Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3, 1593-1620
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i3.5645

An efficient fuzzy logic and artificial intelligence based optimization strategy for bigdata healthcare system

Ravi Kumar[1*], S. Gokulakrishnan[2], S. N. V. J. Devi Kosuru[3], R. Praveen kumar[4], Thota Radha Rajesh[5]
[1]Department of Electronics and Communication Emgineering, Jaypee University of Engineering and Technology, A.B.Road, Raghogarh, Guna, Mathya Pradesh, India; ravikumarresearch890@gmail.com (R.K.).
[2]Department of Computer Science and Engineering, Dayananda Sagar University Bengaluru India.
[3]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India.
[4]Department of Electronics and Communication Emgineering, Easwari Engineering College, Chennai, Tamilnadu, India.
[5]Department Of CSE, Vignan's Foundation for Science, Technology and Research,Guntur, Andhrapradesh India.

Abstract: Digital health has revolutionized patient care by integrating big data analytics for predictive diagnostics, personalized treatment, and real-time health monitoring. However, the rapid generation of healthcare data from IoT devices, electronic health records, and medical imaging poses challenges such as high dimensionality, noise, and real-time processing. Existing methodologies struggle to balance accuracy and efficiency, making them limited for real-time healthcare applications. This paper proposes an optimization technique for big data-based healthcare systems incorporating fuzzy logic and artificial intelligence for predictive decision-making. This framework considers IoT-collected health data, which in itself brings forth problems of high dimensionality, noise, and real-time analytics. An intelligent preprocessing stage encompasses noise reduction and data integration, providing consistency and reliability for the dataset, which uses a fuzzy logic system. For optimal feature selection, an advanced AI model combines Lion optimization with heap-based feature estimation, thus reducing dimensionality while conserving health-relevant information. The optimized features are classified using a Hybrid Golden Eagle-Self-constructing Neural Fuzzy (HGE-SNF) algorithm, which dynamically tunes the weights and biases toward optimal classification performance. This hybrid approach improves predictive accuracy in disease detection and patient management issues and enhances computational efficiency for real-time healthcare applications. Experimental results indicate that it performs better than traditional methods and has great potential to revolutionize big data analytics in health systems.

Keywords: *Big data, Classification, Feature Extraction and COVID-19 database, Healthcare, Optimization.*

1. Introduction

Medical information has progressed in the domain of intellectual ability as a result of the increasing advancement of information technology [1]. Big data in healthcare ensures a proper database for healthcare service knowledge and e-health. For something like intelligent health information, categorization of health care big data is critical [2]. Furthermore, the Internet of Things (IoT), Big Data, as well as Artificial Intelligence (AI) are all connected topics of study that have an impact on the design and implementation of better-customized healthcare systems [3, 4]. The IoT offers a reduction in the worldwide cost of serious illness treatment. These technologies' real-time health information can be analyzed to assist patients through self-administration therapy [5]. Consequently, the COVID-19 (coronavirus) outbreak has been widely regarded as a serious health concern caused by virus infection that has a direct impact on the body's various parts of the body [6]. The viral is conveyed mostly by respiratory droplets such as coughing, sneezing, and coughs, and is passed from person to person [7].

History: Received: 9 January 2025; Revised: 11 March 2025; Accepted: 12 March 2025; Published: 21 March 2025
* Correspondence: ravikumarresearch890@gmail.com

Despite fever as well as coughing being the most common signs and symptoms, also before health disorders including heart disease, kidney disease, diabetes, and cancer can potentially worsen the disease's result [8]. A pandemic-ready healthcare system and a fight against this illness is necessary. To further assess the idea, this article uses prospective data mining algorithms on COVID-19 datasets, heart disease, renal disease, diabetes, as well as cancer [9]. Humans have been suffering from treatable illnesses for decades, with just a few people who are affected by premature foreseeable ailments that can be entirely recovered, while another half of the population receives speedy treatment to improve human survival rates [10]. By establishing a functional healthcare system for individuals to use in screening, prior analysis aids physicians in providing patients with the therapy that prefer. By utilizing research centers, online entrepreneurs and banking industries, and other organizations, the shared health benefit expenditure may be achieved, allowing for the long-term growth of vital medical fields, research, emulation, and implementation, as well as financing and administration [11].

Biosensors and social media platforms play a critical role in introducing a new way to collect patient data for effective healthcare monitoring [12]. Constant process characterized using sensor modules, on the other hand, generates a lot of medical data [13]. Furthermore, user-generated information systems on social media sites are unstructured and come in vast quantities. Current healthcare monitoring devices are ineffective in retrieving relevant data from sensors as well as social media data and also analyzing it properly [14]. Several developed countries have announced a lot of large information systems strategies, aggressively encouraged big data application areas, and saw health care with big data as a critical part of national essential services. Different deep-learning concepts are used to study and uncover patterns in datasets using classification algorithms [15]. The most recent advances in machine learning (ML) based tools and methodologies improve the study of emerging models. Numerous researchers used artificial intelligence techniques and procedures to analyze COVID-19 data and came up with some interesting results [16]. To forecast the diagnosis of the illness, several researchers employed neural network models and deep learning approaches like decision tree (DT) [17]. Multi-Layer Pi-Sigma based Neuron Model (MLPSNM) Ahmed, et al. [18] fuzzy analytic-based hierarchy process (AHP) Nazari, et al. [19] k-nearest neighbor (KNN) Mittal, et al. [20] logistic regression (LR) Manogaran and Lopez [21] and Back-Propagation Based Neural Network (BPNN) Ravindra, et al. [22] while many others utilized optimization-based classification algorithms [23]. Furthermore, typical machine learning algorithms are insufficient for processing healthcare large data to anticipate abnormalities. Nevertheless, such techniques are vulnerable to incorrect thresholds and an unequal distribution of training data classifications, resulting in a low classification performance.

The main contribution of this study is to use a novel classifier with optimum features to classify health information. Healthcare big data analytics can benefit from this examination of specific healthcare data received through IoT. The finest features are selected from the big data for the process of classification is performed by the proposed Lionized Heap Optimizer (LHO) based feature extraction and selection. The LHO method is a combination of Lion and Heap-based optimization combination. The best fitness of Lion is provided by the performance of the heap optimizer in feature estimations. Furthermore, the accurate data of classification is carried out by the proposed Hierarchy Golden Eagle Based Self-Constructing Neural Fuzzy (HGE-SNF) method of classification technologies. The key contribution of this research is articulated as follows:

- Initially, the IoT-based big data is collected from remote areas like various diseases, patient data, etc.
- Then, the pre-processing stage is executed for removing the irreverent data and adding missing information.
- The proposed LHO-based feature extraction and selection method is an aid to finding the optimal features in large datasets.
- Consequently, the finest classification function is achieved by the proposed HGE-SNF classification method.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

- Eventually, the suggested model's performance is assessed using dissimilar metrics and compared with the conventional methods.

The rest of this article is articulated as follows: the recent research related to this research is provided in Section 2. The system model and its problem definition are explained in Section 3. The proposed framework of big data classification in the healthcare system is detailed in Section 4. The result, discussion, and comparison are provided in Section 5. Finally, the article research is concluded in Section 6.

2. Related Work

Some of the current research works related to this idea are articulated as follows: Big data in healthcare ensures a basic information resource for healthcare service knowledge and eHealth. For this reason, Xing and Bei [24] compare the standard KNN technique with an upgraded K-Nearest Neighbor (KNN) algorithm. The classification is done in the request instance neighborhood of a standard KNN classifier, and each classification is given a weight. The approach uses clustering to conduct noise removal filtering and increases the classification performance of the KNN algorithm via increasing the search speed of KNN while retaining the KNN system's accuracy of classification. However, insufficient information can affect the classification performance.

The medical industry can benefit from the proper application of data categorization in the IoT to discover new or undiscovered truths. The Random Forest Classifier (RFC) as well as the MapReduce method is used to build big data technologies on an IoT-based medical system by Lakshmanaprabu, et al. [25]. The e-health data obtained from people suffering from various conditions is taken into consideration for analysis. For enhanced classification, the appropriate characteristics are picked from the collection through the Improved Dragonfly Algorithm (IDA). Furthermore, using optimum attributes, the RFC classifier is applied to categorize the e-health material. Nevertheless, the execution cost of this system is higher.

To properly collect and manage health information, as well as increase prediction performance, a unique smart healthcare architecture provided by the cloud infrastructure and a big data platform is presented by Ali, et al. [26]. Data mining methods, terminologies, and Bidirectional Long Short-Term Memory (Bi-LSTM) are all used in the suggested data analytics engine. To forecast pharmacological side effects and aberrant situations in patients, Bi-LSTM accurately identifies healthcare data. In addition, the suggested system uses healthcare data from diabetes, blood pressure, mental health, and medication evaluations to classify the patient's medical conditions. But the time series for the processing will be more for huge data.

Thanga Selvi and Muthulakshmi [27] presented a large-scale health application structure based on an optimum artificial neural network (OANN) for heart disease diagnostics, which would be the world's deadliest disease. The suggested OANN consists of two key process steps: distance-based misclassified instance removal (DBMIR) and the teaching as well as learning-based optimization (TLBO) method for ANN, together referred to as TLBO-ANN. Yet, the network's existence is undetermined.

Galetsi, et al. [28] did a systematic evaluation of medical big data and analytics due to the significant growth of articles in the healthcare industry. The segmentation of big data categories in healthcare, associated analytical methodologies, produced value for users; infrastructure and technologies for processing large health data, and prospective elements of the subject are all discussed. The results of this study are exciting and give significant information to clinicians, politicians, and academics, as well as pointing users in the direction of future investigation. The conventional machine learning methods reduced the training efficiency.

The investigation of AI for IoT and medical systems, which includes the usage and practice of AI methodology in different fields of healthcare is discussed by Oniani [29] Furthermore, the Internet of Medical Things addresses numerous health conditions such a vital biophysical parameters supervision, diabetes, and medical decision-making support methods. Consequently, Big Data, IoT, and AI are three

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

connected study topics that have a significant influence on the development and deployment of better customized medical systems. High error is possible in this approach due to the uneven classification.

In this case, the need for an IoT-cloud-based healthcare paradigm is critical to making better decisions in the COVID-19 pandemic. Due to this reason, Mukherjee, et al. [30] are to use a classification algorithm learner to do data analytics on the illness. This study suggested the eKNN method, which did not pick the k value arbitrarily. The k value, on the other hand, was calculated using a functional form of the dataset's response rate. The upgraded KNN technique was supplemented by selecting features via Ant Colony Optimization (ACO) in the subsequent research issue. The convergence speed of the classification feature selection is very low.

Various deep learning methods, as well as large data analytic procedures, are employed to aid in the detection and prediction of COVID-19 epidemics throughout the globe. Furthermore, CT and X-ray imaging are used, and an H2O Deep-Learning-inspired approach based on Big Data analytics (DLBD-COV) is presented for early identification of COVID-19 patients by Elghamrawy [31]. For scalability analysis, the suggested diagnostic model is built on the classification algorithm (H2O). The classification accuracies of the Generative Adversarial Networks (GAN), as well as Convolutional Neural Networks (CNNs), are analyzed. Because of the complicated data models, training is exceedingly costly.

Quick transmission of viral illnesses is a growing public health concern throughout the world. COVID-19 is currently regarded as the most dangerous and unique of these infections. The presented study provides by Ahanger [32] a useful approach for tracking and predicting COVID-19 viral infection (C-19VI). For the anticipation and prevention of COVID-19 infection, the suggested framework contains a multiple design. The provided technique is used to encourage a user to check COVID-19-based Fever Measure (C-19FM) regularly and predict it so that preventative steps may be implemented ahead of time. In addition, the presence of C-19VI is detected in more than a geographical region using the self-organized mapping approach.

