International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Hybrid Machine Learning Optimized Algorithm for Type 2 Diabetes Mellitus Prediction

Raja S[1], Dr. Nagarajan. L[2]

Submitted: 05/02/2024 **Revised**: 13/03/2024 **Accepted**: 19/03/2024

Abstract: Investigating a healthcare system employing contemporary computing technology is a prominent field of inquiry within healthcare research. Researchers in the technology and healthcare domain are continuously collaborating to enhance the technological preparedness of these systems. Diabetes is widely recognized as a highly consequential and enduring ailment, giving rise to various complications, including but not limited to visual impairment, limb loss, and cardiovascular disorders. Numerous countries are actively working to mitigate the impact of this disease by implementing early-stage preventive measures, which involve identifying and prognosticating diabetes symptoms through diverse diagnostic approaches. This research activity aims to construct a very accurate model for the early prognosis of Type 2 Diabetes Mellitus (T2-DM). This paper presents a new hybrid clustering model that seeks to improve the integration between the Particle Swarm Optimization (PSO) technique for feature optimization and the Fuzzy Clustering Means (FCM) algorithm for effective clustering. This approach aims to improve the precision, responsiveness, and selectivity of the clustering procedure. The study conducted a comparative analysis of the proposed model's accuracy, sensitivity, and specificity metrics about existing hybrid approaches, such as K means-C4.5 and ANN+FNN, which are considered state-of-the-art in the field.

Keyword- Machine Learning, PSO, FCM, Optimization, Diabetes.

1. Introduction

Diabetes mellitus (DM), colloquially known as diabetes, covers a spectrum of metabolic illnesses distinguished by persistent elevation of blood glucose levels. Symptoms indicative of elevated glucose levels encompass frequent urination, constant thirst, and heightened appetite [1]. Untreated diabetes has the potential to give rise to severe health complications in individuals, including but not limited to diabetic ketoacidosis, hyperosmolar hyperglycemia condition, and mortality. This may result in long-term problems such as cardiovascular disease, cerebral stroke, renal failure, foot ulcers, and ocular issues [2].

Diabetes mellitus arises from insufficient insulin production by the pancreas or impaired use of the generated insulin by tissues and cells in the body. DM manifests in three distinct types [3]:

- Diabetes Mellitus Type-1 (DMT 1) is characterized by inadequate pancreatic insulin synthesis, usually known as "Insulin Dependent Diabetes Mellitus" (IDDM). Individuals afflicted with this type of diabetes mellitus necessitate exogenous insulin administration to compensate for the insufficient endogenous production of this hormone by the pancreatic organ.
- Diabetes Mellitus Type-2 is distinguished by insulin resistance, which refers to an atypical response of body cells to insulin in contrast to their usual activity. The potential outcome of this process is the gradual reduction of insulin levels in the entire organism. This medical disorder is frequently known as "non-insulin dependent diabetes mellitus" (NIDDM) or "type 2 DM". This specific manifestation of diabetes is commonly noticed in those with a higher body mass index (BMI) or those with a predominantly inactive lifestyle.
- Gestational diabetes is a significant condition that is commonly found throughout pregnancy.

Machine learning (ML) is a subfield of artificial intelligence (AI) that enables computers to autonomously acquire knowledge and perform statistical analysis without requiring human interaction [4]. Machine learning models and algorithms are widely utilized and have demonstrated their dependability across various applications. ML has been progressively employed by researchers in the field of healthcare, namely for diagnostics [4, 5] and prognostics [6, 7] and the development of new drugs [8, 9]. Machine

1Research Scholar PG & Research Department of Computer Science Adaikalamatha College Vallam,Thanjavur Affiliated to Bharathidasan UniversityTiruchirappalli, Tamilnadu, India
Mail.id-sk.rajamecse@gmail.com
[2]Assistant Professor PG & Research Department of Computer Science Adaikalamatha College Vallam,Thanjavur Affiliated to Bharathidasan University Tiruchirappalli, Tamilnadu, India
Mail.id-mcadirector@gmail.com

learning begins with data, which can be either structured or unstructured. The second phase, data preparation or preprocessing, is selecting relevant data via a data mining approach and reformatting it into a more usable format [7]. When everything is set up, model validation [8] involves putting the model through its paces by having it assess the accuracy of several trained data sets or run statistical methods. For final confirmation in prediction and classification, hyperparameter tuning is used to optimize or improve the model.

