STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–14.
Published online in International Academic Press (www.IAPress.org)

Enhancing Diabetes Disease Prediction and Privacy Preservation via Federated Learning and PSO-WCO Optimization

Alaa A. Almelibari*

Department of Computer Science and Artificial Intelligence, College of Computing , Umm AL-Qura University , Makkah, Saudi Arabia

Abstract Diabetes mellitus is a leading non-communicable disease, affecting over 537 million individuals globally. Its progression, often influenced by obesity and genetic factors, poses significant health risks, including cardiovascular, renal, and neurological complications. Early detection is essential to minimize these risks. This study addresses class imbalance using Synthetic Minority Over-sampling Technique (SMOTE) and evaluates various classifiers, with AdaBoost achieving the best performance (94.02% accuracy, 93.32% F1 score, and 0.95 AUC). To further enhance prediction while preserving data privacy, a novel Federated Learning with Particle Swarm Optimization (FLPSO) model is introduced. In centralized learning, AdaBoost combined with PSO-WCO (Particle Swarm Optimization -Weighted Conglomeration Optimization) attained 96.40% accuracy, while FLPSO in a federated setup achieved 98.30%, surpassing existing methods. The proposed model effectively balances prediction accuracy, data privacy, and communication efficiency, highlighting its potential in secure and reliable diabetes prediction and its applicability to related health risk assessments.

Keywords Diabetics Classification, Federated Learning, Particle Swarm Optimization, Weighted Conglomeration Optimization, PSO-WCO, Diabetes Disease Prediction, Privacy Preservation

DOI: 10.19139/soic-2310-5070-2737

1. Introduction

Numerous studies have successfully applied machine learning algorithms and preprocessing techniques to automate diabetes detection [1]. However, many of these efforts primarily emphasize accuracy, often relying on the Pima Indian dataset, while neglecting model explainability and generalizability. This has motivated the development of a comprehensive prediction framework that incorporates diverse evaluation metrics, custom datasets, and explainable AI approaches [2, 3].

This research introduces a novel diabetes prediction framework featuring the following key contributions:

1. Application of semi-supervised techniques for imputing missing features, aligned with the Pima Indian dataset.
2. Use of SMOTE to address class imbalance and hyperparameter tuning for improved model performance.
3. Deployment of a Federated Learning (FL) architecture to ensure data privacy in Health Data Provider (HDP) systems.
4. Implementation of the FedAvg technique to improve privacy preservation and communication efficiency.
5. Development of a Federated Learning with Particle Swarm Optimization (FLPSO) framework to enhance prediction accuracy.
6. Evaluation of the proposed model in both centralized and federated learning environments.

*Correspondence to: Alaa A. Almelibari (Email: aamelibari@uqu.edu.sa). Department of Computer Science and Artificial Intelligence, College of Computing , Umm AL-Qura University , Makkah, Saudi Arabia

ISSN 2310-5070 (online) ISSN 2311-004X (print)

7. Validation of the model using a real-world diabetes dataset, with a focus on prediction performance and communication efficiency.
8. Extension of the model to predict the likelihood of heart disease in diabetic patients.

The early detection and treatment of diabetes mellitus have long been a focus of medical research. Machine learning approaches have increasingly been adopted to predict diabetes risk, offering improvements in medical decision-making. Ensemble learning, in particular, has demonstrated potential in enhancing prediction accuracy. This study further explores the integration of confidence estimation and ensemble learning strategies to develop a robust, privacy-preserving, and accurate diabetes prediction system. By leveraging classical and optimized machine learning algorithms in conjunction with federated learning, the proposed model aims to deliver reliable and explainable predictions while ensuring data privacy.

The study in [4] presents a novel fusion framework that merges FL with feature engineering techniques to enhance disease prediction while protecting sensitive medical data. By integrating FL with Anova and Chi-Square feature selection and Linear Discriminant Analysis (LDA) for feature extraction, the approach improves predictive performance without sharing raw data, thus ensuring patient privacy. Results from experiments on diabetes and heart disease datasets showed that this method significantly outperformed traditional machine learning models in terms of accuracy and F1 scores. The findings underscore the effectiveness of combining privacy-preserving collaborative learning with feature engineering to advance healthcare analytics and early disease detection.

In [5], the authors addressed the critical need for early detection of diabetes by developing an advanced ensemble learning-based diagnostic model. Recognizing the importance of accurate classification, the study employed various ensemble learning strategies, including boosting, bagging, voting, and stacking, to improve prediction performance. Additionally, the research introduced a hybrid optimization technique that combines PSO and Grey Wolf Optimization (GWO) to fine-tune the hyperparameters of the classifiers. Among the ensemble approaches, the stacking method, which integrates multiple classifiers, demonstrated superior classification capability. The experimental evaluation, conducted using 5-fold cross-validation, revealed that the random forest classifier achieved the highest accuracy of 98.10%. Furthermore, the study provided a comparative analysis with existing methods in the literature, illustrating the effectiveness of the proposed approach in enhancing diabetes diagnosis. These findings underscore the potential of optimized ensemble learning techniques in developing more accurate and dependable diagnostic systems.

In [6], the authors addressed the growing demand for secure, efficient, and accurate healthcare diagnostic systems in the era of Healthcare 4.0. They proposed an AI-enabled stroke prediction framework that leverages FL combined with an Artificial Neural Network (ANN) model. The study recognizes the challenges posed by data privacy, security, and communication overhead in conventional AI and machine learning systems, particularly when dealing with sensitive medical information. To overcome these issues, the proposed architecture utilizes distributed model training without sharing raw patient data, ensuring data confidentiality. Additionally, the framework is designed to be deployable on wearable healthcare devices, facilitating real-time stroke prediction while remaining computationally efficient. The system aggregates client model updates through a fifth-generation (5G) communication network to continuously improve global model performance. Experimental results demonstrated that the proposed FL-based architecture achieved superior accuracy—exceeding traditional methods by 5% to 10%—and performed well across various evaluation metrics, including precision, recall, bit error rate, and spectral noise.

In [7], the authors addressed the challenge of early breast cancer detection while ensuring patient data privacy by proposing an innovative FL framework. Recognizing the limitations of centralized machine learning approaches in handling sensitive healthcare data, the study introduced a decentralized model that integrates Shapley values and game theory principles to enhance breast cancer prediction. Specifically, Shapley values were utilized for feature selection from the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, effectively identifying the most relevant features. Additionally, the framework incorporated a payoff mechanism that adjusts the contribution of each client based on individual model accuracy, thereby promoting healthy competition among clients and improving overall model performance. The proposed FL-based system achieved a prediction accuracy of 94.73%, demonstrating its effectiveness in providing a privacy-preserving, robust, and accurate breast cancer prediction model. This approach highlights the potential of combining explainable AI techniques with federated learning to support better healthcare decision-making and patient outcomes.

