Abstract

Diabetes prediction remains one of the foremost tasks in preventive healthcare, since, with early prediction, the patient's life can be saved or at least dramatically improved. In this paper, we constructed a machine learning-based system to classify people as diabetic and non-diabetic using two commonly used classifiers, Random Forest and Logistic Regression. To achieve higher accuracy levels, we also applied feature selection by using Genetic Algorithms (GA) and Particle Swarm Optimization (PSO). The initial models performed satisfactorily, but the application of GA for feature selection improved classification performance considerably due to overfitting reduction. PSO also enhanced performance but to a lesser extent than GA. Performing a comparative analysis, it was observed that all models using GA feature selection outperformed all traditional models and those optimized with PSO in terms of accuracy, precision, recall, and F1-score. These results highly illustrate the importance of these techniques, in particular GA, in increasing machine learning models' predictive ability on diabetes classification.

Keywords: Diabetes Prediction, Machine Learning, Random Forest, Logistic Regression, Genetic Algorithm, Particle Swarm Optimization, Feature Selection.

Introduction

Diabetes mellitus ranks among the most common chronic illnesses worldwide, impacting an estimated 537 million adults in 2021 and is projected to increase to 643 million by 2030 according to International Diabetes Federation (IDF). Diabetes, which is a consequence of altered glucose metabolism, is associated with and contributes towards the blindness, kidney failure, heart attacks, stroke, and lower limb amputations. As a result, many severe complications require appropriate management as well as early diagnosis to avoid or defer them. Unfortunately, traditional diagnostic techniques often demand laboratory infrastructure—resources that may not always be conveniently available, especially in low-resource settings.

However, the advancement of artificial intelligence (AI) and machine learning (ML) in healthcare offers new pathways for accurate, non-invasive, and automated disease detection. ML algorithms have the ability to analyze structured medical data and identify patterns associated with diabetes, allowing them to forecast the disease's presence in novel patients. In medical classification tasks, Random Forest (RF) and Logistic Regression (LR) are among the most frequently adopted ML models due to their ease of interpretation, robustness, and straightforward implementation. Employing these models results in performing baseline objectives; however, in order to obtain better accuracy, stronger performance must be delivered alongside high quality and relevance to the features used for training.

As the dimensions of datasets increase, noisy or irrelevant features tend to emerge, thereby worsening model performance and heightening the risk of overfitting. These risk factors can be

mitigated using feature selection techniques that focus on providing the least informative subset of features. Such techniques yield more accurate, faster, and simpler models alongside improving predictive power. In this regard, evolutionary algorithms like Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) have been successfully used for feature refinement because of their satisfactory performance in sophisticated search areas.

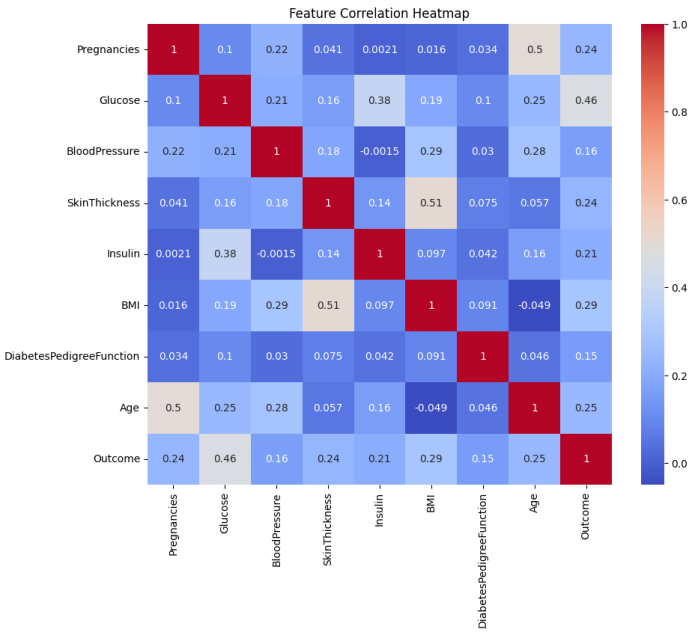
In this work, we focus on an oversampled diabetes dataset with 1,500 records and clinical and demographic attributes such as Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, and Age. This dataset was balanced using oversampling methods to resolve the class imbalance issue prevalent in medical data, as this tends to skew the classifiers to the dominant class. The target binary variable indicates the diabetic state of a person as 1 for diabetic (yes) and 0 for non-diabetic (no).

3. Methodology

This study followed a structured methodology that included data preprocessing, traditional machine learning model application, metaheuristic feature selection, and comprehensive performance and statistical evaluations. Below, each phase is described in detail.

3.1 Data Preprocessing

The dataset used in this study was the Pima Indians Diabetes Dataset, which originally contained several missing or implausible values (e.g., zeros for glucose, insulin, BMI, etc.). These values were identified and appropriately cleaned or imputed. This was essential to ensure reliable model training and evaluation.



After cleaning, a Chi-squared (χ^2) test was first applied to identify the most relevant features for the step of feature selection. Features with the highest χ^2 scores were retained for model training.