2. Literature Survey

The author defines the research purpose as using ML classifiers to assess the risk of diabetes in female Pima Indians. A dataset consisting of 768 female patients was used to determine four classifiers. Across all K values, the results showed that Logistic Regression was the most accurate method. Other classifiers achieved accuracy levels of 0.71, 0.76, and 0.75. With AUC values of 0.83, 0.82, and 0.81, respectively, Logistic Regression, Random Forest, and Naive Bayes were the best-performing models. The use of these models is recommended for making the diagnosis of diabetes in a patient. The objective of this study was to employ cross-validation methodologies to ascertain the optimal machine learning model for the prediction of diabetes. When everything is set up, model validation [1] involves putting the model through its paces by having it assess the accuracy of several trained datasets or run statistical methods. For final verification in prediction and classification, hyperparameter tuning is used to optimize or improve the model.

The primary objective of the study in [2] is to ascertain the incidence of diabetes in India by examining participants' lifestyles and genetic backgrounds. Data from 952 people was acquired through online and paper-and-pencil surveys and analyzed with ML techniques. The objective was to forecast the probability of an individual getting Type 2 diabetes. The Random Forest (RF) Classifier demonstrated the highest overall accuracy among both datasets, suggesting its potential utility in forecasting the probability of getting diabetes. Diabetes and its complications can be avoided via early diagnosis and treatment.

The study in [3] examines the global health concern of diabetes, which is recognized as a prominent contributor to mortality on a worldwide scale. This study examines a range of factors that contribute to the development and progression of diabetes, encompassing lifestyle choices, psychosocial factors, underlying medical disorders, demographic characteristics, and genetic predisposition. The researchers thoroughly assessed 35 ML algorithms to predict Type 2 diabetes. This evaluation utilizes authentic diabetes datasets and incorporates nine feature selection algorithms. This study evaluates the precision, F-measure, and temporal demands associated with the construction and validation of models for both diabetes and non-diabetic populations. The examination of performance offers a more comprehensive comprehension of the efficacy of these models in predicting Type 2 diabetes.

The study's primary aim in [4] is to differentiate and classify different subcategories of Type 2 Diabetes (T2D) to improve prognosis and treatment approaches. Type 2 diabetes (T2D) is a chronic medical disorder characterized by high blood glucose levels due to reduced insulin production and resistance. Researchers employed machine learning techniques to analyze the data obtained from the Pima Indian Diabetes Dataset (PIDD). The model with the highest accuracy, the Generalized Boosted Regression, performed well in both specificity (85.19%) and kappa statistics (78.77%). Several other algorithms, such as Sparse Distance Weighted Discrimination (SDWD), Generalized Additive Model utilizing LOESS (GAML), and Boosted Generalized Additive Models (BGAM), exhibited a notable level of sensitivity (100%), a great area under the receiver operating characteristic curve (AUROC) of 95.26%, and a low gloss value of 30.98%. The research inquiry effectively identified multiple independent variables, namely glucose levels, BMI, diabetic pedigree function, and age, which consistently showed high accuracy in predicting outcomes associated with Type 2 diabetes (T2D). The timely identification of T2D is of utmost importance for predictive purposes and the implementation of productive preventive measures, given the absence of a definitive cure for this condition.

3. Methods and Materials

3.1 Dataset

To carry out the prediction process, a fundamental prerequisite is the availability of a dataset. Numerous academics have performed analyses using various datasets over time. These datasets range from those already available online to those collected by researchers from multiple sources, including hospitals and other medical institutions. To date, there exists a diverse range of datasets on diabetes, each with distinct areas of emphasis. The Pima Indians Diabetes Database (PIDD) is widely recognized as the most commonly utilized dataset among the options provided. This research piece employs the use of PIDD for analysis. Out of 768 instances, 500 were identified as non-diabetic, while the remaining 267 were identified as

diabetic. Table 1 presents a selection of fields derived from the Pima data set.

