Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14,issue.1s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Machine Learning Identification of Oxidative Stress Biomarkers for Early Chronic Kidney Disease Detection

Wajihus shams[1], M. Mustafa khan[2], Girish Chandra Sharma[3], Anupama Sharma[4]*

[1]PhD Scholar Department of Medical Lab Technology, Nims Institute of Allied & Health Care Sciences, Nims University Rajasthan, Jaipur.

[2]Associate Professor, Department of Pathology, Deccan College of Medical Sciences, Hyderabad, India.

[3]Associate Professor, Department of Chemistry, Nims Institute of Engineering and Technology, Nims University Rajasthan, Jaipur, India.

[4]Associate Professor, Department of Biochemistry and Clinical Research, Research and Innovation Cell, Nims University Rajasthan, Jaipur

***Corresponding Author:**

Anupama Sharma

Associate Professor, Department of Biochemistry and Clinical Research, Research and Innovation Cell, Nims University Rajasthan, Jaipur

Cite this paper as Wajihus shams, M. Mustafa khan, Girish Chandra Sharma, Anupama Sharma .(2025) Machine Learning Identification of Oxidative Stress Biomarkers for Early Chronic Kidney Disease Detection. .Journal of Neonatal Surgery, 14, (1s), 1470-1483

ABSTRACT

Oxidative stress plays a pivotal role in chronic kidney disease (CKD) pathogenesis, particularly in hypertensive patients, creating a self-perpetuating cycle of renal damage. This study investigated oxidative stress mechanisms in chronic hypertension and renal impairment while validating related clinical biomarkers through machine learning approaches. A comprehensive analysis was conducted using clinical datasets containing 25 parameters from patients with varying kidney function stages. Three machine learning algorithms i.e. K-Nearest Neighbors, Decision Tree, and AdaBoostwere employed for CKD prediction, with performance evaluated using accuracy, precision, recall, and AUC metrics. The Decision Tree algorithm achieved exceptional performance with 97.5% accuracy and 0.969 AUC, followed by AdaBoost (96.7% accuracy, 0.962 AUC), while KNN showed moderate performance (66.7% accuracy). Significant distributional differences between CKD and non-CKD populations were observed for oxidative stress-related parameters, with blood glucose showing pronounced separation consistent with hyperglycemia-induced oxidative stress. Hemoglobin and albumin levels reflected oxidative damage-associated anemia and compromised antioxidant defenses, respectively. The superior performance of tree-based algorithms suggests discrete oxidative stress thresholds aligning with established pathophysiological mechanisms. These findings demonstrate that machine learning can effectively validate oxidative stress-related biomarkers for CKD prediction, supporting computational approaches for early detection and personalized antioxidant intervention strategies to break the oxidative stress cycle in kidney disease progression..

Keywords: *Oxidative stress, chronic kidney disease, machine learning, hypertension, biomarkers, predictive modeling*.

1. INTRODUCTION

Machine learning has fundamentally changed how industries approach complex problems and make critical decisions. In manufacturing, these algorithms have improved production processes, enhanced quality control, and created maintenance systems that predict equipment failures before they occur, saving companies significant time and resources [1-6]. The financial sector has adopted machine learning for detecting fraudulent transactions, automating trading decisions, and assessing investment risks, while transportation companies use these technologies for developing self-driving vehicles and optimizing delivery routes. However, the most significant and potentially life-saving applications of machine learning have emerged in healthcare.

Healthcare systems worldwide are experiencing a data revolution, with machine learning playing an increasingly important role in medical practice. These computational methods can analyze enormous amounts of medical information to discover patterns that traditional analysis might miss [7-9]. Medical imaging, diagnostic procedures, drug development, and treatment planning have all benefited from machine learning applications that help doctors make better decisions, reduce errors, and improve patient care. Electronic health records, genetic information,

and continuous

patient monitoring now generate massive amounts of data that machine learning algorithms can process to predict how diseases will progress, identify high-risk patients, and customize treatments for individual patients.

Chronic diseases present particular opportunities for machine learning applications because they involve complicated biological processes with multiple contributing factors. Chronic kidney disease (CKD) exemplifies this complexity, where early detection and treatment are essential for preventing progression to complete kidney failure [10-12]. Current diagnostic methods for CKD depend mainly on measuring creatinine levels in blood and calculating kidney filtration rates, but these tests often fail to detect kidney problems until substantial damage has already occurred. This limitation highlights the urgent need for more sensitive and predictive methods that can identify patients at risk before permanent kidney damage develops.

The biological processes underlying CKD, especially when associated with high blood pressure, involve intricate molecular mechanisms that go beyond simple blood flow problems. Growing scientific evidence indicates that oxidative stress plays a fundamental role in how kidney disease begins and worsens, creating a destructive cycle where harmful molecules called reactive oxygen species overwhelm the body's natural protective systems, causing cell damage, inflammation, and progressive kidney destruction. In patients with high blood pressure, this oxidative stress cycle becomes more severe due to the physical stress from elevated blood pressure, blood vessel dysfunction, and activation of hormone systems that control blood pressure and fluid balance.

The connection between chronic high blood pressure and kidney damage represents a complex two-way process where each condition makes the other worse, creating what researchers call the "hypertension-CKD cycle." Oxidative stress acts as both a result and a cause of this disease process, making it an important target for understanding how the disease works and developing prediction models [13-15]. Standard clinical measurements used to assess CKDincluding blood counts, protein levels, blood sugar, and mineral balanceare closely connected to oxidative stress pathways, but their relationships to underlying molecular processes have not been systematically studied using computational methods.

Recent developments in machine learning provide new opportunities to uncover these hidden connections and confirm the clinical importance of oxidative stress markers in predicting CKD. By analyzing patterns in routinely collected medical data, machine learning algorithms can potentially identify oxidative stress patterns that predict disease progression more accurately than traditional risk assessment methods. This computational approach not only improves our ability to predict disease outcomes but also provides insights into disease mechanisms that can guide treatment strategies aimed at interrupting the oxidative stress cycle.

This study addresses this important knowledge gap by examining the role of oxidative stress in chronic high blood pressure and kidney damage while simultaneously testing oxidative stress-related clinical measurements through machine learning-based prediction models. By connecting molecular mechanisms with computational analysis, this research aims to establish a comprehensive framework for understanding and predicting CKD progression, ultimately contributing to more effective early detection strategies and targeted treatments that can break the cycle of oxidative stress-related kidney damage.

2. Methodology

This study employed a cross-sectional analytical design shown in Figure 1 to investigate the relationship between oxidative stress biomarkers and chronic kidney disease progression using machine learning approaches. The research utilized a comprehensive clinical dataset containing demographic, biochemical, and physiological parameters from patients with varying stages of kidney function. The study protocol integrated traditional epidemiological methods with advanced computational techniques to validate oxidative stress-related clinical parameters in CKD prediction.

