



Performance Enhancement of Intelligent Healthcare-Based Recommendation System with IFODNN Model

Gauri Sood^{1,*}, Neeraj Raheja²

¹Research Scholar, MaRaharishi Markandeshwar Engineering College, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, India

²Assoc. Prof, Maharishi Markandeshwar Engineering College, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, India

Emails: er.gauri31sood@gmail.com; neeraj_raheja2003@mmumullana.org

Abstract

Several diseases have been identified as fatal conditions affecting individuals during their middle and old ages worldwide. In recent years, chronic and pulmonary diseases have exhibited the highest mortality rates among all known conditions in this category. Machine Learning (ML) tools have efficiently studied the causes of these harmful diseases, including analyzing large databases. These databases may contain unreliable and redundant features that affect prediction accuracy and speed. Applying the feature-based extraction and selection methods to remove inconsistent components is essential. This article implements a deep neural network (DNN) technique for diagnosis to classify different diseases. However, the DNN model faces a challenge, specifically hallucination, in accurately classifying diseases. To overcome this, a hybrid optimization DNN model has been introduced. This model is useful for recommending treatments based on the diagnosed diseases. The hybrid optimization DNN model, referred to as the Intelligent Healthcare Recommendation (IHCR) model, is designed to predict and recommend treatments for chronic and pulmonary disease patients. The research model effectively extracts features at a specific level and selects valuable features to provide accurate recommendations. This recommendation phase is followed by a statistical analysis based on probability, which evaluates patients' risk levels. Reliant on the data from the risk analysis (RA), patients are given recommendations regarding the severity and performance of the related diseases for early treatment. The proposed work has been estimated using dissimilar databases based on multi-diseases, and the outcomes appear encouraging. This research aims to develop the IHCR model for chronic and pulmonary diseases. The performance of the implemented recommendation models is evaluated using parameters like RMSE, specificity (SP), sensitivity (SN), and accuracy (Acc). The results of the recommendation model show an Acc of 96.81–97%.

Received: November 20, 2024 Revised: January 27, 2025 Accepted: February 27, 2025

Keywords: Intelligent Healthcare recommender model (IHCR); Independent Feature optimized Deep neural network (IFODNN) Classification Model; Statistical analysis-based recommender system; Grey wolf optimizer (GWO); Deep neural network (DNN); Independent component analysis (ICA)

1. Introduction

Currently, most of our actions are saved and recorded as data. This data can be used to anticipate individuals' actions and behaviors by analyzing relevant historical information associated with them. Doctors also treat patients using background data that accounts for distinct clinical parameters. The healthcare system (HCS) organizes information to manage data for specific diseases and make suitable arrangements. The absence or presence of a condition can be determined by detecting precise health-based factors [1]. The main challenge lies in evaluating these parameters and determining the risk level/severity. Once this assessment is made, suitable recommendations can be provided, aiming to either treat the disease effectively or reduce the associated risks. Pulmonary diseases (PDs) and chronic diseases (CDs) affect a day-to-day human lifestyle, adding family history, and diet. One of the advanced research topics requiring enormous effort to reply to CDs and PDs effectively is e-health. E-health

depends on methods implemented for a global variety of patients remotely located and precious by different CDs and PDs [2]. Between CDs and PDs, heart disease (HD) and Lung diseases (LDs) are the key reasons for death globally.

Generally, a huge amount of older individuals are found to be affected by HDs. In recent years, this ratio has significantly shifted, with HDs and LDs now often detected in younger generations. Patients in urban areas can access healthcare and receive treatment for life-threatening diseases, but people in remote regions struggle due to limited access to healthcare services. E-health has developed as a talented explanation to optimize death, costs, and illnesses connected to heart and lung conditions [3]. However, a major limitation of e-health solutions is the challenge of reliably collecting and managing data efficiently. Remote monitoring systems (MSs) are considered robust and effective if they can intelligently analyze patient data and generate reports accordingly [4]. Medical histories are used for preventive actions aimed at minimizing the impact of medical conditions. Many previously isolated health MSs lack advanced analytical capabilities. In the current analysis, health MSs have evolved to be used in a more dynamic way, rather than for simple tasks like tracking sleep hours [5]. Health monitoring (HM) methods are now utilized to procedure records more reliably for the end-user.

In the current analysis, ML methods have been utilized to handle more complex data in key areas such as classification and diagnosis that can also be applied to alarm-based ML methods and administrative processes. These methods are utilized to diagnose input data across distinct categories, now accessible for specific diseases [6]. However, obtaining information in an effective and precise manner that supports personal decision-making for diagnosing diseases remains limited. The primary objective of e-health systems is to reduce risks to human life by classifying numerous diseases at an early-stage and providing accurate classifications and recommendations. Several studies have been conducted to identify specific conditions at different medical stages. However, it is crucial to offer personalized health classification and advice to patients based on their current medical condition [7, 8].

A Recommender System (RS) is a sub-category of data analytics that provides accurate and personalized data for a specific field. Nowadays, when correlated with national emergencies, RSs can play a significant role in enhancing facilities in the medical field. Numerous studies have been conducted in this domain of Healthcare Recommender Systems (HRS), with static systems remaining a key focus of analysis. The primary aim of these studies has been to demonstrate applications in areas such as record tracking, hospital management, health monitoring, and more. However, various challenges still need to be addressed in the healthcare sector, including limited bandwidth, specialized tools, and a lack of supportive infrastructure. Additionally, issues related to data loss remain critical. This research work focuses on the development of an intelligent healthcare recommendation (IHRS) framework that utilizes heart and lung disease diagnosis/classification tools to identify risks associated with these diseases. The system then applies a recommendation method to analyze patients' healthcare records. A Deep Neural Network (DNN) is employed for the illness diagnosis process, followed by statistical analysis that takes input from the diagnosis step and evaluates the risk levels. Finally, the RS advises patients based on the risk levels predicted by the statistical analysis system.

The research method follows these steps:

1. This research work is collecting the database from the online repository (Kaggle) for heart and lung diseases. This database information is labeled by doctors for e-healthcare RS.
2. Feature extraction, selection, and classification are developed for the analysis of several diseases (heart, and lung).
3. A DNN model is used to classify multiple diseases. However, the classification module faces several challenges, such as hallucination (i.e., inaccurate feature extraction and selection), a high error rate, and low accuracy. Further efforts are being made to explore additional data related to this project, then finalize the hybrid optimization DNN model, referred to as the Independent feature-based optimized DNN (IFODNN) and other named IHCR model, which is designed to predict and recommend treatments for chronic and pulmonary disease patients. The DNN model is used for disease classification, followed by the IHCR model which, takes input from the classification procedure and evaluates the risk level. Lastly, the risk of classified disease is evaluated for producing reliable healthcare advice or recommendations.
4. The presentation of the implemented RS is calculated by considering metrics such as training acc, RMSE, SP, SN, F1-Score, etc., and comparing the presentation of the implemented RS with other current RSs.

