of Higher Education, Deemed to be University, Vijayawada 520007, Andhra Pradesh, India
[4] School of Electronics Engineering, VIT-AP University, Amaravati 522241, Andhra Pradesh, India
[5] Engineering Department, College of Engineering and Technology, University of Technology and Applied Sciences, Muscat, Oman
* Correspondence: pavankumar.yv@vitap.ac.in

Citation:
Reddy, P.G.; Rohan, D.; Prakash, P.K.; Kumar, P.V.Y.; Chakravarthi, K.M. Artificaial intelligence-based approach for chronic kidney disease detection. *ASEAN J. Sci. Tech. Report.* **2025**, *28*(5), e258012. https://doi.org/10.55164./ajstr.v28i5.258012.

Article history:
Received: February 22, 2025
Revised: July 10, 2025
Accepted: August 27, 2025
Available online: September 14, 2025

Abstract: Chronic Kidney Disease (CKD) is a long-term medical condition in which the kidneys gradually lose their ability to function properly. Early detection of CKD is crucial for preventing severe complications and improving patient outcomes. Traditionally, CKD diagnosis has relied on manual analysis of clinical parameters and laboratory tests, which often lack scalability and precision. Artificial Intelligence (AI), through machine learning algorithms, has transformed healthcare by enabling automated and accurate disease detection. Datasets play a pivotal role in developing AI-based diagnostic systems, as the quality and balance of data significantly influence model performance. The majority of existing research on CKD detection has focused on balanced datasets, where data samples are evenly distributed across classes, to recommend the most effective classifiers for detection. However, in real-world scenarios, datasets are often imbalanced, with minority classes underrepresented, leading to biased models and poor detection of critical cases. Therefore, adopting suitable techniques to handle these imbalances is necessary. In this context, this paper addresses the issue by evaluating the performance of various classifiers on both slightly imbalanced and severely imbalanced CKD datasets. Through comprehensive experimentation, the research identifies that Gradient Boosting Machine (GBM) demonstrates robust performance across both slightly and severely imbalanced datasets by achieving 99.25% ± 0.68 and 92.26% ± 2.23 testing accuracy, 100% and 90.79% ± 3.9 AUROC, 0.01 ± 0.01 and 0.39 ± 0.15 LR-, 64.98% and 84.54% ± 3.41 H – measure. This work emphasizes the need for adaptable classifiers that reflect real-world data, improving the reliability of AI-based CKD diagnosis.

Keywords: Artificial intelligence in healthcare; Chronic Kidney Disease (CKD); Imbalanced Datasets; Machine Learning for Disease Detection; Gradient Boosting Machine (GBM)

1. Introduction

Chronic Kidney Disease (CKD) is a progressive medical condition characterized by the gradual loss of kidney function over time, preventing the kidneys from effectively filtering waste, excess fluids, and toxins from the blood. This condition affects millions of people globally and poses a significant burden

on healthcare systems. CKD is often referred to as a "silent disease" because symptoms typically do not appear until the disease has progressed to advanced stages, making early detection critical to prevent severe complications such as cardiovascular diseases or kidney failure. The causes of CKD are diverse, with diabetes and hypertension being the most common reasons. Diabetes, characterized by elevated blood sugar levels, can damage the tiny blood vessels in the kidneys, impairing their ability to filter blood. Similarly, high blood pressure puts excessive stress on these delicate vessels, leading to functional decline. Other causes include glomerulonephritis, an inflammatory condition affecting the kidney's filtering units; polycystic kidney disease, a genetic disorder characterized by cyst formation; recurrent kidney infections; and the long-term use of nephrotoxic medications, such as certain painkillers. These factors, alone or in combination, contribute to the progression of CKD. While CKD progresses silently in its initial stages, a range of symptoms may appear as kidney function deteriorates. These include persistent fatigue, swelling in the extremities or face due to fluid retention, nausea, vomiting, changes in appetite, and cognitive difficulties such as trouble concentrating. Changes in urination patterns, including increased frequency, altered color, or the presence of foamy urine, may also occur. Advanced stages may lead to severe complications, including electrolyte imbalances, anemia, and shortness of breath, significantly impacting the patient's quality of life. In addition to medical causes, lifestyle factors such as a high-sodium diet, lack of physical activity, obesity, and smoking significantly increase the risk of CKD. Environmental factors, including exposure to heavy metals, contaminated water, and air pollution, have also been linked to kidney dysfunction, particularly in regions with poor environmental regulations.

CKD is categorized into 5 stages as shown in Figure 1, based on the glomerular filtration rate (GFR), a measure of how well the kidneys filter blood. In Stage 1, kidney function is normal but with mild damage, and the GFR is 90 mL/min or higher. Stage 2 indicates a mild decline in kidney function, with GFR between 60 and 89 mL/min. Stage 3 reflects a moderate decline, with GFR ranging from 30 to 59 mL/min. By Stage 4, the disease is severe, with GFR between 15 and 29 mL/min, and Stage 5 signifies end-stage renal disease (ESRD), where the GFR drops below 15 mL/min, necessitating dialysis or transplantation [1]. Several diagnostic methods are employed to detect and monitor CKD [2]. Blood tests, including measurements of serum creatinine and blood urea nitrogen, are used to assess kidney function. Urine tests are used to detect abnormalities such as proteinuria or hematuria, which can indicate kidney damage. Imaging techniques, such as ultrasound and CT scans, help identify structural abnormalities or obstructions, while kidney biopsies provide a detailed analysis of kidney tissue to determine the extent of damage [3]. Estimates of GFR are critical for staging CKD and monitoring its progression. However, these diagnostic methods have several limitations. Late diagnosis is common due to the asymptomatic nature of early-stage CKD. Some methods, like biopsies, are invasive and carry risks. Cost and accessibility remain barriers, particularly in resource-constrained settings, while manual interpretation of test results introduces subjectivity and potential errors.

Artificial Intelligence (AI) offers promising solutions to these challenges [4]. By leveraging advanced machine learning algorithms, AI systems can analyze large and complex datasets with precision, enabling the early detection of subtle patterns in clinical data. Additionally, Explainable AI (XAI) is growing as a crucial component in healthcare applications, addressing the "black-box" nature of traditional AI models [5]. The Internet of Things (IoT) can play a transformative role by enabling real-time data collection from patients [6]. IoT devices, such as wearable sensors, smart health monitors, and connected medical equipment, can continuously track vital parameters like blood pressure, blood sugar levels, heart rate, and kidney-related biomarkers [7-8]. These devices can transmit the collected data to centralized systems for analysis, allowing for early identification of abnormalities and potential disease progression. The IoT can also facilitate remote monitoring, enabling healthcare providers to track patients' health without requiring frequent hospital visits, which is especially beneficial for individuals in remote or underserved areas. These systems enhance diagnostic accuracy, minimize human error, and facilitate scalable solutions that can handle large patient populations. By leveraging the data collected, AI supports personalized treatment by identifying patient-specific risk factors and predicting disease progression. Moreover, AI-powered tools can address real-world challenges, such as imbalanced datasets, which often result in biased models and poor detection of critical cases [9-10]. In this view, this paper evaluates various classifiers for both slightly imbalanced [11] and severely imbalanced datasets [12].

