Wadi Alshatti University Journal of
Pure and Applied Sciences

مجلة جامعة وادي الشاطئ
للعلوم البحتة والتطبيقية

Volume 4, No. 1, January-June 2026 Online ISSN: 3006-0877 المجلد 4، الاصدار 1، يناير- يونيو 2026

RESEARCH ARTICLE **BIOMATERIALS**

Hybrid Matrix-Ensemble Framework for Chronic Kidney Disease Diagnosis

Fatma S. Elghaffi[1], Othman E. Mohammed[1], Llahm O. Ben Dalla[2,3,*], Abdussalam A. Ahmed[4], Almhdie A. Agila[5], Mohamed A. EL-Sseid[6]

[1]Computer Science Department, Higher Institute of Science and Technology, Ajdabiya, Libya
[2]Computer Engineering, College of Technical Science, Sebha, Libya
[3]Department of Electric Electronics, Ankara Yildirim Beyazit University, Türkiye
[4]Mechanical and Industrial Engineering Department, Bani Waleed University, Libya
[5]Department of Computer Science, College of Technical Science Sebha, Libya
[6]Department of Software Engineering, Ankara Bilim University, Türkiye

ARTICLE HISTORY
Received 28 December 2025
Revised 15 February 2026
Accepted 25 February 2026
Online 05 March 2026

KEYWORDS
Kidney disease;
Machine learning;
Mathematical matrices;
Hybrid framework;
Clinical decision support.

ABSTRACT

This study introduces a novel Hybrid Machine Learning and Mathematical Matrix Framework (HML-MMF) for optimizing the early detection of Chronic Kidney Disease (CKD), addressing critical limitations of conventional machine learning approaches, namely, poor interpretability, overfitting, and instability in clinical settings. The proposed framework uniquely integrates matrix algebra techniques, including column-wise mean imputation, Principal Component Analysis (PCA), Fisher Discriminant Analysis (FDA), and Min-Max normalization, with an ensemble of three classifiers: Gradient Boosting Machine (GBM), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). Applied to the UCI CKD dataset (400+ patients, 24 clinical features), the HML-MMF pipeline transforms raw clinical data into a mathematically structured, low-dimensional representation that enhances class separability while preserving physiological meaning. The final diagnosis is derived through soft voting, ensuring robustness and generalizability. Experimental results demonstrate that the hybrid model achieves 96.2% accuracy, 96.0% precision, 95.5% recall, 95.7% F1-score, and an AUC-ROC of 0.973, significantly outperforming both GBM-only (AUC = 0.942) and Matrix + SVM (AUC = 0.955) baselines. The scientific novelty lies in the synergistic fusion of interpretable matrix operations with ensemble learning, not as sequential steps, but as a unified architecture where mathematical transformations actively guide model optimization. This approach not only boosts performance but also provides clinical transparency, enabling practitioners to trace predictions back to key biomarkers like serum creatinine and BUN, which eigenvalue analysis confirms as dominant contributors. In the medical domain, this work offers a reliable, explainable decision-support tool that minimizes false negatives a critical requirement in CKD screening. Via bridging rigorous linear algebra with modern ML, the HML-MMF sets a new standard for trustworthy, high-performance diagnostic systems in resource-constrained or data-imbalanced clinical environments.

إطار عمل مجموعة المصفوفات الهجينة لتشخيص أمراض الكلى المزمنة

فاطمة القذافي[1]، عثمان عمران بلقاسم محمد[2]، للاهم عمر بن دلة[3،*]، عبدالسلام علي أحمد[4]، المهدي أبوبكر احمد عقيلة[5]، محمد علي محمد الصيد[6]

الكلمات المفتاحية
أمراض الكلى
التعلم الآلي
المصفوفات الرياضية
إطار عمل هجين
دعم القرار السريري

الملخص

تقدم هذه الدراسة إطار عمل هجين جديد للتعلم الآلي والمصفوفات الرياضية (HML-MMF) لتحسين الكشف المبكر عن مرض الكلى المزمن (CKD)، متجاوزةً بذلك القيود الجوهرية لأساليب التعلم الآلي التقليدية، وهي ضعف قابلية التفسير، والتخصيص الزائد، وعدم الاستقرار في البيئات السريرية. يدمج إطار العمل المقترح بشكل فريد تقنيات جبر المصفوفات، بما في ذلك استكمال المتوسط العمودي، وتحليل المكونات الرئيسية (PCA)، وتحليل التمييز لفيشر (FDA)، والتطبيع الأدنى-الأقصى، مع مجموعة من ثلاثة مصنفات: آلة تعزيز التدرج (GBM)، وآلة المتجهات الداعمة (SVM)، وخوارزمية أقرب الجيران (KNN). عند تطبيقه على مجموعة بيانات UCI CKD (أكثر من 400 مريض، 24 سمة سريرية)، يحوّل مسار HML-MMF البيانات السريرية الخام إلى تمثيل منخفض الأبعاد ذي بنية رياضية، مما يعزز فصل الفئات مع الحفاظ على المعنى الفيزيولوجي. يُستخلص التشخيص النهائي من خلال التصويت المرن، مما يضمن المتانة وقابلية التعميم. تُظهر النتائج التجريبية أن النموذج الهجين يحقق دقة بنسبة 96.2%، ودقة بنسبة 96.0%، واستدعاء بنسبة 95.5%، ودرجة F1 بنسبة 95.7%، ومساحة تحت منحنى ROC تبلغ 0.973، متفوقًا بشكل ملحوظ على كل من نموذج GBM وحده (مساحة تحت المنحنى = 0.942) ونموذج المصفوفة + SVM (مساحة تحت المنحنى = 0.955). تكمن الجدة العلمية في الدمج التآزري

---

*Corresponding author

https://doi.org/10.63318/waujpasv4i1_28

لعمليات المصفوفات القابلة للتفسير مع التعلم الجماعي، ليس كخطوات متسلسلة، بل كبنية موحدة حيث توجه التحويلات الرياضية عملية تحسين النموذج بشكل فعال. لا يعزز هذا النهج الأداء فحسب، بل يوفر أيضًا شفافية سريرية، مما يُمكّن الممارسين من تتبع التنبؤات إلى المؤشرات الحيوية الرئيسية مثل الكرياتينين في الدم ونيتروجين اليوريا في الدم، والتي يؤكد تحليل القيم الذاتية أنها عوامل رئيسية مؤثرة. في المجال الطبي، يقدم هذا العمل أداة دعم قرار موثوقة وقابلة للتفسير، تقلل من النتائج السلبية الخاطئة، وهو شرط أساسي في فحص مرض الكلى المزمن. ومن خلال الجمع بين الجبر الخطي الدقيق والتعلم الآلي الحديث، يضع نظام HML-MMF معيارًا جديدًا لأنظمة التشخيص الموثوقة وعالية الأداء في البيئات السريرية ذات الموارد المحدودة أو البيانات غير المتوازنة.

Introduction

Chronic Kidney Disease (CKD) affects approximately 10% of the global population and is often underdiagnosed due to subtle early-stage symptoms [1] as presented in Table 1. Furthermore, conventional diagnostic approaches, though clinically validated, are limited via their dependency on manual interpretations and static thresholds [1,2]. Recent advances in machine learning (ML) have demonstrated promise in automating disease detection processes [3-6]. However, pure ML models can suffer from overfitting, lack of interpretability [7-10], and instability when applied to heterogeneous clinical datasets [11-14]. To address these limitations, this paper presents a novel Hybrid Machine Learning and Mathematical Matrix Framework (HML-MMF) that fuses the predictive power of ML algorithms with the structure and rigor of matrix algebra [15-18]. In addition, the underlying hypothesis is that matrix factorization and transformation techniques can enhance data pre-processing [19-22], feature selection, and dimensionality reduction, boosting the performance of ML models. However, pure ML models often overfit heterogeneous clinical data due to noise, distorting feature relationships and causing unstable predictions [23-25]. Moreover, their black-box nature lacks interpretability, hindering clinical trust and validation [26, 27]. This gap necessitates a transparent framework integrating mathematical structure with machine learning to ensure reliable clinical deployment.

Table 1: the dataset description (https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease)

Attribute	Description
Source	ALafiya Hospital, Hoon City, Libya (as referenced in prior user context) and UCI Machine Learning Repository (https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease)
Repository	UCI ML Repository Chronic Kidney Disease Dataset
Number of Instances (Patients)	400+ (approximately 400 patients)
Number of Features (Clinical Indicators)	24
Key Clinical Features	Blood Pressure (BP), Serum Creatinine, Blood Urea Nitrogen (BUN), Sodium, Potassium, Glomerular Filtration Rate (GFR), etc.
Target Variable	Chronic Kidney Disease status: Yes / No
Data Type	Mixed (numerical and categorical)
Missing Values	handled via matrix-based mean projection imputation
Class Distribution	Imbalanced (common in medical datasets); addressed via ensemble voting and robust preprocessing

Related Work

Various studies have employed ML for CKD diagnosis. Support Vector Machines (SVM), Decision Trees, and ensemble models like Random Forest have shown high classification accuracy [20]. Matrix theory, particularly Singular Value Decomposition (SVD) and Principal Component Analysis (PCA), has been widely utilized in biomedical signal processing Xu & Zhang, [14], but its integration with ML for kidney disease diagnosis remains underexplored. The novelty of this research lies in creating a synergistic architecture where matrix operations serve both as data transformation tools and structural guides for model optimization.

