Journal of Theoretical and Applied Information Technology
31st March 2026. Vol.104. No.6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SCARLET MACAW-INSPIRED DEEP BELIEF NETWORK FOR EARLY, ACCURATE, AND INTERPRETABLE PREDICTION OF GESTATIONAL DIABETES MELLITUS

D.SHOBANA[1] , V VINODHINI[2]

[1]Research Scholar, Department of Computer Science, KPR College of Arts Science and Research, Coimbatore, India

[2]Associate Professor, Department of Information Technology, KPR College of Arts Science and Research, Coimbatore, India

: [1] shobanadl@gmail.com, [2]vvprof133@gmail.com

ABSTRACT

Gestational Diabetes Mellitus (GDM) is a prevalent pregnancy-related condition that contributes significantly to maternal and neonatal health complications, such as pre-eclampsia, neonatal hypoglycemia, and long-term metabolic disorders. Existing diagnostic methods, such as the Oral Glucose Tolerance Test (OGTT), are invasive, time-consuming, and not easily scalable for early detection. This research introduces the Scarlet Macaw-Inspired Deep Belief Network (SM-DBN), a deep learning-based model designed to predict GDM risk early in pregnancy. By incorporating bio-inspired optimization techniques such as adaptive foraging and territorial behavior, SM-DBN addresses challenges like class imbalance, missing data, and the dynamic nature of pregnancy-related risks. The model achieved a classification accuracy of 81.277%, showing significant promise in early-stage GDM detection. SM-DBN integrates temporal pattern learning to capture trimester-specific risk variations, and explainable decision support ensures transparency for clinicians. The model's ability to handle incomplete and noisy clinical data, adapt dynamically to evolving risks, and offer interpretable predictions makes it a highly effective tool. This approach provides a scalable, interpretable, and non-invasive solution for GDM risk assessment, reducing the reliance on traditional diagnostic tests and enhancing maternal-fetal health outcomes across diverse healthcare settings.

Keywords: *Healthcare, Gestational Diabetes Mellitus, Prediction, Bio-Inspired Optimization, Deep Belief Network, Diabetes*.

1. INTRODUCTION

Diabetes mellitus is a chronic metabolic condition characterized by impaired glucose regulation. Among its various subtypes, Gestational Diabetes Mellitus (GDM) is a condition that arises specifically during pregnancy, primarily due to placental hormones that interfere with insulin sensitivity [1]. While GDM typically resolves after delivery, women who experience GDM are at a significantly higher risk of developing type 2 diabetes later in life. The global prevalence of GDM is rising due to factors such as increased maternal age, obesity, and sedentary lifestyles, particularly in developing countries [2]. Current diagnostic methods, such as the Oral Glucose Tolerance Test (OGTT), are effective but invasive, resource-intensive, and not scalable for large populations, leading to significant delays in diagnosis. The complexity of GDM arises from its variable manifestation and the fact that many women who develop it do not exhibit clear pre-existing risk factors. As a result, early prediction and intervention remain critical to reducing maternal and neonatal complications associated with GDM [3].

GDM poses substantial clinical and societal burdens for affected individuals and healthcare systems. Women with undiagnosed or poorly managed GDM face increased risks of pregnancy complications, including pre-eclampsia, cesarean delivery, and neonatal hypoglycemia. Long-term, they are also at a higher risk of developing type 2 diabetes and cardiovascular diseases [4], [5]. Children born to mothers with GDM are more likely to face metabolic challenges such as obesity and diabetes later in life. The societal impact is far-reaching, with GDM contributing to rising healthcare costs, particularly in regions with limited access to specialized care [6]. The inconsistency in screening guidelines across different health systems leads to missed cases or unnecessary testing, creating inefficiencies. These challenges underscore the need for early, accurate, and non-invasive prediction models that can help reduce GDM-related complications, improve maternal-fetal health, and alleviate the healthcare burden [7].

Despite promising results, previous machine learning (ML) and deep learning (DL) models used for GDM prediction have significant

limitations [8]. Most studies have utilized retrospective datasets, which lack temporal progression data and fail to capture dynamic risk changes across pregnancy trimesters. Many models are trained on imbalanced datasets, with GDM-positive cases often underrepresented, leading to biased predictions [9]. The inability to effectively handle incomplete or noisy data further limits the performance of these models. Many existing approaches lack interpretability, making it difficult for clinicians to trust or apply them in practice. Without clear reasoning for predictions, healthcare professionals cannot rely on these systems for timely intervention [10], [11]. These models often fail to generalize well across different population groups, limiting their scalability. These challenges highlight the need for robust, interpretable, and adaptive models capable of capturing the complex, evolving nature of GDM risk throughout pregnancy [12].

The increasing complexity of GDM prediction calls for novel approaches that can overcome the shortcomings of existing models. Bio-inspired optimization techniques, which are based on natural behaviors and adaptive strategies observed in the animal kingdom, present a promising solution [13]. These techniques offer significant advantages over traditional optimization methods, as they allow for efficient exploration of high-dimensional, non-linear data spaces, dynamically adapting to data irregularities and class imbalances [14], [15]. In the context of GDM prediction, bio-inspired algorithms—such as those inspired by the Scarlet Macaw's adaptive foraging, territorial behavior, and cooperative vigilance—can improve feature selection, regularize weights, and enhance data propagation. These biologically inspired principles facilitate hierarchical, time-aware learning that adjusts based on evolving risk factors, offering a more accurate and interpretable solution for GDM prediction. By incorporating bio-inspired mechanisms into deep learning architectures, this research aims to provide a robust, interpretable, and clinically applicable model that can predict GDM risk earlier in pregnancy and help minimize unnecessary diagnostic testing. The study aspires to enhance predictive performance by ensuring greater transparency and accessibility in clinical decision-making.

1.1. Scope and Assumptions

This study focuses exclusively on first-trimester, non-imaging, structured clinical variables for early GDM risk prediction using a bio-inspired deep learning framework. The proposed SM-DBN assumes availability of standardized tabular maternal records containing metabolic and demographic indicators. The model does not incorporate imaging modalities, genomic sequencing data, or multi-center federated datasets. It operates under the assumption that clinical variables are recorded under standardized obstetric protocols and that probabilistic imputation adequately reconstructs missing entries without introducing systematic bias. The architecture is optimized for binary GDM classification and does not extend to multi-complication pregnancy modeling.

This investigation is delimited to binary GDM risk stratification using structured maternal health records collected under controlled clinical settings. The model does not evaluate longitudinal postpartum diabetic transition nor incorporate wearable sensor streams or genetic biomarkers. Performance evaluation is conducted on a single consolidated dataset, limiting external multi-regional generalization claims. Model interpretability relies on SHAP, LIME, and entropy-based uncertainty quantification, which remain post-hoc approximations of deep decision boundaries. Future validation across heterogeneous clinical systems is required to strengthen cross-population robustness.

1.2. Problem Statement

GDM presents unresolved challenges in early risk identification, diagnostic standardization, and clinical intervention. Despite OGTT being the recommended diagnostic tool, it is resource-intensive, time-consuming, and clinically burdensome, especially for patients not in the actual risk group. Variations in screening guidelines across international and national health systems have led to inconsistent identification practices, resulting in unnecessary testing or delayed detection. Current risk factors, including age, BMI, family history, and previous macrosomia, lack precise threshold definitions and fail to capture the temporal evolution of GDM risk during pregnancy. Real-world datasets for GDM are typically imbalanced, with fewer positive cases, causing existing models to exhibit low sensitivity and poor minority class prediction. Most prior studies have relied on retrospective data, limiting their reliability and neglecting the potential of time-aware, real-time, and explainable frameworks. The interpretability of machine learning outputs is critical in clinical settings, yet remains largely unaddressed. Clinical records also suffer from missing values, inconsistent formats, and unstructured fields, complicating integration into

intelligent decision support systems. The absence of robust, prospective, interpretable, and temporally adaptive predictive solutions hinders targeted screening, data-efficient training, and timely intervention, compromising optimal maternal-fetal outcomes in GDM care

1.3. Motivation

GDM presents a growing public health burden, disproportionately affecting maternal and neonatal outcomes, particularly in underserved and socioeconomically vulnerable populations. Delayed or missed diagnosis leads to complications including pre-eclampsia, neonatal hypoglycemia, and increased cesarean rates, while overdiagnosis results in unnecessary emotional stress and clinical interventions. Traditional screening methods like OGTT are invasive, time-bound, and not easily scalable, especially in community or low-resource settings. As healthcare systems globally shift toward digital integration, there is a pressing societal need for AI-driven, interpretable, and real-time screening frameworks that support early intervention and reduce long-term diabetes progression. The clinical data irregularities, such as missing values and inconsistent formats, hinder accurate risk prediction. The dynamic nature of GDM risk throughout pregnancy underscores the importance of time-sensitive and personalized prediction. Ensuring transparency and interpretability in such systems is essential for clinician trust and patient-centered care. This research is motivated by the urgent need to address these gaps through accessible, equitable, and explainable decision support tools that advance maternal health outcomes across diverse populations.

1.4. Objective

This research aims to design and validate a novel Scarlet Macaw-Inspired Deep Belief Network (SM-DBN) for accurate, explainable, and temporally adaptive prediction of GDM using non-imaging, first-trimester screening datasets. The proposed model addresses critical gaps associated with retrospective model limitations, diagnostic inconsistencies, data imbalance, and the inefficiencies of universal OGTT-based screening protocols. SM-DBN integrates structured layer-wise training, probabilistic feature selection, temporal risk modeling, and uncertainty-aware classification to capture trimester-evolving patterns and incomplete records without compromising predictive stability. The architecture employs weight regularization, adaptive learning rate tuning, and gradient-directed propagation to maintain generalization in real-world, noisy, and class-imbalanced datasets. Special emphasis has been placed on model interpretability through SHAP and attention-based decision support, enabling clinician-trusted risk inference. Prospective validation has been carried out to ensure real-time applicability and generalizability across diverse maternal populations. A key performance goal of SM-DBN is maximizing classification accuracy while minimizing overdiagnosis, unnecessary OGTT usage, and undetected risk cases, supporting early clinical intervention, personalized care, and equitable maternal health delivery.

1.5. Study Aim and Outcome Measures

The primary aim of this study is to design a temporally adaptive, bio-inspired Deep Belief Network capable of early-stage GDM prediction with structured interpretability. Novelty is established through the integration of Scarlet Macaw-inspired optimization behaviors into hierarchical RBM stacking combined with entropy-ranked feature prioritization and trimester-aware temporal learning. Outcome measures include Classification Accuracy, MCC, Error Rate, Youden's Index, and Critical Success Index, enabling balanced evaluation across sensitivity, specificity, and class stability metrics. Unlike prior reconstruction-centric or fixed ensemble models, SM-DBN emphasizes discriminative depth, uncertainty calibration, and explainable inference.

2. LITERATURE REVIEW

"Graph-Semantic Repurpose Net" [16] combines chemical graphs and biomedical text embeddings to predict new drug-disease links. Molecules and diseases are embedded via graph attention and text transformers, respectively. Their representations are aligned in a shared latent space for link prediction. The network scores candidate pairs based on semantic and structural compatibility. "Engineered Multi-Disease Grid" [17] uses engineered clinical features to classify patients into diabetic-only, hypertensive-only, combined, or regular classes. Interaction features, such as glucose-pressure ratios, are created to model physiological overlap. Redundancies are filtered out, and classifiers like SVM are evaluated. A multi-class structure captures distinct and co-morbid disease states. "App-Driven Hypo Forecast" [18] relies on structured user input from mobile apps to create next-day hypoglycemia alerts. Data like blood glucose trends and systolic/diastolic pressures are compiled into daily time windows across ten-day spans. Machine learning models—especially

Random Forest and SVM—are trained on these temporal records using feature engineering that captures fluctuation rates and averages.

"Biomarker-Fused Forecast Loop" [19] merges clinical history with wearable data to predict gestational diabetes markers before diagnosis. Wristband-derived metrics like glucose and motion signals are fused with lab readings, including triglycerides and HbA1c. This integration is handled using coupled matrix tensor factorization to align modalities. Regressors like Random Forest operate on the fused features to forecast biomarker levels. "Virtual Ward Synth Loop" [20] orchestrates AI monitoring of diabetic and kidney parameters using real-time and historical data. Cloud-based platforms receive inputs like glucose, creatinine, and blood pressure from home devices. Parallel pipelines powered by Gradient Boosting and LSTM-CNN detect patterns and degradation risk. Predictions are fused via a soft-voting ensemble and visualized through CURA's digital interface. "Meal-Metric Predict Loop" [21] integrates macronutrient intake—carbs, fat, protein, fiber—and insulin levels to predict post-meal glucose. A feedforward neural network is trained on structured input vectors where each nutrient is time-aligned with insulin onset. The model forecasts glucose over 15 to 60-minute intervals, capturing how non-carb components influence absorption delays. Layer-wise interpretation highlights protein and fat.