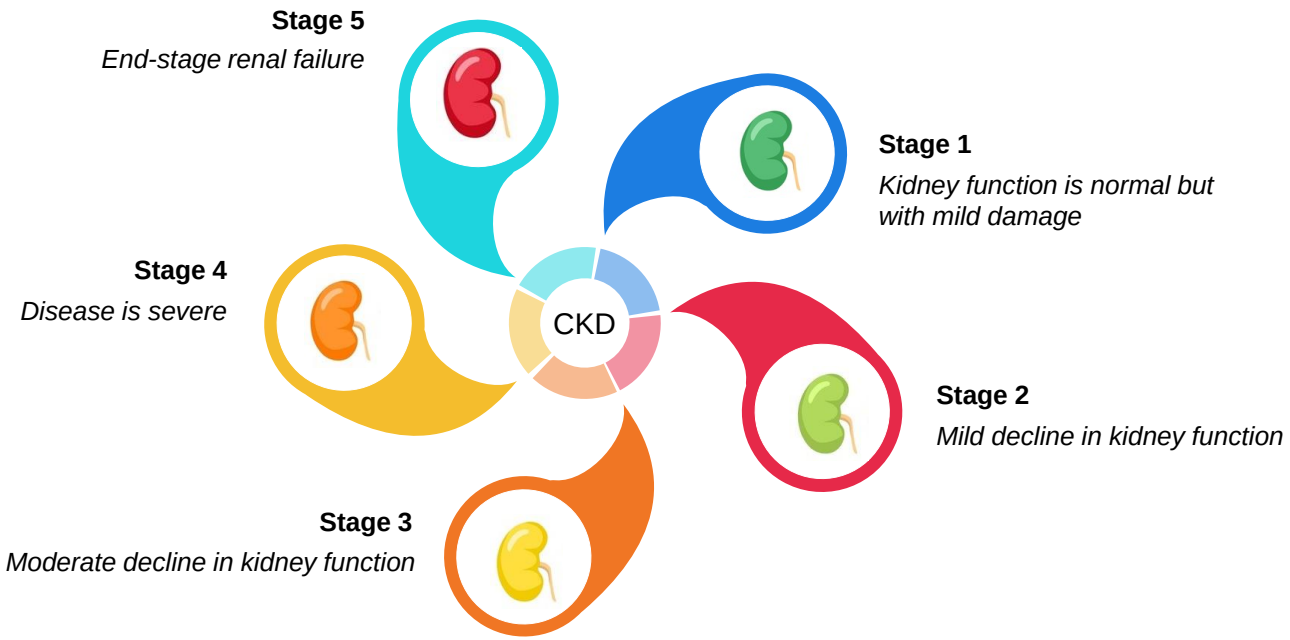

Figure 1. Various stages of CKD.

The rest of this paper is organized as follows: Section 2 provides an overview of related works, and Section 3 describes the proposed methodology in detail. Section 4 presents the results and discussion, analyzing the performance of various classifiers. Finally, Section 5 concludes the paper by summarizing the key findings and suggesting potential avenues for future research in the detection of CKD.

2. Related Work

This section provides a comprehensive review of the existing literature on the detection of CKD using various intelligence-based techniques. It highlights the progress made, key methods and approaches, and significant findings in this domain. A summary of key works in the literature is presented in Table 1. The prediction and early detection of CKD have gained significant attention in the research community due to CKD's high prevalence and associated complications. Various machine-learning models have been explored to enhance prediction accuracy and enable timely interventions [13]. Several comparative analyses have evaluated the performance of machine learning classifiers, such as Random Forest, SVM, XGBoost, and k-Nearest Neighbors [14]. Furthermore, deep learning frameworks, such as convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and ensemble methods, have been proposed for CKD prediction [15-17]. The integration of advanced feature engineering techniques, including recursive feature elimination and optimization strategies like Bayesian tuning, has further enhanced the predictive capabilities of these models. Additionally, several studies have investigated the impact of data augmentation and resampling techniques. In addition to improving prediction accuracy, researchers have also focused on addressing challenges such as data privacy [18]. Privacy-preserving models using encrypted data through the Paillier homomorphic cryptosystem ensure secure predictive analysis without compromising patient confidentiality. Meta-heuristic feature selection techniques, including genetic and bat algorithms, have been used to optimize model performance and reduce computational complexity. Despite significant advances, most studies focus on balanced datasets, where each class is equally represented. However, real-world datasets often exhibit class imbalance, affecting the detection of minority cases. This imbalance affects the classifier's performance and restricts the identification of critical conditions. To address this limitation, this research evaluates classifiers on CKD datasets with both slight and severe class imbalance. The deployment of machine learning models into clinical practice has seen significant progress through frameworks like DEPLOYR [19], which enables seamless integration of real-time predictive models into electronic medical records, aligning model outputs with clinical workflows. Beyond technical integration, real-world

performance monitoring is essential for finding valuable insights about model behavior post-deployment and identifying the need for ongoing surveillance [20]. The SHIELD-RT randomized study further addresses the practical challenges of deploying ML tools in a clinical setting, highlighting barriers such as clinician adoption, workflow alignment, and infrastructure readiness [21]. Multimodal AI frameworks have been proposed to support more generalizable and robust clinical applications by integrating diverse data sources, such as imaging, text, and structured data, into a unified system for decision support [22].

Table 1. Summary of various key works in CKD.

Ref.	Year	Objective	Methods/Approach	Relevance	Key Findings
[23]	2024	To classify and segment kidney MRI images for timely CKD detection and diagnosis.	Utilized Vision Transformers (EfficientNet_b1) for classification and ResNet18-Self-ONN-UNet++ for segmentation with CLAHE preprocessing and STAPLE post-processing.	Achieved 94.38% classification accuracy and superior segmentation performance, enhancing early CKD diagnosis using MRI data.	EfficientNet-b1 achieved 94.38% accuracy; the segmentation model achieved an IoU of 82.34% and a DS of 91.57%.
[24]	2024	To detect CKD using feature selection and hyperparameter tuning strategies for improved model performance.	Introduced Eurygaster Optimization Algorithm (EOA) with an ensemble of LSTM, BiGRU, and BiLSTM models. Applied SFLA for optimal hyperparameter selection.	Demonstrated superior CKD detection performance on benchmark datasets, emphasizing the importance of ensemble techniques.	EOAEDL-CKDD achieved optimal detection rates with improved accuracy and efficiency for early CKD detection.
[25]	2024	To explore non-invasive detection methods for medial vascular calcification (mVC) in CKD patients.	Analyzed peripheral pulse waveforms in the frequency domain, combined with traditional risk factors, to build predictive models using ensemble techniques.	Achieved 93% accuracy and an F-score of 0.96, highlighting the feasibility of non-invasive mVC detection in CKD management.	The ensemble model achieved an AUC ROC of 0.91, demonstrating the potential for non-invasive mVC detection in CKD patients.
[26]	2024	Introduces explainable AI for CKD diagnosis to address the limitations of "black-box" models and enhance interpretability.	Evaluated six machine learning algorithms, with extreme gradient boost (XGB) producing the best performance. SHAP and PDP were used to provide explainable predictions. Developed a graphical user interface (GUI) to assist healthcare professionals.	Pioneers the use of explainable AI in CKD diagnosis, aiding healthcare professionals in making informed decisions with interpretable insights.	XGB achieved the highest accuracy, and the SHAP and PDP methods provided transparent predictions, improving clinical utility and decision-making.

Table 1. Summary of various key works in CKD. (Continue)

Ref.	Year	Objective	Methods/Approach	Relevance	Key Findings
[27]	2023	Investigates the potential of various machine learning approaches for the early detection of CKD.	Predictive modeling was employed in conjunction with 12 machine learning classifiers, including XGBoost, to identify the most relevant features. The model evaluated 25 variables and identified the top 30% using predictive analytics.	Demonstrates the potential of predictive modeling and ML to enhance CKD diagnosis, reducing computational complexity and improving predictive accuracy.	XGBoost achieved the highest performance with 98.3% accuracy, precision, recall, and F1-score. Predictive modeling effectively identified crucial features for CKD diagnosis.
[28]	2023	To develop a novel hybrid deep learning network for early CKD detection and prediction.	Utilized DSCNN and Capsule Network with feature selection by Aquila Optimization. STOA optimized the DSCNN for classification.	Combines deep learning with IoT for enhanced CKD detection.	The proposed model outperformed state-of-the-art methods, demonstrating better classification accuracy with lower computational effort.
[29]	2021	To develop affordable machine learning models for early CKD diagnosis using selective clinical test attributes.	Evaluated models with k-fold cross-validation on optimized datasets derived from low-cost clinical test attributes.	Reduces diagnosis costs for CKD in resource-limited settings.	Optimized datasets and the random forest classifier achieved high accuracy, making it effective for CKD screening.
[30]	2020	To assess the applicability of ML techniques for CKD diagnosis in low-income, developing countries.	Conducted systematic reviews and experiments with ML techniques using k-fold cross-validation on CKD datasets.	Tailors CKD diagnostic approaches to developing countries.	The J48 decision tree achieved 95% accuracy, offering interpretable results suitable for low-resource settings.
[31]	2020	Automate CKD detection using salivary analysis and time-series data.	1-D Correlational Neural Network (CorrNN) and Bidirectional LSTM; tested with CKD sensing module.	Offers a novel salivary-based diagnostic tool for CKD.	Achieved an average accuracy rate of 98.08%, advancing CKD detection methodologies significantly.

