Biomedical Signal Processing and Control 110 (2025) 108160

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Knowledge-aware Attentional Neural Network based healthcare big data analytics optimized with Weighted Velocity-Guided Grey Wolf Optimization Algorithm

N. Vasuki [a,*], C. Anand [b], P. Sukumar [c], V.Suresh Babu [d]

[a] Department of Computer Science and Engineering, Government College of Engineering, Erode, Tamil Nadu, India
[b] Department of Information Technology, Nandha College of Technology, Erode, Tamil Nadu, India
[c] Department of Computer Science and Engineering, Nandha Engineering College, Erode, Tamil Nadu, India
[d] Department of Computer Science and Engineering, Sri Shanmuga College of Engineering and Technology, Sankari, Tamil Nadu, India

ARTICLE INFO

Keywords:
Healthcare big data
Improved Bald Eagle Search Optimization Algorithm
Multiparticle Kalman filter
Knowledge-aware Attentional Neural Network
Weighted Velocity-Guided Grey Wolf Optimization Algorithm

ABSTRACT

A significant increase in data volumes, along with the attractive opportunities and potential arising from data analysis contributes to the idea of Big Data. The existing healthcare big data analytics methods face challenges in handling high-dimensional data, slow convergence and suboptimal feature selection. In this paper, a Knowledge-aware Attentional Neural Network based Healthcare Big Data Analytics optimized with Weighted Velocity-Guided Grey Wolf Optimization Algorithm (KANN-HBA-WVGGWOA) is proposed. Here, the input data are taken from PIMA Indians Diabetes dataset. Then the input data is pre-processed by utilizing Multiparticle Kalman filter (MKF) to calculate every data object value primarily. The feature selection utilizing Improved Bald Eagle Search Optimization Algorithm (IBESOA) to select the optimal features from the dataset. The selected features are given into Knowledge-aware Attentional Neural Network (KANN) to classify the data as diabetes and no diabetes. Finally, Weighted Velocity-Guided Grey Wolf Optimization Algorithm (WVGGWOA) is proposed to optimize the KANN classifier that precisely classifies the diabetes disease. The KANN-HBA-WVGGWOA method is implemented in Python. The proposed KANN-HBA-WVGGWOA method attains 1.28%, 2.22%, and 2.27% higher accuracy; 12.56%, 18.68%, and 19.49% less computational time compared to the existing models: Role of big data analytics for revolutionizing diabetes management including health care decision-making (BDA-LR-RDMH), Map reduce dependent big data framework utilizing associative kruskal poly kernel classifier for diabetic disorder prediction (BDF-MRPK-DDP) and the Implementation of ML approaches with big data along IoT to generate effectual prediction for health informatics (BD-KNN-PHI) respectively.

1. Introduction

Nowadays, Big data is widely used in the medical industry and the corresponding datasets are growing faster, bigger and more difficult for medical professionals to process and comprehend with traditional methods [1,2]. The major aim is to increasing the healthcare efficacy [3,4]. However, prevention, early identification and improved disease control are difficult to achieve due to rapid population growth, an upside-down age pyramid, and a paradigm shift in medical care services [4,5]. Therefore, size of a single dataset does not indicate quality of big data. In addition, speed, heterogeneity, and various forms of data all have an impact on quality in the healthcare field [6,7]. By using the adaptability, diversity, connectedness of data holding technologies, the supplement data is acquired fast rate, decision support available in real-time to preserve growth of novel approaches [8,9]. For example, managing influenza pandemic, diverse data from regulated with random sources analysed, mined, converted into attractive actions for risk management. Data heterogeneity develops in medical care sector owing to integration of several biomedical data sources [10,11]. Dual sorts of data sources: quantitative (gene arrays, sensor data, pictures, lab samples), qualitative (diabetes text, demographics) [12,13]. Main issues in data management is maintaining fundamentals of observational facts when respond to medical care concerns. It addressed by many research involving randomized trials [14,15]. Furthermore, problem of outcome

* Corresponding author.
E-mail addresses: adithyavasuki2@gmail.com (N. Vasuki), canand02@gmail.com (C. Anand), sukumarwin@gmail.com (P. Sukumar), sureshbabu@shanmugha.edu.in (V.Suresh Babu).

https://doi.org/10.1016/j.bspc.2025.108160
Received 17 July 2024; Received in revised form 15 May 2025; Accepted 24 May 2025
Available online 20 June 2025

diabetesization is dependent on the small range of possibilities. This could be rectified by exploiting the merits of big data analytics that have ability to deploy longitudinal research [16]. HBA is a type of big data that significant not just due to huge volume, magnitude, because of complexity, timeliness [17,18]. Medical data includes whole information on a patient, including electronic health records, diagnostic reports, doctor's prescriptions, pharmacy texts, clinical photos and study data from medical journals [19,20]. In current years, critical to digitize data generated by healthcare institutions to improve treatment superiority, undertake primary illness analysis to minimize risk factors, optimally arrange data in hospitals [14]. Descriptive insights emerge from the thorough analysis of clinical data that can improve medical systems and identify performance through sensible care delivery [21,22].

The healthcare sector has various challenges, including manage disease outbreaks and maintaining operational efficacy. Data mining with analytics have developing healthcare applications, but success based upon access to high-quality data. Data analytics cannot be applied to every situation in the same way. As data volumes grow, traditional systems such as relational databases struggle to process and analyse it, posing both advantages and challenges. Big data problems include capturing, storing, sharing and analysing data, ensuring privacy when querying information. These are motivated to do this work. The goal is to analyse the healthcare big data using KANN-HBA-WVGGWOA.

The important contributions of this research are:

- Knowledge-aware Attentional Neural Network based Healthcare Big Data Analytics optimized with Weighted Velocity-Guided Grey Wolf Optimization Algorithm (KANN-HBA-WVGGWOA) is proposed for healthcare classification of diabetes disease.
- To select input data from PIMA Indians Diabetes dataset collection. Using Multi-particle Kalman filter (MKF) method to calculate every data object value primarily. As for better classification, the optimal features are selected under IBESOA method from dataset.
- Reducing training time by classifying only the selected features from dataset using KANN method.
- Improving the classification performance by optimizing the KANN using WVGGWOA.
- The efficiency of the KANN-HBA-WVGGWOA is analysed under some metrics. The performance is compared with existing BDA-LR-RDMH, BDF-MRPK-DDP and BD-KNN-PHI models.

Continuing paper is structured as: section 2 reveals the recent related works, section 3 describes the proposed method, section 4 shows results, section 5 provides conclusion.

2. Literature review

Amongst several research works on DL depend healthcare big data analytics; a few recent works are reviewed here,

Nauman et al., [23] suggested a part of big data analytics in developing diabetes managing together with health care decision-making. Machine Learning (ML) was applied in big data contexts, specially using MLib package in Apache Spark for extracting insights via diabetic health care database. The benefits of incorporating BDA and ML in Big Data for medical decision-making were demonstrated by Behavioural Risk Factor Surveillance System (BRFSS) of the CDC. Logistic Regression (LR) performed better than Naive Bayes (NB), providing insightful information for applications in healthcare. It has greater accuracy and lower precision.

Ramani et al., [24] suggested the map reduce dependent big data structure based upon associative kruskal poly kernel classifier for predicting diabetic disorder. Associative Kruskal Wallis with Map Reduce Poly Kernel (AKW-MRPK) were suggested for initial sickness detection. The relevant characteristics were determined using AKW for feature selection. Based on the important features achieved, it parallelizes polynomial kernel vectors using Map Reduce, resulting in a major computing model to aid in disease early detection. AKW-MRPK attained higher accuracy and less computational time. It has greater recall and lesser f1-score.

Zamani et al., [25] suggested the execution of ML approaches with big data and IoT to develop effectual estimate for health informatics. For processing massive data, the suggetsed system makes use of Map Reduce, where feature extraction was made possible by the map phase and feature selection was made possible by the decrease phase. The necessary medical information was gathered from outside websites. Principal Component Analysis (PCA) and statistical features were extracted. Hybrid Flower Pollination Bumblebees Optimisation Algorithm (HFPBOA) was used to choose the better features while the reduction phase. Ensemble Learning was considered to detect the multiple diseases. The weight function of NN, KNN, fuzzy classifier outputs was averaged to get the final prediction result. It provides high Matthews Correlation Coefficient and low precision.

Hani and Ahmad, [26] presented the Big Data to Predict Young Adult Ischemic vs. Non-Ischemic Heart Disorder Risk Factors: AI dependent Method. Intelligence-dependent Medicine. The artificial intelligence technology was effective in calculating silent trends in data. Machine-learning methods require substantial data to accurately forecast IHD vs. non-IHD. CHAID with high accurateness and area under the curve was a trustworthy approach for determining the link between attributes. When compared to non-IHD, sex was an important factor in identifying IHD. It provides high Cohen's Kappa Score and low accuracy.

