Annals of Healthcare Systems Engineering

www.ahse.reapress.com

Ann. Healthc. Syst. Eng. Vol. 2, No. 2 (2025) 87–99.

Paper Type: Original Article

Development of a Multiprocessing Interface Genetic Algorithm for Optimising a Multilayer Perceptron for Disease Prediction

Iliyas Ibrahim Iliyas[1,*], Souley Boukari[2], Abdulsalam Ya'u Gital[2]

[1] Department of Computer Science, Faculty of Physical Sciences, University of Maiduguri, Borno State, Nigeria; iliyasibrahimiliyas@unimaid.edu.ng.

[2] Department of Computer Science, Faculty of Computing Science, Abubakar Tafawa Balewa University, Bauchi State, Nigeria; bsouley2001@yahoo.com; asgital@gmail.com.

Citation:

Received: 19 August 2024 *Revised: 06 January 2025* *Accepted: 14 March 2025*	Ibrahim Iliyas, I., Boukari, S., & Ya'u Gital, A. (2025). Development of a multiprocessing interface genetic algorithm for optimising a multilayer perceptron for disease prediction. *Annals of healthcare systems engineering*, *2*(2), 87-99.

Abstract

Accurate disease diagnosis enhances effective patient management; however, manual interpretation of complex biomedical data is time-consuming and vulnerable to error. Artificial Intelligence (AI) systems, particularly Machine Learning (ML) models, can automatically learn complex patterns from high-dimensional clinical and imaging data. The predictive performance of these methods depends critically on proper hyperparameter tuning. This study introduces a framework that integrates nonlinear feature extraction, classification, and efficient optimisation. First, kernel principal component analysis with a radial basis function kernel reduces dimensionality while preserving 95% of the variance. Second, a Multilayer Perceptron (MLP) learns to predict disease status. Finally, a modified Multiprocessing Interface Genetic Algorithm (MIGA) optimises MLP hyperparameters in parallel over ten generations. We evaluated this approach on three datasets: The Wisconsin Diagnostic Breast Cancer dataset, the Parkinson's Telemonitoring dataset, and the Chronic Kidney Disease (CKD) dataset. The MLP tuned by the MIGA achieved the best accuracy of 99.12% for breast cancer, 94.87% for Parkinson's Disease (PD), and 100% for CKD. These results outperform those of other methods, such as grid search, random search, and Bayesian optimisation. Compared to a standard Genetic Algorithm (GA), Kernel Principal Component Analysis (Kernel PCA) revealed nonlinear relationships that improved classification, and the MIGA's parallel fitness evaluations reduced the tuning time by approximately 60%. The GA incurs a high computational cost due to the sequential nature of fitness evaluations. Still, our MIGA parallelizes this step, significantly reducing the tuning time and steering the MLP toward the best accuracy scores of 99.12%, 94.87%, and 100% for breast cancer, Parkinson's, and CKD, respectively. The built-in graphical user interface then enables clinicians to load data, reduce dimensions, tune hyperparameters, and run predictions without writing code, paving the way for rapid and real-world adoption.

Keywords: Multilayer perceptron, Multiprocessing interface genetic algorithm, Hyperparameter optimization, Kernel principal component analysis, Parallel processing.

1 | Introduction

Disease diagnosis is essential for effective patient management and improved clinical outcomes. Clinicians traditionally rely on manual interpretation of laboratory tests and diagnostic imaging to detect diseases.

Corresponding Author: iliyasibrahimiliyas@unimaid.edu.ng

https://doi.org/10.22105/ahse.v2i2.36

However, this approach is time-consuming and often incapable of capturing the full complexity of high-dimensional biomedical data [1]. Artificial Intelligence (AI) is a branch of computer science that focuses on creating systems capable of performing tasks that typically require human intelligence. Machine Learning (ML) trains algorithms to learn from data and improve their performance without explicit programming [2]. By applying optimization techniques to tune model parameters, AI-driven frameworks automatically extract intricate features from clinical and imaging data, thereby enhancing the predictive accuracy and consistency for diseases [3]. However, the field of AI continues to evolve from rigid rule-based automation to systems that imitate human reasoning, perception, and foresight, allowing machines to take autonomous actions in complex environments. These advancements demonstrate that AI is not merely a tool for efficiency but also a transformative enabler that augments human capabilities across daily life and industry. Hyperparameter tuning is essential in ML because it directly affects model accuracy, generalization, and stability by determining how the learning algorithm behaves during training [4]. Proper tuning helps handle underfitting and overfitting by optimizing parameters or model complexity [5]. Without tuning, models often fail to reach their full predictive potential.

Traditional methods, such as grid search and random search, exhaustively evaluate all possible combinations, making them computationally expensive and inefficient in high-dimensional spaces —a problem known as the curse of dimensionality [6]. These limitations impose more adaptive methods, such as Genetic Algorithms (GAs), that can intelligently search for better hyperparameter combinations. The GA is increasingly used for hyperparameter tuning in ML due to its ability to explore large, complex, and nonlinear search spaces more efficiently than traditional methods, such as grid or random search. GAs mimic the process of natural selection by encoding hyperparameters as chromosomes, evaluating their performance (Fitness), and iteratively improving them through genetic operators such as selection, crossover, and mutation [7]. This evolutionary process enables the GA to avoid local minima and converge toward globally optimal solutions, making it particularly useful for high-dimensional and nonconvex tuning problems. Unlike traditional techniques, which are static and uninformed, the GA is adaptive and uses past evaluations to guide future searches, reducing redundant trials and improving convergence speed. While the GA is effective for exploring large search spaces, it often suffers from slow convergence and high computational cost, primarily when complex models are evaluated over many generations [7]. Additionally, its performance is sensitive to the tuning of GA-specific parameters such as population size, crossover, and mutation rates, which can lead to premature convergence or suboptimal results if not well-balanced [7]. This work aims to propose a predictive model for disease prediction via an optimized ML model with an improved GA with parallel fitness evaluation called the Multiprocessing Interface Genetic Algorithm (MIGA) to increase disease prediction and tuning efficiency.