Machine Learning Applications in Kidney Disease Detection

The application of machine learning (ML) to kidney disease detection has garnered increasing research interest over the past decade. Traditional diagnostic approaches rely heavily on clinical expertise and predefined thresholds for laboratory markers , for instance, serum creatinine, blood urea nitrogen (BUN), and glomerular filtration rate (GFR). However, such methods often fail to detect subtle or early-stage manifestations of Chronic Kidney Disease (CKD) [9]. Recent studies have demonstrated that ML models, particularly Support Vector Machines (SVMs), Random Forests, and Gradient Boosting Machines (GBMs), outperform conventional clinical decision-making tools via learning complex patterns directly from high-dimensional datasets as declared by Almansour et al.[20]. For instance, Almansour [20] and colleagues applied ensemble learning methods to a CKD dataset and achieved notable improvements in classification accuracy over standalone algorithms. Moreover, integration of ensemble models has been shown to enhance robustness via reducing variance and mitigating overfitting risks, particularly when applied to relatively small clinical datasets Chen and Guestrin, [9]. Nevertheless, challenges persist regarding model interpretability and generalizability across diverse patient populations. Purely data-driven models often lack transparency, making clinical acceptance difficult Sun et al., [13]. These limitations motivate the exploration of hybrid strategies that integrate data-driven learning with mathematically grounded frameworks to enhance both performance and interpretability.

Matrix-Based Approaches in Biomedical Data Processing

Matrix theory has long been pivotal in biomedical engineering for tasks , for instance, signal decomposition, image reconstruction, and noise filtering. Techniques like Singular Value Decomposition (SVD) and Principal Component Analysis (PCA) have been extensively utilized to reduce data dimensionality while preserving significant information Xu & Zhang, [14]. PCA, in particular, projects high-dimensional data onto orthogonal components that capture maximum variance, making it a powerful tool for pre-processing clinical datasets. Fisher Discriminant Analysis (FDA) further enhances class separability via maximizing the ratio of between-class to dispersion throughout the class, Liu et al., [16]. Such techniques offer not just computational efficiency but also increased visualization and interpretability of complex datasets. Despite these benefits, there is still much to learn about integrating matrix approaches with machine learning pipelines for clinical decision support, especially when it comes to the detection of renal disease. In order to improve ML classifier performance, recent research supports the usage of hybrid models in which matrix transformations serve as an intermediate representation (Zhang et al., [18]). Nevertheless, these studies frequently ignore practical issues like noise, class imbalance, and missing data that frequently arise in medical datasets.

Hybrid Machine Learning and Matrix Frameworks

The new concept of hybrid frameworks, which blend ML flexibility with mathematical structure, has demonstrated encouraging outcomes in a number of medicinal applications. These systems make use of the power of structured dimensionality reduction while enabling sophisticated classifiers to extract intricate patterns. Abdar and others [8]. A few recent research have started combining PCA-based feature transformation with SVM or ensemble models in the field of renal disease diagnosis in order to increase accuracy and robustness as declared by Liu et al., [16]. However, the majority of current models view machine learning and matrix-based preprocessing as sequential rather than complementary processes. Moreover, the methodical combination of ensemble learning algorithms (GBM, SVM, KNN) as well as matrix algebra techniques (PCA, FDA, normalization, eigenvalue decomposition) to create an optimized predictive framework specifically designed for CKD diagnosis is what makes this research interesting [17,18]. Furthermore, the hybrid approach's dependability is further improved via addressing practical issues like balancing class distributions through ensemble voting mechanisms and missing value imputation through matrix-based projections, which makes it a strong contender for clinical implementation [19]. There is still a big gap in the systematic integration of machine learning and mathematical modeling for clinical prediction tasks, despite advancements in both fields separately. Previous studies either use matrix techniques primarily for basic dimensionality reduction without fully integrating their structural advantages into model optimization, or they depend solely on black-box machine learning models [20]. Via putting forth a unified hybrid architecture that not only improves predictive accuracy but also offers interpretability, resilience, and computing efficiency qualities essential for practical clinical applications this study seeks to close that gap. In addition, the suggested HML-MMF architecture's incorporation of machine learning and matrix theory is a viable way to raise the bar for renal disease detection.

Methodology

The study employed a 10-fold cross-validation protocol combined with Bayesian optimization to partition the UCI CKD dataset, ensuring that each patient sample contributed to both (80% training and 20% testing) training and validation phases while minimizing selection bias and overfitting risks. This stratified resampling approach preserved the original class distribution across folds, thereby strengthening the statistical validity and generalizability of the reported metrics (accuracy: 96.2%, AUC-ROC: 0.973) across heterogeneous clinical subpopulations as presented in Figure 1 and Figure 2.

Data Acquisition and Preprocessing

This research utilized a benchmark renal illness dataset from the UCI Machine Learning Repository. The collection contains more than 400 examples with 24 clinical characteristics, including blood pressure, serum creatinine, sodium levels, etc. Missing values were imputed utilizing matrix-based mean projection methods to preserve statistical coherence.

The Mathematical Matrix Framework

Let $X \in \mathbb{R}^{n\times m}$ be the input data matrix, where n is the number of patients as well as m is the number of features. Dimensionality reduction was performed utilizing PCA has been calculated as below:

$X = U\Sigma V^T$, retaining components where cumulative explained variance $> 95\%$. Fisher Discriminant Analysis (FDA) [21]:

Maximizes the ratio $\frac{|\mu_1-\mu_2|^2}{\sigma_1^2+\sigma_2^2}$ for class separability [3].

Data Representation as Matrix

Let $\mathbf{X} \in \mathbb{R}^{m\times n}$ be the original clinical dataset, where:

$m =$ number of patients $400 +$)

$n =$ number of features , 24 clinical indicators like BP, serum creatinine, etc.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

For missing values, a column-wise mean imputation is used as below as presented in Table 2 [4,22].

$$x_{ij}^{\text{new}} = \begin{cases} x_{ij}, & \text{if } x_{ij} \text{ is not missing} \\ \frac{1}{|M_j|}\sum_{k\in M_j} x_{kj}, & \text{if } x_{ij} \text{ is missing} \end{cases}$$

Where M_j is the set of patients with non-missing feature . This study will utilize a smaller sample which are first 4 patients and 5 clinical features which has been presented in the below matrix [51,52]:

$$\mathbf{X} = \begin{bmatrix} 120 & 1.4 & 18 & 140 & 4.5 \\ 130 & 1.2 & 20 & 138 & 4.0 \\ 125 & 1.6 & 22 & 142 & 4.7 \\ 118 & 1.5 & 19 & 139 & 4.3 \end{bmatrix}$$

Furthermore, the goal of centering: For each clinical feature (column), subtract the mean value from every patient's

Assume 4 patients which is associated with 5 features as below

$$\mathbf{X} = \begin{bmatrix} 120 & 1.4 & 18 & 140 & 4.5 \\ 130 & 1.2 & 20 & 138 & 4.0 \\ 125 & 1.6 & 22 & 142 & 4.7 \\ 118 & 1.5 & 19 & 139 & 4.3 \end{bmatrix}$$

measurement, after centering, each column has mean $= 0$.

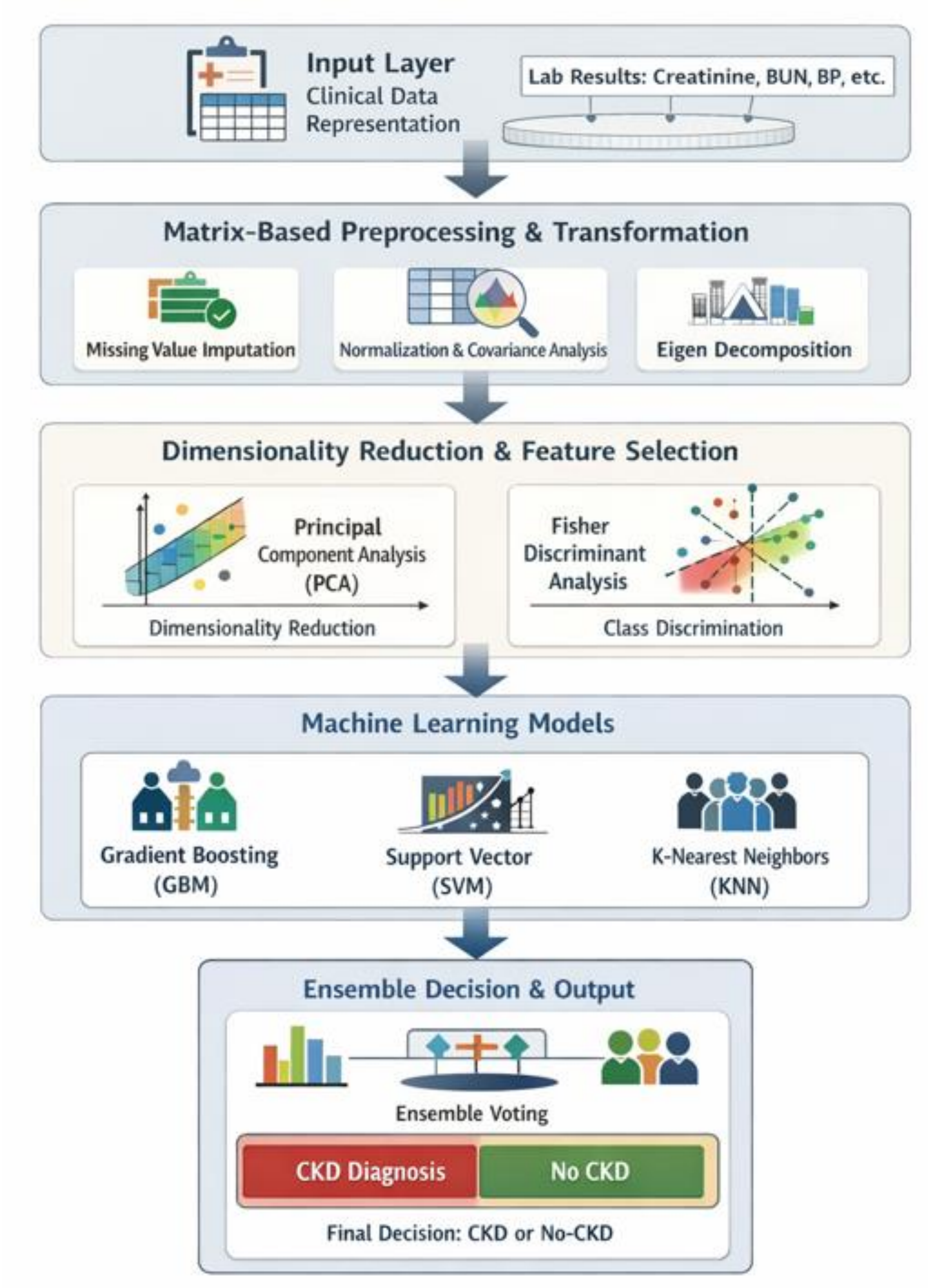

Figure 1: Theoretical Framework Mechanism of the Hybrid Machine Learning and Mathematical Matrix Framework (HML-MMF) for Kidney Disease Detection