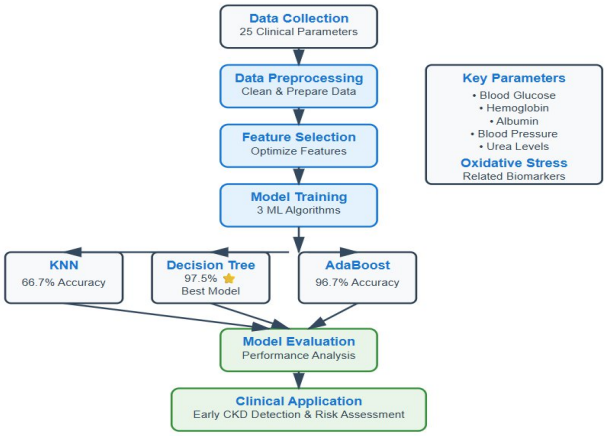

Figure 1. Machine learning workflow for chronic kidney disease (CKD) prediction using oxidative stress-related clinical parameters. The pipeline encompasses data collection of 25 clinical parameters, preprocessing and feature selection to optimize predictive variables, followed by training and evaluation of three machine learning algorithms

The analysis was conducted using a clinical dataset comprising 400 patient records with complete information on 25 clinical parameters. The study population included both patients diagnosed with chronic kidney disease and healthy controls, providing a balanced representation for binary classification analysis. Inclusion criteria encompassed adult patients (≥18 years) with complete clinical data, while exclusion criteria included patients with incomplete records, acute kidney injury, or missing essential laboratory values. The dataset contained comprehensive information on hematological parameters (red blood cell count, white blood cell count, hemoglobin, packed cell volume), biochemical markers (albumin, blood glucose, blood urea, sodium), and clinical indicators (blood pressure status, diabetes presence, specific gravity).

Clinical parameters were selected based on their established relationships to oxidative stress pathways in chronic kidney disease. Blood glucose levels were included as hyperglycemia is known to increase reactive oxygen species production and contribute to diabetic nephropathy. Hemoglobin and packed cell volume were selected as anemia in CKD patients often results from oxidative damage to erythropoietin-producing cells and iron metabolism disruption. Albumin concentration was chosen because serum albumin serves as an important endogenous antioxidant, and its reduction indicates compromised antioxidant defenses. Blood pressure parameters were included as hypertension both contributes to and results from oxidative stress in the kidney. Additional parameters such as blood urea, sodium levels, and specific gravity reflect kidney function status and are influenced by oxidative stress-mediated nephron damage.

Data preprocessing involved several sequential steps to ensure data quality and model reliability. Missing values were identified and handled using appropriate imputation techniques, with median imputation applied for numerical variables and mode imputation for categorical variables. Outlier detection was performed using interquartile range methods, with extreme values investigated for clinical plausibility. Categorical variables were encoded using label encoding techniques to convert text-based classifications into numerical formats suitable for machine learning algorithms. Feature scaling was applied using standardization methods to normalize variables with different measurement scales and ranges, ensuring equal contribution of all parameters to model training.

Feature selection was conducted to identify the most predictive subset of clinical parameters while reducing computational complexity and avoiding overfitting. Multiple feature selection techniques were employed, including correlation analysis to identify highly correlated features, univariate statistical tests to assess individual parameter significance, and recursive feature elimination to systematically remove less important variables. The final feature set was optimized to approximately 30% of the original parameters, representing the most informative variables for CKD prediction while maintaining clinical interpretability.

Three distinct machine learning algorithms were implemented to provide comprehensive model comparison and validation. K-Nearest Neighbors (KNN) was selected as a distance-based algorithm that classifies samples based on similarity to neighboring data points, with hyperparameter tuning performed to optimize the number of neighbors. Decision Tree algorithm was chosen for its interpretability and ability to capture non-linear relationships between variables, with pruning techniques applied to prevent overfitting. AdaBoost (Adaptive Boosting) was implemented as an ensemble method that combines multiple weak learners to create a stronger predictive model, with sequential learning to focus on previously misclassified samples.

The dataset was randomly divided into training and testing subsets using an 80:20 split ratio to ensure adequate data for both model development and evaluation. Cross-validation techniques were employed during training to assess model stability and generalization capability. Hyperparameter optimization was performed using grid search methods to identify optimal parameter combinations for each algorithm. Model training involved iterative processes where algorithms learned patterns from training data while monitoring for overfitting through validation curves and learning curves.

Model performance was comprehensively evaluated using multiple metrics to assess different aspects of predictive capability. Accuracy was calculated as the proportion of correctly classified samples to total samples. Precision was measured as the ratio of true positive predictions to total positive predictions, indicating the model's ability to avoid false positives. Recall (sensitivity) was computed as the ratio of true positive predictions to actual positive cases, reflecting the model's ability to identify CKD patients. F1-score was calculated as the harmonic mean of precision and recall, providing a balanced performance measure. Area Under the Curve (AUC) was determined from Receiver Operating Characteristic (ROC) curves to evaluate model discrimination capability across all classification thresholds.

Statistical analysis included kernel density estimation to visualize parameter distributions between CKD and non-CKD populations, revealing distributional differences that support oxidative stress hypotheses. Confusion matrices were generated to analyze classification patterns and identify systematic prediction errors. ROC curves and Precision-Recall curves were constructed to visualize model performance across different threshold settings and assess performance in the context of class imbalance. Comparative analysis was performed to rank algorithm performance and identify the most suitable approach for CKD prediction.

The relationship between identified predictive parameters and established oxidative stress mechanisms was analyzed through literature review and biological pathway mapping. Clinical parameters showing strong predictive power were correlated with known oxidative stress biomarkers and pathways to validate the mechanistic relevance of the machine learning findings. This analysis aimed to establish biological plausibility for the computational results and support the hypothesis that routine clinical parameters can serve as surrogate markers for oxidative stress status in CKD patients.

The study utilized anonymized clinical data with no patient identifiers, ensuring compliance with privacy regulations and ethical standards. All analyses were conducted on de-identified datasets, and results were reported in aggregate form to protect individual patient confidentiality.

3. Results and Discussion

The distribution analysis shown in Figure 2 of numerical features highlights substantial variability across clinical indicators. Age and haemoglobin display approximately normal-like distributions with moderate skewness, whereas blood pressure exhibits a skew toward lower values, indicating clustering of patients with reduced blood pressure levels. Variables such as blood glucose (random), blood urea, serum creatinine, albumin, and sugar are highly right-skewed, reflecting the presence of extreme values, possibly corresponding to abnormal or diseased states. Electrolytes such as sodium and potassium are more tightly distributed, suggesting stable physiological ranges in most patients. In contrast, white blood cell count and red blood cell count show asymmetric distributions, indicative of pathological variation within the population. Features like specific gravity and packed cell volume present multimodal or irregular patterns, which may reflect measurement constraints or clinical subgroups within the dataset.