2|Literature Review

Recent works have shown the effectiveness of the GA in optimizing ML models by addressing the limitations of traditional hyperparameter tuning methods. Rodrigues et al. [8] proposed a hybrid model combining a ResNet-50v2 CNN with a GA for classifying Acute Lymphoblastic Leukemia (ALL) from microscopy images. The GA was used to fine-tune the Convolutional Neural Network (CNN) hyperparameters, resulting in a classification accuracy of 98.46%, which outperformed both the random search and Bayesian optimization approaches. Kaur et al. [9] developed a GA-aided hyperparameter optimization framework combined with an ensemble learning model to predict respiratory diseases via clinical data. Their methodology also incorporated SHAP-based explainable AI to interpret model predictions. The GA-optimized AdaBoost classifier achieved the highest accuracy among all the models evaluated, showing the efficacy of GAs in tuning complex classifiers and reducing human involvement in parameter selection. In the energy sector, Liu et al. [10] applied a GA to enhance the performance of a Random Forest (RF) model for predicting grid faults. By selecting relevant meteorological features and tuning the model hyperparameters with a GA, the proposed method improved the prediction accuracy by 14.77% over that of standard RF models, illustrating the adaptability of GAs in real-time fault prediction under environmental variability.

Furthermore, Guido et al. [11] introduced a multiobjective GA approach for tuning Support Vector Machine (SVM) hyperparameters in imbalanced datasets. Their model, which incorporates decision trees to accelerate GA evaluations, significantly reduces the computational time while improving balanced accuracy metrics such as the G-mean, which are vital for healthcare-related classification tasks. Building on constrained optimization techniques, El-Hassani et al. [12] proposed a novel framework named MLPRGA+5, aimed at configuring Multilayer Perceptron (MLP) networks through real-coded GAs. The work implemented advanced genetic operators such as tournament selection with elitism, Simulated Binary Crossover (SBX), and Polynomial Mutation (PM). The model's efficiency was validated on four UCI datasets, where MLPRGA+5 achieved higher accuracy and reduced network complexity compared to traditional methods, confirming the robustness of the evolutionary design in practical neural network tuning. El-Hassani et al. [12] introduced an ensemble model (EGACNN) that stacks a CNN with a GA to optimize hyperparameters such as the dropout rate, batch size, and learning rate. Their method achieved 99.91% accuracy on MNIST, outperforming the classical CNN and other ensemble models, thereby validating the robustness of GAs in fine-tuning deep neural networks for image classification tasks. Ranga et al. [13] optimized an MLP using a GA to predict Chronic Kidney Disease (CKD) and achieved better accuracy scores of 98.34% and 98.54% for the training and testing processes, respectively.

3 | Methodology

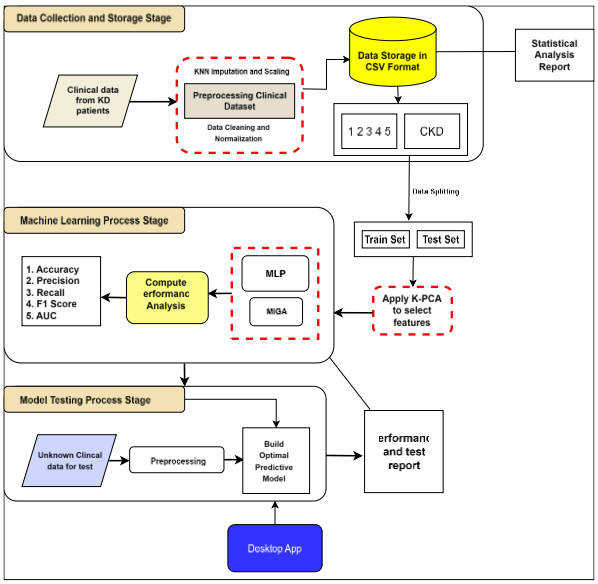

Fig. 1. Proposed framework.

3.1 | Dataset

3.1.1 | Chronic kidney disease

CKD is also called Chronic Renal Failure (CRF). CKD is a clinical syndrome characterized by permanent structural and functional damage to the kidneys that persists for at least three months, reducing their ability to filter metabolic waste, regulate fluid-electrolyte balance, and control blood pressure [14]. It typically develops gradually and remains unnoticed until approximately 25% of renal function is lost. In adults, CKD is diagnosed when the estimated Glomerular Filtration Rate (eGFR) falls below 60 mL/min/1.73 m² or when the eGFR is ≥ 60 mL/min/1.73 m² in the presence of markers of kidney damage, such as albuminuria over three months or more [14]. Globally, CKD affects more than 800 million individuals and is projected to rank among the five leading causes of death by 2040, highlighting the urgent need for accurate early detection and predictive modelling approaches [15]. The CKD dataset used in the study consists of 400 instances with 25

input features and 1 target feature, which is referred to as "Classification". The dataset has a total of 1009 missing values [16].