Mansour et al., [27] presented the brain intracranial hemorrhage e-diagnosis depending on artificial intelligence along big data analytics utilizing CT imageries. IoMT devices were employed for data collection. Graph cut-base segmentation was used by the AIBDA-ICH to recognize damaged regions in CT scans. To maintain massive data, Hadoop Ecosystem along its components were applied. The capsule network (CapsNet) model was used to extract relevant feature vectors. AIBDA-ICH uses the fuzzy deep neural network for categorization. It has higher f1-score and higher computational time.

Samiei et al., [28] suggested the categorization of skin cancer stages depending on AHP fuzzy strategy in big data healthcare. The big data technology improves research quality and speed, leading to better outcomes. It essay organizes the skin cancer stages treatment with respect to the relevant data. To evaluate the effectiveness of SVM multiple class categorization, fuzzy selector, and radial basis function-dependent binary migration classifiers for virtual machines. The links have been categorized. The investigations determined if the tumours were malignant or benign, and the severity of the malignancy. The data set utilized for processing consists of images of skin spots collected from laboratory images. It has high recall and high MSE.

Jaiswal et al., [29] presented the breast cancer risk predication with categorization using ensemble learning including big data fusion. Big data analytics combined with an enhanced XGBoost Ensembling for the diagnosis of breast cancer cells. To attain high accuracy, the identification processes were feature extraction, target role, and data preprocessing. Wisconsin breast cancer diagnostic data was employed to testing. Decision Tree, Random Forest, Naive Bayes, KNN, SVM, Adaboost, XGBoost were compared. It has high Mathews correlation coefficient and low Cohen's kappa score.

Existing illness prediction methods based upon big data analytics and machine learning face several limitations. Nauman et al., [23] achieved high accuracy in diabetes management but lacked precision. Ramani et al., [24] reduced computational time but lacked generalizability. Zamani et al., [25] integrated IoT for multi-disease prediction but suffered from low precision. Hani et al., [26] required large datasets, increasing computational costs. Mansour et al., [27] proposed AI-based haemorrhage detection but faced high computational time. Samiei et al., [28] exhibited high MSE in skin cancer classification, while Jaiswal et al., [29] had low Cohen's Kappa Score in breast cancer classification. In contrast, the proposed KANN-HBA-WVGGWOA method overcomes these shortcomings by integrating an optimized feature selection

mechanism using the IBESOA to enhance relevant feature extraction while reducing computational overhead. KANN ensures robust classification with improved generalization across datasets, while the WVGGWOA fine-tunes the classifier to enhance accuracy, precision and recall. By addressing the inefficiencies of existing approaches, the proposed method significantly improves prediction reliability and computational efficiency.

3. Proposed methodology

The KANN depend HBA for classification of diabetes disease optimized with WVGGWOA (KANN-HBA-WVGGWOA) are deliberated.

Fig. 1 portrays the block diagram of KANN-HBA-WVGGWOA system. This figure comprises pre-processing, feature selection, classification. The input data is collected through PIMA Indians diabetes database. The gathered input data is given to the MKF for the pre-processing stage. From pre-processed data, the optimum features are chosen under IBESOA. Using KANN, the selected features are classified. WVGGWOA optimizes KANN classifier to accurately classify input data.

3.1. Data Acquisition

The input data are gathered from PIMA Indians diabetes database [30]. This database originated from National Institute of Diabetes and Digestive and Kidney Diseases. Certain requirements are established for identifying these occurrences from a large database. The dataset consists of 768 samples, each signifying a female patient of around 21 years age from Pima Indian population. The dataset includes 8 aspects related to medical and physiological factors, along with a binary outcome representing regarding if the patient has diabetes or not. From those, 70 % and 15 % were selected randomly to be used in the training as well as testing set and the remaining 15 % became the validation set. The dataset was chosen due to its widespread use as a benchmark in diabetes prediction research, enabling fair comparisons with existing methods. It contains real patient data, making it highly relevant for developing predictive healthcare models. Its public availability further facilitates reproducibility and research advancements. Table 1 tabulates the features of PIMA Indians Diabetes dataset.

3.2. Pre-processing under Multiparticle Kalman filter

The pre-processing under Multiparticle Kalman filter (MKF) [31] is discussed here. MKF act as pre-processor to calculate every data object

Table 1
Features in PIMA Indians diabetes dataset.

SI. No	Name of Features	Description
1	Pregnancies	No. of times pregnant
2	Glucose	During two-hour oral glucose tolerance test, plasma glucose concentration
3	Blood Pressure	Diastolic blood pressure (mmHg)
4	Skin Thickness	Triceps skin fold thickness (mm)
5	Insulin	2-Hour serum insulin (mu U/ml)
6	BMI	Body mass index (weight in kg/(height in m)^2)
7	Diabetes Pedigree Function	Diabetes pedigree function
8	Age	Age (year)

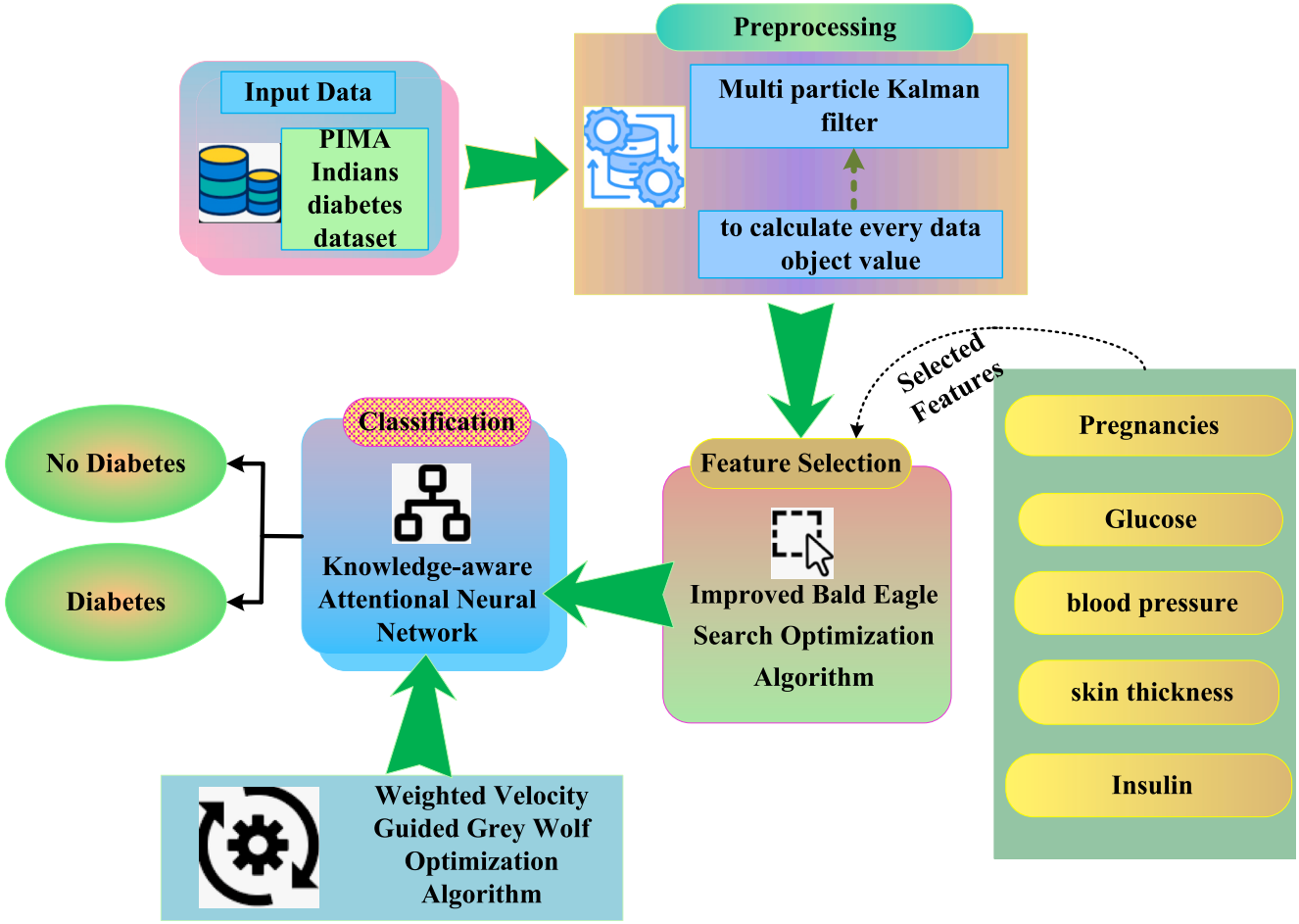

Fig. 1. Block diagram of KANN-HBA-WVGGWOA method.