The research article comprises the subsequent sections. In Section 2, prior work related to prediction, feature extraction, diagnosis/classification, and recommendation is presented. Section 3 discusses an outline of the implemented work and is categorized into two sections. (i) Explain the applied steps. (ii) Explain the research methodology for giving healthcare recommendations. The exploration of database and simulation outcomes is discussed in section 4. Lastly, the conclusion and upcoming improvements are mentioned in Sect 5.

2. Related Work

This section has studied multiple healthcare diseases, such as those affecting people with diabetes, heart disease, cancer, chronic diseases, etc., by using different ML, DL, and AI methods. These technologies are helpful for healthcare disease classification and recommendation. **Youness et al. 2019 [9]** proposed the rapid correlation-based feature selection (CBF) technique to evaluate terminated feature sets to improve the superiority of heart disease classification (HDC). So, it performs a category depending on several classification methods, comprising K-nearest neighbour (KNN), Support vector machine (SVM), Naive Bayes (NB), Random forest (RF), multilayer perceptron (MLP), Artificial neural network (ANN), particle swarm optimization (PSO) merged with ant colony optimization (ACO) methods. The implemented hybrid technique was applied to the heart disease database; the outcomes demonstrate the robustness and efficiency of the implemented mixed model in processing several categories of data for HDC. So, this analyzed several ML methods and compared the outcomes using assessment parameters, like acc, recall, precision, and F1-Score. The highest classification accuracy rate of 99% utilizing the optimized model introduced by FCBF, PSO, and ACO methods. **Badhagouni et al. 2021 [10]** discussed predicting harmful diseases using ML classifier methods. This author works with heart, diabetes, and breast cancer (BC) was involved. This proposed work is practical for the mass public; the defined team conducts a health examination web application (WA) which creates classifications related to several diseases using the concept of ML. Generally, the motive to design a condition classifies WA that utilizes the ML concept based on categories related to several diseases such as BC, heart, etc. **Arumugam et al. 2023 [11]** described ML and data imagined diabetes based on heart illness was a form of heart illness that disturbs people with diabetes. Diabetes disease was considered a protracted disorder that was developed cause of insufficient insulin in the human body and failure to practice the insulin manufactured appropriately. Heart-related disease, frequently recognized as cardiovascular sickness, affects the heart or blood vessels. Although several classification methods were used for forecasting cardiovascular disease, there were insufficient statistics for forecasting heart sickness in a diabetic specific. Due to the decision tree (DT) model reliably hiding the other techniques, such as NB and SVM, and reached better performance. **Baha et al. 2021 [12]** discussed an intelligent healthcare recommendation system (HCRS) that classified and recommended diabetic disease (DD) precisely using optimized ML methods with the data fusion (DF) method on Healthcare databases. Several ML methods have been implemented in the past to classify DD. These systems must be able to handle the massive number of multiple characteristic datasets on DD completely. Depending on deep ML and DF standpoints, a smart HCRS was implemented for DD. Utilizing DF, they can eliminate the redundant load of scheme calculation capabilities and increase the research system's show to classify and recommend this serious disease. In the last, the ensemble ML approach was trained for diabetes prediction. This smart RS was estimated based on a definite diabetes database, and its performance was related to the most present improvements from the review. The research system reached 99 per cent acc, i.e., the maximum compared to the existing deep ML models. So, the research system was better for multidisciplinary DD classification and recommendation. **Fouad H Awad et al. (2023) [13]** described a robust big medical data detection and classification method using LR and YOLO4, resp. It improved the performance of these methods and implemented the progressive parallel K-means clustering method, which verified models and designs in the data. The implemented approach has calculated on different medical databases and defined that it might precisely huge amounts of medical datasets and classify medical images such as X-ray, MRI, etc. **Mahmoud et al. (2023) [14]** presented an involuntary arrhythmia classification method based on the combination of current meta-heuristic optimization (MHO) methods. The part of this optimization is attained outcomes that the presentation of all the analyzed classifiers was enhanced after mixing the MHO method with the average. ECG arrhythmia detection acc. attainment of 96.4% and an SN of 92%. Tables 1 and 2 describe the implemented method, comparative methods, existing research gaps or problems, implemented performance metrics with values, experimental tools used such as MATLAB, PYTHON, etc., and existing implemented technique outcomes.

Table 1: Analysis of the Existing Methods

Ref no.	Classification Methods	Comparison Methods	Problem/ Issues	Simulation Tools	Classifier Values	Outcomes
[9]	Fast Correlation-Based Feature selection (FCBF) technique	SVM, MLP, KNN, RF, NB	<ul style="list-style-type: none"> Clustering and classification problems. 	LibSVM and WEKA data mining tool	<ul style="list-style-type: none"> Accuracy = 99%, Precision = 0.9, Recall = 1, F1-Score = 0.9 	<ul style="list-style-type: none"> Remove irrelevant features with the FCBF technique.

[10]	Logistic Regression (LR)	SVM, KNN	<ul style="list-style-type: none"> Many deaths are owed to the absence of appropriate check-ups for illnesses. 	Python Flask API Framework	<ul style="list-style-type: none"> Acc with HDC = 83.84%, Acc with Diabetes = 77.6%, BC with accuracy = 94.5% 	<ul style="list-style-type: none"> LR method achieved high precision values for the disease classification of diabetes and BC.
[11]	Decision Tree (DT)	SVM, NB	<ul style="list-style-type: none"> Multi-class and large dimensional space for non-separable issues 	Weka	<ul style="list-style-type: none"> Error = 9, Accuracy = 90% 	<ul style="list-style-type: none"> DT method has attained optimized presentation in predicting the chance of HDC in diabetes patients.
[12]	SHRS-M3DP model	KNN, RF, DT, NB, SVM	<ul style="list-style-type: none"> Classification issues 	Computational tool	<ul style="list-style-type: none"> Accuracy = 99%, Precision = 67%, Recall = 55%, F1-measure = 61%, RMSE = 0.25, MAE = 0.50 	<ul style="list-style-type: none"> The implemented method effectiveness is classified and recommends whether the patient is the target of multi-disciplinary DD or not.
[13]	Advanced Parallel k-means clustering method	LR, NB	<ul style="list-style-type: none"> Binary classification issues 	C Programming	<ul style="list-style-type: none"> Accuracy = 96.21%, Precision = 99.7%, Recall = 93.8%, F1 = 97.4% 	<ul style="list-style-type: none"> It defined that the method was capable of efficiently organizing the health images.
[14]	Meta-heuristic optimization (MHO), SVM, GBDT, RF, and KNN	SVM, CNN, RF, KNN, LSTM-CNN	<ul style="list-style-type: none"> Classification issues 	MATLAB	<ul style="list-style-type: none"> Accuracy = 96%, SN = 92% 	<ul style="list-style-type: none"> MPA can efficiently optimize the ML methods and attained performance by the optimization phase was ranked highest one.