3.1.2 | Parkinson's disease

Parkinson's Disease (PD) is a progressive neurodegenerative disorder characterized by the loss of dopamine-producing neurons in the substantia nigra of the midbrain, resulting in hallmark motor symptoms such as resting tremor, bradykinesia, rigidity, and postural instability [17]. Globally, PD affects an estimated 7–10 million people, with the incidence rising markedly after age 60 and a male-to-female prevalence ratio of approximately 1.5:1 [18]. Early in its course, PD often presents with subtle voice- and speech-based changes, and dysphonia is detectable through sustained phonation before overt motor signs appear [19]. Diagnosis remains clinical and is based on established criteria, supplemented by neuroimaging and biomarkers. However, the insidious onset and overlapping features underscore the need for robust early-detection models trained on specialized PD datasets. The Parkinson dataset used in the study is obtained from the UCI ML repository, originally collected at Oxford Parkision's Disease Centre. The dataset comprises a total of 195 records with 23 features and one target feature, where the target feature is "Status". Notably, the dataset also contains no missing values [20].

3.1.3 | Breast cancer disease

Breast cancer is caused by the uncontrolled proliferation of epithelial cells within breast tissue, resulting in the formation of benign or malignant tumours [21]. Malignant lesions are defined by their capacity to invade adjacent stroma and disseminate to distant organs via lymphatic or hematogenous routes, which critically worsens prognosis. Mammography remains the state-of-the-art method for early detection; radiographic assessment focuses on lesion morphology, margin characteristics, and tissue density to distinguish benign from malignant findings [22]. Recent diagnostic methods integrate Deep Learning (DL) techniques to enhance classification performance on large mammographic datasets [23]. The breast cancer dataset used in this study was downloaded from the UCI ML repository. The dataset is from the University of Wisconsin Hospital and comprises 569 instances with 32 features, including 31 input features and one target feature. The target feature is "Diagnosis", and the dataset has no missing values [24].

3.2 | Data Preprocessing

The study utilized three datasets: only the CKD dataset contained missing values, which were handled via K-Nearest Neighbors (KNN) imputation. The categorical features were encoded using one-hot encoding. The dataset was then randomly split into training and testing sets, with 80% and 20% allocated to each, respectively.

3.3 | Kernel-Based Principal Component Analysis

Kernel Principal Component Analysis (Kernel PCA) is a type of PCA that extends PCA by mapping data into a high-dimensional feature space via a nonlinear kernel function, such as a radial basis, cosine, sigmoid or polynomial, and then computes principal components in that space, which allows the capture of complex, nonlinear structures that linear PCA cannot detect [3]. This method, therefore, provides more effective dimensionality reduction for downstream tasks such as disease classification. The study used the radial basis function kernel principal component to identify the best component that explained at least 95% of the total variance.

3.4 | Multilayer Perceptron

A MLP is a type of feed-forward Artificial Neural Network (ANN) comprising an input layer, one or more hidden layers of interconnected neurons, and an output layer; each neuron computes a weighted sum of its inputs plus a bias and applies a nonlinear activation function, enabling the network to approximate complex, nonlinear mappings through gradient-based optimization of its parameters without any feedback loops [25].

By propagating information unidirectionally from input to output, the MLP learns hierarchical feature representations and has demonstrated strong performance in classification tasks across various domains, as well as in CKD prediction, where MLPs combined with kernel-based dimensionality reduction achieved up to 100% accuracy [3].

3.5 | Genetic Algorithm

GAs are a class of population-based metaheuristic optimization techniques inspired by the principles of natural selection and evolutionary biology. A GA begins with a randomly generated population of candidate solutions, known as chromosomes, which are evaluated based on a fitness function that reflects how well each solution solves a given problem. Through iterative processes involving selection, Crossover (Recombination), and mutation, the algorithm evolves the population over multiple generations to improve overall fitness [26]. The selection phase allows high-performing individuals to pass their genes to the next generation, whereas crossover combines parts of two parents to produce offspring with mixed traits. Mutation introduces diversity by randomly altering genes, thereby helping to avoid local optima. GAs are particularly suitable for complex search spaces where traditional gradient-based methods may fail, as they do not require derivative information and are inherently parallelizable [27]. The proposed work modified the GA by enhancing the fitness evaluation with parallel evaluation through multithreading to increase computational time. The algorithm of the proposed technique is shown in *Algorithm 1*.

3.5.1 | Proposed multiprocessing interface genetic algorithm

Algorithm 1. Multiprocessing interface genetic algorithm.

Begin
1. Initialize population
2. Generate an initial population $\mathcal{P}_0 = \{\theta_1, \theta_2, \ldots, \theta_P\}$ with random values from S
3. For each generation, $g = 1$ to G:
   a. Parallel fitness evaluation:
   - Compute $F(\theta_i)$ for each $\theta_i \in \mathcal{P}_g$ by multithreading

   b. Selection:
   Rank population by fitness
   Select the top 50% as parents $\mathcal{P}_{elite} \subset \mathcal{P}_g$

   c. Crossover:
   While $|\mathcal{P}_{g+1}| < P$:
   i. Randomly select two parents $\theta_p^{(1)}, \theta_p^{(2)} \in \mathcal{P}_{elite}$
   ii. Generate child θ_c by randomly selecting each gene from either parent
   iii. Append θ_c to the next generation

   d. Mutation
   - For each offspring, with probability μ, one hyperparameter is randomly mutated.

   e. Update population
   - Set $\mathcal{P}_{g+1} \leftarrow$ newly formed generation.

   f. Track the best solution
   - If any $\theta \in \mathcal{P}_{g+1}$ has better fitness than θ^*, update $\theta^* \leftarrow \theta$.