value primarily. MKF enables more precise estimation and prediction of patient health statuses by combining numerous data sources at the same time. This may result in more precise diagnosis and treatment schedules. Kalman is an intuitive combination of particle filters. This combination aims to create an algorithm that handles unknown initial states similar to the particle filter, also quickens convergence to optimum states, similar to Kalman filter. Because symmetric environs are particularly difficult particle filter, demonstrate the accelerated convergence of the proposed approach is shown in equation (1)

$$\Gamma(K_{s+1}\left|Y_{s+1}^{j+}\right.)=\prod_{i=1}^{L}q(K_{s+1}^{i}\left|Y_{s+1}^{j+}\right.) \tag{1}$$

where K_{s+1}, Y_{s+1}^{j+} denotes the data object, K_{s+1}^{i}, Y_{s+1}^{j+} denotes sample data in i^{th} *and* j^{th} junction. Extended Kalman filter for each generated data in parallel. To compute the updated value of every data object ω_{s+1}^{j}. After that, update weights of then data. Then weights are updated similar way in particle filter. It is shown in equation (2).

$$\omega_{s+1}^{j}\approx\Gamma(K_{s+1}\left|Y_{s+1}^{j+}\right.)\omega_{s}^{j} \tag{2}$$

where ω_{s+1}^{j} denotes data value. A resampling obtained particle states. There modify to resample not only particle Y_{s+1}^{j+} t corresponding data values. Modified resampling process utilized in proposed technique is abridged. It is given in equation (3).

$$\omega_{s+1}^{j}=\frac{1}{N} \tag{3}$$

where N denotes number of data located at j^{th} section. It calculated precisely with low error. The values in the data is analysed and calculate the data values. The value of data that is calculated is shown in equation (4).

$$Y_{s+1}^{j}=g(y_{s=1}^{j+}),0,\eta^{j} \tag{4}$$

where η^{j} denotes the data sample variable,$y_{s=1}^{j+}$ denotes data value calculation,g denotes the random variable. The value of data is calculated successfully by MKF model. Then, the preprocessed data is given into the IBESOA for feature selection.

3.3. Feature selection utilizing improved bald eagle search Optimization Algorithm

This section discusses feature selection under IBESOA [32]. IBESOA model selects the optimal features like pregnancies, glucose, blood pressure, skin thickness and insulin from the input dataset. IBESOA gives better feature selecting capabilities. It successfully extracts the most significant features from big datasets, resulting in better model performance and interpretability. The IBESOA method is an improved version of BESOA. The IBESOA algorithm is revolutionary nature-inspired optimization process simulates intellectual hunting behaviour of bald eagles. BESOA 's hunting procedure has three essential steps likes eagle selects the search region with highest prey density, searches for prey inside it, then attacks the prey using the ideal attack point depend search findings. Once best point of attack is established, each subsequent move focused on location.

3.3.1. Stepwise process of improved bald eagle search Optimization Algorithm

The stepwise procedure is demarcated to get the optimal value of KANN under IBESOA. The IBESOA selecting the features from the dataset.

Step 1: Initialization

During their initialisation phase, bald eagles choose the best hunting locations with lots of prey, it is given in Eq. (5)

$$Q_{new,j}=Q_{best}+s^{*}\alpha(Q_{mean}-Q_{j}) \tag{5}$$

where s signifies random count with 0 to 1 value, α regulates the variations in position, Q_{best} represents best search area that is preferred by bald eagles,Q_{mean} displays that these eagles have consume all knowledge from the earlier point and Q_j represents the area selected for search with bald eagles

Step 2: Random generation

Generate randomly the weight parameters. The optimum fitness values are preferred under obvious hyperparameter condition.

Step 3: Fitness Function

Starlings' initial positional vectors are represented by the solution candidate matrix, which is identified. The fitness function chooses the best features once this matrix is initially allocated to random values in the searchspace is labelled in Eq. (6).

$$Fitness\,function\;=\;[\,Selecting\;\;Optimal\;\;Features\,] \tag{6}$$

Step 4: Exploration phase

Bald eagles search for prey inside a specific search zone using a spiral pattern. Eagles move randomly and are governed by the following Eq. (7).

$$Q_{j,new}=Q_{j}+x(j)*(Q_{i}-Q_{i+1})+y(j)^{*}(Q_{i}-Q_{mean}) \tag{7}$$

where $x(j)$,$y(j)$ shows the search space $Q_{j,new}$ signify the new position of the bald eagle and $Q_i - Q_{i+1}$ denotes location of prey in search space. The search cycles done in the search space is shown in Eq. (8).

$$\theta(j)=b^{*}\Pi^{*}ran\;s(j)=\theta(j)+S^{*}ran \tag{8}$$

where $ran(.)$ denotes the random operation, the variables S and θ represent search cycles, angle among search locations in central point,S runs from 0.5 to 2, and θ ranges from 5 to 10. Bald eagles move about search space in different directions to find the best attack point.

Step 5: Exploitation phase.

Bald eagles focus on attacking prey from the most advantageous location inside the search area during the attack phase, even while other points move in its direction. This procedure can be described numerically in Eq. (9)

$$Q_{j,new}=ran^{*}Q_{best}+x_{1}(j)*(Q_{i}-d_{1}{}^{*}Q_{mean})+y_{1}(j)^{*}(Q_{i}-d_{2}Q_{best}) \tag{9}$$

where d_1 and d_2 denotes the external sources.Q_{best} denotes best search area selected with bald eagles depend on best location was detected through previous search,$d_2, d_1 \in [1,2]$, Q_{mean} displays that such eagles have consumed each knowledge from earlier points,s denotes random count, it has 0–1 value and is to define angle among search for points at central point, then the value lies among 5 and 10.

Step 6: Termination

The optimum features are designated under IBESOA, else step 3 is repeated iteratively until fulfil the halting criterion $Q = Q + 1$. The selected features are feed as input into the diabetes disease classification. Amongst 8 features of PIMA Indians Diabetes database, 5 features are selected by IBESOA. Table 2 depicts the list of selected features using IBESOA.

Table 2
Selected features utilizing IBESOA.

SI. No	Features
1	pregnancy
2	glucose
3	blood pressure
4	skin thickness
5	Insulin

3.4. Classification using Knowledge-aware Attentional Neural Network

The classification using KANN [33] is discussed. KANN allows model to effectively focus on relevant pieces of information inside healthcare data, resulting in more accurate disease diagnoses. The term knowledge-aware in the KANN refers to the model's ability to incorporate domain-specific knowledge into its learning process. Unlike traditional deep learning models that rely purely on data-driven learning, a knowledge-aware model integrates external knowledge sources, structured relationships, and contextual information to improve decision-making and classification performance. The incorporated knowledge includes medical domain knowledge, feature relevance, and historical data patterns to enhance classification accuracy. Medical domain knowledge is derived from clinical studies, prior research on diabetes classification, and expert insights related to healthcare big data, ensuring that the model aligns with established medical understanding. Feature relevance and importance play a crucial role by identifying key medical attributes, like blood glucose levels, BMI, insulin levels, allowing the model to prioritize meaningful features that significantly impact diabetes prediction. The KANN employs an attention mechanism to focus on the most relevant features based on medical domain knowledge, ensuring that key diagnostic factors are given more importance during classification. This model aims to classify health care data based on disease diagnosis and provide a reasonable explanation by modelling the user-data interaction using low-order knowledge features. The architecture diagram of KANN is given in Fig. 2.