Table 2: Feature Extraction and selection methods used by several analyzers.

Ref no.	Dataset	Feature Extraction/Selection	Classifier
[9]	Heart Disease	<ul style="list-style-type: none"> Fast Correlation-based Feature Selection (FCBF) Particle swarm Optimizer (PSO) Ant colony optimization (ACO) 	<ul style="list-style-type: none"> KNN = 99.6% SVM = 83.5% RF = 99.6% NB = 86.15% MLP = 91.65%
[10]	Heart Disease	<ul style="list-style-type: none"> Predefined feature extraction process 	<ul style="list-style-type: none"> LR = 82.50%, SVM = 78.87%, KNN = 83.84%
[11]	Heart Disease	<ul style="list-style-type: none"> ML feature extraction 	<ul style="list-style-type: none"> NB = 77% SVM = 87% DT = 90%

[18]	Heart Disease	<ul style="list-style-type: none"> EO-SSA (Elite Opposition-based squirrel search algorithm) feature selection 	<ul style="list-style-type: none"> Hybrid DL-based heart disease detection and classification (FSHDL-HDDC) = 97%
[19]	Heart Disease	<ul style="list-style-type: none"> Linear Quadratic Discriminant Analysis based feature extraction 	<ul style="list-style-type: none"> Kmean _LQDA_DG_ConvNet = 96%
[20]	Lung Disease	<ul style="list-style-type: none"> PSO 	<ul style="list-style-type: none"> SVM=81%
[21]	Lung Disease	<ul style="list-style-type: none"> Feature ranking methods 	<ul style="list-style-type: none"> SVM = 95.4% KNN=95.2% ANN = 94.6% NB=95% DT = 93.7%
[22]	Lung Disease	<ul style="list-style-type: none"> Weighted Optimized Neural Network MLMR attributes 	<ul style="list-style-type: none"> Ensemble = 91%

3. Research Methodology

This section offers a scheme that can help doctors and patients in making fast medical results relating to disease classification and health treatments. It consists of different phases: (i) a heart and lung disease classification using the DNN model and (ii) the IFODNN recommender model. The introduced model is separated into several sections: (i) data pre-processing, (ii) feature extraction and selection, (iii) diagnosis of diseases, and (iv) Risk recommendations are provided. Figure 1 illustrates the various modules of the proposed recommendation framework. In the IFODNN recommender model, data preprocessing is applied to HD and LD patient data collected from the Kaggle repository to eliminate missing data and handle outliers. Feature extraction and selection techniques are applied to the database to identify reliable features. The selected feature sets are utilized to classify one of the well-defined categories of collective HDs and LDs, including chest pain, angina, blood pressure, smoking status, AreaQ, Alhhol, etc.

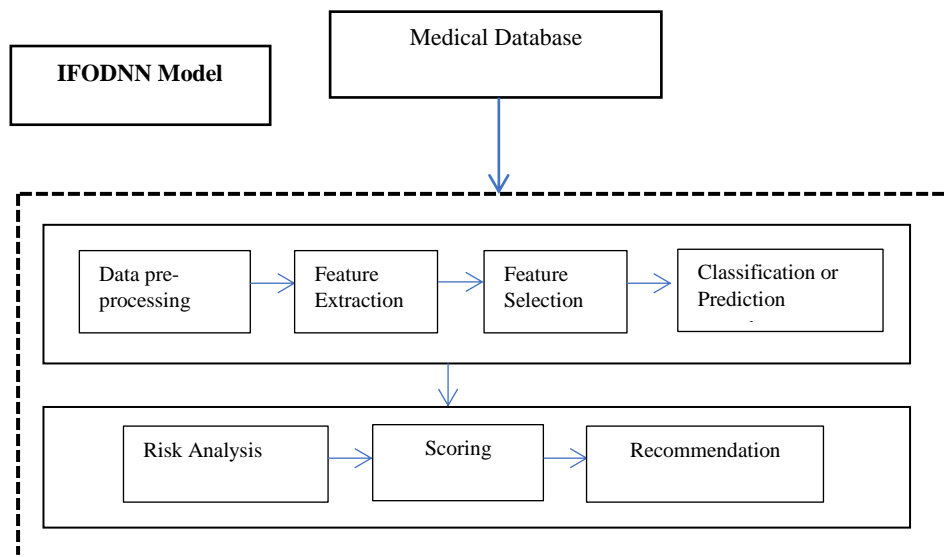


Figure 1. Overview of research methodology flow chart for medical healthcare recommendations

The hybrid recommendation model, using the patient's medical record, is employed by the IFODNN model to provide recommendations for patients with specific illnesses. These recommendations are based on features like categories of predicted disease, risk aspects, probability of disease occurrence, and a detailed analysis of critical parameters derived from the patient's data. Each phase is briefly explained in the following sections.

As shown in Figure 2, we first gathered a database from an online repository called Kaggle [15, 16]. The data pre-processed and selected features influence the target, remove outliers, and split the database into two parts. The training set is utilized for model training, and the testing set is utilized to calculate the method performance.

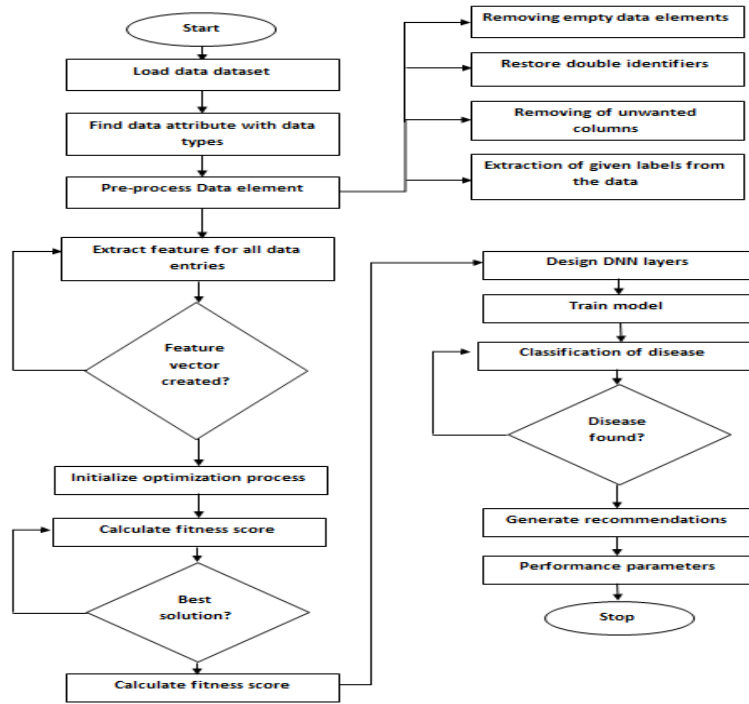


Figure 2. The Block Diagram of Multiple Diseases (Heart and Lung diseases) Classification and Recommendation Model

A. Data Gathering

A brief step utilized in the multiple diseases classification and recommendation method is defined in Figure 2. We have collected a medical healthcare database of numerous disease patients from the Kaggle repository site. The database gathered from [15, 16] was composed of 1,032 and comprised of different databases. It considered 47 characteristics, involving the classification characteristic, but all available investigations mention using a sub-section of 14 of them. The ‘target’ domain denotes the occurrence of HD in the patient. It is numerically valued 0 = no disease present, and 1 = diseased patient data shown in Table 3.

