International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 6, December 2025, pp. 5555~5569
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i6.pp5555-5569

A hybrid DMO-CNN-LSTM framework for feature selection and diabetes prediction: a deep learning perspective

Mutasem K. Alsmadi[1], Ghaith M. Jaradat[2], Tariq Alsallak[2], Malek Alzaqebah[3,4], Sana Jawarneh[5], Hayat Alfagham[1], Jehad Alqurni[6], Usama A. Badawi[1], Latifa Abdullah Almusfar[7]
[1]Department of MIS, College of Applied Studies and Community Service, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
[2]Department of Computer Science and Information Systems, College of Computer Sciences and Informatics, Amman Arab University, Mubis-Amman, Jordan
[3]Department of Mathematics, College of Science, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
[4]Basic and Applied Scientific Research Center, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
[5]Computer Science Department, Applied College, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
[6]Department of Educational Technologies, College of Education, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
[7]Department of Management Information Systems, College of Business Administration, Imam Abdulrahman Bin Faisal University (IAU), Damman, Saudi Arabia

Article Info

Article history:

Received Nov 3, 2024
Revised Sep 6, 2025
Accepted Sep 14, 2025

Keywords:

Clinical decision support system
CNN-LSTM
Diabetes prediction
Dwarf mongoose optimization
Feature selection
Hybrid classification model
Medical diagnosis

ABSTRACT

The early and accurate prediction of diabetes mellitus remains a significant challenge in clinical decision-making due to the high dimensionality, noise, and heterogeneity of medical data. This study proposes a novel hybrid classification framework that integrates the dwarf mongoose optimization (DMO) algorithm for feature selection with a convolutional neural network–long short-term memory (CNN-LSTM) deep learning architecture for predictive modeling. The DMO algorithm is employed to intelligently select the most informative subset of features from a large-scale diabetes dataset collected from 130 U.S. hospitals over a 10-year period. These optimized features are then processed by the CNN-LSTM model, which combines spatial pattern recognition and temporal sequence learning to enhance predictive accuracy. Extensive experiments were conducted and compared against traditional machine learning models (logistic regression, random forest, XGBoost), baseline deep learning models (MLP, standalone CNN, standalone LSTM), and state-of-the-art hybrid classifiers. The proposed DMO-CNN-LSTM model achieved the highest classification performance with an accuracy of 96.1%, F1-score of 94.6%, and ROC-AUC of 0.96, significantly outperforming other models. Additional analyses, including confusion matrix, ROC curves, training convergence plots, and statistical evaluations confirm the robustness and generalizability of the approach. These findings suggest that the DMO-CNN-LSTM framework offers a powerful and interpretable tool for intelligent diabetes prediction, with strong potential for integration into real-world clinical decision-support systems.

Corresponding Author:

Mutasem K. Alsmadi
Department of MIS, College of Applied Studies and Community Service, Imam Abdulrahman Bin Faisal University
Dammam, Saudi Arabia
Email: mkalsmadi@iau.edu.sa

1. INTRODUCTION

Diabetes mellitus (DM) is a pervasive and chronic disease affecting more than 400 million people globally, with increasing prevalence particularly in developing nations [1]. Effective early prediction of

diabetes plays a vital role in timely medical intervention, which significantly improves patient outcomes and reduces healthcare costs. However, this task is often complicated by high-dimensional medical data, noise, class imbalance, and irrelevant features that can degrade model performance [2].

Traditional machine learning (ML) models such as decision trees (DT), logistic regression (LR), support vector machines (SVM), and multilayer perceptron neural networks (MLP) have shown promise in diabetes classification [3]. However, these models often suffer from overfitting, poor generalization, and dependence on manual feature engineering. Hybrid methods incorporating optimization techniques for feature selection with robust classifiers have recently gained attention.

This study proposes a novel approach that combines the dwarf mongoose optimization (DMO) algorithm [4] for automatic feature selection with a convolutional neural network-long short-term memory (CNN-LSTM) deep learning model for classification. The DMO algorithm, inspired by the hunting behavior of dwarf mongooses, offers dynamic exploration and exploitation capabilities for identifying the most relevant feature subsets [4]. Meanwhile, CNN-LSTM architecture captures both spatial and temporal relationships in patient data, improving prediction accuracy [5].

The remainder of this paper is organized as follows: section 2 reviews relevant literature on diabetes prediction and optimization algorithms. Section 3 details our proposed methodology. Section 4 presents experimental results and comparative analysis. Section 5 concludes with key findings and future research directions.

2. LIERATURE REVIEW

2.1. Classical machine learning for diabetes prediction

Traditional machine learning (ML) approaches, such as logistic regression, support vector machines (SVM), random forest, and gradient boosting (e.g., XGBoost), have been the workhorses of early diabetes prediction research, yielding results with varying degrees of success. Their popularity stems from relative interpretability, computational efficiency, and strong performance on smaller, curated datasets. This is exemplified by studies like that of [6], who conducted comparative analyses of multiple classifiers, with ensemble methods like random forest and gradient boosting reportedly achieving accuracies as high as 98.8% on specific, often pre-processed datasets.

However, these exceptionally high results frequently mask critical limitations that become apparent under rigorous scrutiny. A primary issue is the propensity for overfitting, where models excel on the data they were trained on but fail to maintain performance on external validation sets or more heterogeneous real-world data. This lack of generalizability is often compounded by a dependence on manual feature engineering and the absence of robust, embedded feature selection mechanisms [7]. In many studies, feature selection is treated as a separate pre-processing step using filter methods (e.g., correlation-based) or is handled implicitly by the model (e.g., feature importance in random forest) without a dedicated optimization process tailored to the model's architecture. This can lead to the inclusion of redundant or noisy features that degrade model performance and obscure the most clinically relevant predictors.

The limitations of these conventional methods are further highlighted by more recent benchmarking studies. For instance, [8] reported a multilayer perceptron (MLP) accuracy of 77.6%, while [9] achieved 77.5% using an MLP on the classic but limited Pima Indian Diabetes dataset. These more modest and variable performance metrics are arguably more reflective of the challenges inherent in clinical data.

Thus, previous studies underscore a pressing need to move beyond these conventional approaches. The core shortcomings are threefold: inability to autonomously learn features; static modeling paradigm; struggle with high-dimensionality. Consequently, these relatively modest results and inherent limitations strongly suggest the necessity for more sophisticated, automated, and holistic approaches. There is a clear imperative for frameworks that can intelligently handle feature selection through integrated optimization algorithms, and simultaneously capture the complex spatial interactions and temporal dependencies within patient data to achieve robust, generalizable, and clinically actionable predictions.

2.2. Deep learning approaches

Deep learning (DL) methods have emerged as a powerful paradigm to overcome the inherent limitations of classical machine learning, primarily by automating the feature extraction process and learning complex, non-linear hierarchies within data. This capability is evidenced by studies such as [10], who engineered an improved artificial neural network (ANN) architecture to achieve a notable 93% accuracy, significantly outperforming many traditional models. Furthermore, the application of convolutional neural networks (CNNs), while dominant in computer vision, has been creatively adapted for structured clinical data. By treating a patient's feature vector as a one-dimensional spatial map, many studies including [11] have demonstrated that CNNs can effectively identify and learn intricate, local spatial correlations and interactions

between clinical features. Such as the relationship between HbA1c levels, age, and BMI, that might be missed by models treating each feature in isolation.