3. Methodology

The experiment utilized two datasets: slightly imbalanced and severely imbalanced. The slightly imbalanced dataset was taken from the UCI Machine Learning Repository, and the severely imbalanced dataset was taken from Kaggle [11-12]. The slightly imbalanced dataset was collected over 2 months from a hospital, consisting of 24 features and a target variable. The target variable has two classes: 250 positive cases and 150 negative cases. The severely imbalanced dataset was collected from a Hospital in the United Arab Emirates, which consists of 21 features and a target variable with 56 positive and 435 negative cases. The dataset included the samples of 241 men and 250 women. The slightly imbalanced dataset consists of numerical and categorical columns. Imputation techniques were used to handle missing values. Missing numerical values were replaced with random samples from the same column. Missing categorical values were

filled using the mode of the same column. Inconsistent entries in categorical columns were removed and replaced with clean values. At last, label encoding was used to convert categorical columns into numeric values, ensuring that each unique value in a column is mapped to a corresponding numeric label, and to prepare the dataset for machine learning and deep learning algorithms. SMOTE-Tomek was employed to handle the class imbalance in the datasets. SMOTE works by generating synthetic examples for the minority class based on feature space similarities between existing minority samples. This helps to reduce bias in the model's predictions by ensuring better representation of minority instances during training. Tomek links are pairs of instances from opposite classes that are each other's nearest neighbors. The presence of a Tomek link indicates ambiguity or noise at the class boundary. Removing such pairs helps sharpen the decision boundary and improve the classifier's ability to distinguish between classes. The 80% dataset was used for training the classifiers, and the 20% dataset was used for testing the classifiers' performance. Five-fold cross-validation was employed using the StratifiedKFold method from scikit-learn's model_selection module to ensure that each fold maintains approximately the same proportion of class labels. Each fold was indexed and tracked using a variable fold, initialized at 1 and incremented during each iteration. Additionally, a small constant epsilon (= 1e-10) was defined to avoid division by zero or numerical instability in metric calculations, particularly in cases where denominators could become very small.

A total of 16 machine learning and 5 deep learning classifiers were implemented. Machine learning classifiers include Gaussian Naïve Bayes (GNB), Bernoulli's Naïve Bayes (BNB), Complement Naïve Bayes (CNB), Multinomial Naïve Bayes (MNB), Decision Tree (DT), Extra Tree Classifier (ETC), K Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Stochastic Gradient Descent (SGD), Light Gradient Boosting Machine (LGBM), Histogram-based Gradient Boosting (HGB), Adaptive Boosting (ADB), Gradient Boosting Machine (GBM), Categorical Boosting (CB), and Extreme Gradient Boosting (XGB). Deep learning classifiers include Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Gated Recurrent Unit (GRU), Bidirectional GRU (BiGRU), and Recurrent Neural Network (RNN). LSTM, GRU, and RNN architectures share a similar sequential structure. Each classifier consisted of five layers with decreasing numbers of neurons – 128, 64, 32, 16, and 8 - utilizing the ReLU activation function. The five bidirectional layers in each model have decreasing numbers of units – 128, 64, 32, 16, and 8 - with the first layer specifying the input shape and utilizing the tanh activation, while subsequent layers use the ReLU activation. All five deep learning classifiers include a dropout layer with a dropout rate of 0.2 to prevent overfitting, as well as a final dense layer with a sigmoid activation function for predicting the disease. The proposed GBM model's hyperparameters are tabulated in Table 2.

Table 2. Parameters of GBM

S. No.	Hyper-Parameter	Value
1	loss	log_loss
2	learning_rate	0.1
3	n_estimators	200
4	subsample	0.8
5	criterion	friedman_mse
6	max_depth	7
7	min_samples_split	4
8	min_samples_leaf	2
9	max_features	log2
10	random_state	42
11	verbose	1

Thirteen metrics were considered for evaluating the classifiers. These metrics include testing accuracy, precision, sensitivity, F1 score, F2 score, specificity, balanced accuracy, Mathews Correlation Coefficient (MCC), AUROC, AUPRC, Likelihood Ratio Positive (LR+), Likelihood Ratio Negative (LR-), and H-measure. Testing accuracy can be used as a baseline metric, but it is often misleading in imbalanced datasets. So, other metrics must be considered for balanced evaluation. Precision and sensitivity were included to ensure the

classifier does not over-predict the minority class and to measure how well the model captures the minority class. In imbalanced datasets, the F1 score is highly relevant, as it balances both sensitivity and precision. The F2 score gives more preference for sensitivity and is crucial when false negatives are more important than having false positives. While sensitivity focuses on the minority class, specificity evaluates how well the classifier identifies the majority class. Balanced accuracy measures the classifier's performance on both the minority and majority equally by averaging sensitivity and specificity. MCC is useful for imbalanced datasets as it evaluates the correlation between predictions and accurate labels by considering all true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). AUPRC is more useful than AUROC in imbalanced datasets. LR+ helps understand the strength of positive predictions and avoids over-reliance on the majority class bias. LR- ensures the classifier can eliminate cases as the majority class while correctly identifying most minority class cases. The H-measure is especially useful for imbalanced datasets because it allows for the weighting of false positives and negatives based on their respective importance in the specific problem domain. The H-measure was calculated using the trapezoidal rule and then normalized to the range of [0, 1]. For better interpretability, the results were expressed as percentages. It is considered an alternative to AUROC.

4. Results and Discussion

This section analyzes and compares the results based on the proposed methodology to provide key insights. It is divided into two subsections: the first presents the performance of classifiers on a slightly imbalanced CKD dataset, while the second examines their performance on a severely imbalanced dataset. A detailed discussion of these results is provided below.

4.1. Results of classifiers on the slightly imbalanced dataset

GBM achieved the highest testing accuracy of 99.25% ± 0.68, surpassing all other models, including other high-performing classifiers such as HGB, CB, LSTM, and BiLSTM, each of which achieved accuracies of nearly 99%. GBM achieved a superior precision of 99.35% ± 1.44, outperforming models such as CB, LSTM, and BiLSTM, and is significantly higher than that of probabilistic models, including GNB, CNB, and MNB, which achieved precision scores below 88%, as shown in Table 3.

GBM achieved a sensitivity of 98.67% ± 1.83 and outperformed all models in specificity, achieving 99.6% ± 0.89%, which demonstrates its predictive power in classifying negative instances and minimizing false positives correctly. The F1 score of GBM was 98.99% ± 0.92, the highest among all classifiers, and achieved an F2 score of 98.79% ± 1.38. GBM also achieved the highest balanced accuracy of 99.13% ± 0.84, indicating its effectiveness across both positive and negative classes, especially in scenarios with class imbalance. An MCC of 98.42% ± 1.44 was also the highest observed, further reflecting the GBM's strong overall correlation between predicted and actual outcomes. GBM achieved an AUROC and AUPRC value of 100% ± 0.00, outperforming other models such as ETC, SVM, CB, LSTM, and BiLSTM, which reported AUROC and AUPRC scores above 99.8%. GBM achieved the highest H-measure of 64.98% ± 0.00, which was the highest among all models. GBM performed its best in the first fold, as shown in Figure 2.

Table 3. Results of classifiers on the slightly imbalanced dataset (Mean ± Variance).