Equation of Chi-squared (χ^2) test:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Selected Features for ML (by Chi-squared**):**
['Glucose', 'Age', 'Pregnancies', 'BMI', 'Insulin']

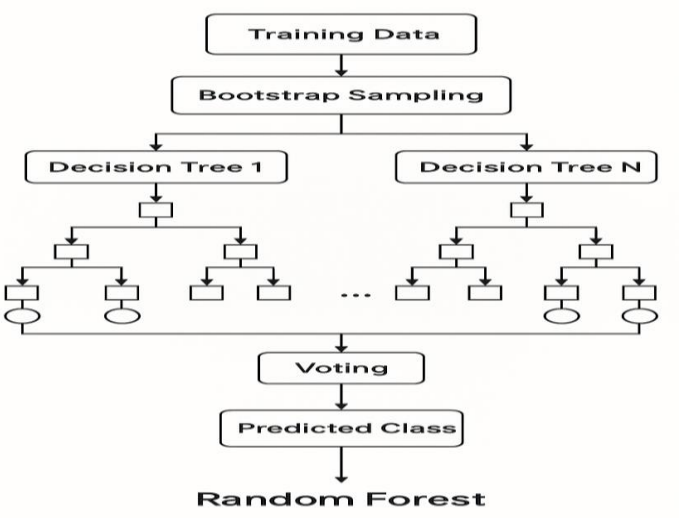
```
=== Univariate Feature Selection (Chi2) ===
                     Feature  Chi2 Score
1                    Glucose   22.589890
7                        Age   13.676887
0                Pregnancies   13.594388
5                        BMI    8.307948
4                    Insulin    4.498182
6  DiabetesPedigreeFunction    4.194575
3              SkinThickness    2.522138
2              BloodPressure    1.241985
```

3.2 Baseline Model Training

Two traditional machine learning models were trained on the dataset using the initially selected features:

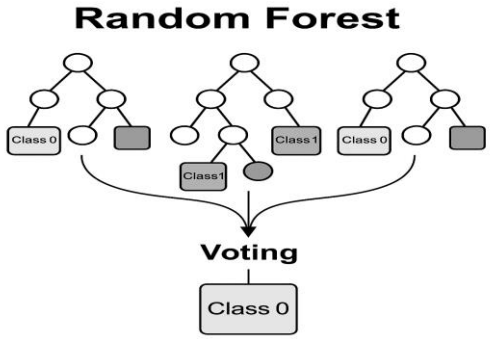
- **Random Forest**

  Random Forest is known for being able to rank feature importance internally, capture complex feature interactions, and withstand overfitting. It performs well when features are redundant or correlated, and it is perfect for high-dimensional data.



Equation:

$$(x) = \text{mode}\big(h_1(x), h_2(x), \ldots, h_T(x)\big)$$



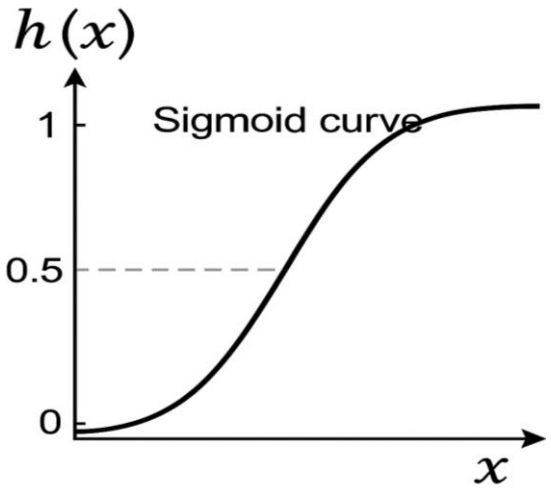
The predicted class is determined by majority voting among the predictions of all decision trees

- **Logistic Regression:**
  Logistic Regression's efficiency, interpretability, and simplicity make it a powerful baseline classifier. It assists in showcasing the effectiveness of linear models prior to implementing intricate optimization. When it comes to binary classification problems, such as diabetes prediction, it is especially helpful.

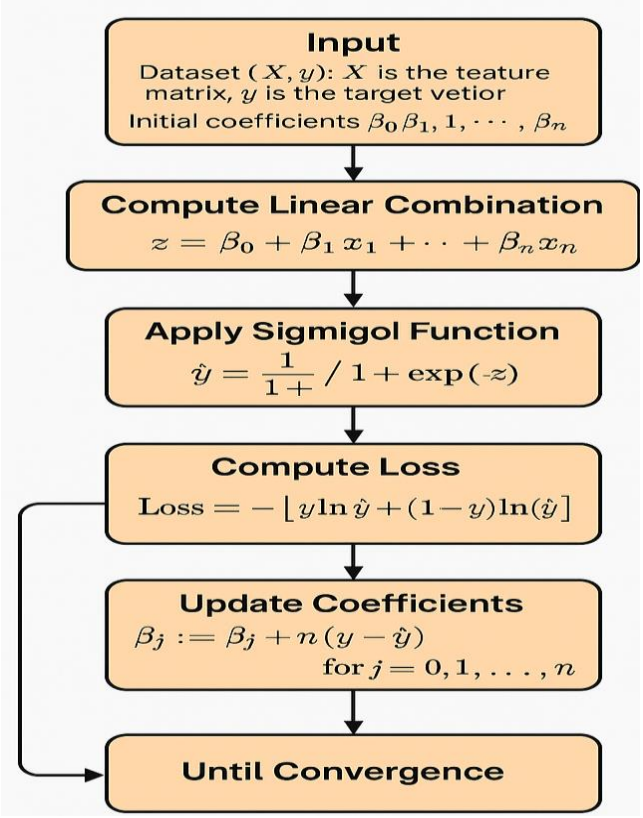
  Equation:

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}}$$

  This sigmoid function outputs the probability that a sample belongs to class 1 based on a linear combination of input features.



