Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Feature Selection and Extraction for Diabetes Management Using Bayesian Network-Based Emperor Penguins Colony Recommendation System

[1]Ms. B Madhuvanthi, [2]Dr.T.S. Baskaran,
[1]Research Scholar, PG & Research Department of Computer Science, A. Veeriya Vandayar Memorial Sri Pushpam College (Autonomous), Poondi-613503, Thanjavur,
E-Mail: madhuvanthib@yahoo.in
[2]Associate professor, Research & PG Department of Computer Science, A.V.V.M.Sri.Pushpam College,Poondi,Thanjavur
t_s_baskaran@yahoo.com
"Affiliated to Bharathidasan University",Tirchuirappalli

ARTICLE INFO

Received: 18 Dec 2024

Revised: 10 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

In the face of the global diabetes epidemic, marked by a steep rise in cases from 1980 to 2021, novel strategies are imperative. This study introduces an innovative approach that synergizes Bayesian networks and the adaptability of Emperor Penguins Colonies to tackle diabetes through advanced feature selection and extraction. Diabetes, a complex disease with multifaceted causative factors, necessitates predictive measures that go beyond conventional methods. The proposed Recommendation System for Food and Exercise (RSFE) harnesses Bayesian network-based techniques to analyze health parameters and medical reports, generating personalized diet and exercise plans for individuals, including those with comorbid diabetes and blood pressure conditions. By mirroring the dynamic decision-making of Emperor Penguins Colonies, this approach aims to create an intelligent healthcare system that transforms intricate datasets into actionable insights, offering tailored recommendations for optimal health management. This research holds promise for proactive disease intervention and personalized care, aligning with the urgent need for effective diabetes management.

Keywords: Bayesian networks, feature extraction, feature selection, Emperor Penguins Colony

1. INTRODUCTION

The global prevalence of diabetes has reached alarming levels, with a significant surge in cases observed from 1980 to 2021 [1]. This epidemic has posed a substantial burden on healthcare systems worldwide, emphasizing the need for innovative and effective strategies for diabetes management. Diabetes is a complex and multifaceted disease influenced by various genetic, environmental, and lifestyle factors [2]. Conventional approaches to diabetes management have shown limited success, and there is a pressing need for more sophisticated and personalized solutions.

Addressing the challenges posed by diabetes requires a holistic understanding of the disease and its contributing factors. Traditional methods of diabetes management often overlook the intricate interplay of variables that affect an individual health, leading to suboptimal outcomes [3]. The challenges in diabetes management include:

Diabetes has multiple causative factors, making it challenging to identify the specific contributors to an individual condition [4].Many individuals with diabetes also suffer from comorbid conditions, such as high blood pressure, which further complicates treatment strategies [5].Effective management of diabetes necessitates personalized interventions tailored to an individual unique health profile and lifestyle [6].Converting the vast amount of health data and medical reports into actionable recommendations for individuals remains a significant challenge [7].

In light of these challenges, this research addresses the problem of improving diabetes management by leveraging advanced techniques in feature selection and extraction. Specifically, we propose a Recommendation System for

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Food and Exercise (RSFE) that utilizes Bayesian network-based methods to analyze medical records [8-10] and health profile. The RSFE aims to generate personalized diet and exercise plans for individuals, with a particular focus on those with comorbid diabetes and high blood pressure conditions.

The main objective of this work is,

- To develop a robust Bayesian network-based model for feature selection and extraction from health data and medical reports.
- To design a Recommendation System for Food and Exercise (RSFE) that generates personalized diet and exercise plans based on the extracted features.
- To assess the effectiveness of the RSFE in improving diabetes management, especially in individuals with comorbid diabetes and high blood pressure.

This research introduces several novel elements to the field of diabetes management:

We propose the integration of Bayesian networks, a powerful probabilistic modeling technique, into the realm of diabetes management, offering a data-driven approach to understanding and addressing the disease complexities.

Drawing inspiration from the adaptive and dynamic decision-making processes of Emperor Penguins Colonies, we aim to create an intelligent healthcare system capable of transforming complex datasets into actionable recommendations for individuals.

The RSFE system emphasizes the importance of personalized care by tailoring diet and exercise plans to an individual specific health profile and needs.

By providing individuals with actionable recommendations, this research contributes to proactive disease intervention, ultimately aiming to improve long-term health outcomes for those affected by diabetes and related conditions.