The linear projection of user's knowledge embedding features i into H knowledge semantic spaces, it is given in equation (10)

$$V_{jth} = V_j X_q \tag{10}$$

where V_{jth} denotes data of user j in space q and X_q denotes big data samples. KANN is used in this work since it is a deep structure, significant knowledge network. Using huge hidden layers and extra layers, data samples were created from input to output layer. It is shown in Eq. (11),12)

$$V_{th}^{inner} = sof\max\left(\frac{P_{ith}L_{iyh}^{S}}{\sqrt{e_m}}\right)V_{ith} \tag{11}$$

$$V_{th}^{outer} = G_{outer}(N_j) \tag{12}$$

where V_{th}^{inner} ,V_{th}^{outer} denotes the inner and outer layer of the KANN.$P_{ith}L_{iyh}^{S}$ denotes the input big data samples,e_m denotes the random variable, G_{outer} denotes the data collected in outer layer and N_j signifies number of data samples at j^{th} layer. Along with KANN method, hidden units categorize system by itself, creates activation function. RBM was implemented to address potential activation function generating challenges. One layer of stochastic hidden units and stochastic visible units make up the RBM class of Markov arbitrary is given in equation (13)

$$v_j = \mathrm{Re}(v_i B_{ji} N_i^{inner} + a_v) \tag{13}$$

where $v_i B_{ji} N_i^{inner}$ denotes the multilayer hidden hints,a_v shows the data samples in hidden layer and Re(.) denotes Markov arbitrary operation. Although the given hidden and visible units are contemporaneous, the hidden unit is better. It intends to route difficult phase. It is shown in equation (14).

$$i = v_j \otimes n_j \tag{14}$$

where $\otimes$ operator denotes data wise product,v_j,n_j denotes health care data of outer attention layer associated with user i,database j. Final classification, decision rating for user i and database j is shown in Eq. (15).

$$H_{sq} = \sum_{j,i}\left(\hat{s}_{ji} - s_{ji}\right)^2 \tag{15}$$

where H_{sq} denotes the classification of selected features,$\left(\hat{s}_{ji} - s_{ji}\right)$ denotes the output values in classification layer. The resulting layer is positioned on top of the DNN to classify data. Three hidden layers and the number of input neurones are used in KANN technique. The weight

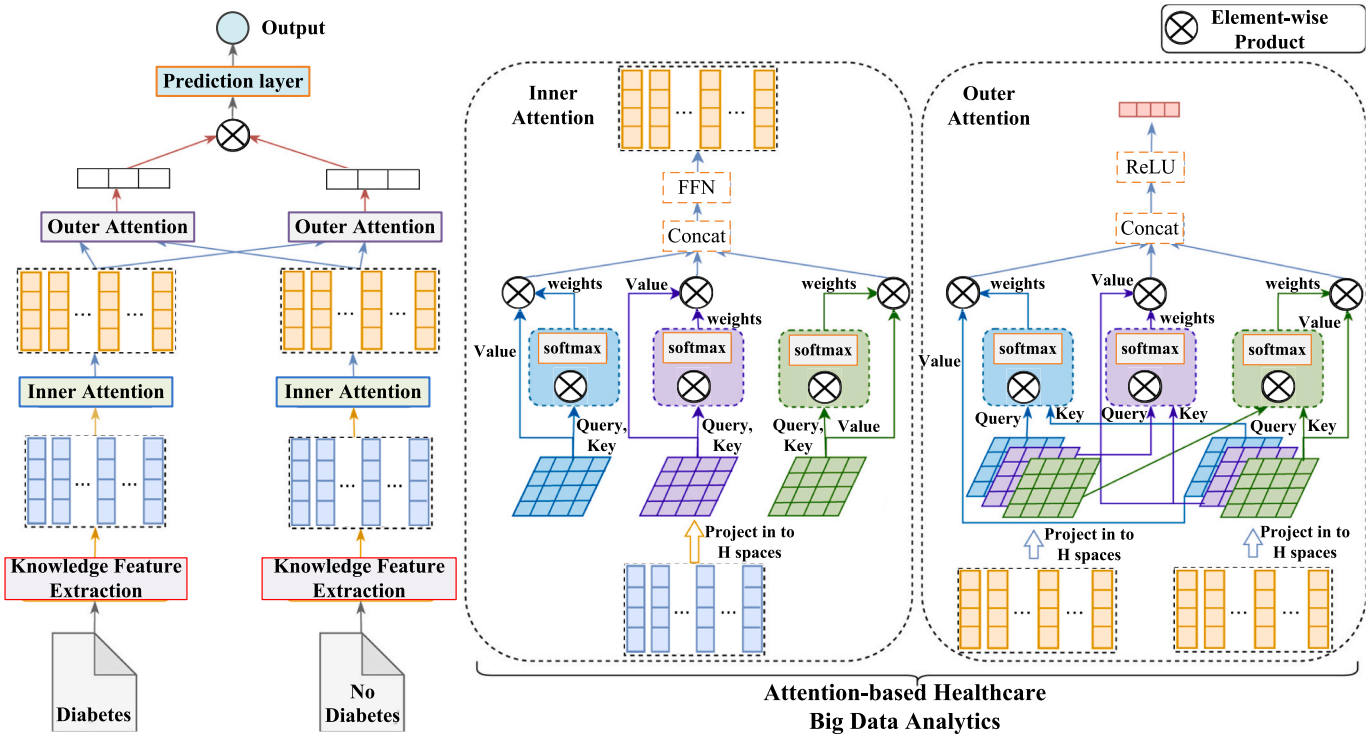

Fig. 2. Architecture of KANN.

enhanced through training phase with the help of training dataset. It is shown in Eq. (16).

$$H_{csq} = -\sum_{j,i} s_{j,i}\log\sigma(\widehat{s}_{ji}) + (1 - s_{ji})\log(1 - \sigma(\widehat{s}_{ji})) \tag{16}$$

where $1 - \sigma(\widehat{s}_{ji})$ denotes the classified data samples for training. The pre-training level weights are used to initialise the BP technique. The lower error value is evaluated while high accuracy of KANN classification is attain utilized enhanced weights. KANN model successfully classified the health care data as diabetes and no diabetes. Here, WVGGWOA is employed to enhance KANN optimum parameter s_{ji} *and* a_v.

3.5. Optimization utilizing Weighted Velocity-Guided Grey Wolf Optimization Algorithm

The weight parameter s_{ji} and a_v of KANN is enhanced using WVGGWOA [34] is discussed here. The weight parameter s_{ji} decrease the computational time, a_v increase the accuracy. The WVGGWOA enhances optimization by improving convergence speed, avoiding local optima, and dynamically adjusting weights for better exploration and exploitation. It balances intensification and diversification, leading to more accurate parameter tuning, reduced computational time, and improved classification performance in complex healthcare big data analytics models. WVGGWOA is preferred for its better balance amid the exploration and exploitation, faster convergence, and dynamic velocity-guided adjustments. It prevents stagnation in local optima, outperforms Particle Swarm Optimization, Genetic Algorithm, Ant Colony Optimization, and enhances feature selection and parameter tuning, ensuring improved classification accuracy and efficiency in healthcare big data analytics. The WVGGWOA imitates the social hierarchy along hunting propensities of grey wolves. WVGGWOA randomly generates a flock of wolves to initialize the optimisation process. In every iteration, alpha, beta, and delta are the three well-suited wolves selected as the leaders. The omega wolves then encircle the strongest wolves in search of the most promising regions in the searchspace. Such wolves serve as search agents searching for solutions. The mathematical mean of the updated positioning of each wolf is lastly chosen as the updated position of the population since it encompasses the three gents in the search space. Due to the involvement of three agents in directing the other agents, this technique improves the algorithm's exploration.

3.5.1. Step-by-step procedure of WVGGWOA

The stepwise process is delineated to obtain the better KANN value depend upon WVGGWOA. First, WVGGWOA creates an equally distributed population to enhance KANN parameters s_{ji} *and* a_v. The optimal solution is provoked by the WVGGWOA.

***Step 1*:** Initialization

The proposed algorithm includes population initialization, population evaluation, changing parameters. The steps of the proposed WVGGWOA is given in Eq. (17).

$$W = \begin{bmatrix} W_{\alpha,i}^{s+1} \\ W_{\beta,i}^{s+1} \\ W_{\delta,i}^{s+1} \end{bmatrix} = L \begin{bmatrix} A_{\alpha,i}^{s} & \cdots & W_{\alpha,i}^{s+1} & \cdots & E_{\alpha,i}^{s+1} \\ \vdots & \ddots & & \cdots & \vdots \\ A_{\beta,i}^{s} & & W_{\beta,i}^{s+1} & & E_{\alpha,i}^{s+1} \\ \vdots & \cdots & & \ddots & \vdots \\ A_{\delta,i}^{s} & \cdots & W_{\delta,i}^{s+1} & \cdots & E_{\alpha,i}^{s+1} \end{bmatrix} \tag{17}$$

where $W_{\delta,i}^{s+1}$,$W_{\beta,i}^{s+1}$ and $W_{\alpha,i}^{s+1}$ denote the search agent's velocity and is computed to improve the position update, $A_{\alpha,i}^{s}$, $A_{\beta,i}^{s}$, and $A_{\delta,i}^{s}$ denote the search agent's acceleration and the term L denotes the random function at i^{th} junction.α, β, δ denotes the best solution of wolf.

***Step 2*:** Random Generation

Generate randomwise the input parameters. Depends on its particular hyper parameter circumstances, the optimum progressive value is chosen.

