Intelligent Detection of Chronic Kidney Disease Using Optimized MLP Models and Feature Selection Techniques on the AP-CKD Dataset

B. Ramesh, Karu Prasada Rao

***Abstract*—Chronic Kidney Disease (CKD) is a substantial health issue worldwide. It is important to have accurate and efficient ways to detect CKD. This can help improve outcomes. For this, we collected CKD and non-CKD patients' clinical data in different clinical centres from the north-coastal Districts of AP, India. The study introduces a novel method for detecting CKD using optimized multi-layer perceptron (MLP) models and advanced feature selection techniques. It highlights correlations between key biomarkers and CKD progression. Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) were utilized to reduce feature space by 43.33% (selected 17 features out of 30) and 46.67% (selected 16 features of 30), respectively. The study assesses the Classification Accuracy (CA) of Multi-Layer Perceptron (MLP)-based ML models for CKD detection using Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) for feature selection. Results show that PSO-selected features are more informative for MLP classification tasks, with the MLP model achieving a CA of 97.79%. The MLP+PSO model achieved the highest accuracy of 99.01%, outperforming. The PSO optimization strategy proved more effective in fine-tuning MLP weights. The study uses machine learning to detect CKD early, enhancing feature selection and optimizing Multi-Layer Perceptron models, providing region-specific insights for tailored healthcare interventions and personalized medicine.**

***Index Terms*—Chronic Kidney Disease, Optimized Multi-Layer Perceptron, Feature Selection Techniques, Particle Swarm Optimization, Firefly Algorithms, Genetic Algorithm, Andhra Pradesh CKD Data.**

I. Introduction

As per WHO reports, kidney disease affects 850 million people around the world and causes millions of deaths each year because treatment is often too expensive. Most of those affected live in low- and lower-middle-income countries, with diabetes and high blood pressure being the leading causes. CKD also puts a heavy financial burden on healthcare systems, and better data is needed. CKD affects 9-10% of the world's population [11], [12]. It is the eleventh leading cause of mortality worldwide. Diabetes, hypertension, lifestyle choices, and environmental exposure are all risk factors. CKD causes about 1.2 million deaths per year and is indirectly related to cardiovascular disease [13], [14]. The prevalence of CKD in Uddanam ranges from 18.3% to 32.2%, with estimates suggesting that it could be as high as 60% among agricultural workers. In Andhra Pradesh, the overall prevalence is around 32.2%, with 17.2% in general populations outside of Uddanam. CKD is a significant health problem in AP [15], [16], [17]. Healthcare costs, poverty, and disparities in access to treatment are all components of the financial impact. By 2040, CKD is expected to be the fifth leading cause of death worldwide. WHO activities promote awareness, early detection, and prevention methods. At the same time, groups like the Society for CKD and Nephrology: Improving Worldwide Outcomes works to improve worldwide CKD care standards. **Table I** provides a detailed overview of the stages of CKD according to GFR, describing symptoms, diagnostic methods, treatment approaches, and preventive measures [18]. It emphasizes early detection and management strategies, guiding the understanding of the progression of CKD and the importance of proactive healthcare interventions.

Machine learning (ML) enhances healthcare by enabling intelligent systems for diagnosing illnesses, predicting outcomes, and planning treatments. However, CKD detection requires thorough optimization to provide high accuracy, resilience, and generalizability. Particle Swarm Optimization, Genetic Algorithm, and Firefly Algorithm successfully solved these difficulties. PSO, GA, and FF are problem-solving techniques. PSO is based on how swarms behave. GA uses principles from natural evolution to improve solutions. FF imitates how fireflies glow to help find solutions. These methods help select important features and tune model parameters. They remove unnecessary features and optimize performance. This study aims to improve the performance of MLP models in detecting CKD using the clinical AP-CKD dataset, ensuring reliable and interpretable results for healthcare professionals using advanced optimization techniques.

II. Literature Review

Chronic Kidney Disease (CKD) is a global health issue. Early detection and treatment are important. Machine learning (ML) models help in diagnosis and prognosis. Common models include Logistic Regression (LR), Decision (DT) Trees, and Random (RF) Forests. High-quality data is necessary for training these models. Data preprocessing techniques are important. These techniques include normalization and feature selection. ML can improve diagnostics [Dutta et al. (2024)[19]. Swain et al. (2023) [20] present a robust approach to CKD classification using machine learning models, including SVM and RF models. RF achieves 98.67% accuracy, outperforming. SMOTE addresses class imbalance, and the chi-squared

Manuscript received April 4, 2025; revised July 31, 2025.

B. Ramesh, Research is a PhD Scholar of the Department of CSE, School of Technology, GITAM Deemed To Be University, Visakhapatnam, Andhra Pradesh, India (phone: +91 99857 81116; fax: 08945 245266; e-mail: ramesh.bandaru15@gmail.com).

Karu Prasada Rao is an Assistant Professor of the Department of CSE, School of Technology, GITAM Deemed to Be University, Visakhapatnam, Andhra Pradesh, India (email: pkaru@gitam.edu).

TABLE I
OVERVIEW OF CHRONIC KIDNEY DISEASE (CKD) STAGES WITH SYMPTOMS, DIAGNOSIS, TREATMENT, AND PREVENTIVE MEASURES

Stages of CKD (GFR)	Symptoms	Diagnosis	Treatment	Precautions and Preventions
Stage 1 (GFR $\geq$ 90)	Often asymptomatic, mild proteinuria	Blood tests, urine tests, blood pressure monitoring	Control BP, manage diabetes	Healthy diet, regular exercise, avoid smoking, control hypertension [1]
Stage 2 (GFR 60–89)	Mild fatigue, slight swelling, proteinuria	Blood and urine tests, imaging tests	Monitor kidney function, manage comorbidities	Maintain healthy weight, reduce salt intake, stay hydrated
Stage 3 (GFR 30–59)	Fatigue, swelling in hands/feet, back pain, urination changes	Blood tests, creatinine levels, electrolyte tests	Medications to manage symptoms, avoid nephrotoxic drugs	Avoid NSAIDs, manage diabetes and hypertension
Stage 4 (GFR 15–29)	Severe fatigue, swelling, nausea, loss of appetite	Blood tests, kidney biopsy, advanced imaging	Kidney transplant, dialysis preparation or manage complications	Follow dietary restrictions, potassium monitoring, regular checkups
Stage 5($GFR <$ 15)	Severe symptoms: vomiting, confusion, shortness of breath	Blood tests, dialysis evaluation, imaging	Kidney transplant, dialysis, supportive care	Early detection in previous stages, adherence to medical advice [2], [3]

TABLE II
EXISTING RESEARCH ON CKD PREDICTION MODELS AND RESULT ANALYSIS

Ref. No.	Author(s) (Year)	Description and Models	Result Analysis
[4]	Chittora et al. (2021)	Models utilized include ANN, C5.0, CHAID, Logistic Regression, LSVM with L1 and L2 penalties, and Random Tree.	Highest Accuracy: 98.86% with LSVM (L2) and SMOTE with Full Features. Second-best: LSVM (L2) with LASSO and SMOTE.
[5]	Debal et al. (2022)	The study evaluated LR, DTs, and RFs for early CKD detection, focusing on categorical predictions and ensemble methods.	LR outperformed with an F1 score of 0.87, precision of 91.49%, and recall of 83.49%, while RF achieved an F1 score of 0.71, precision of 64.29%, and recall of 78.64%.
[6]	Islam et al. (2023)	Used predictive ML models to identify early CKD, selecting 30% of relevant features from 25 variables.	XGBoost performed best with an accuracy of 98.3%, precision of 0.98, recall of 0.98, and F1-score of 0.98.
[7]	Alsuhibany et al. (2021)	Developed an Ensemble DL Clinical Decision Support System (EDL-CDSS) for CKD diagnosis in IoT frameworks. Compared DT, MLP, DBN, CNN-GRU, KELM, FNC, D-ACO, and EDL-CDSS.	EDL-CDSS showed the best performance with sensitivity of 0.9680, specificity of 0.9702, accuracy of 0.9691, and F1-score of 0.9692.
[8]	Chowdhury et al. (2021)	Assessed 10 ML models, including RF, LightGBM, KNN, SVM, DT, GB, XGB, SGD, LR, and GNB for CKD prediction.	RF had the highest accuracy (96%), followed by LightGBM (95%). Other models also performed above 90%.
[9]	Srikanth et al. (2023)	Used ensemble ML techniques for CKD prediction using a dataset of 201 records and 29 attributes. Evaluated RF, DT, SVM, and AdaBoost.	RF and SVM achieved 98.3% accuracy, while DT reached 96.6%.
[10]	Ghosh et al. (2023)	Compared ML algorithms (XGBoost, RF, LR, AdaBoost) and proposed a Hybrid Model for CKD prediction.	Hybrid Model outperformed all with accuracy of 94.99%, precision of 95.21%, recall of 95.11%, F1-score of 95.32%, and AUROC of 95.56%.

test optimizes feature selection. Arif et al. (2023) [21] created an ML framework to detect CKD early. They used advanced preprocessing methods and the Boruta algorithm. By applying the K-NN algorithm and optimizing hyperparameters, they achieved perfect accuracy on the UCI CKD dataset. Bai et al. (2022) [22] used ML models to predict the progression of CKD to end-stage kidney disease (ESKD). They used a longitudinal dataset and tested five ML algorithms, with logistic regression, NBs, RF, DTs, and k-NNs performing well. The study suggests ML models can be helpful for patient screening and early detection but acknowledges the need for external validation and further enhancement. Dritsas et al. (2022)[23] developed ML tools for predicting CKD using class-balancing techniques. They used feature ranking and analysis to identify key variables influencing CKD prediction. The study emphasizes the importance of preprocessing steps in building robust models for medical applications but lacks specific algorithmic comparisons and metric outcomes. Iftikhar et al. (2023) [24] compared ML models for predicting CKD using a case-control dataset from Pakistan. The models were assessed using many criteria, such as specificity, accuracy, sensitivity, and the Diebold-Mariano test. The research indicated kernel-based SVMs were proficient in predicting CKD, highlighting the need for a thorough assessment. Singh et al. (2022) [25] devised a deep neural network (DNN) for the early identification and prediction of CKD, emphasizing diabetes and hypertension. The DNN surpassed traditional classifiers in predicted accuracy and robustness, underscoring the efficacy of deep learning in medical datasets. Rashed-Al-Mahfuz et al. (2022) [26] studied CKD using machine-learning techniques. They employed eight classifiers, such as Logistic Regression and Random Forest, to evaluate the disease's characteristics. They utilized principal component analysis (PCA) for feature extraction to enhance diagnosis accuracy. **Table II** presents a comparative analysis of various CKD prediction models, outlining their descriptions, techniques used, and performance results.