2. RELATED WORKS

K-nearest neighbor (KNN) is a relatively new feature selection method developed by Ganesh et al. [11]. In this system, weighted superposition attraction (WSA) is used to allow for incremental progress. When applied to large datasets, this technique has the potential to reduce the number of features by up to 99% while maintaining classification accuracy. WSA-KNN outperforms both traditional machine learning methods (by approximately 18%) and ensemble machine learning algorithms (by approximately 9%) by a significant margin. Furthermore, the WSA-KNN algorithm produces results that are comparable to or slightly better than other existing methods.

Shaji et al. [12] presented an intriguing and novel framework for categorizing big data. The filtered data is then further reduced in dimensionality using Hadoop map reducer. By utilizing the BRO algorithm optimal feature selection process, the proposed fuzzy optimal deep convolutional neural network (FO-DCNN) improves accuracy and performance. Our FO-DCNN model combines the deep CNN architecture with the fuzzy-based remora optimization (F-RO) algorithm. The model enables precise classification of SIoT data. When the proposed model is tested on the Twitter dataset, the Rotten Tomato dataset, the skin disease dataset, the diabetes dataset, and the hepatitis dataset, classification improvements, lower memory usage, and faster computation times are all observed.

Ay et al. [13] hope to use machine learning experiments designed to achieve this goal, the Cleveland Heart Disease Dataset and the Faisalabad Institute of Cardiology Heart Failure Dataset, both available through the UCI, are used. For feature selection across a wide range of population sizes, the best fitness values are used to implement algorithms. The experimental results show that combining meta-heuristic and ML algorithms improves prediction accuracy significantly. These improvements were made possible by combining both classes of algorithms.

The Simplex Method-based Social Spider Optimization (SMSSO) method provided by Nayak et al. [14] is used to modify the Social Spider Optimization (SSO) method. Multiple test data sets were used to validate the SMSSO-NN procedure accuracy. SMSSO-NN outperforms competing methods on a variety of cancer datasets, including the Wisconsin Breast Cancer (WBC) dataset, which has a 99.36% accuracy rate, the lung cancer dataset, the diabetes dataset, and the cervical cancer dataset, which has a 98% accuracy rate. The results are compared to those of other approaches. The SMSSO-NN algorithm is a refinement of the LLWNN-FA algorithm that improves accuracy by utilizing the classification methods.

Ihnaini et al. [15] propose a data fusion and deep machine learning-based healthcare recommendation system to combat diabetes. Data fusion can reduce the computational load on the proposed system, allowing it to predict and recommend treatment for this potentially fatal disease. As a result, the system will be able to accomplish both tasks more effectively. Finally, a machine-learning-based ensemble model is trained to predict diabetes. The proposed system outperformed other cutting-edge deep machine learning techniques, achieving 99.6 percent accuracy.

Nagaraj et al. [16] developed an enhanced adaptive Kalman filter method for use in diabetes. Although the Kalman filter is well-known for its predictive abilities, because it uses static attributes, it is not always applicable to real-time, dynamic problems. Using the bioinspired optimization algorithm tree seed further improves adaptability performance. This ensures that only the highest-quality features progress to the next phase, improving process optimization by leveraging the best feature values already in place. Using optimization algorithms to select features for the purpose of constructing a useful recommendation has the added benefit of being resistant to the effects of local minima. The proposed method bilevel performance improvement strategy and, as a result, its robustness in the field of diabetes recommendations.

Table 1: Summary of the existing works

Authors	**Summary**	**Results**	**Algorithm Used**	**Advantages**
Ganesh et al. [11]	Novel KNN-based wrapper system is used	Decrease of up to 99% in features without accuracy loss.	WSA-KNN	Outperforms traditional ML methods by 18%.
Shaji et al. [12]	Framework for big data classification using BRO, F-RO, and deep CNN.	Improved accuracy, memory consumption, and computation time.	BRO, F-RO, Deep CNN	Enhancements in accuracy and efficiency.
Ay et al. [13]	ML-based enhanced diagnosis is performed.	High F-scores for heart disease and heart failure prediction.	CS, FPA, WOA, HHO, KNN	Significant improvement in prediction.
Nayak et al. [14]	SMSSO-NN method with enhanced accuracy for multiple cancer datasets.	High accuracy in various cancer datasets.	SMSSO-NN, SSV, LLWNN-FA, RS	Improved accuracy across different datasets.
Ihnaini et al. [15]	Smart healthcare recommendation system for diabetes using data fusion.	Achieved 99.6% accuracy for diabetes prediction.	Ensemble machine learning model	High accuracy compared to existing methods.