Step 3: Fitness Function

Generate the random solution through initialization using Eq. (18)

$$Fitness\,function = Optimize\ \ (s_{ji}\ and\ a_v) \tag{18}$$

Step 4: Update the velocity of the leading agent for optimizing s_{ji}

Exploitation phase allows the search agents to move to global solution. The distance between prey and the wolf is calculated. To enable a proper and dependable shift from exploration to exploitation, this is determined by equation (19).

$$L = 0.9 - 0.5 \times s_{ji}\left(\frac{s}{s_{\max}}\right) \tag{19}$$

where L denotes the optimum parameter,s denotes the search agents. GWO's position update equation is based on the average of three grey wolf positions. Fig. 3 shows the flowchart of WVGGWOA for optimizing KANN.

This approach yields satisfactory results for simple problems, but not for big dimensions multimodal scenarios. It is given in Eq. (20).

$$\left.\begin{aligned} Y_{1,i}^{s+1} &= Y_{\alpha,i}^{s} - W_{\alpha,i}^{s+1} \\ Y_{2,i}^{s+1} &= Y_{\beta,i}^{s} - W_{\beta,i}^{s+1} \\ Y_{3,i}^{s+1} &= Y_{\delta,i}^{s} - W_{\delta,i}^{s+1} \end{aligned}\right\} \tag{20}$$

where $Y_{1,i}^{s+1}$, $Y_{2,i}^{s+1}$, $Y_{3,i}^{s+1}$ denotes the three position update agents. This approach produces satisfactory results for simple problems but unsatisfactory results for large dimensional multimodal situations.

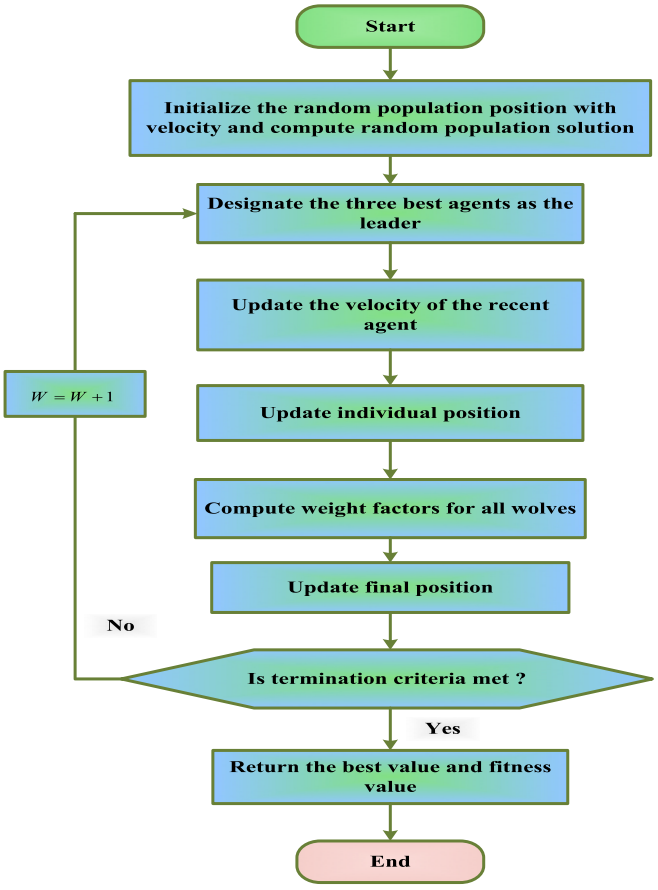

Fig. 3. Flowchart of WVGGWOA for optimizing KANN.

Step 5: Update the individual position for optimizing a_v.

During each algorithm iteration, each updated search agent is compared to its previous optimal location. If the fitness value improves, it is recognized as the new best position. Otherwise, the old best position is used. Iterations improve populations with time. It is shown in Eq. (21).

$$\left.\begin{aligned} s_{ji1} &= A^{s}_{\alpha,i} - D^{s+1}_{\alpha,i} \\ s_{ji2} &= A^{s}_{\beta,i} - D^{s+1}_{\beta,i} \\ s_{ji3} &= A^{s}_{\delta,i} - D^{s+1}_{\delta,i} \end{aligned}\right\} \tag{21}$$

where s_{ji1}, s_{ji2},s_{ji3} denotes the position update equation is weighted in every algorithm iteration. Wolf moves towards the prey based on the three positional information of $D^{s+1}_{\alpha,i}$, $D^{s+1}_{\beta,i}$ and $D^{s+1}_{\delta,i}$.The difficult multi-modal global optimization problem has early convergence and low quality solutions. The weighted distance improves the performance of the GWO. The equations below show that the location update equation is weighted in each algorithm iteration. It is shown in Eq. (22).

$$Y^{s+1}_{j,i} = \frac{a_v s_{ji1} \times \alpha^{s}_{vi} + s_{ji2} \times \alpha^{s}_{vi} + s_{ji3} \times \alpha^{s}_{vi}}{s_{ji1} + s_{ji2} + s_{ji3}} \tag{22}$$

where a^{s}_{vi} the wolf position is updated using the $Y^{s+1}_{j,i}$ vector coefficients. Therefore, the position update equation presented in Eq. (22). This approach is especially valuable in challenging situations where the problem's landscape consists of several steep peaks.

Step 6: Termination

The weight parameter s_{ji} *and* a_v generated from KANN is enhanced by utilizing WVGGWOA; else step 3 will repeat until fulfil the halting criteria $W = W + 1$. Then, WVGGWOA enhances the parameters of KANN. Finally, the KANN classifies the diabetes disease with better accuracy and less computational time.

4. Result and Discussion

The simulation outputs of the KANN-HBA-WVGGWOA are described here. The KANN-HBA-WVGGWOA model is implemented in Python 3.8 using Tensor Flow 2.6, Scikit-learn and NumPy. The experiments are done in a system equipped with Intel Core i9-12900 K processor (3.9 GHz, 16 cores), 32 GB DDR5 RAM, and an NVIDIA RTX 3090 GPU with 24 GB VRAM, running on Ubuntu 20.04. The PIMA Indians Diabetes Dataset was partitioned using stratified sampling to maintain class balance, with 70 % (537 samples) allotted for training, 15 % (115 samples) for validation, 15 % (116 samples) for testing. Table 3 depicts the hyperparameter of the KANN-HBA-WVGGWOA method.

The acquired results of the KANN-HBA-WVGGWOA are compared with the existing BDA-LR-RDMH [23], BDF-MRPK-DDP [24] and BD-KNN-PHI[25] models respectively.

4.1. Performance measures

It is a vital phase for selecting better classifier. The mentioned metrics are used to assess the efficacy of the KANN-HBA-WVGGWOA approach. For that, the following confusion matrix is required.

Table 3
Hyperparameters of the KANN-HBA-WVGGWOA method.

Parameters	Values	Parameters	Values
Number of particles	500	hidden sizes of all layers	1024
Population size	30	number of spaces	4
Mutation rate	0.2	Activation Function	ReLU
Batch size	128	Pack size	20 wolves
Maximum iteration	100	Learning rate	0.001
Count of Hidden Layers	3	Weight decay factor	0.85

- True Positive (TP): positive samples appropriately characterized as positive.
- True Negative (TN): negative samples appropriately characterized as negative.
- False Positive (FP): negative samples imperfectly characterized as positive.
- False Negative (FN): positive samples imperfectly characterized as negative.

4.1.1. Accuracy

It calculates the rate of accurately categorized samples among the overall instances using equation (23),

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{23}$$

4.1.2. Precision

Precision indicates true positive rate events amongst the instances categorized as positive and computed using Eq. (24).

$$precision = \frac{TN}{TN + FP} \tag{24}$$

4.1.3. Recall

It scales true positive rate and original positive instances using Eq. (25),

$$Sensitivity = \frac{TP}{TP + FN} \tag{25}$$

4.1.4. F1-score

This is a harmonic mean for precision along recall, there is only one metric that balances both, and is calculated using Eq. (26)

$$F1 - score = 2 * \frac{Precision * \mathrm{Recall}}{Precision + \mathrm{Recall}} \tag{26}$$

4.1.5. Matthews Correlation Coefficient (MCC)

It measures classification presentation by deeming true positives, true negatives, false positives, and false negatives using Eq. (27)

$$MCC = \frac{(TP + FP)(TP + FN)}{\sqrt{(TN + FP)(TN + FN)(TP \times TN)(FP \times FN)}} \tag{27}$$

4.1.6. Cohen's kappa score

It scales agreement amongst predicted and real labels while considering chance agreement using Eq. (28),

$$\kappa = \frac{S_o - S_e}{1 - S_e} \tag{28}$$

where S_o symbolizes observed agreements and S_e represents expected agreement by chance.