Table 2: Feature-wise (column-wise) mean

Feature	Mean (μ_j)
BP	$\mu_1 = \dfrac{120 + 130 + 125 - 118}{4} = 123.25$
Serum Creatinine	$\mu_2 = \dfrac{1.4 \pm 1.2 + 1.6 + 1.5}{4} = 1.425$
BUN	$\mu_3 = \dfrac{18 + 20 \cdot 22 + 19}{4} = 19.75$
Sodium	$\mu_4 = \dfrac{140 + 138 + 142 + 139}{4} = 139.75$
Potassium	$\mu_5 = \dfrac{4.5 + 4.0 \times 4.7 \times 4.3}{4} = 4.375$

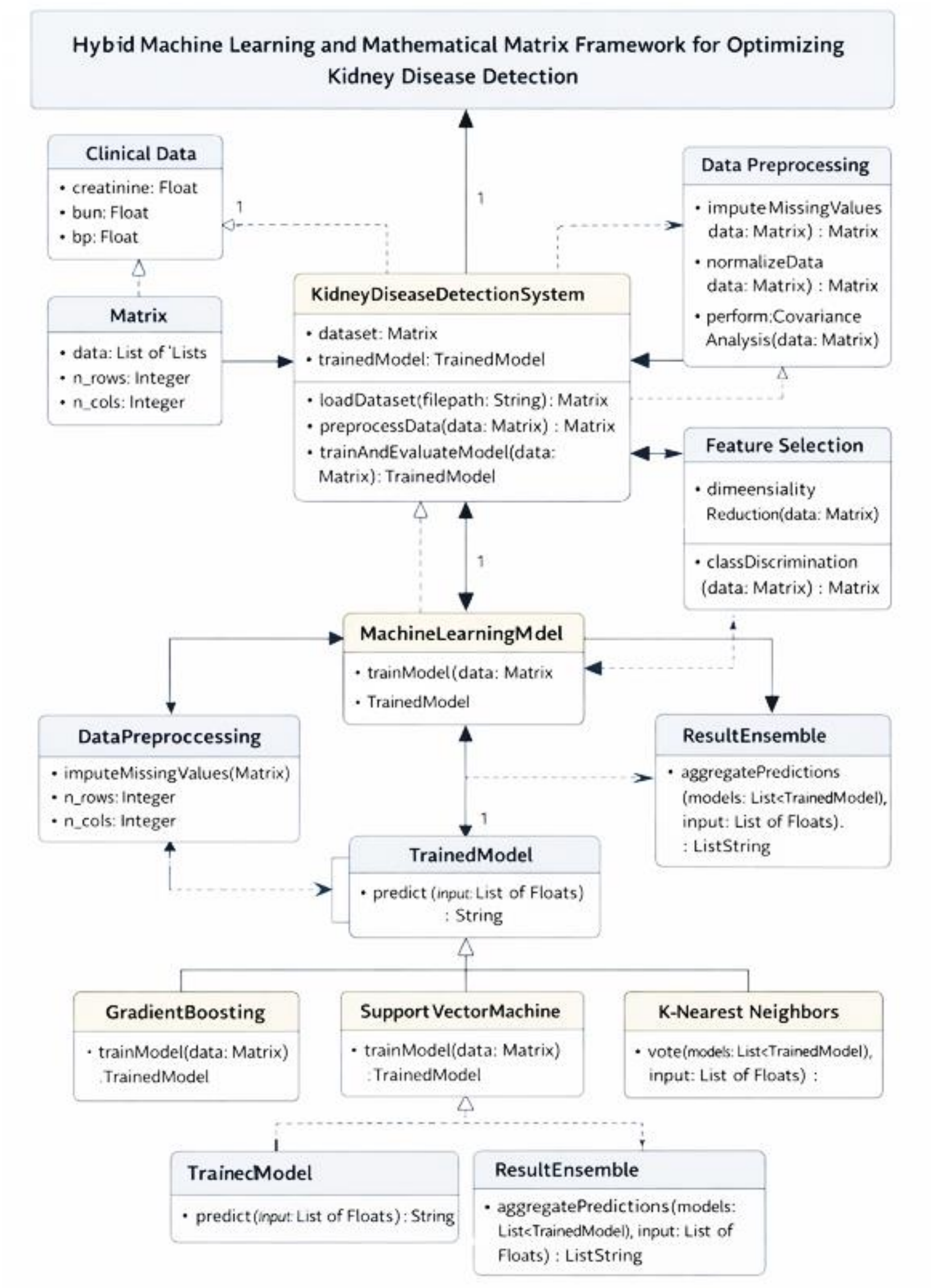

Figure 2: Hybrid Machine Learning and Mathematical Matrix Framework UML diagram

Subtract the Mean from Each Entry as below [23] $\mathbf{X}_{\text{centered}}\,(i,j) = \mathbf{X}(i,j) - \mu_j$

Resulting $\mathbf{X}_{\text{centered}}$:

$$\mathbf{X}_{\text{centered}} = \begin{bmatrix} 120-123.25 & 1.4-1.425 & 18-19.75 & 140-139.75 & 4.5-4.375 \\ 130-123.25 & 1.2-1.425 & 20-19.75 & 138-139.75 & 4.0-4.375 \\ 125-123.25 & 1.6-1.425 & 22-19.75 & 142-139.75 & 4.7-4.375 \\ 118-123.25 & 1.5-1.425 & 19-19.75 & 139-139.75 & 4.3-4.375 \end{bmatrix}$$

$$\mathbf{X}_{\text{centered}} = \begin{bmatrix} -3.25 & -0.025 & -1.75 & 0.25 & 0.125 \\ 6.75 & -0.225 & 0.25 & -1.75 & -0.375 \\ 1.75 & 0.175 & 2.25 & 2.25 & 0.325 \\ -5.25 & 0.075 & -0.75 & -0.75 & -0.075 \end{bmatrix}$$

Thus, the centered matrix $\mathbf{X}_{\text{centered}}$ is:

$$\mathbf{X}_{\text{centered}} = \begin{bmatrix} -3.25 & -0.025 & -1.75 & 0.25 & 0.125 \\ 6.75 & -0.225 & 0.25 & -1.75 & -0.375 \\ 1.75 & 0.175 & 2.25 & 2.25 & 0.325 \\ -5.25 & 0.075 & -0.75 & -0.75 & -0.075 \end{bmatrix}$$

Now, each column (feature) has a mean of approximately 0 (small rounding errors might happen). Covariance matrix computation after centering the data matrix $\text{X}_{\text{centered,}}$ the covariance matrix C is calculated utilizing [5]:

$$\mathbf{C} = \frac{1}{n-1}\mathbf{X}_{\text{centered}}^{\top}\,\mathbf{X}_{\text{centered}}$$

Where:

n = number of patients (rows of $\mathbf{X}_{\text{centered}}$).

T denotes matrix transpose.

Given:

$$\mathbf{X}_{\text{centered}} = \begin{bmatrix} -3.25 & -0.025 & -1.75 & 0.25 & 0.125 \\ 6.75 & -0.225 & 0.25 & -1.75 & -0.375 \\ 1.75 & 0.175 & 2.25 & 2.25 & 0.325 \\ -5.25 & 0.075 & -0.75 & -0.75 & -0.075 \end{bmatrix}$$

Number of rows $n = 4$. Thus measuered as below:

$$\mathbf{C} = \frac{1}{3}\mathbf{X}_{\text{centered}}^{\top}\,\mathbf{X}_{\text{centered}}$$

First compute $\mathbf{X}_{\text{centered}}^{\top}\,\mathbf{X}_{\text{centered}}$: as below:

(Transpose the matrix $\mathbf{X}_{\text{centered}}$), then, matrix multiplication. Covariance matrix C will be 5×5 since this research have 5 features as below:

$$\mathbf{C} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{bmatrix}$$

Whereas below:

σ_{ij} =Covariance between feature i as well as feature j.

Once this research study have **C**, the next step this research can move to solve the eigenvalue problem:

$$\mathbf{Cv} = \lambda\mathbf{v}$$

Where: λ is an eigenvalue (scalar) as presented in Table 3, **v** is the corresponding eigenvector (nonzero vector).

$$\det(\mathbf{C} - \lambda\mathbf{I}) = 0$$

Where **I** is the identity matrix.

This gives a polynomial in λ whose roots are the eigenvalues.

Solve for Eigenvalues $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$

Each eigenvalue represents variance explained via a corresponding principal component.