S. No.	Classifier	Testing Accuracy (%)	Precision (%)	Sensitivity (%)	F1 Score (%)	F2 Score (%)	Specificity (%)	Balanced Accuracy (%)	MCC (%)	AUROC (%)	AUPRC (%)	LR+	LR-	H-measure (%)
1	GNB	94 ± 2.05	87.76 ± 4.81	98 ± 1.83	92.51±2.32	95.7 ± 1.17	91.6 ± 3.85	94.8 ± 1.54	88.03 ± 3.66	94.73 ± 1.61	86.61 ± 4.42	13.91 ± 6.63	0.02 ± 0.01	61.67 ± 0.91
2	BNB	97.75 ± 1.37	94.43 ± 3.21	100 ± 0.00	97.11±1.71	98.82 ± 0.71	96.40 ± 2.19	98.2 ± 1.10	95.41 ± 2.71	99.8 ± 0.38	99.66 ± 0.65	36.66 ± 18.25	0.00 ± 0.00	64.82 ± 0.18
3	CNB	92.5 ± 3.06	83.64 ± 5.67	100 ± 0.00	91.01±3.37	96.17 ± 1.51	88 ± 4.90	94 ± 2.45	85.79 ± 5.29	99.84 ± 0.15	99.76 ± 0.23	9.69 ± 4.41	0.00 ± 0.00	64.9 ± 0.07
4	MNB	92.5 ± 3.06	83.64 ± 5.67	100 ± 0.00	91.01±3.37	96.17 ± 1.51	88 ± 4.90	94 ± 2.45	85.79 ± 5.29	99.84 ± 0.15	99.76 ± 0.23	9.69 ± 4.41	0.00 ± 0.00	64.9 ± 0.07
5	DT	95.75 ± 2.59	93.04 ± 4.45	96 ± 4.35	94.43±3.44	95.35 ± 3.73	95.6 ± 2.97	95.8 ± 2.75	91.12 ± 5.48	96.49 ± 2.05	92.17 ± 4.69	$2.00\times10^{9} \pm 4.47\times10^{9}$	0.04 ± 0.04	62.68 ± 1.47
6	ETC	98.25 ± 1.12	96.25 ± 3.42	99.33 ± 1.49	97.73±1.43	98.67 ± 0.95	97.6 ± 2.19	98.47 ± 0.87	96.39 ± 2.29	99.96 ± 0.04	99.94 ± 0.06	$3.93\times10^{9} \pm 5.38\times10^{9}$	0.00 ± 0.01	64.96 ± 0.01
7	KNN	94.75 ± 2.40	88.39 ± 5.08	99.33 ± 1.49	93.48±2.91	96.89 ± 1.68	92 ± 3.74	95.67 ± 2.01	89.63 ± 4.62	98.39 ± 0.98	95.25 ± 2.95	18.58 ± 17.64	0.00 ± 0.01	63.82 ± 0.89
8	SVM	98.5 ± 1.05	96.79 ± 2.21	99.33 ± 1.49	98.03±1.37	98.81 ± 1.28	98 ± 1.41	98.67 ± 1.03	96.86 ± 2.19	99.93 ± 0.08	99.89 ± 0.14	$2.00\times10^{9} \pm 4.47\times10^{9}$	0.00 ± 0.01	64.95 ± 0.04
9	LR	98 ± 1.43	95.62 ± 3.44	99.33 ± 1.49	97.41±1.82	98.55 ± 1.28	97.2 ± 2.28	98.27 ± 1.23	95.87 ± 2.90	99.93 ± 0.12	99.89 ± 0.18	$2.00\times10^{9} \pm 4.47\times10^{9}$	0.00 ± 0.01	64.95 ± 0.05
10	SGD	98.25 ± 0.68	96.85 ± 3.13	98.67 ± 1.83	97.7±0.85	98.27 ± 0.91	98 ± 2.00	98.33 ± 0.41	96.36 ±1.37	99.91 ± 0.17	99.85 ± 0.29	$3.86\times10^{9} \pm 5.29\times10^{9}$	0.01 ± 0.01	64.93 ± 0.08
11	LGBM	95.75 ± 2.59	93.9 ± 6.15	95.33 ± 5.58	94.41±3.45	94.91 ± 4.16	96 ± 4.24	95.67 ± 2.70	91.25 ± 5.39	99.76 ± 0.29	99.62 ± 0.48	$3.80\times10^{9} \pm 5.20\times10^{9}$	0.04 ± 0.05	64.86 ± 0.14
12	HGB	99 ± 1.63	98.18 ± 4.07	99.33 ± 1.49	98.71±2.07	99.07 ± 1.36	98.8 ± 2.68	99.07 ± 1.36	97.96 ± 3.29	99.89 ± 0.20	99.83 ± 0.33	$7.93\times10^{9} \pm 4.43\times10^{9}$	0.00 ± 0.01	64.93 ± 0.10
13	ADB	96.75 ± 2.59	94.8 ± 3.70	96.67 ± 3.33	95.72±3.42	96.28 ± 3.34	96.8 ± 2.28	96.73 ± 2.72	93.12 ± 5.50	96.73 ± 2.72	92.97 ± 5.36	$2.00\times10^{9} \pm 4.47\times10^{9}$	0.03 ± 0.03	62.67 ± 1.82
14	GBM	99.25 ± 0.68	99.35 ± 1.44	98.67 ± 1.83	98.99±0.92	98.79 ± 1.38	99.6 ± 0.89	99.13 ± 0.84	98.42 ± 1.44	100 ± 0.00	100 ± 0.00	$7.86\times10^{9} \pm 4.40\times10^{9}$	0.01 ± 0.01	64.98 ± 0.00
15	CB	99 ± 1.05	98.1 ± 2.81	99.33 ± 1.49	98.69±1.36	99.07 ± 1.12	98.8 ± 1.79	99.07 ± 0.92	97.92 ± 2.15	99.92 ± 0.18	99.87 ± 0.29	$5.93\times10^{9} \pm 5.41\times10^{9}$	0.00 ± 0.01	64.94 ± 0.08
16	XGB	98 ± 1.43	95.62 ± 3.44	99.33 ± 1.49	97.41±1.82	98.55 ± 1.28	97.2 ± 2.28	98.27 ± 1.23	95.87 ± 2.90	99.87 ± 0.19	99.79 ± 0.30	$2.00\times10^{9} \pm 4.47\times10^{9}$	0.00 ± 0.01	64.91 ± 0.09
17	LSTM	99 ± 1.05	98.1 ± 2.81	99.33 ± 1.49	98.69±1.36	99.07 ± 1.12	98.8 ± 1.79	99.07 ± 0.92	97.92 ± 2.15	99.93 ± 0.12	99.89 ± 0.19	$5.93\times10^{9} \pm 5.41\times10^{9}$	0.00 ± 0.01	64.95 ± 0.05
18	BiLSTM	99 ± 1.05	98.1 ± 2.81	99.33 ± 1.49	98.69±1.36	99.07 ± 1.12	98.8 ± 1.79	99.07 ± 0.92	97.92 ± 2.15	99.93 ± 0.15	99.89 ± 0.24	$5.93\times10^{9} \pm 5.41\times10^{9}$	0.00 ± 0.01	64.95 ± 0.07
19	GRU	98.75 ± 0.88	98.75 ± 2.80	98 ± 1.83	98.34±1.14	98.13 ± 1.20	99.2 ± 1.79	98.6 ± 0.80	97.38 ± 1.81	99.89 ± 0.14	99.83 ± 0.23	$7.80\times10^{9} \pm 4.36\times10^{9}$	0.02 ± 0.01	64.93 ± 0.06
20	BiGRU	99 ± 1.05	98.1 ± 2.81	99.33 ± 1.49	98.69±1.36	99.07 ± 1.12	98.8 ± 1.79	99.07 ± 0.92	97.92 ± 2.15	99.91 ± 0.17	99.85 ± 0.28	$5.93\times10^{9} \pm 5.41\times10^{9}$	0.00 ± 0.01	64.93 ± 0.08
21	RNN	98.5 ± 1.05	98.08 ± 2.82	98 ± 1.83	98.01±1.37	98 ± 1.34	98.8 ± 1.79	98.4 ± 0.98	96.85 ± 2.19	99.91 ± 0.10	99.85 ± 0.16	$5.86\times10^{9} \pm 5.35\times10^{9}$	0.02 ± 0.01	64.93 ± 0.05