Table 3: Heart Disease Databases [15]

Age	Sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
52	1	0	125	212	0	1	168	0	1	2	2	3	0
53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
61	1	0	148	203	0	1	161	0	0	2	1	3	0
62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
58	0	0	100	248	0	0	122	0	1	1	0	2	1
34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
51	0	2	140	308	0	0	142	0	1.5	2	1	2	1
54	1	0	124	266	0	0	109	1	2.2	1	1	3	0
50	0	1	120	244	0	1	162	0	1.1	2	0	2	1

The efficient LD prediction system helps individuals assess their risk with minimal cost and make appropriate decisions based on their disease risk rank. Table 4 presents the complete data on lung diseases. Algorithm 1 provides a brief explanation of the classification model.

Algorithm1. Multi-disease classification model using DNN model

Input: = Patient's medical healthcare database DB.

Output: = Classification of heart and lung disease $C = 1, 2$.

Suppose DB is separated into train DB and test DB.

For each patient's records from DB_train do

If DB_train has absent values then,

Relate data pre-processing phases.

End

Feature Extraction:

1. Data Preparing

N (input features), and C (output class), compose a novel $(N+1)$ -d input DB.

Normalize each feature (mean, SD).

2. Performing ICA

3. Shrinking small_wts.

4. Extracting feature candidates

5. Deleting un-appropriate feature sets.

Feature Selection:

Set the max. no. of iterations I

Start population $x_j(j=1, 2, \dots, N)$

Start a , aa , and c //

Evaluate the fitness of wolves

xa = best extract_feature

xb = twice best extract_feature

xc = 3rd best extract_feature

while ($t < I$) do

update the position of the recent extract_feature

end for

update a , aa , c

evaluate the fitness of all extract_feature

update xa , xb , and xc

$t = t + 1$

end while

return xc .

Classification:

$F.S'$ defines chosen feature set from DB_train

$F.S' = (f_1, f_2, \dots, f_n)$.

End

Train_classifier $TC = (F.S', \text{data_label})$

Disease_classification $C = (TC, \text{DB_test})$.

Return disease_classification C .

Table 4: Lung Disease Databases [16]

Name	Surname	Age	Smokes	AreaQ	Alkhol	Result
John	Wick	35	3	5	4	1
John	Constantine	27	20	2	5	1
Camela	Anderson	30	0	5	2	0
Alex	Telles	28	0	8	1	0
Diego	Maradona	68	4	5	6	1
Cristiano	Ronaldo	34	0	10	0	0
Mihail	Tal	58	15	10	0	0
Kathy	Bates	22	12	5	2	0
Nicole	Kidman	45	2	6	0	0

B. Data Preprocessing

It is an important section in applying DL methods to present fullness so that relevant analysis is achieved on the record and optimized outcomes are attained. Through data cleaning, miscalculations and improbable data are uninvolved. The misplaced data are conserved similarly. Initially, the misplaced values in the data are noticeable when utilizing an integral filter. The distortion-free data is utilized for the feature extraction and selection procedure, and unreliable feature sets are unconcerned from the database, improving the training method's performance.

C. Feature Extraction Method

It is a computational process used to divide multivariate signals into independent components. This process is typically employed for feature extraction to explore the underlying factors that explain the information. In disease analysis, feature extraction focuses on selecting a subset of the most independent features. The procedure aims to reduce dimensionality, filter pre-processed data, and extract independent features relevant to distinct categories.

The database contains 32 features, but only a few are critical for decision-making in disease diagnosis. Therefore, database values are used to extract relevant features and reduce them to a meaningful sample size. This process enhances disease diagnosis and improves performance metrics.

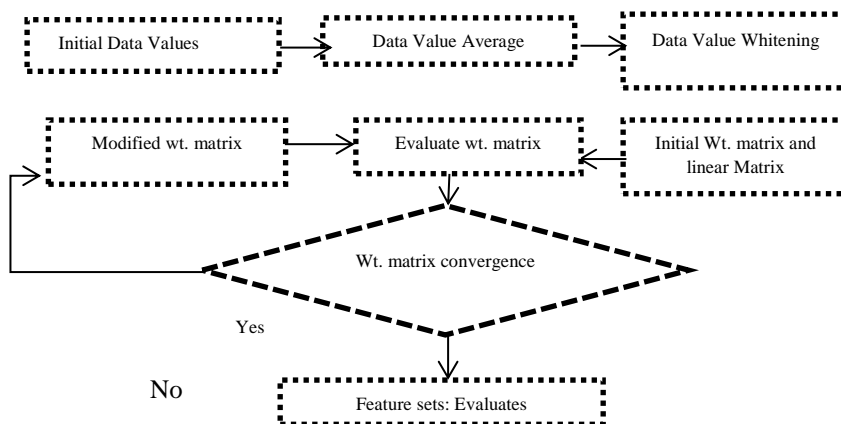


Figure 3. Extracted Features Steps

Figure 3 defines the steps of the feature extraction approach. In the research method, an inherent texture-related feature extraction approach is utilized [17]. In this approach, texture-based features, i.e., the unique characteristics of each feature for the defined class, are evaluated.

D. Feature Selection

For diagnosis of the diseases, feature selection proposes to choose a sub-category of extremely reliable feature sets. The feature selection procedure motivates to choose those reliable features that relate to distinct available categories, such as heart, and lung. The database includes 24 components, but only a subset of them importantly contributes to the decision-making procedure for multi-disease classification. To identify this, feature selection has been applied to decrease the independent feature vectors to a more reliable and valuable size. The feature selection procedure is separated into different portions: (i) An attribute evaluator method is used to assess the reliability of the database features concerning the target output class, and (ii) the Explore approach is employed to search several combinations of feature sets and verify the optimized set for the classification task. In the research model, a reliable feature selection approach is used [17]. In this approach using the grey wolf method to select a feature i.e., the uniqueness of each extracted feature for the defined class is evaluated.

These selected features are determined by a **fitness function**, which evaluates the quality of the selected subset of features. There are steps defined as below:

1. Feature defines

Assume there is an “n” feature in the extracted sets from the database. Individual feature set, which is defined as a binary value: X_i , where:

$X_i = 1$ defines the selected features, and $X_i = 0$ defines the not-selected features. The solution is defined as binary vector $X = [X_1, X_2 \dots \dots X_n]$.