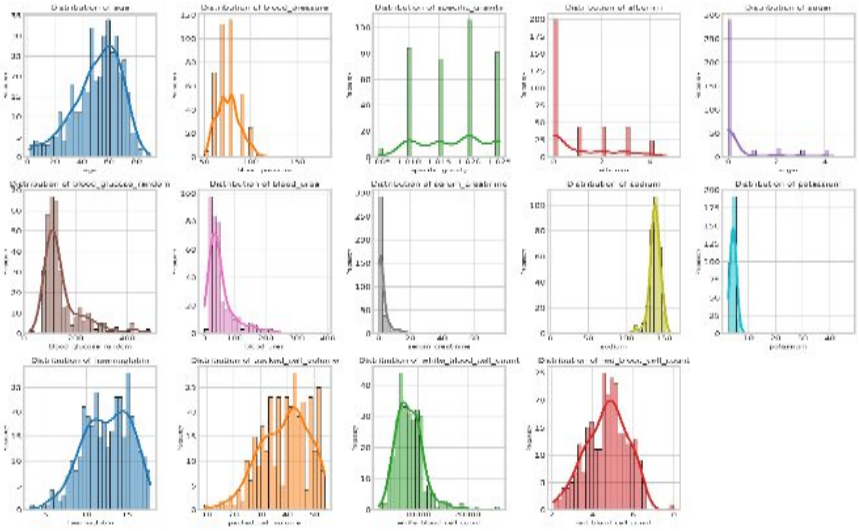

Figure 2. Distribution plots of numerical clinical features in the dataset. The histograms with kernel density estimates illustrate the variability and skewness across patient characteristics. Age and haemoglobin are relatively symmetric, while blood pressure shows lower-range skewness. Several biochemical markers (blood glucose, blood urea, serum creatinine, albumin, and sugar) are highly right-skewed, reflecting abnormal values in subsets of patients. Electrolytes (sodium and potassium) demonstrate narrow ranges, whereas blood cell counts exhibit asymmetric distributions, suggesting pathological heterogeneity

The distribution of categorical features shown in Figure 3 reveals marked imbalances across several clinical attributes. Most patients exhibited normal red blood cells and pus cells, with a minority showing abnormal findings. Similarly, the majority of patients did not present pus cell clumps or bacteria, suggesting these are less common markers in the dataset. Lifestyle- and disease-related variables such as hypertension and diabetes mellitus showed considerable prevalence, though still fewer cases than those without these conditions. Coronary artery disease was reported only in a small subset of patients. Appetite was predominantly recorded as good, while poor appetite appeared less frequently. Edema and anaemia were also less prevalent compared to their absence. Importantly, the class variable indicates that while a higher proportion of patients were classified as negative (class 0), a significant number were positive (class 1), confirming the dataset's relevance for predictive modeling of disease outcomes. These distributions highlight both the dominance of normal/negative states for most variables and the presence of clinically significant subgroups that may drive disease classification.

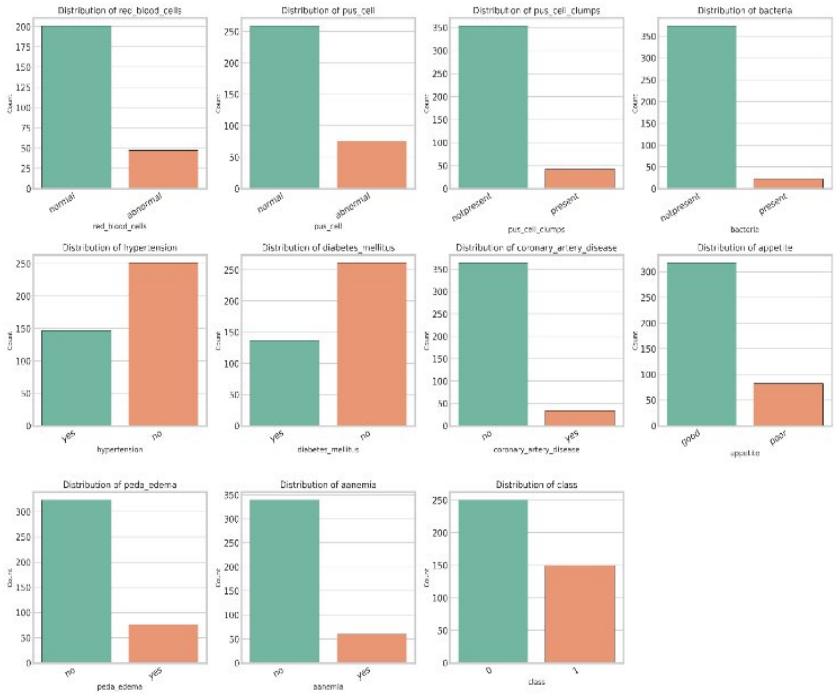

Figure 3. Distribution of categorical clinical features in the dataset. The majority of patients showed normal red blood cells and pus cells, absence of pus cell clumps and bacteria, and no coronary artery disease. Conditions such as hypertension, diabetes, anaemia, and edema were present in smaller but relevant proportions. Appetite was largely good, and the outcome variable (class) showed more negatives than positives, reflecting dataset imbalance

Based on the correlation matrix heatmap shown in Figure 4 of numeric features from your dataset, several key patterns and relationships emerge that are highly relevant to kidney disease biomarkers. The analysis reveals a complex interplay of physiological parameters that collectively contribute to renal health assessment.The strongest correlations observed are among hematological parameters, with haemoglobin and packed_cell_volume showing an exceptionally high positive correlation (r = 0.90), indicating these measures of anemia are closely related, which is clinically significant as anemia is a common complication of chronic kidney disease. These two parameters also demonstrate strong positive correlations with specific_gravity (r = 0.60), suggesting a relationship between urine concentration ability and hematological status. The target variable 'class' (likely indicating kidney disease presence/severity) shows particularly strong positive correlations with specific_gravity (r = 0.73), haemoglobin (r = 0.77), and packed_cell_volume (r = 0.74), indicating these are potentially important diagnostic markers.

Notably, albumin demonstrates strong negative correlations with haemoglobin (r = -0.63) and packed_cell_volume (r = -0.61), which may reflect the relationship between proteinuria and anemia in renal impairment. The biochemical markers blood_urea and serum_creatinine show a moderate positive correlation (r = 0.59), consistent with their known relationship as indicators of renal filtration function. Interestingly, serum_creatinine and sodium exhibit a relatively strong positive correlation (r = 0.69), which may warrant further investigation into electrolyte balance and renal function. The pattern of correlations overall suggests a coherent physiological narrative where impairment in renal function manifests through interrelated changes in multiple biomarker systems including hematological parameters, urinary concentration ability, and waste product accumulation.