3. System Model and Problem Statement

One of the most challenging and exciting topics in recent days is big data-based healthcare monitoring combined with artificial intelligence [3]. The IoT-based approach has been widely utilized throughout the world since it can be accessed at whatever time and from any location. IoT-based biosensors are attached to the human body in a remote medical system and can measure factors including blood, diabetes, cardiac rate, pulse rate, and so on Mbunge and Muchemwa [12]. All health centers irrespective of where the diagnostic was performed, could acquire knowledge for each patient using Big Data technology. Furthermore, the tests would be recorded in real-time, enabling decisions to be made as soon as the patient was tested. Preprocessing, feature extraction, and classification are three of the key purposes of big data analytics in healthcare for disease identification. The system model of big data in the healthcare system is illustrated in Figure 1.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

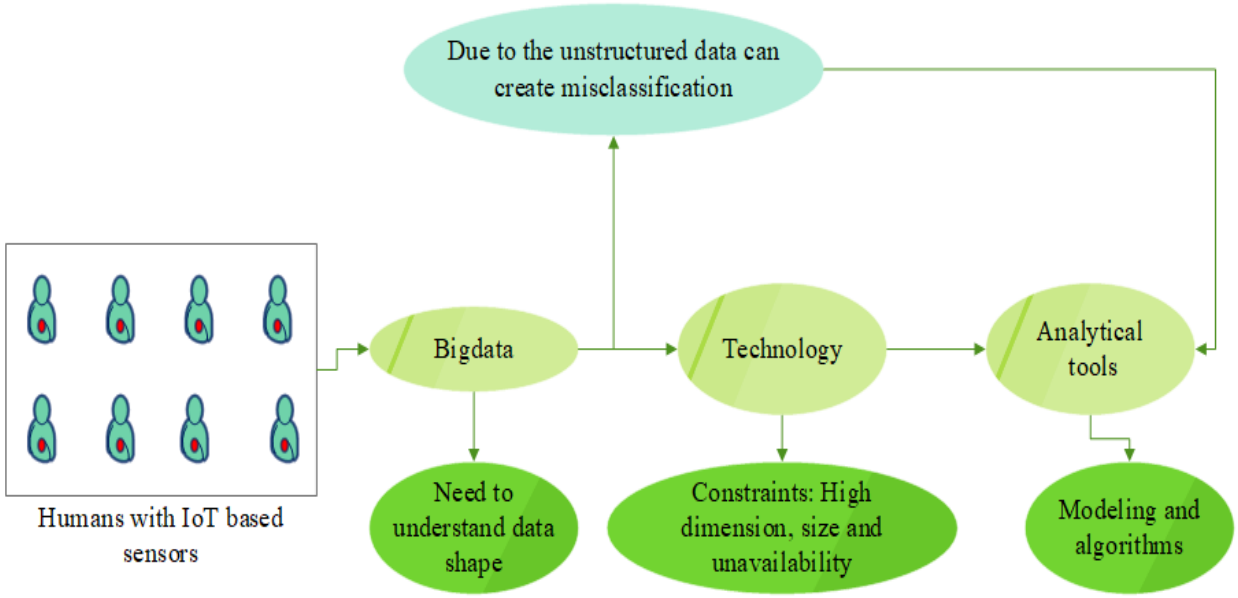


Figure 1.
System model and problem of big data classification.

Diagnosis of sickness and hospitalization, on the other hand, takes huge amounts of energy, money, and time. The provision of excellent treatment is the primary problem in healthcare management (health centers, hospitals) [16]. Lack of medical treatments and a scarcity of professionals might result in a significant number of incorrectly diagnosed situations. As a result, advancement has necessitated the development of quick and efficient prediction techniques.

In recent years, most hospitals have sorted patient information systems to provide health treatment [10]. Typically, these technologies generate vast volumes of data in textual form, photos, statistics, and graphs. However, these statistics are frequently employed in medical decision-making. This behavior leads to mistakes, annoyances, and higher medical expenditures, all of which have an impact on the quality of care provided to patients. As a result, the breadth of regular computing approaches must be expanded to improve healthcare resources.

4. Proposed Framework

The proposed model for the healthcare system's comprehensive working process is depicted in Figure 2. The study's important component is the development of a unique classifier with optimal characteristics for categorizing health data.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

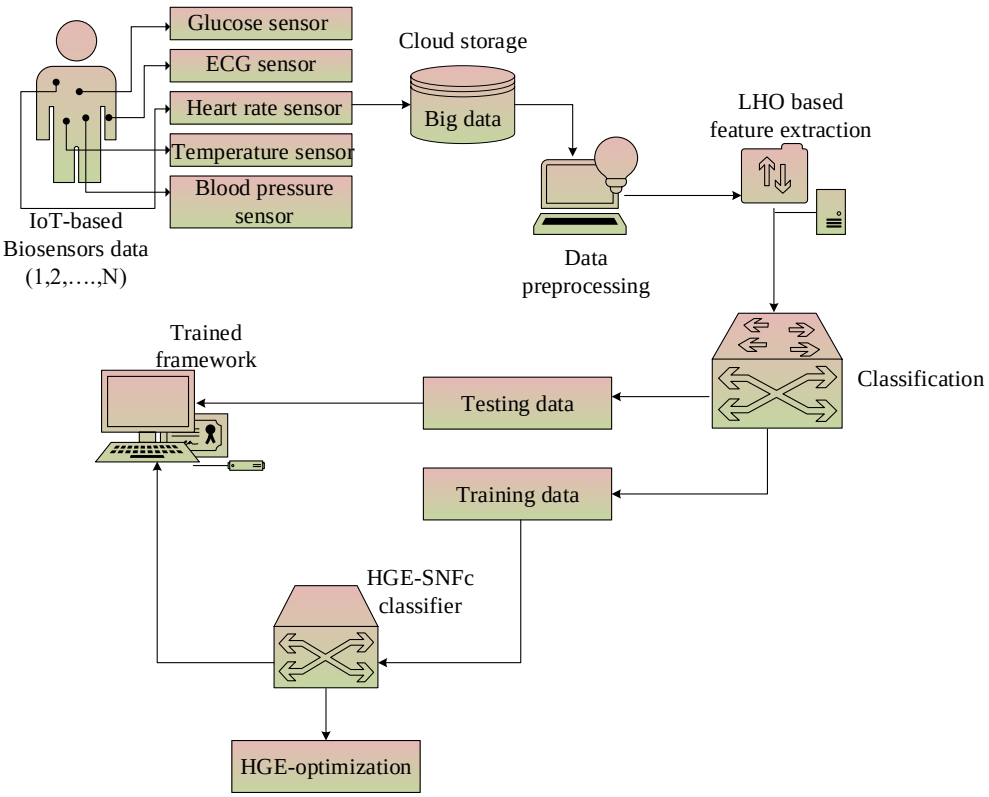


Figure 2.
Proposed framework of big data classification in the healthcare system.

This investigation of specialized healthcare data collected by IoT can improve big data analytics in healthcare. Proposed LHO-based feature extraction and selection is used to choose the best features from the huge data for the classification process. The LHO technique is a hybrid of the Lion and Heap optimization methods. The heap optimizer's performance in feature estimates is given the highest level of Lion fitness. In addition, the suggested HGE-SNF technique of classification technology performs accurate data classification.

4.1. Big Data Collection

The base layer is designed to collect information about a patient's health that is gathered over time from various sensors such as health sensors, meteorological sensors, biological sensors, and geographic sensors. In general, databases include infrastructures connecting systems that are used to improve service, and these records contain certain basic healthcare information like the patient's name, gender, age, type of disease, medications ingested, and so on.

4.2. Data Pre-Processing Stage

The process of converting raw data into a comprehensible format is known as data preparation. The goal of preprocessing is to discover the most relevant collection of features in an attempt to optimize the classifier's performance. Data preparation is separated into four steps to make the work easier: data cleaning, data reduction, data integration, and data transformation. The practice of removing erroneous, inadequate, and misleading data from databases, as well as replacing missing information, is known as data cleaning. Data integration is the process of merging various sources into a particular dataset. Data reduction reduces the amount of data, making processes easier while still producing the same or nearly identical results. Data transformation is the process of changing the format or organization of data.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

4.2. Feature Extraction and Selection

After pre-processing the raw healthcare data, the exact features should be extracted and selected from the dataset for accurate disease classification. The dataset has a large number of characteristics, however just 114 are necessary for this research. As a result, the unique LHO approach is presented for accurate feature extraction. The suggested LHO approach is a type of hybrid optimization that combines lion and heap-based optimization. To extract the necessary characteristics from the database, the fitness of optimization is applied. Patient registration number, age, gender, blood glucose levels, chronic pain category, blood pressure information, blood cholesterol levels, cigarette usage, coronary artery disease, anxiety levels, respiration, chronic pain location, exercise data, and so on are some examples of features for extraction. The pseudo-code of the LHO method is provided in algorithm 1.

4.3.1. Initialization

Set up the featured and unknown data, as well as the algorithm settings. Thus, eqn. (1) arranges data information in a random order,

$$F_d = f_0, f_1 \ldots\ldots f_n \quad (1)$$

Where $F_d = 1,2,\ldots.n$ and the overall amount of data is denoted as n. The fitness value of the proposed system for finest feature selection is expressed in eqn. (2)

$$\text{Fitness}(F\text{eatures}) = \text{f}(F_d) = f(f_0, f_1 \ldots\ldots f_n) \quad (2)$$

Where F_d is the fitness function that provides the optimal selection of features from the database. The initial stage is to produce a random number of population options in the search space. As input data, a proportion of created responses are picked at random. The remainder of the population will be split into S features at random. During the search, each technique records the most often used position. Every feature region is constructed based on these indicated places. As a result, each feature's dataset is made up of marked spots (the most often checked positions) by its individuals.

4.3.2. Features Extraction

The parameters are defined and initialized as follows: Set up general settings such population size (N), number of design variables/dimensions (F), maximum number of iterations (T), and design variable values (b_d, a_d) as well as the algorithm specific parameter $D = T/No.of iteration$. Create a stochastic group of iterations with parameters respectively. Even though this is a tree-shaped data structure, due to its validity, it can be simply implemented using an array.

Feature interaction: In a centralized organizational structure, top rules and regulations are enforced, and followers observe their immediate supervisor. This type of behavior could be simulated by upgrading the position of every iteration of the algorithm f_d about its primary node C using eqn. (3),

$$f_d^s(t+1) = C^s + \alpha\beta^s \left| C^s - f_d^s(t) \right| \quad (3)$$

Where f_d represented as d^{th} feature of the dataset, t is the present iteration, s is the superscript of the vector element, α and β are the noteworthy parameters. Also, after decreasing α linearly from 2 to 0 during iterations, iterations begin to climb again to 2 after reaching 0. Regardless, the parameter determines how many cycles C complete in each repetition.

Consequent feature point interaction: The precise features are represented by feature points that have the same score. In order to perform official tasks, they communicate with one another. In this

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

work, the variables with the same rank are called features, and each search agent $\vec{f}_d$ changes its position based on the expression and its randomly selected features $\vec{D}_f$ using eqn. (4)

$$f_d^s(t+1) = \begin{cases} D_d^s + \alpha\beta^s \left| D_d^s - f_d^s(t) \right| , f(\vec{D}_f) < f(\vec{f}_d(t)) \\ f_d^s + \alpha\beta^s \left| D_d^s - f_d^s(t) \right| \; f(\vec{D}_f) \geq f(\vec{f}_d(t)) \end{cases} \tag{4}$$

where f is the objective function, which is used to calculate the search agent's fitness. Here, in Eq. (4) permits the search agent to examine the area around D_d^s if $f(\vec{D}_f) < f(\vec{f}_d(t))$, and the area about f_d^s else. Exploration as well as exploitation is both promoted by this behavior. The diversity is included through the randomly chosen of coworkers, and the constant quest for good solutions encourages exploitation.