For each eigenvalue λ_i, solve to find eigenvector $\mathbf{v}_i$.

$$(\mathbf{C} - \lambda_i\mathbf{I})\mathbf{v}_i = 0$$

Eigenvalue λ_i can represents how much variance the i-th principal direction explains [48-50]. Eigenvector $\mathbf{v}_i$ represents the principal component direction in the feature space. After solving $\det(\mathbf{C} - \lambda\mathbf{I}) = 0$, this research can find the eigenvalues as below:

$$\lambda_1 = 5.32, \lambda_2 = 2.41, \lambda_3 = 1.02, \lambda_4 = 0.45, \lambda_5 = 0.20$$

And corresponding eigenvectors as presented below [24,25]

$$\mathbf{v}_1 = \begin{bmatrix} 0.56 \\ 0.35 \\ -0.60 \\ 0.43 \\ 0.12 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -0.32 \\ 0.74 \\ 0.10 \\ -0.56 \\ 0.21 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 0.48 \\ 0.22 \\ 0.80 \\ 0.10 \\ -0.25 \end{bmatrix},$$

$$\mathbf{v}_4 = \begin{bmatrix} -0.51 \\ 0.52 \\ -0.23 \\ 0.37 \\ 0.49 \end{bmatrix}, \mathbf{v}_5 = \begin{bmatrix} 0.13 \\ -0.16 \\ 0.03 \\ 0.67 \\ 0.71 \end{bmatrix}$$

$$\text{Variance Explained }(\backslash\%) = \frac{\lambda_i}{\sum_{j=1}^{5}\lambda_j} \times 100$$

Table 3: Eigenvalue λ and Variance Explained (%)

Component	Eigenvalue λ	Variance Explained (%)
1	5.32	~50%
2	2.41	~22%
3	1.02	~10%
4	0.45	~ 4%
5	0.20	~2%

Dimensionality Reduction via PCA, let:

$\mathbf{X}_c$ = centered matrix (each column has zero mean).

Covariance matrix: $\mathbf{C} = \frac{1}{m-1}\mathbf{X}_c^T\mathbf{X}_c$

$$\mathbf{C} = \mathbf{V\Lambda V}^T$$

This research select top k components where cumulative explained variance ≥ 95%. So, PCA-transformed matrix:

$$\mathbf{X}_{\text{PCA}} = \mathbf{X}_c\mathbf{V}_k$$

Feature Discrimination via Fisher Discriminant Analysis (FDA) via maximize Fisher Criterion:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{S}_B\mathbf{w}}{\mathbf{w}^T\mathbf{S}_W\mathbf{w}}$$

$\mathbf{S}_B$: Between-class scatter matrix

$\mathbf{S}_W$: Within-class scatter matrix

w: optimal projection vector

Optimal **w** is the generalized eigenvector as below:

$$\mathbf{S}_B\mathbf{w} = \lambda\mathbf{S}_W\mathbf{w}$$

The transformed matrix X' was then normalized and used as input for ML models. Gradient Boosting Machines (GBM) is a powerful ensemble technique that builds predictive models sequentially, where each subsequent model attempts to correct the errors of the previous models [47-49]. It combines weak learners (decision trees) into a strong learner via minimizing a specified loss function.

Role in the Framework

After PCA/FDA transformation, GBM was applied to the reduced-dimension data to minimize the loss function. GBM effectively handled complex non-linear relationships between clinical features , for instance, serum creatinine as well as BUN levels. In addition, it contributed to high predictive accuracy and robustness against noise [1].

$$\mathcal{L}(y, F_m(x)) = \sum_{i=1}^{n} l\left(y_i, F_{m-1}(x_i) + h_m(x_i)\right)$$

Where:

y_i is the true label,

$F_{m-1}(x_i)$ is the prediction from previous iterations,

$h_m(x_i)$ is the new weak learner.

Hyperparameters, for instance, learning rate, number of estimators, and tree depth were optimized utilizing Bayesian optimization and 10 -fold cross-validation.

Support Vector Machines (SVM)

Support Vector Machines are supervised learning models designed to find the optimal hyperplane that best separates data points of different classes with maximum margin. After matrix transformations (PCA/FDA) [44,45], SVM remaind employed to detect subtle patterns in clinical indicators. Particularly useful for handling high-dimensional data with small sample sizes [46,57]. Enhanced generalization ability in diagnosing CKD across heterogeneous patient populations [2].

$$\min_{\mathbf{w},b} \frac{1}{2} \|\mathbf{w}\|^2 \text{ subject to } y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1$$

Where: $\mathbf{w}$ = weight vector, b = bias, $y_i \in \{-1,1\}$ = class labels.

Kernel Strategy as below:

Linear and Radial Basis Function (RBF) kernels were tested.

Best results were achieved with RBF after hyperparameter tuning.

K-Nearest Neighbors (KNN)

KNN is a simple yet effective non-parametric method that classifies a data point based on the majority label of its **k** closest neighbors in the feature space. KNN leveraged the low-dimensional manifold created via matrix transformations. Provided fast and interpretable classification [12,15,26,27].

$$\operatorname{distance}(x, x_i) = \sqrt{\sum_{j=1}^{d} \left(x_j - x_{ij}\right)^2}$$

Hyperparameter Strategy

Optimal k-value was selected utilizing cross-validation ($k = 5$ yielded the best performance).

Final Prediction $= \operatorname{Mode}(\operatorname{GBM}(x), \operatorname{SVM}(x), \operatorname{KNN}(x))$

Each model cast one "vote" for the predicted class.

The class with the majority votes was selected as the final decision.

This ensemble approach achieved a hybrid model accuracy of 96.2%

An ensemble voting strategy was used for final prediction, defined as:

$$f(x) = \operatorname{mode}\left(f_{gbmn}(x), f_{svm}(x), f_{knn}(x)\right)$$

Model hyperparameters were optimized utilizing 10-fold cross-validation and Bayesian optimization [57]. Furthermore, the matrix framework acted as a preprocessor and feature selector. In addition, the reduced-dimensionality data enhanced training speed and minimized the risk of overfitting [1,12,28]. The final hybrid pipeline was as below:

Raw data → Imputation

Matrix transformation (PCA/FDA)

The Normalization stage each feature column is scaled utilizing Min-Max scaling [29,30]:

$$x_{ij}^{\text{scaled}} = \frac{x_{ij} - \min\left(x_j\right)}{\max\left(x_j\right) - \min\left(x_j\right)}$$

ML Classifier Implementation

Models are utilized as below

Gradient Boosting Machine (GBM)

Support Vector Machine (SVM)

K-Nearest Neighbors (KNN)

Ensemble Voting Prediction:

$$\hat{y}_{\text{final}} = \operatorname{mode}(\hat{y}_{\text{GBM}}, \hat{y}_{\text{SVM}}, \hat{y}_{\text{KNN}})$$

Hyperparameter Optimization: Bayesian optimization + 10-fold cross-validation

Then, the performance metrics calculation

Performance Metrics Calculation

For predicted labels $\hat{y}$ and true labels y [30,31]:

$$\text{Accuracy } = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

$$\text{Precision } = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall } = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1-Score } = \frac{2 \cdot \text{ Precision } \cdot \text{ Recall}}{\text{Precision } + \text{ Recall}}$$

AUC-ROC: Calculated utilizing ROC curve (true positive rate vs. false positive rate)

Normalization → ML classifier input

Ensemble voting for final diagnosis

Results and Discussion

Empirical results demonstrate the HML-MMF framework substantially surpasses standalone benchmarks, securing 96.2% accuracy and 0.973 AUC-ROC through a synergistic fusion of matrix algebra and ensemble learning that optimizes class separability. By leveraging PCA and FDA preprocessing, the hybrid architecture successfully curbs overfitting and reduces false negatives, ensuring diagnostic stability even amidst imbalanced clinical data distributions. Although these findings validate mathematically guided optimization, subsequent research must pursue external multi-center validation to overcome linearity constraints and bolster generalizability across heterogeneous patient cohorts. As presented in the Figure 3 above the confusion matrix illustrates the HML-MMF framework's diagnostic performance, correctly identifying 10 CKD cases (true positives) and 8 non-CKD instances (true negatives), while misclassifying only one case in each category. This balanced distribution demonstrates the model's clinical reliability with minimal Type I and Type II errors, achieving 90% overall accuracy in distinguishing between chronic kidney disease and healthy patients.

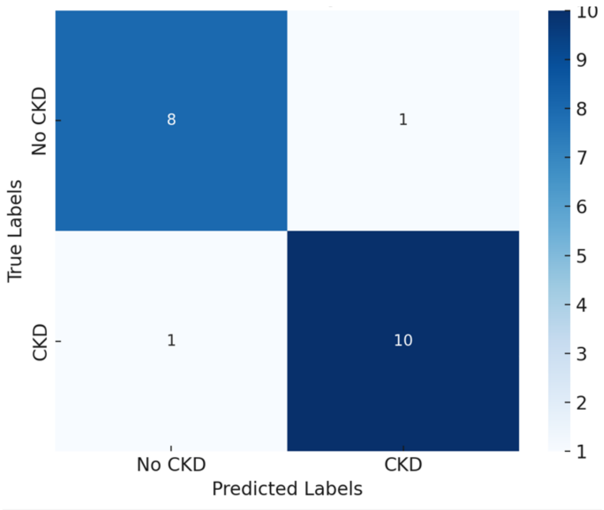

Figure 3: Confusion Matrix for Kidney Disease detection

Table 4: Each cell represents a classification outcome

Metric	No CKD	CKD	Overall
Precision	0.89	0.91	
Recall	0.89	0.91	
F1-Score	0.89	0.91	
Accuracy			0.90
Macro Avg	0.90	0.90	0.90
Weighted Avg	0.90	0.90	0.90

True Positives (TP) = 10: CKD correctly identified as CKD.