Table 1. Sample data from Pima Indian Diabetes Dataset

Preg	Plas	Pres	Skin	Insu	BMI	Pedi	Age	Class
8	151	67	43	346	35.7	0.918	44	Non-Diabetic
12	142	96	36	149	38.7	0.454	55	Diabetic
0	120	82	48	232	44.7	0.751	33	Diabetic
8	103	78	36	0	35.9	0.865	48	Diabetic
6	101	76	29	0	32	0.403	34	Non-Diabetic
3	92	89	44	0	40.2	0.703	29	Diabetic
5	132	74	0	123	24.8	0.477	62	Diabetic
12	108	76	56	0	38.6	0.378	43	Non-Diabetic
2	117	72	32	99	36.7	0.729	35	Diabetic
9	86	78	33	0	39.4	0.657	40	Non-Diabetic

3.2 Machine Learning Classification Algorithm

The traditional approach of the standard Fuzzy C-Means (FCM) algorithm exhibits high sensitivity towards noisy data. One significant limitation is that it will harm the overall effectiveness of the problem-solving process. To address the concern mentioned above, modifying the conventional Fuzzy C-Means (FCM) algorithm is necessary. This adjustment should be accompanied by using preprocessed data as the input. This study presents a novel model that integrates the advantageous characteristics of Particle Swarm Optimization (PSO) with the conventional Fuzzy C-means (FCM) method by adapting each cluster's membership weighting. In the suggested methodology of Particle Swarm Optimization-Fuzzy C-Means (PSO-FCM), each data point in the dataset is assigned a distinct weight corresponding to each cluster. The unique weight is crucial in correctly clustering the noisy data.

3.3 Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) is classified within the domain of evolutionary algorithms. The synchronized movement of birds, fish, or insects distinguishes PSO. In this context, the particles and swarm correspond to the individual entities and the collective population, respectively. Every particle actively seeks the optimal spot by utilizing its intelligence and employing strategic movement patterns. The fitness value is used to calculate a fitness function, which is used to estimate where the particle is. PSO offers significant high-level data points that are valuable during the initial selection process to facilitate further classification. The primary focal point in PSO pertains to the alterations in velocity. The trajectory for each particle is determined by its respective velocity.

Let m_i represent the i[th] particle and V_i denote the velocity of the particle. Similarly, the term P_{best} denotes the location with the best performance inside a local neighborhood, whereas G_{best} represents the position with the best performance globally. The particle positions are updated with the assistance of P_{best} and G_{best}. The position and velocity updates are represented by equations (1) and (2).

$$m_{ij}(p+1) = m_{ij}(p) + V_{ij}(p+1) \quad (1)$$

$$V_{ij}(p+1) = WV_{ij}(p) + C_1R_1\left(P_{ij}(p) - m_{ij}(p)\right) + C_2R_2\left(P_{ij}(p) - m_{ij}(p)\right) \quad (2)$$

Where j=1,2,3….. n, W represents the weight of inertia. C_1 and C_2 represents coefficients of the acceleration and two random sequence elements are denoted by R_1 and R_2.

Algorithm 1: Procedure for PSO

Step 1: The population of particles is initialized by assigning random values to their locations and velocities

Step 2: The personal best position (P_{best}) of each particle is initialized to its current position
Step 3: The global best position (G_{best}) is initialized as the optimal position among all particles
Step 4: Set the maximum number of iterations or a convergence criterion
Step 5: Set parameters: weight (W), cognitive and social coefficient (C1 and C_2)
Step 6: For all possible iterations between 1 and the maximum
For every member of the population:
Calculate the current position fitness value
If the fitness is better than the fitness at P_{best}:
Update P_{best} to the current position
Update G_{best} to the position of the particle with the best P_{best} among all particles
Step 7: For each particle in the population
Generate a random number R_1 and R_2 between 0 and 1
Step 8: The velocity of the particle is updated by equation (2)
Step 9: The particle's position is updated by employing equation (1)
Step 10: Clip new_position to stay within the boundaries of the search space
Step 11: Return G_{best}

3.4 Fuzzy Clustering Means (FCM) Algorithm

The Fuzzy C-Means (FCM) algorithm is a widely employed clustering technique in data analysis and ML. The proposed approach expands upon the conventional K-Means method by enabling data points to be assigned to different clusters with varied membership levels. The FCM algorithm is a fundamental element within the broader field of fuzzy clustering, which involves the application of fuzzy logic to efficiently address the inherent uncertainty and ambiguity inherent in the task of data clustering.