Updating extracted features: A roulette wheel's goal is to balance these probabilities, which are separated into three parts, such as f_1, f_2 and f_3. Furthermore, eqn. (5) is used to update the extracted features,

$$f_d^s(t+1) = \begin{cases} f_d^s & f \leq f_1 \\ C^s + \alpha\beta^s \left| C^s - f_d^s(t) \right| & f > f_1 \text{and} f \leq f_2 \\ D_d^s + \alpha\beta^s \left| D_d^s - f_d^s(t) \right| & f > f_2 \text{and} f \leq f_3 \text{and} f(\vec{D}_f) < f(\vec{f}_d(t)) \\ f_d^s + \alpha\beta^s \left| D_d^s - f_d^s(t) \right| & f > f_2 \text{and} f \leq f_3 \text{and} f(\vec{D}_f) \geq f(\vec{f}_d(t)) \end{cases} \tag{5}$$

Where f is a one-of-a-kind number between 0 and 1 generated at random. As a result, feature points alter their scores on a frequent basis in order to convergence on the optimal global solution based on the previously described algorithms.

4.3.3. Features Selection

Some data check for a feature in a collection to offer classes for their features in each feature. To surround the features and capture it, these hunters use unique tactics. During the search, each data point adjusts its position dependent on its own and the locations of the other members of the group. Because some of these hunters use Opposition depends learning to surround features and strike from contradictory stances while hunting. The group having the largest total fines among its members is referred to as the Center, while another two groups are referred to as the Wings. The center of attention for hunters is a fictitious feature. During searching, predators are chosen at random and assault fictitious characteristics. This technique will be established later as per the class that the picked information relates under. If a hunter increases his or her finesses while pursuing, features will evade the hunter and a new location of features will be gained as follows:

$$f_a' = f_a + r(0,1) \times p \times (f_a - f_d) \tag{6}$$

Where f_a denotes the present location of a feature, f_d denotes a new location hunter who pursues features, and p denotes the percentage progress in the predator's f_d fitness.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

Encircling feature points: The below formulations are given to simulate hunting parties around features. The following are the new spaces of hunters belonging to both the conservative and progressive wings:

$$f'_d = \begin{cases} r((2\times f_a - f_d), f_a), & (2\times f_a - f_d) < f_a \\ r(f_a, (2\times f_a - f_d)), & (2\times f_a - f_d) > f_a \end{cases} \quad (7)$$

Where f_a is the new location of features, f_d is the current position of feature selector and f'_d is the new location of feature selection.

$$f'_d = \begin{cases} r(f_d, f_a), & f_d < f_a \\ r(f_a, f_d), & f_d > f_a \end{cases} \quad (8)$$

Where $r(0,1)$ creates a random value between 0 and 1 in the above formulae, where 0and 1 are top and bottom limits, respectively.

Moves to new features: The selector performance is if the features selection is increases their best place in the last iteration of the LHO The performance of selection at iteration t^{th} in dataset is defined as using eqn. (9),

$$P(f_d, t, D) = \begin{cases} 1 & Best^{t}_{f_d,D} < Best^{t-1}_{f_d,D} \\ 0 & Best^{t}_{f_d,D} = Best^{t-1}_{f_d,D} \end{cases} \quad (9)$$

The best location found by lion f_d, till iteration t is $Best^{t}_{f_d,D}$. The selectors have converged at a place that is far from optimal, based on the large number of performances. Similarly, a low performance rate indicates that the feature selection is circling the best option without making meaningful progress. As a result, this factor may be used to determine the size of a classification. When the value of performance drops, the features size grows, this leads to more variety. As a result, the features size is determined by eqn. (10)

$$F_d^{size} = \max\left(2, c\left(\frac{M_d(u)}{2}\right)\right) \quad d = 1,2,....N \quad (10)$$

Where $M_d(u)$ is the features quantity in the dataset that has the enhancement of fitness function in the final iteration. If the optimal features are selected the process is stopped its criteria otherwise it repeats till the final features are obtain for another iterations.

Algorithm 1: The pseudo-code of LHO feature extraction and selection method

Initialization: Initialize pre-processed data and algorithm parameters.

Compute $\text{Fitness}(Features) = f(F_d) = f(f_0, f_1 ..... f_n)$

Feature Extraction: Set up (N), (F), (T), (b_d, a_d), $D = T/No.of iteration$

`

Compute *Feature interaction by eqn. (3)*

Estimate *Consequent feature point interaction by eqn., (4)*

if $f(\vec{D}_f) < f(\vec{f}_d(t))$

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

Examine the features
else
Moves to another feature points
Updating extracted features using eqn. (5)
Feature Selection: Setup extracted features for optimal selection
Generate features selection location $f_a' = f_a + r(0,1) \times p \times (f_a - f_d)$
For a=1: f_d (f_d is the number of selector)
Move a[th] detector to features based on the related threshold value
If new location of f_d the selector is better than the old points
Features data is neglected
End
End
Encircling feature point using eqn. (7) and (8)
Moves to new features
The performance of selection at iteration t^{th} in dataset is defined as using eqn. (9)
If performance is low
Encircling the new position of feature selection
Else
Optimal feature point is selected
end
Estimate features size for data classification
End
If not optimal features are obtained
Returns back to next iteration

4.4. HGE-SNF Based Classification

The HGE-SNF is made up of nodes, each with a limited number of "fan-in" connections (expressed by weights and biases from all other modules) and "fan-out" connections (defined by weights and biases from all other modules). An incorporation function is linked to a node's fan-in, and it's used to aggregate data, activation, or proof from those other components. A fuzzification layer, four hidden layers, and a defuzzification layer make up the proposed HGE-SNF structure. The proposed SLGENF method is the combination of heap golden eagle optimization with the self-constructing improved fuzzy neural technique. The weights and bias, also the parameters of the intelligent method is optimized by the lionized golden eagle method.

4.4.1. Finest HGE Optimization Tuning

The suggested HGE method is a hybrid of heap and Golden eagle optimization techniques. In the suggested HGE-SNF technique, the fitness of hunting and encircling behaviors is taken into account for optimum tuning selection.

Initialization: All iteration, each parameter a_{yk}, c_{ky} and σ_{ky} picks k the activation of another parameter y at random and revolves about the best position that parameter y has visited thus far. Here, have $y \in$ (1,2,...,PopSize} since the parameter a_{yk}, c_{ky} and σ_{ky} might select to cycle has its memory.

Tuning selection: Each parameter should select a tuning to execute the cruise and assault actions in each cycle. Tuning is modeled in HGE as the best solution found by the swarm of parameters thus far. Every parameter has the ability to remember the best answer it has discovered up to this point. Each

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

cycle, every search agent chooses a specific tuning from the flock's database. The strike and cruise paths for each variable are then computed based on the tuning that has been chosen. The database is updated if the new location determined using attack as well as cruise directives is superior to the prior position in the memory. In HGE, the tuning search algorithm is quite significant. A simple method of selection is for each parameter to pick the tune in the own database. In the current iteration, each parameter chooses its tuning at random from the database of any other swarm component. It's important to remember that the chosen tune isn't always the closest or farthest tuning. The assault and cruise procedures are then performed on the specified tuning by each parameter.

Parameter exploitation: The approach can be represented by a vector that starts at the present state of the parameter and ends at the parameter's location in the classification memory. Furthermore, eqn. (11) may be used to compute the entry point for parameter k.

$$\vec{a}_k \equiv \vec{Y}_y^* - \vec{Y}_k \tag{11}$$

Where $\vec{a}_k$ is parameter k's assault vector, $\vec{Y}_y^*$ is parameter k's best location of tuning thus far y, and $\vec{Y}_k$ is parameter k's current position. The exploitation phase in HGE is highlighted by the exploitation, which directs the population of parameters toward the most frequented places.

Parameter exploration: The exploitation is used to compute the exploration. The exploration is perpendicular to the exploitation and tangential to the circle. The parameter's constant rate relative to the SNF fuzzy can alternatively be viewed of as the exploration. The n-dimensional exploration takes place inside the circle's perpendicular hyperplane. To compute the exploration, we must first estimate the perpendicular hyperplane's solution. A random point from a hyperplane as well as a perpendicular component to that hyperplane, termed the coordinates of the hyperplane, can be used to compute the expression of a hyperplane in n dimensions. In n-dimensional space, Eqn. (12) shows the scalar version of the hyperplane solution.

$$s_1 y_1 + s_2 y_2 + \ldots\ldots\ldots\ldots s_m y_m = r \Rightarrow \sum_{i=1}^{m} s_i y_i = D \tag{12}$$

where $S = [s_1, s_2 \ldots\ldots s_m]$ are the ordinary vector and $Y = [y_1, y_2, \ldots\ldots y_m$ are the decision key vector of i^{th} node. The cruise hyperplane's arbitrary target point is discovered. The generic form of the exploration hyperplane's target destination is shown in Eqn. (13),

$$c_{ky} = \frac{D - \sum_{k;k \neq y} a_y}{a_k} = r \tag{13}$$

New parameter tuning: The movement of the parameters comprises of exploitation and exploration. The step vector for the parameters k in t iteration using eqn. (14)

$$\Delta y_m = \vec{r}_1 \left(q_a^0 + \frac{t}{T} \left| q_a^T - q_a^0 \right| \right) \frac{a_{ky}}{\left\| a_{ky} \right\|} + \vec{r}_2 \left(q_c^0 + \frac{t}{T} \left| q_c^T - q_c^0 \right| \right) \frac{c_{ky}}{\left\| c_{ky} \right\|} \tag{14}$$

Where q_a^T represents the exploration coefficient in T iteration and q_c^T represents the cruise coefficient in T iteration, and control how exploitation and exploration impact. $\vec{r}_1$ and $\vec{r}_2$ are random

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

vectors with entries in the Dash, et al. [1] range. This behavior may be represented by changing the position of each search agent y_m with reference to its starting parameter B using eqn. (15)

$$y_m(t+1) = C^k + \alpha\beta^k \left| C^k - y_m(t) \right| \quad (15)$$

Where y_m represented as m^{th} feature of the dataset, t is the present iteration, α and β are the significant parameters.

Termination: The parameter location in iteration $t+1$ is obtained by adding the step vector from iteration t to the positions from iteration t.

$$y^{t+1} = y^t + \Delta y_k^t \quad (16)$$

If the starting position of the parameter I is more appropriate than the previous position, the database of this parameter is changed to reflect the new position. However, the memory is preserved, but the parameter is relocated. In the next iteration, each parameter picks a variable from the population at arbitrary to cycle it around most-visited location, determines exploitation, exploration, and ultimately the step vector and new role for the next iterative process. This loop continues until one or more of the termination conditions is met.