2. Fitness function (FFn)

The main motive of this approach is to explore the sub-set of the features, which increases classification accuracy and reduces error rate. The fitness function calculates how good a specific feature subset is for the classification task. A standard fitness function (FFn) is the weighted sum of the classification error and the number of selected feature sets.

$$f(X) = \alpha \cdot e(X) + \beta \cdot \frac{\text{no.of selected_features}}{n} \quad (1)$$

Here, Equation (1),

- $f(X)$ = fitness value for the solution of X .
- $e(X)$ = classification error utilizing selected features X .
- α and β are used for weighted metrics, which manage the trade-off between reducing the classification error, and minimizing the selected features.
- n represents the no. of extracted features.

3. Optimized procedure

This method reduces this FFn by managing the binary values X utilizing the search engine. The key features (α, β, δ) help to explore the optimized feature sets by handling their locations.

The grey wolf attributes calculation is a method for choosing features depending on the grey wolf of the feature sets. It is developed to calculate feature sets in the framework of defined class labels surveyed by an FFn step to give a binary list of components. The FFn is utilized in combination with the feature calculation step and binary features by their distinct calculations. FFn to choose and remove the features, i.e., 0.1, to eliminate all features having minimum values than the cut-off value and recollect the entire maximum selected features. A total of 24 feature sets are designated and fed into the classification method for effective decision-making.

E. Prediction

The designated feature sets are applied to the training method that uses these features to predict and classify diseases. The online databases are available with labels. The classification tasks are organized as a multiclass problem, where the medical data is divided into different classes representing specific categories of lung and heart diseases. This method aids in determining the probability of a patient taking a disease depending on the chosen main features. There are dissimilar classification methods that have been used to train the method in this analysis specifically, SVM, deep neural network (DNN), decision tree (DT), and CNN model.

1. SVM

This method mechanism on the rule, which by utilizing a precise function, dual categories of data can be simplified, and divided with hyperplane [17]. This method highest the neighbor distance with positive/negative results. The diseases are then divided based on the cross of the hyperplane they are laying. The general SVM model is a twofold classification method, which takes only dual categories as input data. The real-world issues are not dual, thus SVM is reduced to work for multiple classes. The main motive has different versions of SVM utilized like one vs. one, one vs. all, and so on.

2. CNN

The diagnosis of disease feature sets is applied to the training method, which manages the selected features for disease diagnosis. This is the main procedure where class labels are assigned, and the category of the patient's disease is predicted. A CNN model is employed to identify the disease category for each patient. The CNN architecture [23] is developed using a specific stack of layers from several specialized layers. The initial phase of configuring a CNN focuses on defining the system architecture.

3. DNN

The research work has developed a DNN model based on Deeper Multi-Layer Perceptron (DMLP) architecture with regularization and dropout techniques using deep learning (DL). The implemented method involves a classification model for diagnosing new patient cases using a dataset from the UCI repository, containing records of patients diagnosed with multiple diseases. This model is employed for the classification and diagnosis of multiple diseases. The learning process reflects the self-learning abilities of the machine. Interneuron association strengths, also known as weights (wts.), are used to store information within the network structure. During training, a learning algorithm adjusts the weights of linear and non-linear transfer functions across all neuron units in the network to achieve the design objective—typically minimizing the error rate in an optimized way.

This model manages all weights based on the input-output relationships of the target variable to achieve optimal performance. The training process ensures continuous improvement during each epoch to enhance the model's accuracy.

4. DT

It is an ML method utilized as a classification method. It is a tree-based construction classifier. As represented in Figure 4, it built a tree of decision nodes (DNs) [17] where the inner nodes define the qualities of information edges represent the result principles, and all leaf node (LN) represents the class_label, i.e., target_class label if which route is selected. In a DT, the class forecast process operates as follows:

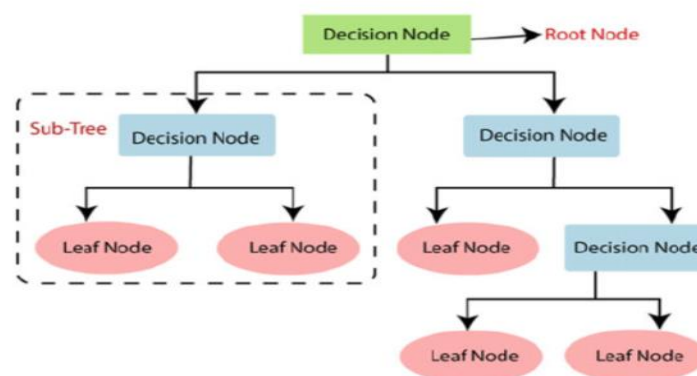


Figure 4. Classification Model: decision tree [17]

The method initializes with the root node (RN) of the DT, following the decision rules to traverse through subsequent nodes until it arrives at an LN. The class_label assigned to the LN represents the predicted class_label for the instance under consideration.

F. Recommendation

After classifying the diseases, then introduced the recommendation model, which is the hybrid DNN model known as the independent feature-based optimized DNN recommender model. The main motive is to give precise and rapid recommendations conferring to the severity of diseases and risks to human life. Different recommendation methods are used for providing advice in this analysis, including IFODNN and Deep Neuro-Fuzzy.

1. IFODNN

The main feature of the researched hybridization is the recommendation model. The main motive is to precise recommendations allowing for the severity of disease and level of risk to human life. An intellectual recommender method surveys the forecast to examine the disease patient's record for retrieving risk, severity, and probability of the incidence of disease that in the shot is utilized to give precise healthcare approvals for further events. The phases shadowed in the researched method are defined in Algorithm 2.

Algorithm 2. Multi-disease recommender model (IFODNN)

Input: Disease classification C . dataset d

Output: Recommendation $R = (1,2,3,4)$.

(1: No recommendation, 2: Need Exercise, 3: Need to visit doctor, and 4: Treatment required).