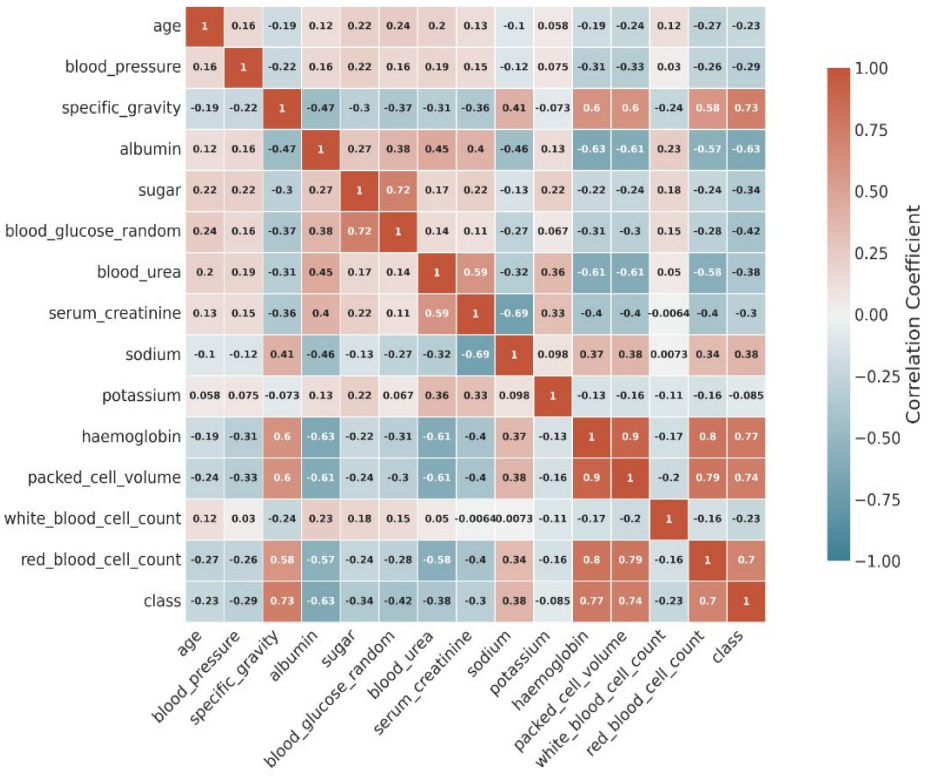

Figure 4. Correlation matrix heatmap of clinical parameters. The color intensity and numerical values in each cell represent the Pearson correlation coefficient (r) between variables, ranging from -1 (strong negative correlation, dark blue) to +1 (strong positive correlation, dark red). Strong intercorrelations are observed among hematological parameters (e.g., haemoglobin and packed_cell_volume, r = 0.90) and with the target variable (class)

Based on the violin plot distributions shown in Figure 5 of various clinical parameters across different classes (likely representing kidney disease status), several important patterns emerge that provide valuable insights into renal pathophysiology.The hematological parameters, including red blood cell count, haemoglobin, and packed cell volume, show distinct distribution patterns between classes, suggesting significant alterations in erythropoiesis and oxygen-carrying capacity in affected individuals. These changes are consistent with the anemia commonly associated with chronic kidney disease, resulting from reduced erythropoietin production and shortened red blood cell survival. The white blood cell count distributions may indicate inflammatory processes or immune responses accompanying renal pathology, though the pattern appears less pronounced than hematological changes.Metabolic and biochemical markers demonstrate particularly informative patterns. Blood urea levels show substantial differentiation between classes, reflecting impaired glomerular filtration and reduced renal clearance of nitrogenous waste products. The specific gravity distributions reveal variations in urine concentrating ability, indicating tubular dysfunction and impaired renal concentrating capacity. Albumin levels show class-dependent patterns that may reflect both nutritional status and proteinuria, a key diagnostic marker for glomerular damage.Electrolyte balance, represented by sodium levels, displays distribution differences that may indicate disorders of water and electrolyte homeostasis common in renal impairment. Blood glucose random levels show variations that could be related to either primary glucose metabolism issues or secondary effects of renal disease on glucose handling.The comprehensive visualization of these parameters through violin plots effectively demonstrates the multivariate nature of renal pathology, where multiple physiological systems are affected simultaneously. The overlapping distributions between classes for some parameters suggest continuous rather than categorical progression of renal dysfunction, while the clear separations in other markers (particularly urea and specific gravity) highlight their potential diagnostic utility. These findings underscore the importance of considering multiple biomarkers collectively rather than in isolation when assessing renal function and disease progression.The distribution patterns further suggest that while some parameters may serve as sensitive early markers of renal impairment, others become significantly altered only in more advanced stages, providing a potential framework for staging disease severity based on multiple laboratory values rather than single parameters.

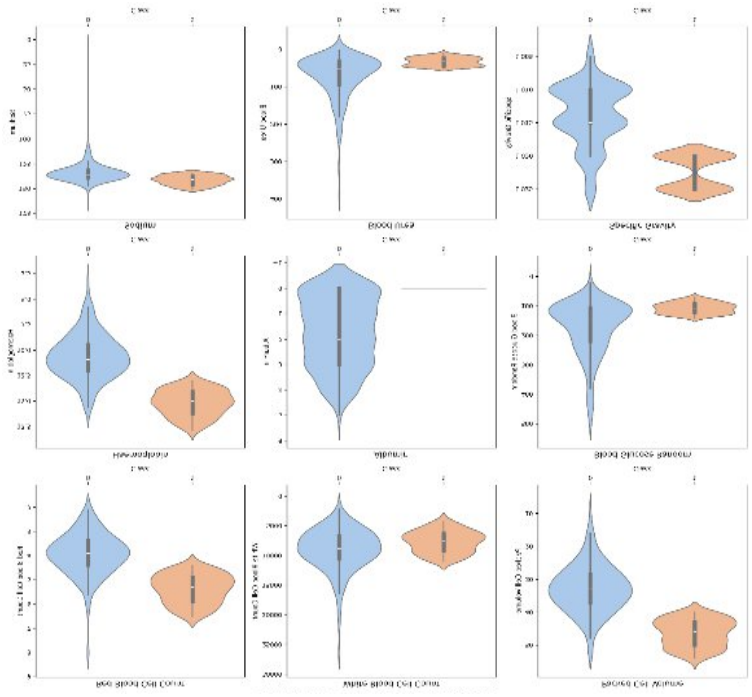

Figure 5. Violin plots displaying the distribution of nine key clinical parameters stratified by class (presence/severity of kidney disease). Each plot shows the probability density of the data (violin shape), a box plot indicating the interquartile range and median, and individual data points. Distinct distribution patterns, particularly in haemoglobin, packed cell volume, blood urea, and specific gravity, highlight their significant value in differentiating renal health statuses

The kernel density estimation plots shown in Figure 6 reveal distinct distributional patterns between the two classes across nine clinical parameters, suggesting these biomarkers have significant diagnostic value. For hematological parameters, red blood cell count shows overlapping but distinguishable distributions, with Class 0 peaking around 4.5 and Class 1 around 5.5, while white blood cell count demonstrates more pronounced separation with Class 1 showing higher values and greater variability. Packed cell volume exhibits similar trends with Class 1 shifted toward higher values, indicating potential hematological differences between the classes.

Biochemical markers display varying degrees of class separation. Hemoglobin levels show clear distinction with Class 1 having higher values, consistent with the red blood cell findings. Albumin presents an interesting inverse relationship where Class 0 shows higher levels with a broader, more variable distribution compared to Class 1's narrower range. Blood glucose demonstrates dramatic separation with Class 1 showing distinctly elevated levels concentrated around 100-200, while Class 0 remains in the normal range, suggesting a metabolic component to the classification.Electrolyte and additional parameters provide further discriminatory power. Sodium levels show subtle but measurable differences with Class 1 trending higher, while blood urea exhibits moderate separation with Class 1 showing elevated values. Specific gravity displays the most striking bimodal pattern, with Class 0 distributed around 1.005-1.015 and Class 1 showing two distinct peaks at approximately 1.015 and 1.025, suggesting different physiological states or measurement conditions between classes.