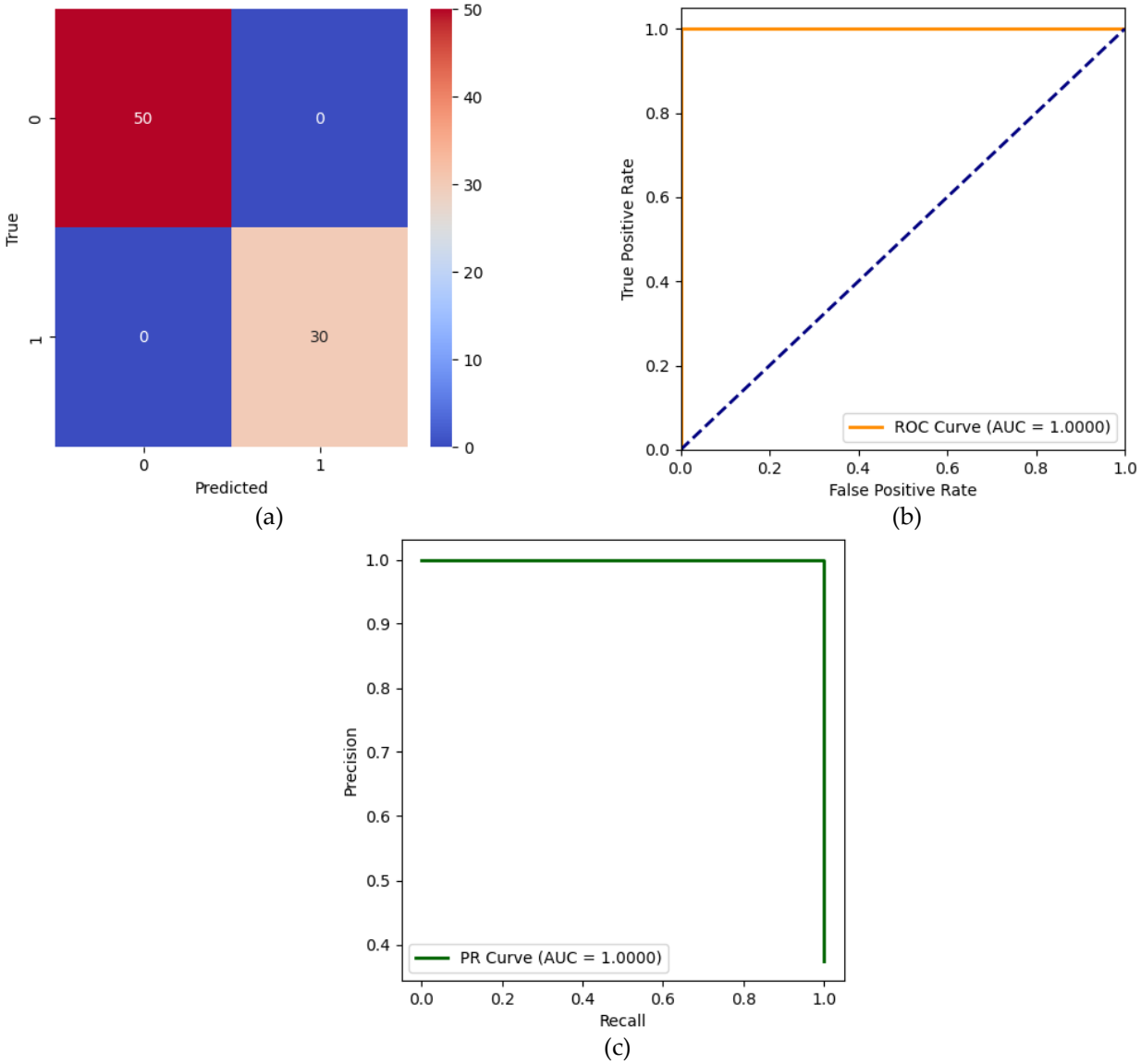

Figure 2. (a) Confusion Matrix, (b) AUROC, (c) AUPRC of GBM on the slightly imbalanced dataset.

GBM demonstrated superior performance across nearly all evaluation metrics. It achieved the highest testing accuracy, precision, F1 score, specificity, balanced accuracy, MCC, AUROC, and AUPRC. GBM also recorded a high LR+ value, indicating its stronger ability to identify true positives with fewer false positives. Additionally, GBM achieved a lower LR- (0.01) compared to traditional models, demonstrating its ability to reduce false negatives effectively. It also achieved the highest H-measure of 64.98%, making it statistically strong.

4.2 Results of classifiers on a severely imbalanced dataset

GBM achieved the highest testing accuracy of 92.26% ± 2.23 among all models, outperforming ensemble methods like ADB, CB, and GRU, as well as deep learning architectures such as BiLSTM and RNN, reflecting its strong generalization capability on unseen data. GBM outperformed other models, including ADB and CB, in terms of precision, particularly in its ability to avoid false positives. GBM outperformed other models in terms of F1 score, followed closely by CB and GRU, highlighting GBM's ability to handle class imbalance. GBM stands out with a specificity of 96.09% ± 2.65, demonstrating its ability to correctly identify negative cases, which is crucial in domains where false positives can be costly. GBM achieved a balanced

accuracy of 79.18% ±7.21 and an MCC of 61.09% ± 11.10, suggesting a decent correlation between predicted and actual labels, even in the presence of class imbalance. In terms of AUROC, GBM scored 90.79% ± 3.90 and achieved the highest AUPRC of 69.11% ± 11.41. GBM achieved an LR+ of 26.2 ± 22.45, the highest among all classifiers, indicating a powerful increase in the odds of disease presence when the test is positive, and an LR- of 0.39, reflecting a decreased probability of disease when the test is negative, as shown in Table 4. Figure 3 depicts the second fold results of GBM, which are the highest among the five folds. Table 3 shows that several models, including CB, LSTM, BiLSTM, and GRU, performed competitively on the slightly imbalanced dataset, achieving near-perfect metrics across accuracy, precision, and AUROC. However, GBM outperformed all classifiers, achieving high AUROC and AUPRC values of 100%, the highest precision of 99.35% ± 1.44, and the highest balanced accuracy of 99.13% ± 0.84. On the severely imbalanced dataset, shown in Table 4, GBM again emerged as the top performer, with the highest testing accuracy of 92.26% ± 2.23, specificity of 96.09% ± 2.65, and AUPRC of 69.11% ± 11.41, along with the highest LR+ of 26.2 ± 22.45. It also maintained strong sensitivity and F1 scores, indicating a reliable balance between correctly detecting both positive and negative cases. From these results, it is evident that GBM performed robustly across different imbalance scenarios and consistently outperformed other classifiers, making it the most suitable model to be proposed.

Table 4. Results of classifiers on severely imbalanced dataset (Mean ± Variance).