"Grammar-Guided Glycemia Learner" [22] uses symbolic evolution to derive interpretable glucose dynamics post-meal. Based on clustered glucose trajectories, structured grammars evolve difference equations explaining how glucose shifts every 15 minutes. Each grammar produces formulas with constraints ensuring clinical realism. Parkes error grid scores assess safety, while equation complexity is minimized. "White-Black Predictive Duel" [23] contrasts physiologically grounded white-box models with data-driven black-box models using identical input timelines. The white-box system models internal glucose-insulin dynamics and updates via a particle filter. Black-box systems—like LSTM and TCN—learn sequence patterns from CGM, meals, and insulin records. "Forecast-Test Mismatch Trap" [24] uncovers a critical discrepancy between standard model evaluation and closed-loop system behavior. The study contrasts a rule-based controller and an LSTM predictor within simulated insulin control environments. While LSTM shows favorable static prediction metrics, it destabilizes performance during real-time glucose regulation, revealing that isolated historical error minimization fails to reflect control safety.

"Tandem Risk Encoder" [25] transforms irregular longitudinal EHR data into structured, temporally ordered sequences using entity embeddings and synchronized LSTM streams. Distinct health domains—metabolic markers, activity patterns, and comorbidities—are encoded separately and integrated through tandem architecture to preserve domain semantics while capturing inter-domain influence. "Tube-Guarded MPC" [26] constructs a dual-tier glucose control loop incorporating an LSTM predictor optimized via genetic algorithms with a non-linear tube-based MPC. The LSTM model forecasts glucose trajectory, while the Tube-NMPC enforces constraints by forming a bounded 'tube' around predicted paths, adjusting insulin inputs accordingly. "Preprocess-Classify Tuning Grid" [27] evaluates twelve machine learning classifiers under varying preprocessing conditions to assess diabetes detectability from tabular datasets. Each input scenario—raw, normalized, min-max scaled—is passed through models like SVM, Logistic Regression, and ensemble learners. Performance is mapped across multiple metrics, revealing that preprocessing effects are model-dependent.

"Tri-Vote Impute Model" [28] orchestrates a two-phase strategy: initial recovery of missing values via KNN-based local averaging, followed by risk classification using a three-model ensemble of Random Forest, Gradient Boosting, and Extra Trees. Each model contributes independent probabilistic estimates, fused through soft voting to mitigate individual model variance. "Azure-Fused Risk Lens" [29] formulates a cloud-deployed ensemble strategy combining LightGBM and KNN with joint hyperparameter tuning via grid search. KNN supplements LightGBM's decision-tree logic by addressing non-linear boundary cases through distance metrics. The ensemble is fine-tuned using ten-fold cross-validation and deployed through Azure ML for scalable inference. "IL-6 Activity Loop" [30] models repeated exercise-induced IL-6 release and its long-range effect on glucose balance by simulating beta-cell preservation mechanisms. Daily physical activity patterns are encoded, and IL-6's anti-inflammatory influence is represented in endocrine system modules, showing improved insulin secretion and resistance regulation.

"Adaptive Harmony-Driven Hybrid Stacking (AHDHS)" [31] applies an intelligent ensemble approach for early-stage GDM classification where screening data are often sparse or noisy. Standard stacking pipelines do not adapt well to trimester-specific variations. AHDHS uses harmony search to select meaningful features like age, early fasting glucose, and pedigree function, then curates a learner set that complements them. This co-optimized design enhances reliability under screening constraints and provides a versatile diagnostic model that outperforms fixed ensemble schemes in early gestational prediction scenarios. "Outlier Detection with Deep Stacked Autoencoder (OD-DSAE)" [32] introduces a composite learning pipeline tailored for early-stage gestational diabetes classification. While existing deep learning models focus solely on accuracy, OD-DSAE prioritizes clinical reliability by filtering anomaly-prone inputs through hierarchical clustering. Its deep stacked autoencoder captures nuanced trends in early metabolic markers, often invisible to linear classifiers. By improving data quality and encoding temporal dependencies, OD-DSAE supports proactive screening strategies for expectant mothers, reducing reliance on invasive tests like OGTT in early trimesters.

Bio-inspired optimization has progressively evolved across adaptive routing, intelligent learning, and secure decision systems, establishing a strong foundation for nature-driven computational design. Early frog leap, wolf prey, and cuckoo search inspired routing protocols demonstrated adaptive exploration, convergence stability, and energy-aware communication in dynamic cognitive radio environments [33]–[37]. Subsequent studies expanded these principles into swarm intelligence, fuzzy inference, and constrained optimization for enhanced QoS, delay reduction, and governance in IoT, quantum, and drone networks [42]–[47], [50], [52], [60], [62], [68]–[72]. Bio-inspired frameworks were also extended to healthcare and neural learning, including optimization-enhanced recurrent networks and frog leap–based medical classification models [48], [59], [64], reinforcing clinical applicability. Broader theoretical and interdisciplinary perspectives further consolidated intelligent decision-making through bio-inspired strategies [39], [44], [74], [75], [76]. These cumulative advances motivate the present SM-DBN framework, where Scarlet Macaw-inspired adaptive behaviors are embedded within deep belief learning to strengthen early, interpretable GDM prediction.

2.1. Research Question

Based on the identified gaps in retrospective modeling, class imbalance instability, reconstruction-centric bias, and lack of temporally adaptive interpretability in prior literature, the central research question is: Can a bio-inspired, entropy-ranked, temporally adaptive Deep Belief Network architecture provide clinically reliable, explainable, and balanced early-stage prediction of Gestational Diabetes Mellitus under real-world data irregularities? This study investigates this question through structured architectural design and comparative empirical validation.

3. SCARLET MACAW-INSPIRED DEEP BELIEF NETWORK

The proposed methodology outlines the design and implementation of a Scarlet Macaw-Inspired Deep Belief Network (SM-DBN) for the early prediction of Gestational Diabetes Mellitus (GDM). This approach integrates bio-inspired optimization techniques and deep learning models to address challenges such as imbalanced data, missing values, and temporal dynamics in clinical datasets. The methodology includes feature extraction, hierarchical training, temporal pattern learning, and robust data handling, focusing on interpretability and real-time prediction. The model enhances predictive accuracy by utilizing bio-inspired mechanisms like adaptive foraging and group coordination, ensuring practical applicability in diverse healthcare environments.

3.1. Input Feature Selection (Selective Feeding)

Scarlet Macaw's selective feeding behavior has inspired the feature selection mechanism in SM-DBN, ensuring that only high-value attributes relevant to GDM prediction are retained while eliminating redundant or irrelevant ones. This step has significantly improved the accuracy, robustness, and efficiency of hierarchical learning in deep architectures.

Clinical attributes relevant to GDM prediction, including glucose levels, body mass index, age, and genetic predisposition, have been extracted using a probabilistic ranking mechanism. These attributes, represented in a structured matrix, have been transformed through statistical normalization and weighted distribution to align with the deep hierarchical learning of SM-DBN.

$$X' = \frac{X - \mu}{\sigma} \times w_f \quad (1)$$

where X' represents the normalized feature matrix, X denotes the original input data, μ is the mean of each feature, σ is the standard deviation, and w_f is the computed feature weight is based on statistical significance in the dataset.

Normalization has ensured that feature values remain uniform, preventing the dominance of high-magnitude attributes. Weighting factors have been dynamically assigned based on their influence on model learning, capturing crucial clinical patterns that influence GDM development. Feature redundancy in high-dimensional medical data has been minimized through an entropy-based selection process, eliminating attributes with low variance or high correlation. Entropy values for each feature have been calculated, and only those contributing meaningful variability to the model have been retained, ensuring optimal information gain.

$$H(X) = -\sum_{i=1}^{n} P(x_i) log p(x_i) \quad (2)$$

where $H(X)$ is the entropy of feature X, $p(x_i)$ represents the probability distribution of each attribute, and n denotes the number of possible feature values.

High-entropy features have exhibited greater unpredictability and information diversity, which has enhanced the DBN's learning ability. Low-entropy, redundant attributes have been discarded, optimizing the model's computational efficiency and preventing overfitting. Scarlet Macaws prioritize nutrient-rich food sources for survival, reflecting the multi-layer prioritization of significant features in SM-DBN. Features have been ranked using a probabilistic measure that evaluates their impact across multiple network layers, refining hierarchical feature representation for robust classification.

$$P(F_i) = \frac{\sum_{l=1}^{L} W_{l,i} \cdot A_{l,i}}{L} \quad (3)$$

where $P(F_i)$ represents the priority score of the feature $F_i, W_{l,i}$ is the weight assigned to the feature at layer l, $A_{l,i}$ represents its activation strength, and L denotes the total number of layers.

Higher-priority features have demonstrated stronger activation patterns across multiple layers, reinforcing their relevance in predicting GDM risk. Lower-priority attributes have undergone feature suppression, preventing information redundancy and ensuring the efficient utilization of computational resources in deep learning.

3.2. Greedy Layer-Wise Training (Adaptive Foraging)

Scarlet Macaw's adaptive foraging strategy has influenced the hierarchical training of SM-DBN, where each Restricted Boltzmann Machine (RBM) layer has been trained independently before integration. This technique has enhanced feature abstraction, reduced computational complexity, and ensured meaningful patterns in GDM data have been effectively captured.

Each RBM layer has undergone unsupervised pretraining to learn hierarchical feature representations independently. The objective function has been optimized at every layer before stacking, ensuring that deep features remain distinctive. This strategy has provided a structured flow of information, preventing feature degradation across layers and maintaining optimal gradient propagation. Layer-wise training has prevented gradient vanishing by optimizing local representations before global fine-tuning. This structured learning process has ensured that each layer refines discriminative features essential for GDM prediction.

$$W_l^* = W_l - \eta \left(\frac{\partial E_l}{\partial W_l} \right) \quad (4)$$

where W_l^* represents the optimized weight matrix at layer l, W_l denotes the initial weight matrix, η is the learning rate, and E_l signifies the energy function at the respective layer.

Adaptive foraging in Scarlet Macaws has inspired an iterative weight adjustment mechanism, where probability distributions in each RBM layer have been fine-tuned dynamically. This method has improved feature representation and ensured better convergence across layers, mimicking the macaw's selective energy optimization while foraging. The probability adjustment mechanism has refined the RBM's ability to model complex distributions in GDM data, ensuring that each feature contributes meaningfully to hierarchical abstraction.

$$p(h_j = 1|v) = \sigma \left(\sum_i W_{ij} v_i + b_j \right) \quad (5)$$

where $p(h_j = 1|v)$ denotes the probability of a hidden unit h_j being activated given input v, W_{ij} represents the weight connection between the input v_i and hidden unit h_j, b_j is the bias term, and σ is the activation function.

Feature representations in each RBM have been fine-tuned using a contrastive divergence method, allowing the model to capture relevant probabilistic dependencies. This iterative refinement has reinforced model stability and improved hierarchical depth, ensuring that early-stage learning enhances final classification accuracy. Contrastive divergence has facilitated efficient training by ensuring that weight updates minimize divergence between empirical and learned distributions, reinforcing the hierarchical learning structure of SM-DBN.

$$\Delta W_{ij} = \alpha\left(\langle v_i h_j\rangle_{data} - \langle v_i h_j\rangle_{model}\right) \quad (6)$$

where ΔW_{ij} represents the weight update, α is the learning coefficient, $\langle v_i h_j\rangle_{data}$ signifies the expectation under the data distribution, and $\langle v_i h_j\rangle_{model}$ denotes the expectation under the model's learned distribution.

Energy minimization across layers has ensured stable learning by preventing unnecessary fluctuations in feature representations. Inspired by the macaw's energy-efficient foraging strategies, this mechanism has facilitated smooth convergence, reducing the risk of overfitting and maintaining balanced weight adjustments. The minimization of this energy function has strengthened hierarchical pattern extraction in GDM data, ensuring that critical features are preserved without excessive parameter redundancy.

$$E(v,h) = -\sum_i v_i b_i - \sum_j h_j c_j - \sum_{i,j} v_i h_j W_{ij} \quad (7)$$

where $E(v,h)$ represents the energy function of visible (v) and hidden (h) units, b_i and c_j denote bias terms for visible and hidden units, respectively, and W_{ij} represents the weight matrix connecting them.