4.4.2 Improved Classification

The proposed model of classifier is illustrated in figure 3.The crisp input from the feature selection is turned into a fuzzy set of values in the fuzzification layer. In eqns. (17) and (18), the input activation function and output of layer were each stated, respectively.

$$Input = f\left[a_1^{(s)}, a_2^{(s)}, \cdots a_n^{(s)}; q_1^{(s)}, q_2^{(s)}, \cdots q_n^{(s)}\right] \quad (17)$$

where the inputs to this unit are $a_1^{(s)}, a_2^{(s)}, \cdots a_n^{(s)}$ and the link weights are $q_1^{(s)}, q_2^{(s)}, \cdots q_n^{(s)}$. The layer number is indicated by the superscript (s) in the above equation. Every node's second activity is to produce an activation value based on its primary input.

$$\text{Output} = O_k^{(s)} = A(\text{input}) = A(f) \quad (18)$$

Where the activation function is denoted as $A(\cdot)$, then the subsequent six layers for the classification is explained as follows:

First input layer: This layer does not do any computations. This layer's nodes, each of which relates to a single input parameter, only send input data straight towards the next layer. That is appropriate using eqn. (19) and the first layer link weight factor is $\left[q_k^{(1)}\right]$ one.

$$f = a_1^{(s)} \text{ and } A^{(1)} = f \quad (19)$$

Second Membership function layer: Each unit in this layer correlates to one of the input parameters in the first layer of linguistic variables (small, medium, large, etc.). Furthermore, the second layer calculates the membership functions, which represents the degree whereby an input data corresponds to fuzzy rules. A Gaussian membership functional is used in this study, which has been demonstrated to be a global probabilistic model of any nonlinear system on a consists largely using eqn. (20)

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

$$f\left[a_{ky}^{(2)}\right]=-\frac{\left[a_k^{(2)}-c_{ky}\right]^2}{\sigma_{ky}^2} \text{ and } A^{(2)}(f)=h^f \quad (20)$$

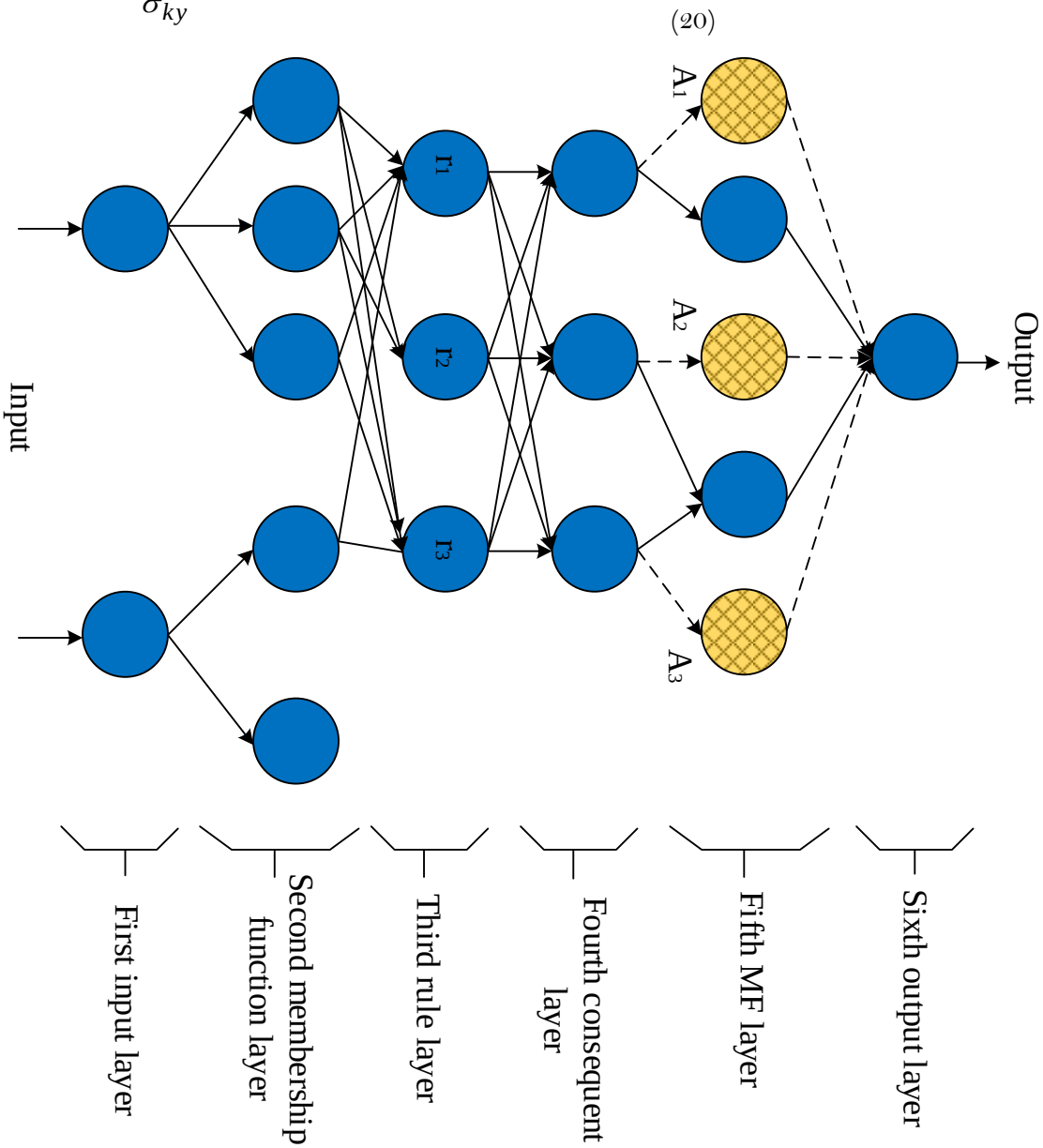


Figure 3.
The proposed model of classification

where c_{ky} are the mean and σ_{ky} are the variance of the Gaussian MF of the y^{th} term of the k^{th} input variable is z_k, respectively. As a result, in this second layer the weight of link may be expressed as c_{ky}. Here, the HGE method is established for the fuzzy set values.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

Third Rule layer: This layer's nodes each contain one fuzzy inference system rule and conduct prerequisite testing. For third layer component, here utilized the AND function as seen below.

$$f\left[a_k^{(3)}\right]=\prod_k a_k^{(3)} = h^{-[E_k(z-c_k)]^T[E_k(z-c_k)]} \text{ and } A^{(3)}(f)=h^f \tag{21}$$

Where the number of second layer is denoted as n involved in the IF portion of the fuzzy rule and the diagonals are denoted as $E_k = d(1/\sigma_{k1},1/\sigma_{k2},.....1/\sigma_{kn})$ and $c_k = d(c_{k1},c_{k2},.....c_{kn})^T$. The third layer weight link $\left\lfloor q_k^{(3)} \right\rfloor$ is one. The firing strength of the associated fuzzy rule is represented by the third layer consequences f.

Fourth Consequent layer: This layer has the same number of components as third layer, and the firing strength determined in third layer is normalized by in this layer via eqn. (22) and similarly, the fourth layer weight link $\left\lfloor q_k^{(4)} \right\rfloor$ is one.

$$f[a_k^{(4)}]=\sum_k a_k^{(4)} \text{ and } A^{(4)}(f)=\frac{a_k^{(4)}}{h^f} \tag{22}$$

Fifth MF layer: The subsequent layer is the term coined to this layer. This layer uses two different modes, which are depicted in Fig. 3 as blank as well as shaded circles, correspondingly. The fundamental node represented by empty circles is a fuzzy set characterized by a Gaussian membership degree of the outcome variable. For the local mean of maximum (LMOM) based defuzzification procedure, only the center of each Gaussian membership value is sent to the subsequent layer, while the width is solely utilized for output grouping. Multiple fourth layer terminals may be linked to the same empty component in the fifth layer, resulting in the same fuzzy numbers being provided for separate rules. The empty node's purpose is to serve as a substitute using eqn. (23)

$$f=\sum_k a_k^5 \text{ and } A^{(5)}(f)=f\cdot A_{0k} \tag{23}$$

Where A_{0k} is represented as c_{0k}, the mean of Gaussian MF. The shaded element is only created when it is required. Each shaded component in the fifth layer corresponds to a component in fourth layer. The result from fourth layer is among the sources to a shaded node, whereas the other potential inputs concepts are the input parameters from first layer. The shaded branch function can be defined that allows you to create a shaded component

$$f=\sum_k a_{yk} z_y \text{ and } A^{(5)}(f)=f\cdot a_k^5 \tag{24}$$

Where a_{yk} is the relevant variable and the summing is over the key terms linked to the shaded node alone. By merging these two components in fifth layer, the entire function provided through this layer may be expressed as

$$A^{(5)}(f)=\left(\sum_k a_{yk} z_y + A_{0k}\right)a_k^5 \tag{25}$$

Sixth output layer: This layer's nodes each relate to a single output variable. The node combines all of fifth layer recommendations and works as a defuzzifier also then classifies the accurate results.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

$$f\left\lfloor a_k^{(6)} \right\rfloor = \sum_k a_k^{(6)} \quad and \quad A^{(6)}(f) = f \tag{26}$$

For optimal classification, the parameters a_{yk}, c_{ky} and σ_{ky} are tuned by the proposed HE algorithm. The flowchart of proposed model in big data classification is illustrated in Figure 4.

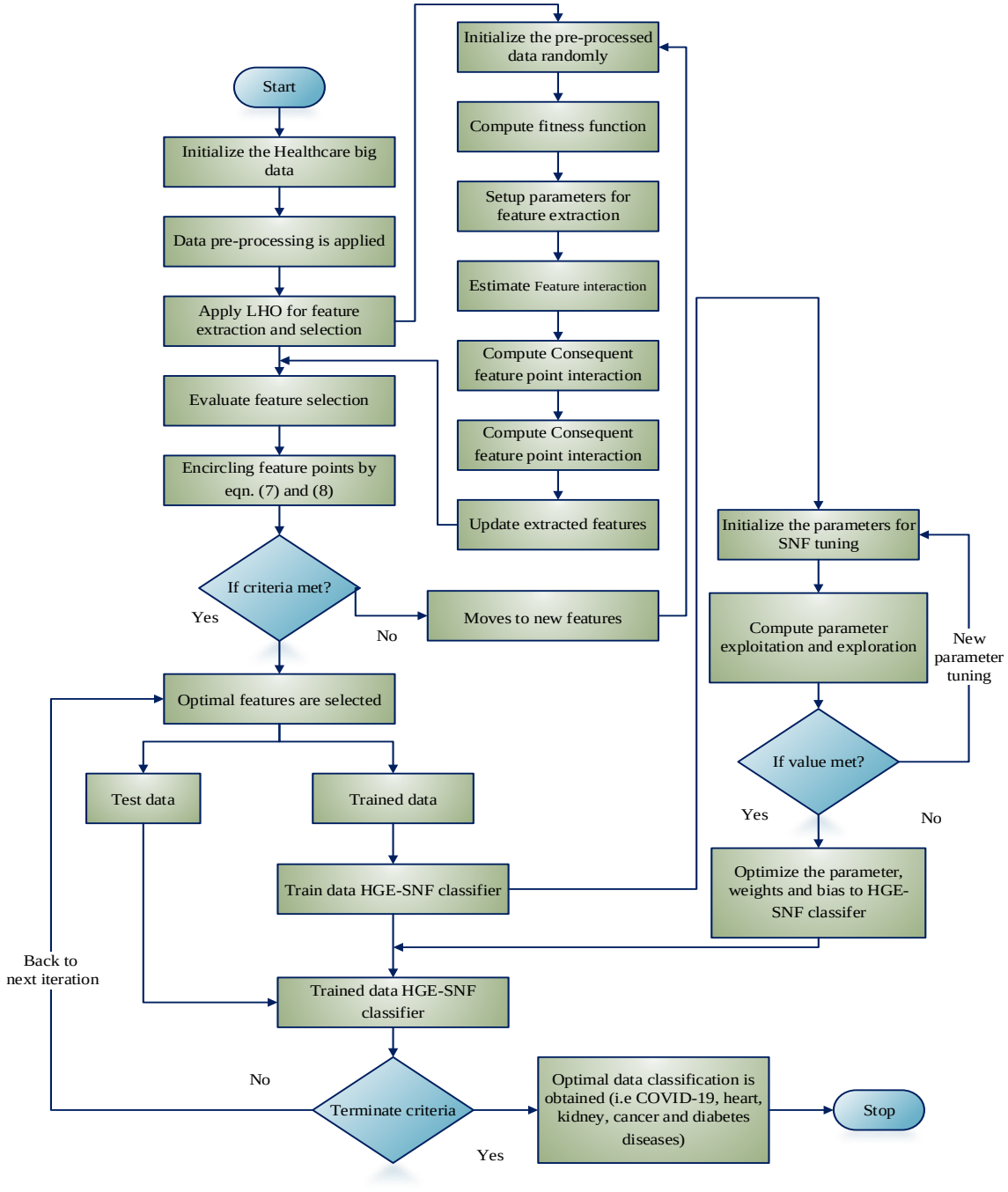


Figure 4.
The flowchart of proposed model in big data classification

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

5. Result and Discussion

This section included research to verify the effectiveness of the recommended technique. The suggested method is implemented with the MATLAB R2018b tool on a Windows 7 platform with 4GB RAM and Intel(R) dual-core. Math Works designed this numerical and technical programming language.