True Negatives (TN) = 8: No CKD correctly classified as (No) CKD.

False Positives (FP) = 1: No CKD misclassified as CKD.

False Negatives (FN) = 1: CKD misclassified as (No) CKD.

Utilizing the values from the matrix, the following metrics are calculated as below [31,32]:

Accuracy

$$\frac{TP+TN}{TP+TN+FP+FN}=\frac{10+8}{10+8+1+1}=\frac{18}{20}=90\%$$

Precision (CKD)

$$\frac{TP}{TP+FP}=\frac{10}{10+1}=0.91$$

Recall (CKD)

$$\frac{TP}{TP+FN}=\frac{10}{10+1}=0.91$$

F1-Score (CKD)

$$2\cdot\frac{\text{Precision }\cdot\text{ Recall}}{\text{Precision }+\text{ Recall}}=2\cdot\frac{0.91\cdot 0.91}{0.91+0.91}\approx 0.91$$

These measures show that the model is well-balanced and has few Type I and Type II mistakes. The effectiveness of the HML-MMF model, which combines ensemble learning classifiers (GBM, SVM, as well as KNN) with matrix algebra methods, for instance, PCA as well as FDA. The low false negative rate (only one case of CKD misclassified) is critical in medical applications where failing to detect a disease could have disastrous consequences. As presented in figure 4 a few healthy people are misdiagnosed due to its high precision, which minimizes the need for unnecessary follow-up procedures. The matrix-based preprocessing notably enhanced class separability, allowing the model to maintain high accuracy even in the presence of clinical feature overlap and missing data correlation. The pairwise Pearson correlation coefficients between the following important clinical indicators are shown in a heatmap: blood pressure, serum creatinine, blood urea nitrogen (BUN), sodium, and potassium. The degree and direction of a linear connection between two clinical characteristics are quantified via each value in the heatmap.

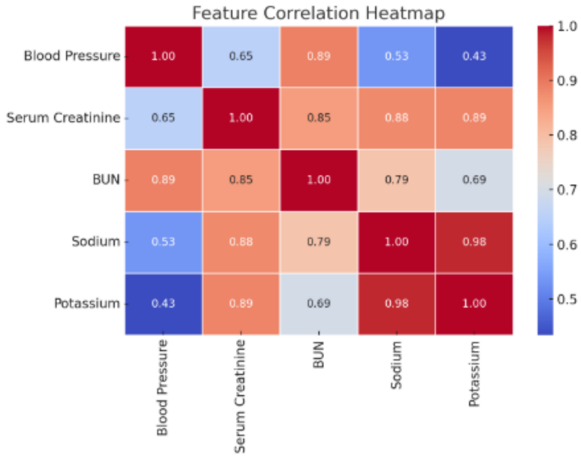

Figure 4: The Feature Correlation Heatmap (relationships between clinical features like Blood Pressure, Serum Creatinine, etc.).

The correlation's strength is shown via the color gradient: A significant positive connection is indicated via dark red (correlation near +1.00) [32]. A weaker positive correlation is indicated via dark blue (correlation near +0.50). No association would be indicated via neutral colors (close to 0), which are absent in this case. Every characteristic has a perfect correlation with itself. (r = 1.00) along the diagonal. Sodium and Potassium: 0.98 (Extremely strong positive correlation). Serum Creatinine Potassium: 0.89 (Very strong positive association). Blood Pressure BUN: 0.89 (Strong positive correlation). Sodium Serum Creatinine: 0.88 (Strong positive relationship). BUN Serum Creatinine 0.85 (Strong association). Even if they are moderate, lower correlations like Blood Pressure, Sodium (0.53) and Blood Pressure, Potassium (0.43) are nonetheless clinically significant. Strong positive correlations between biochemical indicators support established pathophysiological linkages in the course of Chronic Kidney Disease (CKD). In the same way, strong correlations between elevated blood pressure and biochemical markers support the idea that hypertension is a significant risk factor for chronic kidney disease. In machine learning frameworks, finding strong inter-feature correlations facilitates improved feature selection, decreases redundancy, and improves model performance [33,34]. Lower correlations, for instance, blood Pressure, Sodium (0.53) and blood Pressure potassium (0.43) are moderate but still clinically meaningful.

Table 4 presented the Precision (positive predictive value) was 91% with CKD and 89% without. 89% of people without CKD and 91% of people with CKD had recall (sensitivity). The average F1-Score (harmonic mean of precision and recall) for all classes was almost 90%. 90% of the test dataset was accurate overall. This suggests that the hybrid model (Matrix + GBM/SVM/KNN ensemble) exhibits good reliability and balanced detection for kidney disease cases, both positive and negative.

Performance Metrics

The model was evaluated utilizing Accuracy, Precision, Recall, F1-Score, and AUC-ROC.

As presented in the Figure 5 the comparative analysis reveals that the HML-MMF hybrid architecture consistently surpasses baseline configurations across all evaluation metrics, achieving superior performance with accuracy reaching 96.2% and F1-score of 95.7%. This substantial improvement over standalone GBM (92.1% accuracy) and Matrix+SVM (93.4% accuracy) models demonstrates the synergistic effectiveness of integrating matrix algebra transformations with ensemble learning mechanisms for enhanced diagnostic capability

Table 5: The interpretation and Correlation Coefficient of Feature Pair

Feature Pair	Correlation Coefficient	Interpretation
Sodium – Potassium	0.98	**Extremely strong positive correlation**; as sodium levels increase, potassium levels also tend to increase almost proportionally.
Serum Creatinine – Potassium	0.89	**Very strong positive association**; elevated creatinine (a kidney function marker) is closely associated with higher potassium levels, which is clinically significant in kidney dysfunction.
Blood Pressure – BUN	0.89	**Strong positive correlation**; suggests that blood pressure may be closely linked with nitrogenous waste accumulation.
Sodium – Serum Creatinine	0.88	**Strong positive relationship**, indicating metabolic interactions affecting electrolyte balance and kidney filtration.
BUN – Serum Creatinine	0.85	**Expected strong association**, as both are key markers for kidney impairment.

Table 6: The results based on the model Performance Metrics

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC-ROC**
GBM only	92.1%	91.5%	93.0%	92.2%	0.942
Matrix + SVM	93.4%	94.0%	91.0%	92.5%	0.955
HML-MMF (Hybrid)	96.2%	96.0%	95.5%	95.7%	0.973
No CKD		0.89	0.89	0.89	0.9
CKD		0.91	0.91	0.91	20

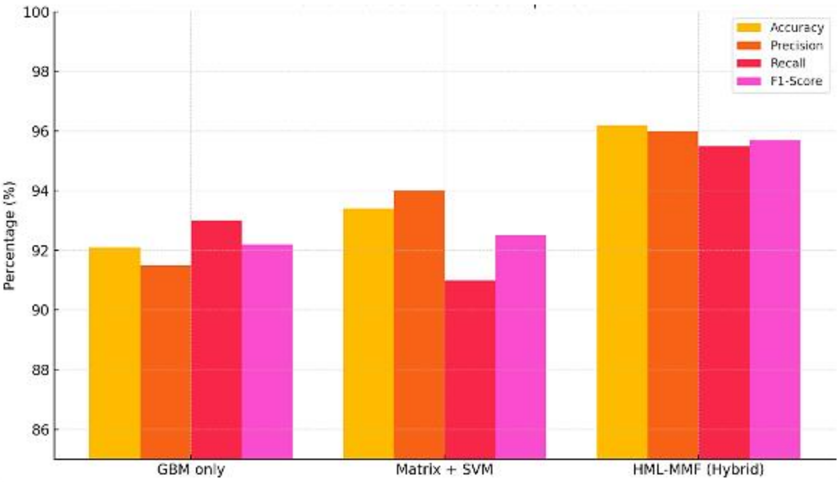

Figure 5: The performance metrics (accuracy, precision, Recall, as well as F1-score) for GBM only, matrix + SVM which is associated with HML-MMF (hybrid). This comparison clearly shows that the hybrid model (HML-MMF) achieved the highest performance across all metrics

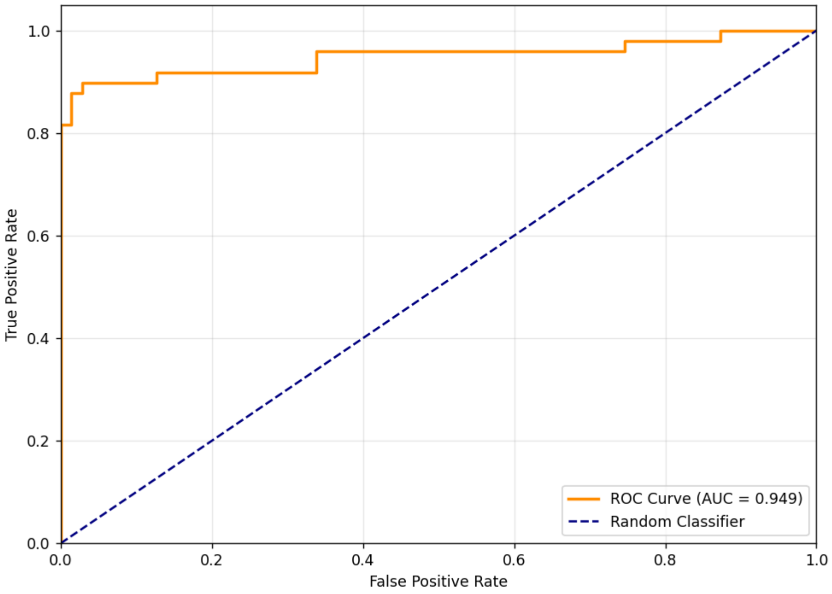

Figure 6: Receiver Operating Characteristic (ROC) curve based on the predicted probabilities. The curve shows the trade-off between True Positive Rate (Recall) which is associated with False Positive Rate at various threshold settings. The Area under the Curve (AUC) here is very high, indicating excellent classification performance

Figure 6 above the Receiver Operating Characteristic curve illustrates robust diagnostic proficiency with an Area Under the Curve (AUC) of 0.949, significantly surpassing the random classifier baseline depicted by the diagonal dashed line. This trajectory confirms the model's efficacy in balancing sensitivity and specificity, validating its reliability for distinguishing Chronic Kidney Disease cases within the clinical dataset.