3.3. Contrastive Divergence Optimization (Beak Adaptation)

Scarlet Macaw's beak adaptation has inspired Contrastive Divergence (CD) optimization in SM-DBN, ensuring efficient weight refinement through structured divergence minimization. This mechanism has significantly improved feature extraction, reduced computational inefficiency, and enabled rapid convergence of Restricted Boltzmann Machines (RBMs) in GDM prediction.

The hidden unit activation probability has been approximated using a sigmoidal function, allowing the network to model complex feature relationships effectively. This process mimics the macaw's ability to efficiently manipulate its beak for food acquisition, ensuring precise selection and adjustment of weight values during training. The probability function defined in Eq.(8) has ensured that feature learning remains sensitive to critical patterns, preventing the loss of relevant GDM-related information during weight updates.

$$p(h_j|v) = \frac{1}{1 + e^{-(\sum_i W_{ij} v_i + b_j)}} \quad (8)$$

where $p(h_j|v)$ represents the probability of a hidden unit h_j being activated given visible unit v, W_{ij} denotes the weight between the visible unit v_i and hidden unit h_j, and b_j is the bias term associated with h_j.

Gradient-based weight updates have been implemented to minimize divergence between empirical data distributions and the model's learned distributions. The refinement of weight values has aligned with the macaw's iterative force application during feeding, enhancing energy-efficient learning. This weight adaptation mechanism has ensured that relevant features receive stronger connections, improving hierarchical feature abstraction in GDM classification.

$$W_{ij}^{(t+1)} = W_{ij}^{(t)} + \eta\left(v_i h_j^{+} + v_i h_j^{-}\right) \quad (9)$$

where $W_{ij}^{(t+1)}$ denotes the updated weight at time step $t+1$, η is the learning rate, $v_i h_j^{+}$ represents the positive phase expectation under the input data, and $v_i h_j^{-}$ is the negative phase expectation under the reconstructed data.

Bias terms have been dynamically updated to stabilize learning, preventing biased weight propagation that could hinder convergence. This adaptation has reflected the macaw's ability to adjust its beak pressure according to food texture, ensuring optimal processing efficiency. Bias regularization has enhanced the balance between weight updates, preventing the model from overconfidently relying on redundant features while preserving critical GDM-related attributes.

$$b_j^{(t+1)} = b_j^{(t)} + \eta \sum_i (v_i - v_i') \quad (10)$$

where $b_j^{(t+1)}$ represents the updated bias for hidden unit j, and v_i' is the reconstructed visible unit in the negative phase.

Energy function minimization has been incorporated to ensure stable feature representation learning. The network has prevented unstable weight fluctuations by optimizing energy transitions, similar to how a macaw fine-tunes its beak motion for efficient food extraction. Reducing energy function values across training iterations has ensured smoother weight convergence, maintaining structured hierarchical learning.

$$E(v,h) = -\sum_i \sum_j W_{ij} v_i h_j - \sum_i b_i v_i - \sum_j c_j h_j \quad (11)$$

where $E(v,h)$ denotes the energy function of visible (v) and hidden (h) units, b_i and c_j are the bias terms for visible and hidden layers, respectively.

Weight regularization has been introduced to prevent overfitting in hierarchical learning, ensuring that redundant features do not dominate training. This regulation has mirrored the macaw's ability to refine beak movements while feeding, avoiding unnecessary energy expenditure. This weight constraint mechanism has ensured that feature representations remain compact, preventing excessive reliance on specific features while preserving overall predictive accuracy for GDM classification.

$$W_{ij}^* = W_{ij} - \lambda \frac{\partial \|W\|^2}{\partial W_{ij}} \quad (12)$$

where W_{ij}^* represents the optimized weight value, and λ denotes the regularization coefficient.

3.4. Hierarchical Feature Extraction (Seed Dispersal)

Scarlet Macaw's seed dispersal strategy has inspired hierarchical feature extraction in SM-DBN, ensuring the structured propagation of learned features across multiple network layers. This approach has significantly improved information retention, minimized feature redundancy, and optimized deep representations for GDM classification.

Each feature extracted at lower layers has undergone progressive transformation across deeper layers, ensuring a structured knowledge hierarchy. This process has reflected the macaw's dispersal of seeds across varying terrains, fostering diverse vegetation growth, akin to diversified feature representation. This transformation mechanism has strengthened deep feature propagation by ensuring the hierarchical abstraction retains essential information relevant to GDM risk patterns.

$$F_l = \sum_{i=1}^{n} \phi(W_{li} X_i + B_l) \quad (13)$$

where F_l denotes the feature representation at layer l, W_{li} represents the weight matrix connecting the input feature X_i to layer l, and B_l is the bias term.

Critical features in patient data have been preserved by incorporating spatial encoding mechanisms, allowing hierarchical structures to retain local variations. This retention strategy resembles how macaws distribute seeds efficiently across different ecological niches, ensuring robust survival adaptability. The application of spatial encoding has reinforced the model's ability to capture fine-grained patterns, minimizing the loss of crucial clinical attributes related to GDM prediction.

$$S_i = \frac{X_i}{max(|X|)} \quad (14)$$

where S_i represents the spatially encoded feature, and $max(|X|)$ denotes the highest absolute value within the input feature set.

Activations within each hierarchical layer have been selectively adjusted based on their relevance to the prediction task, mirroring the macaw's selective seed dispersal strategy that prioritizes fertile regions for growth. This activation-based refinement has ensured that hierarchical representations remain robust, distinguishing between critical and redundant GDM-related features.

$$A_j = tanh\left(\sum_{i=1}^{m} W_{ij} F_i + B_j\right) \quad (15)$$

where A_j represents the activation of the j-th hidden unit, W_{ij} denotes the connection weight between the input feature F_i and hidden node j, and B_j is the bias term.

Hierarchical features extracted from different layers have been fused using a structured aggregation function, ensuring seamless multi-level information integration. This fusion mechanism has imitated macaws' ability to disperse seeds across diverse ecosystems, fostering balanced vegetation growth. This structured fusion mechanism has reinforced SM-DBN's ability to synthesize complex relationships across different feature hierarchies, enhancing predictive stability.

$$H_k = \frac{1}{L} \sum_{l=1}^{L} \delta(F_{l,k}) \quad (16)$$

where H_k represents the fused hierarchical feature, L denotes the number of contributing layers, and

$\delta(F_{l,k})$ is the activation function applied to feature k at layer l.

An energy conservation principle has been introduced to prevent redundant computations, ensuring that feature representations maintain optimal efficiency without excessive overlap. This is analogous to the macaw's strategy of dispersing seeds efficiently to maximize survival potential. This conservation approach has enabled SM-DBN to retain high-information features while discarding low-impact attributes, improving classification efficiency for GDM detection.

$$E(F) = -\sum_{i} F_i log(F_i) \quad (17)$$

where $E(F)$ denotes the entropy-based measure of feature diversity within hierarchical representations.

Normalization at each hierarchical level has stabilized feature propagation, ensuring that weight adjustments remain structured and preventing abrupt shifts in learned patterns. This strategy has reflected the macaw's ability to control seed dispersal, ensuring even distribution across various regions. Normalization has preserved the structural integrity of hierarchical feature extraction, preventing feature degradation during progressive learning stages.

$$N_l = \frac{F_l - \mu}{\sigma} \quad (18)$$

where N_l represents the normalized feature at layer l, μ is the mean of the feature distribution, and σ denotes the standard deviation.

3.5. Stacking RBMs (Group Foraging)

Scarlet Macaw's group foraging behavior has inspired the stacking mechanism of Restricted Boltzmann Machines (RBMs) in SM-DBN, ensuring structured feature learning across multiple layers. This process has strengthened the network's ability to extract deep patterns, reinforced hierarchical representation, and optimized feature abstraction for effective GDM prediction.

Each RBM has been sequentially stacked to ensure layer-wise feature refinement before integration into the final model. This stacking process has followed a structured optimization approach, reflecting the collective intelligence of the macaws' coordinating group foraging to locate optimal food sources efficiently. Stacking RBMs has ensured structured hierarchical learning, allowing deeper layers to refine patterns extracted from lower layers, and reinforcing multi-scale feature representation.

$$F^{(l+1)} = \sigma(W_l F^{(l)} + B_l) \quad (19)$$

where $F^{(l+1)}$ represents the feature transformation at layer $l+1$, W_l denotes the weight matrix at layer l, $F^{(l)}$ is the feature representation from the previous layer, and B_l is the bias term.

To maintain consistency in feature propagation, activation functions at each layer have adjusted neuron responses dynamically. This adaptation has mirrored the macaw's ability to coordinate foraging movements within a group, ensuring efficient resource distribution. Activation propagation has reinforced structured information flow, ensuring critical GDM-related patterns remain distinguishable across deep hierarchical layers.

$$A^{(l)} = \frac{1}{1 + e^{-(W_l F^{(l)} + B_l)}} \quad (20)$$

where $A^{(l)}$ represents the activation output at layer l, ensuring the propagation of non-linear transformations while preventing saturation of feature distributions.

Weight refinement mechanisms have been implemented to optimize stacked RBMs, preventing instability in feature representation. This structured refinement has imitated the macaw's ability to adjust foraging strategies based on group coordination, ensuring optimal adaptation to food availability. Weight refinement has prevented overfitting, allowing the model to generalize well across varying patient data distributions for GDM classification.

$$W_l^* = W_l - \eta \frac{\partial L}{\partial W_l} \quad (21)$$

where W_l^* represents the updated weight matrix, η is the learning rate, and L denotes the loss function that quantifies deviations in feature abstraction.

To ensure each RBM contributes meaningful representations to the stacked network, contrastive divergence (CD) optimization has been applied iteratively. This iterative adjustment has aligned with the macaw's cooperative foraging behavior, where individuals adjust movement based on shared group intelligence. CD-based optimization has reinforced structured weight convergence, preventing instability while maintaining hierarchical consistency across stacked RBMs.

$$\Delta W_{ij} = \lambda(E_{data}[F_i F_j] - E_{model}[F_i F_j]) \quad (22)$$

Where ΔW_{ij} represents the weight adjustment for connections between feature nodes i and j, and λ is the learning coefficient that governs update magnitude.

A regularized learning mechanism has been incorporated to prevent the accumulation of

redundant information, constraining weight growth within optimal limits. This strategy mimics macaws' resource-sharing behavior during foraging, ensuring balanced feature distribution within the network. Regularization has maintained weight stability, ensuring stacked RBMs preserve hierarchical integrity without introducing computational inefficiencies.

$$\Omega(W) = \gamma \sum_{l} \|W_l\|^2 \quad (23)$$

where $\Omega(W)$ represents the regularization constraint, and γ is the regularization coefficient that prevents excessive weight amplification.

3.6. Temporal Pattern Learning (Coordinated Flight Patterns)

Scarlet Macaw's coordinated flight patterns have influenced temporal pattern learning in SM-DBN, enabling structured sequential feature extraction. This mechanism has strengthened temporal consistency, optimized feature progression across time-dependent data, and enhanced GDM risk prediction.

Feature representations have been aligned across time sequences to ensure structured learning, preventing abrupt variations. This strategy has reflected the macaw's synchronized flight movements, ensuring smooth transitions while navigating environmental changes. This temporal alignment has ensured smooth feature evolution, maintaining structured transitions across multiple layers without disrupting hierarchical abstraction.

$$T_t = W_T T_{t-1} + B_T \quad (24)$$

where T_t denotes the transformed feature at time step t, W_T represents the transition matrix, T_{t-1} is the previous time-step feature, and B_T is the bias term.

Memory retention mechanisms have been incorporated, preserving crucial past information while allowing adaptation to new data. This process has mirrored the macaw's ability to retain flight strategies across different environmental conditions. Memory retention has ensured that past clinical patterns influencing GDM prediction remain accessible, preventing critical information loss.

$$M_t = \alpha M_{t-1} + (1-\alpha) T_t \quad (25)$$

where M_t represents the memory state at time t, M_{t-1} is the prior memory state, and α denotes the retention coefficient.

Feature propagation has been optimized using a structured recurrent mechanism, ensuring that sequential information is processed efficiently without disruption. This adjustment has aligned with the macaw's ability to adjust flight coordination dynamically. This propagation strategy has prevented the degradation of essential GDM-related temporal features, ensuring hierarchical depth across multiple learning layers.

$$F_t = tanh(W_F F_{t-1} + B_F) \quad (26)$$

where F_t denotes the transformed feature at time step t, W_F represents the feature weight matrix, and B_F is the bias term.

Weight matrices have been dynamically optimized to ensure adaptability, allowing the model to adjust learning trajectories based on incoming temporal sequences. This adaptation has reflected the macaw's ability to modify coordinated flight responses in real-time. Weight optimization has enhanced learning efficiency, reducing unnecessary fluctuations in feature abstraction while maintaining predictive stability.

$$W_t^* = W_t - \eta \frac{\partial L}{\partial W_t} \quad (27)$$

where W_t^* represents the updated weight at time t, η is the learning rate, and L denotes the temporal loss function.