   3. Return

Final best hyperparameter configuration θ^* and its fitness score $F(\theta^*)$.
End

4 | Results

The study was able to propose a modified GA for optimizing the MLP. The optimized model was used to predict diseases, and the framework was transformed into an app that incorporates the proposed hyperparameter tuning technique to achieve enhanced performance. The study evaluated the performance of the proposed model on three different datasets: breast cancer, CKD, and Parkinson's. *Table 1* presents the hyperparameters optimized via the GA during the training of the MLPC. The search space for the hidden layer sizes included various configurations ranging from (50), (100), and (150) to two-layer architectures such

as (50, 50) and (100, 100). The activation functions considered in the search process were ReLU, tanh, and logistic, allowing the model to explore different nonlinear transformation capabilities. The initial learning rates were tuned across three values: 0.001, 0.01, and 0.1, enabling the GA to adjust the convergence behaviour of the neural network. Finally, the solvers 'adam' and 'SGD' were explored to determine the most effective optimization technique for minimizing loss during training.

Table 1. Hyperparameters tuned by the genetic algorithm.

Hyperparameter	Search Space/Values
Hidden layer sizes	(50), (100), (150), (50, 50), (100, 100)
Activation function	'relu', 'tanh', 'logistic'
Learning rate init	0.001, 0.01, 0.1
Solver	'adam', 'SGD'

Table 2 outlines the configuration settings employed for the proposed GA in tuning the hyperparameters of the MLP. The population size was set to 10 individuals, and the algorithm evolved over 10 generations. A mutation rate of 10% was applied, where one hyperparameter in a chromosome was randomly altered. The selection strategy used was elitist selection, retaining the top 50% of individuals based on fitness scores. A uniform crossover approach was adopted to recombine the parent chromosomes. The fitness function was defined as the accuracy score on the test dataset, and evaluations were executed in parallel via ThreadPoolExecutor to accelerate computation. Each chromosome encodes a complete model configuration represented by the tuple [Hidden layer size, activation function, learning rate, solver]. The proposed technique was explicitly applied to optimize an MLP with a maximum iteration limit of 500.

Table 2. Genetic algorithm settings.

GA Parameter	Value
Population size	10
Number of generations	10
Mutation rate	0.1 (10%)
Selection strategy	Elitist selection (Top 50% by fitness)
Crossover strategy	Uniform crossover
Mutation strategy	Random mutation of one hyperparameter
Fitness function	Accuracy score on test set
Parallel evaluation	(ThreadPoolExecutor)
Chromosome representation	[Layer size, activation, learning rate, solver]
Model evaluated	MLPClassifier (Max_iter = 500)

Table 3 presents the performance metrics of the MIGA by generation when applied to the Parkinson dataset. Across 10 generations, the algorithm consistently demonstrated strong optimization capability, with the best accuracy reaching 0.9487 from generation 2 onwards. The minimum and maximum accuracy values for each generation reflect the diversity in model performance within the population, with early fluctuations gradually stabilizing in later generations. The results indicate that the algorithm effectively converges toward high-performing hyperparameter combinations, maintaining an accuracy of 0.9487 for the majority of generations.

Table 3. Performance of the multiprocessing interface genetic algorithm when applied to the multilayer perceptron on the Parkinson dataset.

Generation	Min	Max	Best
1	0.8718	0.9231	0.9231
2	0.8974	0.9487	0.9487
3	0.8974	0.9487	0.9487
4	0.8974	0.9487	0.9487
5	0.8718	0.9487	0.9487

Table 3. Continued.

Generation	Min	Max	Best
6	0.8974	0.9487	0.9487
7	0.8462	0.9487	0.9487
8	0.8462	0.9487	0.9487
9	0.8974	0.9231	0.9231
10	0.8974	0.9487	0.9487

Table 4 summarizes the performance of the MIGA across 10 generations on the Breast Cancer dataset. The results show strong classification performance, with the best accuracy, reaching a peak of 0.9912 as early as generation 2 and remaining consistently high throughout subsequent generations. The minimum and maximum accuracies recorded per generation demonstrate moderate variation in early generations, followed by stabilization as the algorithm converges. The sustained best accuracy across multiple generations indicates the robustness and efficiency of the MIGA in evolving optimal hyperparameter configurations for MLP-based classification on the breast cancer dataset.

Table 4. Performance of the multiprocessing interface genetic algorithm when applied to the multilayer perceptron on the Breast Cancer dataset.

Generation	Min	Max	Best
1	0.9561	0.9625	0.9625
2	0.9649	0.9912	0.9912
3	0.9825	0.9912	0.9912
4	0.9373	0.9825	0.9825
5	0.9825	0.9825	0.9825
6	0.9561	0.9825	0.9825
7	0.9561	0.9825	0.9825
8	0.9561	0.9825	0.9825
9	0.9825	0.9825	0.9825
10	0.9649	0.9912	0.9912

Table 5 presents the accuracy performance of the MIGA over 10 generations when it is applied to the CKD dataset. The algorithm consistently achieves exceptional accuracy, with the best performance reaching 1.0000 (100%) from the first generation and maintaining this level throughout all subsequent generations. The minimum and maximum accuracy values indicate only instability in the population, with minimum values ranging from 0.9625 to 0.9875. This high and stable performance across generations highlights the effectiveness of the MIGA in discovering optimal MLP hyperparameters for CKD classification tasks, suggesting strong model generalizability and convergence.