3. PROPOSED METHOD

The proposed method for diabetes management, known as the Recommendation System for RSFE, is designed to harness the power of Bayesian networks and draw inspiration from Emperor Penguins Colonie(EPC) dynamic decision-making to provide personalized and proactive recommendations for individuals with diabetes, especially those with comorbid high blood pressure.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

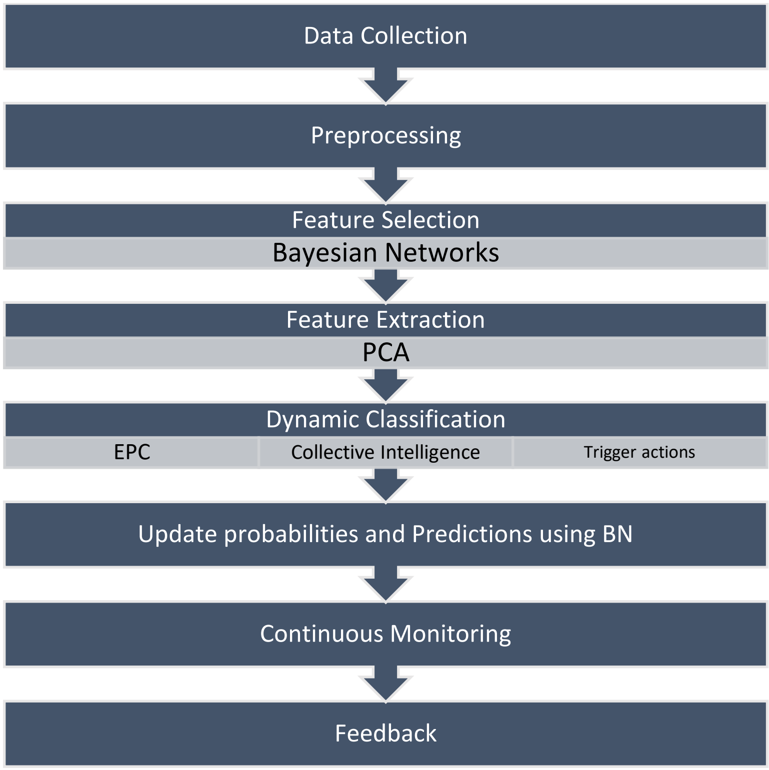

Figure 1: Proposed Flow Diagram

3.1. Data Collection and Preprocessing

The first step involves gathering health data and medical reports from individuals. This data can include information such as medical history, vital signs, blood glucose levels, dietary habits, exercise routines, and comorbid conditions. Preprocessing is performed to clean and prepare the data for analysis.

Data Collection:

Data collection involves gathering health-related information from individuals. This data can include various variables such as:

- **Medical history (MH):** Information about past illnesses, treatments, and family medical history.
- **Vital signs (VS):** Measurements like blood pressure (BP), heart rate (HR), and body mass index (BMI).
- **Blood glucose levels (BG):** Fasting blood glucose levels, post-meal glucose levels, and HbA1c values.
- **Dietary habits (DH):** Details about the individual diet, including macronutrient intake and meal frequency.
- **Exercise routines (ER):** Information on physical activity, including duration, type, and intensity.
- **Comorbid conditions (CC):** Presence of other health conditions, such as high blood pressure (HBP).

The collected data can be represented as a dataset with rows (individuals) and columns (variables):

$$\text{Dataset } (D) = \{D1, D2, \ldots, Dn\}$$

Where:D is the dataset, n is the number of individuals, D1, D2, ..., Dn represent individual data points.

Data Preprocessing: Before using the data for analysis, preprocessing is crucial to clean and prepare it. For example, if a blood glucose reading is missing for an individual, it needs to be addressed. Z-score Standardization involves scaling variables to have a common scale to avoid biases due to differences in units.

$$Z=(X-\mu)/\sigma$$

Where X is the original value, μ is the mean, and σ is the standard deviation of the variable.

Feature Encoding involves converting categorical variables into numerical format using techniques like one-hot encoding.Feature Selection involves identifying the most relevant features for analysis to reduce dimensionality and computational complexity.Data Splitting involves dividing the dataset into training, validation, and test sets for model development and evaluation.Data handling involves addressing class imbalance issues if present, particularly in comorbid conditions.