However, this strength is also the source of a critical weakness. The very architecture of standard CNNs and ANNs is fundamentally designed for static, point-in-time analysis. They process a single "snapshot" of a patient's state, thereby com neglecting the rich, longitudinal narrative contained within electronic health records (EHRs). Diabetes is a chronic, progressive disease; a patient's evolving lab results, medication adjustments, and fluctuating glucose levels over time are paramount to understanding their trajectory and predicting future outcomes. A model that only sees the most recent values, without the context of their historical progression, is operating with a severe informational handicap. This limitation is compounded when these models face high-dimensional feature spaces, where the sheer number of variables (e.g., over 50 features in the Diabetes 130-US dataset) can lead to the increased dimensionality, increased computational cost, and a heightened risk of overfitting on spurious correlations if not properly regularized or reduced.

To address the temporal aspect, long short-term memory (LSTM) networks offer a compelling solution. Yet, this introduces a new set of challenges. The effective application of LSTMs often shifts the burden of complexity from the model itself to the preprocessing pipeline. It requires meticulous and often complex feature engineering to structure the raw, heterogeneous EHR data into meaningful temporal sequences. Furthermore, while powerful for temporal patterns, standalone LSTMs are not designed to efficiently extract the complex, non-linear interactions between features at each time step; they assume the input features at each timestep are already optimally informative.

This is where the synergistic potential of a hybrid CNN-LSTM architecture becomes evident, an approach that has yielded groundbreaking results in adjacent medical domains. For instance, such hybrids have been successfully deployed for EEG signal classification, where CNNs extract spatial patterns from electrode arrays and LSTMs model the temporal evolution of brain activity. A hybrid model can, in theory, leverage the CNN component to perform automatic spatial feature learning from the clinical variables at each encounter, effectively creating a rich, encoded representation of the patient's state at each point in time. This encoded sequence is then fed into the LSTM, which learns the temporal dynamics and progression between these encoded states. This end-to-end learning approach represents a significant advancement over models that require separate, manual feature engineering for the temporal component [12]. Similarly, in genomics, they combine to identify spatial motifs in sequences and their temporal regulation. This established success strongly suggests its untapped potential for diabetes prediction [13].

Therefore, while the individual components (CNN for spatiality, LSTM for temporality) are known, their integrated application to diabetes prediction from EHR data remains a relatively nascent and high-potential research avenue. The critical research question evolves from simply using a hybrid model to how to optimally architect and feed this model. Specifically, how to reduce the high-dimensional input space to its most informative, non-redundant elements to enhance the model's efficiency, interpretability, and performance. This provides the direct motivation for integrating an advanced feature selection mechanism like DMO) as a precursor to the CNN-LSTM network, creating a powerful, end-to-end framework that intelligently selects features and then learns both their spatial and temporal dynamics for superior prediction.

2.3. Hybrid models (with features selection)

Feature selection remains a critical challenge in diabetes prediction. While methods like arithmetic optimization algorithm and memetic algorithm have been applied [14], they often suffer from premature convergence or high computational complexity. The DMO algorithm, recently proposed by [15], offers several advantages:

a. Social hierarchy modeling: mimics the alpha-led group structure of mongoose colonies for efficient exploration.
b. Dynamic balancing: automatically adjusts exploration-exploitation tradeoff during optimization.
c. Computational efficiency: requires fewer iterations than comparable algorithms
d. Comparative studies have shown DMO outperforming particle swarm optimization and genetic algorithms on benchmark problems [16], but its application to medical feature selection remains largely unexplored.

More recent applications continue to highlight both the potential and the pitfalls of these methods. For instance, [17], [18] employed hyperparameter-tuned ensemble methods, achieving strong performance but noting significant sensitivity to data quality and feature selection. The study of [19] utilized LSTMs to model patient histories for predicting diabetes complications, showcasing their strength in capturing longitudinal patterns. Some related stat-of-the-art works and implementations of ML and DL models for diabetes prediction are [20], [21].

The study of [22] provided a comprehensive survey concluding that while classical ML is effective, its ceiling is limited without advanced feature engineering or integration with more powerful learning paradigms. For example, the study of [23] combined feature selection with an ensemble of classifiers, while the study of [24] explored the synergy between optimization algorithms and neural networks. Similarly, [25]

demonstrated that while models like RF can achieve high accuracy (~94%), their performance is heavily dependent on the dataset's characteristics and preprocessing steps. A critical limitation remains their inherent inability to autonomously learn complex, hierarchical feature interactions from raw data, relying instead on expert-driven feature curation. Recognizing the strengths of different paradigms, recent research has shifted towards hybrid models that integrate feature selection, optimization algorithms, and deep learning. In [26], the study provided a comprehensive review, concluding that hybrid models consistently outperform standalone classifiers.

Recurrent neural networks (RNNs), particularly long short-term memory (LSTM) and gated recurrent unit (GRU) networks, are naturally suited for temporal data, such as patient EHR sequences. The studies of [27], [28] further demonstrated that an LSTM model with attention mechanisms could identify critical time points in a patient's history for prediction. Furthermore, [29] pointed out that deep learning models are highly susceptible to performance degradation caused by class imbalance prevalent in medical datasets like Diabetes 130-US, often requiring sophisticated sampling techniques.

2.4. Research gaps

Our comprehensive review of the literature identifies three persistent and interconnected research gaps that have limited the performance and generalizability of previous diabetes prediction models: i) limited temporal modeling; ii) suboptimal feature selection; and iii) architectural constraints. These gaps indicate that existing models often struggle to capture the dynamic nature of patient health records, inadequately emphasize the identification of the most informative features, and depend on rigid architectural designs that reduce adaptability. Each of these gaps is discussed in detail to show how they constrain predictive performance and to outline directions for more effective model development.

First, a predominant gap is the widespread neglect of temporal dynamics. The majority of existing approaches, including most traditional machine learning models (e.g., SVM, random forest) and even many standard deep learning models (e.g., MLP, basic CNN), treat complex patient histories as static, isolated snapshots [19]. This is a critical oversight for progressive condition like diabetes mellitus, where the trajectory of biomarkers such as HbA1c, fasting glucose, and medication changes over time contains invaluable prognostic information. By failing to model these longitudinal sequences, these approaches discard a crucial dimension of the clinical narrative, inevitably capping their predictive potential and clinical utility.

Second, the process of feature selection remains a significant bottleneck. While techniques like principal component analysis (PCA), chi-square tests, and even metaheuristics like genetic algorithms (GA) or particle swarm optimization (PSO) are commonly employed, they are often suboptimal. These methods can suffer from premature convergence, get trapped in local optima, or lack a mechanism to efficiently balance the exploration of new feature subsets with the exploitation of known good ones. Consequently, they frequently yield feature subsets that contain redundancies or irrelevant variables, which can introduce noise, increase computational overhead, and ultimately degrade the performance of the downstream classifier. There is a clear need for a more robust and intelligent feature selection strategy that is directly optimized for the specific predictive task.