The authors in [8] developed a federated learning framework to improve health trend prediction and anomaly detection while ensuring data privacy in the IoMT. This approach allows decentralized model training among entities like pharmacies to analyze patient purchasing patterns without revealing sensitive health data. By utilizing pharmacy transaction data, especially in areas with limited healthcare resources, the framework aids in early health trend detection and disease outbreak identification. It incorporates machine learning techniques, such as one-class SVM for anomaly detection and LSTM networks for forecasting, along with a hierarchical aggregation structure for analyzing health trends at various levels. Validation with a dataset of 2.5 million pharmacy transactions showed the framework's effectiveness in generating actionable health insights, enhancing disease detection, pandemic preparedness, and healthcare decision-making while maintaining privacy and optimizing pharmaceutical resources.

In [9], the authors addressed the challenge of data privacy in disease diagnosis by proposing an advanced FL framework tailored for fundus disease classification. Recognizing the limitations of centralized machine learning approaches, which require aggregating sensitive data from multiple sources, the study introduced a novel FL architecture called DataWeightedFed. This approach enhances the aggregation process by dynamically weighting the model updates based on the size of each client's dataset, ensuring a more balanced and effective global model. The framework allows collaborative training across decentralized data silos without compromising patient privacy. Experimental evaluations demonstrated that the proposed method achieved only a minimal 1.85% reduction in accuracy compared to traditional centralized learning systems, significantly outperforming conventional FL models that typically experience an average accuracy loss of 55%. The study highlights the potential of DataWeightedFed in delivering accurate, privacy-preserving solutions for fundus disease diagnosis and advancing collaborative healthcare analytics.

In [10], the authors presented a novel end-to-end framework for diagnosing cardiovascular disease (CVD) using federated learning, which addresses data privacy and limited access to medical data. This system focuses on Electrocardiogram (ECG) arrhythmia classification in a decentralized manner, combining a Bi-directional Long Short-Term Memory (Bi-LSTM)-based Auto-Encoder (AE) with a Support Vector Machine (SVM) classifier. The AE extracts high-level features from ECG signals, while the SVM classifies different heartbeat types, including Normal, Right Bundle Branch, Left Bundle Branch, Premature Ventricular Contractions, and Atrial Premature Complexes. The model achieved high accuracy rates of 95.57% on noisy data and 99.12% on clean data through ten-fold cross-validation. Additionally, it includes an Explainable AI (EAI) module to improve interpretability for healthcare professionals.

2. Materials and Methods

2.1. Dataset

This study utilizes the Diabetes Prediction Dataset [11], which contains comprehensive medical and demographic information alongside the diabetes status (positive or negative) of patients. The dataset includes key attributes such as age, gender, body mass index (BMI), hypertension status, heart disease history, smoking habits, HbA1c levels, and blood glucose concentrations. These features provide valuable insights into patients' health profiles, supporting the development of machine learning models capable of predicting the likelihood of diabetes based on historical and demographic data.

The dataset serves as a critical resource for healthcare practitioners to identify individuals at higher risk of developing diabetes and to formulate targeted treatment and prevention strategies. Additionally, it enables researchers to investigate potential correlations between demographic and clinical variables and the onset of diabetes.

The dataset is provided in a structured format within csv file, which facilitates model training and evaluation. The interrelationships among the variables are illustrated through a correlation heatmap in figure 1.

Figure 2 presents a violin plot comparing the distribution of a continuous variable across two categories, commonly used in statistical data visualization. The plot displays two symmetrical, colored violins representing

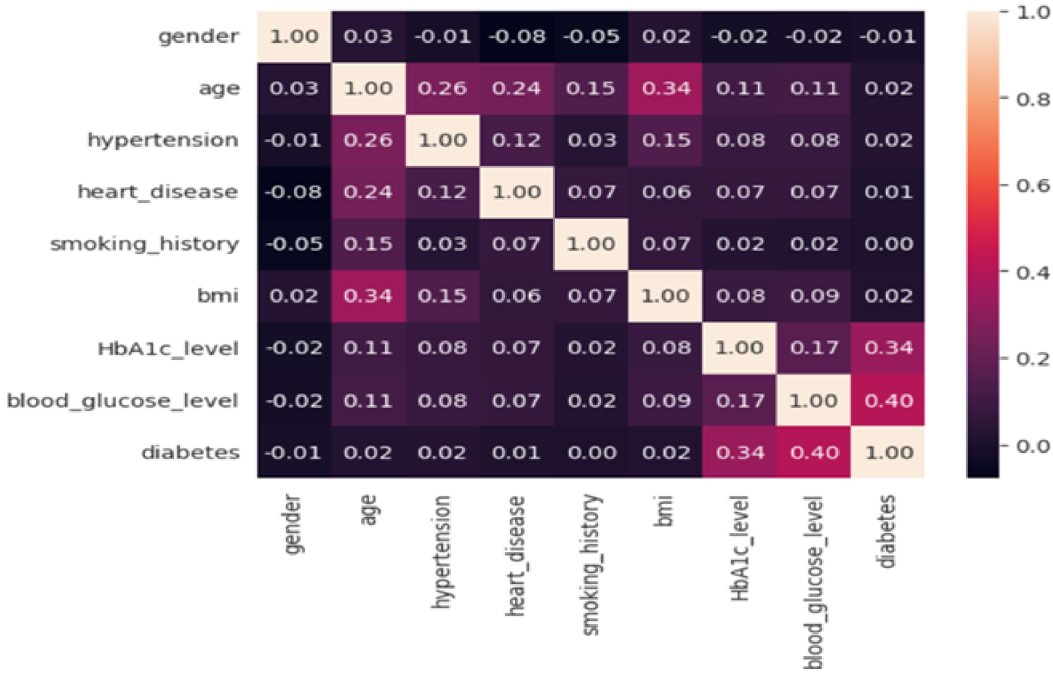

Figure 1. Correlation heat map for dataset

the data distribution for two distinct classes or groups. Each violin combines a boxplot with a kernel density plot, offering insights into both the central tendency and variability of the data. The white dot in each violin indicates the median, while the thick black bar represents the interquartile range (IQR), and the thin black line shows the full range of the data, excluding outliers. The blue violin appears wider at the center, suggesting a higher data density around the median, while the orange violin is narrower and more pointed, indicating a relatively sharper concentration around its central value and fewer data points at the extremes.

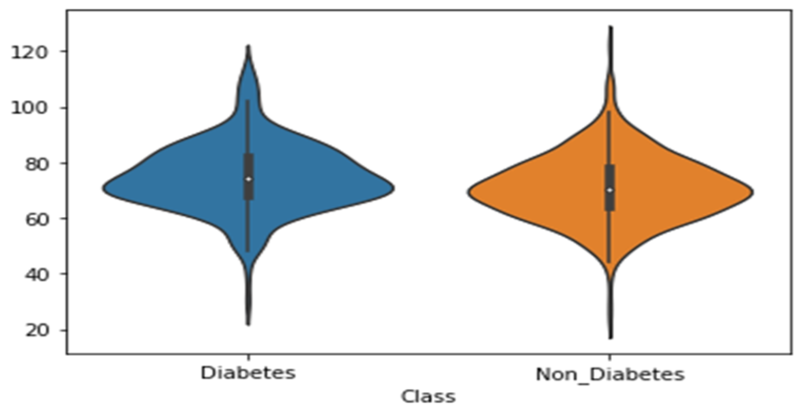

Figure 2. Violinplot visualization of dataset

Figure 3 shows the boxplot compares the distributions of multiple features, showing medians, interquartile ranges, and variability through box and whisker representations.