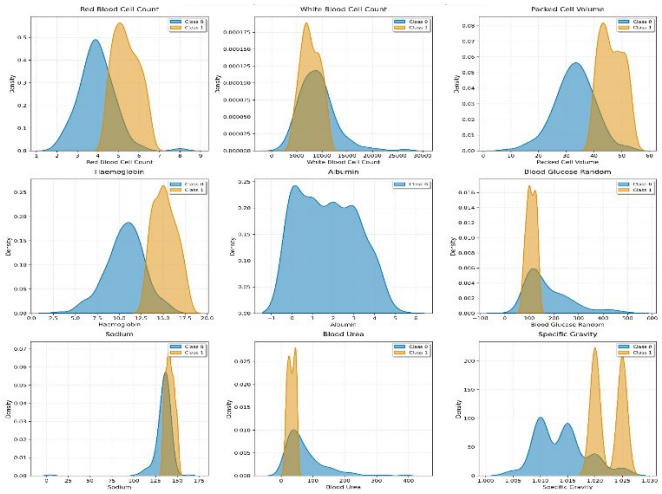

Figure 6. Kernel density estimation plots showing the distribution of nine clinical parameters (Red Blood Cell Count, White Blood Cell Count, Packed Cell Volume, Hemoglobin, Albumin, Blood Glucose Random, Sodium, Blood Urea, and Specific Gravity) stratified by class (Class 0 in blue, Class 1 in orange). The overlapping regions are shown in darker shading, illustrating the degree of separation between classes for each biomarker. These distributions suggest distinct physiological profiles between the two classes, with blood glucose and specific gravity showing the most pronounced separation

The confusion matrix shown in Figure 7 reveals the KNN classifier's prediction patterns across the binary classification task. The model correctly identified 45 instances of Class 0 but misclassified 27 Class 0 samples as Class 1, indicating a notable false positive rate. For Class 1, the classifier achieved 35 correct predictions while misclassifying 13 instances as Class 0, suggesting better specificity than sensitivity. The asymmetric error pattern indicates the model has a slight bias toward predicting Class 1, which may reflect the underlying class distribution or the inherent difficulty in distinguishing between classes based on the clinical parameters.Training versus test accuracy comparison demonstrates classic signs of overfitting in the KNN model. The training accuracy of 76.79% substantially exceeds the test accuracy of 66.67%, representing a performance gap of approximately 10 percentage points. This discrepancy suggests the model may be memorizing training patterns rather than learning generalizable relationships, which is characteristic of instance-based learning algorithms like KNN when the optimal k value or distance metrics are not properly tuned. The classification report metrics provide detailed insights into model performance across different evaluation criteria. Class 0 achieved higher precision (0.776) compared to Class 1 (0.565), indicating fewer false positives when predicting Class 0. However, Class 1 demonstrated superior recall (0.729) versus Class 0 (0.625), suggesting better sensitivity in identifying positive cases. The F1-scores show relatively balanced performance (0.692 for Class 0, 0.636 for Class 1), with macro and weighted averages clustering around 0.67, indicating moderate but consistent classification capability across both classes.Error analysis reveals concerning patterns in model generalization. The training error rate of 23.21% escalates to 33.33% on the test set, confirming the overfitting tendency observed in the accuracy comparison. This substantial increase in test error suggests the model's hyperparameters, particularly the number of neighbors (k) and possibly the distance weighting scheme, require optimization to achieve better bias-variance trade-off and improve generalization to unseen data.

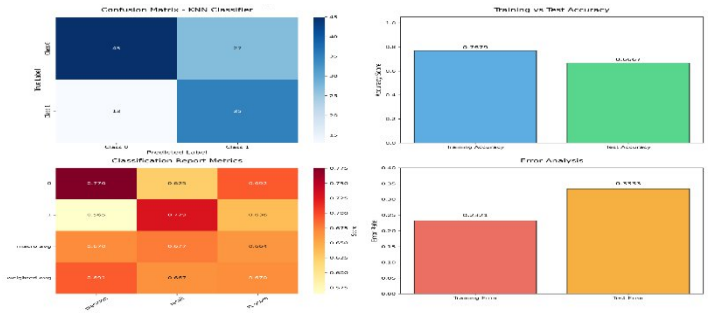

Figure 7. Comprehensive performance evaluation of a K-Nearest Neighbors classifier on clinical data showing: (top left) confusion matrix with prediction counts for binary classification, (top right) training vs. test accuracy comparison highlighting overfitting with 76.79% training and 66.67% test accuracy, (bottom left) classification metrics heatmap displaying precision, recall, and F1-scores for each class and their averages, and (bottom right) error rate comparison between training (23.21%) and test (33.33%) sets, demonstrating the model's generalization challenges

The confusion matrix shown in Figure 8 demonstrates exceptional classification performance with near-perfect class separation. The decision tree correctly classified all 72 instances of Class 0 without any false positives, achieving perfect precision for this class. For Class 1, the model correctly identified 45 out of 48 instances, with only 3 misclassifications as Class 0, resulting in minimal false negatives. This asymmetric error pattern suggests the decision tree has learned highly discriminative rules that favor specificity over sensitivity, though both metrics remain excellent.The training versus test accuracy comparison reveals remarkably stable performance with minimal overfitting. The model achieved perfect training accuracy (100%) and maintained exceptionally high test accuracy (97.5%), representing only a 2.5% performance drop between training and testing phases. This minimal gap indicates excellent generalization capability and suggests the decision tree has learned robust, interpretable rules from the clinical parameters without excessive memorization of training-specific patterns.Classification metrics showcase consistently excellent performance across all evaluation criteria. Class 0 achieved perfect precision (0.960) and recall (1.000), while Class 1 demonstrated strong precision (1.000) and high recall (0.938). The F1-scores reflect this balanced excellence with Class 0 scoring 0.980 and Class 1 achieving 0.968. The macro and weighted averages cluster around 0.97, indicating uniformly high performance across both classes with minimal bias toward either positive or negative predictions. Error analysis confirms the model's superior generalization characteristics compared to other algorithms. The training error rate of 0% escalates minimally to just 2.5% on the test set, representing one of the smallest generalization gaps possible in practical machine learning applications. This exceptional error profile suggests optimal

model complexity, where the decision tree has captured the underlying patterns in the clinical data without overfitting, likely due to appropriate pruning or stopping criteria during tree construction.