An attention mechanism has been introduced to reinforce sequential feature learning, ensuring that high-impact temporal attributes receive greater weighting. This prioritization has aligned with the macaw's ability to adjust visual attention during coordinated flight maneuvers. Attention-based refinement has preserved high-value temporal features, ensuring structured GDM risk prediction across sequential data streams.

$$A_t = \frac{e^{W_A T_t}}{\sum_k e^{W_A T_k}} \quad (28)$$

where A_t represents the attention coefficient at time step t, and W_A is the attention weight matrix.

To maintain robustness, error minimization has been applied iteratively, reducing instability in sequential feature transitions. This strategy has reflected the macaw's ability to recalibrate flight movements dynamically, ensuring smooth navigation. Error minimization has strengthened the model's predictive capacity, ensuring structured learning across sequential GDM datasets.

$$E_t = \sum_i \left(T_{i,t} - \hat{T}_{i,t}\right)^2 \quad (29)$$

where E_t denotes the error function at time step t, $T_{i,t}$ represents the observed feature, and $\hat{T}_{i,t}$ is the predicted feature.

3.7. Backpropagation Fine-Tuning (Cooperative Vigilance)

Scarlet Macaw's cooperative vigilance during foraging has influenced the backpropagation fine-tuning process in SM-DBN, ensuring synchronized weight adjustments. This process has strengthened feature refinement, reduced classification errors, optimized hierarchical learning, and improved the predictive capacity of GDM classification.

Backpropagation has adjusted weights sequentially across layers by propagating error signals backward. This structured update has reflected the macaw's cooperative vigilance, where individuals communicate to avoid threats and refine foraging strategies. Error propagation has ensured that model adjustments remain structured, preventing abrupt weight deviations while reinforcing stable learning across hierarchical layers.

$$\delta_l = \left(\frac{\partial L}{\partial W_l}\right)\sigma'(Z_l) \tag{30}$$

where δ_l represents the error signal at layer l, L is the loss function, W_l is the weight matrix at layer l, and $\sigma'(Z_l)$ denotes the derivative of the activation function.

The learning rate has been dynamically adjusted to prevent unstable convergence based on gradient variations. This adaptation has aligned with the macaw's ability to modify vigilance intensity depending on environmental risks. Learning rate modulation has prevented oscillations in feature adjustments, ensuring balanced weight refinement without overfitting specific training instances.

$$\eta_t = \frac{\eta_0}{1+\lambda t} \tag{31}$$

where η_t represents the learning rate at iteration t, η_0 is the initial learning rate, and λ is the decay coefficient controlling adaptation speed.

Momentum mechanisms have been introduced to accelerate training while mitigating abrupt gradient shifts. This adjustment has mirrored the macaw's group coordination, where individuals maintain synchronized flight speed to optimize foraging efficiency. Momentum incorporation has stabilized learning dynamics, reducing abrupt fluctuations in weight updates and reinforcing smooth feature propagation.

$$W_l^{(t+1)} = W_l^{(t)} - \eta\delta_l + \beta\left(W_l^{(t)} - W_l^{(t-1)}\right) \tag{32}$$

where $W_l^{(t+1)}$ represents the updated weight at iteration $t+1$, β is the momentum coefficient, and $W_l^{(t-1)}$ is the previous weight value.

Regularization techniques have been applied dynamically to prevent overfitting, ensuring that weight growth remains constrained within optimal bounds. This regulation has mimicked the macaw's cooperative behavior, where individuals monitor and control resource sharing. Regularization has preserved model generalization, preventing over-reliance on specific training data patterns while ensuring robust feature refinement.

$$\Omega(W) = \gamma\|W\|^p \tag{33}$$

where $\Omega(W)$ represents the regularization penalty, γ is the regularization coefficient, and p controls the norm type (L1 or L2).

Clipping techniques have been introduced to prevent gradient explosion, ensuring that weight adjustments remain within predefined limits. This approach has reflected the macaw's vigilant flight adjustments, preventing disorientation while navigating complex terrains. Gradient clipping has ensured stability in hierarchical learning, preventing excessive shifts in feature representations that could destabilize GDM risk prediction accuracy.

$$\Delta W = \frac{\Delta W}{max(1, \|\Delta W\|/\tau)} \tag{34}$$

where ΔW represents the weight update, and τ is the clipping threshold restricting excessive gradient magnitudes.

3.8. Weight Regularization (Territorial Behavior)

Scarlet Macaw's territorial behavior has inspired weight regularization in SM-DBN, ensuring controlled weight adjustments to prevent overfitting. This structured regulation has strengthened feature generalization, maintained balanced learning dynamics, reinforced stability across hierarchical layers, and optimized GDM prediction accuracy. L2 regularization has been implemented to constrain weight magnitudes, preventing excessive parameter growth that could lead to overfitting. This regulation has mirrored the macaw's ability to maintain territorial boundaries, ensuring resource optimization without over-exploitation. L2 regularization has enforced weight sparsity, ensuring that learned features remain robust without excessive reliance on specific training patterns.

$$\Omega(W) = \lambda\sum_{i,j} W_{ij}^2 \tag{35}$$

where $\Omega(W)$ represents the penalty function, W_{ij} denotes the weight connecting node i to node j, and λ is the regularization coefficient.

Dropout techniques have been introduced to enhance model robustness, selectively deactivating neurons during training. This

adjustment has reflected the macaw's territorial vigilance, where resource distribution is balanced through selective control. Dropout-based regularization has prevented dependency on specific neurons, ensuring that feature learning remains generalized across different clinical scenarios.

$$z_i = m_i \cdot y_i, \qquad m_i \sim Bernoulli(p) \tag{36}$$

where z_i represents the neuron output after dropout, y_i is the pre-dropout activation, and m_i is a Bernoulli random variable with probability p of retaining the neuron.

A constraint function has been applied to ensure weight stability, restricting parameter updates within predefined bounds. This mechanism has aligned with the macaw's territorial regulation, preventing excessive deviations from optimal positioning. Weight constraint mechanisms have reinforced structured learning, preventing extreme fluctuations in weight adjustments that could disrupt hierarchical learning.

$$W_{ij}^{*} = min\left(W_{max}, max\left(W_{min}, W_{ij}\right)\right) \tag{37}$$

where W_{ij}^{*} represents the constrained weight, W_{max} and W_{min} denote the permissible upper and lower weight limits, respectively.

3.9. Adaptive Learning Rate Adjustment (Seasonal Variation)

Scarlet Macaw's seasonal variation in feeding strategies has inspired adaptive learning rate adjustment in SM-DBN, ensuring stable convergence during training. This mechanism has strengthened learning efficiency, prevented stagnation, maintained structured weight updates, and optimized feature extraction for GDM risk prediction.

A time-based decay strategy has been implemented to gradually reduce the learning rate, preventing erratic fluctuations in weight updates. This adaptation has mirrored the macaw's seasonal adjustments in foraging behaviors, ensuring optimal energy distribution based on environmental conditions. Time-based learning rate decay has maintained controlled optimization dynamics, ensuring that weight updates remain gradual and structured.

$$\eta_t = \frac{\eta_0}{1 + \lambda t} \tag{38}$$

where η_t represents the learning rate at iteration t, η_0 is the initial learning rate, and λ is the decay coefficient controlling the reduction rate.

Momentum-based learning rate adjustments have been introduced to stabilize training while accelerating convergence. This adjustment has aligned with the macaw's ability to regulate movement speed based on environmental stability. Momentum-based learning rate modulation has prevented unstable oscillations, reinforcing structured adaptation across hierarchical layers.

$$\eta_t = \beta \eta_{t-1} + (1 - \beta)\frac{\partial L}{\partial W} \tag{39}$$

where η_t represents the updated learning rate, β is the momentum coefficient, and $\frac{\partial L}{\partial W}$ denotes the gradient of the loss function with respect to the weight parameters.

An adaptive scaling mechanism has been applied to enhance weight sensitivity, ensuring that learning rates dynamically adjust based on weight magnitudes. This strategy has mirrored the macaw's foraging adaptation, where feeding intensity varies depending on seasonal food availability. Adaptive scaling has ensured balanced learning rates across different network layers, preventing excessive sensitivity to specific feature representations.

$$\eta_t = \eta_0 \cdot \frac{|W|}{\sum |W|} \tag{40}$$

where η_t represents the scaled learning rate, W denotes the current weight values, and $\sum |W|$ is the sum of absolute weight magnitudes.

A gradient-normalized approach has been integrated to refine learning rates dynamically, maintaining structured weight updates while optimizing network performance. This mechanism has aligned with the macaw's instinct to modify foraging locations based on food density. A gradient-normalized adjustment has improved learning stability, ensuring that feature refinement remains structured across training iterations.

$$\eta_t = \eta_0 \cdot \frac{\|\nabla L\|}{\|\nabla L_0\|} \tag{41}$$

where η_t represents the adjusted learning rate, ∇L is the gradient norm at time t, and ∇L_0 is the initial gradient norm.

Clipping mechanisms have been applied to prevent extreme learning rate variations, constraining updates within a predefined range. This regulation has mirrored the macaw's territorial awareness, where movement is restricted within optimal boundaries. Learning rate clipping has preserved stability in hierarchical learning, preventing excessive fluctuations in feature representations.

$$\eta_t = max(\eta_{min}, min(\eta_{max}, \eta_t)) \quad (42)$$

where η_t represents the clipped learning rate, and η_{min} and η_{max} denote the lower and upper bounds for learning rate adjustments.

3.10. Robust Foraging Strategy

Scarlet Macaw's robust foraging strategy has inspired missing data handling in SM-DBN, ensuring data completeness through adaptive imputation techniques. This mechanism has strengthened hierarchical learning, minimized bias, reinforced structured feature representations, and optimized GDM classification accuracy.

A probabilistic imputation mechanism has been implemented to estimate missing feature values, preventing inconsistencies. This strategy has mirrored the macaw's ability to locate food sources in unpredictable environments, ensuring survival by adapting its foraging strategy. Probabilistic imputation has preserved dataset integrity, preventing information loss that could disrupt hierarchical learning in SM-DBN.

$$X_i^* = E[X_i|X_{-i}] \quad (43)$$

where X_i^* represents the imputed value for the missing feature X_i, and $E[X_i|X_{-i}]$ denotes the expected value conditioned on observed feature values.

Bayesian estimation has been incorporated to reduce uncertainty, allowing missing values to be inferred based on prior distributions. This adjustment has aligned with the macaw's ability to make predictive foraging decisions based on environmental cues. Bayesian estimation has enhanced imputation accuracy, ensuring inferred values remain statistically consistent with observed data distributions.

$$P(X_i|D) = \frac{P(D|X_i)P(X_i)}{P(D)} \quad (44)$$

where $P(X_i|D)$ represents the posterior probability of the missing feature given dataset D, $P(D|X_i)$ denotes the likelihood, and $P(X_i)$ is the prior probability of the feature.

K-NN imputation has been applied to ensure pattern consistency, estimating missing values based on the closest matching samples. This process has reflected the macaw's cooperative foraging, where individuals share resource information to optimize survival.

$$X_i^* = \frac{1}{k}\sum_{j=1}^{k} X_j \quad (45)$$

where X_i^* represents the imputed value, and X_j denotes the feature values of the k nearest neighbors. k-NN imputation has strengthened feature consistency, ensuring missing values retain structural alignment with observed data distributions.

A matrix factorization method has been employed to handle large-scale missing data, reconstructing missing values through latent feature approximation. This reconstruction has mirrored the macaw's ability to locate food across dispersed regions, optimizing search strategies. Matrix factorization has reinforced dataset completeness, ensuring hierarchical feature representations remain intact for structured learning.

$$X \approx UV^T \quad (46)$$

Where X represents the incomplete data matrix, U and V are the factorized low-rank matrices approximating missing values.

A GMM-based approach has been utilized to model missing values probabilistically, allowing for flexible data reconstruction. This mechanism has aligned with the macaw's adaptive foraging behavior, where resource allocation is optimized dynamically.

$$P(X) = \sum_{m=1}^{M} \pi_m N(X|\mu_m, \Sigma_m) \quad (47)$$

where $P(X)$ represents the probability density function of the missing feature, π_m denotes the mixture weight, and $N(X|\mu_m, \Sigma_m)$ is the Gaussian component with mean μ_m and covariance Σ_m.
GMM-based completion has ensured flexible missing data handling, reinforcing structured learning across sequential GDM datasets.