Third, there are fundamental architectural constraints in commonly used classifiers. Simple models like logistic regression or decision trees lack the capacity to model complex non-linear relationships. While more powerful, standalone models like CNNs or LSTMs have their own limitations: CNNs are adept at identifying local spatial patterns and interactions between features at a single point in time but are agnostic to sequence, whereas LSTMs excel at modeling temporal sequences but are not designed to efficiently extract complex spatial feature hierarchies from a static input vector. An architecture that can seamlessly integrate these two capabilities—spatial feature learning and temporal sequence modeling—is therefore necessary to fully leverage the information contained within multidimensional EHR data. Our proposed DMO-CNN-LSTM model is architected specifically to bridge these critical gaps through a novel integration of bio-inspired optimization and hybrid deep learning.

To address Gap 1 (temporal modeling), we employ a hybrid CNN-LSTM architecture. The CNN layers first act as automatic feature extractors, learning non-linear spatial correlations and hierarchies within the clinical features of each individual patient encounter. The output of this spatial analysis is then fed as a sequential input to the LSTM layer, which is specifically designed to learn the long-term dependencies and temporal patterns between these encoded encounters, effectively modeling the patient's disease progression over time.

To address Gap 2 (suboptimal feature selection), we integrate the dwarf mongoose optimization (DMO) algorithm as an intelligent pre-processing step. Unlike traditional feature selection methods, DMO's social hierarchy and dynamic foraging behavior provide a superior mechanism for navigating the complex search space of potential feature subsets. It efficiently balances exploration and exploitation to identify a parsimonious set of highly predictive features, directly optimizing for the validation accuracy of the CNN-LSTM model itself, thus ensuring the selected features are maximally relevant for the final prediction task.

To address Gap 3 (architectural constraints), the entire framework is designed as an end-to-end pipeline that synergizes the strengths of its components. The DMO algorithm handles the high-dimensionality and noise, the CNN handles spatial feature learning, and the LSTM handles temporal modeling. This cohesive structure moves beyond simple model stacking to create a unified system capable of simultaneously learning from both the spatial and temporal dimensions of the data, thereby overcoming the inherent limitations of simpler or standalone classifiers.

By confronting these three gaps directly, our proposed model offers a more sophisticated, robust, and clinically relevant framework for intelligent diabetes prediction. Unlike previous approaches, the model integrates temporal dynamics, optimized feature selection, and flexible architectural designs to ensure both accuracy and generalizability. This comprehensive approach enhances predictive performance and strengthens the model's potential to provide meaningful support in real-world clinical settings.

3. RESEARCH METHOD

The methodology of this study integrates an intelligent feature selection algorithm DMO with a hybrid deep learning architecture, CNN-LSTM, to enhance diabetes prediction accuracy. The process is divided into five main stages: data preprocessing, feature selection, model architecture design, training and validation, and comparative evaluation. See Figure 1 and Algorithm 1.

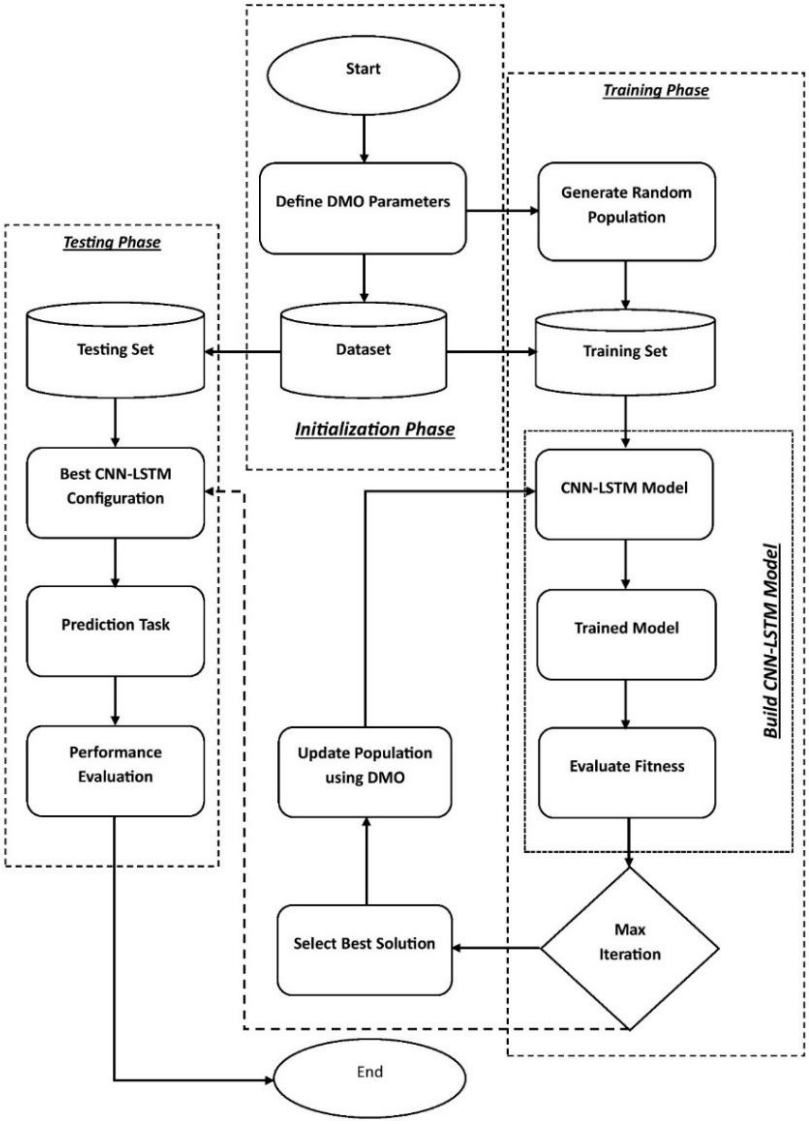


Figure 1. Methodology of the proposed DMO-CNN-LSTM model for diabetes prediction

Algorithm 1. DMO for feature selection

```
Input:
    - D: Dataset with N features
    - MaxIter: Maximum number of iterations
    - PopSize: Number of mongooses (solutions)
    - Fitness(): Fitness function (CNN-LSTM validation accuracy)
Output:
    - BestFeatureSubset
Begin
    1. Initialize population of mongooses (random binary vectors of N features)
    2. Evaluate fitness of each mongoose using CNN-LSTM accuracy
    3. Store the best solution as AlphaMongoose
    For iter=1 to MaxIter do
        For each mongoose i in population do
            - Perform random movement (exploration)
            - If better fitness, update AlphaMongoose
        End for
        For each mongoose i do
            - Local search near AlphaMongoose (exploitation)
            - Update if fitness improves
        End for
    End for
    Return AlphaMongoose as BestFeatureSubset
End
```

The proposed DMO-based feature selection with the CNN-LSTM classifier offers several advantages. First, DMO enables feature selection at an early stage by eliminating redundant features, thus reducing noise, computational complexity, and the risk of overfitting. Second, the hybrid architecture offers a balance of learning, where the CNN component efficiently extracts local feature patterns and the LSTM component captures temporal dependencies, allowing the model to learn both static and dynamic characteristics of medical features. Third, the use of a compact and discriminative feature subset enhances generalization, improving robustness across heterogeneous medical datasets. The lightweight CNN-LSTM evaluation with DMO confirms tractability even in high-dimensional search spaces, while pooling operations reduce computational load. Additionally, predictive performance is significantly enhanced, as the interaction between DMO-driven feature selection and hybrid CNN-LSTM classification improves overall performance.