1. Suppose FV feature set, $FV = (FV1, FV2, \dots)$
 2. Suppose V presents a severity range of FV, V
 3. Suppose X defines weights of FV.
 4. Domain_knowledge (DK) = FV, X, V. // X (weight), and V (severity range).
 5. for each disease C and data from DK,
 6. Evaluate prob. $\Pr(C)$ and $\Pr(C')$.
 7. $\Pr(C)$ - disease present and $\Pr(C')$ – disease not present
 8. Risk calculate $R = \Pr(C)/\Pr(C')$
 9. If $R > 0.5$ and for $C < 2$:
 10. Identify severity score SS and evaluate final_sc;
 11. Final_sc = sim_heart (m,i=1) (SSi(Xi))
 12. else
 13. Identify severity score SS and evaluate final_sc;
 14. Final_sc = sim_lungs (m,i=1) (SSi(Xi))
 15. Endif
 16. Final_sc = (-1 to 2);
 17. $-1 < \text{Final_sc} < 0.5$; $R = 1$. (No recommendation)
 18. $0.5 < \text{Final_sc} < 1$; $R = 2$: Need Exercise
 19. $1 < \text{Final_sc} = 1.5$: $R = 3$: Need to visit doctor
 20. $1.5 < \text{Final_sc} < 2$: $R = 4$ Treatment
 21. EndFor
 22. return R
-

Recommendations are given utilizing an input dataset i.e., collected with the assistance of heart and lung diseases. It supports addressing the significance and importance of health metrics that play a main part in classifying precise HD and LD. The IFODNN recommender model is effective in optimizing time, cost, and workload for both patients and clinical experts. This would importantly optimize the commercial load on patients relating to isolated physical positions. The indication of the IFODNN model based on the heart and lung diseases RS is defined in Figure 5. In addition, described in detail in the following sub-module.

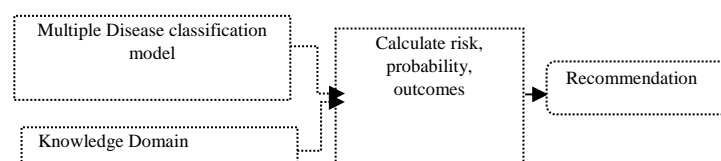


Figure 5. IFODNN on multiple disease recommender system

The initial phase of IFODNN is to generate a database from patient medical records. This database contains medical test data components or features and their relevance in predicting the category of specific heart diseases (HD) and lung diseases (LD). The database is constructed through consultations with clinical doctors and data collected from theoretical analysis. The modules involved in generating the database are illustrated in Figure 5. Key components that contribute to the early diagnosis of multiple diseases are selected. These components include cholesterol (chol), blood pressure (BP), oxygen levels, blood sugar, alcohol consumption, etc. Each selected clinical metric is assigned a significance weight based on its importance in detecting multiple diseases. This significant weight, in turn, determines the patient's criticality. For example, a patient requiring oxygen support is prioritized for healthcare and given greater importance over others.

3. Experimental Result Analysis

This section represents the simulation analysis for data classification and makes recommendations to the patients. The research model is designed using MATLAB 2021a executing a mechanism with an INTEL i7 CPU and 8GB RAM capacity, and the operating system (OS) is Windows 10 with 64 bits. Multi-disease databases are publicly available and can be downloaded from the Kaggle source [15,16]. The database attributes described in the HD and LD database are given in Tables 3 and 4. The research work to detect disease severity in different phases included classification and recommendation. Several parameters are used in this proposed work, like accuracy, sensitivity (SN), specificity (SP), time, F1-score, RMSE, and train accuracy. There are a few terms utilized for these parameters: TP (true positive), TN (true negative), FP (false positive), and FN (false negative). When it is shown that a disease is defined in a patient and the diagnostic analysis has also defined the presence of that disease, in such a scenario, the outcome of the diagnostic investigation is measured as TP. When the disease is absent in a patient, and the diagnostic analysis outcomes also advise that the condition is missing, the test outcomes are TN. As defined, mathematical accuracy, time, RMSE, SP, SN, etc., consider examining the direct comparison of patient classifications with their disease category. The RMSE rate is normally utilized to assess mathematical accuracy. These terms are intended by using the succeeding mathematical prescriptions;

- **Accuracy:** It is a vital degree used to estimate the performance of the classification method. It is the ratio of the sum of total correct TP and TN and the full sum prediction of TP, TN, FP, and FN by the classifier to the whole data points. It represents as (2).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (2)$$

- **Specificity:** It is the ability to regulate healthy cases reliably and evaluate the quality or ratio of TN and FP. It represents (3).

$$Specificity = \frac{TN}{TN+FP} \dots\dots\dots (3)$$

- **Sensitivity:** It is defined as the ability to fix patient cases appropriately and evaluate the fraction or ratio of TP and the amount of FN and TP. It represents as (4).

$$Sensitivity = \frac{TP}{TP+FN} \dots\dots\dots (4)$$

- **F1-Score:** It is equivalent to the harmonic mean of Precision and Recall value. It fetches a suitable balance between these. It correctly estimates the model's performance in classifying specific values. It is the most significant computation to evaluate the model. It is also defined as the multiplication of two with the product of precision (pre) and recall (rec) values separated by the sum of pre and rec values. It represents as (5).

$$F1 - score = \frac{2(precision*recall)}{Precision+recall} \dots\dots (5)$$

- **RMSE:** It calculates the difference between forecast and real values for each data point, computes the residual norm for every data point, computes the mean of residuals, and precedes the square root of that mean. It represents as (6).

$$RMSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{N} \dots\dots (6)$$

Table 5: Proposed IFODNN model Performance metrics

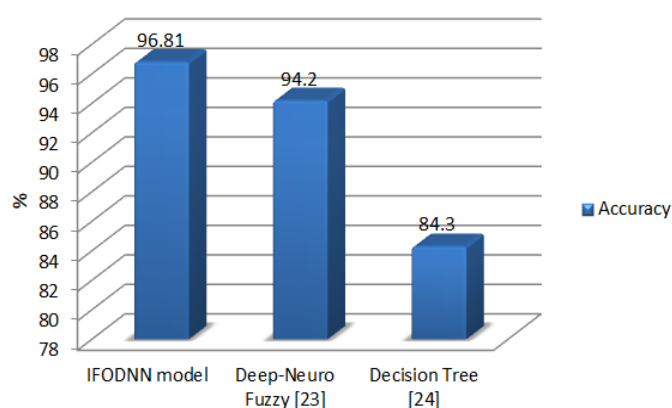
Parameters	Values
Accuracy (%)	96.81
RMSE	0.102
SN(%)	98.65
SP(%)	99.16
F1-score (%)	93.0
Time(sec)	2.44
Train accuracy (%)	96.70

4. Discussions

Table 6 compares proposed and existing models with dissimilar performance parameters, like SN, SP, time, accuracy, F1-score, etc. The proposed model has achieved maximum accuracy in SP, SN, F1-score, train accuracy, and mitigation of the total sum of error and time consumption compared to the existing models, such as CNN with fuzzy and Decision Tree. As defined in Table 5, the implemented IFODNN model is a hybrid of the GWO and CNN methods to classify multiple diseases, such as lung and heart diseases. It also incorporates feature selection to provide necessary recommendations based on the condition. This approach ensures better performance and efficiency in the calculation scenario related to the disease category; it has deep neurofuzzy [23], and decision tree (DT) [24] prediction and recommendation models.