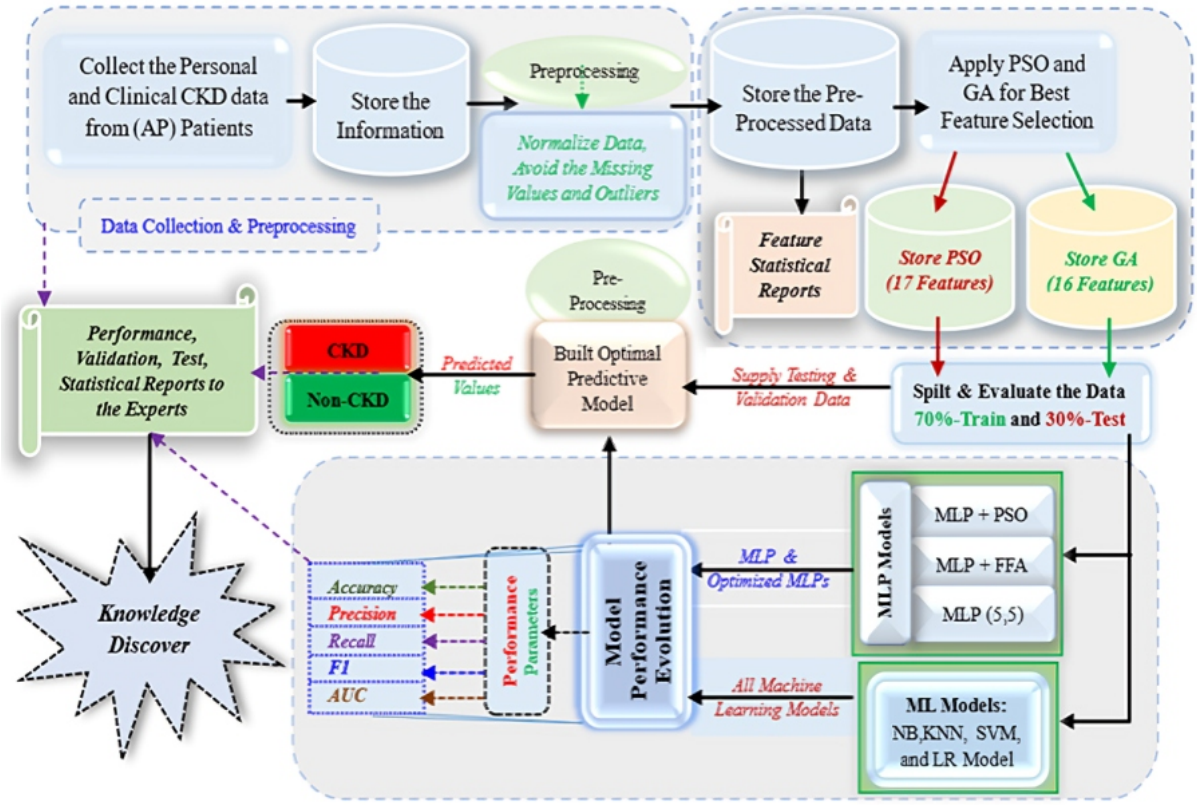

Fig. 1. Framework of the Proposed CKD Detection Model Using AP CKD Dataset

Venkatrao et al. (2023) [27] created HDLNet, a hybrid DL network model for early detection and classification of CKD using advanced architectures and optimization techniques. The model demonstrated superior performance on the UCI CKD dataset. Terlapu et al. (2023) [28] developed a hybrid diagnostic model for Uddanam nephropathy, a localized CKD variant in Andhra Pradesh, India. The model, which combined Principal Component Analysis (PCA) with Genetic Algorithm (GA) and Multi-Layer Perceptron (MLP), outperformed standard classification methods in testing accuracy of 98.54%. Deepika et al. (2023) [29] propose a hybrid IoMT platform for early chronic kidney disease detection and diagnosis, using HMANN and FFOA algorithms for enhanced accuracy and data reduction. The methodology includes preprocessing steps like segmenting kidney regions in ultrasound images. Hosseinzadeh et al. (2021) [30] developed a diagnostic model for chronic kidney disease using IoT-based multimedia data, enhancing predictive performance and reducing computational overhead. The Decision Tree classifier outperforms other models, achieving 97% accuracy and 95% specificity. Further validation on diverse datasets is needed. Gogoi et al. (2024) [31] review ML techniques for CKD prediction and diagnosis, highlighting progress and challenges like small datasets, lack of stage-specific models, and interpretability and privacy concerns. They propose solutions like Generative AI, SHAP and LIME, and privacy-preserving methods like homomorphic encryption and federated learning. Akter et al. (2024) [32] developed CKD.Net, a DL hybrid model that predicts the five stages of CKD using a balanced dataset with 27 features. The model achieves outstanding classification performance and noninvasive prediction of eGFR and creatinine levels with high confidence. This advancement in AI-driven clinical diagnostics is significant for early CKD detection. Vital et al. (2021) [33] conducted a study on Uddanam Chronic Kidney Disease (UCKD) in the coastal Srikakulam region of Andhra Pradesh, India. They used statistical analysis and machine learning techniques to analyse CKD datasets, including Naïve Bayes, k-NN, Logistic Regression, C4.5, SVM, and Probabilistic Neural Networks (PNN). The PNN model showed superior performance, enabling early diagnosis and supporting resource-constrained efforts. Elkholy et al. (2023) [34] propose a ”DFS-ODBN” framework for early and exact detection of CKD. The framework uses a deep belief network optimized by the Grasshopper Optimization Algorithm and a Density-based Feature Selection algorithm. The model outperforms alternative methods with reasonable accuracy, sensitivity, and specificity, advancing medical diagnostics.

III. Models and Materials

The section discusses the clinical AP-CKD dataset from Andhra Pradesh, India, and its optimized MLP models, utilizing advanced feature selection techniques like PSO, GA, and Firefly Algorithm for enhanced detection performance.

1) Proposal Model: **Fig. 1** shows the framework of a proposed model for detecting chronic kidney disease (CKD). It uses the AP CKD dataset to make accurate predictions and analysis. The CKD detection model is designed to predict CKD early using advanced technology and machine learning (ML) methods. We collected the patient's clinical data from various clinical centers and hospitals in Andhra Pradesh.

The data is cleaned and prepared to ensure its quality. Important features needed to detect CKD are chosen using Particle Swarm Optimization (PSO) and Genetic Algorithms (GA). The data is divided into 70% for training and 30% for accuracy testing. Different ML models, such as Naive Bayes, K-Nearest Neighbours, Support Vector Machine, and Logistic Regression, are trained and compared to see which predicts CKD best. The model's performance is evaluated using precision, accuracy, F1 score, recall, and AUC, which help understand how well each model works. After finding the best model, it is tested for reliability. Experts check the model to ensure it meets clinical standards for real-life use. Finally, detailed reports on the model's performance are created for stakeholders, ensuring transparency and helping them make informed decisions.

2) Dataset Description: The CKD dataset was collected from different clinical centres and hospitals from North coastal districts (Srikakulam, Vizainagaram, and Visakhapatnam) of Andhra Pradesh, India. The AP CKD dataset includes 1,348 patient records. After eliminating missing values and outliers, there are 1,150 patient records for analysis, including 29 features and one target variable. These features include personal details like age and gender, necessary health measurements such as blood pressure, and urine test results showing Specific Gravity and Albumin levels. It includes biochemical indicators like haemoglobin, serum creatinine, and blood urea, as well as immune system markers like neutrophils and lymphocytes. It classifies individuals as CKD (1) or non-CKD (0). It is helpful for ML and statistical models. The **Table III** shows in detailed information about each feature attributes (A00 to A29) and one class attribute (T01) with range of each attribute.

3) Firefly Algorithm: The Firefly (FF) Algorithm is an optimization algorithm based on how fireflies flash. This method is effective for resolving complex challenges. Fireflies use flashlights to attract mates or prey. The light dims as they travel away. A dim firefly will go towards a brighter one. If there are no brighter fireflies nearby, they will travel at random. The efficient global search method is easy to use and adaptable. It works well for complex problems with many dimensions. It helps avoid getting stuck in local optima and can be applied to different types of problems. The FF treats all fireflies as the same gender. Their attractiveness depends on how bright they are. If a firefly doesn't see a brighter one nearby, it will move randomly.

The objective function f(x) is to be minimized or maximized where x is a solution vector in d dimensions.

$$f(x), \quad x \in \mathbb{R}^d \tag{1}$$

An objective function defines a firefly's brightness at a particular position. In a maximization problem, increasing brightness improves the objective function. In a minimization problem. I(x) is the intensity of light determined as

$$I(x) \propto f(x) \tag{2}$$

$$I(x) = \frac{1}{1 + f(x)} \tag{3}$$

The attractiveness of a firefly is determined by its beta value.

$$\beta(r) = \beta_0 e^{-\gamma r^2} \tag{4}$$

β_0 : Initial attractiveness at a distance $r = 0$.
γ : Light absorption coefficient.
r : Distance between two fireflies.

$$r = \|\mathbf{x}_i - \mathbf{x}_j\|^2 = \sqrt{\sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2} \tag{5}$$

The formula for attracting a less bright firefly i to a brighter firefly j is:

$$x_i = x_i + \beta_0 e^{-\gamma r^2}(x_j - x_i) + \alpha \dot{o}^1 \tag{6}$$

xi represents the current position of a firefly. xj indicates the current position of another firefly, which is used for attraction. α is the randomization parameter that determines the step size. ϵ is mentioned as random number vector drawn from a Gaussian or Uniform distribution If a firefly doesn't see a brighter one, it will move in a random direction as follows:

$$x_i = x_i + \alpha(\text{rand} - 0.5) \tag{7}$$

Fig. 2 shows the detailed analysis of FF+MLP fine-tuned algorithms An MLP is a feedforward ANN with input, hidden, and output layers. It learns through weighted edges, and its performance is significantly influenced by its hyper parameters. The Firefly Optimization Algorithm (FOA) is a method that uses the bioluminescent behaviour of fireflies to attract them based on their brightness, aiming to find optimal solutions. The tuning of hyper parameters in a Multi-Layer Perceptron (MLP) is focused on reducing the error when validating the model. Key factors in this optimization process include Learning rate, the no. of neurons and layers that are hidden in each layer, activation functions, weight initialization methods, dropout rates, and batch size. Implementing a ML model entails designing the MLP architecture, configuring the FOA parameters, integrating FOA with MLP training, executing the FOA algorithm, and assessing the final model. The model is trained using the optimal hyper parameters and then evaluated on a different dataset to determine its performance. This method guarantees precise and efficient ML.

The FF (shown in **Algorithm 1**) is a technique of optimization. It simulates how fireflies flash to discover the optimum option. The algorithm adjusts their placements over time to come closer to the best solution.

Algorithm 1 Firefly Algorithm

Initialization: Randomly initialize the firefly population
Step 1: Evaluate Brightness: Calculate brightness using the objective function.
Step 2: Firefly Movement: For each firefly

- Move towards the brighter firefly based on attractiveness.
- If no brighter firefly is found, move randomly.

Step 3: Update Solutions: Adjust the fireflies' positions and recalculate brightness.
Step 4: Repeat: Iterate until the stopping criterion (maximum iterations or convergence) is met.
Step 5: Return the Best Solution: Output the optimal solution. =0

TABLE III
CKD DATASET DESCRIPTION

(Code) Attribute	Data Description	(Code) Attribute	Data Description
(A00) Age	Continuous	(A16) Potassium	Continuous
(A01) Sex	Discrete	(A17) Chloride	Continuous
(A02) BP	Discrete	(A18) Haemoglobin	Continuous
(A03) Specific-Grvy	Continuous	(A19) PackedCell	Continuous
(A04) Albumin	Continuous	(A20) WBCells	Continuous
(A05) Urinal Sugar	Continuous	(A21) Neutrophils	Continuous
(A06) MCV	Continuous	(A22) Lymphocytes	Continuous
(A07) Platelet-Count	Continuous	(A23) Eosinophils	Continuous
(A08) RedBloodCells	Continuous	(A24) Monocytes	Continuous
(A09) Pus-Cell	Discrete	(A25) Basophils	Continuous
(A10) PusCellClumps	Discrete	(A26) Uric Acid	Continuous
(A11) Bacteria	Discrete	(A27) Bilirubin	Continuous
(A12) Blood-Glucose	Continuous	(A28) Hypertension	Discrete
(A13) Blood-Urea	Continuous	(A29) Diabetic	Discrete
(A14) Serum Creatine	Continuous	(T01) CKD/NON-CKD	Discrete
(A15) Sodium	Continuous		

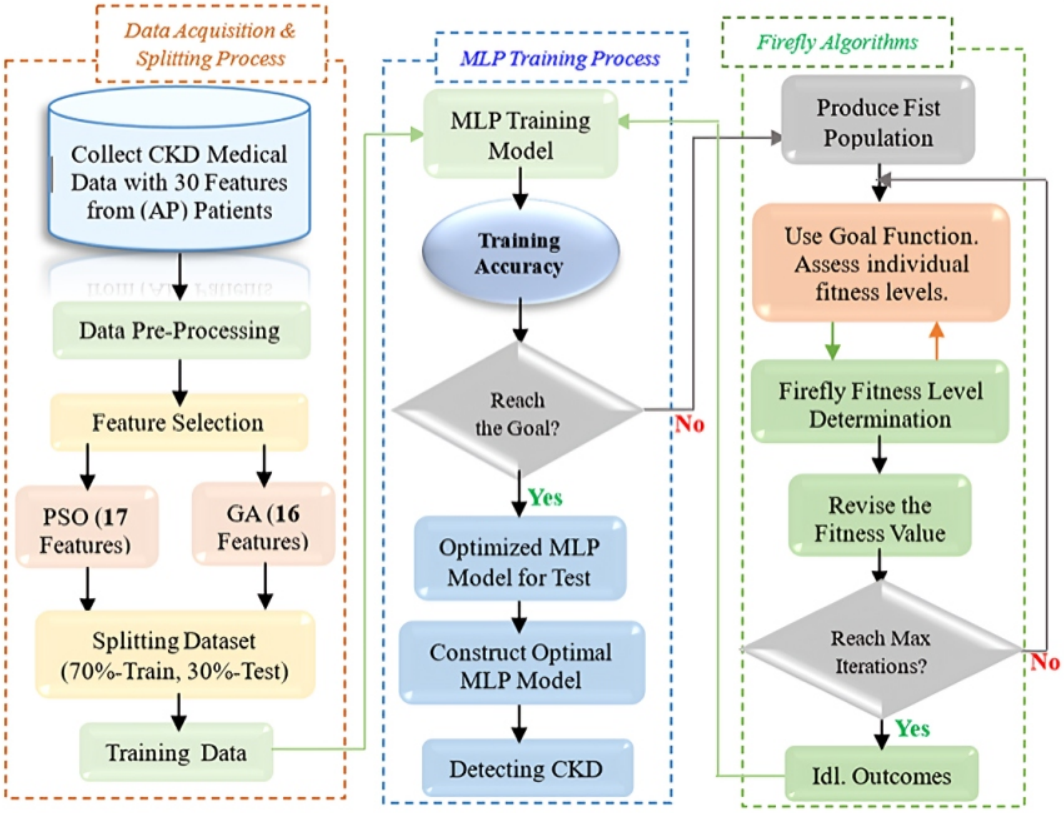

Fig. 2. Framework of the MLP Hyper Parameters Tuning using Firefly Algorithm

4) PSO Algorithm: Russell Eberhart and James Kennedy developed the Particle Swarm Optimization (PSO) method 1995.The algorithm is based on the premise that birds in a flock communicate information to alter their location while seeking for food, assessing their personal and global best experiences. PSO is a search space model that provides a possible solution inside a swarm of particles. PSO is easy to implement because it has few parameters to adjust. It works well for high-dimensional problems and converges quickly for continuous optimization. Additionally, it is flexible and can be adapted for various optimization tasks. It entails computing each particle's position (xi), velocity (vi), personal best (pbesti), and global best (gbesti) depending on its location in the search space. The objective function to be minimized or maximized should be

$$f(x), \quad x \in \mathbb{R}^d \tag{8}$$

The d-dimensional solution vector represents a particle's position. The particle position and velocity are updated using specific equations.

$$v_j^{(t+1)} = wv_j^{(t)} + c_1 r_1 (pbest_j - x_j^{(t)}) + c_2 r_2 (gbest - x_j^{(t)}) \tag{9}$$