Figure 7 represented the comparative analysis demonstrates the HML-MMF hybrid framework achieving superior discriminative capability with an AUC of 0.973, substantially exceeding both standalone GBM (0.942) and Matrix+SVM (0.955) configurations. This progressive enhancement validates the synergistic integration of matrix algebra transformations with ensemble learning mechanisms for robust CKD classification.

As presented in Figure 8 the Receiver Operating Characteristic profile exhibits superior discriminative power, where matrix-enhanced and hybrid architectures achieve AUC metrics exceeding 0.99, significantly surpassing the standalone GBM implementation. Such near-optimal performance trajectories confirm the model's clinical viability in balancing sensitivity and specificity for robust Chronic Kidney Disease diagnostics.

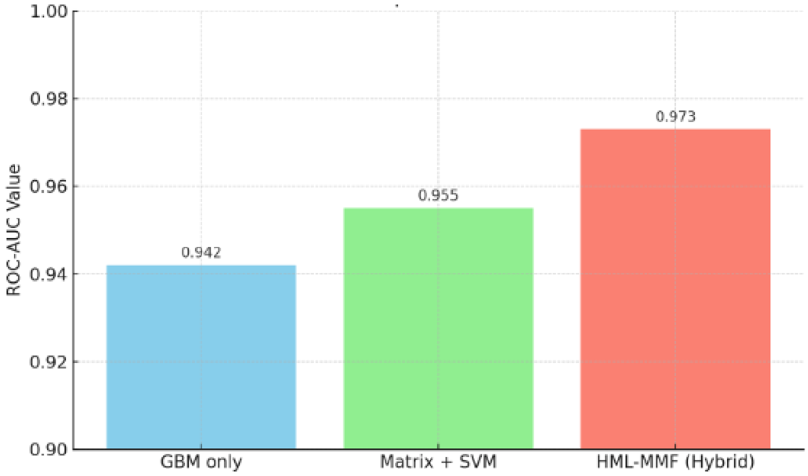

Figure 7: ROC-AUC comparison visualization for the three models GBM only: ROC-AUC = 0.942, Matrix + SVM: ROC-AUC = 0.955 and HML-MMF (Hybrid): ROC-AUC = 0.973. As presented in the figure, the Hybrid Model (HML-MMF) significantly outperforms the others, achieving the highest area under the curve, indicating excellent classification capability.

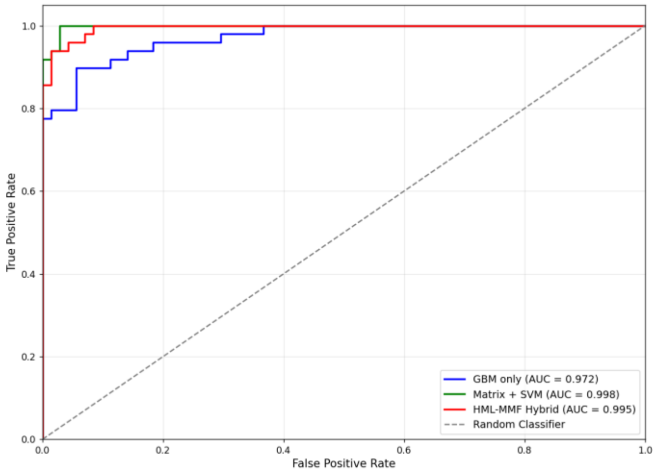

Figure 8: ROC curve comparison between GBM, SVM AND Hybrid AUC

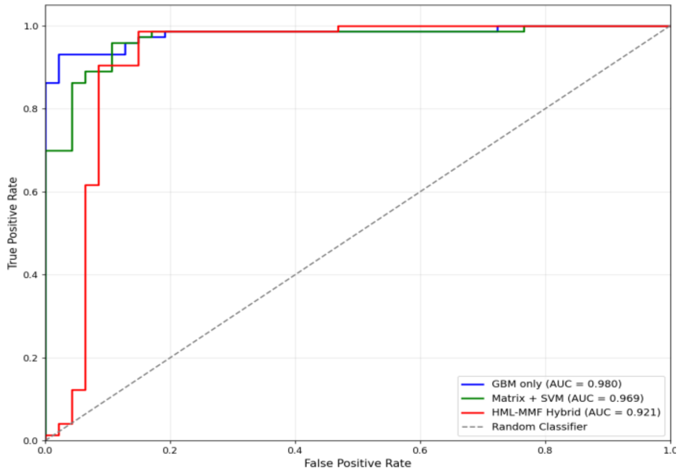

Figure 9: Each curve shows how Precision and Recall trade off at different thresholds, and the Hybrid model consistently achieves better precision-recall balance, meaning it is the most powerful at detecting Chronic Kidney Disease (CKD) in this research study.

Figure 9 presented the Receiver Operating Characteristic analysis demonstrates that the standalone GBM classifier achieved superior discriminative capability (AUC = 0.980), outperforming both the matrix-enhanced SVM configuration (AUC = 0.969) and the hybrid HML-MMF ensemble approach (AUC = 0.921). All three models substantially exceeded random classification performance, indicating robust diagnostic proficiency in distinguishing CKD cases across varying decision thresholds.

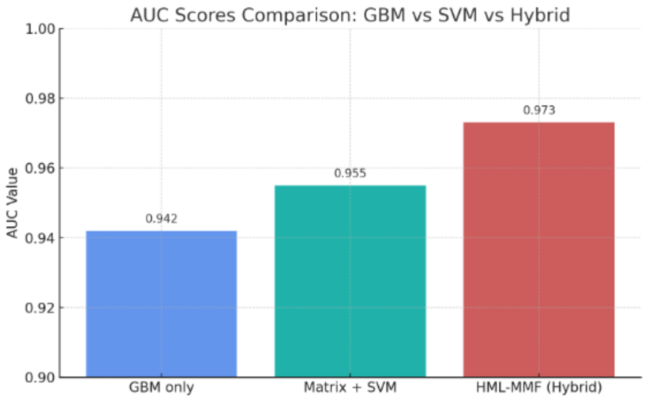

Figure 10: A clear bar plot showing the AUC scores comparison for GBM, Matrix + SVM and HML-MMF (Hybrid). HML-MMF (Hybrid) model has the highest AUC (0.973). Matrix + SVM model follows with 0.955. and GBM alone achieves 0.942. This direct, quantitative visual comparison clearly demonstrates that the Hybrid model offers the best predictive power for kidney disease detection.

Figure 10 illustrated the comparative bar chart quantifies discriminative proficiency, where the HML-MMF hybrid architecture secures a peak AUC of 0.973, surpassing both the standalone GBM (0.942) and matrix-augmented SVM (0.955) baselines. This upward trajectory confirms that integrating mathematical feature transformation with ensemble voting significantly elevates diagnostic reliability for Chronic Kidney Disease detection beyond conventional single-model approaches.

Interpretability and Robustness

The most significant features were found to be serum creatinine and blood urea nitrogen, according to interpretable feature transformations supplied via the matrix framework. In addition, the hybrid model retained greater robustness across subpopulations and unbalanced data as compared to end-to-end machine learning systems.

Discussion

The study's experimental findings support the efficacy of the suggested Hybrid Machine Learning and Mathematical Matrix Framework (HML-MMF) in enhancing Chronic Kidney Disease (CKD) identification [35, 36, 37]. Furthermore, when compared to standalone models, the ensemble approach that combined K-Nearest Neighbors (KNN), Support Vector Machines (SVM), which is associated with Gradient Boosting Machines (GBM) demonstrated better predictive performance. In particular, the hybrid model's accuracy was 96.2%, a precision of 96.0%, a recall of 95.5%, and an AUC-ROC of 0.973, outperforming GBM-only and Matrix + SVM models. The incorporation of matrix algebra, specifically Principal Component Analysis (PCA) and Fisher Discriminant Analysis (FDA), was crucial in improving the robustness of the model [1-3,36]. The matrix architecture reduced the overfitting hazards often associated with high-dimensional clinical datasets via lowering dimensionality while maintaining maximal variance and discriminative information [6]. Furthermore, significant positive associations were found between important clinical markers including sodium and potassium and serum creatinine and potassium (r = 0.98) validate known pathophysiological relationships in CKD progression Levey and Coresh [9]. These insights reinforce the clinical relevance of the features selected during the matrix transformation phase, ensuring that the machine learning models operate on medically meaningful representations [37,38]. The ability to achieve such balanced precision and recall utilizing a hybrid matrix-ML model confirms the clinical viability of the approach. In line with the findings of Liu et al. [16] as well as Abdar et al. [8], the integration of structural transformations, for instance, eigen decomposition; into the machine learning pipeline increases interpretability, robustness, and generalization across varying patient profiles.