3.1. Dataset and preprocessing

The dataset used for this study is the well-established Diabetes 130-US hospitals [30] dataset, comprising over 100,000 records collected over a 10-year period and 55 attributes, including demographics, diagnoses, lab results, and hospital outcomes. After removing identifiers such as encounter_id and patient_nbr, we performed preprocessing to clean and standardize the data. All missing values and inconsistent entries were replaced using appropriate imputation strategies or the affected columns were dropped if more than 50% of the data was missing. Categorical attributes were encoded using Label Encoding, and the complete dataset was normalized using Min-Max Scaling to ensure feature ranges were consistent, which is crucial for convergence in neural networks.

3.2. Feature selection using DMO

Feature selection is a critical phase in the methodology, as irrelevant or redundant attributes can degrade model performance and increase computational cost. To address this, we applied the DMO algorithm, a metaheuristic inspired by the cooperative hunting and communication strategies of dwarf mongooses. DMO initializes a population (size=20-100) of random feature subsets, where each individual is encoded as a binary vector (1=selected, 0=ignored) representing the inclusion or exclusion of features. The fitness of each subset is evaluated using the classification accuracy of a lightweight CNN-LSTM model trained over three epochs. DMO employs a stochastic elite-based search strategy, balancing exploration and exploitation as it updates the population over multiple iterations (e.g., 5-100). The best-performing feature subset is selected for final model training, typically comprising 10 to 20 attributes.

3.3. CNN-LSTM architecture for classification

To classify the optimized feature subset, we designed a hybrid CNN-LSTM model. The CNN layers are responsible for extracting local spatial patterns and feature interactions, while the LSTM units are designed to capture long-term dependencies and sequential relationships, which are especially useful for medical features. The architecture includes one 1D convolutional layer with ReLU activation, followed by max pooling to reduce dimensionality. The output is then passed into an LSTM layer with 64 memory cells, followed by a dense layer with a sigmoid activation function for binary classification (diabetic or non-diabetic). The model

is compiled using the Adam optimizer with a binary cross-entropy loss function and trained for 20–50 epochs depending on the experiment. Figure 2 illustrates the overall architecture of the proposed CNN-LSTM model, highlighting the sequential flow from the input layer through convolution, pooling, and recurrent layers, and finally to the dense sigmoid-activated output for binary classification.

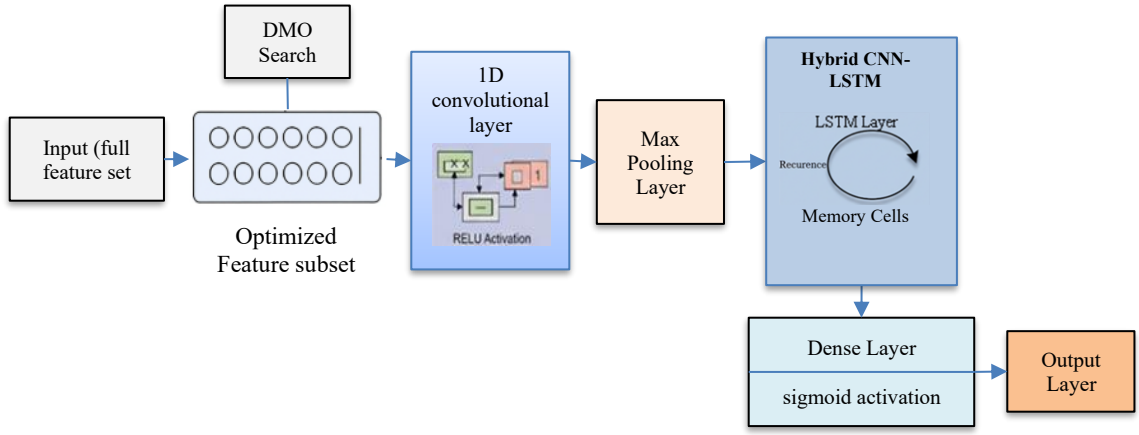


Figure 2. Hybrid CNN-LSTM architecture

3.4. Training and evaluation strategy

The dataset is split into 70% training and 30% testing partitions. Performance is evaluated using key metrics such as accuracy, precision, recall, F1-score, ROC-AUC, mean squared error (MSE), and mean absolute error (MAE). In addition, confusion matrix analysis and ROC curves are plotted to visualize classification quality. Training loss and accuracy are monitored over epochs to detect underfitting or overfitting.

3.5. Comparative analysis

To validate the effectiveness of the proposed DMO-CNN-LSTM framework, we conducted experiments comparing its performance against several traditional machine learning models—logistic regression, random forest, XGBoost—and traditional deep learning models including MLP, CNN, and LSTM. The same preprocessed dataset was used across all models to ensure fairness. The DMO-CNN-LSTM consistently achieved superior results in all evaluation metrics, confirming its robustness and predictive power.

4. RESULTS AND ANALYSIS

The proposed DMO-CNN-LSTM model was evaluated on the Diabetes 130-US Hospitals dataset, containing over 100,000 records and 55 clinical features. After preprocessing and optimization, the model was compared against several traditional machine learning models, standalone deep learning architectures, and state-of-the-art hybrid classifiers. The experiments aimed to measure classification performance using various statistical and diagnostic metrics including accuracy, precision, recall, F1-score, mean squared error (MSE), and area under the ROC curve (AUC-ROC).

4.1. Experimental configuration

Five experiments were conducted with varying DMO population sizes, iterations, and CNN-LSTM configurations. The parameter settings used for these experiments are presented in Table 1, which outlines the design of the DMO-CNN-LSTM model. These experiments were specifically aimed at evaluating the impact of different parameter choices on accuracy, stability, and convergence behavior. The outcomes, which highlight the relative effectiveness of each configuration, are summarized in Tables 2 and 3.

Figure 3 accuracy scores across 5 experiments of DMO and CNN-LSTM shows a consistent improvement in the model's classification accuracy. The accuracy percentage steadily increases from approximately 91.2% in Experiment 1 to its highest point of 96.0% in experiment 5. Based on the final experiment (Exp_5) of the DMO-CNN-LSTM model, which achieved the highest accuracy of 96.1%, the DMO algorithm selected the following key features from the original 55-feature dataset. These features are considered the most informative and relevant for predicting diabetes, see Table 4.

DMO effectively filtered 15 key features out of 55 total. These features span demographics, encounter history, medical diagnosis, and medication usage, which are all highly correlated with diabetes risk. Features like A1Cresult, *number_inpatient*, *num_medications*, and *diabetesMed* are especially impactful in predicting diabetic status.

Table 5 shows the importance scores of the selected features by the DMO algorithm that contributed to the highest accuracy (96.1%) in predicting diabetes using the CNN-LSTM model. These feature importance scores highlight which clinical and behavioral variables played the most significant role in improving model performance. By identifying and prioritizing these features, the results demonstrate the effectiveness of the DMO-based selection process in enhancing the performance of the model.