$v_j^{(t+1)}$ is the velocity of particle i has been updated at iteration t+1. Inertia weight w, self-confidence c1, swarm

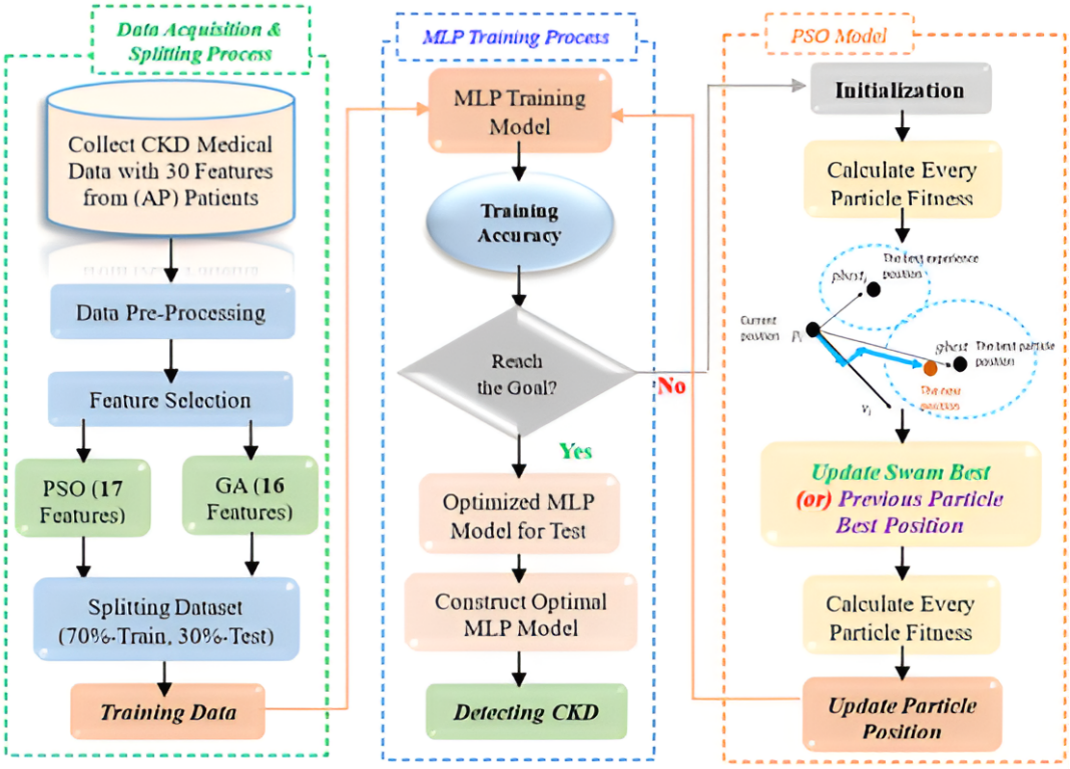

Fig. 3. Framework of the MLP Hyper Parameters Tuning using PSO Model

confidence c2, r1 and r2 are Random numbers between [0 1].

$$x_i^{(t+1)} = x_i^{(t)} + v_i^{(t+1)} \tag{10}$$

The position update is $x_j^{(t+1)}$ the updated position i of the particle during the iteration. Inertia weight adjustment is a technique that maintains the balance between exploration and exploitation by reducing it over iterations.

$$w = w_{max} - \frac{(w_{max} - w_{min})}{t_{max}} \cdot t \tag{11}$$

The **Fig. 3** shows the PSO+MLP fine-tuned models. The

TABLE IV
ANALYSIS OF CONFUSION MATRIX STRUCTURE

	Class	Predicted Values		Total
		NCKD (0)	CKD (1)	
Actual Values	NCKD (0)	(0,0)	(0,1)	T3
	CKD (1)	(1,0)	(1,1)	T4
	Total	T1	T2	T

Particle Swarm Optimization (PSO) **(Algorithm 2)** technique optimizes the parameters of an MLP, improving performance by automating the search for hyper parameters such as learning rate, number of neurons, activation functions, and weight initialization schemes. This population-based optimization method is critical for improving the MLP's performance. An MLP is a neural network with three layers: input, hidden, and output. It is used for tasks like classification and regression. The performance of an MLP depends on the correct hyper parameters. PSO is an optimization method inspired by birds and fish. It involves particles that explore a search space to find optimal solutions, adjusting their positions based on personal and collective experiences. To tune MLP hyper parameters, we aim to reduce errors on a validation dataset. It's essential to establish the range for different hyper parameters, such as the learning rate, the number of hidden layers, the number of neurons in each layer, the activation functions, regularization parameters, and batch size.

Algorithm 2 Particle Swarm Optimization (PSO)

1: **Initialization:** Randomly initialize the positions x_i and velocities v_i of all particles in the search space. Set parameters: w (inertia weight), c_1 (cognitive coefficient), c_2 (social coefficient).
2: **Step 1:**
   a. **Evaluate Fitness:** Evaluate the fitness of each particle using the objective function $f(x_i)$. Update $pbest$ (personal best) for each particle.
   b. **Update *gbest* :** Update the global best ($gbest$) based on the best fitness value across the swarm.
3: **Step 2: Update Velocity and Position**
   a. **Update Velocity:** Update each particle's velocity using the velocity update equation: $v_i \leftarrow wv_i + c_1 r_1(pbest_i - x_i) + c_2 r_2(gbest - x_i)$, where r_1, r_2 are random numbers in $[0, 1]$.
   b. **Update Position:** Update each particle's position using the position update equation: $x_i \leftarrow x_i + v_i$.
4: **Step 3: Check Stopping Criteria** Iterate until the stopping criterion (e.g., maximum iterations or convergence) is met.
5: **Step 4: Return the Best Solution** Output the global best position ($gbest$) and its corresponding fitness value $f(gbest)$. =0

TABLE V
CKD DATASET CONTINUOUS FEATURES STATISTICAL ANALYSIS

Attribute	CKD		NON-CKD		CKD & NON-CKD (Total)	
	Mean	Median	Mean	Median	Mean	Median
(A00) Age	51.74 ± 15.68	54	46.72 ± 16.15	46	49.86 ± 16.04	51
(A03) Specific-Grvy	1.01 ± 0	1.015	1.02 ± 0	1.02	1.02 ± 0.01	1.015
(A04) Albumin	3.35 ± 0.99	3.36	0.32 ± 0.77	0	2.21 ± 1.73	2.65
(A05) Urinal Sugar	1.47 ± 1.73	1	0.17 ± 0.47	0	0.98 ± 1.53	0
(A06) MCV	82.12 ± 10.63	83	80.91 ± 9.34	81	81.67 ± 10.18	82.5
(A07) Platelet-Count	2.25 ± 0.9	2.1	2.45 ± 0.88	2.4	2.33 ± 0.9	2.2
(A08) RedBloodCells	3.27 ± 1	3.21	5.1 ± 0.82	5.2	3.96 ± 1.29	3.885
(A12) Blood-Glucose	181.63 ± 88.36	154.5	115.63 ± 33.03	111.5	156.95 ± 79.46	129
(A13) Blood-Urea	73.2 ± 55.63	57	35.68 ± 16.83	34	59.17 ± 48.7	45
(A14) Serum Creatinine	7.4 ± 3.49	7.81	2.86 ± 1.24	2.91	5.7 ± 3.61	4.82
(A15) Sodium	138.46 ± 12.12	139.2	141.25 ± 5.21	141	139.51 ± 10.19	139.6
(A16) Potassium	6.02 ± 7.98	5.14	4.44 ± 0.68	4.6	5.43 ± 6.37	4.8
(A17) Chloride	106.52 ± 8.73	105.8	103.51 ± 4.86	102.45	105.4 ± 7.66	104.3
(A18) Haemoglobin	9.18 ± 2.84	8.5	14.71 ± 1.86	14.9	11.25 ± 3.68	10.85
(A19) Packed Cell	26.25 ± 7.3	24.75	44.5 ± 5.73	45	33.07 ± 11.12	31.2
(A20) WBCells	7608.36 ± 4159.03	6800	7008.13 ± 2683.69	7200	7383.93 ± 3687.5	7000
(A21) Neutrophils	60.43 ± 13.38	62	60.07 ± 13.54	61.9	60.29 ± 13.43	61.9
(A22) Lymphocytes	29.47 ± 11.35	28	30.41 ± 11.15	29.1	29.82 ± 11.28	28.4
(A23) Eosinophils	4.1 ± 4.02	2.8	3.69 ± 3.74	2.6	3.94 ± 3.92	2.7
(A24) Monocytes	5.29 ± 3.38	4.75	5.28 ± 3.22	4.7	5.29 ± 3.32	4.7
(A25) Basophils	0.16 ± 0.26	0.1	0.15 ± 0.25	0.1	0.16 ± 0.26	0.1
(A26) Uric Acid	6.82 ± 2.03	6.815	6.51 ± 2.16	6.46	6.71 ± 2.08	6.78
(A27) Bilirubin	0.76 ± 1.22	0.53	0.74 ± 1.17	0.5	0.75 ± 1.2	0.52

TABLE VI
CKD DATASET DISCRETE FEATURES STATISTICAL ANALYSIS.

Attribute	CKD	NCKD
(A01) **Sex**	Male-515(71.53%)	Male-323(75.12%)
	Female-205(28.47%)	Female-107(23.49%)
(A02) **Blood pressure**	Normal-127(17.64%)	Normal-139(32.33%)
	High-295(40.97%)	High-23(5.35%)
	Low-298(41.39%)	Low-269(62.56%)
(A09) **Pus cell**	Normal-378(52.50%)	Normal-377(87.67%)
	Abnormal-342(47.50%)	Abnormal-53(12.33%)
(A10) **Pus cell clumps**	Present-478(66.39%)	Present-26(6.05%)
	Not Present-242(33.61%)	Not Present-404(93.95%)
(A11) **Bacteria**	Present-245(34.03%)	Present-23(5.35%)
	Not Present-475(65.97%)	Not Present-407(94.65%)
(A28) **Hypertension**	Present-421(58.47%)	Present-37(8.6%)
	Not Present-299(41.53%)	Not Present-393(91.4%)
(A29) **Diabetic**	Present-441(61.25%)	Present-13(3.02%)
	Not Present-278(38.61%)	Not Present-417(96.98%)

5) Confusion Matrix and Performance Metrics: The confusion matrix (**Table IV**) assesses a model's accuracy in predicting CKD and NCKD, encompassing True Positives (TP), False Negatives (FN), False Positives (FP), and True Negatives (TN). The matrix also includes totals for predicted and actual cases of both diseases.

Performance matrices

ML metrics include accuracy (ACC), precision (PRE), recall (REC), and F1-score. Accuracy shows how many predictions were correct. Precision indicates the percentage of true positives among all positive predictions. The recall is a measure of how many TPs were recognized. The F1 score combines ACC and REC to provide a fair evaluation. Equations (1) to (10) shows the whole metrics for classes CKD(1) and NCKD(0).

$$Accuracy = \frac{TP(CKD) + TP(NCKD)}{Total(CKD) + Total(NCKD)} \tag{12}$$

$$Precision(CKD) = \frac{TP(CKD)}{TP(CKD) + FP(CKD)} \tag{13}$$

$$Precision(NCKD) = \frac{TP(NCKD)}{TN(NCKD) + FP(NCKD)} \tag{14}$$

$$OverallPrecision = \frac{Precision(CKD) + Precision(NCKD)}{2} \tag{15}$$

$$Recall(CKD) = \frac{TN(CKD)}{TN(CKD) + FN(CKD)} \tag{16}$$

$$Recall(NCKD) = \frac{TP(NCKD)}{TP(NCKD) + FalseNegative(NCKD)} \tag{17}$$

$$OverallRecall = \frac{Recall(CKD) + Recall(NCKD)}{2} \tag{18}$$

$$F1Score(CKD) = 2\times\frac{Precision(CKD) \times Recall(CKD)}{Precision(CKD + Recall(CKD)} \tag{19}$$

$$F1Score(NCKD) = 2\times\frac{Precision(NCKD)\times Recall(NCKD)}{Precision(NCKD)+Recall(NCKD)} \quad (20)$$

$$OverallF1Score = \frac{F1Score(CKD)+F1Score(NCKD)}{2} \quad (21)$$

IV. Result Analysis

The Result Analysis section evaluates proposed MLP models using CA and AUC metrics, highlighting the effectiveness of PSO, GA, and the Firefly Algorithm for feature selection. The MLP+PSO model demonstrates superiority in CKD detection.

A. Statistical Analysis

The CKD dataset (**Table V**) describes significant differences between CKD and non-CKD groups across various biomarkers. CKD patients have a higher average age, suggesting increased susceptibility to kidney-related disorders. Specific gravity decreases in CKD patients, indicating reduced kidney function regulating urine concentration. Albumin levels are significantly higher in CKD patients, indicating impaired kidney function in preventing protein leakage into urine. Urine sugar levels are also elevated in CKD patients, indicating a higher prevalence of glucose-related metabolic disruptions. Mean corpuscular volume and platelet counts are slightly elevated in CKD patients, suggesting anemia-related complications. Blood glucose levels are higher in CKD patients, suggesting a link between diabetes and kidney dysfunction. Elevated blood urea and serum creatinine confirm impaired kidney filtration. Electrolyte imbalances in CKD patients are also observed. Hemoglobin levels are reduced in CKD patients, supporting anemia as a hallmark complication. Uric acid levels are slightly elevated in CKD patients, indicating reduced clearance efficiency by compromised kidneys. The CKD dataset (**Table VI**) shows apparent differences between people with CKD and non-CKD. In the CKD group, most patients are male, making up 71.53% of the total. In contrast, the non-CKD group has an even higher percentage of males, 75.12%, and 23.49% of females. Blood pressure patterns show that hypertension is prevalent in CKD patients, with 17.64% having normal blood pressure and 40.97% having high blood pressure. Pus cell analysis reveals that CKD patients' urinary tract or renal infections are more common. Pus cell clumps indicate severe kidney infection or damage, while bacterial presence is strongly associated with CKD pathology. Hypertension is both a cause and an effect of CKD, with 58.47% having a history of hypertension and 41.53% not having it. Diabetes is one of the leading causes of CKD, with 61.25% having it and 38.61% not having it.