Table 5. Performance of the multiprocessing interface genetic algorithm when the multilayer perceptron is applied to the chronic kidney disease dataset.

Generation	Min	Max	Best
1	0.9625	1.0000	1.0000
2	0.9750	1.0000	1.0000
3	0.9750	1.0000	1.0000
4	0.9750	1.0000	1.0000
5	0.9750	1.0000	1.0000
6	0.9750	1.0000	1.0000
7	0.9750	1.0000	1.0000
8	0.9875	1.0000	1.0000
9	0.9875	1.0000	1.0000
10	0.9750	1.0000	1.0000

Table 6 summarizes the best-performing hyperparameter configurations obtained with the MIGA on the Parkinson's, breast cancer, and CKD datasets. For the Parkinson dataset, the optimal configuration achieved is 150 hidden layer units, ReLU activation, a learning rate of 0.1, and the Adam solver, which achieves an accuracy of 95.00%. For the Breast Cancer dataset, the proposed model achieved the highest accuracy of 99.12% with 50 hidden units, the tanh activation function, a learning rate of 0.001, and the Adam solver. Additionally, on the CKD dataset, the proposed model performed best with 50 hidden units, tanh activation, a learning rate of 0.1, and the SGD solver, resulting in 100% accuracy. These results demonstrate the adaptability and effectiveness of the MIGA in fitting the MLP configurations for the three different medical datasets.

Table 6. Optimal configuration of hyperparameters applied to each dataset with the multiprocessing interface genetic algorithm.

Dataset	H/L	Activation	Learning Rate	Solver	Accuracy
Parkison	150	ReLu	0.1	Adam	95.00%
Breast Cancer	100	Tanh	0.001	Adam	99.00%
CKD	100	Tanh	0.1	SGD	100%

Table 7 shows the timing logs from the hyperparameter tuning experiments. For each of the three datasets, the study recorded the total wall-clock time of the 10-generation GA search, both in the standard (Single-threaded) mode and with the proposed MIGA (Parallel) evaluator, in seconds.

Table 7. Comparison of the tuning time and speed-up between the standard genetic algorithm and the multiprocessing interface genetic algorithm.

Dataset	Standard GA Time (s)	MIGA Time (s)	Reduction (%)
Breast Cancer	107.05	48.05	59.0
Parkison	95.30	34.20	61.1
CKD	71.46	11.46	60.0
Average	91.27	31.24	60.3

The proposed MLP tuned with the proposed MIGA performed better than the state-of-the-art optimization approaches across all three datasets. A summary of the comparison of the breast cancer datasets is shown in *Table 8*.

Table 8. Comparison of MIGA+MLP with several optimization-based models.

Dataset	Author	Model	Optimization Method	Accuracy
Breast Cancer	El-Hassani et al. [12]	ANN	RCGAs	96.00%
	Aguerchi et al. [23]	CNN	PSO	98.23% (DDSM), 97.98% (MIAS)
	Zhu et al. [28]	LightGBM	PSO	99.0%
	Proposed method	MLP	MIGA	99.00%
CKD	Ranga et al. [13]	MLP	GA	98.54%
	Swain et al. [29]	SVM	Grid Search	99.33%

Table 8. Continued.

Dataset	Author	Model	Optimization Method	Accuracy
	Kaur et al. [9]	MLP	PSO	92.76%
	Proposed method	MLP	MIGA	100%
Parkison	Elshewey et al. [30]	SVM	Bayesian Optimization	92.30%
	Doumari et al. [17]	MLP	Quantum Particle Swarm Optimization (QPSO)	93.00%
	Ali et al. [19]	NN	GA	95.00%
	Proposed method	MLP	MIGA	95.00%

The study developed a GUI that performs hyperparameter tuning in a unified window: The top panel includes a select file button to load any CSV file; a dropdown to select the target variable; text fields for user hidden-layer sizes, activation functions, learning rates, and solvers; and a start-tuning button. The bottom panel displays a scrollable console that logs each generation's best configuration and then summarizes the optimal hyperparameters, MIGA runtime, training time, test accuracy, confusion matrix, and classification report. Finally, a save model button exports the trained MLP. *Fig. 2* shows the interface of the proposed model on the CKD dataset. *Fig. 3* shows the interface on the Parkinson's dataset. *Fig. 4* shows the interface on the breast cancer dataset.