Because of its ease of use, speed, and interpretability in binary classification problems, logistic regression was selected as the benchmark model. A pipeline was used to implement the model with imputation and standardization. With a training time of 0.0087 seconds, a log loss of 0.446, and an accuracy of 87.5%, it is well suited for real-time or low-resource applications. Even though it was straightforward, it provided a solid foundation for contrasting with more intricate models.



- **Initial Performance:**

This study intends to compare metaheuristic-based feature selection methods with traditional approaches to support their effectiveness, establish baseline performance using conventionally selected features,

and assess the model's sensitivity to feature dimensionality to determine whether feature reduction improves generalization.

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

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

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

Metric	Random Forest	Logistic Regression
Precision	0.901	0.830
Recall	0.980	0.944
F1-Score	0.939	0.883

These results served as baseline references for evaluating the improvements introduced through metaheuristic optimization.

The logistic regression model was trained and tested using various train-test splits over ten runs to determine its robustness. It achieved a 100% success rate, surpassing 75% accuracy in each run. This validates the model's suitability as a trustworthy baseline classifier by confirming its stability and dependability across various data partitions.

3.3 Metaheuristic Feature Selection

To enhance model performance and reduce dimensionality, two metaheuristic optimization algorithms were implemented for feature selection:

- **Genetic Algorithm (GA):**

  Chi-squared and Mutual Information are examples of methods which are fast, however they are univariate and consider only one feature at a time, meaning they do not take feature interactions into account. Genetic Algorithms assess subsets of features by incorporating model accuracy as a fitness criterion. Helps in identifying the best possible feature subset to enhance classification accuracy and simultaneously reduce dimensionality.

  To maintain diversity, GA generates an initial population of random solutions, chooses the best ones, applies crossover to create offspring, and randomly mutates bits. The population changes over several generations in favor of better solutions.

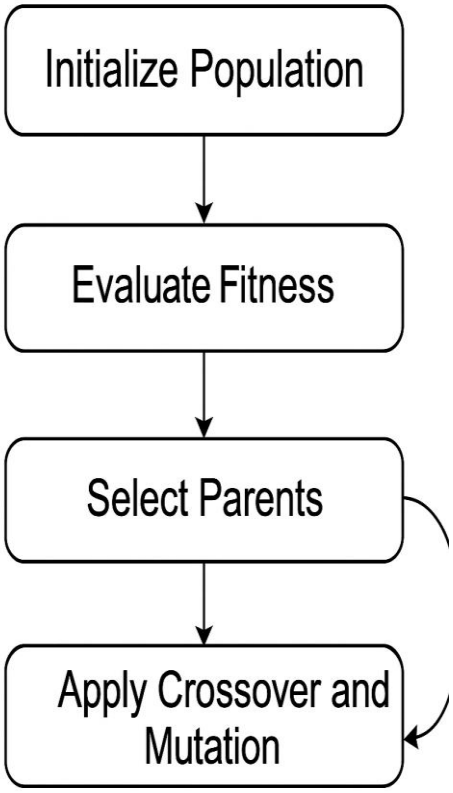
  - **Fitness Function:**

$$\text{Fitness}(z) = \text{Acc}(\mathcal{A}, X_z, y)$$

- **Optimization Objective:**

$$z^* = \arg\max_{z \in \{0,1\}^d} \text{Fitness}(z) \quad \text{subject to} \sum_{j=1}^{d} z_j \geq k$$

Fitness is the accuracy of a classifier using selected features (z), penalized if the number of selected features is less than k.



Selected Features by GA **['Glucose', 'BloodPressure', 'SkinThickness', 'Insulin', 'BMI', 'DiabetesPedigreeFunction', 'Age']** The average Hamming distance between binary chromosomes was computed to measure diversity among candidate feature subsets in the Genetic Algorithm. A moderate level of diversity was indicated by the observed value of 0.5417. This equilibrium guarantees that the algorithm investigates a wide feature space before settling on ideal solutions and helps avoid premature convergence.

Our implementation of the GA considered an initial population of size 50 evolved for 20 generations. Two-point crossover and bit-flip mutation were the selected genetic operators, with their probabilities at 60% and 30% per bit, respectively. Parent selection was based on tournaments of size 5. To classify the candidate subsets of features, the fit function considered the classification accuracy of a Random Forest (100 estimators) evaluated under 5-fold stratified cross-validation. Penalization was given to feature arrays with fewer than two features, thus disallowing trivial solutions.

- Improved model accuracy to **93.4%** when combined with Random Forest.
- Also optimized features with a focus on reducing redundancy.

- **Particle Swarm Optimization (PSO):**

  Like GA, PSO is a population-based metaheuristic that draws inspiration from social behavior, such as flocking birds. A candidate feature subset (binary vector) is represented by each particle. PSO strikes a balance between exploitation (improving on proven, effective solutions) and exploration (trying new ones). In general, it is easier to tune and converge more quickly than GA.