B. Feature Selection using PSO Algorithm

The Particle Swarm Optimization (PSO) algorithm has enhanced ML models for detecting chronic kidney disease (CKD). The algorithm **(Fig. 4)** selects 17 key features, such as age, sex, albumin, and more, which correlate with CKD risk factors. This results in a 43.33% reduction in feature space while retaining the most influential attributes. The PSO-based feature selection improves model performance, interpretability, and scalability, enabling faster training with fewer resources, improved prediction accuracy, and more interpretable results. The PSO algorithm reduced feature space by 43.33% while retaining key features for CKD detection, selecting 17 from the CKD dataset, including age, sex, albumin, MCV, pus-cell, bacteria, blood-glucose, blood-urea, sodium, potassium, packed cell volume, white blood cell count, neutrophils, lymphocytes, monocytes, basophils, and hypertension.

PSO selected CKD Features Correlation Analysis: The correlation heatmap (**Fig. 5**) of the AP-CKD dataset shows important relationships among features enhanced using PSO. There are strong positive correlations, such as T01-F28 with a correlation of 0.49, T01-F12 with a correlation of 0.40, and T01-F13 with a correlation of 0.37. These findings indicate that features like urea, creatinine, and other markers are closely linked to CKD. Conversely, negative correlations, such as T01-F19 (-0.79) and T01-F09 (-0.36), correspond to clinical signs of anaemia in CKD. These findings are reinforced by statistical differences in the CKD and non-CKD groups for haemoglobin, creatinine, and urea. PSO feature selection focuses on important features such as F28, F12, F13, F11, F00, F09, and F19 for CKD. Features like F01, F06, and F22, which are redundant or weakly correlated, are less important unless they show non-linear patterns.

Statistical Feature Alignment: Linking correlation patterns to medical statistics (Table V): Creatinine (A14): Very high in CKD (7.4 vs. 2.86). Mapped to F13/F12. Positively correlated with T01 (CKD) and F28. Blood Urea (A13): Same pattern (73.2 vs. 35.68). Strongly supports positive correlation with CKD labels. Haemoglobin (A18): Very low in CKD (9.18 vs. 14.71). Mapped to F19, which has strong negative correlation with CKD. RBC (A08): 3.27 in CKD vs. 5.10 in Non-CKD $\rightarrow$ confirms negative correlation (T01–F09). Potassium (A16): High variability in CKD (6.02±7.98) $\rightarrow$ appears in moderate relationships (e.g., F16). Age (A00): Older patients more likely to have CKD (51.74 vs. 46.72) $\rightarrow$ justifies F00–F28 positive correlation. The analysis shows that T01 (the presence of CKD) is closely linked to features F28, F12, F13, and F11, indicating that these features are important for predicting CKD. Additionally, features F19 and F09 may be connected to important health factors like haemoglobin and red blood cell count, emphasizing their combined relevance in CKD diagnosis.

C. Feature Selection using GA Algorithm

The Genetic Algorithm was used to study the CKD dataset. It helped reduce the number of features while keeping important ones for predicting the disease. The study identified 16 key features that improved accuracy and efficiency. It highlighted how advanced feature selection methods, like Best First Search, can enhance the performance of MLP and ML models. The Genetic Algorithm (GA) (**Fig. 6**) has been used to identify 16 key features for detecting CKD, reducing the original feature space by 46.67%. These

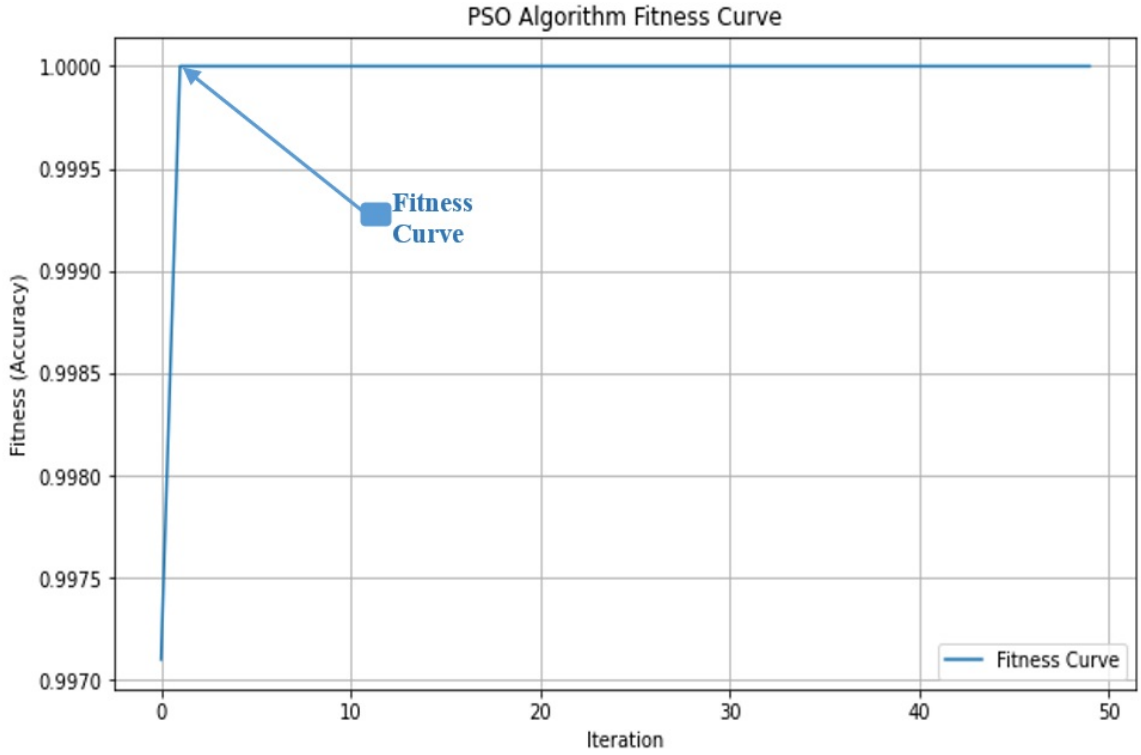

Fig. 4. PSO Algorithm Selected Features from CKD Dataset Fitness Curve with fitness value 1

features include age, gender, kidney function, albumin, anemia, red blood cells, pus-cells, blood glucose, high blood urea, sodium imbalance, packed cell volume, white blood cell count, neutrophils, monocytes, basophils, and hypertension.

GA selected CKD Features Correlation Analysis (Fig. 7): The GA-based feature correlation heatmap helps us understanding the relationship between clinical features and CKD status in the AP-CKD dataset. Strong positive correlations with features F28, F12, and F13, indicate important biomarkers, such as urea and creatinine, that increase as CKD progresses. Conversely, negative correlations, particularly with F19 (haemoglobin) and F08/F09 (red blood cells), reflect the anaemia associated with CKD patients. F28 stands out as a central feature, closely connected to both the target and other markers. The heatmap shows Pearson correlation coefficients between GA-selected features and target class in the AP-CKD dataset. Positive correlation indicates a direct relationship, negative correlation indicates an inverse link, and near-zero correlation indicates weak or no linear relationship. The study shows that important features such as F28, F12, F13, F11, F00, F09, and F19 should be prioritized with higher fitness values in GA optimization and kept in MLP modelling. On the other hand, features like F01, F06, and F22 should be excluded unless justified by non-linear patterns.

Statistical Feature Alignment: As per Table V and VI, connecting key medical statistics to feature correlations: Creatinine (A14): Extremely elevated in CKD (7.4 vs. 2.86) → aligns with F13 and T01/F28 positive correlations. Blood Urea (A13): Sharp rise in CKD (73.2 vs. 35.68) → possibly F12, positively tied with CKD. Haemoglobin (A18): Significant drop in CKD (9.18 vs. 14.71) → strongly negative with F19. RBC (A08): Lower in CKD (3.27 vs. 5.10) → correlates negatively with T01 and F08/F09. Potassium (A16): High variance (6.02±7.98) → weakly correlates, appears moderately linked to F16. Age (A00): Higher in CKD group (51.74 vs. 46.72) → weakly but positively related to F00–F28.

D. Comparison with GA Feature Selection to PSO Feature selection

The comparison (**Table VII**) examines GA and PSO for feature selection on the CKD dataset, focusing on their computational complexity, accuracy, and interpretability differences. GA selected 16 out of 30 features, reducing the feature space by 46.67%, while PSO selected 17 with a 43.33% reduction. GA is slightly better for dimensionality reduction, and both methods retain essential attributes. GA focuses on key biomarkers like Age and Albumin, while PSO emphasizes Age and Sodium. PSO is faster and more accurate for large datasets, while GA is better for feature reduction.

TABLE VII
COMPARISON OF GA AND PSO FEATURE SELECTION

Aspect	GA	PSO
Number of Features	16 out of 30 attributes	17 out of 30 attributes
Reduction %	46.67%	43.33%
Top Features	Age, Albumin, Hypertension	Age, Sodium, Lymphocytes
Convergence Speed	Moderate	Faster
Accuracy	High	Slightly Higher

E. ML Models Analysis for the GA CKD Dataset Features

The confusion matrices (**Table VIII**) for four ML models (Logistic Regression, Naive Bayes, KNN, and SVM-RBF) provide insights into their performance in predicting CKD

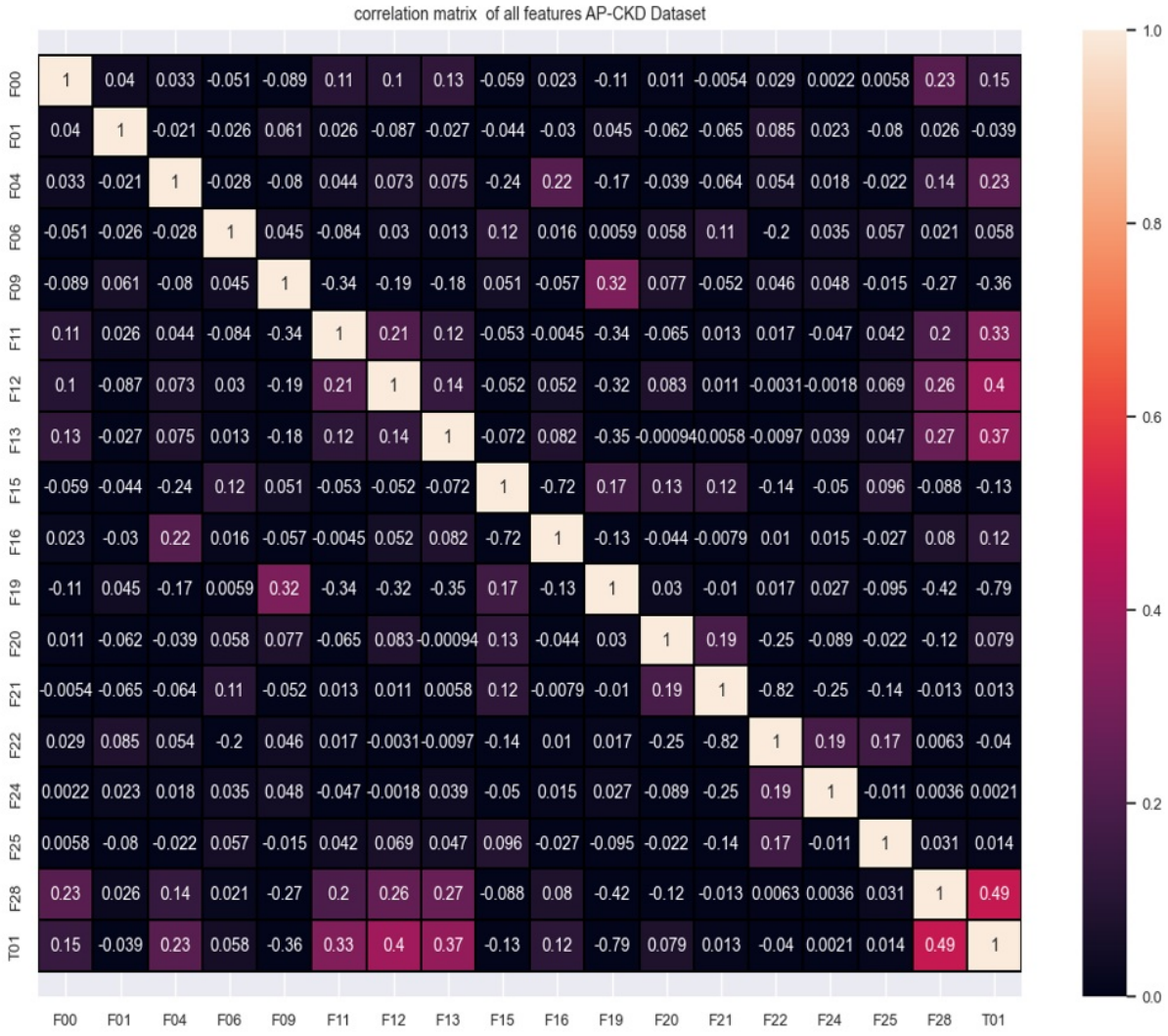

Fig. 5. PSO Features Heat-Map (Corelation Matrix)

(Class-1) and Non-CKD (Class-0) cases. The analysis shows that Logistic Regression correctly predicted 114 non-CKD cases out of 122 actual non-CKD cases, with eight non-CKD cases wrongly classified as CKD. The model correctly predicted 214 out of 223 cases for CKD cases, resulting in a True Positive Rate (TPR) (Recall) of 95.96%. Naive Bayes correctly predicted 112 non-CKD cases out of 122 actual cases, with 10 non-CKD cases wrongly classified as CKD. It correctly predicted 219 out of 223 CKD cases for CKD, yielding a TPR (Recall) of 98.21%. Only 4 CKD cases were missed, leading to an FNR of 1.79%. key Insights: Accuracy: 95.94%, slightly better than LR. **Fig. 8** shows the performance analysis of ML models trained on GA-selected CKD dataset features, highlighting the effectiveness of feature selection in improving model outcomes. K-Nearest Neighbours (KNN) correctly predicted 108 non-CKD cases out of 122, with an FPR of 11.48%. However, 14 non-CKD cases were misclassified as CKD. For CKD, 182 out of 223 cases were correctly predicted, resulting in a TPR (Recall) of 81.61%. A significant 41 CKD cases were misclassified as non-CKD, leading to a FNR of 18.39%. SVM-RBF correctly predicted only 4 non-CKD cases out of 122, resulting in an FPR of 96.72%. 118 non-CKD cases were misclassified as CKD. The model correctly predicted 219 out of 223 cases for CKD, yielding a TPR (Recall) of 98.21%.