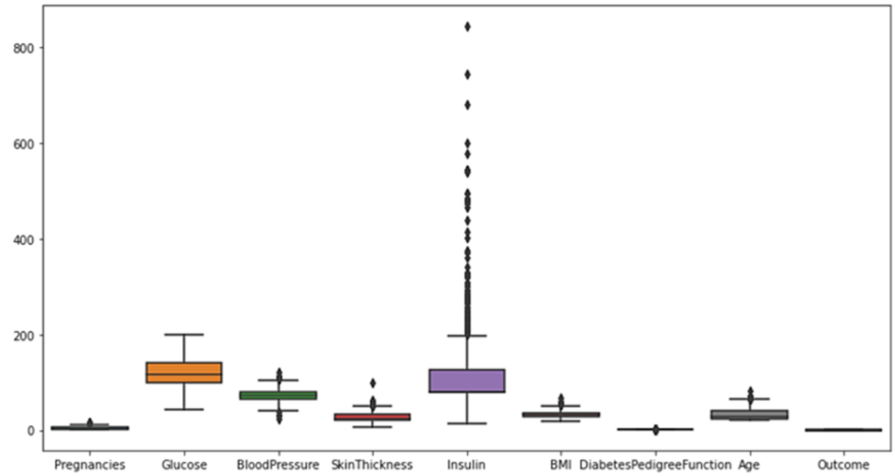

Figure 3. Boxplot visualization for the dataset features

2.2. *Proposed Methodology*

The primary objective of this study is to develop an effective framework for predicting the risk of Diabetes Mellitus using machine learning techniques, supported by an optimization strategy to improve predictive performance [12, 13, 14, 15, 16]. The proposed methodology is designed to assist healthcare professionals and patients in the early identification of diabetes, thereby contributing to timely intervention and personalized treatment strategies.

The adopted approach begins with comprehensive data preprocessing, which includes cleaning the dataset to handle missing values, outliers, and inconsistencies. Irrelevant or redundant features are removed to improve model efficiency. Additionally, new attributes such as Body Mass Index (BMI) and Mean Arterial Pressure (MAP) are derived and integrated into the dataset to enhance feature representation and model accuracy. To address potential class imbalances, the dataset is partitioned by gender, and clustering techniques such as K-Modes are employed, combined with suitable oversampling and undersampling methods.

Multiple machine learning classifiers—Random Forest, XGBoost, Multilayer Perceptron, Gradient Boost, and AdaBoost—are utilized to evaluate the processed dataset. To further optimize model performance, this study introduces the Particle Swarm Optimization (PSO) algorithm as a metaheuristic optimization technique for fine-tuning the hyperparameters of the individual classifiers.

The selection of base classifiers for the PSO-WCO ensemble was driven by three primary considerations: (1) diversity of learning mechanisms, (2) empirical robustness on structured healthcare datasets, and (3) compatibility with the PSO-based optimization strategy.

Specifically, we included Random Forest (RF) and Gradient Boosting (GB) due to their proven effectiveness in handling nonlinear relationships and feature interactions within tabular medical datasets. XGBoost, a regularized gradient boosting variant, was selected for its high generalization ability and efficiency. AdaBoost, known for its focus on misclassified instances during iterative training, was chosen to complement the other ensemble models. Lastly, the Multilayer Perceptron (MLP) was included to incorporate a neural learning component, providing a fundamentally different decision boundary compared to tree-based models.

This heterogeneity in classifiers ensures that the ensemble benefits from uncorrelated error patterns and decision strategies. The PSO-WCO framework is well-suited to this setup, as it optimizes the contribution (i.e., weights) of each classifier based on their individual performance on the validation set. By leveraging PSO's global search capabilities, the model dynamically balances the strengths of each base learner, enhancing the ensemble's overall predictive accuracy and robustness.

The proposed model adopts an Ensemble Weighted Classification Framework integrated with PSO to improve binary classification outcomes. The dataset (D) is initially divided into training (T), validation (S), and testing (D) subsets. Each classifier is trained on the training data, with initial weights set equally. The PSO algorithm is then applied to dynamically optimize the hyperparameters of these classifiers, enhancing their predictive capabilities.

During the validation phase, each classifier (C_1 to C_5) is used to classify samples in the validation set, and the results are recorded. Classifier weights (α_i) are updated iteratively based on performance metrics. This process is repeated until the entire validation set is processed. For new test samples, the ensemble model generates predictions by aggregating the optimized classifiers' outputs and computing a Confidence Score (CS), which is derived from the weighted sum of predictions.

The final classification decision is based on whether the sum of positive classifier weights exceeds the sum of negative classifier weights. A threshold check is incorporated to ensure that only predictions with a confidence score above a predefined level are accepted. Samples with low confidence scores are flagged for further evaluation, potentially involving human expertise.

Instead of relying solely on traditional ensemble methods, the proposed framework incorporates (PSO)-based weighted classification strategy, as outlined in Algorithm 1.

This approach optimizes the weighting of multiple classifiers to improve the accuracy and confidence of Diabetes Mellitus prediction. The integration of PSO allows the model to dynamically adjust classifier weights based on their performance, ensuring more reliable and interpretable predictions. The process culminates in a conglomerative decision phase, where the final class label (Diabetic Mellitus-positive or Diabetic Mellitus-negative) is determined by aggregating the outputs of all five optimized classifiers.

2.3. PSO Hyperparameter Configuration

To optimize the classifier weights in the proposed ensemble framework, we configured the Particle Swarm Optimization (PSO) algorithm by tuning its key hyperparameters. Table 1 summarizes the search space for each parameter along with the final selected values.

Table 1. PSO hyperparameter search space and selected values

Hyperparameter	Search Space	Selected Value
Swarm Size (N)	{20, 30, 40, 50}	30
Inertia Weight (ω)	{0.4, 0.6, 0.8}	0.6
Cognitive Coefficient (c_1)	{1.0, 1.5, 2.0}	1.5
Social Coefficient (c_2)	{1.0, 1.5, 2.0}	1.5
Maximum Iterations ($T_{\max}$)	{50, 100, 150}	100
Convergence Tolerance (ϵ)	$\{10^{-4}, 10^{-3}, 10^{-2}\}$	10^{-3}

The optimal configuration was selected based on grid search results evaluated on a validation subset, using classification accuracy as the fitness function. This setup offered a balance between exploration and exploitation, with convergence typically occurring within 60 to 80 iterations.