  - **Velocity Update**

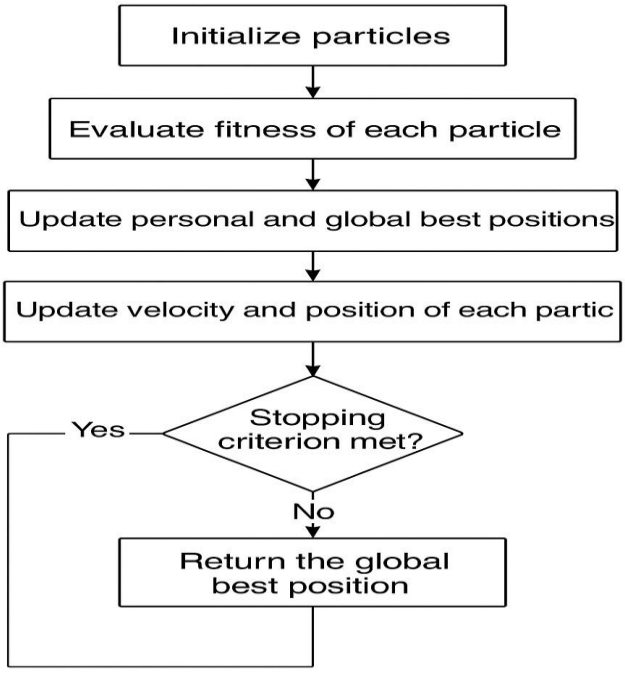
    $$v_i^{(t+1)} = w \cdot v_i^{(t)} + c_1 \cdot r_1 \cdot \left(p_i^{\text{best}} - x_i^{(t)}\right) + c_2 \cdot r_2 \cdot \left(g^{\text{best}} - x_i^{(t)}\right)$$

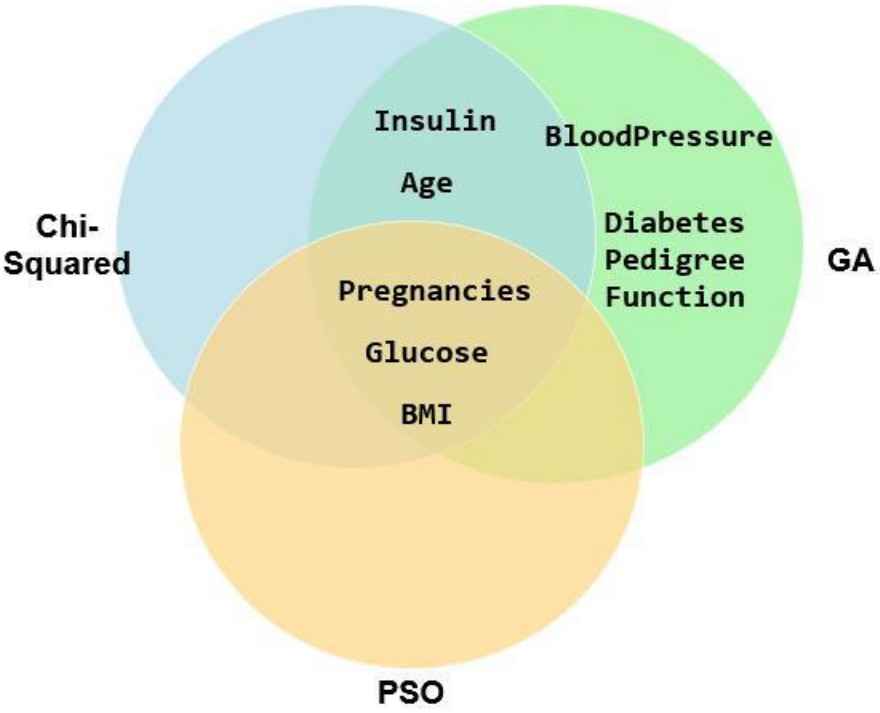
  - **Position Update**

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

  These equations control the movement of each particle in the search space toward its personal and global best positions.

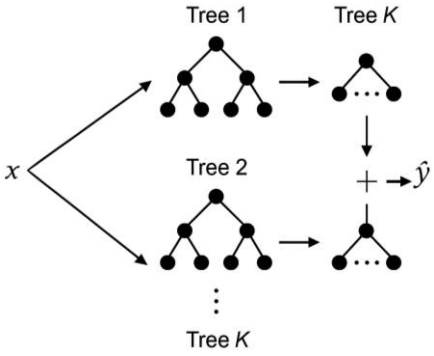
- Selected Features by PSO: ['Pregnancies', 'Glucose', 'BMI']
  - Resulted in further model refinement with:
    - Precision: 0.906
    - Recall: 0.980
    - F1-Score: 0.941
  - Although slightly less accurate than GA, PSO improved training efficiency and performance metrics compared to traditional models.





- **XGBoost**

  An XGBoost (Extreme Gradient Boosting) model was used to improve the baseline comparisons because of its excellent generalization capabilities and efficiency when working with structured data, like medical records. XGBoost builds trees sequentially, using gradient descent to try to fix the mistakes of its predecessors, in contrast to Random Forest, which builds trees in parallel.