TABLE VIII
CONFUSION MATRIX ANALYSIS FOR GA CKD DATASET FEATURES

Model	TNs(non-CKD)	FPs(non-CKD)	TNs(CKD)	FNs(CKD)
LR	114	8	214	9
Naive Bayes	112	10	219	4
KNN	108	14	182	41
SVM (RBF)	4	118	219	4

F. ML Models Analysis for the PSO CKD Dataset Features

The confusion matrices (Table IX) for four ML models (Logistic Regression, Naive Bayes, KNN, and SVM-RBF) provide insights into their performance in predicting CKD (Class-1) and Non-CKD (Class-0) cases. The analysis shows that Logistic Regression correctly predicted 115 non-CKD cases out of 122 actual non-CKD cases, with eight non-CKD cases wrongly classified as CKD. The model correctly predicted 214 out of 223 cases for CKD cases, resulting in a True Positive Rate (TPR) (Recall) of 95.96%. Naive Bayes correctly predicted 112 non-CKD cases out of 122 actual cases, with 10 non-CKD cases wrongly classified as CKD. It correctly predicted 219 out of 223 CKD cases

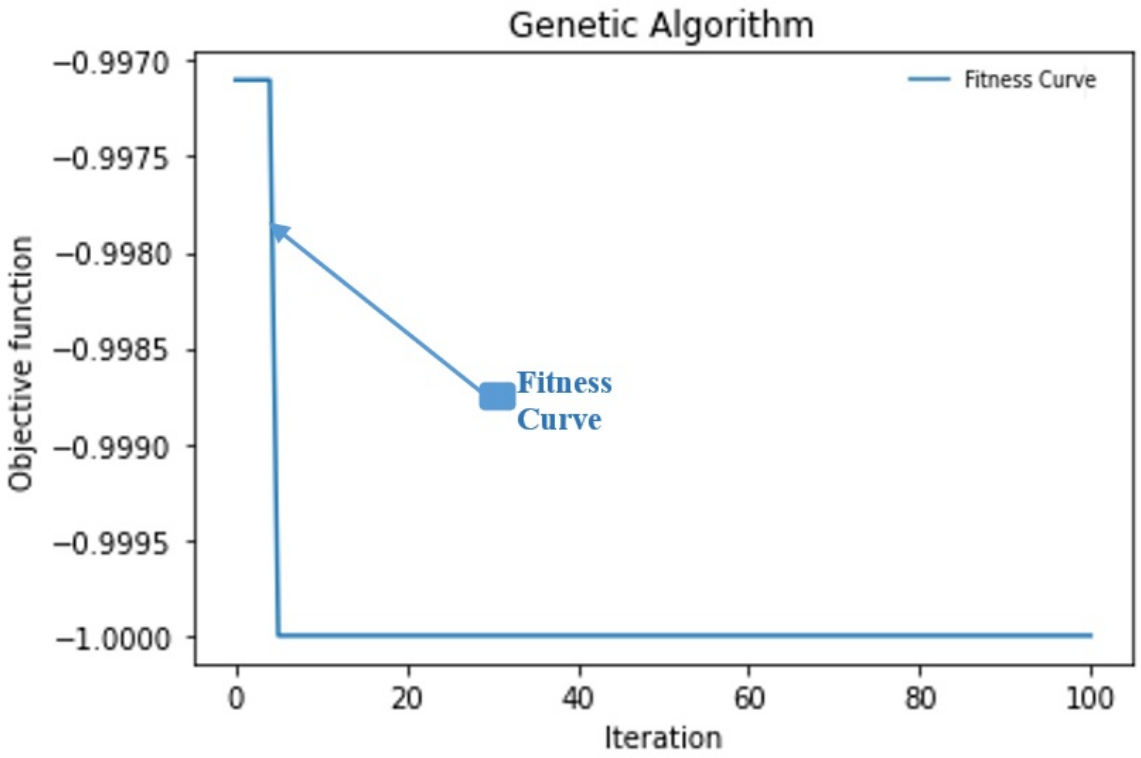

Fig. 6. GA Algorithm Selected Features from CKD Dataset

for CKD, yielding a TPR (Recall) of 98.21%. Only 4 CKD cases were missed, leading to an FNR of 1.79%—key Insights: Accuracy: 95.94%, slightly better than LR.

TABLE IX
CONFUSION MATRIX ANALYSIS FOR PSO CKD DATASET FEATURES

Model	TNs(non-CKD)	FPs(non-CKD)	TNs(CKD)	FNs(CKD)
LR	115	7	214	9
Naive Bayes	112	10	219	4
KNN	108	14	182	41
SVM (RBF)	4	118	219	4

TABLE X
PERFORMANCE PARAMETERS ANALYSIS FOR PSO CKD DATASET FEATURES

ML Model	AUC	CA	F1 Score	Precision	Recall
LR Model	0.9867	0.9507	0.9045	0.9641	0.9596
Naive Bayes	0.9927	0.9594	0.9691	0.9563	0.9821
KNN	0.9121	0.8406	0.8687	0.9286	0.8161
SVM (RBF)	0.6804	0.6464	0.7822	0.6499	0.9821

Table X presents the performance parameters of ML models

TABLE XI
CONFUSION MATRIX ANALYSIS FOR GA CKD DATASET FEATURES

Model	TNs(non-CKD)	FPs(non-CKD)	TNs(CKD)	FNs(CKD)
MLP model	118	4	215	8
MLP+PSO	121	1	218	5
MLP+FF	120	2	218	5

using PSO-selected CKD dataset features, comparing metrics such as AUC, CA, F1-score, precision, and recall. The study evaluated four ML models on the CKD dataset using five key performance metrics: AUC, Accuracy, F1 Score, Precision, and Recall. Naive Bayes and Logistic Regression demonstrated excellent performance distinguishing CKD from non-CKD cases, while KNN and SVM struggled with complex decision boundaries. Classification Accuracy (CA) showed Naive Bayes and Logistic Regression had excellent overall accuracy, while SVM had significant limitations in prediction reliability. The F1 Score balanced Precision and Recall, with Naive Bayes and Logistic Regression delivering the most balanced performance. Precision-measured the accuracy of optimistic predictions, with Logistic Regression achieving the highest Precision of 96.41%. Naive Bayes and KNN had relatively good Precision, while SVM (RBF) performed poorly, with a Precision of 64.99%. Recall showed Naive Bayes and SVM had excellent performance, with Logistic Regression showing a strong recall of 95.96%. **Fig. 9** shows ROC-AUC analysis of ML models on GA-selected CKD dataset features, revealing Naive Bayes and Logistic Regression's superior performance in classification tasks.

G. Confusion matrix Analysis for PSO CKD Dataset Features Versus to GA CKD Dataset Features

The confusion matrix (Table IX) analysis shows the performance of various models in classifying CKD and non-CKD cases. The Logistic Regression (LR) model achieved a high True Positives count and a low False Positives count, indicating reliable performance in identifying non-CKD cases. The Naive Bayes model showed high sensitivity towards CKD detection, while the KNN model displayed weaknesses with a high False Negative count and frequent misclassification of CKD cases. The SVM model performed poorly, with a TN count of 4 and a substantial FP count of 118. The PSO-selected features slightly enhanced the LR model's performance, reducing

Fig. 7. GA Features Heat-Map (Corelation Matrix)

False Positives from 8 to 7 while maintaining consistent True Positives and False Negatives. The Naive Bayes model maintained its robustness, while the KNN model struggled with a high False Negative count and inability to classify CKD and non-CKD patients effectively. **Fig. 10** presents the

TABLE XII
PERFORMANCE PARAMETERS ANALYSIS FOR GA CKD DATASET FEATURES

Model	AUC	CA	F1	Precision	Recall
MLP model	0.9929	0.9653	0.9654	0.9657	0.9653
MLP+PSO	0.9963	0.9826	0.9862	0.9847	0.9821
MLP+FF	0.9988	0.9798	0.9798	0.9798	0.9798

performance analysis of ML models for the PSO-selected CKD dataset features. The study evaluated ML models' performance across GA and PSO-selected features. Naive Bayes emerged as the top performer, with an AUC of 0.9927 and a Recall of 0.9821, indicating excellent ability to distinguish between CKD and non-CKD patients. The Logistic Regression model followed closely, with an AUC of 0.9847 and a Precision score of 0.9683. K-Nearest Neighbors (KNN) displayed moderate performance, while SVM (RBF) underperformed significantly. Naive Bayes remained the strongest contender for PSO-selected features, with an AUC of 0.9927 and a Recall of 0.9821. The Logistic Regression model also performed exceptionally well, achieving an AUC of 0.9867. KNN showed moderate results, while SVM failed to deliver satisfactory results. PSO showed marginal improvements in Logistic Regression, specifically in reducing false positives. The research examines how different MLP models work to find chronic kidney disease (CKD). It compares the basic Multilayer Perceptron (MLP) model with improved versions that use techniques called Particle Swarm Optimization (PSO) and the Firefly Algorithm. The study emphasizes AUC, accuracy, F1 score, precision, and recall measures.

H. Performance Parameters Analysis for CKD Dataset GA Selected Features for Optimized MLPs

Fig. 11 shows the performance evaluation of improved MLP models for the GA CKD dataset characteristics, contrasting the regular MLP model, the MLP augmented with PSO, and the MLP integrated with FF. The research examines how different MLP models work to find chronic kidney disease (CKD) using selected . It compares the basic Multilayer Perceptron (MLP) model with improved versions that use techniques called Particle Swarm Optimization (PSO) and the Firefly Algorithm. The study emphasizes