Table 6: Comparison Analysis between proposed and existing Methods

Models	Accuracy	RMSE	SN	SP	F1-score	Time	Train Accuracy
IFODNN model	96.81	0.102	98.65	99.16	93.0	2.44	96.70
Deep-Neuro Fuzzy [23]	94.2	0.20	96.8	97.4	89.0	4.75	92.28
Decision Tree [24]	84.30	0.314	90.9	93.3	84.0	7.35	87.84

**Figure 7.** Accuracy of the IFODNN model for test samples

The maximum accuracy rate of the implemented IFODNN model defines the full ability of this approach, depending on the DNN method in the implemented technique. Accordingly, the accuracy rate diagram of the implemented IFODNN model for test samples using classification and recommendation techniques is defined in Figure 7. As defined in Figure 7, the implemented IFODNN model, using hybridization of the DNN and optimum feature selection, has performed well in terms of accuracy rate in multiple diseases (heart and lung). Another case calculated in the implemented model is the standard of SN, which is helpful as a percent of the accurate recommendations given to the complete recommendations defined in the implemented approach.

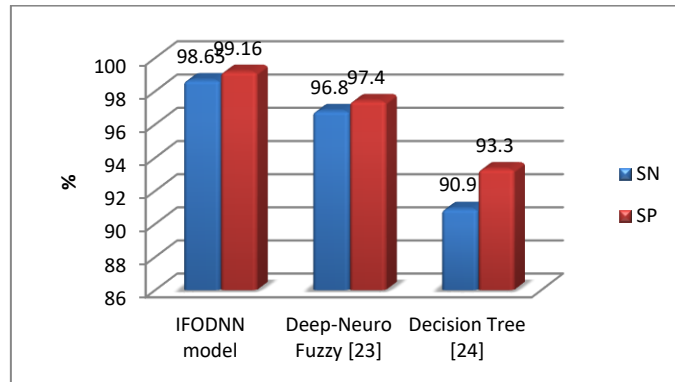


Figure 8. Comparison analysis between IFODNN and existing methods with SN and SP rates

Figure 8 represents the hybridization of the KNN method and GWO feature selection; more than 95% of the recommendations defined to patients are accurate, which explains the maximum ability of the implemented statistical analysis of the heart and lung disease recommender system. Another case used in the implemented IFODNN model is accuracy, the percent of accurate recommendations defined to the complete recommendations.

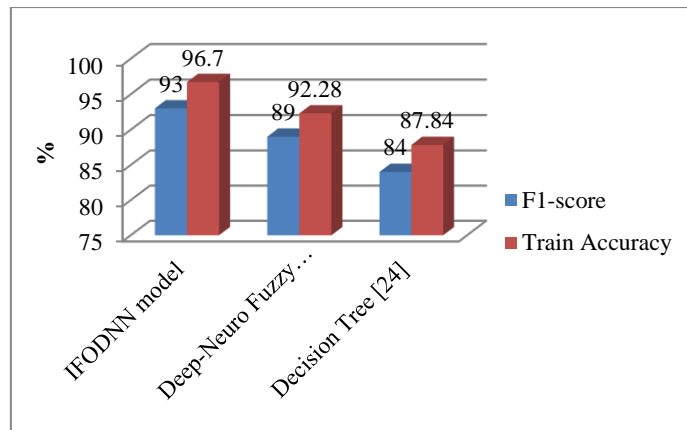


Figure 9. Comparison analysis between IFODNN and existing methods with F1-score and train accuracy

Figure 9, in the hybridization of the DNN method and the GWO feature selection, about 90% of the recommendations given are correct. Another case utilized in the implemented IFODNN model is the f1-score, i.e., the hybridization of two cases of accuracy and SN. Figure 8 represents the accurateness diagram of the implemented model.

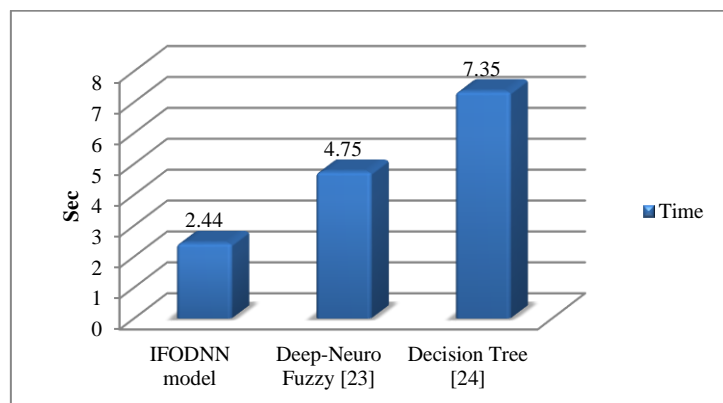


Figure 10. Comparison analyses between proposed and existing methods with time (Second)

Figure 10 represents the time execution metric in the hybridization of the DNN model with the GWO method as the minimum execution time as compared with the previous methods.

6. Conclusion

In this research work, we have implemented a multiple-level-based architecture for the healthcare RSs. This implemented model addresses the distinct features of the patient record to classify them into multi-disease classes using DNN. Afterward, the IFODNN model is utilized to risk of disease, approximate the probability of disease incidence, and lastly, a recommendation method has developed recommendations to the patients based on the severity staging of the diseases and medical data by aiding doctors in making rapid decisions by designing a crossbreed classification and recommendation model. Generally, the implemented method mainly focused on predicting heart, and lung disease, while the implemented method depends on the multiple-label classification of different heart and lung disease metrics. These metrics are discussed in the above section 3.1. The proposed system chooses reliable independent features and classifies conditions using significant filtered feature sets. Between the classification methods, DNN achieves the maximum accuracy rate, and based on classified diseases and information collected from doctors, a statistical analysis module of IFODNN is performed for authenticating and giving accurate recommendations to the patients. Different methods might be applied with the IFODNN model to improve the outcomes. The introduced method will improve more with the help of a hybrid feature selection approach with the novel optimization method. This proposed model can mainly use feature selection for incomplete and missing data. This novel method will assist in improving performance metrics, like accuracy, precision, etc.

Conflicts of Interest: The Authors have no conflicts of interest to declare.