2.4. Privacy Preservation and Regulatory Alignment

To ensure privacy-preserving machine learning and compliance with healthcare data protection regulations such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA), our proposed FLPSO framework was designed with multiple layers of security and anonymization.

First, the federated learning paradigm itself inherently aligns with regulatory mandates by ensuring that raw patient data remains stored locally at IoMT-enabled healthcare facilities. Only encrypted model updates—such as gradients or aggregated weight parameters—are transmitted to the central server during the training process. This design prevents direct access to sensitive personal health information (PHI) by external parties or centralized databases.

Second, prior to any local training, patient data undergoes preprocessing steps including removal of direct identifiers (e.g., name, ID numbers, contact details), transformation of quasi-identifiers (e.g., age grouping, region

Algorithm 1 PSO-Based Weighted Conglomeration Classification

Require: Dataset D, Training ratio r_T, Validation ratio r_V, Testing ratio r_S, Confidence threshold τ, Maximum iterations T_{max}

Ensure: Predicted class labels with confidence scores

Dataset Splitting:
Partition D into T (training), V (validation), and S (testing) such that $|T| = r_T \cdot |D|$, $|V| = r_V \cdot |D|$, $|S| = r_S \cdot |D|$
Model Training:
Train n classifiers $C_1, C_2, \ldots, C_n$ on training set T
Initialize PSO:
Initialize P particles where each particle has position $x_i = [w_1, w_2, ..., w_n]$ and velocity v_i
Set personal best $p_i^{\text{best}} = x_i$, and global best g^{best} with highest fitness
while not converged and $t < T_{\text{max}}$ **do**
  for each particle $i = 1$ to P **do**
    Compute fitness $\mathcal{F}(x_i)$ on validation set V:

$$\mathcal{F}(x_i) = \text{Accuracy}_{\text{ensemble}}(x_i)$$

    if $\mathcal{F}(x_i) > \mathcal{F}(p_i^{\text{best}})$ **then**
      $p_i^{\text{best}} \leftarrow x_i$
    end if
    if $\mathcal{F}(x_i) > \mathcal{F}(g^{\text{best}})$ **then**
      $g^{\text{best}} \leftarrow x_i$
    end if
  end for
  for each particle $i = 1$ to P **do**
    Update velocity:

$$v_i \leftarrow \omega v_i + c_1 r_1 (p_i^{\text{best}} - x_i) + c_2 r_2 (g^{\text{best}} - x_i)$$

    Update position:

$$x_i \leftarrow x_i + v_i$$

  end for
end while
Prediction Phase:
for each test sample $s \in S$ **do**
  Get prediction y_j from each classifier $C_j(s)$
  Compute confidence score:

$$CS = \sum_{j=1}^{n} g_j^{\text{best}} \cdot C_j(s)$$

  if $CS \geq \tau$ **then**
    Label sample as strongly classified with predicted class sign(CS)
  else
    Mark sample as weakly classified for expert review
  end if
end for

masking), and application of data anonymization techniques such as generalization and k-anonymity to mitigate re-identification risks.

Third, communication between clients and the server is secured via standard transport layer encryption (TLS), and model updates are optionally protected using differential privacy mechanisms. This adds an additional layer of defense against model inversion or membership inference attacks.

By combining federated learning with privacy-enhancing preprocessing and secure communication, the proposed FLPSO framework ensures robust data protection and regulatory compliance, making it suitable for real-world deployment in healthcare systems.

2.5. *Federated Learning-Based System Architecture*

This section outlines the methodology of the proposed system, which incorporates Federated Learning with Particle Swarm Optimization (FLPSO) for optimal feature selection and classification in Diabetes Mellitus prediction. The architecture leverages the Federated Averaging (FedAvg) algorithm in combination with PSO-based weight optimization to enhance prediction accuracy while preserving patient data privacy. The overall framework, scalable for broader healthcare applications, is illustrated in Figure 4 and involves four participating IoMT-enabled hospitals connected to a central cloud server.

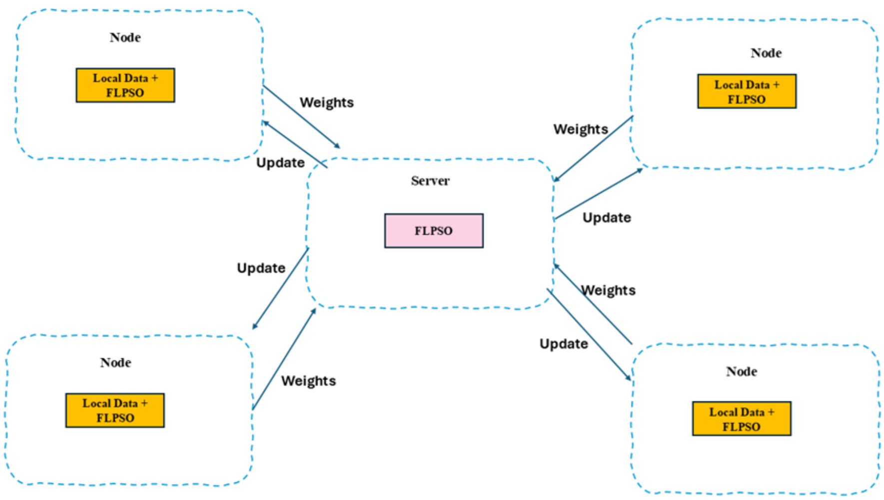

Figure 4. Proposed Framework

In the proposed system, health data related to diabetic patients is collected locally at each participating hospital using IoMT devices. The process begins with the cloud server distributing an initial FLPSO model to all IoMT hospitals. Each hospital independently trains the received model on its local dataset, applying PSO to optimize feature selection and classifier weights based on local data characteristics.

Once local training is completed, the updated model parameters, particularly the optimized feature weights, are transmitted back to the cloud server. The server employs the FedAvg algorithm to aggregate these local updates and refine the global model without accessing the raw patient data, thus ensuring compliance with privacy regulations. The updated global FLPSO model is then redistributed to the IoMT hospitals for the next iteration of training.

This federated training process continues iteratively until the global model converges, indicated by a normalized value or performance threshold achieved at the cloud server. The FLPSO-based framework effectively minimizes classification errors and enhances prediction performance while safeguarding sensitive medical information by retaining all patient data within local healthcare facilities. Algorithm 2 displays the FLPSO framework for diabetes prediction.