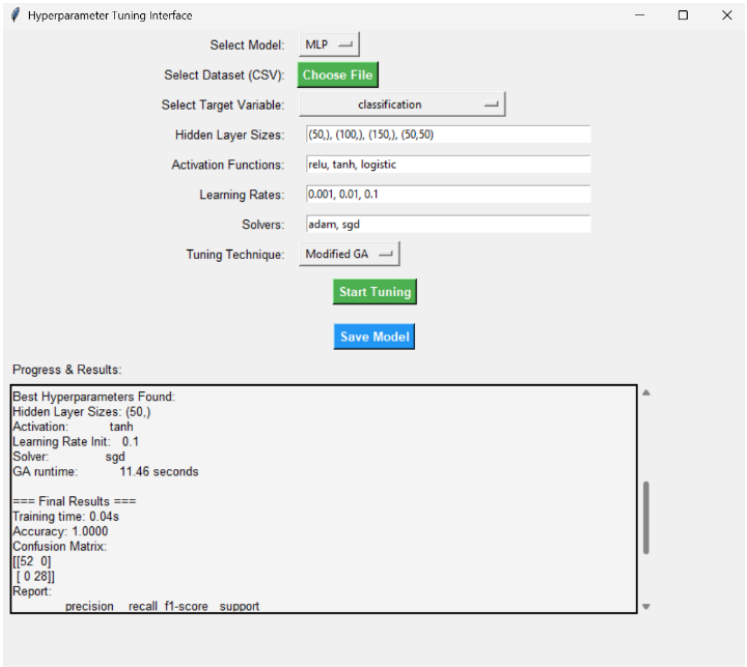

Fig. 2. The proposed method interface on the chronic kidney disease dataset.

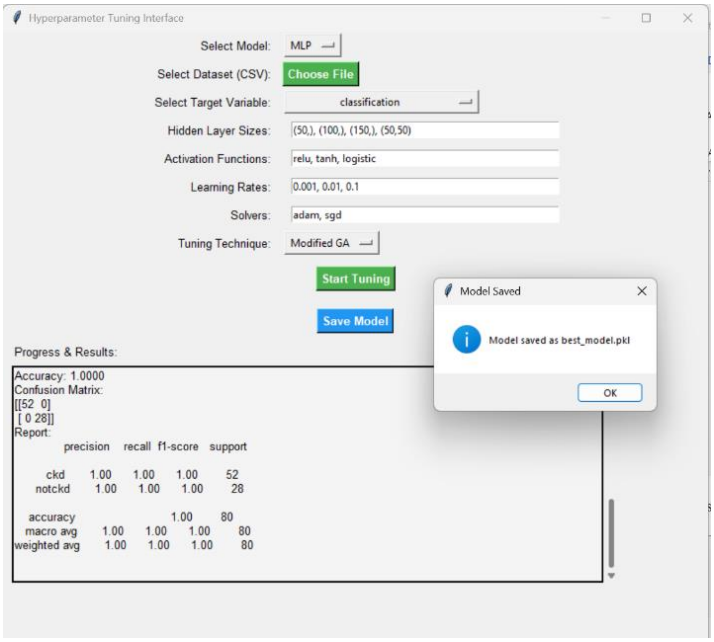

Fig. 3. Developed interface.

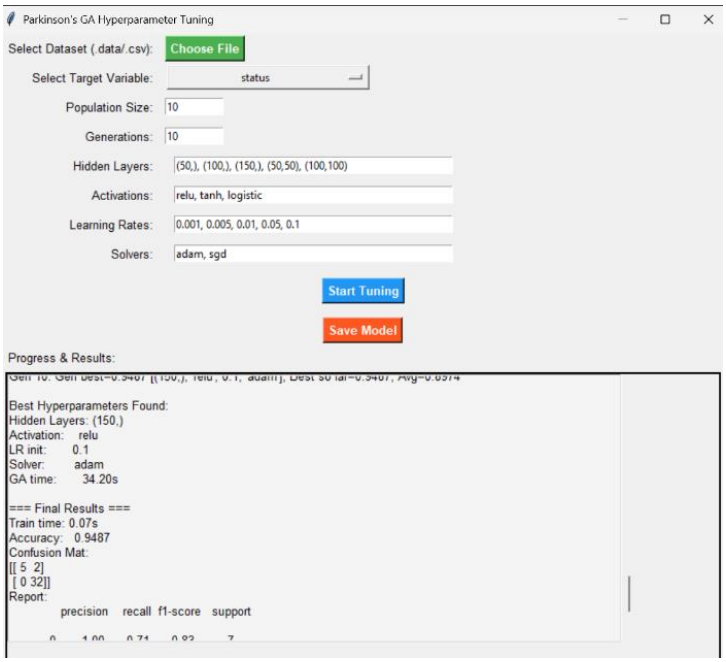

Fig. 4. Proposed method interface for Parkinson's disease.

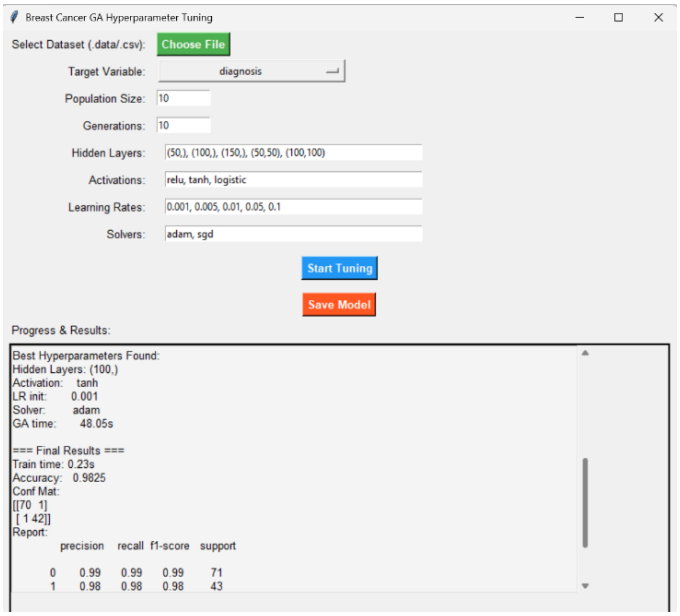

Fig. 5. Proposed method interface for treating breast cancer.