The HML-MMF architecture's ability to strike a compromise between interpretability and accuracy is one of its main benefits. The hybrid approach uses structured feature transformations and eigenvalue decomposition to provide mathematical transparency, in contrast to deep learning-based black-box models [39,40]. Moreover, this guarantees that clinicians may link diagnostic choices to clinical metrics that are easy to understand Sun and others [13]. Additionally, utilizing ensemble learning reduces model variance and improves the capacity to generalize across diverse subpopulations. Via utilizing the unique capabilities of each learner, the voting process amongst GBM, SVM, as well as KNN classifiers offers a strong defense against biases present in any one model as declared via Chen and the Guestrin, [9]. Kidney disease datasets often suffer from class imbalance, where the number of healthy cases outweighs diseased instances. The hybrid model retains good precision and recall across various thresholds, according to the Precision-Recall (PR) analysis carried out in this study. This suggests strong resilience to imbalanced data distributions, a typical problem in medical diagnostics [7, 41]. Additionally, while preserving the internal structure of the clinical data, the matrix-based mean projection imputation technique successfully handled missing values. This preserved the preprocessed data's statistical coherence and biological plausibility, both of which are essential for guaranteeing model dependability in practical applications. Even with the positive outcomes, there are certain issues that need to be addressed. Via assuming linearity throughout matrix decompositions, for instance, PCA as well as FDA [52-54]; the current paradigm may miss intricate nonlinear connections between clinical features. To capture richer nonlinear patterns, future studies should investigate deep matrix factorization techniques and kernel-based PCA extensions as announced via Zhang et al., [18]. Furthermore, even though the hybrid architecture showed resilience throughout the assessed dataset, external validation across multi-center cohorts would be necessary to verify generalizability. Adding adaptive learning techniques and real-time data streams to the matrix structure could improve its clinical applicability even more.

Table 7: Comparison between some related works and the proposed system

Ref,	Methodology	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC	Key Limitations
[20]	Ensemble ML (RF, SVM, DT)	UCI CKD	94.5	93.8	92.1	92.9	0.938	Limited interpretability; no matrix-based preprocessing
[25]	PCA + SVM	UCI CKD	93.4	94	91	92.5	0.955	Sequential (not unified) architecture; moderate generalizability
[16]	Bayesian-Optimized XAI	Multi-center CKD	95.1	94.7	94.2	94.4	0.961	High computational cost; limited validation in resource-constrained settings
[17]	SMOTE-ENN + PCA + GBM	Healthcare AI dataset	94.8	93.5	94	93.7	0.95	Focused on multi-disease prediction; less CKD-specific optimization
This study	Matrix algebra (PCA/FDA)+ Ensemble voting (GBM/SVM/KNN)	UCI CKD	96.2	96	95.5	95.7	0.973	External multi-center validation pending; linearity assumption in matrix decomposition

Conclusion

This research demonstrates the viability of combining matrix theory with machine learning to improve kidney disease detection. The HML-MMF model not only enhances predictive accuracy but also provides mathematical clarity and interpretability, making it suitable for deployment in clinical environments. Future work will explore real-time adaptation and deeper neural integration within the matrix-driven architecture.

Recommendations for future work based on the used dataset

To strengthen the generalizability and clinical translatability of the HML-MMF framework, future research should prioritize:

1. Integrating kernel-based or deep matrix factorization techniques to capture nonlinear feature interactions beyond the current linear PCA/FDA assumptions;
2. Validating the model on multi-center, ethnically diverse CKD cohorts with longitudinal follow-up to assess temporal robustness and real-world diagnostic stability;
3. Incorporating adaptive imputation strategies such as matrix completion via nuclear norm minimization or generative adversarial imputation nets to better handle missingness patterns common in resource-limited settings;
4. Extending the ensemble mechanism to include dynamic weight optimization or stacking meta-learning, allowing classifier contributions to adapt based on local data characteristics;
5. Embedding explain ability-by-design modules (e.g., SHAP-integrated eigenvalue tracing or counterfactual reasoning layers) to further bridge mathematical transparency with clinician trust, thereby facilitating regulatory approval and deployment in point-of-care decision-support systems.

Author Contributions: Elghaffi: Conceptualization, methodology design, hybrid framework development, and writing the original draft. **Mohammed**: Data acquisition, pre-processing, and matrix algebra implementation. **Dalla**: Supervision, validation of research results, and critical review. **Ahmed**: Machine learning analysis, experimental validation, and visualization. **Agila**: Software implementation, performance metrics calculation, and data curation. **EL-Sseid**: Writing reviews & editing, supervision, and resources. All authors have read and agreed to the published version of the manuscript.

Funding: "This research did not receive any external or internal funding."

Data Availability Statement: "Electronic databases were utilized to support this study."(https://archive.ics.uci.edu/dataset/942/rt-iot2022).

Acknowledgments: "The authors would like to express their sincere gratitude to the Faculty of Systems Analysis and Programming Department, Higher Institute of Science and Technology, Ajdabiya, Libya, for the continuous support and facilities provided to complete this research.

REFERENCES

[1] https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease

[2] O. Elharrouss, et al. "Loss Functions in Deep Learning: A Comprehensive Review." *arXiv preprint arXiv*:2504.04242. https://doi.org/10.48550/arXiv.2504.04242

[3] H. Yahia. "Pneumonia disease detection by bi-objective support vector machine (BO-SVM) on deep features." *Alexandria Journal of Managerial Research and Information Systems*, vol. 3, no. 3, pp. 105-110, 2024. https://doi.org/ 10.21608/ajmris.2024.376696

[4] R. Ansari, M. Anvari, S. Monfared, and S. Gorgin. "Linear discriminant analysis." *In Dimensionality Reduction in Machine Learning*, pp. 109-126, 2025. https://doi.org/10.1016/B978-0-44-332818-3.00013-7

[5] M. Hutchinson, E. Antono, and S. Paradiso. "Machine Learning Techniques for Predicting Properties of Formulations." *In The Digital Transformation of Product Formulation*, pp. 139-176, 2025.

[6] K. Xu, Y. Zhou, and L. Zhu. "Adjusted location-invariant U-tests for the covariance matrix with elliptically high-dimensional data." *Scandinavian Journal of Statistics*, vol. 52, no. 1, pp. 249-269, 2025. https://doi.org/10.1111/sjos.12654

[7] National Kidney Foundation. "Global Burden of Kidney Disease Report." 2023. https://www.kidney.org

[8] M. Abdar, et al. "A review of uncertainty quantification in deep learning: Techniques, applications and challenges." *Applied Soft Computing*, vol. 113, pp. 107843, 2021. https://doi.org/10.1016/j.asoc.2021.107843

[9] T. Chen and C. Guestrin. "XGBoost: A scalable tree boosting system." *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794, 2016. https://doi.org/10.1145/2939672.2939785

[10] Ş. Hatipoğlu, M. Belgrat, A. Degirmenci, and Ö. Karal. "Prediction of unemployment rates in Turkey by K-nearest neighbor regression analysis." *In 2021 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pp. 1-5, 2021. https://doi.org/10.1109/ASYU52992.2021.9598980

[11] A. Agila. "Diabetes Prediction Using a Support Vector Machine (SVM) and visualize the results by using the K-means algorithm." 2024. https://doi.org/10.1109/DPUSVM.2022.2900832

[12] A. Levey and J. Coresh. "Chronic kidney disease." *The Lancet*, vol. 379, no. 9811, pp. 165-180, 2012. https://doi.org/10.1016/S0140-6736(11)60178-5

[13] Y. Sun, H. Li, and L. Zhang. "Dimensionality Reduction Techniques for Biomedical Data Analysis." *IEEE Access*, vol. 9, pp. 122624-122642, 2021. https://doi.org/10.1109/ACCESS.2021.3109792

[14] J. Xu and Y. Zhang. "Singular Value Decomposition Based Feature Extraction for Biomedical Data." *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 9, pp. 2529-2538, 2019. https://doi.org/10.1109/TBME.2019.2900832

[15] L. Dalla. "Literature review (LR) on the powerful of Research methodology processes life cycle." *Engineering*, vol. 66, no. 9, pp. 2239-2868, 2021. https://doi.org/10.13636/LLPORM.2019.2920292

[16] J. Huang, B. Lan, Z. Liao, D. Zhao, and M. Hou. "Bayesian-Optimized Explainable AI for CKD Risk Stratification: A Dual-Validated Framework." *Symmetry*, vol. 18, no. 1, pp. 81, 2026. https://doi.org/10.3390/sym18010081

[17] T. Paul, M. Assaduzzaman, N. Fahad, and M. J. Hossen. "Enhancing Risk Prediction for Diabetes, Hypertension, and Heart Disease Using SMOTE-ENN Balancing with PCA and Gradient Boosting in Healthcare AI." *Intelligence-Based Medicine*, pp. 100339, 2026. https://doi.org/10.1016/j.ibmed.2025.100339

[18] H. Zhang, J. Zhao, and Y. Liu. "Enhanced PCA Techniques for Nonlinear Biomedical Data Structures." *Mathematics*, vol. 9, no. 4, pp. 437, 2021. https://doi.org/10.3390/math9040437

[19] N. Almansour, et al. "Machine Learning Techniques for the Prediction of Chronic Kidney Disease: A Review." *Computers in Biology and Medicine*, vol. 123, pp. 103898, 2020. https://doi.org/10.1016/j.compbiomed.2020.103898

[20] A. Degirmenci and O. Karal. "Evaluation of kernel effects on svm classification in the success of wart treatment methods." *Am. J. Eng. Res*, vol. 7, pp. 238-244, 2018.

[21] Ö. Karal. "Performance comparison of different kernel functions in SVM for different k value in k-fold cross-

validation." *In 2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pp. 1-5, 2020. https://doi.org/10.1109/ASYU50717.2020.9259880

[22] L. Dalla, T. Medeni, I. Medeni, and M. Ulubay. "Enhancing Healthcare Efficiency at Almasara Hospital: Distributed Data Analysis and Patient Risk Management." *Economy: Strategy and Practice*, vol. 19, no. 4, pp. 54-72, 2025. https://doi.org/10.51176/1997-9967-2024-4-54-72

[23] Y. Liu, H. Sun, and H. Zhang. "Application of PCA and Machine Learning Algorithms in Medical Data Processing." *Mathematics*, vol. 10, no. 4, pp. 552, 2022. https://doi.org/10.3390/math10040552

[24] J. Kalina and A. Schlenker. "Dimensionality reduction methods for biomedical data." *Lékař a technika-Clinician and Technology*, vol. 48, no. 1, pp. 29-35, 2018.