5.1. Dataset

Many IoT sources were obtained and compiled for the sake of our research. COVID-19, diabetes, cancer, heart, and renal disease data, as well as data from healthy people, are all included in this collection. First, for COVID-19, we examined data from archives such as the Khorshid COVID Cohort (KCC) study, which included 1634 patients with COVID-19 and comparable features that had a negative RT-PCR and chest CT scan. Furthermore, certain valuable websites, such as https://github.com/HarshCasper/Vyaadhi/tree/master/, include primary health care datasets that are regarded to be the most popular datasets. Datasets are a category of diseases that may be utilized to train the suggested method differentiate COVID-19. The normal dataset will be utilized as the final dataset in this investigation.

5.2. Performance Analysis

The proposed method of big data classification in healthcare system is executed using MATLAB 2018b software. In terms of determining the performance of the classifier of such recommended model in a big data classification problem, a variety of well-known assessment criteria can be used. The symbols $\vec{TP}$ and $\vec{TN}$ signify the numbers of correctly classified favourably and negatively data. The symbols $\vec{FP}$ and $\vec{FN}$ denote the erroneously categorised negative and positive data. The most often used classification performance parameter is accuracy. It is expressed as a percentage and reflects the proportion of incidents that have been correctly classified. For better classification results, the classification performance should be at or above 100%.

$$\text{Accuracy} = \frac{\vec{TN}+\vec{TP}}{\vec{TP}+\vec{TN}+\vec{FP}+\vec{FN}} \tag{27}$$

This exhibits the ability to find a patient who may be at risk for a range of illnesses. The ratio of true positives to the sum of true positives and false negatives, which is calculated using eqn. (28), is known as sensitivity.

$$\text{Sensitivity/Recall} = \frac{\vec{TP}}{\vec{TP}+\vec{FN}} \tag{28}$$

Specificity is defined as the ratio of a proportion of actual negatives to the sum of true negatives as well as false positives. The most specificity is 1.0, whereas the least specificity is 0.0.

$$\text{Specificity} = \frac{\vec{TN}}{\vec{TN}+\vec{FP}} \tag{29}$$

The F-measure is used to assess the efficacy of the validation procedure. It's a weighted average that takes both high accuracy and recall into consideration.

$$\text{F}-\text{measure} = \frac{2\vec{TP}}{2\vec{TP}+\vec{FP}+\vec{TN}} \tag{30}$$

This is the potential that a patient with a positive diagnostic test is infected with the illness.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

$$\text{Precision/ PPV} = \frac{\vec{TP}}{\vec{TP} + \vec{FP}} \quad (31)$$

The proportion of negative test outcomes analyzed is known as the Negative Predictive Value (NPV). This indicates the possibility of discovering a patient who is resistant to all diseases.

$$\text{NPV} = \frac{\vec{TN}}{\vec{TN} + \vec{FN}} \quad (32)$$

Furthermore, the Matthews' correlation coefficient (MCC) is defined as the proposed classifier's prediction capacity and ranges from -1 to +1. The result is positive if the classification correctly diagnoses the condition at the MCC level; otherwise, the result is negative, indicating that the classifier misidentified the ailment. The classifier produces an incorrect prediction when the MCC is close to zero. The MCC computation is evaluated using eqn. (33)

$$\text{MCC} = \frac{\vec{TP} \times \vec{TN} - \vec{FP} \times \vec{FN}}{\sqrt{\left(\vec{TP} + \vec{FP}\right)\left(\vec{TP} + \vec{FN}\right)\left(\vec{TN} + \vec{FP}\right)\left(\vec{FN} + \vec{TN}\right)}} \quad (33)$$

For a given analysis, the FPR is used to determine the probability of wrongly testing the null hypothesis. The FPR is calculated by dividing the proportion of false positives by the overall number of negatives.

$$\text{FPR} = \frac{\vec{FP}}{\vec{FP} + \vec{TN}} \quad (34)$$

The FNR is calculated by dividing the number of incorrect negative diagnoses by the total amount of negativity.

$$\text{FNR} = \frac{\vec{FN}}{\vec{FN} + \vec{TP}} \quad (35)$$

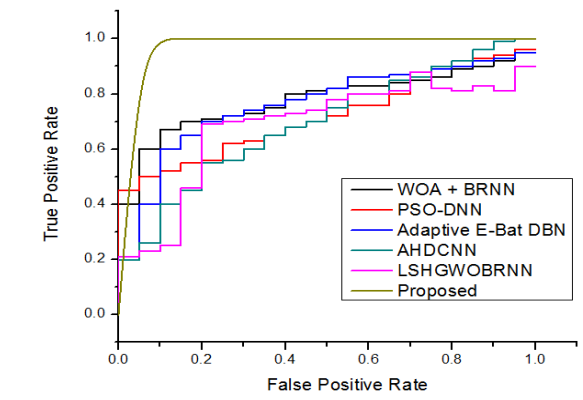
The AUC, which will be used to measure performance in this part of the experiment, is one of the most commonly utilized metrics in the situation of unbalanced class populations. Balancing performance is the AUC metric for a binary problem described by a certain point on the ROC curve.

$$\text{AUC} = \frac{\text{Sensitivit y} + \text{Specificit y}}{2} \quad (36)$$

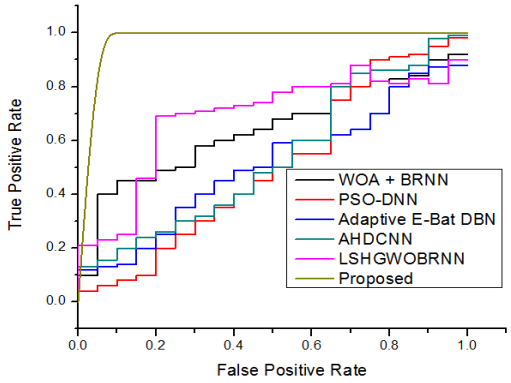
5.3. Comparative Analysis

The proposed approaches in big data classification outcomes in healthcare applications are compared to traditional methods such as WOA + BRNNHassib, et al. [33] PSO-DNN [34]. Adaptive E-Bat DBN[35]. AHDCNN Chen, et al. [36] and LSHGWOBRNN Hassib, et al. [37] in terms of different metrics. This study created Receiver Operating Characteristic (ROC) curves for each of the a) COVID-19, b) Diabetes, c) Cancer, d) Heart disease and e) Kidney disease are presented in Fig 5 (a-e). AUC includes the level or degree of reparability, whereas ROC is a probability curve.

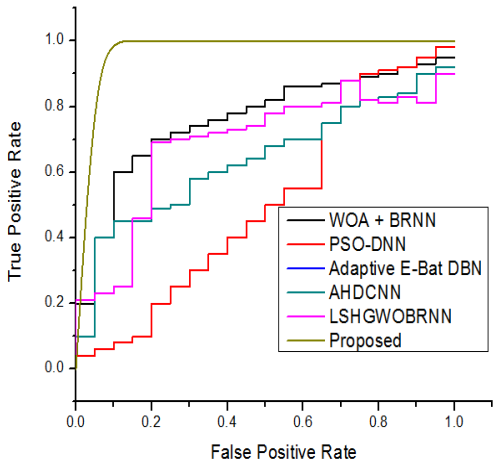
Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645



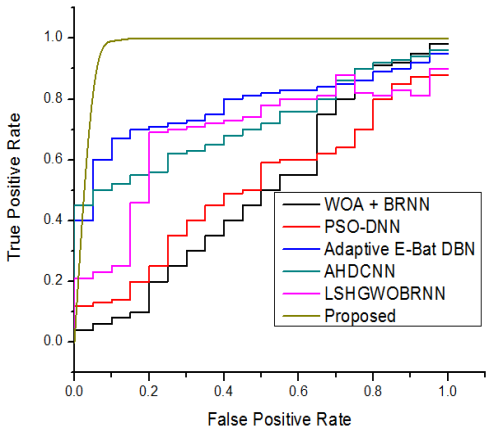
(a)



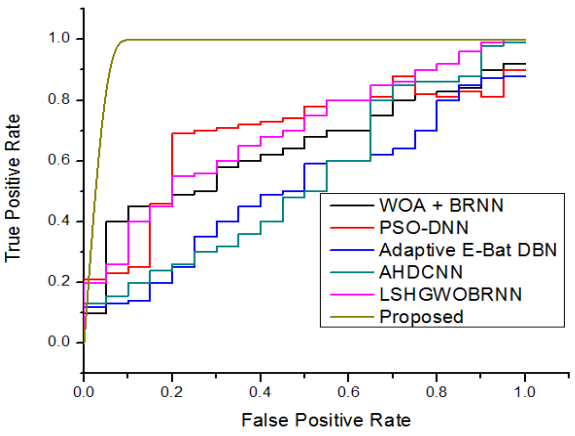
(b)



(c)



(d)



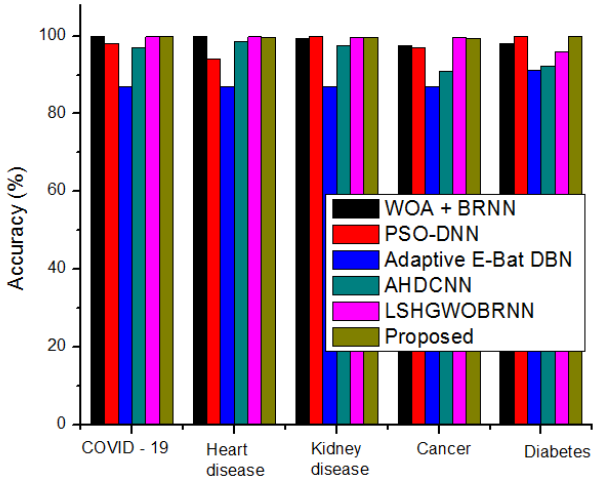
(e)

Figure 5.
ROC curve such a) COVID-19, b) Diabetes, c) Cancer, d) Heart disease and e) Kidney disease

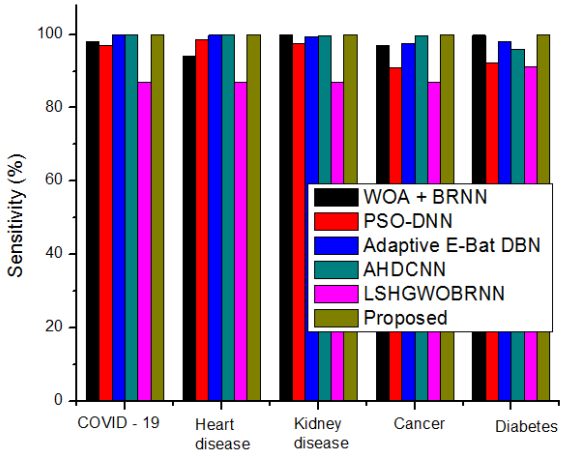
Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

It indicates how well the model can discriminate between classes. The AUC indicates how well the model predicts 0 classes as 0 and 1 course as 1. The higher the AUC, the superior the method is used to predict 0 categories as 0 and 1 category as 1. By comparison, the better the AUC, the finer the algorithm distinguishes between patients who have the disease and those who are not. Thus, the comparison shows that the proposed method has achieved superior performance over the conventional methods. Moreover, Sensitivity, specificity, accuracy, Area under Curve (AUC), Negative Predictive Value (NPV), False Positive Rate (FPR), F-Measure, False Negative for different disease classification by the proposed method has achieved finest performance over the conventional methods, which is illustrated in Figure 6 (a-e) and Figure 7(a-e).