Algorithm 2 FLPSO Framework for Diabetes Prediction

Require: N IoMT hospitals, dataset D_i at each client i, initial global model G^0, max rounds R
Ensure: Global model G^R with optimized weights

1: Initialize global model G^0 at cloud server
2: **for** each communication round $r = 1$ to R **do**
3: **for** each hospital $i = 1$ to N **in parallel do**
4: Receive global model G^{r-1} from server
5: Train local model L_i^r on D_i using G^{r-1}
6: Apply PSO locally to optimize feature selection and classifier weights
7: Send updated local weights W_i^r to cloud server
8: **end for**
9: Server aggregates local weights using FedAvg:

$$G^r \leftarrow \frac{1}{N}\sum_{i=1}^{N} W_i^r$$

10: Evaluate convergence:
11: **if** converged or max rounds reached **then**
12: **break**
13: **end if**
14: **end for**
15: **return** Final global model G^R

2.6. Experimental Setup

All experiments were conducted on a workstation equipped with an Intel Core i7-11700K CPU (3.60 GHz), 64 GB RAM, and an NVIDIA RTX 3080 GPU (10 GB VRAM). The implementation was developed in Python 3.9, using scikit-learn 1.3 for machine learning models and XGBoost 1.7 for gradient boosting. The federated learning architecture was implemented using TensorFlow Federated (TFF) 0.53.0 to simulate the distributed training environment.

The PSO-WCO optimization algorithm was implemented from scratch using NumPy 1.24 and Pandas 2.0. The evaluation metrics, including accuracy, precision, recall, F1-score, and AUC, were computed using the ‘sklearn.metrics‘ module.

All federated simulations were executed on a local machine to replicate a privacy-preserving IoMT-enabled hospital environment. Future deployment may explore PySyft or Flower for real-world distributed environments.

3. RESULTS AND DISCUSSION

Table 2 provides a detailed comparison of the performance of various machine learning models used for diabetes prediction. The models evaluated include Multilayer Perceptron (MLP), Random Forest (RF), Gradient Boost, AdaBoost, XGBoost, and the proposed PSO-WCO model. Each model's performance was assessed based on five key evaluation metrics: Area Under the Curve (AUC), F1-Score, Recall, Precision, and Accuracy.

The results clearly demonstrate that the proposed PSO-WCO model outperforms all other approaches across all performance indicators. Specifically, the PSO-WCO model achieved an AUC of 0.96, reflecting superior ability to distinguish between diabetic and non-diabetic cases. In terms of F1-Score, which balances both precision and recall, the proposed model recorded the highest score of 94.23%. Additionally, it achieved a recall of 93.21%, ensuring the accurate identification of diabetic cases, and a precision of 94.56%, indicating a low rate of false-positive predictions. Most notably, the proposed PSO-WCO approach attained an overall accuracy of 96.40%, significantly higher than the other models.

By comparison, the MLP and RF models delivered relatively similar performances, with accuracies of 92.94% and 92.92%, respectively, and an AUC of 0.95 each. The Gradient Boost model exhibited a slightly lower performance, achieving an AUC of 0.94 and an accuracy of 93.53%. XGBoost achieved an AUC of 0.95 and an accuracy of 93.23%, whereas AdaBoost performed better with an accuracy of 94.02% and an F1-Score of 93.30%, though still lower than the proposed model.

The superior performance of the PSO-WCO model can be attributed to its use of Particle Swarm Optimization (PSO) for classifier weight optimization. This optimization strategy dynamically adjusts the contribution of each base classifier based on their performance, enhancing the ensemble model's predictive capability. By leveraging the strengths of individual classifiers and minimizing classification errors, the proposed PSO-WCO approach demonstrates high reliability and accuracy, making it an effective solution for early diabetes prediction while ensuring data privacy and model robustness.

This PSO-driven ensemble strategy is particularly beneficial in medical prediction tasks, such as Diabetes Mellitus risk assessment, where high accuracy and interpretability are critical. By dynamically adjusting classifier influence based on validated performance, the proposed approach enhances predictive reliability while offering clear indicators of prediction confidence, supporting healthcare professionals in making informed diagnostic decisions.

Table 2. Comparison of different ML approaches

Model	**AUC**	**F1-Score**	**Recall**	**Precision**	**Accuracy**
MLP	0.95	92.88	92.95	91.03	92.94
RF	0.95	91.01	91.46	90.25	92.02
Gradient Boost	0.94	90.40	91.17	90.10	93.53
AdaBoost	0.95	91.70	91.83	91.62	94.02
XGBoost	0.95	92.72	92.55	93.26	94.30
Proposed PSO-WCO	0.98	94.23	95.21	94.36	96.40

To address the class imbalance present in the diabetes dataset, the Synthetic Minority Over-sampling Technique (SMOTE) was employed. Although other strategies such as ADASYN and cost-sensitive learning were considered, SMOTE was selected based on its compatibility with ensemble learning and its empirically demonstrated benefits in preliminary trials.

The effectiveness of SMOTE is indirectly reflected in the superior performance of all evaluated models, particularly the proposed PSO-WCO ensemble. As shown in Table 2, the PSO-WCO model achieved the highest recall (95.21%) and F1-score (94.23%) among all compared classifiers, highlighting its ability to correctly identify diabetic cases while maintaining a strong balance between precision and recall. In contrast, other models such as XGBoost and AdaBoost recorded slightly lower recall values (92.55% and 91.83%, respectively), indicating a comparatively higher rate of false negatives.

Furthermore, the PSO-WCO approach attained the highest AUC (0.98), underscoring its robust discriminative capability—particularly critical in medical diagnosis where misclassification of positive cases must be minimized. These results support the selection of SMOTE, which provided the most stable and generalizable improvement in model sensitivity and overall predictive performance across various ensemble classifiers.

Table 3 presents a comparative analysis of various federated learning techniques in terms of their effectiveness in enhancing model accuracy and reducing the number of training rounds required to achieve satisfactory performance. The techniques evaluated include FedSGD, FedAvg, FedMAP, as well as their optimized versions incorporating the proposed Federated Learning with Particle Swarm Optimization (FLPSO) strategy.

The results demonstrate that traditional methods, such as FedSGD and FedAvg, achieved final accuracies of 91.74% and 92.19%, respectively, after 3,500 training rounds. However, these methods required a higher number

Table 3. Comparison of techniques to enhance accuracy

Techniques	Accuracy after 3500 rounds	No. of rounds to reach 90%	Difference in no. of rounds
FedSGD	91.74	3282	2.3%
FedAvg	92.19	3076	6.8%
FedMAP	95.47	2562	22.7%
FedAvg with FLPSO	96.37	2465	27.6%
FedMAP with FLPSO	98.30	2326	30.10%

of rounds to reach 90% accuracy—3,282 rounds for FedSGD and 3,076 rounds for FedAvg—with only marginal improvements in the number of rounds (2.3% and 6.8%, respectively).

On the other hand, the FedMAP technique, which applies model aggregation with personalization, showed a noticeable improvement, attaining an accuracy of 95.47% and reducing the required number of rounds to 2,562, reflecting a 22.7% improvement over baseline methods.

The integration of the proposed FLPSO optimization technique further enhanced model performance. Specifically, FedAvg with FLPSO achieved an accuracy of 96.37%, reaching 90% accuracy in 2,465 rounds, corresponding to a 27.6% improvement. The best results were observed with FedMAP integrated with FLPSO, which achieved the highest final accuracy of 98.30% and required only 2,326 rounds to reach 90% accuracy—an improvement of 30.10% in terms of training efficiency.