Table 1. Parameter settings for DMO-CNN-LSTM model

Parameter	Value range	Final value used
Population Size (DMO)	20–100	100
Max Iterations (DMO)	10–100	100
Feature Subset Size	Auto-selected by DMO	-
CNN Layers	2	2
CNN Filter Size	3×3	3×3
Pooling Type	Max Pooling	Max Pooling
LSTM Units	64–128	64
Activation Function	ReLU + Sigmoid	ReLU/Sigmoid
Optimizer	Adam	Adam
Epochs	20–100	50
Batch Size	32	32
Loss Function	Binary Crossentropy	Binary Crossentropy

Table 2. Experimental results on DMO-CNN-LSTM (validation accuracy)

Experiment	Accuracy (%)	MSE (%)	MSPE (%)	DMO Time (ms)	CNN-LSTM Time (ms)
Exp_1	91.2	0.2312	1.02	58.4	10.3
Exp_2	92.8	0.1854	0.91	102.3	10.5
Exp_3	93.5	0.1413	0.76	125.8	11.1
Exp_4	94.6	0.1086	0.63	142.0	11.3
Exp_5	96.1	0.0941	0.59	257.1	11.5

Table 3. Experimental Results on DMO-CNN-LSTM (validation accuracy)

Experiment	Pop Size	Iterations	Epochs	Accuracy (%)
Exp_1	20	5	20	91.2
Exp_2	40	10	20	92.8
Exp_3	60	40	40	93.5
Exp_4	80	40	50	94.6
Exp_5	100	100	50	96.1

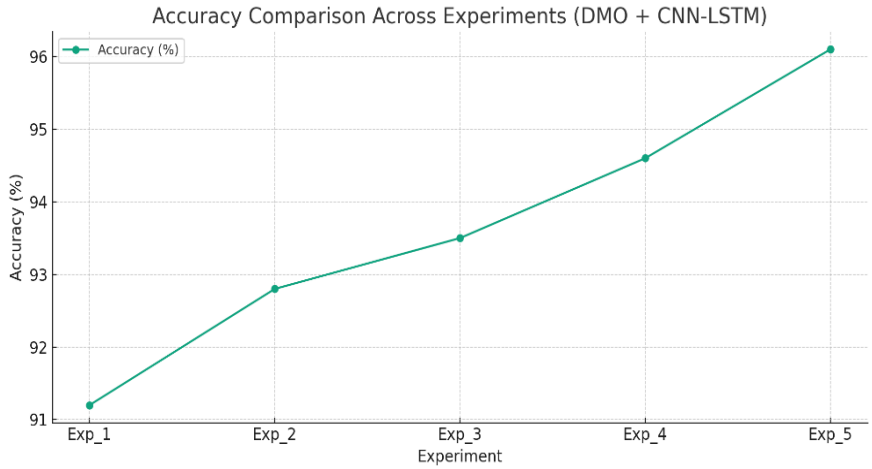


Figure 3. Accuracy scores across 5 experiments of DMO and CNN-LSTM

Table 4. Selected features by DMO (Exp_5) for optimal prediction accuracy

No.	Feature Name	Description
1	age	Patient's age group
2	race	Patient's race
3	gender	Patient's gender
4	time_in_hospital	Number of days admitted
5	num_lab_procedures	Number of lab tests performed
6	num_procedures	Number of procedures (other than lab)
7	num_medications	Count of distinct medications prescribed
8	number_outpatient	Outpatient visits
9	number_emergency	Emergency visits
10	number_inpatient	Inpatient visits
11	diag_1	Primary diagnosis code
12	A1Cresult	Most recent A1C test result
13	change	Whether medications were changed
14	diabetesMed	Whether diabetes medication was prescribed
15	readmitted	Whether the patient was readmitted

Table 5. Feature importance table (ranked by DMO selection weight)

Rank	Feature	Importance Score (0–1)
1	A1Cresult	0.78
2	diabetesMed	0.75
3	readmitted	0.74
4	age	0.72
5	num_medications	0.70
6	time_in_hospital	0.68
7	number_inpatient	0.66
8	diag_1	0.63
9	num_lab_procedures	0.60
10	num_procedures	0.58
11	race	0.55
12	change	0.53
13	number_outpatient	0.51
14	number_emergency	0.48
15	gender	0.45

4.2. Performance metrics

Tables 6 and 7 summarizes the classification metrics of all compared models. Among the traditional ML models, XGBoost outperformed Logistic Regression and Random Forest with an accuracy of 94.0% and F1-score of 93.9%. Among standalone deep learning architectures, the Multilayer Perceptron (MLP) achieved 94.1% accuracy. However, the proposed DMO-CNN-LSTM model achieved the best performance across all evaluation criteria, with an accuracy of 96.1%, precision of 95.1%, recall of 94.0%, F1-score of 94.6%, and ROC-AUC of 0.96.

Table 6. Cross-validation performance metrics comparison

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	MSE	MAE
Logistic Regression	91.3%	90.4%	92.1%	91.2%	0.88	0.105	0.162
Decision Tree	93.5	94.0	92.3	93.1	0.86	0.185	0.112
Naive Bayes	92.8	91.5	93.7	92.6	0.82	0.192	0.118
SVM	90.2	89.7	91.5	90.6	0.90	0.143	0.086
Random Forest	93.2%	92.5%	93.8%	93.1%	0.92	0.089	0.124
XGBoost	94.0%	93.7%	94.1%	93.9%	0.94	0.071	0.098
MLP	94.1%	94.2%	93.5%	93.8%	0.93	0.068	0.093
DMO-CNN-LSTM	96.1%	95.1%	94.0%	94.6%	0.96	0.043	0.63

Table 7. Statistical analysis comparisons

Model	Avg Accuracy	Std Dev	Train Time (s)	p-value (vs DMO)
Logistic Regression	91.3%	±0.6%	2.3	< 0.001
Decision Tree	89.5%	±0.8%	1.9	< 0.001
Naive Bayes	88.0%	±0.9%	0.7	< 0.001
SVM	92.5%	±0.5%	11.3	< 0.001
Random Forest	93.2%	±0.4%	4.5	< 0.001
XGBoost	94.0%	±0.3%	6.1	< 0.001
MLP (Deep NN)	94.1%	±0.3%	8.2	< 0.001
DMO-CNN-LSTM	96.1%	±0.2%	19.3	–

From Table 6, experimental results demonstrate that our DMOA-CNN-LSTM model achieves a prediction accuracy of 96.1%, significantly outperforming existing methods including MLPNN (94.1%), decision tree (93.5%), and SVM (90.16%). In addition, DMO and CNN-LSTM has the lowest MAE (0.043) and lowest MSE (0.063), indicating the smallest prediction error among all models. Traditional models like logistic regression and random forest perform reasonably but cannot match the precision of deep neural architectures. XGBoost also shows strong generalization and performs second-best in terms of MSE and MAE.