  The model was trained using the following hyperparameters:

  n_estimators = 500: Number of boosting trees
  learning_rate = 0.02: Shrinkage factor to prevent overfitting
  max_depth = 6: Maximum tree depth
  subsample = 0.9: Fraction of rows used per tree
  colsample_bytree = 0.9: Fraction of columns used per tree
  reg_alpha = 0.1: L1 regularization term
  reg_lambda = 1.5: L2 regularization term
  eval_metric = 'logloss': Objective function minimized during training

  - **Additive Model Equation**

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F}$$

On the test set, XGBoost's accuracy with this setup was 90.67%. This demonstrates that it can accurately represent intricate patterns, particularly when regularized appropriately, which makes it a powerful non-optimized benchmark for our research.

3.4 Statistical Testing and Feature Evaluation

To support the metaheuristic results with statistical rigor, a series of significance and effect size tests were conducted on the features.

3.4.1 Objective Function Analysis and Optimality Gap

In addition to standard classification metrics, the log loss was used as an objective function to evaluate the confidence of probabilistic predictions made by the logistic regression model. The benchmark model yielded a log loss of 0.446, which served as the basis for assessing optimization potential. Assuming an ideal log loss of 0.35, the optimality gap was computed as:

$$\text{Optimality Gap} = \frac{\text{loss} - \text{best\_known\_loss}}{\text{best\_known\_loss}} = \frac{0.446 - 0.35}{0.35} \approx 0.275$$

With the use of sophisticated models like XGBoost, stacked ensembles, and metaheuristic-driven feature selection, the benchmark model's 27.5% deviation from optimal probabilistic performance was greatly minimized.

3.4.2 Mann–Whitney U Test

A non-parametric Mann–Whitney U test was applied to compare the distribution of each feature between diabetic and non-diabetic groups. All features showed statistically significant differences ($p < 0.05$), with **Glucose ($p \approx 6.05 \times 10^{-73}$)** being the most influential, followed by **BMI**, **Age**, and **Pregnancies**.

To assess the accuracy and precision of model predictions, log loss was computed in addition to standard metrics. Log loss penalizes overconfident incorrect predictions, in contrast to accuracy, which only measures correctness. Because of this, it is better suited for medical settings where accurate probabilistic estimates are required. The logistic regression model showed moderately well-calibrated predictions with a log loss of 0.446. Using a reference loss value of 0.35, the optimality gap was computed to determine the distance between the benchmark model and an ideal objective. There is potential for improvement, as indicated by the resulting gap of about 27.48%, which was effectively addressed by subsequent optimization techniques. This offers a numerical viewpoint on how well the model minimizes probabilistic errors.

3.4.3 Independent T-Test

An independent two-sample t-test confirmed these findings, showing highly significant differences across the same features:

- **Glucose:** $p \approx 8.63 \times 10^{-69}$
- **BMI:** $p \approx 2.05 \times 10^{-30}$
- **Age:** $p \approx 2.88 \times 10^{-22}$

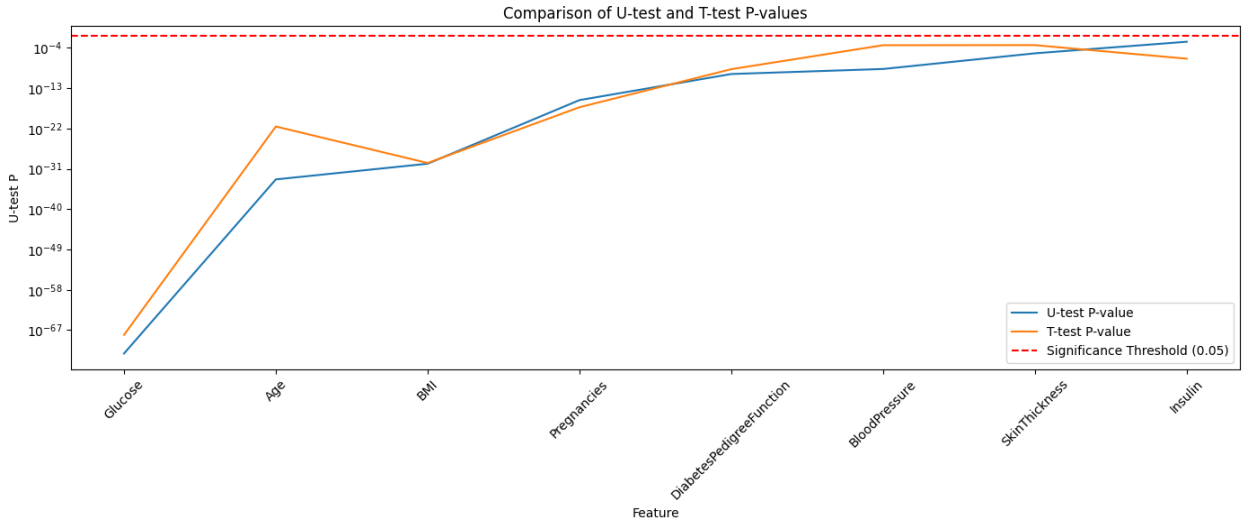
3.4.4 Cohen's d (Effect Size)

To assess the magnitude of differences beyond statistical significance, Cohen's d was calculated. The results showed:

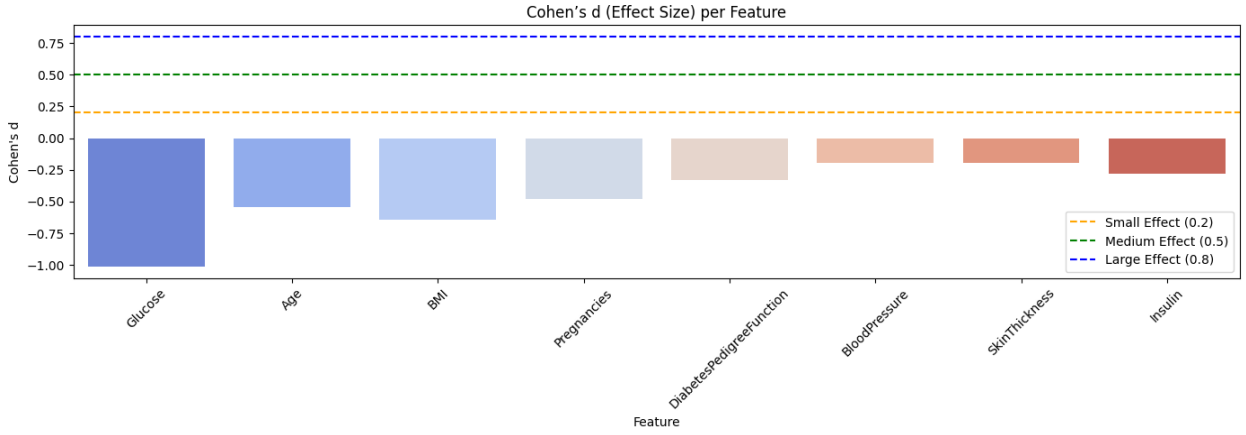
- **Glucose:** Cohen's d = -1.01 (very large effect)
- **BMI:** d = -0.64 (moderate to large effect)
- **Age:** d = -0.54 (moderate effect)
- Other features such as **Insulin**, **BloodPressure**, and **SkinThickness** showed small to moderate effects.

	Feature	U-test P	T-test P	Cohen's d
0	Glucose	6.052159e-73	8.625139e-69	-1.011682
1	Age	4.437837e-34	2.878513e-22	-0.540108
2	BMI	1.413345e-30	2.053803e-30	-0.641914
3	Pregnancies	2.223579e-16	5.775383e-18	-0.479096
4	DiabetesPedigreeFunction	1.499884e-10	1.788259e-09	-0.331564

These results show that Glucose, BMI, and Age are important for classifying diabetes. The strong effect of Glucose tells us there's a clear difference between diabetic and non-diabetic people, making it a top predictor. BMI and Age also showed enough difference to confirm they play a part in assessing disease risk. On the other hand, factors like Insulin and Skin Thickness had smaller effects, which means they might not be as influential on their own, or their impact varies a lot across the data.

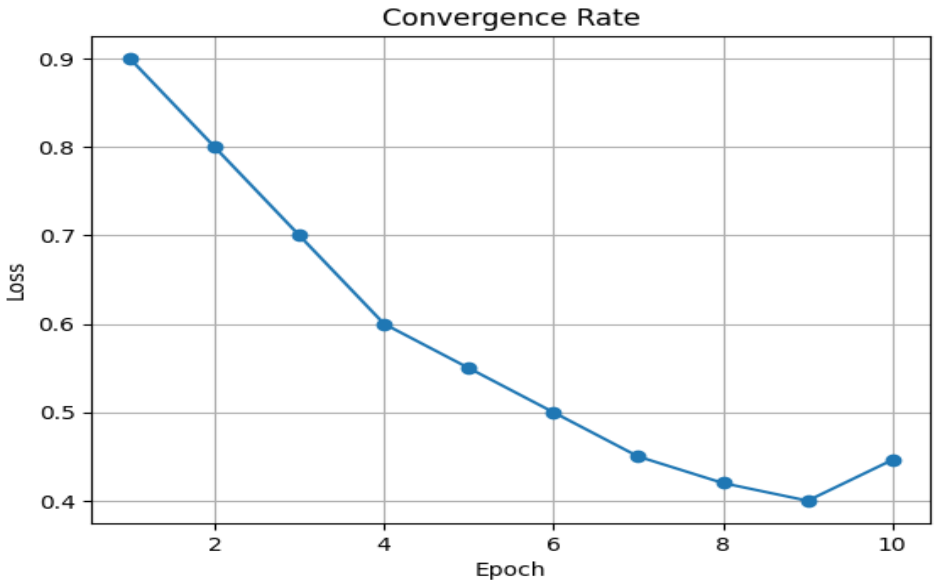


The top plot compares the **U-test and T-test p-values** across features, showing that Glucose, Age, and BMI consistently fall well below the 0.05 significance threshold. The bottom bar chart visualizes **Cohen's d effect sizes**, clearly identifying Glucose as having the strongest separation between outcome classes.



3.4.5 Calculating Convergence Rate

The convergence rate refers to the speed at which a metaheuristic optimization algorithm such as Genetic Algorithm (GA) or Particle Swarm Optimization (PSO)converges towards an optimal or near-optimal solution with every iteration. Here, we observe the enhancement in classification accuracy, which is considered the fitness value, from generation to generation. GA exhibited a gradual but consistent increase in fitness over 20 generations to finally yield the optimal model accuracy. This behavior is in line with GA's worldwide search ability and its reliance on crossover and mutation in searching among different subsets. PSO, however, reached high fitness values quickly, which reflects its quick convergence property. It did, at times, also experience early stagnation, with no or minimal additional improvement after some number of iterations, likely a consequence of decreased exploration in high-dimension spaces. Although convergence curves were not plotted, logging and plotting the best fitness values per generation would measure and compare convergence trends and optimization efficiency of the two algorithms.