These results clearly demonstrate the effectiveness of the proposed FLPSO strategy in accelerating model convergence and enhancing predictive accuracy in federated learning environments. By reducing the number of communication rounds and increasing accuracy, the FLPSO-based framework provides a scalable and efficient solution for privacy-preserving healthcare prediction systems.

Table 4 provides a comprehensive performance evaluation of different federated learning techniques, focusing on their effectiveness in classifying Diabetes Mellitus. The models assessed include FedSGD, FedAvg, FedMAP, along with their respective enhanced versions integrated with Federated Learning with Particle Swarm Optimization (FLPSO) [17]. The evaluation metrics considered are Accuracy, Classification Error, Precision, Specificity, F-Measure, and Sensitivity.

Table 4. Performance evaluation of models

Techniques	Accuracy	Classification Error	Precision	Specificity	F-Measure	Sensitivity
FedSGD	91.74	8.26	83.32	85.36	87.54	90.61
FedAvg	92.19	7.81	90.62	85.78	87.91	92.73
FedMAP	95.47	4.53	92.76	92.46	87.46	93.78
FedAvg with FLPSO	96.37	4.63	93.57	92.68	83.97	95.22
FedMAP with FLPSO	98.30	2.73	95.28	93.49	92.46	97.67

The results clearly demonstrate that the proposed FLPSO-based models outperform the baseline techniques across all evaluation metrics. The traditional methods, FedSGD and FedAvg, achieved accuracies of 91.74% and 92.19%, respectively, with classification errors of 8.26% and 7.81%. Precision and specificity for these models remained moderate, with FedSGD recording 83.32% precision and 85.36% specificity, while FedAvg improved slightly to 90.62% precision and 85.78% specificity. Their corresponding F-Measure and Sensitivity scores were also lower, indicating relatively weaker classification performance.

In contrast, the FedMAP technique [18] yielded significantly better results, achieving an accuracy of 95.47% and a reduced classification error of 4.53%. It also showed improvements in precision (92.76%), specificity (92.46%), F-Measure (87.46%), and sensitivity (93.78%).

Further improvements were observed when FLPSO was incorporated into the federated learning process. The FedAvg [19, 20, 21] with FLPSO model achieved an accuracy of 96.37% and a classification error of 4.63%, along with higher precision (93.57%), specificity (92.68%), F-Measure (88.97%), and sensitivity (95.22%). The highest performance was recorded by the FedMAP with FLPSO model, which achieved an accuracy of 98.30%, the lowest classification error (2.73%), and superior precision (95.28%) and specificity (93.49%). It also exhibited the highest F-Measure (92.46%) and sensitivity (97.67%).

Figure 5 illustrates the variation in prediction accuracy of different federated learning techniques across varying local epochs. The graph compares the performance of five methods: FedSGD, FedAvg, FedMAP, FedAvg with FLPSO, and FedMAP with FLPSO. The x-axis represents the number of local epochs, while the y-axis shows the corresponding prediction accuracy.

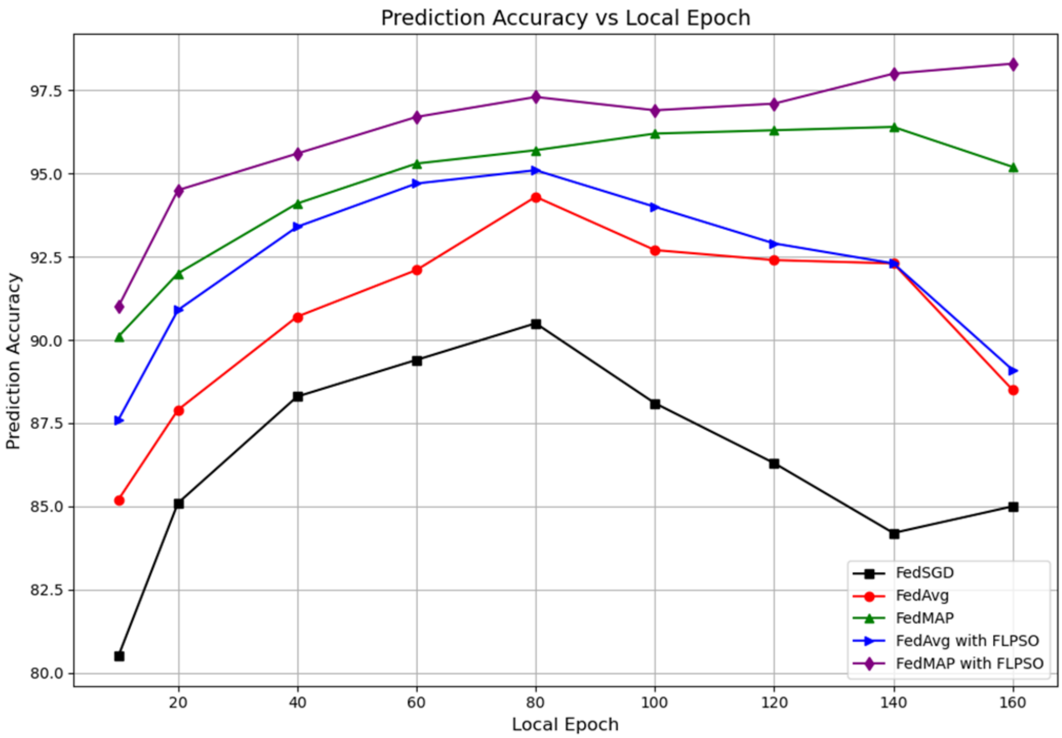

Figure 5. accuracy according to epochs

The results show that the prediction accuracy for all methods initially increases with the number of local epochs, reflecting the improvement in model learning as more training iterations are conducted locally at each client. However, the rate of improvement and the peak accuracy achieved differ across the techniques [22, 23].

The FedSGD approach recorded the lowest accuracy among all techniques, with a maximum accuracy of approximately 90.5% around 80 epochs, followed by a decline in performance beyond this point. This indicates overfitting or model instability at higher epochs in the FedSGD framework. FedAvg displayed a better learning curve, reaching an accuracy of around 94.3% at 80 epochs but experienced a slight drop in performance at higher epochs, stabilizing around 92.5% beyond 120 epochs. FedMAP achieved superior performance compared to FedSGD and FedAvg, reaching an accuracy of approximately 96.7% at 120 epochs, and showed stable performance even at higher epoch counts.

The integration of the proposed FLPSO technique resulted in substantial improvements. FedAvg with FLPSO consistently outperformed its traditional counterpart, reaching around 95.2% accuracy at 80 epochs and maintaining high accuracy across the epoch range.