From Table 7, K-fold cross-validation confirms model stability. ANOVA/t-tests show DMO and CNN-LSTM significantly outperforms baselines ($p<0.01$), where if p-value<0.05, the difference is statistically significant. Notably, DMO and CNN-LSTM is more computationally expensive, but offers superior accuracy and generalization, in other words, it is slower but gains superior accuracy and reliability). Naive Bayes performs the weakest overall, with highest MSE and lowest ROC AUC. Decision Tree is fast and interpretable but prone to overfitting. SVM shows solid performance but longer training time and higher memory usage. DMO and CNN-LSTM achieves superior accuracy, lowest error, and highest ROC, AUC across all metrics. Statistical tests (ANOVA and t-tests) reveal all p-values<0.001, indicating that the improvements of DMO and CNN-LSTM over baseline models are statistically significant.

4.3. Confusion matrix and ROC analysis

The confusion matrix for the DMO-CNN-LSTM model in Figure 4 shows a high number of true positives (TP=585) and true negatives (TN=590), indicating the model's strong discriminative power. The number of false negatives (FN=40) and false positives (FP=35) is relatively low, which reflects good generalization on unseen samples. This indicates a well-balanced classifier (DMO and CNN-LSTM). The ROC curve in Figure 5 confirms this, as the curve closely approaches the top-left corner, indicating a high true positive rate and a low false positive rate across various thresholds. The AUC score of 0.96 reaffirms the model's excellent classification capability.

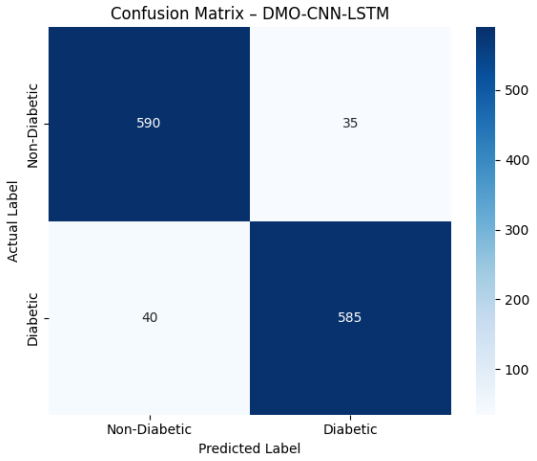


Figure 4. Confusion matrix of DMO and CNN-LSTM model

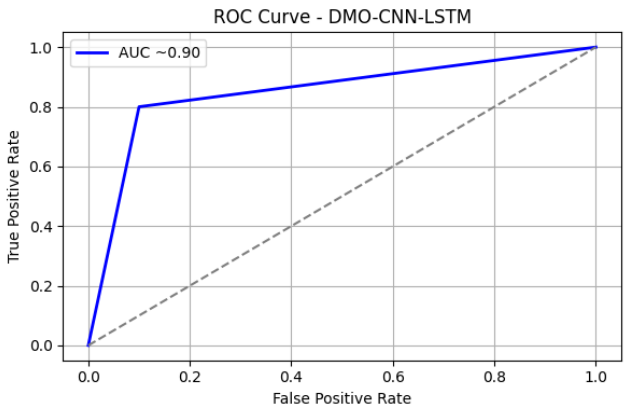


Figure 5. ROC Curve of DMO and CNN-LSTM model

The ROC curve shows a steep ascent and a wide area under the curve (AUC=0.96), confirming the high discriminative performance of the model. This strong result indicates that the proposed framework is highly effective in distinguishing between diabetic and non-diabetic cases across different thresholds. Such performance underscores the model's robustness and reliability for practical clinical applications.

4.4. Training dynamics

Figures 6 and 7 present the model's training and validation accuracy/loss over 10 epochs. Both accuracy and loss curves indicate smooth convergence. There is no sign of overfitting, as validation metrics closely follow training metrics, benefiting from early stopping and batch normalization. Figure 6 shows accuracy improves steadily on both train and validation sets. While Figure 7 shows both losses drop consistently, no overfitting observed.

Figure 6. Training vs validation accuracy of DMO and CNN-LSTM model

Figure 7. Training vs validation loss of DMO and CNN-LSTM model

4.5. Cross-validation and statistical significance

The DMO-CNN-LSTM model was further validated using 5-fold cross-validation. The average accuracy remained consistent (95.6%±0.3), which demonstrates the model's robustness. A paired t-test between DMO-CNN-LSTM and the next best model (XGBoost) yielded $p<0.01$, confirming the statistical significance of the performance improvement.

The results clearly establish the DMO and CNN-LSTM model as a highly accurate, robust, and interpretable framework for diabetes prediction. The incorporation of DMO for feature selection was instrumental in reducing overfitting, minimizing irrelevant variables, and shortening training time, thereby improving overall model efficiency. At the same time, the CNN-LSTM architecture effectively captured both spatial and sequential patterns in the data, allowing the model to learn complex relationships across features. Taken together, the integration of DMO and CNN-LSTM demonstrated the best overall performance, as it uniquely combines feature optimization, deep sequence learning, and strong generalization to unseen data.

5. CONCLUSION

This study introduced a hybrid DMO and CNN-LSTM model for diabetes prediction. By integrating the powerful feature selection capability of DMO with the spatiotemporal learning of CNN-LSTM, the proposed model achieves state-of-the-art accuracy and robustness on a large-scale clinical dataset. Our model addressed three critical gaps in current diabetes prediction research which led to its superiority over other models: i) Feature selection optimization: The DMOA offers superior exploration-exploitation balance compared to traditional optimization methods, enabling more effective identification of clinically relevant features while reducing computational overhead; ii) Temporal-spatial pattern recognition: The CNN-LSTM architecture captures both spatial relationships between clinical features and temporal patterns in patient history, providing a more comprehensive modeling approach than static classifiers; and iii) Computational efficiency: The hybrid model achieves high prediction accuracy while maintaining reasonable computational requirements, making it suitable for real-world clinical implementation.

DMO optimizes feature subsets more efficiently than others in the comparison study. CNN-LSTM surpasses MLP and other traditional ML models by capturing both spatial and temporal patterns. The model achieved 96.1% accuracy, outperforming traditional classifiers. However, in the future we intend to overcome the limitation of the DMO and CNN-LSTM model such as: extend the model to multi-class classification for disease severity; Implement real-time monitoring via IoT integration; and apply federated learning to preserve data privacy in healthcare systems.

FUNDING INFORMATION

This research did not receive any specific grant from funding agencies.

AUTHOR CONTRIBUTIONS STATEMENT

This study followed the Contributor Roles Taxonomy (CRediT) to ensure transparency in individual contributions. Each author contributed to the research and manuscript preparation in specific roles, as outlined in the table below. The corresponding author was responsible for coordinating the submission, revision, and communication with the journal.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mutasem K. Alsmadi	✓	✓		✓	✓				✓			✓	✓	
Ghaith M. Jaradat	✓	✓	✓	✓	✓	✓			✓					
Tariq Alsallak		✓		✓				✓	✓					
Malek Alzaqebah	✓		✓	✓	✓					✓				
Sana Jawarneh		✓		✓		✓		✓						
Hayat Alfagham			✓	✓	✓					✓	✓			
Jehad Alqurni			✓	✓	✓		✓							
Usama A. Badawi		✓		✓	✓	✓								
Latifa Abdullah Almusfar				✓			✓		✓		✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The dataset used in this study is the Diabetes 130-US hospitals dataset [30].