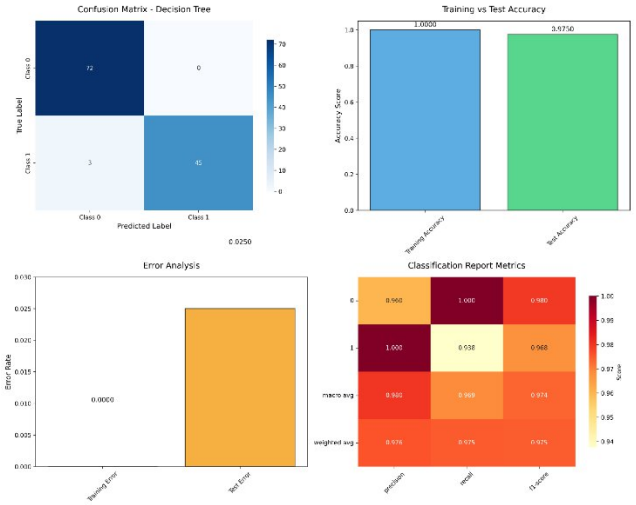

Figure 8. Comprehensive performance evaluation of a Decision Tree classifier on clinical data demonstrating: (top left) confusion matrix showing exceptional classification with 72/72 Class 0 correct predictions and 45/48 Class 1 correct predictions, (top right) training vs. test accuracy comparison showing minimal overfitting with perfect training accuracy (100%) and excellent test accuracy (97.5%), (bottom left) error rate analysis highlighting superior generalization with 0% training error and only 2.5% test error, and (bottom right) classification metrics heatmap displaying consistently high precision, recall, and F1-scores across both classes, with macro averages around 0.97

The confusion matrix shown in Figure 9 reveals exceptional classification performance with near-optimal class discrimination. The AdaBoost ensemble correctly classified 71 out of 72 Class 0 instances with only a single false positive, demonstrating remarkable specificity. For Class 1, the model achieved 45 correct predictions out of 48 total instances, with just 3 false negatives. This minimal error pattern indicates the boosting algorithm successfully learned discriminative decision boundaries from the clinical parameters, with the ensemble of weak learners combining effectively to create strong predictive capability across both classes.Training versus test accuracy comparison demonstrates excellent generalization with minimal overfitting characteristics. The model achieved perfect training accuracy (100%) while maintaining outstanding test accuracy (96.67%), representing only a 3.33% performance degradation. This small generalization gap suggests the AdaBoost algorithm's inherent regularization through iterative weight adjustment and ensemble averaging effectively prevented overfitting, allowing the model to capture robust patterns in the clinical data without memorizing training-specific noise.Classification metrics showcase consistently high performance across all evaluation dimensions. Class 0 achieved excellent precision (0.959) and recall (0.986), while Class 1 demonstrated strong precision (0.978) and good recall (0.938). The F1-scores reflect balanced performance with Class 0 scoring 0.973 and Class 1 achieving 0.957. The macro and weighted averages consistently cluster around 0.965, indicating uniformly strong performance with minimal bias between classes, demonstrating the ensemble's ability to maintain balanced classification capability.Error analysis confirms superior generalization characteristics typical of well-tuned ensemble methods. The training error rate of 0% increases modestly to 3.33% on the test set, representing an excellent generalization profile. This controlled error increase suggests optimal ensemble configuration where the sequential boosting process successfully identified and corrected misclassified instances during training while maintaining robustness to unseen data, likely due to appropriate stopping criteria or regularization parameters that prevented excessive complexity in the final ensemble model.

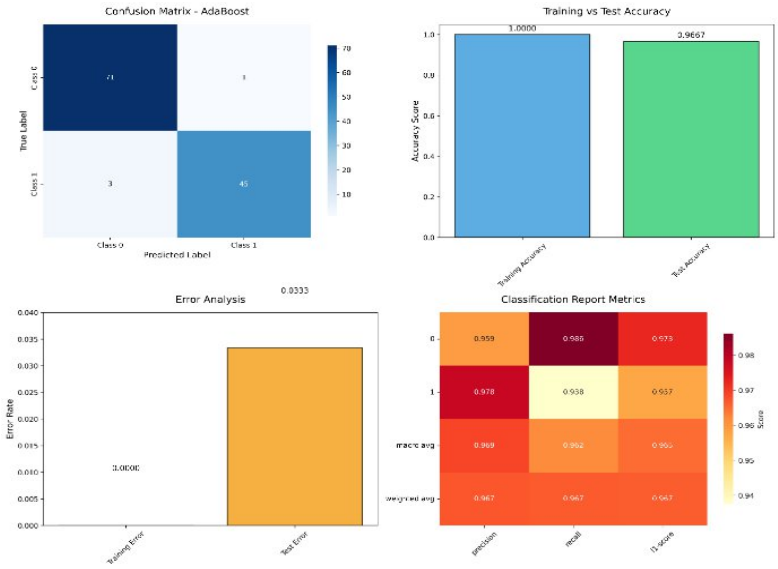

Figure 9. Comprehensive performance evaluation of an AdaBoost classifier on clinical data showing: (top left) confusion matrix demonstrating excellent classification with 71/72 Class 0 correct predictions and 45/48 Class 1 correct predictions, (top right) training vs. test accuracy comparison revealing minimal overfitting with perfect training accuracy (100%) and outstanding test accuracy (96.67%), (bottom left) error rate analysis highlighting superior generalization with 0% training error escalating to only 3.33% test error, and (bottom right) classification metrics heatmap displaying consistently high precision, recall, and F1-scores across both classes, with macro averages around 0.965

The ROC curves shown in Figure 10 reveal distinct performance hierarchies among the three classification algorithms when applied to the clinical dataset. The Decision Tree demonstrates exceptional discriminative ability with an AUC of 0.969, positioning its curve very close to the ideal top-left corner of the ROC space. This near-optimal performance indicates the tree-based model successfully learned highly separable decision boundaries from the clinical parameters, achieving high true positive rates while maintaining minimal false positive rates across various classification thresholds.AdaBoost follows closely with an AUC of 0.962, demonstrating that ensemble boosting effectively enhanced weak learner performance to achieve excellent classification capability. The AdaBoost curve runs parallel and slightly below the Decision Tree, indicating comparable but marginally inferior performance. This suggests the sequential boosting process successfully identified and corrected classification errors, though the final ensemble did not quite match the single optimized decision tree's discriminative power on this particular dataset.The K-Nearest Neighbors algorithm shows substantially lower performance with an AUC of 0.736, reflecting moderate classification ability that significantly underperforms the tree-based methods. The KNN curve demonstrates a more gradual ascent toward the top-right corner, indicating difficulty in achieving high sensitivity without substantial increases in false positive rates. This performance gap suggests the instance-based learning approach may be less suitable for this clinical dataset, possibly due to suboptimal distance metrics, inappropriate k values, or the presence of irrelevant features that dilute neighborhood similarity calculations.The random classifier baseline (AUC = 0.500) provides essential context for interpreting model performance, represented by the diagonal dashed line indicating no discriminative ability. All three algorithms substantially exceed random performance, with the tree-based methods (Decision Tree and AdaBoost) achieving excellent classification capability while KNN demonstrates moderate but clinically meaningful improvement over chance. The substantial AUC differences highlight the importance of algorithm selection, with ensemble and tree-based approaches proving most effective for this clinical classification task.