The FCM algorithm assigns membership values to individual data points, indicating the degree of association with each cluster. In contrast to the K-Means algorithm, which gives each data point exclusively to a single cluster, the FCM algorithm offers a more adaptable and nuanced approach to clustering. FCM permits data points to possess partial memberships in several cluster, providing a more flexible clustering outcome. The optimization of membership values and cluster centroids is accomplished through an iterative process.

The cluster numbers in Fuzzy C-Means (FCM) are predetermined. Due to the absence of rigid cluster boundaries, the FCM algorithm is a more suitable approach than K-Means clustering techniques. The efficacy of this strategy is primarily contingent upon the appropriate initialization of the clustering center. The initialization of clustering centers can be achieved in several ways, which then enhances the finding of the optimal clustering by refining the centroids. The proposed approach can be classified as an iterative optimization technique, with its objective function being mathematically represented by Equation 3.

$$K_m = \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}\left|x_i - c_j\right| \quad (3)$$

$$c_j = \sum_{i=1}^{c}\mu_{ij}^{m}x_i / \sum_{i=1}^{c}\mu_{ij}^{m} \quad (4)$$

$$\mu_{ij} = \sum_{k-1}^{c}\left(\frac{\left||x_i - c_j|\right|}{||x_i - c_k||}\right)\text{^} - 2/(m-1) \quad (5)$$

Algorithm 2: Fuzzy Clustering Means

Step 1: Choose K Clusters
Step 2: Initialize the fitness parameter m ($1 < m \leq \infty$)
Step 3: Initialize cluster centroid randomly
Step 4: For each data point i
For each Cluster j
Calculate the membership function degree for data point i to j clusters by using eqn (3)
Step 5: Update the cluster centroid by using Eqn (4)
Step 6: calculate highest membership degree value and updae it using Eqn (5)
Step 7: Repeat the above steps until convergence or the maximum number of iterations is reached
Step 8: Return the cluster centroid and membership function

3.5 Hybrid of PSO and FCM

The FCM algorithm is widely recognized for its high efficiency and effectiveness in addressing various clustering challenges. One significant limitation associated with the exclusive use of FCM is its susceptibility to converging toward the local optimal point. This study paper proposes a novel technique that integrates the salient characteristics of PSO and FCM to mitigate the issue of premature convergence. This approach encompasses the advantages of FCM as well as PSO. Figure 1 illustrates the visual representation of the proposed predictive model. The procedural procedures of the Particle Swarm Optimization - Fuzzy C-means (PSO-FCM) method are elaborated in Algorithm 3.

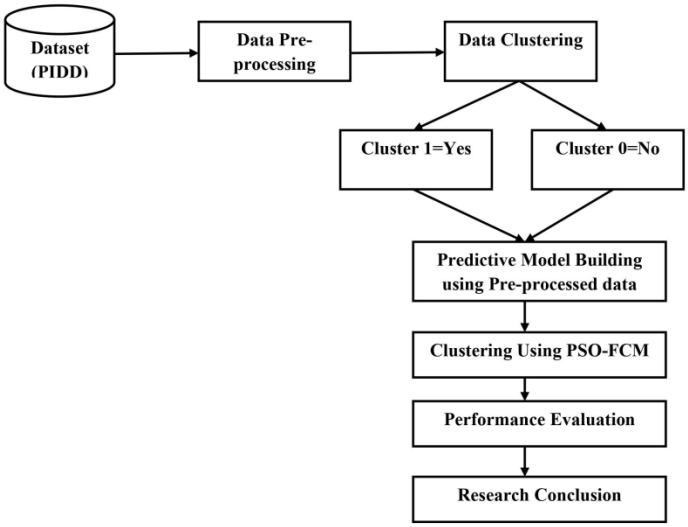


Figure 1: Block diagram of proposed work

Developing a hybrid algorithm that integrates PSO and FCM techniques to predict type 2 diabetes is a multifaceted and specialized endeavor. A hybrid approach can be devised to optimize the selection of features, model parameters, or both. The PSO algorithm is employed to optimize the process of feature selection. Please define a fitness function that assesses the efficacy of feature subsets in predicting diabetes. The PSO algorithm aims to identify the most optimal subset of features.