S. No.	Classifier	Testing Accuracy (%)	Precision (%)	Sensitivity (%)	F1 Score (%)	F2 Score (%)	Specificity (%)	Balanced Accuracy (%)	MCC (%)	AUROC (%)	AUPRC (%)	LR+	LR-	H-measure (%)
1	GNB	82.08 ± 1.89	34.92 ± 3.23	67.73 ± 16.73	45.81 ± 6.57	56.7 ± 11.11	83.91 ± 3.04	75.82 ± 7.35	39.49 ± 9.04	85.18 ± 4.95	58.13 ± 9.60	4.2 ± 0.68	0.38 ± 0.18	79.66 ± 4.21
2	BNB	82.89 ± 3.28	37.48 ± 5.22	76.67 ± 8.41	50.19 ± 5.73	63.21 ± 6.49	83.22 ± 3.69	79.94 ± 4.37	45.1 ± 7.01	89.73 ± 2.92	63.83 ± 6.34	4.74 ± 1.11	0.28 ± 0.10	83.62 ± 2.46
3	CNB	75.16 ± 4.28	28.56 ± 5.12	76.97 ± 13.14	41.55 ± 6.92	57.28 ± 9.18	74.94 ± 4.33	75.96 ± 7.15	35.82 ± 9.85	84.6 ± 3.00	54.72 ± 6.74	3.16 ± 0.84	0.3 ± 0.18	79.15 ± 2.52
4	MNB	75.16 ± 4.28	28.56 ± 5.12	76.97 ± 13.14	41.55 ± 6.92	57.28 ± 9.18	74.94 ± 4.33	75.96 ± 7.15	35.82 ± 9.85	84.6 ± 3.00	54.72 ± 6.74	3.16 ± 0.84	0.3 ± 0.18	79.15 ± 2.52
5	DT	83.92 ± 5.60	39.16 ± 13.35	62.42 ± 16.40	47.72 ± 13.98	55.4 ± 14.93	86.67 ± 5.11	74.55 ± 9.57	40.66 ± 16.81	80.37 ± 8.86	38.72 ± 13.03	5.53 ± 2.99	0.43 ± 0.20	74.47 ± 6.31
6	ETC	90.02 ± 2.20	55.38 ± 9.49	71.21 ± 17.73	61.47 ± 9.96	66.71 ± 13.70	92.41 ± 2.65	81.81 ± 8.28	57.03 ± 11.61	90.66 ± 3.38	63.74 ± 10.05	10.46 ± 4.79	0.30 ± 0.18	84.44 ±2.98
7	KNN	82.49 ± 2.89	34.14 ± 7.14	56.97 ± 12.38	42.58 ± 8.66	50.13 ± 10.38	85.75 ± 2.65	71.36 ± 6.49	34.62 ± 10.43	79.6 ± 3.69	37.02 ± 8.77	4.13 ± 1.21	0.5 ± 0.14	73.42 ± 4.62
8	SVM	84.94 ± 2.94	39.81 ± 6.68	58.64 ± 22.29	45.84 ± 9.63	52.21 ± 15.78	88.28 ± 5.29	73.46 ± 9.03	39.67 ± 10.89	86.21 ± 4.90	54.31 ± 15.76	5.29 ± 1.65	0.45 ± 0.22	80.58 ± 4.24
9	LR	85.75 ± 3.84	41.68 ± 12.37	62.27 ± 23.81	49.1 ± 14.94	55.92 ± 19.01	88.74 ± 4.25	75.5 ± 11.37	42.99 ± 17.45	87.02 ± 5.37	56.57 ± 17.81	6.12 ± 3.31	0.42 ± 0.26	81.29 ± 4.63
10	SGD	86.15 ± 2.09	42.29 ± 7.41	65.61 ± 21.93	50.97 ± 11.60	58.64 ± 16.50	88.74 ± 2.74	77.17 ± 10.15	45.15 ± 14.08	87.48 ± 6.11	57.47 ± 15.17	5.86 ± 1.67	0.38 ± 0.24	81.67 ± 5.36
11	LGBM	86.56 ± 2.28	44.78 ± 6.13	73.18 ± 20.35	54.78 ± 8.49	64.17 ± 13.85	88.28 ± 3.49	80.73 ± 9.07	50.05 ± 11.04	89.38 ± 5.54	63.69 ± 16.08	6.47 ± 1.79	0.29 ± 0.22	83.32 ± 4.82
12	HGB	89.62 ± 2.19	55.28 ± 9.53	62.27 ± 20.05	56.75 ± 10.85	59.58 ± 16.06	93.1 ± 3.15	77.69 ± 9.08	52.2 ± 11.41	87.96 ± 4.31	62.43 ± 15.49	10.36 ± 4.32	0.40 ± 0.20	82.09 ± 3.73
13	ADB	91.45 ± 1.13	66.16 ± 9.11	56.97 ± 12.38	59.8 ± 6.48	57.87 ± 10.27	95.86 ± 2.09	76.42 ± 5.36	56.08 ± 5.01	91.06 ± 3.78	63.24 ± 9.49	17.09 ± 8.75	0.44 ± 0.11	84.79 ± 3.38
14	GBM	92.26 ± 2.23	70.13 ± 14.06	62.27 ± 15.38	64.24 ± 11.31	62.78 ± 13.62	96.09 ± 2.65	79.18 ± 7.21	61.09 ± 11.10	90.79 ± 3.9	69.11 ± 11.41	26.2 ± 22.45	0.39 ± 0.15	84.54 ± 3.41
15	CB	91.04 ± 2.65	59.89 ± 9.52	65.91 ± 17.75	62.05 ± 12.10	64.15 ± 15.26	94.25 ± 2.15	80.08 ± 8.87	57.54 ± 13.44	90.91 ± 4.47	68.65 ± 14.15	12.48 ± 4.84	0.36 ± 0.18	84.65 ± 3.90
16	XGB	84.94 ± 3.56	39.4 ± 9.40	60.45 ± 24.03	46.75 ± 12.87	53.72 ± 18.04	88.05 ± 4.78	74.25 ± 10.87	40.41 ± 15.29	86.89 ± 5.49	56.72 ± 18.60	5.32 ± 2.13	0.44 ± 0.25	81.17 ± 4.74
17	LSTM	88.38 ± 3.53	49.06 ± 15.58	61.97 ± 25.45	53.72 ± 17.29	58.07 ± 21.26	91.72 ± 2.98	76.85 ± 12.54	48.35 ± 19.94	88.09 ± 7.28	60.43 ± 14.56	8.64 ± 5.02	0.41 ± 0.27	82.19 ± 6.42
18	BiLSTM	89.2 ± 2.78	52.08 ± 10.52	63.94 ± 17.42	56.97 ± 12.27	60.83 ± 14.91	92.41 ± 2.09	78.18 ± 8.85	51.53 ± 14.14	89.14 ± 4.73	65.35 ± 14.03	9 ± 3.32	0.39 ± 0.18	83.11 ± 4.20
19	GRU	91.24 ± 2.35	61.35 ± 11.26	63.94 ± 17.42	61.88 ± 11.86	62.93 ± 14.83	94.71 ± 1.92	79.33 ± 8.60	57.43 ± 13.35	89.33 ± 4.64	61.31 ± 13.94	13.88 ± 6.83	0.38 ± 0.18	83.28 ± 4.06
20	BiGRU	87.39 ± 4.17	47.77 ± 9.53	62.27 ± 15.38	53.18 ± 9.57	57.93 ± 11.79	90.57 ± 5.17	76.42 ± 6.97	47.3 ± 10.96	90.41 ± 4.03	63.11 ± 11.45	7.59 ± 3.08	0.41 ± 0.15	84.21 ± 3.44
21	RNN	90.63 ± 3.35	60.48 ± 19.75	54.85 ± 19.82	56.58 ± 16.92	55.33 ± 18.20	95.17 ± 2.49	75.01 ± 10.18	52.02 ± 19.13	88.23 ± 5.65	60.04 ± 11.88	17.92 ± 17.20	0.47 ± 0.21	82.33 ± 5.01