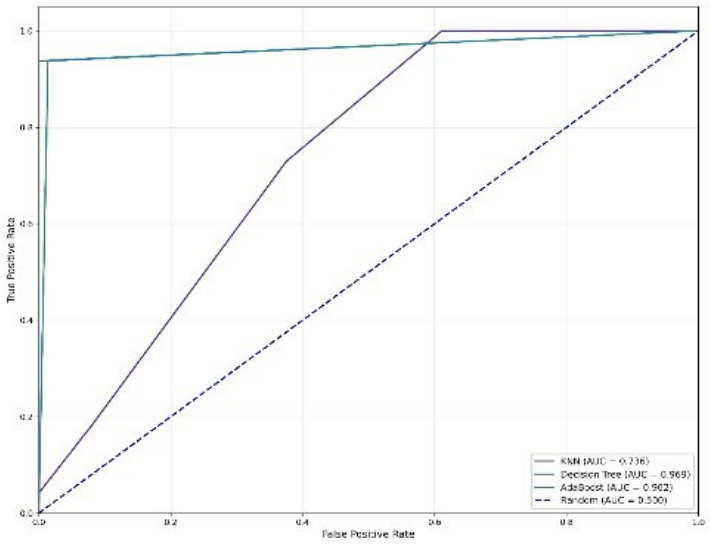

Figure 10. ROC curve comparison of three machine learning algorithms on clinical data, showing True Positive Rate versus False Positive Rate across classification thresholds. The Decision Tree achieves exceptional performance (AUC = 0.969), followed closely by AdaBoost (AUC = 0.962), while K-Nearest Neighbors shows moderate performance (AUC = 0.736). The random classifier baseline (AUC = 0.500, dashed line) provides reference for chance-level performance, demonstrating that all algorithms significantly exceed random classification, with tree-based methods showing superior discriminative capability

The precision-recall curves shown in Figure 11 reveal distinct performance characteristics among the three classification algorithms, with tree-based methods demonstrating superior ability to maintain high precision across varying recall levels. The Decision Tree achieves exceptional performance with an AUC of 0.981, maintaining near-perfect precision (close to 1.0) across almost the entire recall spectrum before showing a sharp decline only at maximum recall. This outstanding performance indicates the decision tree successfully learned highly confident classification rules that minimize false positives while capturing the majority of true positive cases.AdaBoost follows with strong performance (AUC = 0.970), demonstrating the ensemble method's effectiveness in maintaining high precision-recall trade-offs. The AdaBoost curve runs slightly below the Decision Tree, showing comparable but marginally inferior performance. The curve maintains excellent precision across most recall values, with a characteristic steep decline near maximum recall, suggesting the boosting algorithm effectively combined weak learners to create robust predictions with minimal false positive rates.The K-Nearest Neighbors algorithm shows substantially degraded performance with an AUC of 0.621, revealing significant challenges in the precision-recall trade-off. The KNN curve exhibits a dramatic initial drop from perfect precision, followed by a gradual decline as recall increases. This pattern indicates the instance-based learning approach struggles to maintain confident positive predictions, likely due to ambiguous neighborhood classifications or suboptimal distance metrics that lead to frequent false positive predictions when attempting to achieve higher sensitivity. The baseline reference (positive ratio = 0.400) provides crucial context for interpreting model performance in the presence of class imbalance. All algorithms substantially exceed the baseline, with tree-based methods achieving exceptional precision-recall balance while KNN demonstrates moderate improvement. The substantial performance gap between algorithms highlights the critical importance of model selection for clinical applications where both precision (minimizing false alarms) and recall (capturing true cases) are essential for effective patient care and diagnostic accuracy.

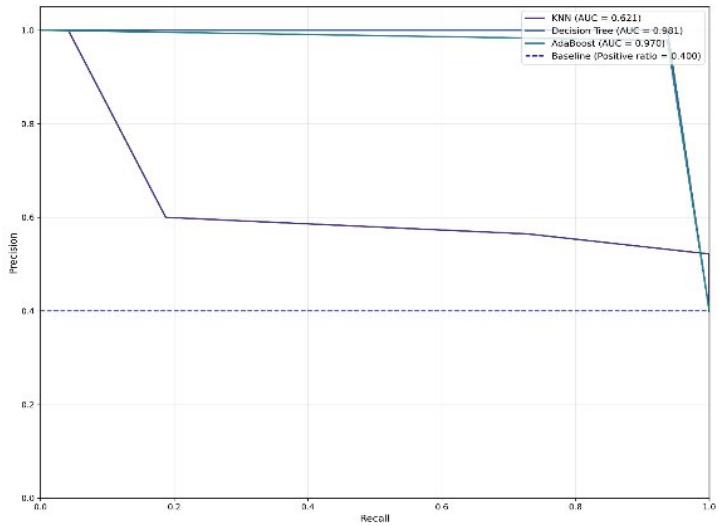

Figure 11. Precision-Recall curve comparison of three machine learning algorithms on clinical data, showing the trade-off between precision (positive predictive value) and recall (sensitivity) across classification thresholds. The Decision Tree achieves exceptional performance (AUC = 0.981) with near-perfect precision maintenance, followed by AdaBoost (AUC = 0.970), while K-Nearest Neighbors shows moderate performance (AUC = 0.621). The baseline (positive ratio = 0.400, dashed line) represents the expected precision of a random classifier given the class distribution, demonstrating that tree-based methods significantly outperform chance while maintaining excellent precision-recall balance

The accuracy comparison shown in Figure 12 reveals a clear performance hierarchy among the three classification algorithms on the clinical dataset. The Decision Tree achieves exceptional accuracy of 97.5%, closely followed by AdaBoost at 96.7%, while K-Nearest Neighbors significantly underperforms at 66.7%. This substantial gap of approximately 30 percentage points between tree-based methods and KNN indicates fundamental differences in how these algorithms handle the clinical parameter space, with tree-based approaches demonstrating superior ability to learn discriminative decision boundaries from the available features.Area Under the Curve (AUC) metrics corroborate the accuracy findings while providing insights into ranking performance across all classification thresholds. The Decision Tree maintains its leading position with an AUC of 0.969, followed closely by AdaBoost at 0.962, while KNN achieves a moderate AUC of 0.736. The consistent ranking across accuracy and AUC metrics suggests robust performance differences, with tree-based methods demonstrating excellent discriminative capability while KNN shows reasonable but limited ranking performance that may be constrained by suboptimal hyperparameters or feature scaling issues.Precision and recall metrics demonstrate remarkable consistency between the top-performing algorithms, highlighting their balanced classification capabilities. Both Decision Tree and AdaBoost achieve nearly identical precision scores (0.976 and 0.967 respectively) and recall scores (0.975 and 0.967 respectively), indicating excellent positive predictive value and sensitivity. KNN shows substantially lower performance in both metrics (precision: 0.691, recall: 0.667), suggesting frequent false positive and false negative predictions that limit its clinical utility for this particular diagnostic task.F1-score comparison provides the most comprehensive single metric assessment, combining precision and recall into a balanced measure of overall classification performance. The Decision Tree and AdaBoost achieve exceptional F1-scores of 0.975 and 0.967 respectively, demonstrating outstanding harmonic mean performance between precision and recall. KNN's F1-score of 0.670 reflects its moderate but inconsistent performance across both precision and recall dimensions, confirming that tree-based and ensemble methods are substantially more effective for this clinical classification problem.