Algorithm 3: Hybridization of PSO-FCM

Step 1: Initialize a PSO Parameters

Step 2: Initialize PSO Particle i in N

Initialize a random binary feature selection vector

Step 3: Initialize global best (G_{best}) and particle best (P_{best}) positions

Step 4: For iteration in 1 to max_iter

For each particle i in N

Calculate fitness based on FCM clustering quality using selected features:

Perform FCM clustering on selected features

Update fitness based on the clustering quality

Step 5: Update P_{best} if the current fitness is better than the previous P_{best}

Step 6: Update G_{best} based on the best Pbest among all particles

Step 7: For each particle i in N

Update particle velocity and position using PSO equations:

Update velocity based on current velocity, P_{best}, and G_{best}

Update position by applying velocity

Step 8: End of iterations

Step 9: Select the best particle based on fitness among all particles

Step 10: Return the selected features from the best particle as the feature subset

Step 11: Perform FCM clustering on the selected feature subset

Step 12: Initialize cluster centroids randomly or using another method

Step 13: Iterate FCM until convergence based on selected features and parameters (K, m)

Step 14: Assign data points to clusters based on their fuzzy memberships

Step 15: Evaluate the clustering results

4. Performance analysis

The primary objective of early diagnosis of diabetes is to enhance longevity and improve overall life expectancy. Numerous supervised and unsupervised methodologies have been developed to identify diabetes. This section analyzes the performance of the Particle Swarm Optimization-Fuzzy C-Means (PSO-FCM) algorithm and shows the outcomes according to various parameters. The research was conducted using MATLAB R2019b, loaded on a computer system with 8GB of RAM, running the Windows 10 operating system, and equipped with an Intel Core i3 processor. In this instance, a comprehensive examination of all eight qualities and one class is conducted. Using the confusion matrix, the algorithm's performance is evaluated.

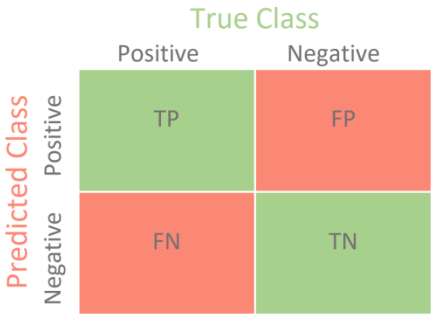
The execution has been conducted for multiple occurrences, and the outcomes acquired for Sensitivity (Sen) and

Specificity (Spe) are presented in Table 2. The hybrid model shown in this study demonstrates greater values of Specificity (96.3) and Sensitivity (96.6) compared to the other methodologies selected for analysis. As previously stated, a higher kappa statistic value suggests that the constructed model is more reliable for predicting the occurrence of this particular disease.

The kappa statistic was computed to be 0.9183, suggesting that the suggested hybrid model PSO-FCM is a viable and effective model for early disease prediction. Additionally, the precision values were determined to be 0.9657.

The accuracy of the problem is assessed using a confusion matrix, which includes the variables True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Table 3 presents the commonly used confusion matrix. In addition to accuracy, the performance of the constructed hybrid model is influenced by characteristics such as Sensitivity (Sen), Specificity (Spe), and Kappa value.

Table 3: Template of Confusion Matrix



4.1 Sensitivity (Sen)

The term "sensitivity" (Sen) represents the True Positive Rate (TPR), and its mathematical expression is provided as follows

$$Sen = \frac{TP}{TP+FN} \quad (6)$$

4.2 Specificity (Spe)

The term "sensitivity" (Sen) represents the True Positive Rate (TPR), and its mathematical expression is provided as follows.

$$Spe = \frac{TN}{TN+FP} \quad (7)$$

4.3 Accuracy (Acc)

Accuracy (Acc) is a metric that quantifies the proportion of samples that are properly categorised in relation to the overall number of samples

$$Accuracy\ (Acc) = \frac{TN+TP}{TP+TN+FP+FN} \quad (8)$$

4.4 Kappa Statistic

The kappa statistic is a measure that can be used to assess a model's level of agreement or consistency. The kappa statistic is a measure that spans from 0 to 1. The model demonstrates applicability when the value approaches 1, while a value of "0" indicates its incapability. The parameter is denoted by the equation (9).