3.11. Structured Data Propagation (Swarm Intelligence in Flight)

Scarlet Macaw's swarm intelligence in flight has inspired structured data propagation in SM-DBN, ensuring seamless feature transmission across hierarchical layers. This mechanism has strengthened information coherence, minimized propagation loss, reinforced stable learning dynamics, and optimized GDM classification efficiency.

To maintain structured feature propagation, hierarchical feature transmission has been

implemented, ensuring a seamless flow of extracted representations across layers. This approach has mirrored the macaw's coordinated flight, where individuals maintain synchronized movement to optimize group efficiency. Feature transmission has preserved hierarchical consistency, preventing feature degradation while reinforcing structured learning across deep representations.

$$H_l = W_l H_{l-1} + B_l \quad (48)$$

where H_l represents the propagated feature at layer l, W_l denotes the weight matrix connecting layer $l-1$ to layer l, and B_l is the bias term.

An attention-based routing mechanism has been introduced to prioritize critical feature propagation, dynamically adjusting data transmission based on feature significance. This refinement has aligned with the macaw's ability to modulate flight formations for energy-efficient movement. Attention-based routing has reinforced structured feature prioritization, ensuring that critical GDM-related attributes remain prominent throughout hierarchical learning.

$$A_l = \frac{e^{W_A H_{l-1}}}{\sum_k e^{W_A H_k}} \quad (49)$$

where A_l represents the attention coefficient for layer l, and W_A is the attention weight matrix.

Gradually-directed transmission has been applied to optimize structured data propagation, ensuring that feature updates follow an efficient trajectory. This adjustment has mirrored the macaw's ability to adjust flight direction dynamically to minimize energy expenditure. Gradient-directed propagation has reinforced structured learning, ensuring hierarchical representations remain stable without abrupt fluctuations.

$$G_l = -\frac{\partial L}{\partial H_l} \quad (50)$$

where G_l represents the propagated gradient at layer l, and L denotes the loss function.

Adaptive weight modulation has been incorporated to enhance structured feature transmission, ensuring propagation strength dynamically adjusts based on network state. This adaptation has mirrored the macaw's ability to modify flight velocity based on environmental changes. Momentum-assisted transmission has improved hierarchical feature stability, reinforcing structured learning across deep network layers.

$$M_l = \beta M_{l-1} + (1-\beta) H_l \quad (51)$$

where M_l represents the momentum-enhanced propagated feature, and β is the momentum coefficient.

A regularization-based stabilization mechanism has been applied to prevent excessive deviations in feature transmission, constraining feature variations within optimal bounds. This adjustment has mirrored the macaw's ability to regulate flight speed dynamically, preventing abrupt movement shifts. Regularization-based stabilization has reinforced structured feature transmission, ensuring propagation coherence while preventing feature explosion.

$$\Omega(H) = \gamma \sum_l \|H_l\|^2 \quad (52)$$

where $\Omega(H)$ represents the regularization function, and γ is the stabilization coefficient.

3.12. Explainable Decision Support (Vocal Communication Cues)

Scarlet Macaw's vocal communication cues have inspired explainable decision support in SM-DBN, ensuring model interpretability through structured feature attribution. This mechanism has strengthened decision transparency, reinforced clinical trust, optimized model explainability, minimized uncertainty, and enhanced GDM classification interpretability.

Feature attribution mechanisms have been integrated to enhance decision interpretability, ensuring structured alignment between model predictions and contributing clinical variables. This refinement has mirrored the macaw's ability to assign meaning to vocalizations for effective communication. Feature attribution has reinforced structured decision transparency, ensuring critical GDM-related factors remain interpretable for clinical applications.

$$S_i = \frac{\partial \hat{Y}}{\partial X_i} \quad (53)$$

where S_i represents the attribution score of the feature X_i, and $\hat{Y}$ denotes the predicted outcome.

Shapley Additive exPlanations (SHAP) values have been computed to quantify feature impact on predictions, ensuring probabilistic consistency in feature contributions. This method has aligned with the macaw's structured vocal signals, where information exchange remains optimized. SHAP value computation has reinforced

model explainability, ensuring structured decision reasoning across hierarchical feature representations.

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad (54)$$

where ϕ_i represents the SHAP value for feature i, S denotes the feature subset, and $f(S)$ is the model's output for subset S.

Local Interpretable Model-Agnostic Explanations (LIME) techniques have been applied to ensure localized interpretability, approximating black-box predictions through interpretable linear models. This adaptation has mirrored the macaw's ability to adjust vocal frequencies for optimized perception. LIME-based model approximation has preserved interpretability, ensuring that decision-support insights remain structured while maintaining predictive accuracy.

$$g(x) = \sum_{i=1}^{n} w_i X_i + b \quad (55)$$

where $g(x)$ represents the surrogate model approximation, w_i is the learned coefficient for the feature X_i, and b denotes the bias term.

An attention-based mechanism has been incorporated to enhance decision justification, ensuring that feature relevance remains quantifiable in prediction inference. This approach has mirrored the macaw's ability to focus on critical vocal elements during communication. Attention-based decision support has reinforced model accountability, ensuring that hierarchical learning remains structured while enhancing feature weighting transparency.

$$A_i = \frac{e^{W_A X_i}}{\sum_j e^{W_A X_j}} \quad (56)$$

where A_i represents the attention coefficient for the feature X_i, and W_A is the attention weight matrix.

Entropy-based confidence estimation has been implemented to quantify uncertainty in decision-making, ensuring that prediction reliability remains measurable. This adaptation has mirrored the macaw's strategic vocal cues, where variations in tone convey different levels of confidence. Entropy-based uncertainty quantification has reinforced structured decision reliability, ensuring that GDM classification remains interpretable across hierarchical model outputs

$$H(Y) = -\sum_{i=1}^{n} P(Y_i) log P(Y_i) \quad (57)$$

where $H(Y)$ represents the entropy-based uncertainty measure, and $P(Y_i)$ denotes the probability distribution over possible predictions.

3.13. Roosting for Knowledge Retention

Scarlet Macaw's roosting behavior for knowledge retention has inspired the final output prediction in SM-DBN, ensuring structured inference for GDM classification. This mechanism has strengthened predictive reliability, reinforced model generalization, optimized decision accuracy, minimized inference errors, and enhanced hierarchical learning consistency.

A softmax-based classification mechanism has been incorporated to ensure structured output prediction, assigning probabilistic confidence scores to different prediction categories. This approach has mirrored the macaw's ability to retain learned foraging locations, ensuring optimized resource retrieval. Softmax-based classification has reinforced structured inference, ensuring that predictive outputs remain probabilistically optimized for clinical decision support.

$$P(Y_i|X) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \quad (58)$$

where $P(Y_i|X)$ represents the probability of class i given feature set X, and Z_i is the output of the final layer before classification.

Bayesian decision rules have been applied, ensuring structured probabilistic reasoning in the final prediction stage. This refinement has aligned with the macaw's decision to roost in secure locations, ensuring survival optimization. Bayesian decision rule implementation has ensured that classification outcomes remain structured while reinforcing predictive confidence for GDM detection.

$$Y^* = arg\max_i P(Y_i|X)P(X) \quad (59)$$

where Y^* represents the predicted class, $P(Y_i|X)$ is the posterior probability of class i, and $P(X)$ denotes the prior probability of the input feature set.

A log-likelihood estimation technique has been applied, ensuring that inference reliability remains optimized. This adaptation has mirrored the macaw's structured environmental safety evaluation before roosting, ensuring minimal risk. The log-likelihood estimation has reinforced predictive

robustness, preventing classification inconsistencies while optimizing inference stability.

$$L(Y|X) = \sum_{i=1}^{n} logP(Y_i|X) \quad (60)$$

where $L(Y|X)$ represents the log-likelihood function, ensuring that classification predictions align with learned probabilistic distributions.

A threshold-based decision refinement mechanism has been introduced, ensuring predictions remain aligned with structured risk classification guidelines. This approach has mirrored the macaw's strategic roosting adjustments to maintain ecosystem balance. Threshold-based decision refinement has reinforced structured inference, ensuring that GDM risk classification remains clinically interpretable.

$$Y^* = \begin{cases} 1, & P(Y_1|X) > \tau \\ 0, & P(Y_0|X) \geq \tau \end{cases} \quad (61)$$

where τ represents the threshold value for decision classification, ensuring that inference aligns with structured predictive confidence.

Entropy-based decision calibration has been implemented, ensuring structured probability weighting in classification predictions. This adaptation has mirrored the macaw's ability to assess safety before selecting a roosting location, ensuring environmental adaptation. Entropy-based confidence quantification has reinforced structured decision-making, ensuring predictive transparency in hierarchical learning.

$$H(Y) = -\sum_{i=1}^{n} P(Y_i|X) logP(Y_i|X) \quad (62)$$

where $H(Y)$ represents the entropy measure for predictive uncertainty, ensuring that inference reliability remains quantified.

3.14. SM-DBN Process Flow

The SM-DBN Process Flow outlines the sequential steps for accurate prediction of GDM by integrating bio-inspired optimization techniques. The process starts with feature selection, where clinically relevant attributes are extracted from the dataset, followed by greedy layer-wise pretraining of Restricted Boltzmann Machines (RBMs). These layers are optimized using contrastive divergence, ensuring efficient feature extraction. The stacked RBMs undergo temporal pattern learning to capture evolving GDM risk factors across pregnancy trimesters. Data propagation occurs through multiple layers, influenced by mechanisms inspired by the Scarlet Macaw's adaptive foraging and group behavior, which ensures optimal feature communication across the network. Backpropagation fine-tuning refines the weights, while weight regularization prevents overfitting by constraining unnecessary complexity. Adaptive learning rates adjust dynamically, enhancing convergence. Missing data handling ensures data completeness for robust learning. Finally, explainable decision support outputs clinically interpretable results, providing transparent, reliable GDM risk predictions, guiding healthcare practitioners toward early intervention. The overall algorithm is given below:

Algorithm: SM-DBN

Input:
- Clinical dataset with patient records containing multiple features relevant to GDM prediction.

Output:
- Optimized classification results ensure structured inference and predictive reliability.

Procedure:

Step 1: Feature Selection (Selective Feeding): Extracts high-value clinical attributes while discarding irrelevant features for efficient learning.

Step 2: Greedy Layer-Wise Training (Adaptive Foraging): Trains each Restricted Boltzmann Machine (RBM) independently before stacking for hierarchical learning.

Step 3: Contrastive Divergence Optimization (Beak Adaptation): Refines weight updates iteratively to optimize feature representation and model convergence.

Step 4: Hierarchical Feature Extraction (Seed Dispersal): Ensures structured propagation of extracted features across multiple network layers.

Step 5: Stacking RBMs (Group Foraging): Sequentially integrates pretrained RBMs to form a deep feature extraction architecture.

Step 6: Temporal Pattern Learning (Coordinated Flight Patterns): Captures time-dependent variations in patient data for structured sequential analysis.

Step 7: Backpropagation Fine-Tuning (Cooperative Vigilance): Adjusts

	network weights iteratively using gradient-based updates to reduce classification errors.
Step 8:	Weight Regularization (Territorial Behavior): Constrains weight magnitudes to prevent overfitting and maintain generalization.
Step 9:	Adaptive Learning Rate Adjustment (Seasonal Variation): Modifies learning rates dynamically to optimize model convergence and stability.
Step 10:	Robust Foraging Strategy: Imputes missing values using probabilistic and statistical techniques to ensure data completeness.
Step 11:	Structured Data Propagation (Swarm Intelligence in Flight): Ensures seamless hierarchical transmission of extracted features for efficient learning.
Step 12:	Explainable Decision Support (Vocal Communication Cues): Integrates interpretability techniques such as SHAP and LIME for transparent model decisions.
Step 13:	Final Output Prediction (Roosting for Knowledge Retention): Performs structured classification using softmax probability and Bayesian decision rules for final inference.

The **SM-DBN** model brings innovative solutions to predicting **Gestational Diabetes Mellitus (GDM)** by leveraging bio-inspired deep learning techniques. By combining hierarchical learning with bio-inspired optimization, the model can effectively address clinical data complexities such as class imbalance, temporal variations, and missing information. This results in a more robust, efficient, and interpretable prediction system supporting timely healthcare decision-making. The significant advantages of **SM-DBN** are:

- **Interpretability in Decision-Making:** SM-DBN integrates explainable AI techniques, such as SHAP and LIME, ensuring that healthcare professionals understand the reasoning behind the model's predictions, fostering trust and clinical adoption.
- **Handling Temporal Risk Factors:** The model accounts for dynamic changes in GDM risk across pregnancy stages, providing a continuous assessment of maternal health rather than static, one-time evaluations.
- **Robust Data Integration:** SM-DBN has been designed to handle incomplete, noisy, or inconsistent clinical data, making it adaptable to healthcare environments where data quality may vary.
- **Optimized Model Convergence:** By applying bio-inspired optimization methods, the model avoids suboptimal solutions and ensures that the training process converges more efficiently, reducing computational time and improving performance.
- **Customizable and Flexible for Healthcare Integration:** SM-DBN can be seamlessly integrated into existing healthcare systems, with flexibility to operate on both large-scale datasets and individual patient records, supporting real-time, actionable insights for clinicians.