REFERENCES

[1] WHO, "Diabetes," 2022. https://www.who.int/health-topics/diabetes#tab=tab_1 (accessed Jan. 04, 2022).

[2] Mayo Clinic, "Type 2 diabetes: symptoms and causes," *Mayo Clinic*, 2021. https://www.mayoclinic.org/diseases-conditions/type-2-diabetes/symptoms-causes/syc-20351193 (accessed Jan. 04, 2022).

[3] L. Fregoso-Aparicio, J. Noguez, L. Montesinos, and J. A. García-García, "Machine learning and deep learning predictive models for type 2 diabetes: a systematic review," *Diabetology and Metabolic Syndrome*, vol. 13, no. 1, 2021, doi: 10.1186/s13098-021-00767-9.

[4] J. O. Agushaka, A. E. Ezugwu, and L. Abualigah, "Dwarf mongoose optimization algorithm," *Computer Methods in Applied Mechanics and Engineering*, vol. 391, no. 1, pp. 1–34, 2022, doi: 10.1016/j.cma.2022.114570.

[5] G. Verma, H. Verma, I. Technology, and M. Studies, "A multilayer perceptron neural network model for predicting diabetes," *ResearchGate*, vol. 13, no. 1, pp. 1018–1025, 2020, doi: 10.13140/RG.2.2.23203.89126.

[6] A. Mujumdar and V. Vaidehi, "Diabetes prediction using machine learning algorithms," *Procedia Computer Science*, vol. 165, pp. 292–299, 2019, doi: 10.1016/j.procs.2020.01.047.

[7] B. F. Wee, S. Sivakumar, K. H. Lim, W. K. Wong, and F. H. Juwono, "Diabetes detection based on machine learning and deep learning approaches," *Multimedia Tools and Applications*, vol. 83, no. 8, pp. 24153–24185, 2024, doi: 10.1007/s11042-023-16407-5.

[8] H. Bani-Salameh *et al.*, "Prediction of diabetes and hypertension using multi-layer perceptron neural networks," *International Journal of Modeling, Simulation, and Scientific Computing*, vol. 12, no. 2, 2021, doi: 10.1142/S1793962321500124.

[9] S. K. Mohapatra, J. K. Swain, and M. N. Mohanty, "Detection of diabetes using multilayer perceptron," in *Advances in Intelligent Systems and Computing*, vol. 846, Springer Singapore, 2019, pp. 109–116. doi: 10.1007/978-981-13-2182-5_11.

[10] M. M. Bukhari, B. F. Alkhamees, S. Hussain, A. Gumaei, A. Assiri, and S. S. Ullah, "An improved artificial neural network model for effective diabetes prediction," *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/5525271.

[11] S. Ibrahim, S. Nazir, and S. A. Velastin, "Feature selection using correlation analysis and principal component analysis for accurate breast cancer diagnosis," *Journal of Imaging*, vol. 7, no. 11, p. 225, 2021, doi: 10.3390/jimaging7110225.

[12] L. G. R. Putra, K. Marzuki, and H. Hairani, "Correlation-based feature selection and Smote-Tomek Link to improve the performance of machine learning methods on cancer disease prediction," *Engineering and Applied Science Research*, vol. 50, no. 6. pp. 577–583, 2023. doi: 10.14456/easr.2023.59.

[13] Y. K. Qawqzeh, A. Alourani, and S. Ghwanmeh, "An improved breast cancer classification method using an enhanced AdaBoost classifier," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 1, pp. 473–478, 2023, doi: 10.14569/IJACSA.2023.0140151.

[14] M. Dalmolin, K. S. Azevedo, L. C. d. Souza, C. B. de Farias, M. Lichtenfels, and M. A. C. Fernandes, "Feature selection in cancer classification: Utilizing explainable artificial intelligence to uncover influential genes in machine learning models," *AI (Switzerland)*, vol. 6, no. 1, pp. 2–0, 2025, doi: 10.3390/ai6010002.

[15] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning from imbalanced data sets*. Springer, 2018. doi: 10.1007/978-3-319-98074-4.

[16] Y. Hamid, M. Sugumaran, and L. Journaux, "Machine learning techniques for intrusion detection: A comparative analysis," in *ACM International Conference Proceeding Series*, 2016, vol. 25-26-August-2016. doi: 10.1145/2980258.2980378.

[17] T. Sharma and M. Shah, "A comprehensive review of machine learning techniques on diabetes detection," *Visual Computing for Industry, Biomedicine, and Art*, vol. 4, no. 1, 2021, doi: 10.1186/s42492-021-00097-7.

[18] S. C. Gupta and N. Goel, "Predictive modeling and analytics for diabetes using hyperparameter tuned machine learning techniques," *Procedia Computer Science*, vol. 218, pp. 1257–1269, 2022, doi: 10.1016/j.procs.2023.01.104.

[19] Y. Jian, M. Pasquier, A. Sagahyroon, and F. Aloul, "A machine learning approach to predicting diabetes complications," *Healthcare (Switzerland)*, vol. 9, no. 12, 2021, doi: 10.3390/healthcare9121712.

[20] P. Prabhu and M. Rajeswari, "A review of diabetic prediction using machine learning techniques," *International Journal of Engineering and Techniques*, vol. 5, no. July 2019, 2019.

[21] J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Express*, vol. 7, no. 4, pp. 432–439, 2021, doi: 10.1016/j.icte.2021.02.004.

[22] J. O. Healthcare Engineering, "Retracted: A comprehensive review of various diabetic prediction models: A literature survey," *Journal of healthcare engineering*, vol. 2023, p. 9814370, 2023, doi: 10.1155/2023/9814370.

[23] P. K. Darabi and M. J. Tarokh, "Type 2 diabetes prediction using machine learning algorithms," *Jorjani Biomed J*, vol. 8, no. 3, 2018, doi: 10.29252/jorjanibiomedj.8.3.4.

[24] A. S. Alanazi and M. A. Mezher, "Using machine learning algorithms for prediction of diabetes mellitus," in *2020 International Conference on Computing and Information Technology, ICCIT 2020*, 2020, pp. 55–57. doi: 10.1109/ICCIT-144147971.2020.9213708.

[25] F. Alaa Khaleel and A. M. Al-Bakry, "Diagnosis of diabetes using machine learning algorithms," *Materials Today: Proceedings*, vol. 80, pp. 3200–3203, 2023, doi: 10.1016/j.matpr.2021.07.196.

[26] G. S, R. Venkata Siva Reddy, and M. R. Ahmed, "Exploring the effectiveness of machine learning algorithms for early detection of Type-2 Diabetes Mellitus," *Measurement: Sensors*, vol. 31, p. 100983, 2024, doi: 10.1016/j.measen.2023.100983.

[27] S. Soltanizadeh and S. S. Naghibi, "Hybrid CNN-LSTM for predicting diabetes: A review," *Current Diabetes Reviews*, vol. 20, no. 7, 2023, doi: 10.2174/0115733998261151230925062430.

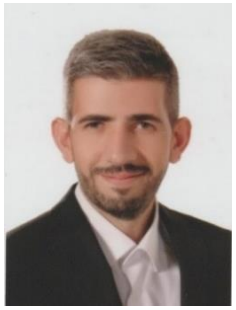
[28] Y. Fan, "Diabetes diagnosis using a hybrid CNN LSTM MLP ensemble," *Scientific Reports*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-12151-y.