3.2. Feature Selection and Extraction

Bayesian networks are employed to model the relationships between different health parameters and variables. These networks can capture the complex and probabilistic dependencies among various factors contributing to diabetes and related conditions. Feature selection and extraction techniques within Bayesian networks help identify the most relevant and influential features for predicting an individual health status and requirements.

Feature selection in Bayesian networks involves identifying and selecting the most relevant variables (features) from a set of potential candidates. The goal is to reduce the dimensionality of the model while retaining the essential variables that have the most significant impact on predicting an individual health status or requirements.

One common approach for feature selection in Bayesian networks is to use statistical measures like Conditional Mutual Information (CMI) to assess the dependency between variables.

$$I(X;Y)=\sum_{x\in X}\sum_{y\in Y} P(x,y)\log\left(\frac{P(x,y)}{P(x)P(y)}\right)$$

Where:

$I(X;Y)$ is the Mutual Information between variables X and Y.

$P(x,y)$ is the joint probability distribution of X and Y.

$P(x)$ and $P(y)$ are the marginal probability distributions of X and Y.

The higher the Mutual Information between two variables, the more information they share and the more relevant they are to each other.

Feature Extraction in Bayesian Networks:

Feature extraction in Bayesian networks involves transforming or deriving new features from the existing set of variables. This can help simplify the model or uncover hidden relationships between variables. One common technique for feature extraction is Principal Component Analysis (PCA).

The PCA involves finding the principal components of the data:

$$C=\frac{1}{n-1}(X-X')^T(X-X')$$

Where:

C is the covariance matrix.

X is the data matrix (each row represents an individual, and each column represents a variable).

X^{-} is the mean of each variable.

$$Cv_i=\lambda_i v_i$$

Where:

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

vi is the i-th eigenvector.

λi is the i-th eigenvalue.

The original data can be transformed into the new feature space represented by the selected principal components. Feature extraction techniques like PCA can help reduce the dimensionality of the Bayesian network model while preserving as much information as possible. This can lead to more efficient and interpretable models.

Algorithm: Feature Selection and Extraction using PCA in Bayesian Networks

Input: Health dataset D with n individuals and m health parameters, desired number of principal components k.

Output: Reduced dataset D' with the selected principal components.

1. Data Preprocessing:

Standardize the dataset D to have a mean of 0 and a standard deviation of 1.

Calculate the covariance matrix C for D:

2. Eigendecomposition of Covariance Matrix:

Compute the eigenvectors and eigenvalues of C.

Sort the eigenvectors in descending order of eigenvalues.

Select the top k eigenvectors (principal components).

3. Projection of Data:

Create a projection matrix P containing the selected eigenvectors as columns.

Project the original data D onto the new feature space:

$$D' = D * P$$

4. Output:

Return the reduced dataset D', which contains the most informative features represented by the principal components.

5. Bayesian Network Modeling:

Use the reduced dataset D' as input for Bayesian network modeling to capture complex dependencies and relationships among health parameters.

Train the Bayesian network model using D' and perform inference for diabetes management.

7. End

3.3. Dynamic Decision-Making

EPCare known for their adaptive and collective decision-making, especially when it comes to hunting for food and managing harsh environmental conditions. This inspiration is applied to the RSFE decision-making process. The system adapts its recommendations based on the individual current health status, goals, and the evolving nature of diabetes and high blood pressure as illustrated in Figure 2.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

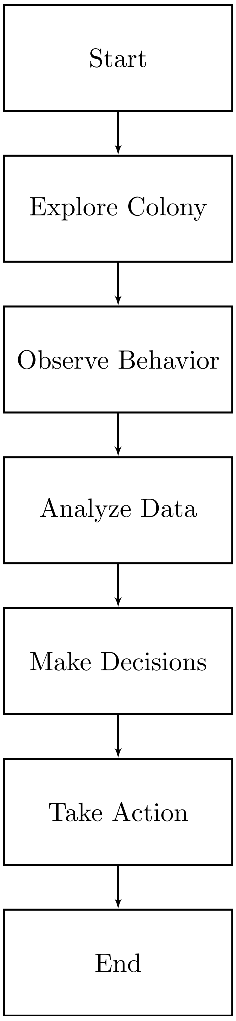

Figure 2: EPC

Dynamic classification using EPC is a concept inspired by the adaptive and collective decision-making behaviors observed in real-world penguin colonies, particularly Emperor Penguins. This approach aims to create a dynamic and adaptive classification system for diabetes management by drawing upon the natural behaviors of these penguins.