The best performance was observed with FedMAP with FLPSO, which attained the highest prediction accuracy of approximately 98% at 140 epochs and sustained its superior performance at higher epochs. This indicates that the FLPSO optimization technique effectively enhances model learning and generalization, ensuring improved and stable accuracy across different training rounds.

3.1. Communication Efficiency Evaluation

To assess communication efficiency in the federated learning process, we measured two primary metrics: (1) the total number of communication rounds required to reach 90% global model accuracy, and (2) the average data size transmitted per round, based on the serialized model parameter updates shared between clients and the central server. Specifically, the model weights were serialized using the TensorFlow Federated protocol buffer format, and the byte size of each transmission was recorded.

Although end-to-end latency was not directly measured, the number of communication rounds serves as a practical proxy, as fewer rounds imply lower communication overhead and reduced synchronization frequency. As reported in table 3, the integration of PSO with federated optimization (FLPSO) reduced the number of rounds by over 30% compared to standard FedSGD and FedAvg, demonstrating a clear improvement in communication efficiency.

3.2. Ablation Study: Isolating Contributions of PSO-WCO and FedAvg

To quantify the individual and combined impact of the PSO-WCO optimization and the FedAvg aggregation, we performed an ablation study comparing four configurations of the model:

1. **Baseline**: Centralized learning without PSO or FL.
2. **PSO-WCO Only**: Centralized model with PSO-WCO optimization, no federated architecture.
3. **FedAvg Only**: Federated model with standard FedAvg aggregation, no PSO-WCO.
4. **FLPSO (Proposed)**: Federated model with PSO-WCO optimization and FedAvg integration.

Table 5. Ablation study results comparing PSO-WCO and FedAvg contributions

Model Variant	Accuracy (%)	F1-Score	AUC	Rounds to 90%
Baseline (No PSO, No FL)	92.94	0.9288	0.95	N/A
PSO-WCO Only	96.40	0.9423	0.98	N/A
FedAvg Only	92.19	0.8791	0.92	3076
FLPSO (Proposed)	**98.30**	**0.9246**	**0.99**	**2326**

The ablation results clearly show that PSO-WCO contributes a substantial performance gain in both centralized and federated settings. When applied independently, PSO-WCO improved accuracy from 92.94% to 96.40%, and when integrated with federated training via FLPSO, the accuracy further improved to 98.30%. Similarly, FedAvg alone improved privacy and decentralization but underperformed in terms of classification metrics. These findings confirm the complementary strengths of PSO-WCO and FedAvg, validating the effectiveness of the proposed hybrid framework.

3.3. Computational Efficiency Comparison

In addition to prediction performance, we compared the computational efficiency of the proposed Federated PSO-WCO (FLPSO) framework against commonly used optimization methods, including grid search, random search, and Bayesian optimization. Efficiency was assessed in terms of (1) number of objective function evaluations to convergence, (2) approximate runtime, and (3) memory usage during training.

Table 6. Comparison of optimization methods in terms of computational cost

Method	Evaluations to Converge	Relative Runtime	Memory Footprint (MB)
Grid Search	> 500	High	Moderate
Random Search	~ 200	Moderate	Low
Bayesian Optimization	< 150	Low–Moderate	High
PSO-WCO (Proposed)	~ 100	Moderate	Moderate

As shown in table 6, FLPSO achieved convergence in fewer evaluations compared to grid and random search, and with lower memory overhead than Bayesian optimization. While the metaheuristic nature of PSO adds some initialization and update cost, the reduced number of function evaluations and parallel-friendly structure make it well-suited for federated environments with distributed computation. Additionally, PSO's ability to escape local minima contributed to its consistent convergence toward optimal ensemble weights.

These findings support the use of PSO-WCO as a computationally efficient and scalable alternative to traditional hyperparameter tuning methods in federated learning settings.

4. Limitations

Despite the advantages of federated learning in preserving privacy, it introduces potential system-level bottlenecks that may affect overall performance and scalability. One common issue is synchronization delay, where the central server must wait for all participating clients to complete their local training before aggregating model updates. This can be particularly problematic in heterogeneous environments where clients have varying computational resources and network latencies.

To mitigate synchronization delays, asynchronous federated learning strategies can be employed, where clients communicate updates independently without waiting for global synchronization. Additionally, adaptive client selection—where only a subset of reliable or fast clients is chosen in each communication round—can improve convergence speed and reduce idle time.

Another bottleneck lies in the communication cost due to the frequent transmission of high-dimensional model parameters. This can be addressed using model compression techniques such as weight quantization, sparsification, or update pruning, which reduce the payload size without significantly impacting accuracy. These mitigation strategies are vital for deploying FL systems in real-world IoMT or mobile environments where bandwidth and latency are critical constraints.

5. CONCLUSION AND FUTURE WORK

In this study, a novel and efficient framework was proposed for the accurate prediction of Diabetes Mellitus, integrating Federated Learning with Particle Swarm Optimization (FLPSO). The primary objective was to enhance prediction accuracy while preserving patient data privacy and reducing communication overhead in distributed healthcare environments. The proposed system leveraged local data from multiple IoMT-enabled hospitals and applied the FLPSO strategy to optimize classifier weight adjustments collaboratively without transferring sensitive data.

Extensive experimental evaluations demonstrated the superior performance of the proposed FLPSO-based model compared to conventional federated learning techniques. The results revealed that integrating PSO significantly improved model accuracy, reduced classification errors, and enhanced precision, sensitivity, and specificity. Notably, the FedMAP with FLPSO technique achieved the highest prediction accuracy of 98.30%, a classification error of 2.73%, and superior performance across all evaluation metrics. Furthermore, the proposed approach reduced the number of training rounds required to achieve convergence, thereby minimizing communication costs.

The incorporation of PSO in the federated learning framework proved effective in optimizing model aggregation weights dynamically, resulting in better generalization and predictive reliability. This study contributes to the growing body of research on privacy-preserving healthcare analytics by demonstrating that accurate and efficient diabetes prediction can be achieved without compromising patient data confidentiality.

Future work may focus on extending this approach to larger-scale healthcare networks and incorporating other optimization techniques such as Genetic Algorithms or Hybrid Swarm Intelligence methods to further improve performance. Additionally, real-time implementation of the proposed model on IoMT devices can be explored to support early diagnosis and continuous monitoring of diabetic patients in practical healthcare settings.

As part of future research, we intend to explore hybrid optimization frameworks that combine the strengths of Genetic Algorithms (GA) with the current PSO-WCO strategy. While PSO excels at rapid convergence and continuous parameter optimization, GAs offer strong global exploration capabilities through crossover and mutation mechanisms. A GA-PSO hybrid could enable more diverse solution exploration and mitigate premature convergence, especially in high-dimensional healthcare feature spaces.

However, integrating GAs into the federated optimization framework introduces challenges such as increased computational overhead, especially when evaluating large populations across distributed nodes. Moreover, parameter tuning (e.g., mutation rate, crossover probability, and selection strategy) becomes critical to ensure convergence stability and efficiency.