References

- [1] M. R. Cowie, J. Bax, N. Bruining, J. G. Cleland, F. Koehler, M. Malik, F. Pinto, and P. Vardas, "e-Health: A position statement of the European Society of Cardiology," *Eur. Heart J.*, vol. 37, no. 1, pp. 63–66, 2016.
- [2] D. Lloyd-Jones, R. J. Adams, T. M. Brown, M. Carnethon, S. Dai, and J. Wylie-Rosett, "Heart disease and stroke statistic: A report from the American Heart Association," *Circulation*, vol. 121, no. 7, pp. e46–e215, 2010.
- [3] L. He, J. Wang, L. Zhang, X. Zhang, W. Dong, and H. Yang, "Decreased fractal dimension of heart rate variability is associated with early neurological deterioration and recurrent ischemic stroke after acute ischemic stroke," *J. Neurol. Sci.*, vol. 396, pp. 42–47, 2019.
- [4] S. Chen, X. Zhu, S. Zhang, and J. A. Wang, "A framework for massive data transmission in a remote real-time health monitoring system," in *Proc. 18th Int. Conf. Autom. Comput. (ICAC)*, 2012, pp. 1–5.
- [5] T. Huang, L. Lan, X. Fang, P. An, J. Min, and F. Wang, "Promises and challenges of big data computing in health sciences," *Big Data Res.*, vol. 2, no. 1, pp. 2–11, 2015.
- [6] M. M. Baig and H. Gholamhosseini, "Smart health monitoring systems: An overview of design and modeling," *J. Med. Syst.*, vol. 37, pp. 1–14, 2013.
- [7] J. D. Ochoa, O. Csiszár, and T. Schimper, "Medical recommender systems based on continuous-valued logic and multi-criteria decision operators, using interpretable neural networks," *BMC Med. Inform. Decis. Mak.*, vol. 21, pp. 1–15, 2021.
- [8] A. Mustaqeem, S. M. Anwar, A. R. Khan, and M. A. Majid, "Statistical analysis-based recommender model for heart disease patients," *Int. J. Med. Inform.*, vol. 108, pp. 134–145, 2017.
- [9] Y. Khourdifi and M. Baha, "Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 1, 2019.
- [10] M. Indukuri, K. S. Kumar, and B. S. Reddy, "An approach to detect multiple diseases using machine learning algorithm," in *J. Phys.: Conf. Ser.*, vol. 2089, no. 1, p. 01:2009, 2021.

- [11] K. Arumugam, M. Naved, P. P. Shinde, O. Leiva-Chauca, A. Huaman-Osorio, and T. Gonzales-Yanac, "Multiple disease prediction using machine learning algorithms," *Mater. Today: Proc.*, vol. 80, pp. 3682–3685, 2023.
- [12] B. Ihnaini, M. A. Khan, T. A. Khan, S. Abbas, M. S. Daoud, M. Ahmad, and M. A. Khan, "A smart healthcare recommendation system for multidisciplinary diabetes patients with data fusion based on deep ensemble learning," *Comput. Intell. Neurosci.*, 2021.
- [13] F. H. Awad, M. M. Hamad, and L. Alzubaidi, "Robust classification and detection of big medical data using advanced parallel k-means clustering, YOLOv4, and logistic regression," *Life*, vol. 13, no. 3, p. 691, 2019.
- [14] M. Hassaballah, Y. M. Wazery, I. E. Ibrahim, and A. Farag, "ECG heartbeat classification using machine learning and metaheuristic optimization for smart healthcare systems," *Bioengineering*, vol. 10, no. 4, p. 429, 2023.
- [15] D. Lapp, "Heart disease dataset," *Public health dataset on Kaggle*, 2019. [Online]. Available: <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>. [Accessed: 19-Jul-2023].
- [16] Y. Dede, "Lung cancer dataset," *Kaggle*, 2018. [Online]. Available: <https://www.kaggle.com/datasets/yusufdede/lung-cancer-dataset>. [Accessed: 19-Jul-2023].
- [17] S. R. Vijaya, K. Gajavelly, N. A. Kousar, R. Vasavi, and A. R. Reddy, "Heart disease prediction using decision tree and SVM," in *Proc. 2nd Int. Conf. Adv. Comput. Eng. Commun. Syst. (ICACECS)*, Singapore: Springer Nature, 2021, pp. 69–78.
- [18] S. Kallimani, W. Jagadish, and B. Ranjan, "A novel feature selection with hybrid deep learning-based heart disease detection and classification in the e-healthcare environment," *Comput. Intell. Neurosci.*, vol. 105, p. 748–759, 2022.
- [19] K. V. Saikumar, G. S. Rajesh, J. Chun, and L. Wei, "Heart disease detection based on Internet of Things data using linear quadratic discriminant analysis and a deep graph convolutional neural network," *Front. Comput. Neurosci.*, vol. 1, no. 6, p. 964686, 2022.
- [20] BioMed Research International, "Retracted: Lung cancer prediction from text datasets using machine learning," *BioMed Res. Int.*, 2023.
- [21] E. Dritsas and M. Tzani, "Lung cancer risk prediction with machine learning models," *Big Data Cogn. Comput.*, vol. 6, no. 4, p. 139, 2022.
- [22] J. A. ALzubi, B. S. Bharathikannan, R. M. Tanwar, A. Khanna, and C. Thaventhiran, "Boosted neural network ensemble classification for lung cancer disease diagnosis," *Appl. Soft Comput.*, vol. 80, pp. 579–591, 2019.
- [23] D. Sharma, A. G. Singh, and R. Bajaj, "Deep neuro-fuzzy approach for risk and severity prediction using recommendation systems in connected healthcare," *Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 7, pp. 41–59, 2021.
- [24] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in e-healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, 2020.

Appendix I

SRNO.	ABBREVIATIONS	DESCRIPTION
	Acc.	Accuracy
2.	ANN	Artificial Neural Network
3.	AF	Activation Function
4.	ACO	Ant Colony Optimization
5.	BP	blood pressure
6.	BC	Breast Cancer
7.	CDs	Chronic Diseases
8.	CNN	Convolutional Neural Network
9.	DD	Diabetic Disease
10.	DL	Deep Learning
11.	DF	Data Fusion
12.	DT	Decision Tree
13.	DB	Database
14.	FCBF	Fast Correlation-based Feature Selection
15.	FBS	fasting blood sugar
16.	FV	Feature Vector
17.	FS	feature selection
18.	FIS	Fuzzy Interference System
19.	FL	Fuzzy Logic
20.	FP	False Positive
21.	FN	False Negative
22.	GWs	Grey Wolves
23.	GWO	Grey Wolf Optimizer
24.	HDC	Heart Disease Classification
25.	HCRS	Healthcare Recommendation System
26.	HDs	Heart Diseases
27.	HM	Health Monitoring

28.	IFODNN	Independent Feature-based Optimized DNN
29.	ICA	Independent Component Analysis
30.	KNN	K-nearest Neighbor
31.	LN	Leaf Node
32.	LD	Lung diseases
33.	MHO	Meta-Heuristic Optimization
34.	MS	Monitoring Systems
35.	MLP	Multilayer Perceptron
36.	ML	Machine Learning
37.	NB	Naïve Bayes
38.	OS	Operating System
39.	PC	pus cell
40.	PVC	Packed Cell Volume
41.	PSO	particle swarm optimization
42.	PD	Pulmonary diseases
43.	RA	Risk Analysis
44.	RF	Random Forest
45.	RN	Root Node
46.	SA	Statistical Analysis
47.	SF	Sigmoid Function
48.	SVM	Support Vector Machine
49.	SS	Severity Scores
50.	SP	Specificity
51.	SN	Sensitivity
52.	TN	True Negative
53.	TP	True Positive
54.	WA	Web Application