The comparative analysis of the proposed accuracy over existing methods is illustrated in Figure 6 a). The proposed method achieved accuracy has been diagnosis with different diseases like Covid-19, heart disease, kidney diseases, cancer and diabetes, which is compared with the conventional methods like WOA + BRNN, PSO-DNN, Adaptive E-Bat DBN, AHDCNN, and LSHGWOBRNN. The COVID diseases diagnosis of the proposed approach has obtained higher accuracy as 99.8% over the existing WOA + BRNN, PSO-DNN, Adaptive E-Bat DBN, AHDCNN, and LSHGWOBRNN methods as 100%, 98.05%, 86.9%, 97%, 99.8%, and 99.8%. The heart diseases diagnosis of the proposed approach has obtained higher accuracy as 99.76% over the existing methods 99.8%, 94.058%, 87%, 98.5% and 99.7%. The kidney diseases diagnosis of the proposed approach has obtained higher accuracy as 99.9% over the existing methods as 97.6%, 97%, 86.9%, 91%, and 99.7%. The cancer diseases diagnosis of the proposed approach has obtained higher accuracy as 99.94% over the existing methods as 98.05%, 99.8%, 91.05%, 92.3%, and 96%. The diabetes diseases diagnosis of the proposed approach has obtained higher accuracy as 99.85 % over the existing methods as 98.05%, 99.8%, 91.05%, 92.3%, and 96%.



(a)



(b)

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

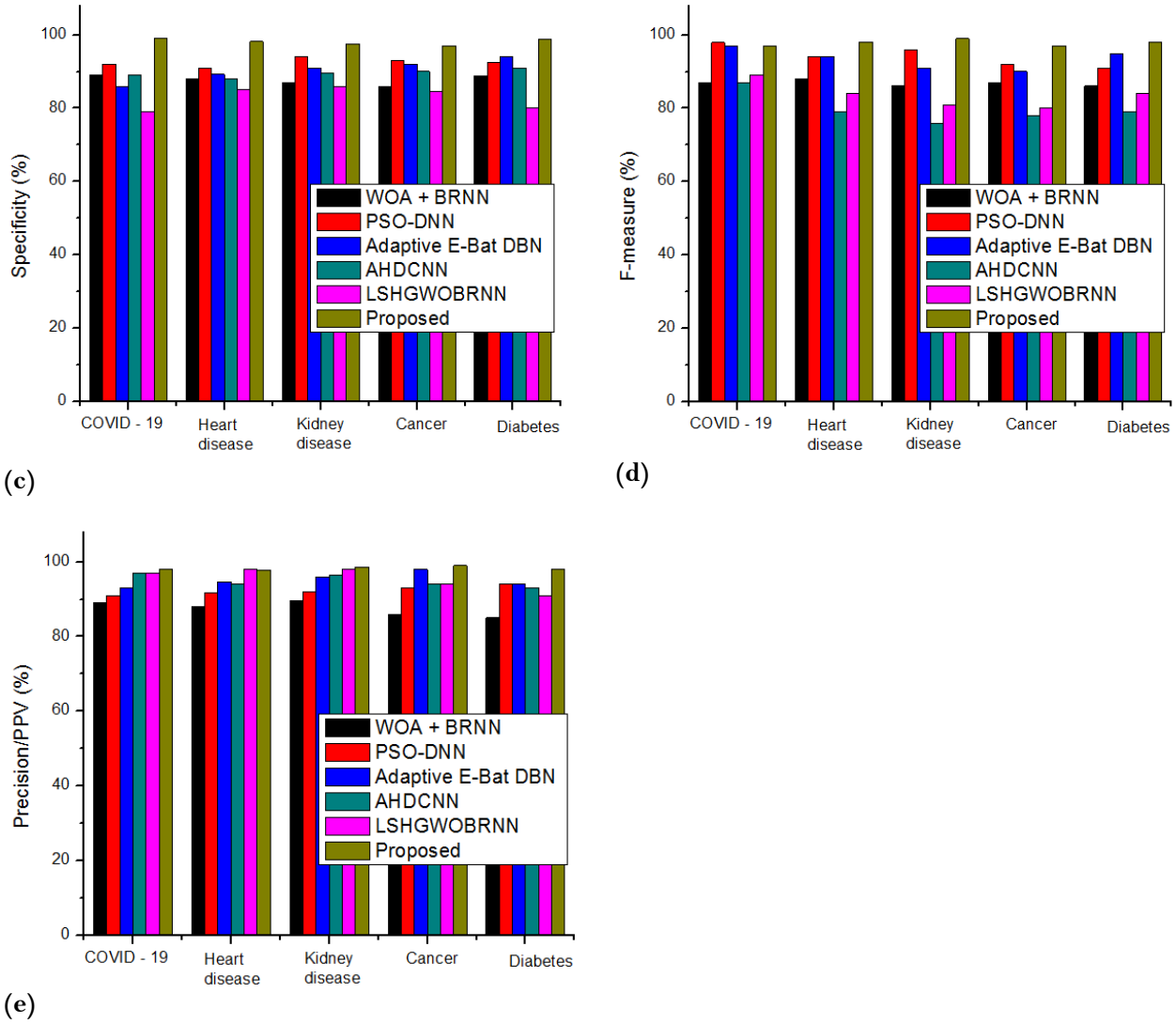


Figure 6. Comparative analysis of different metrics like a) Accuracy, b) Sensitivity/recall, c) Specificity, d) Precision/PPV and e) F-measure.

Figure 6 b) depicts a comparison of the suggested sensitivity/recall strategy with existing methods. In comparison to existing approaches, the suggested strategy provides a % sensitivity rate for COVID illness diagnosis. In comparison to existing approaches, the suggested strategy provides a greater sensitivity of 99.9% for heart disease detection. The suggested method for diagnosing kidney disorders has a sensitivity of 100 %, which is greater than existing approaches. In comparison to existing approaches, the suggested methodology provides 100 % sensitivity for cancer illness detection. The suggested method for diagnosing diabetic illnesses has a sensitivity of 99.98 %, which is greater than existing approaches. Furthermore, Figure 6 c) depicts a comparative examination of the suggested specificity with existing approaches. The suggested strategy for diagnosing COVID disorders has a better specificity rate of 99.9 % when compared to existing approaches. The proposed strategy for diagnosing cardiac problems has a greater specificity of 100 % when compared to existing approaches. The suggested method for diagnosing kidney disorders has a greater specificity of 100 % when compared to existing approaches. The suggested method for diagnosing cancer disorders has a greater

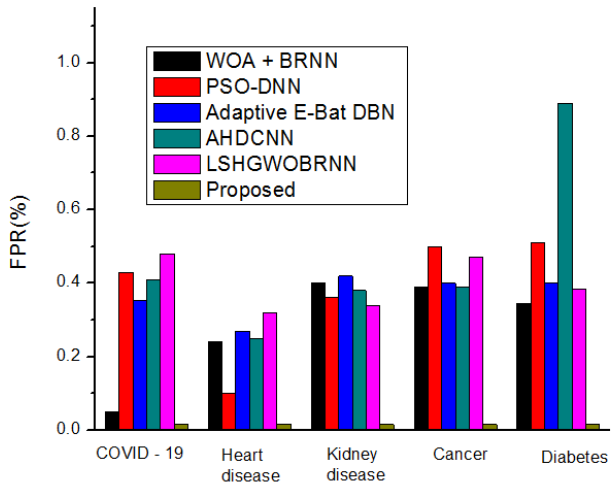
Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

specificity of 99.98% when compared to existing approaches. The suggested method for diagnosing diabetic illnesses has achieved a sensitivity of 100 % when compared to existing approaches.

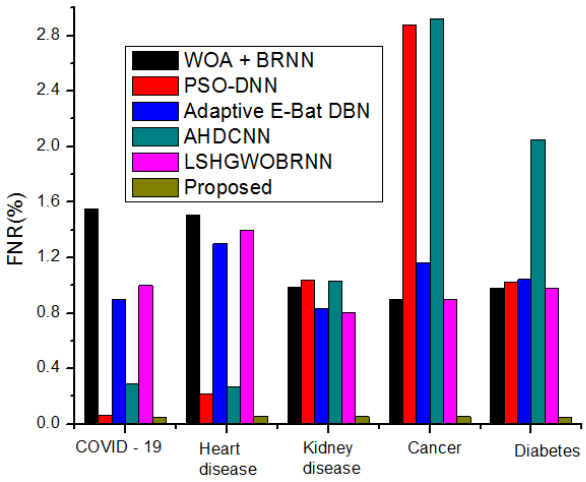
The comparative analysis of the proposed precision/PPV over existing methods is illustrated in figure 6 d). The COVID diseases diagnosis of the proposed approach has obtained higher precision rate as 100% over the existing methods. The heart diseases diagnosis of the proposed approach has obtained higher precision/PPV as **99.9**%, over the existing methods. The kidney diseases diagnosis of the proposed approach has obtained higher precision/PPV as 99.9% over the existing methods. The cancer diseases diagnosis of the proposed approach has obtained higher precision/PPV as 95.06% over the existing methods. The diabetes diseases diagnosis of the proposed approach has obtained higher precision/PPV as 99.9 % over the existing methods.

Figure 6 e) depicts a comparison of the proposed F-measure to existing approaches. In comparison to existing approaches, the suggested strategy has a higher F-measure rate of 97 % for COVID illness diagnosis. In comparison to existing approaches, the suggested strategy offers a higher F-measure of 98.06 % for heart disease diagnosis. In comparison to existing approaches, the suggested strategy offers a higher F-measure of 99 % for renal disease detection. In comparison to existing methodologies, the suggested strategy has a higher F-measure of 97.089 % for cancer illness diagnosis. In comparison to existing approaches, the suggested strategy has a higher F-measure of 99.09 % for diabetic illness diagnosis.

The proposed method achieved FPR has been diagnosis with different diseases like Covid-19, heart disease, kidney diseases, cancer and diabetes, which is compared with the conventional methods like WOA + BRNN, PSO-DNN, Adaptive E-Bat DBN, AHDCNN, and LSHGWOBRNN. The comparative analysis of the proposed FPR, FNR, MCC, AUC and NPV over existing methods is illustrated in figure 7 (a-e). In comparison to existing approaches, the suggested strategy provides a 0.0155 % FPR rate for Covid-19, heart disease, kidney diseases, cancer and diabetes illness diagnosis, which is lower than existing approaches.