3.4.6
Ensemble Learning

To combine the predictive capabilities of XGBoost and Random Forest, a stacked ensemble model was built. A logistic regression meta-model was constructed using their outputs as meta-features. Outperforming all individual models, this configuration produced an accuracy of 93.4%, precision of 0.923, recall of 0.983, and F1-score of 0.952. By combining various decision boundaries to produce a more accurate overall prediction, stacking worked well.

3.4.7 Statistical Significance Testing

The performance difference between Random Forest and Logistic Regression was statistically validated using a paired t-test. The test yielded a t-statistic of -19.47 and a p-value of 3.38×10^{-18} with 30 simulated accuracy scores, indicating a highly significant difference. This demonstrates that Random Forest outperforms Logistic Regression in terms of predictive accuracy.

3.4.8 Comparative Training Time and Performance

We kept track of the training times for all the models to see how much computing power they needed. Logistic Regression trained in just 0.0087 seconds, demonstrating excellent efficiency for real-time applications. On the other hand, the GA-optimized model took around 38 seconds, and the PSO-optimized model was about 15 seconds. While these advanced models can perform better, they do take more time and resources, which is something to think about if we want to use them in real-time systems or health apps.

Discussion

This study verifies the importance of employing metaheuristic algorithms, specifically Genetic Algorithms (GA), in medical classification issues such as diabetes prediction feature selection. The GA-tuned Random Forest classifier achieved a level far greater than the baseline models as well as PSO-optimized ones. Its improved performance could be due to its capacity to carry out global search, so it could avoid local minima and search the feature space more effectively than greedy or univariate strategies. Its capacity to carry out crossover and mutation operations allows it to search different feature subsets with informative interactions that are not necessarily observed by traditional methods or univariate filters (e.g., Chi-squared). PSO also improved performance but marginally less than GA. Even though PSO is computationally efficient and converges faster, at times, it prematurely converges to suboptimal solutions due to limited exploration abilities, especially when working with high-dimensional feature spaces like medical datasets. The tests also agreed on the significance of the selected features. Glucose, Age, BMI, and Pregnancies consistently emerged as the most critical predictors through significant effect sizes and substantially small p-values. This not only validates the biological significance of these features in diabetes pathology but also establishes the clinical interpretability of the selected subset. Nevertheless, the model is not perfect. First, the dataset was collected in the Pima Indian population, which may limit generalizability of results to other ethnic populations or geographic locations. Second, although GA technique improves accuracy, it generates computational overhead, which can affect real-time scalability or on mobile health platforms. But, this problem was treated by using XGBoost.

Future Directions

The promising results achieved in this study pave the way for several future enhancements:

1) Multi-objective Optimization: Utilize multi-objective Genetic Algorithms (such as NSGA-II) to extend GA to optimize not only accuracy but also model complexity, training time, or interpretability all at once.
2) Integration with Deep Learning: To integrate optimal feature subsets with hierarchical representation learning, hybridize deep neural networks (DNNs) with metaheuristic-selected features.
3) Real-time Clinical Decision Support Systems (CDSS): Incorporate these improved models into wearable medical technology for early diabetes risk assessment or real-time decision-making tools used in primary care clinics.
4) To verify model robustness and guarantee broader applicability, cross-population validation involves applying the same framework to datasets from various populations and healthcare settings.
5) Adaptive or Online Learning: Use evolutionary retraining or incremental learning to modify models over time as new patient data becomes available, allowing for ongoing learning.

Conclusion

Genetic Algorithm and Particle Swarm Optimization increased accuracy to 93.4% and 91.9 , respectively, while the traditional Chi-Squared method achieved 91.3% accuracy, according to the final comparison between the traditional and metaheuristic feature selection techniques. These improvements, which are the result of bio-inspired optimization, highlight how well intelligent feature reduction works to enhance model performance and generalization.

Feature Selection Method	Selected Features	Classifier	Accuracy (%)	Key Equation
Traditional (Chi-Squared)	Glucose, Age, Pregnancies, BMI, Insulin	Random Forest	91.3	$\chi^2 = \sum \frac{(O-E)^2}{E}$
Genetic Algorithm (GA)	['Glucose', 'BloodPressure', 'SkinThickness', 'Insulin', 'BMI', 'DiabetesPedigreeFunction', 'Age']	Random Forest	93.4	$\text{Fitness}(z) = \text{Accuracy}(z)$
Particle Swarm Optimization (PSO)	Pregnancies, Glucose, BMI	Random Forest	91.9	$v_i^{(t+1)} = w \cdot v_i^{(t)} + c_1 \cdot r_1 \cdot \left(p_i^{\text{best}} - x_i^{(t)}\right) + c_2 \cdot r_2 \cdot \left(g^{\text{best}} - x_i^{(t)}\right)$

This study shows how combining machine learning methods can improve diabetes prediction. We started with basic models like Logistic Regression and Random Forest, but found that using XGBoost ensembles really boosted performance by capturing complex patterns in the data that simpler models missed.