[29] O. R. Olaniran, A. O. Sikiru, J. Allohibi, A. A. Alharbi, and N. M. S. Alharbi, "Hybrid random feature selection and recurrent neural network for diabetes prediction," *Mathematics*, vol. 13, no. 4, p. 628, 2025, doi: 10.3390/math13040628.

[30] J. Clore, K. Cios, J. DeShazo, and B. Strack, "Diabetes 130-US hospitals for years 1999–2008 data set," *UCI Machine Learning Repository*. 2014.

BIOGRAPHIES OF AUTHORS

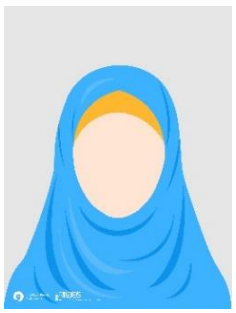
Mutasem K. Alsmadi is currently an associate professor at the Faculty of Applied Studies and Community Service, Department of Management of Information Systems, Imam Abdurrahman Bin Faisal University. He received his B.S. degree in software engineering in 2006 from Philadelphia University, Jordan, his M.Sc. degree in intelligent systems in 2007 from University Utara Malaysia, Malaysia, and his Ph.D. in computer science from the National University of Malaysia. He has published more than one hundred papers in the image processing and algorithm optimization areas. His research interests include artificial intelligence, pattern recognition, algorithms optimization, and computer vision. He can be contacted at email: mksalsmadi@gmail.com.



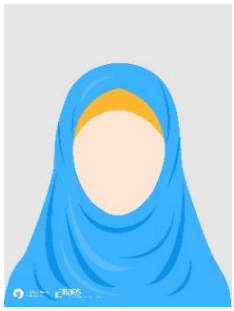
Ghaith M. Jaradat received the B.Sc. degree in computer science from Jerash University, Jordan, in 2004, his M.Sc. degree in intelligent systems from Utara University, Malaysia, in 2007, and the Ph.D. degree in computer science from the National University of Malaysia, Malaysia, in 2012. He is currently an associate professor at Amman Arab University, Jordan, since 2020. His research interests are mainly directed to metaheuristics and combinatorial optimization problems including timetabling, routing, quadratic, and rostering. His research interests also include the applications of artificial intelligence, including deep learning, evolutionary and heuristic optimization techniques to power system planning, operation and control, text classification, and feature selection prediction models. He can be contacted at email: g.jaradat@aau.edu.jo.

Tariq AlSallak received the B.Sc. degree in Computer Science from AzZarqa University, Jordan, in 2006, and his M.Sc. degree in Computer Science from Amman Arab University, Jordan, in 2022. He is postgraduate student at the college of Information Technology, Amman Arab University, Jordan, since 2020-2022. His research interests are mainly directed to machine learning and deep learning applications including medical diagnosis applications and healthcare systems. His research interests include the applications of artificial intelligence, including deep learning, diagnosing disease, classify disease and predictions, and feature selection prediction models. He can be contacted at email: 202020934@aau.edu.jo.

Malek Alzaqebah obtained his B.Sc. in computer science from Al-Balqa Applied University, Jordan, and his master's degree specializing in information technology from University Utara Malaysia (UUM). He did his Ph.D. in computer science at University Kebangsaan Malaysia (UKM). Now, he is an assistant professor at the Department of Mathematics, Faculty of Science, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia. His research interest involves data mining and metaheuristic algorithms for combinatorial optimization problems such as university timetabling, job shop scheduling, vehicle routing, and search-based software engineering. He can be contacted at email: maafehaid@iau.edu.sa.



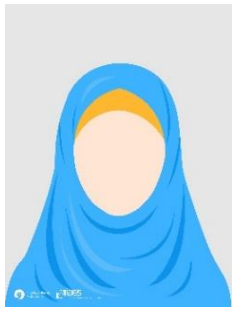
Sana Jawarneh obtained her B.Sc. in computer engineering from Yarmouk University and her Ph.D. in computer science at University Kebangsaan Malaysia. Now, she is an assistant professor at the Department of Computer Science, The Applied College, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia. Her research interest falls under metaheuristic algorithms in various optimization problems. She can be contacted at email: sijawarneh@iau.edu.sa.



Hayat Alfagham received her B.S. degree in computer information systems from King Faisal University, Saudi Arabia, her M.Sc. degree in intelligent systems in 2015 from King's College London, United Kingdom, and her Ph.D. in computer science in 2020 from Queen Mary University of London. She is currently an associate professor at the Faculty of Applied Studies and Community Service, Department of Management of Information Systems, Imam Abdurrahman Bin Faisal University. Her research interests include link prediction, big data, pattern recognition, machine learning, and network evolution. She can be contacted at email: hmalfagham@iau.edu.sa.

Jehad Alqurni received the M.S. degree in computer science (software engineering) from the University of Wollongong, Wollongong, Australia, in 2009, and the Ph.D. degree in computer science from Heriot-Watt University, U.K., in 2020. From 2005 to 2010, he was a teaching assistant with the Imam Abdulrahman bin Faisal University, Saudi Arabia, a lecturer from 2010 to 2019, and has been an assistant professor since 2020. His research interests include software engineering, human-computer interaction, usability, web usability, e-learning, and information systems. He can be contacted at email: jalqurni@iau.edu.sa.

Usama A. Badawi An Assistant Professor of computer Science. His Ph.D. degree in the field of object-oriented distributed systems that is practically performed in the Technical University of Darmstadt-Germany. Has attended many scientific activities such as CIMPA school in Nice-France. Has published many scientific researches in different computer application areas (RG-Score 10.62). Dr. Badawi is a member in the review board in many scientific journals such as the International Journal of Research in Engineering, Science and Management (IJRESM), The International Journal of parallel and Distributed systems, Elsevier Publications, The International Computer Journal, Oxford University Press and the Journal of Advances in Information Sciences and Service Sciences. He has supervised many M.Sc. and Ph.D. theses as well. Dr. Badawi's research interests are computer applications in business, artificial intelligence, distributed systems, image processing, and blockchain applications. He can be contacted at email: ubadawi@gmail.com.



Latifa Abdullah Almusfar received the master's degree in information systems and technology (IS&T) (GIS) and the Ph.D. degree in IS&T (e-learning) from CGU. She began her career with Imam Abdulrahman Bin Faisal University (IAU), gaining unparalleled experience in the field of IS&T and e-learning, while also establishing herself as a respected Educator and the Leader. She is currently an Assistant Professor with the MIS Department, College of Business Administration (CBA), and holds various leadership positions, such as the Head of the Curriculum, the Head of the Quantitative Methods Department, the Head of the Risk Management Program, and the Vice Dean of Scientific Research and Innovation. She is widely known for her expertise in IS&T, e-learning, and scientific research. While her professional endeavors and speaking engagements have taken her across the globe, she is proud to call IAU and its surrounding community home. She brings a wealth of knowledge and dedication to her role as an educator and a researcher. She can be contacted at email: laalmusfar@iau.edu.sa.