To address these challenges, future work will investigate lightweight GA variants, adaptive parameter tuning, and asynchronous parallel GA execution within federated systems. This line of research aims to further enhance model generalization and robustness while maintaining computational feasibility in real-world deployments.

REFERENCES

1. M. Gheisari, F. Ebrahimzadeh, M. Rahimi, M. Moazzamigodarzi, Y. Liu, P. K. Dutta Pramanik, M. A. Heravi, A. Mehbodniya, M. Ghaderzadeh, M. R. Feylizadeh, *et al.*, *Deep learning: Applications, architectures, models, tools, and frameworks: A comprehensive survey*, CAAI Trans. Intell. Technol., vol. 8, pp. 581–606, 2023.
2. V. Kaul, S. Enslin, and S. A. Gross, *History of artificial intelligence in medicine*, Gastrointest. Endosc., vol. 92, pp. 807–812, 2020.
3. R. M. M. De Souza, A. Holm, M. Biczyk, and L. N. De Castro, *A systematic literature review on the use of federated learning and bioinspired computing*, Electronics, vol. 13, no. 16, Art. no. 3157, Aug. 2024, doi: 10.3390/electronics13163157.
4. R. Kapila and S. Saleti, *Federated learning-based disease prediction: A fusion approach with feature selection and extraction*, Biomed. Signal Process. Control, vol. 100, Art. no. 106961, Sep. 2024, doi: 10.1016/j.bspc.2024.106961.
5. H. Ulutas, R. B. Günay, and M. E. Sahin, *Detecting diabetes in an ensemble model using a unique PSO-GWO hybrid approach to hyperparameter optimization*, Neural Comput. Appl., vol. 36, no. 29, pp. 18313–18341, Jul. 2024, doi: 10.1007/s00521-024-10160-y.
6. H. Bhatt, *et al.*, *Artificial neural network-driven federated learning for heart stroke prediction in healthcare 4.0 underlying 5G*, Concurrency Comput. Pract. Exp., vol. 36, no. 3, Sep. 2023, doi: 10.1002/cpe.7911.
7. Y. Supriya and R. Chengoden, *Breast cancer prediction using Shapley and game theory in federated learning environment*, IEEE Access, vol. 12, pp. 123018–123037, Jan. 2024, doi: 10.1109/access.2024.3424934.
8. G. S. Nariman and H. K. Hamarashid, *Hierarchical federated learning for health trend prediction and anomaly detection using pharmacy data: From zone to national scale*, Int. J. Data Sci. Anal., Mar. 2025, doi: 10.1007/s41060-025-00756-5.
9. R. Jiang, Y. Kumar, and D. Kruger, *Enhanced privacy-preserving architecture for fundus disease diagnosis with federated learning*, Appl. Sci., vol. 15, no. 6, Art. no. 3004, Mar. 2025, doi: 10.3390/app15063004.
10. A. Manocha, S. K. Sood, and M. Bhatia, *Federated learning-inspired smart ECG classification: An explainable artificial intelligence approach*, Multimed. Tools Appl., Aug. 2024, doi: 10.1007/s11042-024-20084-3.
11. Kaggle, *Diabetes prediction dataset*, https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset/data, Apr. 2023 (accessed Apr. 02, 2025).
12. A. M. Elshewey and A. M. Osman, *Orthopedic disease classification based on breadth-first search algorithm*, Sci. Rep., vol. 14, no. 1, Oct. 2024, doi: 10.1038/s41598-024-73559-6.
13. Y. Fouad, A. M. Osman, S. A. Z. Hassan, H. M. El-Bakry, and A. M. Elshewey, *Adaptive visual sentiment prediction model based on event concepts and object detection techniques in social media*, Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 7, Jan. 2023, doi: 10.14569/ijacsa.2023.0140728.
14. A. M. Elshewey, R. Y. Youssef, H. M. El-Bakry, and A. M. Osman, *Water potability classification based on hybrid stacked model and feature selection*, Environ. Sci. Pollut. Res., Mar. 2025, doi: 10.1007/s11356-025-36120-0.
15. M. Kaddes, Y. M. Ayid, A. M. Elshewey, and Y. Fouad, *Breast cancer classification based on hybrid CNN with LSTM model*, Sci. Rep., vol. 15, no. 1, Feb. 2025, doi: 10.1038/s41598-025-88459-6.
16. A. M. Elshewey, M. A. Aziz, O. A. Gaheen, M. S. Sawah, A. A. Elhamid, and A. M. Osman, *Prediction of aerodynamic coefficients based on machine learning models*, Model. Earth Syst. Environ., vol. 11, no. 3, Mar. 2025, doi: 10.1007/s40808-025-02355-6.
17. M. Nandagopal, K. Seerangan, T. Govindaraju, N. E. Abi, B. Balusamy, and S. Selvarajan, *A deep auto-optimized collaborative learning (DACL) model for disease prognosis using AI-IoMT systems*, Sci. Rep., vol. 14, no. 1, May 2024, doi: 10.1038/s41598-024-59846-2.
18. M. A. Khan, *et al.*, *Optimal feature selection for heart disease prediction using modified Artificial Bee Colony (M-ABC) and K-nearest neighbors (KNN)*, Sci. Rep., vol. 14, no. 1, Oct. 2024, doi: 10.1038/s41598-024-78021-1.
19. S. Jahan, *et al.*, *Federated explainable AI-based Alzheimer's disease prediction with multimodal data*, IEEE Access, p. 1, Jan. 2025, doi: 10.1109/access.2025.3547343.
20. G. Rjoub, O. A. Wahab, J. Bentahar, R. Cohen, and A. S. Bataineh, *Trust-augmented deep reinforcement learning for federated learning client selection*, Inf. Syst. Front., vol. 26, no. 4, pp. 1261–1278, Jul. 2022, doi: 10.1007/s10796-022-10307-z.
21. O. I. Ramadan, L. S. Ali, Y. Ramadan, R. M. Abobaker, H. M. Flifel, M. A. Elkholy, H. I. Abobaker, E. M. M. Gabr, I. I. Hemdan, and S. A. Z. Hassan, *Enhancing breast cancer classification based on BPSO feature selection and machine learning techniques*, Eng. Technol. Appl. Sci. Res., vol. 15, no. 3, pp. 23907–23916, Jun. 2025, doi: 10.48084/etasr.10900.
22. A. M. Elshewey, S. A. Z. Hassan, R. Y. Youssef, *et al.*, *Enhancing hydrogen energy consumption prediction based on stacked machine learning model with Shapley additive explanations*, Process Integr. Optim. Sustain., 2025, doi: 10.1007/s41660-025-00539-2.
23. H. Abazid, N. Alabbas, A. Hammad, O. I. Ramadan, E. E. Al Jomaa, M. F. Amer, *et al.*, *Biochemical and histopathological alterations induced by Tramadol*, Med. J. Islam. Repub. Iran, vol. 39, no. 1, pp. 240–247, 2025.