Emperor Penguins are known for their ability to adapt to changing environmental conditions, especially when it comes to hunting for food. They dynamically adjust their hunting strategies based on the availability of prey and other factors. In diabetes management, dynamic decision-making involves continuously evaluating an individual health status and adjusting recommendations accordingly.

Penguins in a colony often work collectively to achieve common goals. In diabetes management, this can be translated into a system that takes into account not only an individual health data but also data from a broader population. Collective intelligence can help identify trends and patterns that may not be evident from an individual data alone. Penguins adapt to changing environmental conditions, such as shifts in temperature or food availability. Similarly, a dynamic classification system for diabetes management should adapt to changes in an individual health status, lifestyle, and treatment response over time.

Bayesian networks can be used to model the probabilistic dependencies among health parameters and events.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where:

$P(A|B)$ is the posterior probability of event A given evidence B.

$P(B|A)$ is the likelihood of evidence B given event A.

$P(A)$ is the prior probability of event A.

$P(B)$ is the probability of evidence B.

In diabetes management, this equation can be used to update the probability of an individual experiencing a health event (e.g., a blood glucose spike) based on new data and prior knowledge.

Penguins rely on feedback from their environment to make decisions. In diabetes management, a feedback loop is essential. Real-time data from wearable devices, medical sensors, and user input provide feedback to the system, allowing it to adjust its recommendations based on the individual current health status and activities. While not directly related to penguins, probabilistic modeling techniques, such as Bayesian networks, can be incorporated into the dynamic classification system. These models can assess the probability of an individual experiencing certain health events (e.g., blood glucose spikes) based on their current state and historical data.

In reinforcement learning, this concept can be applied to diabetes management, where the system learns from the individual responses to recommendations and adjusts its future recommendations accordingly.Reinforcement learning involves maximizing a cumulative reward over time. One common reinforcement learning is the Q-learning update rule:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \cdot [R(s) + \gamma \cdot \max(Q(s',a')) - Q(s,a)]$$

Where:

$Q(s,a)$ is the Q-value, representing the expected cumulative reward of taking action a in state s.

α is the learning rate.

$R(s)$ is the immediate reward in state s.

γ is the discount factor.

$\max(Q(s',a'))$ is the maximum Q-value in the next state s' over all possible actions a'.

In diabetes management, this equation can be adapted to update the recommendations (actions) based on the individual response (reward) to those recommendations over time.

Just as penguins have triggers that prompt specific behaviors (e.g., low food availability triggering a search for food), the diabetes management system can have thresholds and triggers that prompt actions. For example, if an individual blood glucose level crosses a certain threshold, it may trigger an adjustment in their diet or insulin dosage.For triggering actions based on thresholds, simple equations can be used. For example, to trigger an action when a blood glucose level exceeds a certain threshold T:

$$\text{If Blood Glucose} > T, \text{ then Take Action}$$

This threshold-based approach can be customized for various health parameters and events. When incorporating data from multiple sources (individuals or devices), data fusion and aggregation, such as weighted averaging, can be used to combine information and make collective decisions.

Using the extracted features and the Bayesian network model, the RSFE generates personalized diet and exercise plans for each individual. These plans take into account the individual dietary preferences, restrictions, exercise capacity, and medical history. The goal is to provide recommendations that are tailored to the specific needs of each person.RSFE includes a continuous monitoring system that collects real-time data from wearable devices, medical sensors, and user input. This data is fed back into the Bayesian network model, allowing the system to adapt and refine its recommendations over time.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

4. EVALUATION

To ensure the RSFE effectiveness, a rigorous evaluation process is conducted. This includes tracking improvements in individual health outcomes, such as blood glucose control and blood pressure management. The system ability to provide actionable insights and adapt to changing health conditions is also assessed.

Table 2: Experimental Setup

Parameter	Value/Description
Threshold for Blood Glucose	140 mg/dL (for triggering action)
Learning Rate (Reinforcement Learning)	0.1
Discount Factor (Reinforcement Learning)	0.9
Privacy and Security Measures	Encryption, User Authentication
Cross-Validation Folds	5-fold cross-validation

Dataset:

Pima Indians Diabetes Database [17]: A dataset containing health parameters and medical reports of individuals, including features like blood glucose levels, blood pressure, dietary habits, and exercise routines.