4.1.7. Mean squared error (MSE)

It scales average squared variance of real and estimated values using equation (29),

$$MSE = \frac{1}{P}\sum_{a=1}^{P}(b_a - \widehat{b}_a)^2 \tag{29}$$

where b_a epitomizes actual value and $\widehat{b}_a$ epitomizes predicted value.

4.2. Performance analysis

Figs. 4-8 display the experimental outputs of the KANN-HBA-WVGGWOA. Here, the proficiency of the KANN-HBA-WVGGWOA is compared with the existing BDA-LR-RDMH, BDF-MRPK-DDP and BD-

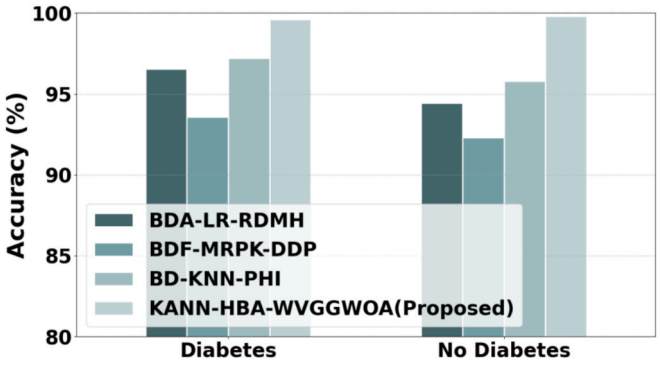

Fig. 4. Performance of accuracy.

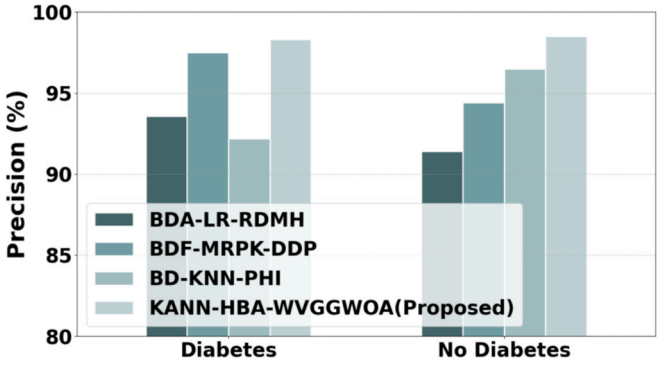

Fig. 5. Precision analysis.

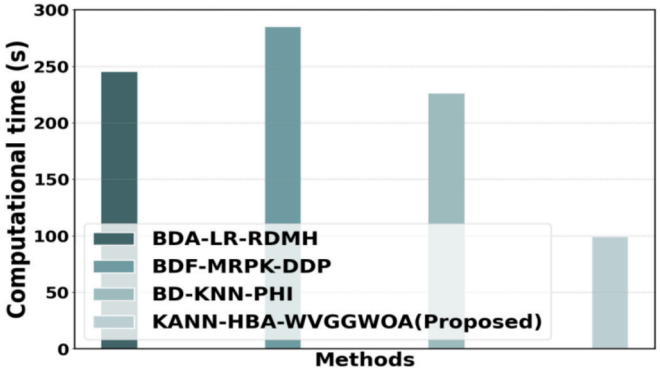

Fig. 6. Analysis of computational time.

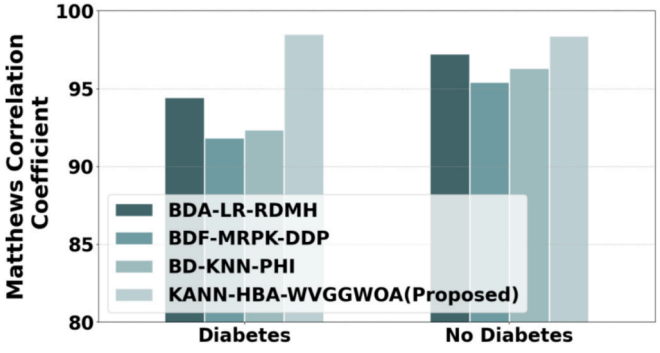

Fig. 7. Performance of Mathews Correlation Coefficient.

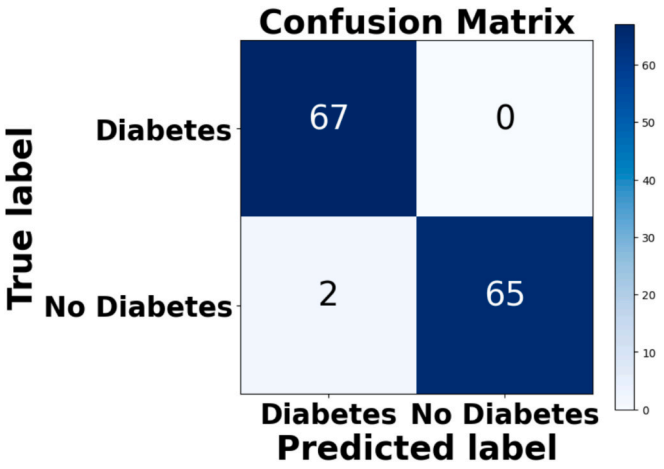

Fig. 8. Confusion matrix.

KNN-PHI techniques.

Fig. 4 portrays accuracy analysis. KANN-HBA-WVGGWOA's high accuracy in diabetes classification can be attributed to KANN, an enhanced attention mechanism, effective WVGGWOA, and strong healthcare big data analytics. These features work together to increase the model's capacity to diagnose diabetes properly. Thus, the proposed KANN-HBA-WVGGWOAattains 1.28 %, 2.22 %, and 2.27 % higher accuracy for diabetes; 2.34 %, 1.31 %, and 1.26 % higher accuracy for no diabetes estimated to the existing techniques, such as BDA-LR-RDMH, BDF-MRPK-DDP and BD-KNN-PHI.

Fig. 5 portrays precision assessment. The KANN-HBA-WVGGWOA model ensures high precision in diabetes classification by leveraging KANN and WVGGWOA. Accuracy alone is insufficient due to class imbalance, while precision minimizes false positives, preventing misdiagnosis. WVGGWOA fine-tunes hyperparameters, enhancing IBE-SOA and ensuring reliable healthcare big data analytics for effective diabetes detection. Thus, the proposed KANN-HBA-WVGGWOAachieves8.26 %, 5.22 %, 2.27 % greater precision for diabetes; 7.26 %, 8.59 %, 6.35 % greater precision for no diabetes than the existing BDA-LR-RDMH, BDF-MRPK-DDP and BD-KNN-PHI models.

Fig. 6 indicates computational time analysis. The KANN-HBA-WVGGWOA achieves low computational time by using attention mechanisms for feature selection, optimizing hyperparameters dynamically, and accelerating convergence. The WVGGWOA minimizes redundant computations, while KANN enhances parallel processing, ensuring efficient big data analytics for diabetes classification with reduced processing overhead and improved model accuracy. Thus, the proposed KANN-HBA-WVGGWOA model attains 12.56 %, 18.68 %, and 19.49 % lower computational time estimated to the current technique such as BDA-LR-RDMH, BDF-MRPK-DDP and BD-KNN-PHI respectively.

Fig. 7 shows performance of MCC. The high MCC of KANN-HBA-WVGGWOA in diabetes classification is due to its knowledge-aware attention mechanism, which prioritizes critical medical features, and WVGGWOA optimization, which fine-tunes model parameters. This ensures balanced predictions, reducing false positives and negatives, making it robust against class imbalance and enhancing classification reliability in healthcare big data analytics. Thus, the proposed KANN-HBA-WVGGWOA model attains 7.55 %, 9.18 %, and 7.73 % higher MCCfor diabetes; 8.16 %, 4.89 %, and 6.35 % higher MCC for no diabetes estimated to the current technique such as BDA-LR-RDMH, BDF-MRPK-DDP and BD-KNN-PHI respectively.

Fig. 8 displays confusion matrix. It shows a count of successfully and erroneously classified instances. Here, 67 instances are accurately identified as diabetes, whereas none are wrongly predicted as no diabetes and 65 instances are accurately predicted as no diabetes, with just two wrongly predicted as diabetes. The model has excellent accuracy

and relatively few misclassifications, demonstrating strong performance in identifying diabetes from non-diabetic cases. The color gradient represents the number of incidents in each category.

Table 4 tabulates the recall and F1-score analysis. KANN-HBA-WVGGWOA obtains high recall in diabetes classification by combining knowledge-aware attention with WVGGWOA optimization and excellent feature extraction. It reduces false negatives, handles class imbalance, and processes multi-modal healthcare data, resulting in strong generalization and accurate diabetic case detection in big data analytics. Thus, the proposed KANN-HBA-WVGGWOA model attains 9.76 %, 8.52 % and 4.59 higher recall for diabetes; 9.46 %, 4.69 and 7.17 % higher recall for no diabetes estimated to the existing BDA-LR-RDMH, BDF-MRPK-DDP and BD-KNN-PHI respectively.