(a)



(b)

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

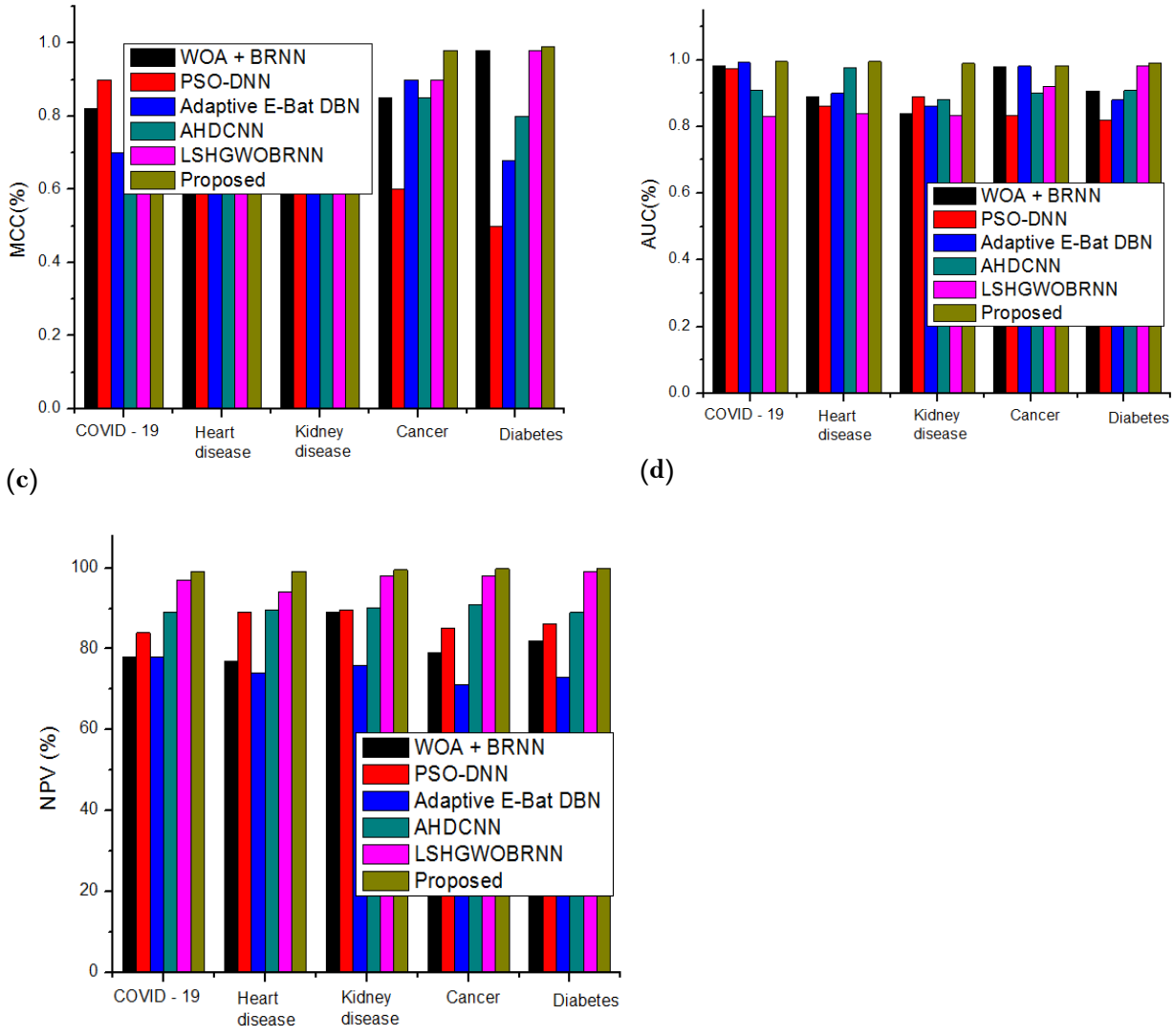


Figure 7.
Comparative analysis of different metrics like a) FPR, b) FNR, c) MCC d) AUC and e) NPV

In comparison to existing approaches, the suggested strategy provides a 0.995 % FNR rate for Covid-19, heart disease, kidney diseases, cancer (0.983%) and diabetes (0.99%) illness diagnosis, which is lower than existing approaches. Furthermore, the MCC and AUC values are also improved by the proposed approach for the optimal classification; the suggested strategy provides a 0.995 % MCC rate for Covid-19, heart disease (0.995%), kidney diseases (0.995%), cancer (0.98%) and diabetes (0.99%) illness diagnosis, which is superior to existing approaches. Consequently, the classicization performance of AUC obtained as for Covid-19, heart disease (0.99%), kidney diseases (0.985%), cancer (0.98%) and diabetes (0.99%) illness diagnosis, which is lower than existing approaches. This consequence shows the effective performance of proposed approach in healthcare applications.

5.4. Discussion

The suggested LHO with HGE-SNF approach techniques have achieved the best performance in healthcare system Big data classification with regard to different training stages and illnesses, according

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

to the findings from Tables 1, Table 2, Table 3, Table 4 and Table.5. When compared to traditional methodologies, the suggested method has attained the highest positive rates for all of the performance indicators. The simulation results demonstrate that the LHO with HGE-SNF model outperformed the degree of competition a superior solution for COVID-19 disease classification, with accuracy of 100%, precision of 100%, sensitivity of 100%, specificity of 99.9%, AUC of 0.995 %, F-Measure of 97 % , FPR of 0.0155 %, FNR of 0.995 %, MCC of 0.98 %, and NPV of 99 % are shown in table.1.

For Heart disease classification, Accuracy of 99.7%, Precision of 99.9%, Sensitivity of 99.9%, Specificity of %, AUC of 0.995 % , F-Measure of 98.06 %, FPR of 0.0155 %, FNR of 0.995 %, MCC of 0.99 %, and NPV of 99.1 % were also obtained in Table.2, which is compared to the conventional methods and shows the effective performance of proposed model in heart diseases classification.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

Table 1.
Performance analysis of big data classification for COVID-19 diseases

Methods	Accuracy	Precision	Sensitivity	Specificity	AUC	F-Measure	FPR	FNR	MCC	NPV
WOA + BRNN Hassib, et al. [33]	99.8	98.05	98.05	94.058	0.982	87	0.05	0.982	0.82	78
PSO-DNN Lakshmanaprabu, et al. [34]	98.05	97	97	98.5	0.974	98	0.43	0.974	0.9	84
Adaptive E-Bat DBN Mujeeb, et al. [35]	86.9	100	100	99.8	0.992	97	0.354	0.992	0.7	78
AHDCNN Chen, et al. [36]	97	99.8	99.8	99.76	0.908	87	0.41	0.908	0.82	89
LSHGWOBRNN Hassib, et al. [37]	99.8	86.9	86.9	87	0.83	89.05	0.48	0.83	0.85	97.08
Proposed	100	100	100	99.9	0.995	97	0.0155	0.995	0.98	99

Table 2.
Performance analysis of big data classification for heart diseases

Methods	Accuracy	Precision	Sensitivity	Specificity	AUC	F-Measure	FPR	FNR	MCC	NPV
WOA + BRNN Hassib, et al. [33]	99.01	94.058	94.058	97	0.89	88	0.24	0.89	0.64	76.9
PSO-DNN Lakshmanaprabu, et al. [34]	94.058	98.5	98.5	91	0.86	94	0.1	0.86	0.8	89
Adaptive E-Bat DBN Mujeeb, et al. [35]	87	99.8	99.8	97.6	0.9	94	0.27	0.9	0.63	74
AHDCNN Chen, et al. [36]	98.5	99.76	99.76	99.7	0.975	79	0.25	0.975	0.64	89.6
LSHGWOBRNN Hassib, et al. [37]	99.4	87	87	86.9	0.84	84	0.32	0.84	0.65	94.06
Proposed	99.7	99.9	99.9	100	0.995	98.06	0.0155	0.995	0.99	99.1

Table 3.
Performance analysis of big data classification for kidney disease

Methods	Accuracy	Precision	Sensitivity	Specificity	AUC	F-Measure	FPR	FNR	MCC	NPV
WOA + BRNN Hassib, et al. [33]	99.4	99.81	99.81	99.05	0.84	86.3	0.4	0.84	0.8	89
PSO-DNN Lakshmanaprabu, et al. [34]	99.1	97.6	97.65	97	0.89	96	0.36	0.89	0.76	89.6
Adaptive E-Bat DBN Mujeeb, et al. [35]	86.93	99.4	99.41	100	0.86	91	0.42	0.86	0.79	76
AHDCNN Chen, et al. [36]	97.6	99.7	99.7	99.8	0.88	76	0.38	0.88	0.8	90

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

LSHGWOBRNN Hassib, et al. [37]	99	86.93	86.13	86.9	0.834	81	0.34	0.834	0.79	98.08
Proposed	99.6	99.9	100	100	0.99	99	0.0155	0.995	0.985	99.5

Table 4.
Performance analysis of big data classification for cancer

Methods	Accuracy	Precision	Sensitivity	Specificity	AUC	F-Measure	FPR	FNR	MCC	NPV
WOA + BRNN Hassib, et al. [33]	97.6	84.06	97	99.8	0.98	97	0.39	0.98	0.85	79
PSO-DNN Lakshmanaprabu, et al. [34]	97	92.06	91	92.3	0.832	92	0.5	0.832	0.6	85
Adaptive E-Bat DBN Mujeeb, et al. [35]	86.9	90.68	97.6	98.05	0.981	90	0.4	0.981	0.9	71
AHDCNN Chen, et al. [36]	91	87.05	99.7	96	0.9	78.06	0.39	0.9	0.85	91.02
LSHGWOBRNN Hassib, et al. [37]	99.1	78.06	86.9	91.05	0.92	80.06	0.47	0.92	0.9	98
Proposed	99.9	95.06	100	99.98	0.983	97.089	0.0155	0.983	0.98	99.84

Table 5.
Performance analysis of big data classification for diabetes

Methods	Accuracy	Precision	Sensitivity	Specificity	AUC	F-Measure	FPR	FNR	MCC	NPV
WOA + BRNN Hassib, et al. [33]	98.05	81.06	99.8	99.81	0.907	86.058	0.345	0.907	0.98	82
PSO-DNN Lakshmanaprabu, et al. [34]	99.8	86.06	92.3	97.6	0.82	91	0.51	0.82	0.5	86.05
Adaptive E-Bat DBN Mujeeb, et al. [35]	91.05	88	98.05	99.4	0.88	95	0.4	0.88	0.68	73
AHDCNN Chen, et al. [36]	92.3	82	96	99.7	0.908	79.08	0.8	0.908	0.8	89
LSHGWOBRNN Hassib, et al. [37]	96	76.9	91.05	86.93	0.981	84.06	0.385	0.981	0.98	99
Proposed	99.85	99.9	99.98	100	0.99	98.09	0.0155	0.99	0.99	99.84

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

Then, for renal disease classification and comparative analysis expressions are provided in Table.3, accuracy of 99.6%, precision of 99.9%, sensitivity of 100%, specificity of 100%, AUC of 0.99 %, F-Measure of 99 % , FPR of 0.0155 %, FNR of 0.995 %, MCC of 0.985 % , and NPV of 99.5 % are obtained.

Furthermore, for cancer disease classification and its comparative analysis is expressed in Table.4, accuracy of 99.4 % , precision of 95.06 % , sensitivity of 100 % , specificity of 99.98 % , AUC of 0.983 % , F-Measure of 97.089 % , FPR of 0.0155 % , FNR of 0.983 % , MCC of 0.98 % , and NPV of 99.84 % were achieved. Finally, for diabetes classification comparative analysis for the proposed and existing models are detailed in Table.5, accuracy of 99.85%, precision of 99.9%, sensitivity of 99.98%, specificity of 100%, AUC of 0.99 % , F-Measure of 98.09 % , FPR of 0.0155 % , FNR of 0.99 % , MCC of 0.99 % , and NPV of 99.84 % were obtained.