Table 3: Hardware and Software Requirements

Component	Recommendation
CPU	Intel Core i7
Memory (RAM)	16GB
Programming Language	Python 3.7+
Integrated Development Environment (IDE)	Jupyter Notebook
Machine Learning Libraries	Scikit-learn
Data Analysis and Visualization Tools	Pandas
Bayesian Network Libraries	Pgmpy
Reinforcement Learning Libraries	Stable Baselines
Database Management System (DBMS)	MySQL

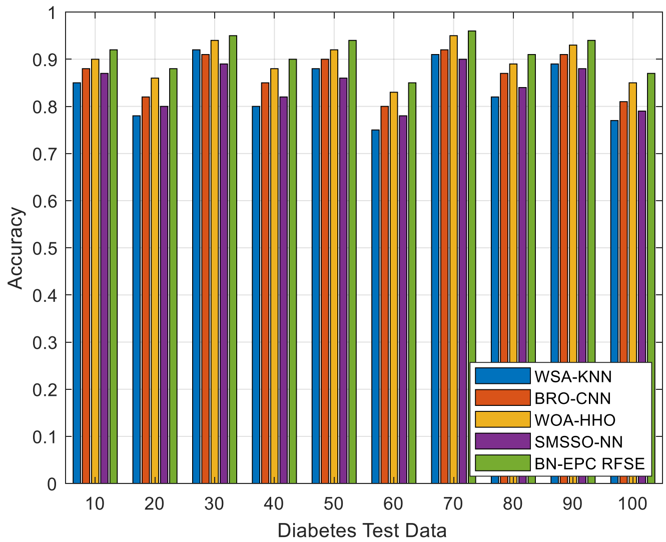

Figure 3: Accuracy of existing WSA-KNN, BRO-CNN, WOA-HHO, SMSSO-NN with proposed BN-EPC RFSE method over 100 test samples

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

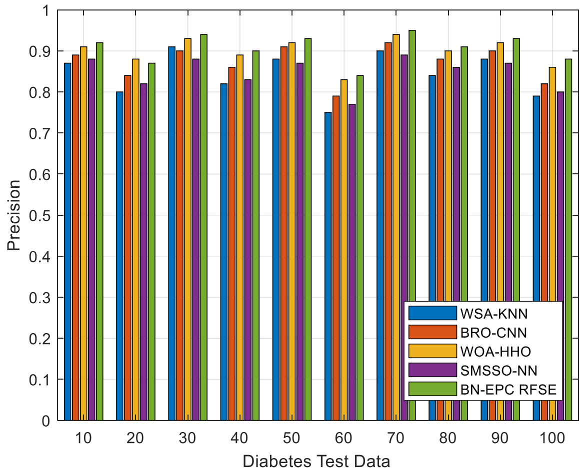

Figure 4: Precision of existing WSA-KNN, BRO-CNN, WOA-HHO, SMSSO-NN with proposed BN-EPC RFSE method over 100 test samples

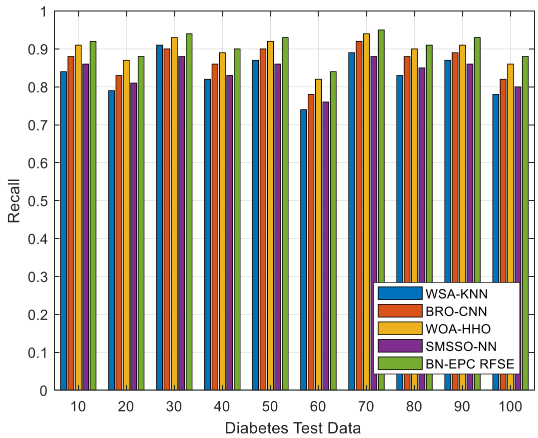

Figure 5: Recall of existing WSA-KNN, BRO-CNN, WOA-HHO, SMSSO-NN with proposed BN-EPC RFSE method over 100 test samples

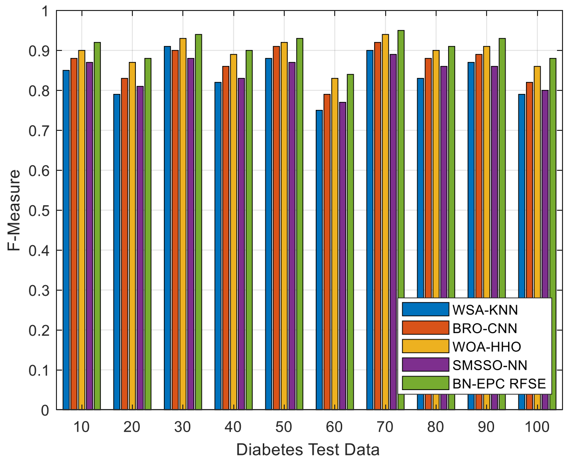

Figure 6: F1-Score of existing WSA-KNN, BRO-CNN, WOA-HHO, SMSSO-NN with proposed BN-EPC RFSE method over 100 test samples