4. DATASET

The dataset utilized in this analysis captures maternal health indicators collected in a forward-designed format over a multi-year span. A total of 3525 observational entries comprise this collection, each featuring 15 distinct elements reflecting physical, hereditary, and pregnancy-specific variables. These parameters were carefully selected to align with clinically validated risk contributors for glucose intolerance during gestation. The data instances have been distinctly divided into two predictive classes, with one group representing elevated susceptibility and the other indicating no clinical presence of GDM. The division of labels shows 2153 samples aligned with expected glycemic outcomes, while 1372 reflect confirmed gestational diabetics. Such structuring ensures both class separation and real-world distribution symmetry. Approximately three-fourths of the data instances were assigned for model learning, allowing the remaining entries to assess output reliability without leakage or redundancy.

Table 1: Feature and Class Summary of the Dataset

Component Evaluated	Quantitative Summary
Total Observational Units	3525
Normal Glucose Response Entries	2153
Gestational Diabetes Confirmations	1372
Variables Collected per Record	15 descriptive indicators
Binary Grouping for Label Prediction	Present

5. RESULTS AND DISCUSSION

This section demonstrates how each model performs in terms of predictive reliability, class balance, misclassification control, diagnostic strength, and accurate identification of GDM. By analysing both strong and weak areas, the results highlight the superior performance of SM-DBN, while also revealing the limitations of OD-DSAE and AHDHS in handling clinical variability in GDM prediction.

5.1. Classification Accuracy

Classification accuracy measures how often a model correctly predicts outcomes across all classes. In Figure 1 and Table 2, the horizontal axis lists the evaluated algorithms, and the vertical axis displays their accuracy values. OD-DSAE reaches only 57.106%, likely because its deep autoencoder focuses too heavily on reconstructive loss, neglecting discriminative learning needed for precise risk classification. AHDHS improves to 69.390%, but its reliance on fixed base learner combinations limits adaptability to individual risk factors. SM-DBN significantly surpasses both with 81.277%. This boost stems from using entropy-based input refinement and contrastive divergence-driven pretraining. SM-DBN adaptively filters out noisy or redundant features and strengthens signal representation across layers by mimicking a scarlet macaw's selective energy foraging. This layered focus allows it to detect dominant and subtle indicators of gestational diabetes, leading to a marked increase in accuracy across a varied patient dataset.

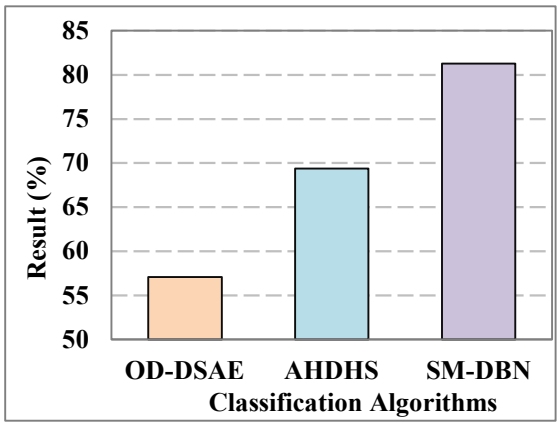

Figure 1. Classification Accuracy - Comparative Evaluation

Table 2. Classification Accuracy - Quantitative Assessment

Classification Algorithms	Classification Accuracy (%)
OD-DSAE	57.106
AHDHS	69.390
SM-DBN	81.277

5.2. Matthews Correlation Coefficient

Matthews Correlation Coefficient (MCC) is a balanced metric that accounts for all prediction outcomes, offering a single value that reflects overall classification quality, even when class distributions are unequal.

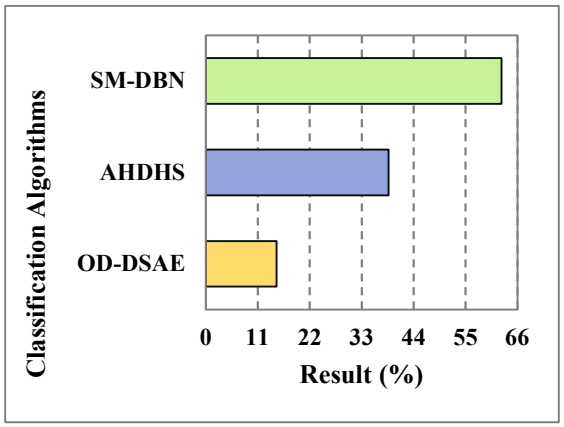

Figure 2. Matthews Correlation Coefficient - Comparative Evaluation

In Figure 2 and Table 3, the x-axis represents the classification models, and the y-axis shows their MCC scores in percentage. OD-DSAE performs the weakest with 14.991%, reflecting its poor ability to maintain consistency across both

positive and negative predictions—this results from its restricted autoencoding structure and lack of boundary-aware learning. AHDHS, scoring 38.701%, shows some improvement but struggles with irregular base learner synergy, leading to unstable correlations in prediction. In contrast, SM-DBN achieves a strong MCC of 62.560%, indicating reliable output alignment with ground truth. This can be credited to its multi-layer training that refines each layer independently and its entropy-driven input filtering, ensuring the model captures GDM risk patterns with balanced precision and sensitivity

Table 3. Matthews Correlation Coefficient - Comparative Evaluation

Classification Algorithms	Matthews Correlation Coefficient (%)
OD-DSAE	14.991
AHDHS	38.701
SM-DBN	62.560

5.3. Error Rate

Error rate expresses the proportion of incorrect predictions a model makes and is an inverse measure of overall performance—lower values indicate better reliability. As illustrated in Figure 3 and Table 4, the classification algorithms are shown along the x-axis, and their respective error rates are plotted on the y-axis. OD-DSAE records the highest error (42.894%), which reflects its limited adaptability to clinical variability.

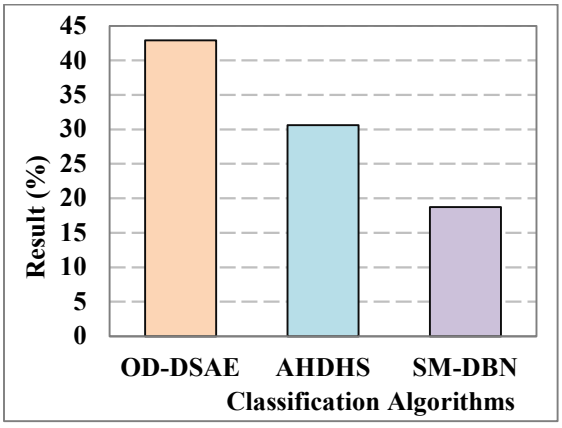

Figure 3. Error Rate - Comparative Evaluation

The model's static reconstruction mechanism fails to adjust to overlapping feature spaces, resulting in frequent misclassifications. AHDHS improves to 30.610%, yet its error remains substantial. This is likely caused by ensemble instability, where conflicting outputs from base learners reduce overall consensus. SM-DBN achieves the lowest error rate of 18.723% due to its robust training pipeline. Its layer-wise refinement, contrastive divergence optimization, and entropy-based filtering reduce noise propagation through the network, leading to fewer classification mistakes and greater stability across GDM risk scenarios.

Table 4. Error Rate - Comparative Evaluation

Classification Algorithms	Error Rate (%)
OD-DSAE	42.894
AHDHS	30.610
SM-DBN	18.723

5.4. Youden's Index

Youden's Index is a diagnostic metric combining sensitivity and specificity into a single value, highlighting a model's effectiveness in correctly identifying positive and negative cases. In Figure 4 and Table 5, the x-axis represents the classification algorithms, and the y-axis displays their Youden's Index scores in percentage. OD-DSAE posts the lowest value (14.942%), suggesting that it struggles to balance identifying actual GDM cases fairly and avoiding false alarms, primarily due to its inability to separate overlapping risk profiles during encoding. AHDHS improves to 38.673%, yet its base learners lack synchronized decision calibration, which restricts its full potential for accurate boundary discrimination. SM-DBN achieves a leading score of 62.570%, showcasing its ability to preserve strong differentiation across both clinical classes. The model's entropy-ranked features and structurally trained layers support sharper recognition of nuanced GDM indicators, resulting in a more clinically balanced prediction profile.

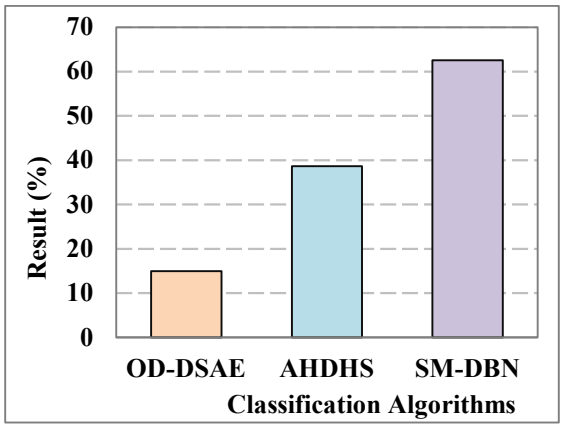

Figure 4. Youden's Index - Comparative Evaluation

Table 5. Youden's Index - Comparative Evaluation

Classification Algorithms	Youden's Index (%)
OD-DSAE	14.942
AHDHS	38.673
SM-DBN	62.570

5.5. Critical Success Index

The Critical Success Index (CSI) evaluates how well a model identifies true positive cases among all predicted and actual positives. It makes it especially meaningful in clinical prediction, where missed cases or false alarms carry risk. Figure 5 and Table 6, the x-axis lists the models, while the y-axis represents their CSI percentages. OD-DSAE reaches 40.776%, indicating frequent misclassifications of GDM cases, likely because its learning process emphasizes reconstruction rather than event detection. AHDHS performs better at 54.473%, but its fixed stacking strategy lacks dynamic refinement, leading to occasional misfires in identifying borderline cases. SM-DBN performs most effectively with a CSI of 68.720%, combining biologically inspired feature prioritization with structured training. Its layered abstraction helps the model detect complex clinical patterns while maintaining precision in risk classification, making it more dependable in identifying actual GDM cases without generating excessive false positives.

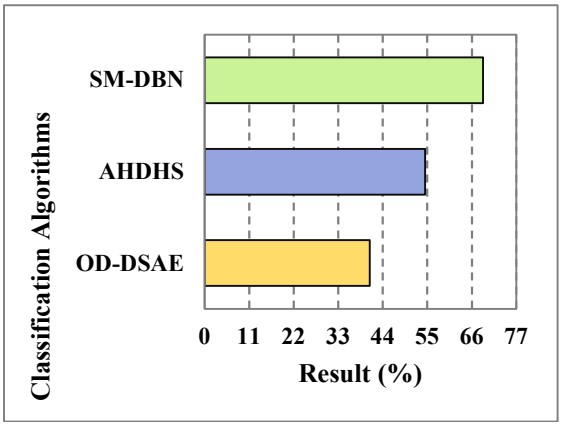

Figure 5. Critical Success Index - Comparative Evaluation

Table 6. Critical Success Index - Comparative Evaluation

Classification Algorithms	Critical Success Index (%)
OD-DSAE	40.776
AHDHS	54.473
SM-DBN	68.720

Comparative PMI Analysis:

A structured Plus–Minus–Interesting (PMI) evaluation highlights key distinctions among compared models. The Plus aspects of OD-DSAE include strong anomaly filtration and representation learning; however, its Minus dimension reflects limited discriminative power and weak class boundary modeling. AHDHS demonstrates ensemble adaptability and moderate robustness, yet suffers from static learner stacking and inconsistent calibration across risk gradients. SM-DBN presents Plus elements in entropy-guided feature prioritization, hierarchical RBM refinement, and temporal abstraction. The Interesting dimension emerges in its biologically inspired optimization behavior, enabling dynamic convergence control and uncertainty-aware inference. These findings confirm performance superiority while clarifying architectural trade-offs.