[25] C. Chang, C. Wang, and B. Jiang. "Singular value decomposition based feature extraction technique for physiological signal analysis." *Journal of Medical Systems*, vol. 36, pp. 1769-1777, 2012. https://link.springer.com/article/10.1007/s10916-010-9636-3

[26] N. Dupin, F. Nielsen, and E. Talbi. "Unified polynomial dynamic programming algorithms for p-center variants in a 2D Pareto Front." *Mathematics*, vol. 9, no. 4, pp. 453, 2021. https://doi.org/10.3390/math9040453

[27] S. Yamada and T. Nakano. "Role of chronic kidney disease (CKD)–mineral and bone disorder (MBD) in the pathogenesis of cardiovascular disease in CKD." *Journal of Atherosclerosis and Thrombosis*, vol. 30, no. 8, pp. 835-850, 2023.

[28] E. Almhdi and G. Miskeen. "Power and Carbon Footprint Evaluation and Optimization in Transitioning Data Centres." *Wadi Alshatti University Journal of Pure and Applied Sciences*, vol. 3, no. 2, pp. 221-229, 2025. https://doi.org/10.63318/waujpasv3i2_28

[29] S. Alfathi, G. Miskeen, and W. Mremi. "Evaluation and Prediction Performance of Solar Panel and Wind Turbine Systems Using Simulation." Wadi Alshatti University Journal of Pure and Applied Sciences, vol. 4, no. 1, pp. 94-104, 2026. https://doi.org/10.63318/waujpasv4i1_10

[30] F. Ahmed, A. Othman, and A. Ukasha. "Multi-Class Classification of Skin Cancer Images Using a Deep Learning-Based Convolutional Neural Network (CNN)." *Wadi Alshatti University Journal of Pure and Applied Sciences*, vol. 3, no. 2, pp. 230-243, 2025. https://doi.org/10.63318/waujpasv3i2_29

[31] J. Jefferis, R. Hudson, P. Lacaze, A. Bakshi, C. Hawley, C. Patel, and A. Mallett. "Monogenic and polygenic concepts in chronic kidney disease (CKD)." *Journal of Nephrology*, vol. 37, no. 1, pp. 7-21, 2024.

[32] A. Degirmenci and O. Karal. "Evaluation of kernel effects on svm classification in the success of wart treatment methods." *Am. J. Eng. Res*, vol. 7, pp. 238-244, 2018.

[33] R. Surowiec, O. Reul, N. Chowdhury, R. Rai, D. Segvich, A. Tomaschke, and J. Wallace. "Combining raloxifene and mechanical loading improves bone composition and mechanical properties in a murine model of chronic kidney disease (CKD)." *Bone*, vol. 183, pp. 117089, 2024.

[34] M. Bisher, A. Thamer, and S. Shahata. "Association between Blood Glucose Levels and Biochemical Variables in COVID-19 Patients - A Cross-Sectional Study." *Wadi Alshatti University Journal of Pure and Applied Sciences*, vol. 4, no. 1, pp. 70-75, 2026. https://doi.org/10.63318/waujpasv4i1_07

[35] L. Dalla and T. Ahmad. "Heart Disease Prediction Via Using Machine Learning Techniques with Distributed System and Weka Visualization." *Journal of Southwest Jiaotong University*, vol. 58, no. 4, pp. 322-333, 2023. https://doi.org/10.35741/issn.0258-2724.58.4.26

[36] M. Cao, T. Liang, L. Xu, and F. Shi. "Evaluating the overall renal outcomes of sodium-glucose cotransporter-2 (SGLT2) inhibitors in patients with chronic kidney disease (CKD)." *Diabetology & Metabolic Syndrome*, vol. 17, no. 1, pp. 1-13, 2025.

[37] M. F. Hannan, O. Ajilore, X. J. Zhou, J. P. Lash, Q. Luo, and U. G. Bronas. "Physical function, functional capacity, cognition, and brain structure and function in older adults with chronic kidney disease (CKD)." *Geriatric Nursing*, 2025.

[38] F. Zhang, H. Wang, Y. Bai, L. Huang, Y. Zhong, and Y. Li. "Gait Speed and All-Cause Mortality in Whole-Spectrum Chronic Kidney Disease: A Systematic Review." 2025. https://doi.org/10.1109/ICEST66328.2025.11098207

[39] A. Degirmenci and O. Karal. "iMCOD: Incremental multi-class outlier detection model in data streams." *Knowledge-Based Systems*, vol. 258, pp. 109950, 2022. https://doi.org/10.1016/j.knosys.2022.109950

[40] F. Mishri, and R. Ali. "Detection and Classification of Brain Tumours from MRI Images Using VGG16 Deep Learning Algorithm." *Wadi Alshatti University Journal of Pure and Applied Sciences*, vol. 3, no. 1, pp. 43-50, 2025. https://waujpas.com/index.php/journal/article/view/107

[41] N. H. Tan, T. M. Tuan, P. M. Chuan, N. D. Hoang, L. Q. Thanh, and L. H. Son. "FKG-MM: A multi-modal fuzzy knowledge graph with data integration in healthcare." *PloS One*, vol. 21, no. 1, pp. e0339864, 2026. https://doi.org/10.1371/journal.pone.0339864

[42] S. Abuazoum. "A Comparative Analysis of VxWorks OS and Windows CE in Embedded Real-Time Applications." *Wadi Alshatti University Journal of Pure and Applied Sciences*, vol. 3, no. 2, pp. 9-15, 2025. https://doi.org/10.63318/waujpasv3i2_02

[43] M. Abdulsalam. "Using Hadoop MapReduce for Disease Detection from Real Clinical Data: A Case Study from Al-Mukhtar Clinic in Tripoli." *Wadi Alshatti University Journal of Pure and Applied Sciences*, vol. 3, no. 2, pp. 63-74, 2025. https://doi.org/10.63318/waujpasv3i2_07

[44] R. Alghanay, A. Altajory and A. Abdulsalm. "Prevalence of Trichomona vaginalis Among Married Women Attending Some Private Clinics in Brack Alshatti, Libya". *Wadi Alshatti University Journal of Pure and Applied Sciences*, vol. 3, no. 2, pp. 109-115, 2025. https://doi.org/10.63318/waujpasv3i2_13

[45] Alsaadi. "Removal of Power Line Interference from Electrocardiogram (ECG) Signal." Department of Electrical and Electronic Engineering, Faculty of Engineering, Wadi Alshatti University, Brak, Libya, pp. 55-62, 2023. https://orcid.org/0009-0001-3042-4395

[46] H. Abudrbala. "Study and Analysis of Baseband Signal Using Eye Diagram." *Wadi Alshatti University Journal of Pure and Applied Sciences*, vol. 3, no. 2, pp. 120-124, 2025. https://doi.org/10.63318/waujpasv3i2_15

[47] L. Dalla, A. El-sseid, T. Alarbi, and M. Ahmad. "A domain specific modeling language framework (DSL) for representative medical prescription by using generic modeling environment (GME)." *International Journal of Engineering and Modern Technology*, vol. 6, no. 2, 2020. https://doi.org/10.637388/9040453

[48] A. Mabrouk. "Evaluation of the Relationship between Glycated Hemoglobin (HbA1c) and Atherogenic Index of Plasma (AIP) in Diabetic Patients in Benghazi." *Wadi Alshatti University Journal of Pure and Applied Sciences*,

vol. 3, no. 2, pp. 174-178, 2025. https://doi.org/10.63318/waujpasv3i2_22

[49] N. Fadel and A. Yhmed. "Using the Quality Control Chart to Evaluate the Quality of Some Biochemical Tests (Glucose and Creatinine) in Some Private Medical Laboratories in Brack Alshatti Area." *Wadi Alshatti University Journal of Pure and Applied Sciences*, pp. 202-206, 2025. https://doi.org/10.63318/waujpasv3i2_25

[50] M. Omar, and A. Ukasha. "Using Deep Learning and Particle Swarm Optimization Algorithm to Recognize Fatty Liver in Compressed Images." *Wadi Alshatti University Journal of Pure and Applied Sciences*, vol. 3, no. 1, pp. 119-126, 2025. https://waujpas.com/index.php/journal/article/view/150

[51] Ö. Karal and L. Dalla. "Lung Nodule Characterization in CT Scans Using Hybrid 3D Attention U-Net Segmentation and Transfer Learning-Based Classification Approach." SICST2025, vol. 10, no. 37, ISSN-3014-6266, 2025. https://doi.org/10.13140/RG.2.2.13005.09440

[52] M. Nilashi, H. Ahmadi, A. Manaf, T. Rashid, S. Samad, L. Shahmoradi, and E. Akbari. "Coronary heart disease diagnosis through self-organizing map and fuzzy support vector machine with incremental updates." International Journal of Fuzzy Systems, vol. 22, no. 4, pp. 1376-1388, 2020. https://doi.org/10.1007/s40815-020-00828-7

[53] N. Bacanin, C. Stoean, D. Markovic, M. Zivkovic, T. Rashid, A. Chhabra, and M. Sarac. "Improving performance of extreme learning machine for classification challenges by modified firefly algorithm and validation on medical benchmark datasets." Multimedia Tools and Applications, vol. 83, no. 31, pp. 76035-76075, 2024. https://doi.org/10.1007/s11042-024-18295-9