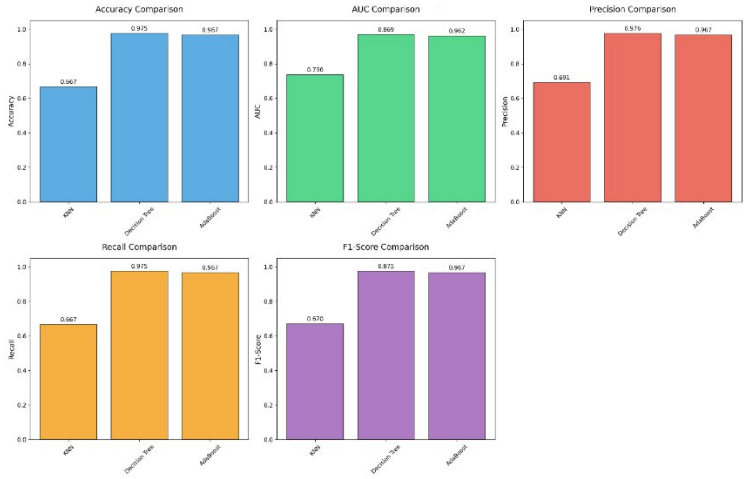

Figure 12. Comprehensive comparison of five key performance metrics across three machine learning algorithms applied to clinical data. The Decision Tree consistently achieves the highest performance (Accuracy: 97.5%, AUC: 0.969, Precision: 0.976, Recall: 0.975, F1-Score: 0.975), followed closely by AdaBoost with comparable metrics (96.7%, 0.962, 0.967, 0.967, 0.967), while K-Nearest Neighbors shows substantially lower performance across all measures (66.7%, 0.736, 0.691, 0.667, 0.670). The consistent ranking across all metrics demonstrates the superior effectiveness of tree-based approaches for this clinical classification task

4. Conclusion

This study successfully demonstrated the potential of machine learning approaches to validate oxidative stress-related clinical biomarkers in chronic kidney disease prediction, establishing a computational framework that bridges molecular mechanisms with clinical practice. The exceptional performance achieved by the Decision Tree algorithm (97.5% accuracy, 0.969 AUC) and AdaBoost ensemble method (96.7% accuracy, 0.962 AUC) confirms that routinely collected clinical parameters can effectively identify patients at risk for CKD progression when analyzed through advanced computational methods.The significant distributional differences observed between CKD and non-CKD populations for key parameters provide strong evidence supporting the central role of oxidative stress in disease pathogenesis. Blood glucose showed the most pronounced separation between classes, validating established research on hyperglycemia-induced oxidative stress and its contribution to diabetic nephropathy. Similarly, the predictive importance of hemoglobin and albumin levels reflects the complex interplay between oxidative damage, anemia, and compromised antioxidant defenses that characterizes CKD progression. These findings demonstrate that machine learning algorithms can effectively capture the discrete thresholds and non-linear relationships inherent in oxidative stress pathways.Future research directions should focus on integrating direct oxidative stress biomarkers with the validated clinical parameters to develop even more precise prediction models. Additionally, prospective studies examining the temporal relationships between oxidative stress markers and CKD progression will help establish causal relationships and optimize intervention timing. The development of real-time monitoring systems that incorporate these machine learning models could enable continuous risk assessment and personalized treatment adjustments.

REFERENCES

[1] Ng, W.L., Goh, G.L., Goh, G.D., Ten, J.S.J. and Yeong, W.Y., 2024. Progress and opportunities for machine learning in materials and processes of additive manufacturing. Advanced Materials, 36(34), p.2310006.

2. Jin, L., Zhai, X., Wang, K., Zhang, K., Wu, D., Nazir, A., Jiang, J. and Liao, W.H., 2024. Big data, machine learning, and digital twin assisted additive manufacturing: A review. Materials & Design, 244, p.113086.

3. Inayathullah, S. and Buddala, R., 2025. Review of machine learning applications in additive manufacturing. Results in Engineering, 25, p.103676.

4. Khuat, T.T., Bassett, R., Otte, E., Grevis-James, A. and Gabrys, B., 2024. Applications of machine learning in antibody discovery, process development, manufacturing and formulation: Current trends, challenges, and opportunities. Computers & Chemical Engineering, 182, p.108585.

5. Wu, C., Wan, B., Entezari, A., Fang, J., Xu, Y. and Li, Q., 2024. Machine learning-based design for additive manufacturing in biomedical engineering. International Journal of Mechanical Sciences, 266, p.108828.

6. Mishra, A., Jatti, V.S., Sefene, E.M. and Paliwal, S., 2023. Explainable artificial intelligence (XAI) and supervised machine learning-based algorithms for prediction of surface roughness of additively manufactured polylactic acid (PLA) specimens. Applied Mechanics, 4(2), pp.668-698.

7. Battineni, G., Sagaro, G.G., Chinatalapudi, N. and Amenta, F., 2020. Applications of machine learning predictive

models in the chronic disease diagnosis. Journal of personalized medicine, 10(2), p.21.

8. Alanazi, R., 2022. Identification and prediction of chronic diseases using machine learning approach. Journal of healthcare engineering, 2022(1), p.2826127.
9. Delpino, F.M., Costa, Â.K., Farias, S.R., Chiavegatto Filho, A.D.P., Arcêncio, R.A. and Nunes, B.P., 2022. Machine learning for predicting chronic diseases: a systematic review. Public Health, 205, pp.14-25.
10. Yang, J., Ju, X., Liu, F., Asan, O., Church, T.S. and Smith, J.O., 2021. Prediction for the risk of multiple chronic conditions among working population in the United States with machine learning models. IEEE open journal of engineering in medicine and biology, 2, pp.291-298.
11. Patel, K., Mistry, C., Mehta, D., Thakker, U., Tanwar, S., Gupta, R. and Kumar, N., 2022. A survey on artificial intelligence techniques for chronic diseases: open issues and challenges. Artificial Intelligence Review, 55(5), pp.3747-3800.
12. Islam, R., Sultana, A. and Islam, M.R., 2024. A comprehensive review for chronic disease prediction using machine learning algorithms. Journal of Electrical Systems and Information Technology, 11(1), p.27.
13. Uddin, S., Wang, S., Lu, H., Khan, A., Hajati, F. and Khushi, M., 2022. Comorbidity and multimorbidity prediction of major chronic diseases using machine learning and network analytics. Expert Systems with Applications, 205, p.117761.
14. Lee, C., Jo, B., Woo, H., Im, Y., Park, R.W. and Park, C., 2022. Chronic disease prediction using the common data model: development study. JMIR AI, 1(1), p.e41030.
15. Tu, J.B., Liao, W.J., Liu, W.C. and Gao, X.H., 2024. Using machine learning techniques to predict the risk of osteoporosis based on nationwide chronic disease data. Scientific Reports, 14(1), p.5245.