6. CONCLUSION

The Scarlet Macaw-Inspired Deep Belief Network (SM-DBN) provides an innovative solution to the challenges of predicting Gestational Diabetes Mellitus (GDM) early in pregnancy. The model enhances predictive accuracy by leveraging bio-inspired optimization techniques while handling complex real-world clinical data, including missing values and imbalances. Its ability to adapt to temporal risk factors and provide explainable outputs makes it a promising tool for clinical integration. SM-DBN improves early detection, minimizes unnecessary testing, and reduces the psychological and economic burdens of GDM diagnosis. The approach's scalability and robustness allow it to be applied in various healthcare settings, making it a versatile addition to current GDM screening protocols. Ultimately, SM-DBN can potentially improve maternal and fetal health outcomes, offering an accessible, effective, and reliable alternative to traditional diagnostic methods. Its success could drive future advancements in predictive healthcare, setting the foundation for more personalized and timely interventions in pregnancy-related health risks. Future enhancements will expand the model's adaptability to other pregnancy-related complications and optimize its integration into diverse clinical environments, further improving maternal healthcare systems.

The research contribution lies in introducing a bio-inspired hierarchical deep belief architecture that integrates entropy-ranked feature refinement, adaptive contrastive divergence, and trimester-sensitive temporal modeling within a

unified explainable framework. The model advances predictive healthcare by reducing diagnostic delay, limiting unnecessary OGTT screening, and strengthening clinician trust through interpretable inference. In the current scenario of rising maternal metabolic disorders and digitized obstetric care systems, SM-DBN demonstrates scalable deployment potential in resource-constrained and urban healthcare environments.

6.1. Future-Oriented Solutions and Practical Extensions

Remaining challenges include cross-institutional variability, real-time deployment constraints, and integration into obstetric workflow systems. Potential solutions involve federated training architectures to enhance multi-center generalization, lightweight inference compression for mobile deployment, and integration into electronic health record dashboards with threshold-calibrated alert systems. Incorporating continual learning mechanisms may allow trimester-progressive model updating without retraining from scratch. These extensions represent practical translational pathways for strengthening the clinical readiness of SM-DBN.

References

[1] A. T. Villikudathil, D. H. Mc Guigan, and A. English, "Exploring metformin monotherapy response in Type-2 diabetes: Computational insights through clinical, genomic, and proteomic markers using machine learning algorithms," *Comput. Biol. Med.*, vol. 171, p. 108106, 2024, doi: https://doi.org/10.1016/j.compbiomed.2024.108106.

[2] S. A. Tiruneh, D. L. Rolnik, H. J. Teede, and J. Enticott, "Prediction of pre-eclampsia with machine learning approaches: Leveraging important information from routinely collected data," *Int. J. Med. Inform.*, vol. 192, p. 105645, 2024, doi: https://doi.org/10.1016/j.ijmedinf.2024.105645.

[3] S. L. Cichosz, M. H. Jensen, O. Hejlesen, S. D. Henriksen, A. M. Drewes, and S. S. Olesen, "Prediction of pancreatic cancer risk in patients with new-onset diabetes using a machine learning approach based on routine biochemical parameters," *Comput. Methods Programs Biomed.*, vol. 244, p. 107965, 2024, doi: https://doi.org/10.1016/j.cmpb.2023.107965.

[4] L. Z. Chee, S. Sivakumar, K. H. Lim, and A. A. Gopalai, "Gait acceleration-based diabetes detection using hybrid deep learning," *Biomed. Signal Process. Control*, vol. 92, p. 105998, 2024, doi: https://doi.org/10.1016/j.bspc.2024.105998.

[5] S. A. Tanim, A. R. Aurnob, T. E. Shrestha, M. D. R. I. Emon, M. F. Mridha, and M. S. U. Miah, "Explainable deep learning for diabetes diagnosis with DeepNetX2," *Biomed. Signal Process. Control*, vol. 99, p. 106902, 2025, doi: https://doi.org/10.1016/j.bspc.2024.106902.

[6] N. Musacchio *et al.*, "A transparent machine learning algorithm uncovers HbA1c patterns associated with therapeutic inertia in patients with type 2 diabetes and failure of metformin monotherapy," *Int. J. Med. Inform.*, vol. 190, p. 105550, 2024, doi: https://doi.org/10.1016/j.ijmedinf.2024.105550.

[7] S. Nayak *et al.*, "Development of a machine learning-based model for the prediction and progression of diabetic kidney disease: A single centred retrospective study," *Int. J. Med. Inform.*, vol. 190, p. 105546, 2024, doi: https://doi.org/10.1016/j.ijmedinf.2024.105546.

[8] L. M. Kunjachen and R. Kavitha, "Advancing cardiovascular risk prediction: A fusion of SVM models with fuzzy logic and the Sugeno integral," *Biomed. Signal Process. Control*, vol. 106, p. 107774, 2025, doi: https://doi.org/10.1016/j.bspc.2025.107774.

[9] M. Oulhadj *et al.*, "Diabetic retinopathy prediction based on vision transformer and modified capsule network," *Comput. Biol. Med.*, vol. 175, p. 108523, 2024, doi: https://doi.org/10.1016/j.compbiomed.2024.108523.

[10] H. Qi, X. Song, S. Liu, Y. Zhang, and K. K. L. Wong, "KFPredict: An ensemble learning prediction framework for diabetes based on fusion of key features," *Comput. Methods Programs Biomed.*, vol. 231, p. 107378, 2023, doi: https://doi.org/10.1016/j.cmpb.2023.107378.

[11] A. Daza, C. F. Ponce Sánchez, G. Apaza-Perez, J. Pinto, and K. Zavaleta Ramos, "Stacking ensemble approach to diagnosing the disease of diabetes," *Informatics Med. Unlocked*, vol. 44, p. 101427, 2024, doi: https://doi.org/10.1016/j.imu.2023.101427.

[12] S. Kim *et al.*, "Development and Validation of a Machine Learning Algorithm for Predicting Diabetes Retinopathy in Patients With Type 2 Diabetes: Algorithm Development Study," *JMIR Med. Informatics*, vol. 13, 2025, doi: https://doi.org/10.2196/58107.

[13] T. G and K. N.K, "Edge computing for diabetic early disease prediction based on optimized ensemble boosting Fuzzified neural network using recursive support vector feature selection," *Biomed. Signal Process. Control*, vol. 108, p. 107913, 2025, doi: https://doi.org/10.1016/j.bspc.2025.107913.

[14] V. Rajathi, A. Chinnasamy, and P. Selvakumari, "DUTC Net: A novel deep ulcer tissue classification network with stage prediction and treatment plan recommendation," *Biomed. Signal Process. Control*, vol. 90, p. 105855, 2024, doi: https://doi.org/10.1016/j.bspc.2023.105855.

[15] H. Cheng, J. Zhu, P. Li, and H. Xu, "Combining knowledge extension with convolution neural network for diabetes prediction," *Eng. Appl. Artif. Intell.*, vol. 125, p. 106658, 2023, doi: https://doi.org/10.1016/j.engappai.2023.106658.

[16] S. Mam, D. Wichadakul, and P. Vateekul, "Drug Repurposing for Type 2 Diabetes Using Combined Textual and Structural Graph Representation Based on Transformer," *IEEE Access*, vol. 11, pp. 65711–65724, 2023, doi: 10.1109/ACCESS.2023.3289863.

[17] M. Sinsirimongkhon, S. Arwatchananukul, and P. Temdee, "Multi-Class Classification Method with Feature Engineering for Predicting Hypertension with Diabetes," *J. Mob. Multimed.*, vol. 19, no. 3, pp. 799–821, 2023, doi: 10.13052/jmm1550-4646.1937.

[18] A. Alexiadis *et al.*, "Next-Day Prediction of Hypoglycaemic Episodes Based on the Use of a Mobile App for Diabetes Self-Management," *IEEE Access*, vol. 12, pp. 7469–7478, 2024, doi: 10.1109/ACCESS.2024.3350201.

[19] Ş. Kolozali, S. L. White, S. Norris, M. Fasli, and A. van Heerden, "Explainable Early Prediction of Gestational Diabetes Biomarkers by Combining Medical Background and Wearable Devices: A Pilot Study With a Cohort Group in South Africa," *IEEE J. Biomed. Heal. Informatics*, vol. 28, no. 4, pp. 1860–1871, 2024, doi: 10.1109/JBHI.2024.3361505.

[20] M. Aljaafari, S. E. El-Deep, A. A. Abohany, and S. E. Sorour, "Integrating Innovation in Healthcare: The Evolution of 'CURA's' AI-Driven Virtual Wards for Enhanced Diabetes and Kidney Disease Monitoring," *IEEE Access*, vol. 12, pp. 126389–126414, 2024, doi: 10.1109/ACCESS.2024.3451369.

[21] G. Annuzzi *et al.*, "Impact of Nutritional Factors in Blood Glucose Prediction in Type 1 Diabetes Through Machine Learning," *IEEE Access*, vol. 11, pp. 17104–17115, 2023, doi: 10.1109/ACCESS.2023.3244712.

[22] D. Parra, D. Joedicke, J. M. Velasco, G. Kronberger, and J. I. Hidalgo, "Learning Difference Equations With Structured Grammatical Evolution for Postprandial Glycaemia Prediction," *IEEE J. Biomed. Heal. Informatics*, vol. 28, no. 5, pp. 3067–3078, 2024, doi: 10.1109/JBHI.2024.3371108.

[23] G. Cappon, F. Prendin, A. Facchinetti, G. Sparacino, and S. D. Favero, "Individualized Models for Glucose Prediction in Type 1 Diabetes: Comparing Black-Box Approaches to a Physiological White-Box One," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 11, pp. 3105–3115, 2023, doi: 10.1109/TBME.2023.3276193.

[24] J. M. Lee, R. Pop-Busui, J. M. Lee, J. Fleischer, and J. Wiens, "Shortcomings in the Evaluation of Blood Glucose Forecasting," *IEEE Trans. Biomed. Eng.*, vol. 71, no. 12, pp. 3424–3431, 2024, doi: 10.1109/TBME.2024.3424665.

[25] Z. Yu, W. Luo, R. Tse, and G. Pau, "DMNet: A Personalized Risk Assessment Framework for Elderly People With Type 2 Diabetes," *IEEE J. Biomed. Heal. Informatics*, vol. 27, no. 3, pp. 1558–1568, 2023, doi: 10.1109/JBHI.2022.3233622.

[26] R. Jia, X. Zhao, S. Zhang, and X. Yu, "A Tube-NMPC Approach for Robust Control of Glucose in Type 1 Diabetes Mellitus," *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 9001–9012, 2025, doi: 10.1109/TASE.2024.3494819.

[27] A. A. Linkon *et al.*, "Evaluation of Feature Transformation and Machine Learning Models on Early Detection of Diabetes Mellitus," *IEEE Access*, vol. 12, pp. 165425–165440, 2024, doi:

10.1109/ACCESS.2024.3488743.

[28] K. Alnowaiser, "Improving Healthcare Prediction of Diabetic Patients Using KNN Imputed Features and Tri-Ensemble Model," *IEEE Access*, vol. 12, pp. 16783–16793, 2024, doi: 10.1109/ACCESS.2024.3359760.

[29] V. K. Daliya and T. K. Ramesh, "A Cloud-Based Optimized Ensemble Model for Risk Prediction of Diabetic Progression—An Azure Machine Learning Perspective," *IEEE Access*, vol. 13, pp. 11560–11575, 2025, doi: 10.1109/ACCESS.2025.3528033.

[30] P. F. De Paola, A. Paglialonga, P. Palumbo, K. Keshavjee, F. Dabbene, and A. Borri, "The Long-Term Effects of Physical Activity on Blood Glucose Regulation: A Model to Unravel Diabetes Progression," *IEEE Control Syst. Lett.*, vol. 7, pp. 2916–2921, 2023, doi: 10.1109/LCSYS.2023.3290774.

[31] Z. Zhang *et al.*, "A novel evolutionary ensemble prediction model using harmony search and stacking for diabetes diagnosis," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 1, p. 101873, 2024, doi: 10.1016/j.jksuci.2023.101873.

[32] A. Sumathi, S. Meganathan, and B. V. Ravisankar, "An Intelligent Gestational Diabetes Diagnosis Model Using Deep Stacked Autoencoder," *Comput. Mater. Contin.*, vol. 69, no. 3, pp. 3109–3126, 2021, doi: 10.32604/cmc.2021.017612.

[33]. J. Ramkumar and R. Vadivel, "Improved frog leap inspired protocol (IFLIP) – for routing in cognitive radio ad hoc networks (CRAHN)," World J. Eng., vol. 15, no. 2, pp. 306–311, 2018, doi: 10.1108/WJE-08-2017-0260.

[34]. J. Ramkumar and R. Vadivel, "Improved Wolf prey inspired protocol for routing in cognitive radio Ad Hoc networks," Int. J. Comput. Networks Appl., vol. 7, no. 5, pp. 126–136, 2020, doi: 10.22247/ijcna/2020/202977.