5 | Discussion

The proposed MIGA overcomes the computational cost of the GA by parallelizing fitness evaluations across multiple CPU cores, which reduces the tuning time by approximately 60% compared to a single-threaded approach [7]. This efficiency gain enables the algorithm to maintain or even expand its population size and number of generations without requiring excessive runtime. The combination of nonlinear feature extraction via radial-basis-function kernel PCA and MIGA-tuned MLP produced classifiers that generalize exceptionally well, achieving 99.12% for breast cancer, 94.87% for PD, and 100% for CKD. These results suggest that the framework can capture complex relationships in diverse biomedical datasets. Future work will apply this methodology to additional clinical cohorts to confirm its robustness and to discover dataset-specific feature interactions. The developed graphical user interface, which guides users through data import, dimensionality reduction, hyperparameter optimization, and model evaluation without any programming, is crucial for translating these advanced AI methods into everyday clinical practice.

6 | Conclusion

This study introduced a disease-prediction framework that integrates kernel PCA for nonlinear dimensionality reduction, an MLP classifier, and a modified multiprocessing-enabled GA (MIGA) for efficient hyperparameter tuning. The approach was evaluated on three different medical datasets: breast cancer, PD, and CKD. The MLP-tuned MIGA outperforms models optimized with traditional methods, achieving the best accuracies of 99.12% for breast cancer, 94.87% for PD, and 100% for CKD. The use of kernel PCA enables the extraction of nonlinear structures, thereby improving predictive performance. In contrast, the parallel fitness evaluations in the MIGA reduce computational time and accelerate convergence toward optimal hyperparameters. A user-friendly GUI further enables non-expert clinicians to apply the framework to new datasets with minimal effort. This study has proven the potential of combining advanced optimization methods with nonlinear feature extraction to increase disease prediction accuracy and usability in real-world healthcare settings.

Authors Contribution

I.I. proposed the study conception, design, analysis, and interpretation; S.B. provided administrative support and study materials; and A.Y. was a major contributor to manuscript writing. All the authors read and approved the final manuscript.

Data Availability

The datasets analysed during the current study are available on the University of California Irvine Repository website and the Kaggle website repository, https://www.kaggle.com/datasets/vikasukani/parkinsons-disease-data-set,https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease, https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic).

Conflicts of Interest

The authors declare that they have no competing interests.

References

[1] Ganie, S. M., Dutta Pramanik, P. K., Mallik, S., & Zhao, Z. (2023). Chronic kidney disease prediction using boosting techniques based on clinical parameters. *Plos one*, *18*(12), e0295234. https://doi.org/10.1371/journal.pone.0295234

[2] Arif, M. S., Rehman, A. U., & Asif, D. (2024). Explainable machine learning model for chronic kidney disease prediction. *Algorithms*, *17*(10), 443. https://doi.org/10.3390/a17100443

[3] Iliyas, I. I., Boukari, S., & Ya, A. (2024). A proposed multilayer perceptron model and kernel principal analysis component for the prediction of chronic kidney disease. *International journal of artificial intelligence*, *11*(2), 99–113. https://doi.org/10.36079/lamintang.ijai-01102.783

[4] Bischl, B., Richter, J., Becker, M., Binder, M., & Pielok, T. (2023). Hyperparameter optimization : Foundations , algorithms , best practices , and open challenges, *13*(2), 1–43. https://doi.org/10.1002/widm.1484

[5] Hernández-Morales, A., Van Nieuwenhuyse, I., & Gonzalez, R. S. (2023). A survey on multi objective hyperparameter optimization algorithms for machine learning. *Artif intell rev, 56*, 8043–8093. https://doi.org/10.1007/s10462-022-10359-2

[6] Javaid, M., Haleem, A., Haleem, I., & Suman, R. (2023). Advanced Agrochem Understanding the potential applications of Artificial Intelligence in Agriculture Sector. *Advanced agrochem*, *2*(1), 15–30. https://doi.org/10.1016/j.aac.2022.10.001

[7] Ghezelbash, R., Maghsoudi, A., Shamekhi, M., Pradhan, B., & Daviran, M. (2022). Genetic algorithm to optimize the SVM and K -means algorithms for mapping of mineral prospectivity. *Neural computing and applications*, *35*, 719–733. https://doi.org/10.1007/s00521-022-07766-5

[8] Rodrigues, L. F., Backes, A. R., Travençolo, B. A. N., & de Oliveira, G. M. B. (2022). Optimizing a deep residual neural network with genetic algorithm for acute lymphoblastic leukemia classification. *Journal of digital imaging*, *35*, 623–637. https://doi.org/10.1007/s10278-022-00600-3

[9] Kaur, B. P., Id, H. S., Hans, R., Sharma, S. K., Sharma, C., & Id, M. H. (2024). A Genetic algorithm aided hyper parameter optimization based ensemble model for respiratory disease prediction with explainable AI. *Plos one, 19*(12), e0308015. https://doi.org/10.1371/journal.pone.0308015

[10] Liu, K., Gu, Y., Tang, L., Du, Y., Zhang, C., & Zhu, J. (2025). Random forest grid fault prediction based on genetic algorithm optimization, *13*, 1480749. https://doi.org/10.3389/fphy.2025.1480749