The high F1-score in KANN-HBA-WVGGWOA for diabetes classification stems from its KANN, and WVGGWOA-based hyperparameter tuning. By reducing false positives and negatives, balancing class distribution, and handling large healthcare datasets efficiently, it ensures high precision and recall, leading to superior classification performance. Thus, the proposed KANN-HBA-WVGGWOA model attains 9.18 %, 4.56 %, and 7.56 % higher F1-scorefor diabetes; 4.56 %, 8.19 %, and 5.18 % higher F1-score for no diabetes estimated to the existing BDA-LR-RDMH, BDF-MRPK-DDP and BD-KNN-PHI respectively.

Table 5 displays the mean squared error together with Cohen's kappa score analysis. The low MSE in KANN-HBA-WVGGWOA for diabetes classification is due to its efficient healthcare data processing. The model uses adaptive learning to prioritize essential characteristics, balance exploration and exploitation, and minimize errors, resulting in high accuracy, reduced variance, and enhanced generalization in big data analytics. Thus, the proposed KANN-HBA-WVGGWOA model attains 22.56 %, 18.45 % and 17.49 % lower MSE for diabetes; 24.38 %, 18.56 % and 18.49 % lower MSE for no diabetes estimated to the existing BDA-LR-RDMH, BDF-MRPK-DDP and BD-KNN-PHI respectively.

The high Cohen's kappa score for KANN-HBA-WVGGWOA in diabetes classification is due to its optimized IBESOA and effective handling of big healthcare data. These factors ensure a high level of agreement among predicted and actual classifications, beyond what would occur by chance. The proposed KANN-HBA-WVGGWOA model attains 2.33 %, 3.45 % and 7.28 % higher cohens kappa score for diabetes; 6.87 %, 7.58 % and 6.79 % higher cohens kappa score for no diabetes estimated to the current technique such as BDA-LR-RDMH, BDF-MRPK-DDP and BD-KNN-PHI respectively.

Table 6 shows the ablation study of the KANN-HBA-WVGGWOA. The baseline KANN model achieves 84.34 % accuracy. Adding MKF and IBESOA improves accuracy to 86.72 %. Combining KANN with WVGGWOA leads to 91.78 % accuracy. Integrating IBESOA with KANN and WVGGWOA further boosts performance to 95.29 %. Finally, the proposed KANN-HBA-WVGGWOA method achieves better accuracy of 99.67 %, with precision and recall at 98.33 % and 98.24 %, shows the effectiveness of the combined approach.

Table 7 tabulates statistical significance testing utilizing ANOVA for healthcare big data analytics based diabetes disease categorization. The three-degree-of-freedom variability between groups has a mean square of 115.23 and a sum of squares 345.67. With 20 degrees of freedom, the variability within groups has 153.32 SS, which results in 7.67 MS. F-statistic of 7.06, less p-value 0.0003 show the variation of the groups. More variant across the groups than would be predicted by chance only shows the independent variable affects the dependent variable when it comes to the big data analytics classification of diabetic condition.

Table 8 shows the comparative analysis using existing methods. The KANN-HBA-WVGGWOA model demonstrates superior performance in several areas. It reaches the better accuracy of 99.67 %, representing its strong ability to properly classify samples. It exhibits high precision 98.33 %, recall 98.24 %, F1-score 98.48 %, suggesting a balanced performance in lessening false positives and false negatives. The high MCC value 98.55 % further supports its robust predictive ability. However, the model shows a relatively higher Cohen's kappa score of 99.35 % and a lower MSE of 0.48 % compared to some other models. This indicates that the method outperforms in overall accuracy in terms of agreement between predicted and actual classifications and reducing prediction errors. In conclusion, the KANN-HBA-WVGGWOA model presents a promising approach for diabetes disease classification with its high accuracy and strong performance across various metrics.

Table 4
Recall and F1-score analysis.

Models	Recall (%)		F1-score (%)	
	Diabetes	No Diabetes	Diabetes	No Diabetes
BDA-LR-RDMH [23]	90.9	94.5	91.48	96.2
BDF-MRPK-DDP [24]	92.45	96.65	96.86	92.45
BD-KNN-PHI [25]	96.73	91.56	93.45	95.55
KANN-HBA-WVGGWOA (Proposed)	98.22	98.1	98.3	98.6

Table 5
Performance analysis of mean squared error and Cohen's kappa score.

Methods	Mean Squared Error (%)		Cohen's kappa score (%)	
	Diabetes	No Diabetes	Diabetes	No Diabetes
BDA-LR-RDMH [23]	0.57	0.85	93.15	96.4
BDF-MRPK-DDP [24]	0.76	0.52	97.24	92.33
BD-KNN-PHI [25]	0.81	0.78	92.3	95.53
KANN-HBA-WVGGWOA (Proposed)	0.55	0.48	98.33	98.65

Table 6
Ablation Study.

Methods	Accuracy (%)	Precision (%)	Recall (%)
KANN (Baseline)	84.34	82.67	80.47
MKF + IBESOA	86.72	84.19	82.49
KANN + WVGGWOA	91.78	88.39	87.39
KANN + IBESOA + WVGGWOA	95.29	93.15	91.37
KANN-HBA-WVGGWOA (Proposed)-with all methods	99.67	98.33	98.24

Table 7
ANOVA based statistical significance test.

Source of variation	Sum of Square(SS)	Degree of Freedom (DF)	Mean Square (MS)	F-Statics	P-value
Between Groups	345.67	3	115.23	7.06	0.0003
within Groups	153.32	20	7.67	–	–
total	499.00	23	–	–	–

4.3. Discussion

In this work, the KANN classifier based on the WVGGWOA is used to process the input data after it has been gathered from the database. Based on the WVGGWOA, the relevant and helpful elements are efficiently chosen. The function with the lowest error value is identified as the optimal answer when the fitness measure is assessed using the error value. By optimizing KANN, the proposed WVGGWOA is utilized to train the KANN classifier. The performance of KANN-HBA-WVGGWOA is evaluated under the mentioned metrics. The proposed model analyses healthcare big data to classify diabetes disease by combining the KANN and IBESOA techniques. The pre-processing data is performed by MKF.

Table 8
Comparative assessment using existing models.

Author & Year	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)	Cohens kappa score (%)	MSE (%)	Computational Time (sec)
Nauman, M., (2025) [23]	96.54	93.57	90.91	91.48	94.45	93.15	0.57	245.55
Ramani, R., (2025) [24]	93.57	97.55	92.45	96.86	91.84	97.24	0.76	285.32
Zamani, A.S., (2024) [25]	97.22	92.27	96.73	93.45	92.35	92.33	0.81	226.44
Hani, S.B., (2025) [26]	94.45	91.46	94.58	96.26	97.25	96.41	0.85	215.48
Mansour, R.F., (2023) [27]	92.33	94.29	96.65	92.45	95.43	92.38	0.52	207.56
Samiei, M., (2023) [28]	95.34	96.56	91.56	95.55	96.34	95.53	0.78	246.78
Jaiswal, V., (2023) [29]	96.45	95.38	94.29	93.19	95.29	95.56	0.82	234.56
KANN-HBA-WVGGWOA (Proposed)	99.67	98.33	98.24	98.48	98.55	98.67	0.48	99.35

Using IBESOA model, the feature selection is carried out from the pre-processing data then the classification is done by the KANN classifier. The training, testing and validation ratios are 70:15:15 produce the desired results. The outcomes are compared with current methods, such as BDA-LR-RDMH, BDF-MRPK-DDP and BD-KNN-PHI. The proposed approach obtained 99.67 % accuracy, 98.33 % precision, 98.24 % recall, 98.48 % F1-score, and 98.55 % MCC. It also obtained 99.35 % computational time. When compared with other classifiers, the proposed HRNN model is far better. Selection and filtering are crucial steps in big data analysing that assist the classifier to produce better classification outcomes.

5. Conclusion

In this work, Knowledge-aware Attentional Neural Network based Healthcare Big Data Analytics for Mobile Networks optimized with Weighted Velocity-Guided Grey Wolf Optimization Algorithm was implemented successfully. The proposed method enables a complete analysis of such structures based on the data, obtaining the best implementation IBESOA structure for feature selection, KANN model for classification. The KANN-HBA-WVGGWOA was implemented in Python utilizing PIMA Indians Diabetes dataset. The KANN-HBA-WVGGWOA approach achieved 23.26 %, 29.22 %, and 17.27 % higher specificity and 22.26 %, 30.22 %, and 27.27 %, with higher scalability ratio compared to the existing methods like BDA-LR-RDMH, BDF-MRPK-DDP and BD-KNN-PHI respectively.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] Y. Himeur, M. Elnour, F. Fadli, N. Meskin, I. Petri, Y. Rezgui, F. Bensaali, A. Amira, AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives, Artif. Intell. Rev. 56 (6) (2023) 4929–5021.