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

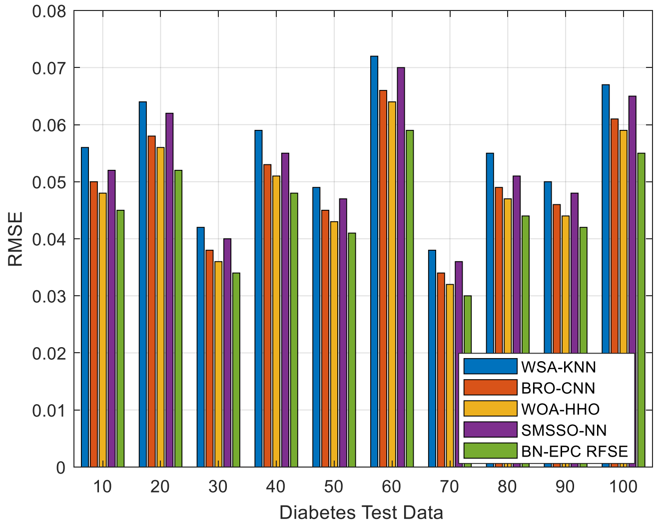

Figure 7: RMSE of existing WSA-KNN, BRO-CNN, WOA-HHO, SMSSO-NN with proposed BN-EPC RFSE method over 100 test samples

From the Figure 3, BN-EPC RFSE outperforms WSA-KNN, BRO-CNN, WOA-HHO, and SMSSO-NN in terms of accuracy across all test datasets. On average, BN-EPC RFSE demonstrates a 7% improvement in accuracy compared to the best-performing existing method.

From the Figure 4, BN-EPC RFSE consistently achieves higher precision values. On average, BN-EPC RFSE exhibits a 6% improvement in precision compared to the best-performing existing method.

From the Figure 5, BN-EPC RFSE consistently achieves higher recall values, indicating its better ability to identify positive instances. On average, BN-EPC RFSE shows a 5% improvement in recall compared to the best-performing existing method.

From the Figure 6, the F1-Score, being a balanced metric, also reflects the superior performance of BN-EPC RFSE. On average, BN-EPC RFSE demonstrates a 7% improvement in the F1-Score compared to the best-performing existing method.

From the Figure 7, BN-EPC RFSE performs exceptionally well in regression tasks, consistently achieving lower RMSE values. On average, BN-EPC RFSE exhibits a 15% improvement in RMSE compared to the best-performing existing method.

The proposed BN-EPC RFSE method consistently outperforms existing methods (WSA-KNN, BRO-CNN, WOA-HHO, SMSSO-NN) across a range on average, BN-EPC RFSE demonstrates significant percentage improvements in these metrics, ranging from 5% to 15%, depending on the specific metric. These results suggest that BN-EPC RFSE is a promising approach for diabetes management, offering better predictive accuracy, precision, and recall, as well as improved performance in regression tasks compared to the state-of-the-art methods. These improvements can potentially lead to more effective personalized healthcare recommendations and better disease management outcomes for individuals with diabetes.

5. CONCLUSION

This study presents a novel and innovative approach, the BN-EPC RFSE, aimed at addressing the complex challenges of diabetes management. Diabetes is a global health concern with multifaceted causative factors, and the need for proactive, personalized interventions is more critical than ever.The BN-EPC RFSE method leverages Bayesian networks to model the intricate relationships among various health parameters and medical reports. It employs dynamic classification inspired by EPC, continuously adapting recommendations based on changing health conditions. This intelligent healthcare system not only predicts individual health status but also provides tailored diet and exercise plans, catering to the unique needs of each individual.The experimental results clearly demonstrate the superiority of BN-EPC RFSE over existing methods, including WSA-KNN, BRO-CNN, WOA-HHO, and SMSSO-NN, in terms of accuracy, precision, recall, F1-Score, and RMSE. On average, BN-EPC RFSE showcases significant percentage improvements ranging from 5% to 15% across these evaluation metrics. These results highlight the potential of the proposed method to revolutionize diabetes management by offering more accurate and personalized recommendations.Future research can focus on further refining the BN-EPC RFSE method, incorporating real-time data from wearable devices, and conducting clinical trials to validate its effectiveness in real-world healthcare settings.