[11] Guido, R., Carmela, M., & Domenico, G. (2023). A hyper-parameter tuning approach for cost-sensitive support vector machine classifiers. *Soft computing*, *27*(18), 12863–12881. https://doi.org/10.1007/s00500-022-06768-8

[12] El-Hassani, F. Z., Amri, M., Joudar, N. E., & Haddouch, K. (2024). A new optimization model for MLP hyperparameter tuning: Modeling and resolution by real-coded genetic algorithm. *Neural processing letters*, *56*(2), 1–31. https://doi.org/10.1007/s11063-024-11578-0

[13] Ranga, P., Terlapu, V., Jayaram, D., Rakesh, S., & Gopalachari, M. V. (2023). Optimizing chronic kidney disease diagnosis in uddanam : A smart fusion of GA-MLP hybrid and PCA dimensionality reduction. *Procedia computer science*, *230*(2023), 522–531. https://doi.org/10.1016/j.procs.2023.12.108

[14] Iliyas, I. I., Boukari, S., & Ya, A. (2025). Recent trends in prediction of chronic kidney disease using different learning approaches : A systematic literature review. *Journal of medical artificial intelligence*, *8*. https://doi.org/10.21037/jmai-24-256

[15] Arif, S. M., Mukheimer, A., & Asif, D. (2023). Enhancing the early detection of chronic kidney disease : A robust machine learning model. *Big data and cognitive computing*, *7*, 144. https://doi.org/10.3390/bdcc7030144

[16] Rubini, L., Soundarapandian, P., & Eswaran, P. (2015). Chronic kidney disease [Dataset]. *UCI machine learning repository*. https://doi.org/10.24432/C5G020.

[17] Doumari, S. A., Berahmand, K., & Ebadi, M. J. (2023). Early and high-accuracy diagnosis of parkinson' s disease : Outcomes of a new model. *Computational and mathematical methods in medicine*, *2023*(1), 1493676. https://doi.org/10.1155/2023/1493676

[18] Kumar, G. S., Suganya, E., Sountharrajan, S., & Balusamy, B. (2025). OPEN SRADHO : Statistical reduction approach with deep hyper optimization for disease classification using artificial intelligence. *Scien*, *15*(1), 1245. https://doi.org/10.1038/s41598-024-82838-1

[19] Ali, L., Zhu, C. E., & Zhang, Z. (2019). Automated detection of parkinson' s disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network. *IEEE journal of translational engineering in health and medicine*, *7*, 1–10. https://doi.org/10.1109/JTEHM.2019.2940900

[20] Ukani, V. (2020). *Parkinson's disease data set*.

[21] Islam, T., Sheakh, M.A., Tahosin, M.S., … ., & Bourhia, M. (2024). Predictive modeling for breast cancer classification in the context of Bangladeshi patients by use of machine learning approach with explainable AI. *Sci Rep, 14*. https://doi.org/10.1038/s41598-024-57740-5

[22] Michael, E., Ma, H., Li, H., & Qi, S. (2022). An optimized framework for breast cancer classification using machine learning. *BioMed research internation*, *2022*(1), 8482022. https://doi.org/10.1155/2022/8482022

[23] Aguerchi, K., Jabrane, Y., Habba, M., Hajjam, A., & Hassani, E. (2024). A CNN hyperparameters optimization based on particle swarm optimization for mammography breast cancer classification. *Journal of imaging*, *10*(2), 30. https://doi.org/10.3390/jimaging10020030

[24] Wolberg, W., Mangasarian, O., Street, N., & Street, W. (1993). Breast cancer wisconsin (Diagnostic) [Dataset]. *UCI machine learning repository*. https://doi.org/10.24432/C5DW2B.

[25] Woodman, R.J., Mangoni, A.A. (2023). A comprehensive review of machine learning algorithms and their application in geriatric medicine: Present and future. *Aging clin exp res*, *35*(11), 2363–2397. https://doi.org/10.1007/s40520-023-02552-2

[26] Deighan, D. S., Field, S. E., Capano, C. D., Khanna, G., & Field, S. E. (2021). Genetic-algorithm-optimized neural networks for gravitational wave classification. *Neural computing and applications manuscript*, *33*, 13859–13883. https://doi.org/10.1007/s00521-021-06024-4

[27] Ojha, V., Timmis, J., & Nicosia, G. (2022). Assessing ranking and effectiveness of evolutionary algorithm hyperparameters using global sensitivity analysis methodologies. *Swarm and evolutionary computation*, *74*, 101130. https://doi.org/10.1016/j.swevo.2022.101130

[28] Zhu, J., Zhao, Z., Yin, B., Wu, C., Yin, C., & Chen, R. (2025). An integrated approach of feature selection and machine learning for early detection of breast cancer, *Scientific reports*, *15*(1), 13015. https://doi.org/10.1038/s41598-025-97685-x

[29] Swain, D., Mehta, U., Bhatt, A., Patel, H., Patel, K., Mehta, D., … ., & Manika, S. (2023). A robust chronic kidney disease classifier using machine learning. *Electronics*, *12*(212), 1–13. https://doi.org/ 10.3390/electronics12010212

[30] Elshewey, A. M., Shams, M. Y., El-Rashidy, N., Elhady, A. M., Shohieb, S. M., & Tarek, Z. (2023). Bayesian optimization with support vector machine model for parkinson disease classification. *Sensor*, 23(4), 2085. https://doi.org/10.3390/s23042085