[2] M. Azadi, S. Yousefi, R.F. Saen, H. Shabanpour, F. Jabeen, Forecasting sustainability of healthcare supply chains using deep learning and network data envelopment analysis, J. Bus. Res. 1 (154) (2023 Jan) 113357.

[3] P. Kumar, R. Kumar, G.P. Gupta, R. Tripathi, A. Jolfaei, A.N. Islam, A blockchain-orchestrated deep learning approach for secure data transmission in IoT-enabled healthcare system, J. Parallel Distrib. Comput. 1 (172) (2023 Feb) 69–83.

[4] A.T. Keleko, B. Kamsu-Foguem, R.H. Ngouna, A. Tongne, Health condition monitoring of a complex hydraulic system using deep neural network and DeepSHAP explainable XAI, Adv. Eng. Softw. 1 (175) (2023 Jan) 103339.

[5] J. Vasa, A. Thakkar, Deep learning: differential privacy preservation in the era of big data, J. Comput. Inf. Syst. 63 (3) (2023 May 4) 608–631.

[6] W. Li, Y. Chai, F. Khan, S.R. Jan, S. Verma, V.G. Menon, F. Kavita, X. Li, A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system, Mobile Networks Appl. 26 (2021 Feb) 234–252.

[7] D. Rajesh Kumar, K. Rajkumar, K. Lalitha, V. Dhanakoti, Bigdata in the management of diabetes mellitus treatment, Internet of Things for Healthcare Technologies (2021) 293–324.

[8] I. Ahmed, M. Ahmad, G. Jeon, F. Piccialli, A framework for pandemic prediction using big data analytics, Big Data Res. 15 (25) (2021 Jul) 100190.

[9] C. Guo, J. Chen, Big data analytics in healthcare, in: Knowledge Technology and Systems: toward Establishing Knowledge Systems Science, Singapore, Springer Nature Singapore, 2023, pp. 27–70.

[10] S.J. Alsunaidi, A.M. Almuhaideb, N.M. Ibrahim, F.S. Shaikh, K.S. Alqudaihi, F. A. Alhaidari, I.U. Khan, N. Aslam, M.S. Alshahrani, Applications of big data analytics to control COVID-19 pandemic, Sensors 21 (7) (2021) 2282.

[11] A. Armghan, J. Logeshwaran, S.M. Sutharshan, K. Aliqab, M. Alsharari, S.K. Patel, Design of biosensor for synchronized identification of diabetes using deep learning, Results Eng. 20 (2023) 101382.

[12] M.A. Uddin, M.M. Islam, M.A. Talukder, M.A.A. Hossain, A. Akhter, S. Aryal, M. Muntaha, Machine learning based diabetes detection model for false negative reduction, Biomed. Mater. Devices 2 (1) (2024) 427–443.

[13] A. Rehman, S. Naz, I. Razzak, Leveraging big data analytics in healthcare enhancement: trends, challenges and opportunities, Multimedia Syst. 28 (4) (2022) 1339–1371.

[14] Kaul, D., Raju, H. and Tripathy, B.K., 2022. Deep learning in healthcare. Deep Learning in Data Analytics: Recent Techniques, Practices and Applications, pp.97-115.

[15] A. Alanazi, Using machine learning for healthcare challenges and opportunities, Inf. Med. Unlocked 30 (2022) 100924.

[16] K. Batko, A. Ślęzak, The use of big data analytics in healthcare, Journal of Big Data 9 (1) (2022) 3.

[17] M. El Khatib, S. Hamidi, I. Al Ameeri, H. Al Zaabi, R. Al Marqab, Digital disruption and big data in healthcare-opportunities and challenges, ClinicoEconomics and Outcomes Research (2022) 563–574.

[18] S. Bag, S. Gupta, T.M. Choi, A. Kumar, Roles of innovation leadership on using big data analytics to establish resilient healthcare supply chains to combat the COVID-19 pandemic: a multimethodological study, IEEE Trans. Eng. Manag. 71 (2021) 13213–13226.

[19] Muneeswaran, V., Nagaraj, P., Dhannushree, U., Ishwarya Lakshmi, S., Aishwarya, R. and Sunethra, B., 2021. A framework for data analytics-based healthcare systems. In Innovative Data Communication Technologies and Application: Proceedings of ICIDCA 2020 (pp. 83-96). Springer Singapore.

[20] V. Ravindran, G. Jeevanandan, V.P. Veeraraghavan, M. El-Sherbiny, A.L. Bohairi, Y.M. Mardini, O.S. Badughaysh, R.A. Found, A.M. Ibrahim, Comparative evaluation of physical and antimicrobial properties of metronidazole incorporated formulation of mineral trioxide aggregate-an invitro study, J Int Dent Med Res 17 (2024) 127–135.

[21] M. Au-Yong-Oliveira, A. Pesqueira, M.J. Sousa, F. Dal Mas, M. Soliman, The potential of big data research in healthcare for medical doctors' learning, J. Med. Syst. 45 (1) (2021) 13.

[22] D. Mukundan, G. Jeevanandan, S. Vishwanathaiah, S. Panda, T. Dawood, A. Abutaleb, P.C. Maganur, Comparative evaluation of the efficacy of 1% and 3% Sodium hypochlorite in reducing the microbial counts in primary teeth root canals using Bioluminometer – A randomized clinical trial, The Saudi Dental, Journal 36 (8) (2024) 1123–1127.

[23] M. Nauman, A.S. Almadhor, M. Albekairi, A.R. Ansari, M.A. Fayyaz, R. Nawaz, The role of big data analytics in revolutionizing diabetes Management and Healthcare Decision-making, IEEE Access (2025).

[24] R. Ramani, S.E. Raja, D. Dhinakaran, S. Jagan, G. Prabaharan, MapReduce based big data framework using associative kruskal poly kernel classifier for diabetic disease prediction, MethodsX (2025) 103210.

[25] A.S. Zamani, A.H.A. Hashim, A.S.A. Shatat, M.M. Akhtar, M. Rizwanullah, S.S. I. Mohamed, Implementation of machine learning techniques with big data and IoT to create effective prediction models for health informatics, Biomed. Signal Process. Control 94 (2024) 106247.

[26] S.B. Hani, M. Ahmad, Using big data to predict young adult ischemic vs, An Artificial intelligence Based Model. Intelligence-Based Medicine, Non-Ischemic Heart Disease Risk Factors, 2025.

[27] R.F. Mansour, J. Escorcia-Gutierrez, M. Gamarra, V.G. Díaz, D. Gupta, S. Kumar, Artificial intelligence with big data analytics-based brain intracranial hemorrhage e-diagnosis using CT images, Neural Comput. & Applic. 35 (22) (2023) 16037–16049.
[28] M. Samiei, A. Hassani, S. Sarspy, I.E. Komari, M. Trik, F. Hassanpour, Classification of skin cancer stages using a AHP fuzzy technique within the context of big data healthcare, J. Cancer Res. Clin. Oncol. 149 (11) (2023) 8743–8757.
[29] V. Jaiswal, P. Saurabh, U.K. Lilhore, M. Pathak, S. Simaiya, S. Dalal, A breast cancer risk predication and classification model with ensemble learning and big data fusion, Decision Analytics Journal 8 (2023) 100298.
[30] https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database.
[31] R. Korkin, I. Oseledets, A. Katrutsa, Multiparticle Kalman filter for object localization in symmetric environments, Expert Syst. Appl. 237 (2024) 121408.
[32] H. Youssef, S. Kamel, M.H. Hassan, L. Nasrat, F. Jurado, An improved bald eagle search optimization algorithm for optimal home energy management systems, Soft. Comput. 28 (2) (2024) 1367–1390.
[33] Y. Liu, J. Miyazaki, Knowledge-aware attentional neural network for review-based movie recommendation with explanations, Neural Comput. & Applic. 35 (3) (2023) 2717–2735.
[34] M. Premkumar, N. Shankar, R. Sowmya, P. Jangir, C. Kumar, L. Abualigah, B. Derebew, A reliable optimization framework for parameter identification of single-diode solar photovoltaic model using weighted velocity-guided grey wolf optimization algorithm and Lambert-W function, IET Renew. Power Gener. 17 (11) (2023) 2711–2732.