$$K = \frac{[P(A)-P(B)}{1-P(B)} \quad (9)$$

Here $$P(A) = \frac{(TP+TN)}{N} \quad (10)$$

$$P(B) = \frac{[(TP+FN)(TP+TN)(TN+FN)]}{N^2} \quad (11)$$

Table 2. Performance metric comparison of Proposed Vs Existing

	Hybrid PSO-FCM	K-Means + C4.5	ANN +FNN
Sensitivity	96.5	93.8	81.3
Specificity	93.5	92.3	83.7

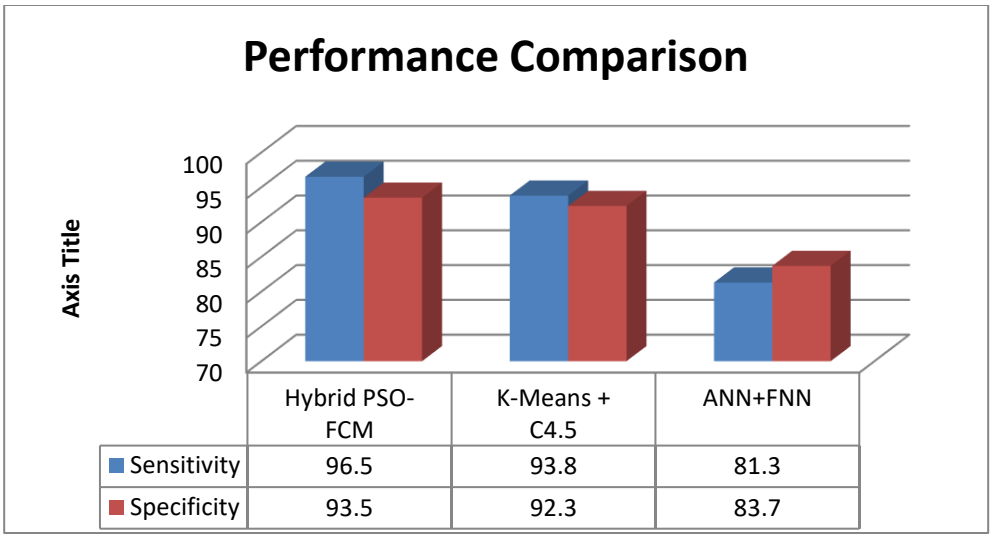


Figure 2: Performance Comparison

Based on the data shown in Table 2, it can be deduced that the hybrid model formulated in this work exhibits a 5.22% and 15.3% enhancement in the SN value compared to the K-Means+C4.5 and ANN + FNN techniques, respectively. Similarly, the value of SP experiences an increase of 2.01% and 8% compared to K-Means + C4.5 and ANN+FNN, respectively. Figure 2 depicts the comparison between Sen and Spe.

Table 3. Accuracy comparison of Proposed Vs Existing

Method	Accuracy
Hybrid PSO-FCM	94.5
K-Means+M-Tree (Tomar and Manjhvar, 2017)	90.3
K-Means (Tomar and Manjhvar, 2017)	63.7
SVM+FCM (Sanakal and Jayakumari, 2014)	93.4
ANN+FNN (Humar and Novruz, 2008)	85.4
K-Means + C4.5 (Patilet et al., 2010)	93.2