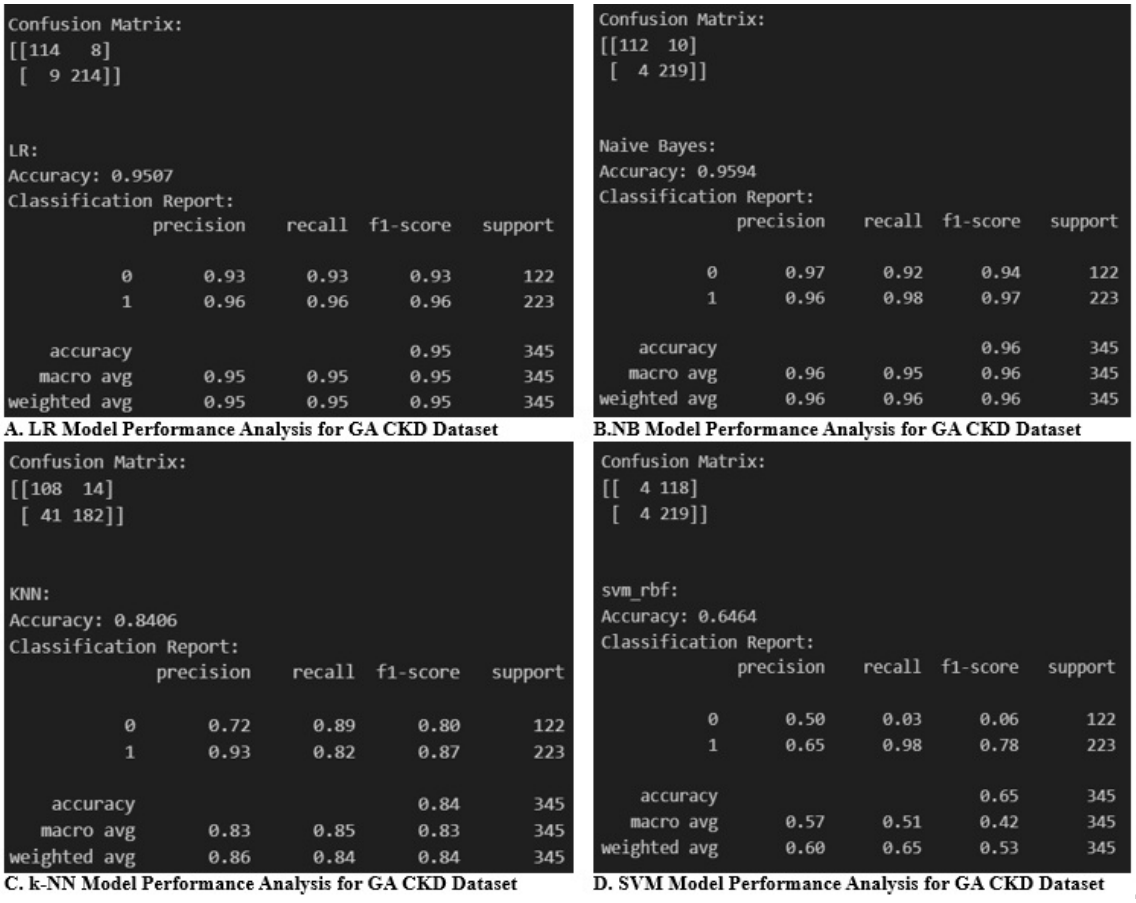

Fig. 8. ML Models Performance Analysis for the GA CKD Dataset Features

AUC, accuracy, F1 score, precision, and recall measures. The MLP model for the GA CKD dataset performed very well. It had an AUC of 0.9867. The accuracy was 96.53%. The precision, F1 score, and recall were all 96.5%. The confusion matrix (Table XI) and performance analysis (Table XII) showed 118 true negatives and 215 true positives. Only four false positives and eight false negatives indicated balanced detection of both CKD and non-CKD cases. The MLP model with PSO optimization improved performance in the GA CKD dataset, achieving the highest AUC of 0.9927, indicating excellent discrimination between CKD and non-CKD cases. Accuracy increased to 98.26%, and both F1 score, and precision were close to 98.6%, indicating superior model reliability. The model correctly identified 121 true negatives and 218 true positives, with only one false positive and five false negatives, showcasing the effectiveness of PSO optimization in minimizing errors and enhancing predictive capabilities. The Firefly Algorithm optimized MLP (MLP+FF) demonstrated robust performance with an accuracy of 97.98% and consistent precision, recall, and F1 scores at 97.98%. **Fig. 12** shows ROC-AUC analysis of Optimized MLP models, including PSO and FF-based optimization, on GA CKD dataset features, demonstrating their high classification performance.

I. Performance Parameters Analysis for CKD Dataset PSO Selected Features for Optimized MLPs

The Study (**Fig. 13**) compares the performance of Multi-Layer Perceptron (MLP) models using PSO selected features AP-CKD Dataset. In categorizing chronic kidney disease (CKD) and non-CKD patients using optimization approaches, comparing the baseline model to two improved versions: MLP+PSO and MLP+FF. The MLP+PSO model outperforms the baseline MLP and MLP+FF in identifying non-CKD and CKD cases, with only one misclassification (Table XIII). It also performs best in identifying 220 instances with only three missed cases, while MLP+FF shows slightly less accuracy, missing four CKD cases. The reduced false positive and false negative counts demonstrate its superior classification reliability. The MLP+PSO model outperforms the baseline MLP and MLP+FF in terms of accuracy, precision, and recall. It achieves the highest AUC

TABLE XIII
CONFUSION MATRIX ANALYSIS FOR PSO CKD DATASET FEATURES

Model	TNs(non-CKD)	FPs(non-CKD)	TNs(CKD)	FNs(CKD)
MLP model	119	3	218	5
MLP+PSO	121	1	220	3
MLP+FF	119	3	219	4

TABLE XIV
OPTIMIZED MLP USING PSO AND FF MODELS PERFORMANCE ANALYSIS FOR THE PSO CKD DATASET FEATURES

Model	AUC	CA	F1	Precision	Recall
MLP model	0.9985	0.9779	0.9776	0.9774	0.9778
MLP+PSO	0.9989	0.9901	0.9902	0.9903	0.9901
MLP+FF	0.9981	0.9804	0.9804	0.9801	0.9807

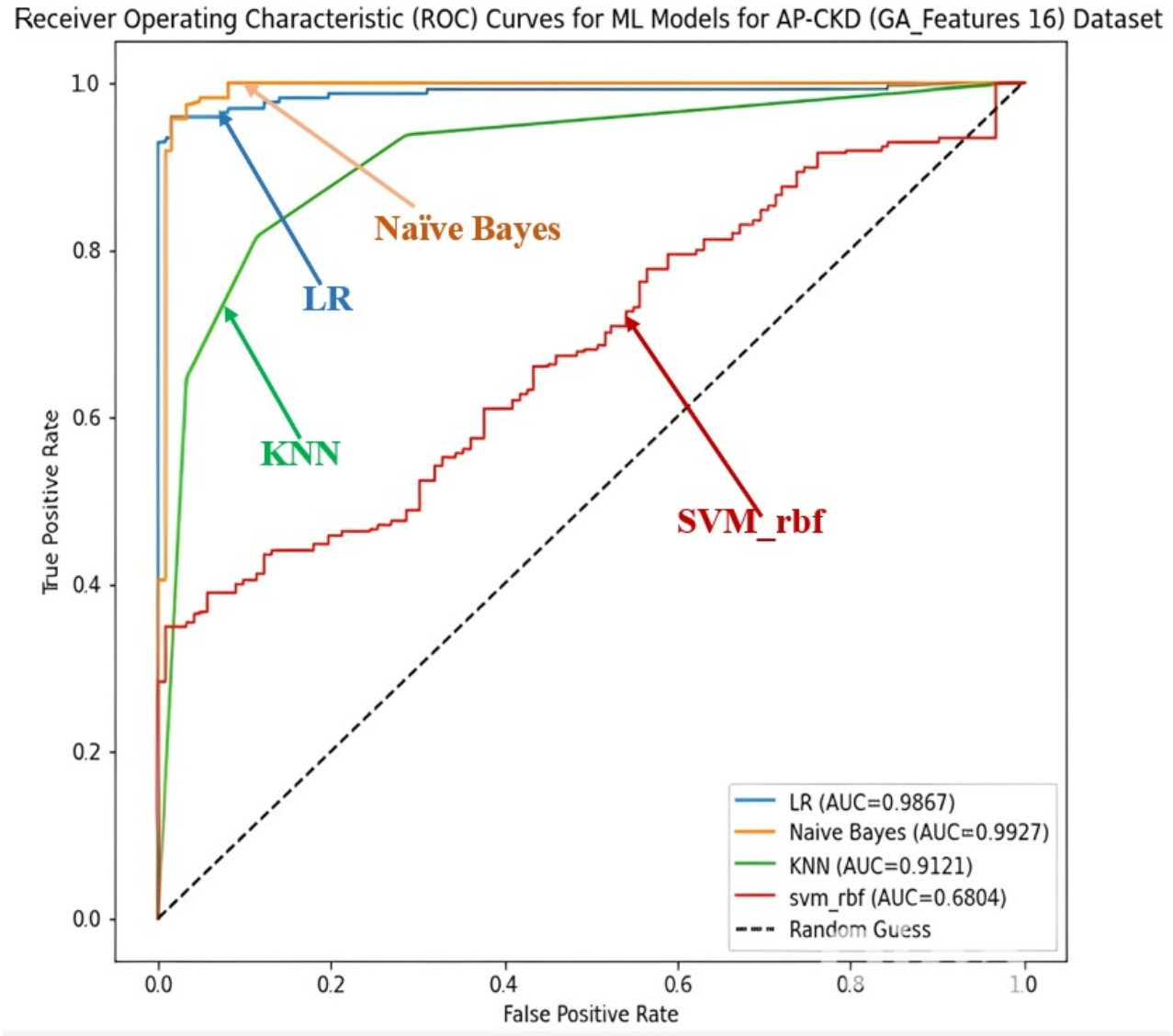

Fig. 9. ML Models ROC-AUC Analysis for the GA CKD Dataset Features

(0.9989), with a near-perfect AUC value indicating excellent model discrimination ability. MLP+PSO outperforms the baseline MLP with 99.01% accuracy, while MLP+FF improves with 98.04% accuracy. It also demonstrates the best balance between precision and recall, with an F1 score of 0.9902, leading in both precision (0.9903) and recall (0.9901), indicating fewer false positives and false negatives. The study compares the performance of Multi-Layer Perceptron (MLP) models using PSO selected features AP-CKD Dataset. In categorizing chronic kidney disease (CKD) and non-CKD patients using optimization approaches, comparing the baseline model to two improved versions: MLP+PSO and MLP+FF. The MLP+PSO model outperforms the baseline MLP and MLP+FF in identifying non-CKD and CKD cases, with only one misclassification. It also performs best in identifying 220 instances with only three missed cases, while MLP+FF shows slightly less accuracy, missing four CKD cases. The reduced false positive and false negative counts demonstrate its superior classification reliability. The MLP+PSO model outperforms the baseline MLP and MLP+FF in terms of accuracy, precision, and recall. It achieves the highest AUC (0.9989), with a near-perfect AUC value indicating excellent model discrimination ability. MLP+PSO outperforms the baseline MLP with 99.01% accuracy, while MLP+FF improves with 98.04% accuracy. It also demonstrates the best balance between precision and recall, with an F1 score of 0.9902, leading in both precision (0.9903) and recall (0.9901), indicating fewer false positives and false negatives. Table XIV evaluates the effectiveness of feature selection via PSO Algorithms and optimization techniques like PSO and FF in enhancing MLP-based CKD classification. The baseline MLP model shows strong classification performance, but higher false positive and false negative rates indicate room for improvement. MLP+PSO delivers the best overall performance, achieving near-perfect metrics across all evaluation parameters. MLP+FF significantly enhances the baseline model's performance, improving accuracy (98.04%) and achieving an AUC of 0.9981. However, its slightly higher FN count indicates marginally less reliability in identifying CKD cases. **Fig. 14** shows the ROC-AUC analysis of Optimized MLP models on the PSO CKD dataset, highlighting their classification efficiency and effective class distinction.

```
Confusion Matrix:
[[115   7]
 [  9 214]]


LR:
Accuracy: 0.9536
Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.94      0.93       122
           1       0.97      0.96      0.96       223

    accuracy                           0.95       345
   macro avg       0.95      0.95      0.95       345
weighted avg       0.95      0.95      0.95       345
```

A. LR Model Performance Analysis for PSO CKD Dataset

```
Confusion Matrix:
[[112  10]
 [  4 219]]


Naive Bayes:
Accuracy: 0.9594
Classification Report:
              precision    recall  f1-score   support

           0       0.97      0.92      0.94       122
           1       0.96      0.98      0.97       223

    accuracy                           0.96       345
   macro avg       0.96      0.95      0.96       345
weighted avg       0.96      0.96      0.96       345
```

B. NB Model Performance Analysis for PSO CKD Dataset

```
Confusion Matrix:
[[108  14]
 [ 41 182]]


KNN:
Accuracy: 0.8406
Classification Report:
              precision    recall  f1-score   support

           0       0.72      0.89      0.80       122
           1       0.93      0.82      0.87       223

    accuracy                           0.84       345
   macro avg       0.83      0.85      0.83       345
weighted avg       0.86      0.84      0.84       345
```

C. k-NN Model Performance Analysis for PSO CKD Dataset

```
Confusion Matrix:
[[  4 118]
 [  4 219]]


svm_rbf:
Accuracy: 0.6464
Classification Report:
              precision    recall  f1-score   support

           0       0.50      0.03      0.06       122
           1       0.65      0.98      0.78       223

    accuracy                           0.65       345
   macro avg       0.57      0.51      0.42       345
weighted avg       0.60      0.65      0.53       345
```

D. SVM Model Performance Analysis for PSO CKD Dataset

Fig. 10. ML Models Performance Analysis for the PSO CKD Dataset Features

```
Confusion Matrix for MLP:
[[118   4]
 [  8 215]]

Classification Report for MLP:
              precision    recall  f1-score   support

           0       0.94      0.97      0.95       122
           1       0.98      0.96      0.97       223

    accuracy                           0.97       345
   macro avg       0.96      0.97      0.96       345
weighted avg       0.97      0.97      0.97       345
```

A. MLP Model Performance Analysis for GA CKD Dataset

```
Confusion Matrix for MLP+PSO:
[[121   1]
 [  5 218]]

Classification Report for MLP+PSO:
              precision    recall  f1-score   support

           0       0.96      0.99      0.98       122
           1       1.00      0.98      0.99       223

    accuracy                           0.98       345
   macro avg       0.98      0.98      0.98       345
weighted avg       0.98      0.98      0.98       345
```

B. MLP + PSO Model Performance Analysis for GA CKD Dataset

```
Confusion Matrix for MLP+FF:
[[120   2]
 [  5 218]]

Classification Report for MLP+FF:
              precision    recall  f1-score   support

           0       0.96      0.98      0.97       122
           1       0.99      0.98      0.98       223

    accuracy                           0.98       345
   macro avg       0.98      0.98      0.98       345
weighted avg       0.98      0.98      0.98       345
```

C. MLP + FF Model Performance Analysis for GA CKD Dataset

Fig. 11. Optimized MLP Models Performance Analysis for the GA CKD Dataset Features