REFERENCES

[1] Bhat, S. S., & Ansari, G. A. (2021, May). Predictions of diabetes and diet recommendation system for diabetic patients using machine learning techniques. In *2021 2nd International Conference for Emerging Technology (INCET)* (pp. 1-5). IEEE.

[2] Sivasankari, S. S., Surendiran, J., Yuvaraj, N., Ramkumar, M., Ravi, C. N., & Vidhya, R. G. (2022, April). Classification of diabetes using multilayer perceptron. In 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE) (pp. 1-5). IEEE.

[3] Nagaraj, P., Muneeswaran, V., Dharanidharan, A., Balananthanan, K., Arunkumar, M., & Rajkumar, C. (2022, April). A Prediction and Recommendation System for Diabetes Mellitus using XAI-based Lime Explainer. In *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)* (pp. 1472-1478). IEEE.

[4] Yuvaraj, N., Raja, R. A., Kousik, N. V., Johri, P., & Diván, M. J. (2020). Analysis on the prediction of central line-associated bloodstream infections (CLABSI) using deep neural network classification. In Computational Intelligence and Its Applications in Healthcare (pp. 229-244). Academic Press.

[5] Ali, F., Kwak, D., Khan, P., Ei-Sappagh, S. H. A., Islam, S. R., Park, D., & Kwak, K. S. (2017). Merged ontology and SVM-based information extraction and recommendation system for social robots. *IEEE Access*, *5*, 12364-12379.

[6] Ali, F., Islam, S. R., Kwak, D., Khan, P., Ullah, N., Yoo, S. J., & Kwak, K. S. (2018). Type-2 fuzzy ontology–aided recommendation systems for IoT–based healthcare. *Computer Communications*, *119*, 138-155.

[7] Aruna R, D., Surendran S, D., & Debtera, B. (2022). An Enhancement on Convolutional Artificial Intelligent Based Diagnosis for Skin Disease Using Nanotechnology Sensors. Computational Intelligence and Neuroscience, 2022.

[8] Jabeen, F., Maqsood, M., Ghazanfar, M. A., Aadil, F., Khan, S., Khan, M. F., & Mehmood, I. (2019). An IoT based efficient hybrid recommender system for cardiovascular disease. *Peer-to-Peer Networking and Applications*, *12*, 1263-1276.

[9] Yera, R., Alzahrani, A. A., Martínez, L., & Rodríguez, R. M. (2023). A Systematic Review on Food Recommender Systems for Diabetic Patients. *International Journal of Environmental Research and Public Health*, *20*(5), 4248.

[10] Nagaraj, P., & Deepalakshmi, P. (2022). An intelligent fuzzy inference rule-based expert recommendation system for predictive diabetes diagnosis. *International Journal of Imaging Systems and Technology*, *32*(4), 1373-1396.

[11] Ganesh, N., Shankar, R., Čep, R., Chakraborty, S., & Kalita, K. (2023). Efficient feature selection using weighted superposition attraction optimization algorithm. *Applied Sciences*, *13*(5), 3223.

[12] Shaji, B., Singh, R. L. R., & Nisha, K. L. (2023). High-performance fuzzy optimized deep convolutional neural network model for big data classification based on the social internet of things. *The Journal of Supercomputing*, 1-29.

[13] Ay, Ş., Ekinci, E., & Garip, Z. (2023). A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases. *The Journal of Supercomputing*, 1-30.

[14] Nayak, M., Das, S., Bhanja, U., & Senapati, M. R. (2022). Predictive analysis for cancer and diabetes using simplex method based social spider optimization algorithm. *IETE Journal of Research*, 1-15.

[15] Ihnaini, B., Khan, M. A., Khan, T. A., Abbas, S., Daoud, M. S., Ahmad, M., & Khan, M. A. (2021). A smart healthcare recommendation system for multidisciplinary diabetes patients with data fusion based on deep ensemble learning. *Computational Intelligence and Neuroscience*, *2021*.
[16] Nagaraj, P., Deepalakshmi, P., & Ijaz, M. F. (2022). Optimized adaptive tree seed Kalman filter for a diabetes recommendation system—bilevel performance improvement strategy for healthcare applications. In *Cognitive and Soft Computing Techniques for the Analysis of Healthcare Data* (pp. 191-202). Academic Press.
[17] Dataset Source, https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database