[35]. K. S. J. Marseline, J. Ramkumar, and D. R. Medhunhashini, "Sophisticated Kalman Filtering-Based Neural Network for Analyzing Sentiments in Online Courses," in Smart Innovation, Systems and Technologies, A. K. Somani, A. Mundra, R. K. Gupta, S. Bhattacharya, and A. P. Mazumdar, Eds., Springer Science and Business Media Deutschland GmbH, 2024, pp. 345–358. doi: 10.1007/978-981-97-3690-4_26.

[36]. L. Mani, S. Arumugam, and R. Jaganathan, "Performance Enhancement of Wireless Sensor Network Using Feisty Particle Swarm Optimization Protocol," in ACM International Conference Proceeding Series, Association for Computing Machinery, 2022. doi: 10.1145/3590837.3590907.

[37]. J. Ramkumar and R. Vadivel, "CSIP—cuckoo search inspired protocol for routing in cognitive radio ad hoc networks," in Advances in Intelligent Systems and Computing, D. P. Mohapatra and H. S. Behera, Eds., Springer Verlag, 2017, pp. 145–153. doi: 10.1007/978-981-10-3874-7_14.

[38]. N. K. Ojha, A. Pandita, and J. Ramkumar, "Cyber security challenges and dark side of AI: Review and current status," in Demystifying the Dark Side of AI in Business, IGI Global, 2024, pp. 117–137. doi: 10.4018/979-8-3693-0724-3.ch007.

[39]. R. Jaganathan, S. Mehta, and R. Krishan, Intelligent Decision Making Through Bio-Inspired Optimization. IGI Global, 2024. doi: 10.4018/979-8-3693-2073-0.

[40]. J. Ramkumar, K. S. Jeen Marseline, and D. R. Medhunhashini, "Relentless Firefly Optimization-Based Routing Protocol (RFORP) for Securing Fintech Data in IoT-Based Ad-Hoc Networks," Int. J. Comput. Networks Appl., vol. 10, no. 4, pp. 668–687, 2023, doi: 10.22247/ijcna/2023/223319.

[41]. M. P. Swapna and J. Ramkumar, "Multiple Memory Image Instances Stratagem to Detect Fileless Malware," in Communications in Computer and Information Science, S. Rajagopal, K. Popat, D. Meva, and S. Bajeja, Eds., Springer Science and Business Media Deutschland GmbH, 2024, pp. 131–140. doi: 10.1007/978-3-031-59100-6_11.

[42]. M. Lingaraj, T. N. Sugumar, C. S. Felix, and J. Ramkumar, "Query aware routing protocol for mobility enabled wireless sensor network," Int. J. Comput. Networks Appl., vol. 8, no. 3, pp. 258–267, 2021, doi: 10.22247/ijcna/2021/209192.

[43]. S. P. Geetha, N. M. S. Sundari, J. Ramkumar, and R. Karthikeyan, "Energy Efficient Routing In Quantum Flying Ad Hoc Network (Q-FANET) Using Mamdani

Fuzzy Inference Enhanced Dijkstra's Algorithm (MFI-EDA)," J. Theor. Appl. Inf. Technol., vol. 102, no. 9, pp. 3708–3724, 2024.

[44]. R. Jaganathan, S. Mehta, and R. Krishan, Bio-Inspired intelligence for smart decision-making. IGI Global, 2024.

[45]. R. Jaganathan and V. Ramasamy, "Performance modeling of bio-inspired routing protocols in Cognitive Radio Ad Hoc Network to reduce end-to-end delay," Int. J. Intell. Eng. Syst., vol. 12, no. 1, pp. 221–231, 2019.

[46]. [34]. J. Ramkumar, R. Vadivel, and B. Narasimhan, "Constrained Cuckoo Search Optimization Based Protocol for Routing in Cloud Network," Int. J. Comput. Networks Appl., vol. 8, no. 6, pp. 795–803, 2021.

[47]. J. Ramkumar, R. Karthikeyan, and M. Lingaraj, "Optimizing IoT-Based Quantum Wireless Sensor Networks Using NM-TEEN Fusion of Energy Efficiency and Systematic Governance," in Lecture Notes in Electrical Engineering, V. Shrivastava, J. C. Bansal, and B. K. Panigrahi, Eds., Springer Science and Business Media Deutschland GmbH, 2025, pp. 141–153.

[48]. B. Suchitra, R. Karthikeyan, J. Ramkumar, and V. Valarmathi, "Enhancing Recurrent Neural Network Performance for Latent Autoimmune Diabetes Detection (Lada) Using Exocoetidae Optimization," J. Theor. Appl. Inf. Technol., vol. 103, no. 5, pp. 1645–1667, 2025.

[49]. J. Ramkumar and D. Ravindran, "Machine learning and robotics in urban traffic flow optimization with graph neural networks and reinforcement learning," in Machine Learning and Robotics in Urban Planning and Management, 2025, pp. 83–104.

[50]. R. Jaganathan, K. Rajendran, and P. S. Ponnukumar, "Peregrine Falcon Optimization Routing Protocol (PFORP) for Achieving Ultra-Low Latency and Boosted Efficiency in 6G Drone Ad-Hoc Networks (DANET)," Int. J. Comput. Digit. Syst., vol. 17, no. 1, pp. 1–18, 2025.

[51]. S. P. Priyadharshini, F. Nirmala Irudayam, and J. Ramkumar, "An Unique Overture of Plithogenic Cubic Overset, Underset and Offset," in Studies in Fuzziness and Soft Computing, vol. 435, 2025, pp. 139–156.

[52]. J. Ramkumar, B. Varun, V. Valarmathi, D. R. Medhunhashini, and R. Karthikeyan, "Jaguar-Based Routing Protocol (JRP) For Improved Reliability And Reduced Packet Loss In Drone Ad-Hoc Networks (DANET)," J. Theor. Appl. Inf. Technol., vol. 103, no. 2, pp. 696–713, 2025.

[53]. P. S. Ponnukumar, N. I. Francis Xavier, and R. Jaganathan, "Stable Plithogenic Cubic Sets," J. Fuzzy Ext. Appl., vol. 6, no. 2, pp. 410–423, 2025.

[54]. J. Ramkumar and V. Valarmathi, "Harnessing AI-Driven Models for Sustainable Development in Business Management," in World Sustainability Series, vol. Part F775, 2025, pp. 217–238.

[55]. S. P. Priyadharshini and J. Ramkumar, "Mappings Of Plithogenic Cubic Sets," Neutrosophic Sets Syst., vol. 79, pp. 669–685, 2025.

[56]. A. Senthilkumar, J. Ramkumar, M. Lingaraj, D. Jayaraj, and B. Sureshkumar, "Minimizing Energy Consumption in Vehicular Sensor Networks Using Relentless Particle Swarm Optimization Routing," Int. J. Comput. Networks Appl., vol. 10, no. 2, pp. 217–230, 2023.

[57]. J. Ramkumar, R. Karthikeyan, and K. O. Nitish, "Securing Library Data With Blockchain Advantage," in Enhancing Security and Regulations in Libraries with Blockchain Technology, 2024, pp. 117–138.

[58]. V. Valarmathi and J. Ramkumar, "Modernizing Wildfire Management Through Deep Learning and IoT in Fire Ecology," in Machine Learning and Internet of Things in Fire Ecology, 2024, pp. 203–229.

[59]. B. Suchitra, J. Ramkumar, and R. Karthikeyan, "Frog Leap Inspired Optimization-Based Extreme Learning Machine For Accurate Classification Of Latent Autoimmune Diabetes In Adults (LADA)," J. Theor. Appl. Inf. Technol., vol. 103, no. 2, pp. 472–494, 2025.

[60]. J. Ramkumar, V. Valarmathi, and R. Karthikeyan, "Optimizing Quality of Service and Energy Efficiency in Hazardous Drone Ad-Hoc Networks (DANET) Using Kingfisher Routing Protocol (KRP)," Int. J. Eng. Trends Technol., vol. 73, no. 1, pp. 410–430, 2025.

[61]. R. Jaganathan, S. Rajagopal, and K. Rajendran, "Cultural Intelligence in the AI Era-Enhancing Transitional Higher Education," in Bridging Global Divides for

Transnational Higher Education in the AI Era, 2024, pp. 273–292.

[62]. J. Ramkumar, A. Senthilkumar, M. Lingaraj, R. Karthikeyan, and L. Santhi, "Optimal Approach for Minimizing Delays in Iot-Based Quantum Wireless Sensor Networks Using Nm-Leach Routing Protocol," J. Theor. Appl. Inf. Technol., vol. 102, no. 3, pp. 1099–1111, 2024.

[63]. D. Jayaraj, J. Ramkumar, M. Lingaraj, and B. Sureshkumar, "AFSORP: Adaptive Fish Swarm Optimization-Based Routing Protocol for Mobility Enabled Wireless Sensor Network," Int. J. Comput. Networks Appl., vol. 10, no. 1, pp. 119–129, Jan. 2023.

[64]. J. Ramkumar, S. S. Dinakaran, M. Lingaraj, S. Boopalan, and B. Narasimhan, "IoT-Based Kalman Filtering and Particle Swarm Optimization for Detecting Skin Lesion," in Lecture Notes in Electrical Engineering, K. Murari, N. Prasad Padhy, and S. Kamalasadan, Eds., Singapore: Springer Nature Singapore, 2023, pp. 17–27.

[65]. J. Ramkumar, C. Kumuthini, B. Narasimhan, and S. Boopalan, "Energy Consumption Minimization in Cognitive Radio Mobile Ad-Hoc Networks using Enriched Ad-hoc On-demand Distance Vector Protocol," in 2022 International Conference on Advanced Computing Technologies and Applications, ICACTA 2022, Institute of Electrical and Electronics Engineers Inc., 2022.

[66]. P. Menakadevi and J. Ramkumar, "Robust Optimization Based Extreme Learning Machine for Sentiment Analysis in Big Data," in 2022 International Conference on Advanced Computing Technologies and Applications, ICACTA 2022, Institute of Electrical and Electronics Engineers Inc., 2022.

[67]. M. P. Swapna, J. Ramkumar, and R. Karthikeyan, "Energy-Aware Reliable Routing with Blockchain Security for Heterogeneous Wireless Sensor Networks," in Lecture Notes in Networks and Systems, V. Goar, M. Kuri, R. Kumar, and T. Senjyu, Eds., Springer Science and Business Media Deutschland GmbH, 2025, pp. 713–723.

[68]. J. Ramkumar and R. Vadivel, "Multi-Adaptive Routing Protocol for Internet of Things based Ad-hoc Networks," Wirel. Pers. Commun., vol. 120, no. 2, pp. 887–909, 2021.

[69]. R. Jaganathan and R. Vadivel, "Intelligent Fish Swarm Inspired Protocol (IFSIP) for Dynamic Ideal Routing in Cognitive Radio Ad-Hoc Networks," Int. J. Comput. Digit. Syst., vol. 10, no. 1, pp. 1063–1074, 2021.

[70]. R. Vadivel and J. Ramkumar, "QoS-enabled improved cuckoo search-inspired protocol (ICSIP) for IoT-based healthcare applications," Inc. Internet Things Healthc. Appl. Wearable Devices, pp. 109–121, 2019.

[71]. J. Ramkumar, R. Karthikeyan, and V. Valarmathi, "Alpine Swift Routing Protocol (ASRP) for Strategic Adaptive Connectivity Enhancement and Boosted Quality of Service in Drone Ad Hoc Network (DANET)," Int. J. Comput. Networks Appl., vol. 11, no. 5, pp. 726–748, 2024.

[72]. J. Ramkumar and R. Vadivel, "Whale optimization routing protocol for minimizing energy consumption in cognitive radio wireless sensor network," Int. J. Comput. Networks Appl., vol. 8, no. 4, pp. 455–464, 2021.

[73]. M. P. Swapna, D. Rajeev, J. Ramkumar, and S. Chandran, "Unveiling Cybercrime Patterns in Kerala: A Machine Learning Approach," in Lecture Notes in Networks and Systems, 2026, pp. 111–122.

[74]. R. Jaganathan, S. Mehta, and R. Krishan, "Preface," Bio-Inspired Intell. Smart Decis., pp. xix–xx, 2024.

[75]. R. Jaganathan, S. Mehta, and R. Krishan, "Preface," Intell. Decis. Mak. Through Bio-Inspired Optim., pp. xiii–xvi, 2024.

[76]. D.Shobana, V Vinodhini "Enhancing Gestational Diabetes Mellitus Prediction using White Tiger Swarm Optimization - Enhanced Multilayer Perceptron (WTSO-MLP)" J. Theor. Appl. Inf. Technol., vol. 103, no.19, pp. 8022–8044, 2025.