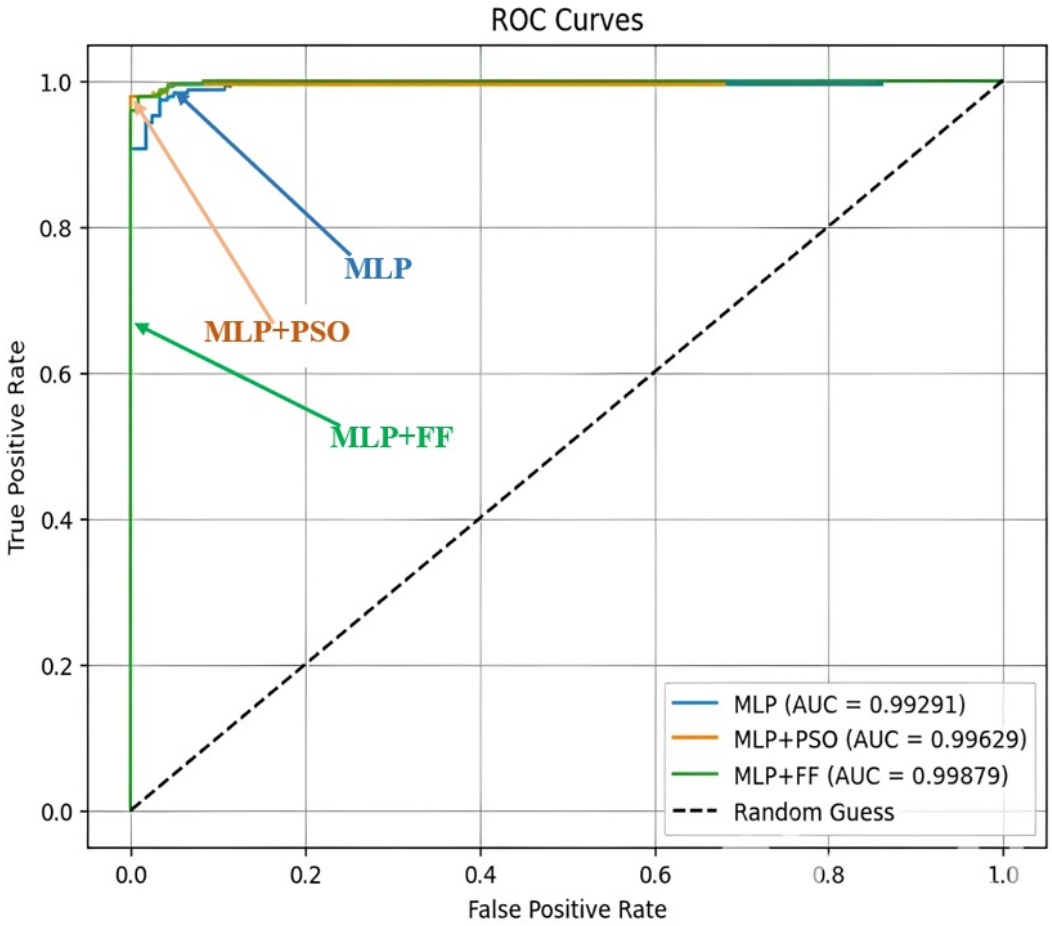

Fig. 12. Optimized MLP by PSO and FF Models ROC-AUC analysis for the GA CKD Dataset Features

```
Confusion Matrix for MLP:
[[118   4]
 [  8 215]]

Classification Report for MLP:
              precision    recall  f1-score   support

           0       0.94      0.97      0.95       122
           1       0.98      0.96      0.97       223

    accuracy                           0.97       345
   macro avg       0.96      0.97      0.96       345
weighted avg       0.97      0.97      0.97       345
```

A. MLP Model Performance Analysis for GA CKD Dataset

```
Confusion Matrix for MLP+PSO:
[[121   1]
 [  5 218]]

Classification Report for MLP+PSO:
              precision    recall  f1-score   support

           0       0.96      0.99      0.98       122
           1       1.00      0.98      0.99       223

    accuracy                           0.98       345
   macro avg       0.98      0.98      0.98       345
weighted avg       0.98      0.98      0.98       345
```

B. MLP + PSO Model Performance Analysis for GA CKD Dataset

```
Confusion Matrix for MLP+FF:
[[120   2]
 [  5 218]]

Classification Report for MLP+FF:
              precision    recall  f1-score   support

           0       0.96      0.98      0.97       122
           1       0.99      0.98      0.98       223

    accuracy                           0.98       345
   macro avg       0.98      0.98      0.98       345
weighted avg       0.98      0.98      0.98       345
```

C. MLP + FF Model Performance Analysis for GA CKD Dataset

Fig. 13. Optimized MLP Models Performance Analysis for the PSO CKD Dataset Features

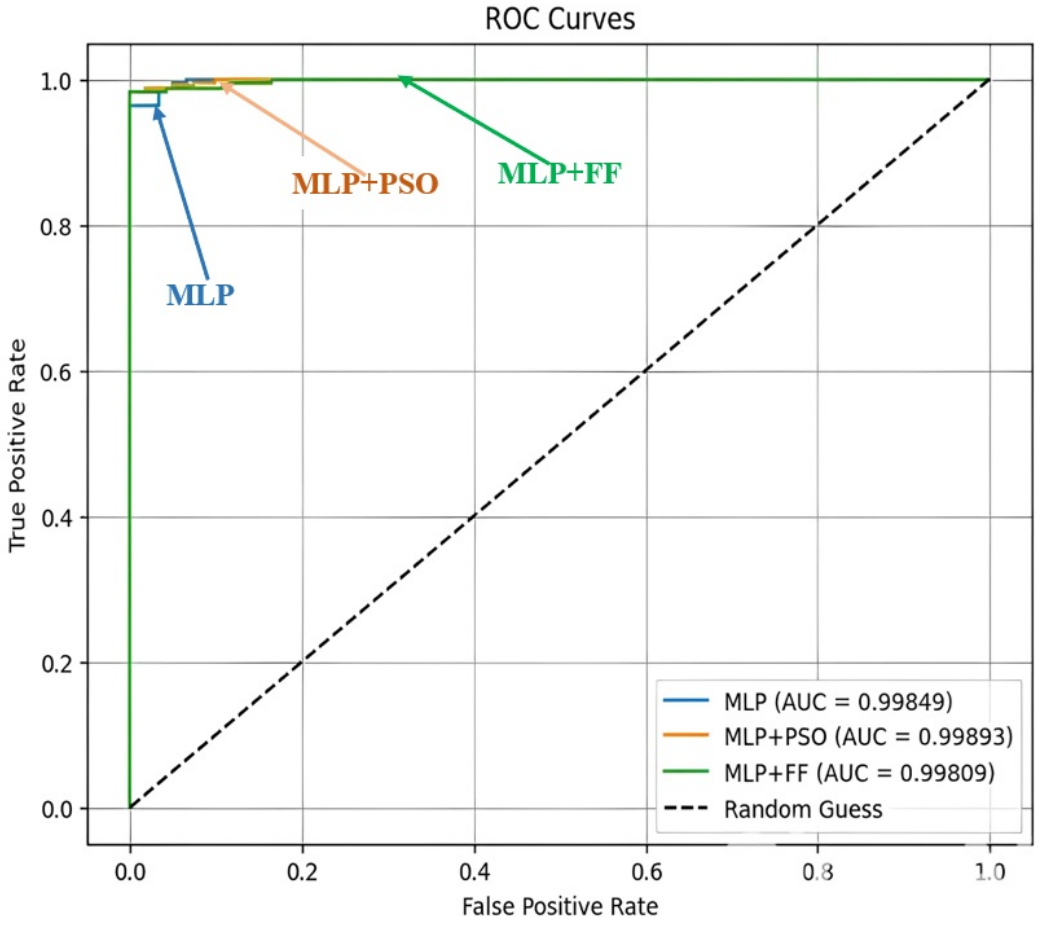

Fig. 14. Optimized MLP using PSO and FF Models ROC-AUC Analysis for the PSO CKD Dataset Features

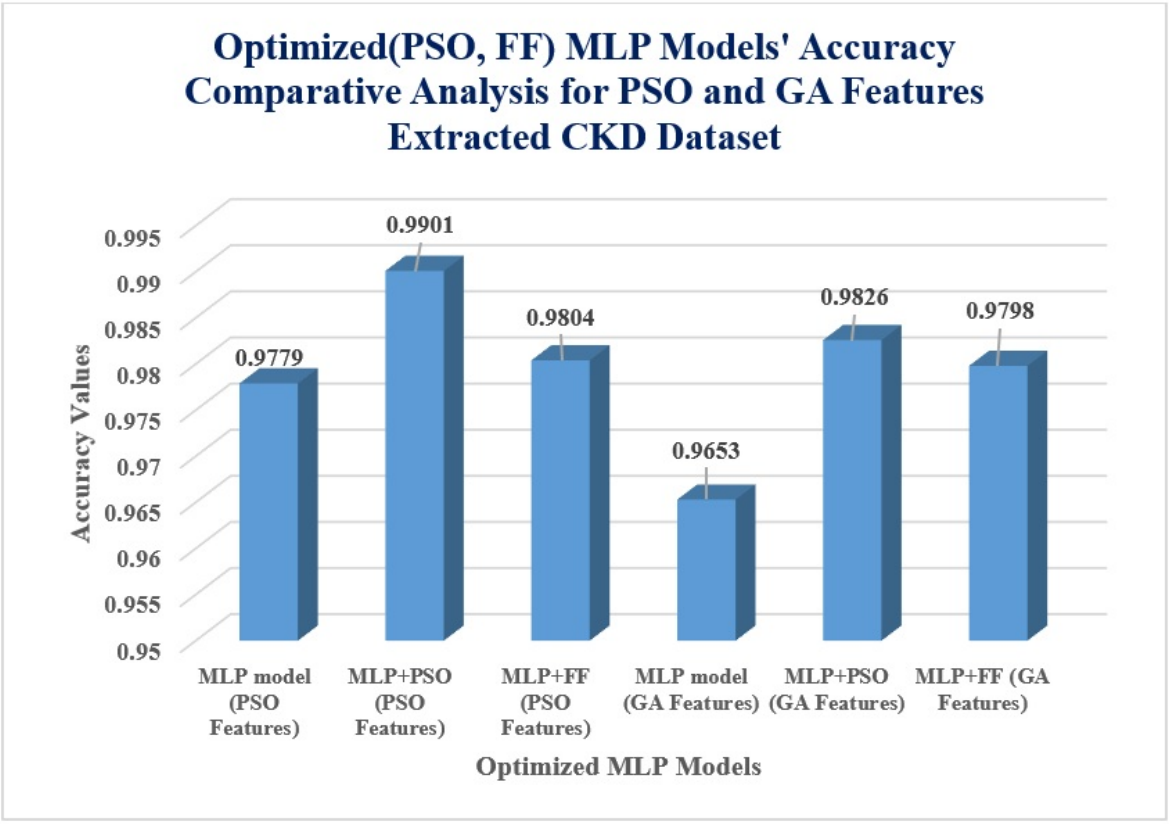

Fig. 15. Comparative Analysis on Classification Accuracy

TABLE XV
SUMMARY OF RELATED WORKS ON CKD PREDICTION MODELS

Ref. No.	Author(s) (Year)	Description and Models	Result Analysis
[26]	Rashed-Al-Mahfuz (2021)	Developed ML models for CKD diagnosis using optimized datasets and classifiers (RF, GB, XGB, LR, SVM) with k-fold cross-validation.	RF achieved 97.75% accuracy on 'DB-II' dataset, sensitivity 96.12%, specificity 98.82%, F1-score 96.88%.
[35]	Saif et al. (2024)	Developed a framework for early CKD prediction using DL and ensemble methods (CNN, LSTM, LSTM-BLSTM) with Adam/Adamax optimizers.	Ensemble model reached 98% accuracy (six months) and 97% (twelve months), improving early diagnosis.
[36]	Ramu et al. (2025)	Proposed a hybrid CNN-SVM model for early CKD detection using 10 medical indicators. CNN extracts features, SVM classifies.	Hybrid model achieved 96.8% accuracy, outperforming standalone SVM (94.8%) and RF (94.6%).
[37]	Rehman et al. (2023)	Introduced a hybrid CKD prediction model using LR, odds ratio analysis, and MRI features with 5-fold cross-validation.	LR outperformed LDA, MLP, identifying key features (serum creatinine, albumin, diabetes), achieving 98.5% train and 97.5% test accuracy.
[38]	Ashafuddula et al. (2023)	Developed a fully automated ML method for early-stage CKD prediction using ensemble classifiers (AdaBoost, LR, Passive Aggressive).	Achieved 96.48% accuracy on Bangladeshi CKD data, reducing prediction time using FS and FSR techniques.
Present Study	**Our Study (2025)**	CKD Detection in North-Coastal Andhra Pradesh, India. Uses clinical data, optimized multi-level perceptron models with PSO and GA for feature selection.	**MLP+PSO outperforms standalone MLP and MLP+FF models with AUC 0.9989 and CA 99.01%.**

V. DISCUSSIONS

In this section assesses the effectiveness of optimized MLP models for CKD detection, comparing them with existing methods and highlighting the MLP+PSO model's superior accuracy and reliability. Comparing MLP models with existing methods. Highlighting MLP+PSO model's superior accuracy and reliability.

A. Comparative Analysis of Classification Accuracy for MLP Models Using GA and PSO Selected Features

The analysis compares the **Fig. 15**. Classification Accuracy (CA) of different MLP-based ML models on a CKD dataset using Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) feature selection methods. The MLP model with GA features achieves a CA of 96.53%, but its accuracy lags those enhanced by optimization techniques. The MLP+PSO model achieves the highest accuracy (98.26%) among those using GA features, demonstrating the effectiveness of PSO in fine-tuning MLP weights. The MLP+FF model outperforms the baseline model with a CA of 97.98%, highlighting the utility of the Firefly Algorithm. The MLP model with PSO features achieves a CA of 97.79%, suggesting that PSO-selected features are more informative for the MLP's classification tasks. The MLP+PSO model achieves the highest accuracy overall with a CA of 99.01%, showcasing its effectiveness in both phases. The MLP+FF model shows an improvement over the baseline and is competitive with MLP+PSO, but its performance is slightly inferior to MLP+PSO, indicating that PSO as an optimization strategy has a slight edge over FF. The study shows that PSO-selected features are better than GA-selected features. Models using PSO perform better in all strategies. The MLP+PSO model is stronger than MLP+FF models. The baseline MLP model has 97.79% accuracy with PSO features, while GA features yield 96.53%. The best model for CKD classification is MLP+PSO, achieving 99.01% accuracy. PSO is the best optimization strategy, followed by the Firefly Algorithm. Effective feature selection and optimization improve ML models for medical datasets.