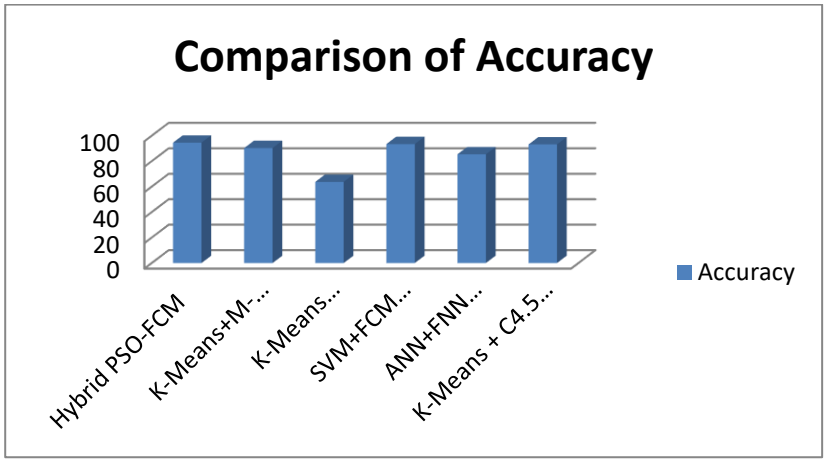


Figure 3: Accuracy Comparison

Based on the data provided in Table 3, it is evident that the hybrid model proposed in this study exhibits an accuracy of 95.42%. This finding signifies a notable enhancement in comparison to the approaches employed in prior research. Based on the obtained outcomes, it is apparent that the created model outperforms the other projected systems/models described. The analysis of statistical measures such as Kappa, Sensitivity, and Specificity can demonstrate the efficacy of the proposed model in predicting individuals with diabetes. Figure 3 depicts the comparative evaluation of accuracy compared to other algorithms described in the study.

Based on the information provided in Table 3, it is apparent that the accuracy of the hybrid model offered demonstrates a 1.12% enhancement in comparison to the SVM + FCM model, 3.04% compared to the K-Means + C4.5 model, 3.388% compared to the K-Means-M-Tree model, 10.92% compared to the ANN + FNN model, and 28.102% compared to the K-Means model.

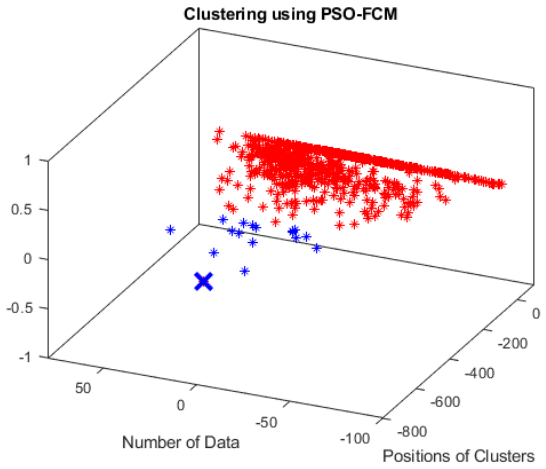


Figure 4: Formation of Cluster and cluster head

The graphical representation of the cluster and its center is depicted in Figure 4, generated using MATLAB software. The classification of individuals as either diabetes or non-diabetic is visually represented using red and blue within a single cluster. The cluster centers are denoted by a symbol in the shape of an 'x' for clusters that have been framed.

5. Conclusion

This study introduces a novel model, PSO-FCM, which combines the advantageous characteristics of both PSO and FCM. Following an extensive examination of previously published literature, a novel approach has been devised that integrates Particle Swarm Optimization (PSO) with Fuzzy C-Means (FCM) algorithms. One notable aspect of Particle Swarm Optimization (PSO) is the utilization of variables that can select values from the particle space based on their respective positions and velocities. Additionally, it necessitates fewer parameters and a limited number of iterations. The Fuzzy C-Means (FCM) algorithm provides stability, increased density, and adequate cluster characterization by employing a limited number of variables.

The P_{best} and G_{best} values were first determined and utilized in FCM to enhance the clustering process. While the fuzzy c-means (FCM) technique is generally adequate for clustering, its accuracy can be diminished as a result of early convergence.

The widely recognized optimization method PSO has been integrated to enhance efficiency and preserve the advantageous characteristics of FCM. As a result of this, the occurrence of early convergence has been prevented, leading to improved accuracy and kappa statistics values. This substantiates the efficacy of the formulated model in early-stage prediction of diabetes disease.

The sensitivity, specificity, and accuracy achieved using the suggested hybrid model were 95.6%, 95.3%, and 95.42%, respectively. In future research endeavors, applying this methodology to a real-time dataset obtained from hospitals encompassing male and female records is recommended. This is because the current dataset utilized for the PIDD just comprises documents of women.