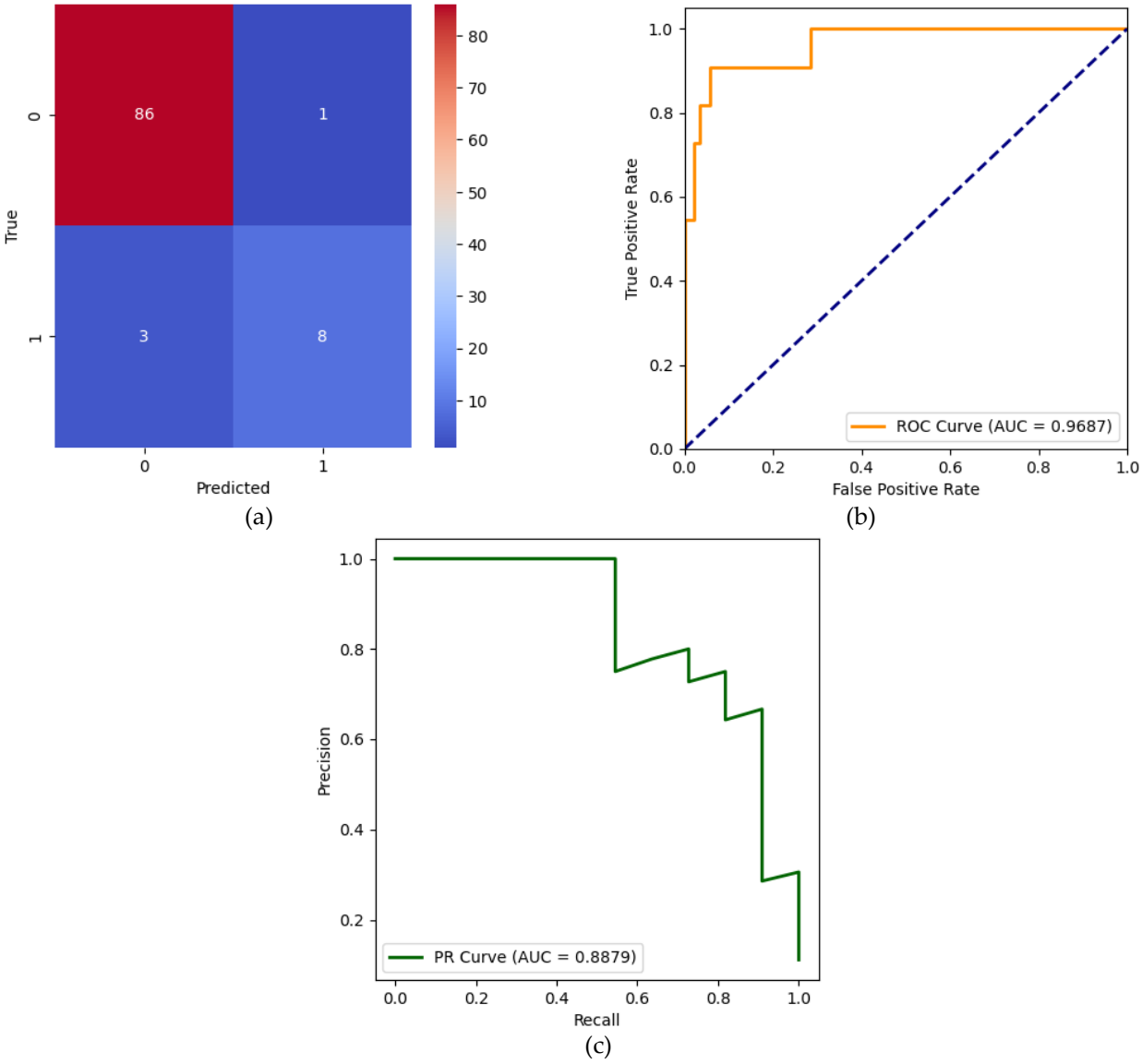

Figure 3. (a) Confusion Matrix, (b) AUROC, (c) AUPRC of GBM on a severely imbalanced dataset.

5. Conclusions

Chronic Kidney Disease (CKD) is a pervasive health issue that requires timely and accurate detection to mitigate severe health complications. While Artificial Intelligence (AI) has significantly advanced CKD diagnostic capabilities, addressing data imbalance remains essential for achieving accurate results. Although many AI techniques perform well on balanced datasets, a thorough investigation of methods that excel in handling imbalanced datasets has not been conducted extensively. This research highlights the importance of addressing data imbalance to improve the performance of AI models in real-world scenarios, where instances of the minority class are often underrepresented. Through comprehensive experimentation, this research evaluates the performance of multiple classifiers on both slightly imbalanced and severely imbalanced CKD datasets. The findings highlight that GBM exhibits superior performance. With specificities of 99.6% ± 0.89 and 96.09% ± 2.65 for slightly and severely imbalanced datasets, respectively, the GBM model demonstrated significant performance in predicting true negatives. Furthermore, metrics such as AUROC (100% and 90.79% ± 3.9),

LR+ (7.86 ×10^9 ± 4.40 ×10^9 and 26.2 ± 22.45), LR- (0.01 ± 0.01 and 0.39 ± 0.15), and H-measure (64.98% and 84.54% ± 3.41) highlight the model's efficacy in effectively distinguishing between classes.

5.1 Limitations

Although the proposed model has shown decent performance across all evaluation metrics, including a testing accuracy of 99.25%, an AUROC of 100%, and a balanced accuracy of 99.13%, some limitations remain. The datasets utilized for the experiment may not represent the demographic characteristics of patients across the globe. Model's generalizability may be affected by variations in diagnosis practices and demographic characteristics. There is a possibility of a decline in performance when the model is faced with missing or noisy data. The proposed model was trained using the past data, which might not have taken into consideration changes in clinical patterns over time.

5.2 Clinical and Healthcare Implications

The performance of GBM shows its potential in clinical practice for the early detection of CKD. It can be used as a support system for nephrologists and physicians for identifying high-risk patients at the early stages to reduce the likelihood of progression of the disease. GBM-based decision support tools can be used in electronic healthcare record (EHR) systems to identify abnormal patients during check-ups. The predictive power of the GBM makes it a promising system for reducing diagnostic delays and improving long-term patient outcomes. By continuously analyzing patient data, GBM can help medical practitioners identify patients who are suffering from CKD, leading to more effective disease management, which is beneficial in primary care settings.

5.3 Feasibility of Real-World Implementation

Due to the efficiency and relatively low computational demands at inference time, the practical implementation of GBM in real-world healthcare settings is promising and feasible. It can be integrated into clinical software systems, even in resource-limited environments such as rural or under-funded hospitals. It can also be operated on cloud-based platforms that connect to local health centers. Integration with IoT-based monitoring systems to enable continuous monitoring of CKD-relevant patterns through wearable or home-based devices can support long-term patient monitoring. In addition, GBM can be deployed using scalable architectures that support modular updates, making it easier to improve and adapt the system over time as more patient data becomes available to ensure that the model remains accurate and relevant across diverse patient populations and clinical scenarios.

5.4 Future Scope

Future research should include validating the proposed model using larger datasets and focus on integrating explainable AI (XAI) techniques that can help clinicians better understand the predictions, improving healthcare trust. Additionally, federated learning can be employed to train models across healthcare institutions without compromising patient privacy. The incorporation of attention mechanisms in sequence-based deep learning models like LSTM or Transformer architectures can allow the system to focus on clinically significant temporal patterns, such as gradual changes. Exploring hybrid models that combine machine learning techniques with deep learning architectures may yield both accuracy and interpretability, balancing computational efficiency with robustness.

Author Contributions: Conceptualization, G.P.R. and D.R.; methodology, K.P.P. and G.P.R.; software, K.P.P. and M.K.C.; validation, Y.V.P.K. and M.K.C.; formal analysis, Y.V.P.K.; investigation, D.R.; resources, G.P.R.; data curation, Y.V.P.K.; writing—original draft preparation, K.P.P.; writing—review and editing, D.R.; visualization, K.P.P.; supervision, Y.V.P.K.; project administration, M.K.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Takkavatakarn, K.; Oh, W.; Cheng, E.; Nadkarni, G. N.; Chan, L. Machine Learning Models to Predict End-Stage Kidney Disease in Chronic Kidney Disease Stage 4. *BMC Nephrol* **2023**, *24*(1), 376. https://doi.org/10.1186/s12882-023-03424-7

[2] Nishanth, A.; Thiruvaran, T. Identifying Important Attributes for Early Detection of Chronic Kidney Disease. *IEEE Rev. Biomed. Eng.* **2018**, *11*, 208–216. https://doi.org/10.1109/RBME.2017.2787480

[3] Hodneland, E.; Keilegavlen, E.; Hanson, E. A.; Andersen, E.; Monssen, J. A.; Rorvik, J.; Leh, S.; Marti, H.-P.; Lundervold, A.; Svarstad, E.; Nordbotten, J. M. In Vivo Detection of Chronic Kidney Disease Using Tissue Deformation Fields From Dynamic MR Imaging. *IEEE Trans. Biomed. Eng.* **2019**, *66*(6), 1779-1790. https://doi.org/10.1109/TBME.2018.2879362

[4] Pradeep Reddy, G.; Rohan, D.; Kumar, Y. V. P.; Prakash, K. P.; Srikanth, M. Artificial Intelligence-Based Effective Detection of Parkinson's Disease Using Voice Measurements. *Engineering Proceedings* **2024,** *82*(1), 28. https://doi.org/10.3390/ecsa-11-20481

[5] Reddy, G. P.; Kumar, Y. V. P. Explainable AI (XAI): Explained. In 2023 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream); *IEEE*: Vilnius, Lithuania, **2023**; pp 1-6. https://doi.org/10.1109/eStream59056.2023.10134984

[6] Ma, F.; Sun, T.; Liu, L.; Jing, H. Detection and Diagnosis of Chronic Kidney Disease Using Deep Learning-Based Heterogeneous Modified Artificial Neural Network. *Future Generation Computer Systems* **2020**, *111*, 17-26. https://doi.org/10.1016/j.future.2020.04.036

[7] Ali, M. M.; Islam, M. S.; Uddin, M. N.; Uddin, Md. A. A Conceptual IoT Framework Based on Anova-F Feature Selection for Chronic Kidney Disease Detection Using Deep Learning Approach. *Intelligence-Based Medicine* **2024**, *10*, 100170. https://doi.org/10.1016/j.ibmed.2024.100170

[8] Chen, G.; Ding, C.; Li, Y.; Hu, X.; Li, X.; Ren, L.; Ding, X.; Tian, P.; Xue, W. Prediction of Chronic Kidney Disease Using Adaptive Hybridized Deep Convolutional Neural Network on the Internet of Medical Things Platform. *IEEE Access* **2020**, *8*, 100497-100508. https://doi.org/10.1109/ACCESS.2020.2995310