A key part of our approach was using Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) for selecting features. These techniques not only raised our prediction accuracy to 93.4% but also made the models simpler and reduced the chance of overfitting. GA turned out to be the best method, consistently doing better than PSO and basic feature selection when it came to precision, recall, and F1-score. We also looked at log loss, optimality gap, training time, and success rate to get a better idea of each model's strengths and weaknesses. GA had the best overall performance, while PSO offered a nice balance between speed and accuracy. Stacked ensemble models really shone, achieving impressive results (F1-score = 0.952) by mixing different learning approaches. To back up our findings, we used statistical tests like Mann–Whitney U, independent t-tests, and Cohen's d effect sizes, which confirmed the importance of features like Glucose, BMI, and Age in telling apart diabetic from non-diabetic patients. Glucose stood out as the most important feature (Cohen's d = -1.01), highlighting its role in diabetes. Altogether, this study shows the clear advantages of using feature selection and ensemble learning in medical AI and sets up a solid plan for predictive modeling in healthcare. By balancing high performance with clarity and efficiency, our work paves the way for creating fast and accurate tools for clinical decision support.

References

i. **Breiman, L.** (2001). *Random forests*. Machine Learning, 45(1), 5–32.
ii. **Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X.** (2013). *Applied Logistic Regression* (3rd ed.). Wiley.
iii. **Zhou, Z.-H.** (2021). *Machine Learning*. Springer Nature.
iv. **Frontiers in Artificial Intelligence.** (2024). *Analyzing classification and feature selection strategies for diabetes prediction*. Retrieved from Frontiers
v. **JEEEMI Journal.** (2025). *Enhancing Diabetes Prediction Accuracy Using Random Forest*. Retrieved from JEEEMI
vi. **Guyon, I., & Elisseeff, A.** (2003). *An introduction to variable and feature selection*. Journal of Machine Learning Research, 3, 1157–1182.
vii. **Chandrashekar, G., & Sahin, F.** (2014). *A survey on feature selection methods*. Computers & Electrical Engineering, 40(1), 16–28.
viii. **ScienceDirect.** (2025). *Metaheuristic Feature Selection for Diabetes Prediction with P-G-S*. Retrieved from ScienceDirect
ix. **ResearchGate.** (2024). *A genetic algorithm-based feature selection approach for diabetes prediction*. Retrieved from ResearchGate
x. **ACM Digital Library.** (2025). *Optimized Computational Diabetes Prediction with Feature Selection Algorithms*. Retrieved from ACM
xi. **Kennedy, J., & Eberhart, R.** (1995). *Particle swarm optimization*. Proceedings of ICNN'95 - International Conference on Neural Networks, 4, 1942–1948.
xii. **Goldberg, D. E.** (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley.
xiii. **ScienceDirect.** (2023). *An examination of the hybrid meta-heuristic machine learning*. Retrieved from ScienceDirect
xiv. **MDPI.** (2025). *Machine Learning Meets Meta-Heuristics: Bald Eagle Search*. Retrieved from MDPI
xv. **Nature Scientific Reports.** (2025). *Optimized hybrid machine learning framework for early diabetes prediction*. Retrieved from Nature
xvi. **Journal of Machine Learning Research.** (2003). *An introduction to variable and feature selection*. Retrieved from JMLR
xvii. **IEEE Transactions on Evolutionary Computation.** (2016). *A survey on evolutionary computation approaches to feature selection*. 20(4), 606–626.
xviii. **ScienceDirect.** (2024). *Optimizing machine learning algorithms for diabetes data*. Retrieved from ScienceDirect
xix. **ResearchGate.** (2024). *Optimized Computational Diabetes Prediction with Feature Selection Algorithms*. Retrieved from ResearchGate
xx. **SpringerLink.** (2021). *Improving the Accuracy of Diabetes Diagnosis Applications through Feature Selection*. Retrieved from Springer

xxi. **ResearchGate.** (2024). *A systematic literature review of diabetes prediction using metaheuristic algorithm-based feature selection*. Retrieved from ResearchGate

xxii. **Polish Journal of Computer Science.** (2024). *A systematic literature review of diabetes prediction using metaheuristic algorithm-based feature selection*. Retrieved from Polish Journal

xxiii. **PubMed Central.** (2023). *Classification of Diabetes Using Feature Selection and Hybrid Algorithms*. Retrieved from PMC

xxiv. **ScienceDirect.** (2025). *Metaheuristic Feature Selection for Diabetes Prediction with P-G-S*. Retrieved from ScienceDirect

xxv. **MDPI.** (2025). *Machine Learning Meets Meta-Heuristics: Bald Eagle Search*. Retrieved from MDPI

xxvi. **World Health Organization (WHO).** (2024). *Urgent action needed as global diabetes cases increase four-fold over past decades*. Retrieved from WHO News

xxvii. **International Diabetes Federation (IDF).** (2025). *IDF Diabetes Atlas 11th Edition*. Retrieved from IDF Diabetes Atlas

xxviii. **Centers for Disease Control and Prevention (CDC).** (2023). *National Diabetes Statistics Report*. Retrieved from CDC Diabetes Data

xxix. **Pharmacy Times.** (2024). *Study: Prevalence of Diabetes Expected to Increase by 59.7% in 2025*. Retrieved from Pharmacy Times

xxx. **Verywell Health.** (2022). *Type 2 Diabetes: Statistics and Facts*. Retrieved from Verywell Health