B. Comparative Analysis Existing work and Proposal model

The research highlights advancements in CKD detection methods, including ML, hybrid approaches, and ensemble classifiers, with high accuracy rates, using techniques like CNN-SVM, logistic regression, and feature selection. The study compares CKD detection methods using DL, hybrid approaches, and ensemble classifiers, achieving high accuracies. The MLP+PSO model outperforms others with a CA of 99.01% and AUC of 0.9989, demonstrating superior performance in CKD detection tailored to specific datasets **(Table XV)**.

VI. CONCLUSION

The study presents a robust and efficient method for early detection of CKD using optimized MLP models and advanced feature selection techniques, focusing on clinical data from north-coastal districts of Andhra Pradesh, India. Particle Swarm Optimization (PSO) and Genetic Algorithm were found to significantly reduce feature space in CKD diagnosis models, retaining key biomarkers. PSO-selected features were more informative, leading to superior model performance. The MLP+PSO model achieved the highest

Classification Accuracy (CA) of 99.01%, outperforming other optimization strategies like Firefly Algorithm and GA. The study shows that PSO is effective for feature selection and weight fine-tuning in medical diagnostic models. This improves the accuracy and reliability of CKD detection. It also supports precision medicine by offering a scalable framework for other medical datasets. Future work can build on this study. It can explore hybrid optimization techniques. Researchers should validate the model on larger and more diverse datasets. This will help improve diagnostic accuracy and generalizability.

REFERENCES

[1] H. A. Fink, A. Ishani, B. C. Taylor, N. L. Greer, R. MacDonald, D. Rossini, and T. J. Wilt, "Screening for, monitoring, and treatment of chronic kidney disease stages 1 to 3: A systematic review for the us preventive services task force and for an american college of physicians clinical practice guideline," *Annals of Internal Medicine*, vol. 156, no. 8, pp. 570–581, 2012.

[2] I. Soveri, U. B. Berg, J. Björk, C. G. Elinder, A. Grubb, I. Mejare, and S. G. R. Group, "Measuring gfr: A systematic review," *American Journal of Kidney Diseases*, vol. 64, no. 3, pp. 411–424, 2014.

[3] Y. C. Ma, L. Zuo, C. L. Zhang, M. Wang, R. F. Wang, and H. Y. Wang, "Comparison of 99mtc-dtpa renal dynamic imaging with modified mdrd equation for glomerular filtration rate estimation in chinese patients in different stages of chronic kidney disease," *Nephrology Dialysis Transplantation*, vol. 22, no. 2, pp. 417–423, 2007.

[4] P. Chittora, S. Chaurasia, P. Chakrabarti, G. Kumawat, T. Chakrabarti, Z. Leonowicz, and V. Bolshev, "Prediction of chronic kidney disease-a machine learning perspective," *IEEE Access*, vol. 9, pp. 17 312–17 334, 2021.

[5] D. A. Debal and T. M. Sitote, "Chronic kidney disease prediction using machine learning techniques," *Journal of Big Data*, vol. 9, no. 1, p. 109, 2022.

[6] M. A. Islam, M. Z. H. Majumder, and M. A. Hussein, "Chronic kidney disease prediction based on machine learning algorithms," *Journal of Pathology Informatics*, vol. 14, p. 100189, 2023.

[7] S. A. Alsuhibany, S. Abdel-Khalek, A. Algarni, A. Fayomi, D. Gupta, V. Kumar, and R. F. Mansour, "Ensemble of deep learning-based clinical decision support system for chronic kidney disease diagnosis in medical internet of things environment," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, p. 4931450, 2021.

[8] N. H. Chowdhury, M. B. I. Reaz, F. Haque, S. Ahmad, S. H. M. Ali, A. A. A Bakar, and M. A. S. Bhuiyan, "Performance analysis of conventional machine learning algorithms for identification of chronic kidney disease in type 1 diabetes mellitus patients," *Diagnostics*, vol. 11, no. 12, p. 2267, 2021.

[9] V. Srikanth, "Chronic kidney disease prediction using machine learning algorithms," pp. 106–109, 2023.

[10] B. P. Ghosh, T. Imam, N. Anjum, M. T. Mia, C. U. Siddiqua, K. S. Sharif, and M. Z. Hossain, "Advancing chronic kidney disease prediction: Comparative analysis of machine learning algorithms and a hybrid model," *Journal of Computer Science and Technology Studies*, vol. 6, no. 3, pp. 15–21, 2024.

[11] C. P. Kovesdy, "Epidemiology of chronic kidney disease: an update 2022," *Kidney International Supplements*, vol. 12, no. 1, pp. 7–11, 2022.

[12] A. Francis, M. N. Harhay, A. Ong, S. L. Tummalapalli, A. Ortiz, A. B. Fogo, and V. Jha, "Chronic kidney disease and the global public health agenda: an international consensus," *Nature Reviews Nephrology*, pp. 1–13, 2024.

[13] V. A. Luyckx, Z. Al-Aly, A. K. Bello, E. Bellorin-Font, R. G. Carlini, J. Fabian, and J. Stanifer, "Sustainable development goals relevant to kidney health: an update on progress," *Nature Reviews Nephrology*, vol. 17, no. 1, pp. 15–32, 2021.

[14] V. A. Luyckx, M. Tonelli, and J. W. Stanifer, "The global burden of kidney disease and the sustainable development goals," *Bulletin of the World Health Organization*, vol. 96, no. 6, p. 414, 2018.

[15] Y. M. K. Farag, K. Karai Subramanian, V. A. Singh, R. R. Tatapudi, and A. K. Singh, "Occupational risk factors for chronic kidney disease in andhra pradesh: 'uddanam nephropathy'," *Renal Failure*, vol. 42, no. 1, pp. 1032–1041, 2020.

[16] V. S. Prasad and M. Subrahmanyam, "Chronic kidney disease prevalent in small villages of north coastal, andhra pradesh, india, uddanam nephropathy," *Journal of Evidence-based Medicine and Healthcare*, vol. 2, no. 21, pp. 3205–3207, 2015.

[17] B. Gummidi, O. John, A. Ghosh, G. K. Modi, M. Sehgal, O. P. Kalra, V. Kher, J. Muliyil, J. S. Thakur, L. Ramakrishnan, C. M. Pandey, V. Sivakumar, R. S. Dhaliwal, T. Khanna, A. Kumari, G. Prasadini, J. C. Reddy, J. Reddy, and V. Jha, "A systematic study of the prevalence and risk factors of ckd in uddanam, india," *Kidney International Reports*, vol. 5, no. 12, pp. 2246–2255, 2020.

[18] K. B. Anusha, T. P. R. Vital, and K. Sangeeta, "Machine learning models and neural network techniques for predicting uddanam ckd," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 2, 2019.

[19] S. Dutta, R. Sikder, M. R. Islam, A. Al Mukaddim, M. A. Hider, and M. Nasiruddin, "Comparing the effectiveness of machine learning algorithms in early chronic kidney disease detection," *Journal of Computer Science and Technology Studies*, vol. 6, no. 4, pp. 77–91, 2024.

[20] D. Swain, U. Mehta, A. Bhatt, H. Patel, K. Patel, D. Mehta, and S. Manika, "A robust chronic kidney disease classifier using machine learning," *Electronics*, vol. 12, no. 1, p. 212, 2023.

[21] M. S. Arif, A. Mukheimer, and D. Asif, "Enhancing the early detection of chronic kidney disease: A robust machine learning model," *Big Data and Cognitive Computing*, vol. 7, no. 3, p. 144, 2023.

[22] Q. Bai, C. Su, W. Tang, and Y. Li, "Machine learning to predict end stage kidney disease in chronic kidney disease," *Scientific Reports*, vol. 12, no. 1, p. 8377, 2022.

[23] E. Dritsas and M. Trigka, "Machine learning techniques for chronic kidney disease risk prediction," *Big Data and Cognitive Computing*, vol. 6, no. 3, p. 98, 2022.

[24] H. Iftikhar, M. Khan, Z. Khan, F. Khan, H. M. Alshanbari, and Z. Ahmad, "A comparative analysis of machine learning models: A case study in predicting chronic kidney disease," *Sustainability*, vol. 15, no. 3, p. 2754, 2023.

[25] V. Singh, V. K. Asari, and R. Rajasekaran, "A deep neural network for early detection and prediction of chronic kidney disease," *Diagnostics*, vol. 12, no. 1, p. 116, 2022.

[26] M. Rashed-Al-Mahfuz, A. Haque, A. Azad, S. A. Alyami, J. M. Quinn, and M. A. Moni, "Clinically applicable machine learning approaches to identify attributes of chronic kidney disease (ckd) for use in low-cost diagnostic screening," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 9, pp. 1–11, 2021.

[27] K. Venkatrao and S. Kareemulla, "Hdlnet: A hybrid deep learning network model with intelligent iot for detection and classification of chronic kidney disease," *IEEE Access*, 2023.

[28] P. R. V. Terlapu, D. Jayaram, S. Rakesh, M. V. Gopalachari, B. V. Ramana, N. Tangudu, and K. R. Kalidindi, "Optimizing chronic kidney disease diagnosis in uddanam: A smart fusion of ga-mlp hybrid and pca dimensionality reduction," *Procedia Computer Science*, vol. 230, pp. 522–531, 2023.

[29] J. Deepika, P. Selvaraju, M. K. Thota, M. Tiwari, D. Venu, K. Manjulaadevi, and N. G. Lakshmi, "Efficient classification of kidney disease detection using heterogeneous modified artificial neural network and fruit fly optimization algorithm," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 31, no. 3, pp. 1–12, 2023.

[30] M. Hosseinzadeh, J. Koohpayehzadeh, A. O. Bali, P. Asghari, A. Souri, A. Mazaherinezhad, and R. Rawassizadeh, "A diagnostic prediction model for chronic kidney disease in internet of things platform," *Multimedia Tools and Applications*, vol. 80, pp. 16 933–16 950, 2021.

[31] P. Gogoi and J. A. Valan, "Machine learning approaches for predicting and diagnosing chronic kidney disease: Current trends, challenges, solutions, and future directions," *International Urology and Nephrology*, pp. 1–24, 2024.

[32] S. Akter, M. Ahmed, A. Al Imran, A. Habib, R. U. Haque, M. S. Rahman, and S. Mahjabeen, "Ckd.net: A novel deep learning hybrid model for effective, real-time, automated screening tool towards prediction of multi stages of ckd along with egfr and creatinine," *Expert Systems with Applications*, vol. 223, p. 119851, 2023.

[33] T. P. Vital, "Empirical study on uddanam chronic kidney diseases (uckd) with statistical and machine learning analysis including probabilistic neural networks," in *Handbook of Computational Intelligence in Biomedical Engineering and Healthcare*. Academic Press, 2021, pp. 283–314.

[34] S. Elkholy, A. Rezk, and A. A. E. F. Saleh, "Enhanced optimized classification model of chronic kidney disease," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, 2023.

[35] D. Saif, A. M. Sarhan, and N. M. Elshennawy, "Early prediction of chronic kidney disease based on ensemble of deep learning models and optimizers," *Journal of Electrical Systems and Information Technology*, vol. 11, no. 17, 2024.

[36] K. Ramu, S. Patthi, Y. N. Prajapati, J. V. N. Ramesh, S. Banerjee, K. B. Rao, and S. I. Alzahrani, "Hybrid cnn-svm model for enhanced early

detection of chronic kidney disease," *Biomedical Signal Processing and Control*, vol. 100, p. 107084, 2025.
[37] A. Rehman, T. Saba, H. Ali, N. Elhakim, and N. Ayesha, "Hybrid machine learning model to predict chronic kidney diseases using handcrafted features for early health rehabilitation," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 31, no. 6, pp. 951–968, 2023.
[38] N. I. Md. Ashafuddula, B. Islam, and R. Islam, "An intelligent diagnostic system to analyze early-stage chronic kidney disease for clinical application," *Applied Computational Intelligence and Soft Computing*, vol. 2023, no. 1, p. 3140270, 2023.

Ramesh Bandaru is a Research Scholar in the Department of Computer Science and Engineering at GITAM (Deemed to be University), Visakhapatnam, Andhra Pradesh, India. He obtained his Bachelor's degree in Computer Science and Engineering from Jawaharlal Nehru Technological University (JNTU), Hyderabad, Andhra Pradesh, India. He later completed his Master of Technology in Computer Science and Engineering from Aditya Institute of Technology and Management, Andhra Pradesh, India. His research interests include note, Deep Learning, Image Processing, and Web Technologies.

Dr. Karu Prasada Rao hold a B.Tech. degree in Computer Science and Engineering from GMRIT in 2004, and a Master's degree in Computer Science and Engineering from JNTU Kakinada (2009). I earned my Ph.D. in Computer Vision and Image Processing from Acharya Nagarjuna University, Guntur. With 20 years of teaching experience, I have published more than 20 articles in International and National journals. Currently, I am working as an Assistant Professor in the Department of Computer Science and Engineering at GITAM (Deemed To Be) University, Visakhapatnam, Andhra Pradesh, India – 530027.