[9] Ramu, K.; Patthi, S.; Prajapati, Y. N.; Ramesh, J. V. N.; Banerjee, S.; Rao, K. B. V. B.; Alzahrani, S. I.; Ayyasamy, R. Hybrid CNN-SVM Model for Enhanced Early Detection of Chronic Kidney Disease. *Biomedical Signal Processing and Control* **2025**, *100*, 107084. https://doi.org/10.1016/j.bspc.2024.107084

[10] Bhaskar, N.; Manikandan, S. A Deep-Learning-Based System for Automated Sensing of Chronic Kidney Disease. *IEEE Sens. Lett.* **2019**, *3(10)*, 1–4. https://doi.org/10.1109/LSENS.2019.2942145

[11] Rubini, L.; Soundarapandian, P.; Eswaran, P. *Chronic Kidney Disease*. **2015**. https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease (accessed 2025 18/02/2025)

[12] AI-Shamsi, S.; Regmi, D.; Govender, R. D.; Chicco, D. *Chronic kidney disease EHRs Abu Dhabi*. **2021**. https://www.kaggle.com/datasets/davidechicco/chronic-kidney-disease-ehrs-abu-dhabi (accessed 2025 18/02/2025)

[13] Norouzi, M.; Kahriman, E. A. A Machine Learning-Based Early Diagnosis Model for Chronic Kidney Disease Using SPegasos. *Netw Model Anal Health Inform Bioinforma* **2024**, *13*(1), 20. https://doi.org/10.1007/s13721-024-00457-2

[14] Debal, D. A.; Sitote, T. M. Chronic Kidney Disease Prediction Using Machine Learning Techniques. *J Big Data* **2022**, *9*(1), 109. https://doi.org/10.1186/s40537-022-00657-5

[15] Saif, D.; Sarhan, A. M.; Elshennawy, N. M. Deep-Kidney: An Effective Deep Learning Framework for Chronic Kidney Disease Prediction. *Health Inf Sci Syst* **2023**, *12*(1), 3. https://doi.org/10.1007/s13755-023-00261-8

[16] Saif, D.; Sarhan, A. M.; Elshennawy, N. M. Early Prediction of Chronic Kidney Disease Based on Ensemble of Deep Learning Models and Optimizers. *Journal of Electrical Systems and Inf Technol* **2024**, *11*(1), 17. https://doi.org/10.1186/s43067-024-00142-4

[17] Al-Jamimi, H. A. Synergistic Feature Engineering and Ensemble Learning for Early Chronic Disease Prediction. *IEEE Access* **2024**, *12*, 62215–62233. https://doi.org/10.1109/ACCESS.2024.3395512

[18] Gogoi, P.; Valan, J. A. Privacy-Preserving Predictive Modeling for Early Detection of Chronic Kidney Disease. *Netw Model Anal Health Inform Bioinforma* **2024**, *13*(1), 16. https://doi.org/10.1007/s13721-024-00452-7

[19] Corbin, C.K.; Maclay, R.; Acharya, A.; Mony, S.; Punnathanam, S.; Thapa, R.; Kotecha, N.; Shah, N.H.; Chen, J.H. DEPLOYR: A Technical Framework for Deploying Custom Real-Time Machine Learning Models into the Electronic Medical Record. *Journal of the American Medical Informatics Association* **2023**, *30*, 1532-1542, doi: https://doi.org/10.1093/jamia/ocad114

[20] Thomas, L.; Hyde, C.; Mullarkey, D.; Greenhalgh, J.; Kalsi, D.; Ko, J. Real-World Post-Deployment Performance of a Novel Machine Learning-Based Digital Health Technology for Skin Lesion Assessment and Suggestions for Post-Market Surveillance. *Front. Med.* **2023**, *10*, 1264846, doi: https://doi.org/10.3389/fmed.2023.1264846

[21] Hong, J.C.; Eclov, N.C.W.; Stephens, S.J.; Mowery, Y.M.; Palta, M. Implementation of Machine Learning in the Clinic: Challenges and Lessons in Prospective Deployment from the System for High Intensity EvaLuation During Radiation Therapy (SHIELD-RT) Randomized Controlled Study. *BMC Bioinformatics* **2022**, *23*, 408, doi: https://doi.org/10.1186/s12859-022-04940-3

[22] Soenksen, L.R.; Ma, Y.; Zeng, C.; Boussioux, L.; Villalobos Carballo, K.; Na, L.; Wiberg, H.M.; Li, M.L.; Fuentes, I.; Bertsimas, D. Integrated Multimodal Artificial Intelligence Framework for Healthcare Applications. *npj Digit. Med.* **2022**, *5*, 149, doi: https://doi.org/10.1038/s41746-022-00689-4

[23] Islam, Md. S. B.; Sumon, Md. S. I.; Sarmun, R.; Bhuiyan, E. H.; Chowdhury, M. E. H. Classification and Segmentation of Kidney MRI Images for Chronic Kidney Disease Detection. *Computers and Electrical Engineering* **2024**, *119*, 109613. https://doi.org/10.1016/j.compeleceng.2024.109613

[24] Awad Yousif, S. M.; Halawani, H. T.; Amoudi, G.; Osman Birkea, F. M.; Almunajam, A. M. R.; Elhag, A. A. Early Detection of Chronic Kidney Disease Using Eurygasters Optimization Algorithm with Ensemble Deep Learning Approach. *Alexandria Engineering Journal* **2024**, *100*, 220-231. https://doi.org/10.1016/j.aej.2024.05.011

[25] Bialonczyk, U.; Debowska, M.; Dai, L.; Qureshi, A. R.; Söderberg, M.; Lindholm, B.; Stenvinkel, P.; Poleszczuk, J. Detection of Medial Vascular Calcification in Chronic Kidney Disease Based on Pulse Wave Analysis in the Frequency Domain. *Biomedical Signal Processing and Control* **2024**, *94*, 106250. https://doi.org/10.1016/j.bspc.2024.106250

[26] Dharmarathne, G.; Bogahawaththa, M.; McAfee, M.; Rathnayake, U.; Meddage, D. P. P. On the Diagnosis of Chronic Kidney Disease Using a Machine Learning-Based Interface with Explainable Artificial Intelligence. *Intelligent Systems with Applications* **2024**, *22*, 200397. https://doi.org/10.1016/j.iswa.2024.200397

[27] Islam, Md. A.; Majumder, Md. Z. H.; Hussein, Md. A. Chronic Kidney Disease Prediction Based on Machine Learning Algorithms. *Journal of Pathology Informatics* **2023**, *14*, 100189. https://doi.org/10.1016/j.jpi.2023.100189

[28] Venkatrao, K.; Kareemulla, S. HDLNET: A Hybrid Deep Learning Network Model With Intelligent IOT for Detection and Classification of Chronic Kidney Disease. *IEEE Access* **2023**, *11*, 99638-99652. https://doi.org/10.1109/ACCESS.2023.3312183

[29] Rashed-Al-Mahfuz, Md.; Haque, A.; Azad, A.; Alyami, S. A.; Quinn, J. M. W.; Moni, M. A. Clinically Applicable Machine Learning Approaches to Identify Attributes of Chronic Kidney Disease (CKD) for Use in Low-Cost Diagnostic Screening. *IEEE J. Transl. Eng. Health Med*. **2021**, *9*, 1-11. https://doi.org/10.1109/JTEHM.2021.3073629

[30] Sobrinho, A.; Queiroz, A. C. M. D. S.; Dias Da Silva, L.; De Barros Costa, E.; Eliete Pinheiro, M.; Perkusich, A. Computer-Aided Diagnosis of Chronic Kidney Disease in Developing Countries: A Comparative Analysis of Machine Learning Techniques. *IEEE Access* **2020**, *8*, 25407-25419. https://doi.org/10.1109/ACCESS.2020.2971208

[31] Bhaskar, N.; Suchetha, M.; Philip, N. Y. Time Series Classification-Based Correlational Neural Network With Bidirectional LSTM for Automated Detection of Kidney Disease. *IEEE Sensors J.* **2021**, *21*(4), 4811-4818. https://doi.org/10.1109/JSEN.2020.3028738