References

[1] Gopi Battineni, Getu Gamo Sagaro, Chintalapudi Nalini, Francesco Amenta and Seyed Khosrow

Tayebati, “Comparative Machine-Learning Approach: A Follow-Up Study on Type 2 Diabetes Predictions by Cross-Validation Methods”, Machines 2019, 7, 74; doi:10.3390/machines7040074.

[2] Neha Prerna Tigga, Shruti Garga, “Prediction of Type 2 Diabetes using Machine learning Classification Methods”, International Conference on Computational Intelligence and Data Science (ICCIDS 2019), Available online at www.sciencedirect.com.

[3] Leila Ismail, Huned Materwala, Maryam Tayefi, Phuong Ngo, Achim P. Karduck, “Type 2 Diabetes with Artificial Intelligence Machine Learning: Methods and Evaluation”, Archives of Computational Methods in Engineering (2022) 29:313–333, https://doi.org/10.1007/s11831-021-09582-x.

[4] Koushik Chandra Howlader, Md. Shahriare Satu, Md. Abdul Awal, Md. Rabiul Islam, Sheikh Mohammed Shariful Islam, Julian M. W. Quinn and Mohammad Ali Moni, “Machine learning models for classification and identification of significant attributes to detect type 2 diabetes”, Health Information Science and Systems (2022) 10:2, https://doi.org/10.1007/s13755-021-00168-2.

[5] Marwan Al-Tawil, Basel A. Mahafzah, Arar Al Tawil and Ibrahim Aljarah, “Bio-Inspired Machine Learning Approach to Type 2 Diabetes Detection”, https://www.mdpi.com/journal/symmetry, Symmetry **2023**, 15, 764. https://doi.org/10.3390/sym15030764.

[6] Yach, D.; Hawkes, C.; Gould, C.L.; Hofman, K.J. The Global Burden of Chronic DiseasesOvercoming Impediments to Prevention and Control. JAMA 2004, 291, 2616–2622.

[7] Vaishali, R.; Sasikala, R.; Ramasubbareddy, S.; Remya, S.; Nalluri, S. Genetic algorithm based feature selection and MOE Fuzzy classification algorithm on Pima Indians Diabetes dataset. In Proceedings of the 2017 International Conference on Computing Networking and Informatics (ICCNI), Lagos, Nigeria, 29–31 October 2017; pp. 1–5.

[8] Khanam, J.J.; Foo, S.Y. A comparison of machine learning algorithms for diabetes prediction. ICT Express 2021, 7, 432–439.

[9] Khalid, S.; Khalil, T.; Nasreen, S. A survey of feature selection and feature extraction techniques in machine learning. In Proceedings of the 2014 Science and Information Conference, London, UK, 27–29 August 2014; pp. 372–378.

[10] Swapna, G.; Vinayakumar, R.; Soman, K.P. Diabetes detection using deep learning algorithms. ICT Express 2018, 4, 243–246.

[11] Yu, L.; Liu, H. Efficient feature selection via analysis of relevance and redundancy. J. Mach. Learn. Res. 2004, 5, 1205–1224.

[12] Ismail, L.; Materwala, H.; Tayefi, M.; Ngo, P.; Karduck, A.P. Type 2 Diabetes with Artificial Intelligence Machine Learning: Methods and Evaluation. Arch. Comput. Methods Eng. 2021, 29, 313–333.

[13] Yusta, S.C. Different metaheuristic strategies to solve the feature selection problem. Pattern Recognit. Lett. 2009, 30, 525–534.

[14] Gandomi, A.H.; Yang, X.-S.; Alavi, A.H. Cuckoo search algorithm: A metaheuristic approach to solve structural optimization problems. Eng. Comput. 2013, 29, 17–35.

[15] Yang, X.-S. Nature-Inspired Metaheuristic Algorithms; Luniver Press: London, UK, 2010.

[16] Negi, A.; Jaiswal, V. A first attempt to develop a diabetes prediction method based on different global datasets. In Proceedings of the 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), Waknaghat, India, 22–24 December 2016; pp. 237–241.

[17] Tigga, N.P.; Garg, S., “Prediction of Type 2 Diabetes using Machine Learning Classification Methods”. Procedia Comput. Sci. 2020, 167, 706–716.