6. Conclusion

This work provided an effective big data health application system based on LHO and HGE-SNF classifier models. To assure the outcome of the LHO with HGE-SNF approach employing a multivariate disease sample, a detailed experimentation procedure is carried out. The simulation results revealed that the LHO with HGE-SNF model outperformed the competition by providing a superior solution as Accuracy of 100%, Precision of 100%, Sensitivity of 100%, Specificity of 99.9%, AUC of 0.995%, F-Measure of 97%, FPR of 0.0155%, FNR of 0.995%, MCC of 0.98% and NPV of 99% is achieved for COVID-19 disease classification. Also, Accuracy of 99.7%, Precision of 99.9%, Sensitivity of99.9%, Specificity of 100%, AUC of 0.995%, F-Measure of 98.06%, FPR of 0.0155%, FNR of 0.995%, MCC of 0.99% and NPV of 99.1% is achieved for Heart disease classification. Then, Accuracy of 99.6%, Precision of 99.9%, Sensitivity of100%, Specificity of 100%, AUC of 0.99%, F-Measure of 99%, FPR of 0.0155%, FNR of 0.995%, MCC of 0.985% and NPV of 99.5% is achieved for kidney disease classification. Furthermore, Accuracy of 99.9%, Precision of 95.06%, Sensitivity of100%, Specificity of 99.98%, AUC of 0.983%, F-Measure of 97.089%, FPR of 0.0155%, FNR of 0.983%,

MCC of 0.98% and NPV of 99.84% is achieved for cancer disease classification. Finally, Accuracy of 99.85%, Precision of 99.9%, Sensitivity of 99.98%, Specificity of 100%, AUC of 0.99%, F-Measure of 98.09%, FPR of 0.0155%, FNR of 0.99%, MCC of 0.99% and NPV of 99.84% is achieved for diabetes classification. While compared to all this performance metrics analysis shows that the proposed method has achieved superior function in healthcare big data classification. This approach can store and analyses large amounts of healthcare data, extract useful features, and offer semantic interpretations for those characteristics in order to enhance health condition categorization performance. This paradigm might be beneficial in the healthcare industry for monitoring prolonged patients utilizing data from multiple sources in this context. The suggested big data analytics engine finds important data, extracts relevant characteristics from healthcare data, minimizes data dimensionality, and classifies illnesses automatically. Apart from these disorders, the HGE-SNF model given here might be applied to the diagnosis of other diseases in the future.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

References

[1] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: Management, analysis and future prospects," *Journal of Big Data*, vol. 6, no. 1, pp. 1-25, 2019. https://doi.org/10.1186/s40537-019-0217-0

[2] V. Palanisamy and R. Thirunavukarasu, "Implications of big data analytics in developing healthcare frameworks–A review," *Journal of King Saud University-Computer and Information Sciences*, vol. 31, no. 4, pp. 415-425, 2019. https://doi.org/10.1016/j.jksuci.2017.12.007

[3] A. Banerjee, C. Chakraborty, A. Kumar, and D. Biswas, "Emerging trends in IoT and big data analytics for biomedical and health care technologies," *Handbook of Data Science Approaches for Biomedical Engineering*, pp. 121-152, 2020.

[4] Z. Allam and Z. A. Dhunny, "On big data, artificial intelligence and smart cities," *Cities*, vol. 89, pp. 80-91, 2019.

[5] V. K. Chattu, "A review of artificial intelligence, big data, and blockchain technology applications in medicine and global health," *Big Data and Cognitive Computing*, vol. 5, no. 3, p. 41, 2021.

[6] R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, "Artificial Intelligence (AI) applications for COVID-19 pandemic," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 14, no. 4, pp. 337-339, 2020. https://doi.org/10.1016/j.dsx.2020.04.012

[7] S. Kumar, R. D. Raut, and B. E. Narkhede, "A proposed collaborative framework by using artificial intelligence-internet of things (AI-IoT) in COVID-19 pandemic situation for healthcare workers," *International Journal of Healthcare Management*, vol. 13, no. 4, pp. 337-345, 2020. https://doi.org/10.1080/20479700.2020.1810453

[8] A. Kishor and C. Chakraborty, "Artificial intelligence and internet of things based healthcare 4.0 monitoring system," *Wireless Personal Communications*, vol. 127, no. 2, pp. 1615-1631, 2022. https://doi.org/10.1007/s11277-021-08708-5

[9] S. Kaur *et al.*, "Medical diagnostic systems using artificial intelligence (ai) algorithms: Principles and perspectives," *IEEE Access*, vol. 8, pp. 228049-228069, 2020. https://doi.org/10.1109/access.2020.3042273

[10] N. Norori, Q. Hu, F. M. Aellen, F. D. Faraci, and A. Tzovara, "Addressing bias in big data and AI for health care: A call for open science," *Patterns*, vol. 2, no. 10, p. 100347, 2021.

[11] C. Chhabra and S. Meghna, "Machine learning, deep learning and image processing for healthcare: A crux for detection and prediction of disease," in *Proceedings of Data Analytics and Management. Springer, Singapore, 2022. 305-325*, 2022.

[12] E. Mbunge and B. Muchemwa, "Towards emotive sensory Web in virtual health care: Trends, technologies, challenges and ethical issues," *Sensors International*, vol. 3, p. 100134, 2022.

[13] J. B. Awotunde, "Artificial intelligence and an edge-iomt-based system for combating covid-19 pandemic intelligent interactive multimedia systems for e-healthcare applications." Singapore: Springer, 2022, pp. 191-214.

[14] R. Wang *et al.*, "Artificial intelligence for prediction of COVID-19 progression using CT imaging and clinical data," *European Radiology*, vol. 32, pp. 205-212, 2022. https://doi.org/10.1007/s00330-021-08049-8

[15] A. Sujith, G. S. Sajja, V. Mahalakshmi, S. Nuhmani, and B. Prasanalakshmi, "Systematic review of smart health monitoring using deep learning and Artificial intelligence," *Neuroscience Informatics*, vol. 2, no. 3, p. 100028, 2022.

[16] S. Triberti, "Artificial intelligence in healthcare practice: How to tackle the "human" challenge handbook of artificial intelligence in healthcare." Cham: Springer, 2022, pp. 43-60.

[17] M. Chugh, J. Rahul, and G. Anmol, *MATHS: Machine Learning techniques in healthcare system international conference on innovative computing and communications*. Singapore: Springer, 2022.

[18] H. Ahmed, E. M. Younis, A. Hendawi, and A. A. Ali, "Heart disease identification from patients' social posts, machine learning solution on Spark," *Future Generation Computer Systems*, vol. 111, pp. 714-722, 2020.

[19] S. Nazari, M. Fallah, H. Kazemipoor, and A. Salehipour, "A fuzzy inference-fuzzy analytic hierarchy process-based clinical decision support system for diagnosis of heart diseases," *Expert Systems with Applications*, vol. 95, pp. 261-271, 2018. https://doi.org/10.1016/j.eswa.2017.11.001

[20] K. Mittal, G. Aggarwal, and P. Mahajan, "Performance study of K-nearest neighbor classifier and K-means clustering for predicting the diagnostic accuracy," *International Journal of Information Technology*, vol. 11, pp. 535-540, 2019.

[21] G. Manogaran and D. Lopez, "Health data analytics using scalable logistic regression with stochastic gradient descent," *International Journal of Advanced Intelligence Paradigms*, vol. 10, no. 1-2, pp. 118-132, 2018. https://doi.org/10.1504/ijaip.2018.10010530

[22] B. V. Ravindra, N. Sriraam, and M. Geetha, "Chronic kidney disease detection using back propagation neural network classifier," presented at the International Conference on Communication, Computing and Internet of Things (IC3IoT), 2018.

[23] S. Kaul and K. Yogesh, "Nature-inspired metaheuristic algorithms for constraint handling: Challenges, issues, and research perspective constraint handling in metaheuristics and applications." Singapore: Springer, 2021, pp. 55-80.

[24] W. Xing and Y. Bei, "Medical health big data classification based on KNN classification algorithm," *IEEE Access*, vol. 8, pp. 28808-28819, 2019.

[25] S. Lakshmanaprabu, K. Shankar, M. Ilayaraja, A. W. Nasir, V. Vijayakumar, and N. Chilamkurti, "Random forest for big data classification in the internet of things using optimal features," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 10, pp. 2609-2618, 2019.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645

[26] F. Ali *et al.*, "An intelligent healthcare monitoring framework using wearable sensors and social networking data," *Future Generation Computer Systems*, vol. 114, pp. 23-43, 2021.
[27] R. Thanga Selvi and I. Muthulakshmi, "RETRACTED ARTICLE: An optimal artificial neural network based big data application for heart disease diagnosis and classification model," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 6, pp. 6129-6139, 2021. https://doi.org/10.1007/s12652-020-02181-x
[28] P. Galetsi, K. Katsaliaki, and S. Kumar, "Big data analytics in health sector: Theoretical framework, techniques and prospects," *International Journal of Information Management*, vol. 50, pp. 206-216, 2020.
[29] S. Oniani, "Artificial intelligence for internet of things and enhanced medical systems bio-inspired neurocomputing." Singapore: Springer, 2021, pp. 43-59.
[30] R. Mukherjee *et al.*, "IoT-cloud based healthcare model for COVID-19 detection: an enhanced k-Nearest Neighbour classifier based approach," *Computing*, pp. 1-21, 2023. https://doi.org/10.1007/s00607-021-00951-9
[31] S. Elghamrawy, "An H 2 O's deep learning-inspired model based on big data analytics for coronavirus disease (covid-19) diagnosis big data analytics and artificial intelligence against covid-19: innovation vision and approach." Cham: Springer, 2020, pp. 263-279.
[32] T. A. Ahanger, "A novel IoT–fog–cloud-based healthcare system for monitoring and predicting COVID-19 outspread," *The Journal of Supercomputing* pp. 1-24, 2021.
[33] E. M. Hassib, A. I. El-Desouky, L. M. Labib, and E.-S. M. El-Kenawy, "WOA+ BRNN: An imbalanced big data classification framework using Whale optimization and deep neural network," *Soft Computing*, vol. 24, no. 8, pp. 5573-5592, 2020. https://doi.org/10.1007/s00500-019-03901-y
[34] S. Lakshmanaprabu, S. N. Mohanty, S. Krishnamoorthy, J. Uthayakumar, and K. Shankar, "Online clinical decision support system using optimal deep neural networks," *Applied Soft Computing*, vol. 81, p. 105487, 2019.
[35] S. M. Mujeeb, R. Praveen Sam, and K. Madhavi, "Adaptive exponential bat algorithm and deep learning for big data classification," *Sādhanā*, vol. 46, no. 1, p. 15, 2021. https://doi.org/10.1007/s12046-020-01521-z
[36] G. Chen *et al.*, "Prediction of chronic kidney disease using adaptive hybridized deep convolutional neural network on the internet of medical things platform," *IEEE Access*, vol. 8, pp. 100497-100508, 2020.
[37] E. M. Hassib, A. I. El-Desouky, E.-S. M. El-Kenawy, and S. M. El-Ghamrawy, "An imbalanced big data mining framework for improving optimization algorithms performance," *IEEE Access*, vol. 7, pp. 170774-170795, 2019.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1593-1620, 2024
DOI: 10.55214/25768484.v9i3.5645