



## **BFCNN: Multi-Disease Classification Bee Inspired CNN Model**

**Lakhvinder Singh<sup>1,2\*</sup> Dalip<sup>3</sup>**

<sup>1</sup>*M.M. Institute of Computer Technology & Business Management, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, 133207, India*

<sup>2</sup>*Department of Computer Science, Government College, Chhachhrauli (Yamuna Nagar), Haryana, India*

<sup>3</sup>*M.M. Institute of Computer Technology & Business Management, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, 133207, India*

\* Corresponding author's Email: lakhvindersinghahuja@gmail.com

---

**Abstract:** An intelligent information technology-based system offering practical health advice can improve the superiority of lifespan for individuals with chronic illnesses. This study develops a BFCNN (Bees featured CNN) model for the classification or prediction of heart, liver, and kidney diseases. The main objective of the proposed classification model is to accurately diagnose a patient's disease. The research model aims to construct a highly efficient deep-learning system for patients with heart, liver, and kidney diseases. A novel approach for detecting multiple diseases in medical datasets using a BEE-featured CNN model. The BEE-featured CNN model is a modified version of the traditional CNN model that incorporates the behavior of bee swarms to enhance the accuracy of the model. The bee-inspired algorithm used in the model involves the use of multiple artificial bees that simulate the behavior of real bees to search for optimal solutions. The proposed model was trained and tested using a dataset consisting of medical datasets of patients with different diseases. The images were collected from various sources and labeled with the corresponding disease type. The dataset was split into training and validation sets, with 80% of the datasets used for training and 20% for validation. The results of the experiments showed that the proposed BEE-featured CNN model outperformed the traditional CNN model in detecting multiple diseases in medical datasets. The results of the prediction model demonstrate outstanding performance, with an accuracy of 96.7%, specificity of 99%, and Sensitivity of 97%. The research system demonstrates promising results in classification accuracy and holds potential for usage in medical and e-health applications.

**Keywords:** Healthcare system, BFCNN classification model, Heart, Liver, Kidney diseases.

---

### **1. Introduction**

In the present day, the majority of our behaviors are captured as data and can be classified by extracting relevant information from our background [1]. Medical professionals also use this background data to provide patient care, considering various clinical aspects. Healthcare systems use this data to maintain disease records and take necessary activities. The presence or absence of a disease can be classified by observing specific medical metrics. The test lies in determining the risk or severity level and making appropriate approvals, both for curing the disease or reducing the risk factors associated with it.

Chronic diseases (CDs) are impacted by various factors in daily life, such as diet and family history. E-Health is a cutting-edge research topic that has made significant progress in providing effective solutions for CDs. E-Health techniques are recycled to treat a general collection of patients, with those in remote locations who are affected by multiple CDs.

Heart disease (HD) is one of the prominent sources of loss globally among CDs. Historically, heart illnesses were more prevalent in elderly adults, but this trend has changed in recent years with heart diseases being increasingly seen in younger individuals [2]. Urban residents have greater access to clinical treatment and can better protect themselves from life-threatening diseases. On the other hand,

people in remote areas often struggle to access medical services and are at a greater risk. E-Health is seen as a promising approach to reducing illness, mortality, and costs associated with heart disorders [3].

Health monitoring (HM) systems have advanced in recent years and are now used in more complex ways, beyond just measuring sleep hours [4]. Methods [5] for HM process data in a way that is more beneficial for the manipulator. Modern training has utilized progressive machine learning (ML) techniques to handle complex data in three different tasks, including prediction, anomaly detection (AD), and decision-making procedures. These techniques are used to predict data in multiple groups already available for a specific disease [6] and [7]. Despite these advancements, the challenge remains in acquiring data through effective and highly precise analytical methods that support personalized decisions for disease analysis. The primary objective of e-health structures is to reduce the human danger by perceiving illnesses early and providing trustworthy recommendations. There has been a significant effort to identify a specific sickness at various clinical phases [8]. However, it is still essential to offer patients customized healthcare recommendations based on their current medical situation.

This paper presents a hybrid estimation framework using the BFCNN model for diagnosing and treating patients with multiple diseases, such as heart, liver, and kidney diseases. The main objective is to propose an improved classification model to classify patients with multiple diseases. The research method offers a BFCNN classification model to predict the patients into one of the three main diseases: liver, heart, or kidney. While much of the previous research in this field has focused on identifying a single disease [9], this approach takes into account the risk analysis of a patient's medical information to provide medical advice for those with multiple diseases. The system's results indicate that it will play a significant role in multiple diseases (MD) prediction, risk assessment (RA), identifying anomalies associated with MD patients, and organizing disease illness. The outcomes of the research method show a strong prediction accuracy rate. The proposed model makes the following contributions:

- (i) A review of recently developed classification and recommendation systems in the health care domain.
- (ii) A feature extraction method using KPCA to extract reliable feature vectors.
- (iii) An implementation of a model to classify different diseases in the health care system using the

BFCNN model to overcome the time consumption problem with classification techniques.

This article is prepared into the subsequent sections: Section 2 offers a review of the literature on the classification and feature extraction methods in the domain. Section 3 defines the research problems and, in Section 4, defines the research methodology. Section 5 defines and details the dataset and simulation results. Lastly, the conclusion and further work are described in Section 6.

## 2. Literature review

Ramasamy et al. (2021) [10] proposed a classification-based model for detecting multiple diseases. The neural network (NN) was developed as an innovative solution for real-world classification problems. Classification methods were used to enhance the training progression and improve accuracy and error rates. The proposed technique was deployed for multiple diseases and the HOT (half of the threshold) algorithm, integrated with Strassen's matrix multiplication algorithm (SMM), was used for computational efficiency and a hybrid approach to train and validate the NN-based Strassen HOT approach. The experimental results of the proposed Strassen HOT algorithm were highly accurate and achieved a 7% to 54% improvement in the time range and a 3% to 15% increase in efficiency across multiple datasets. Wang et al., (2022) [11] described a classification model for predicting multiple diseases from unbalanced medical data. The goal was to diagnose and treat infectious diseases predictably. The proposed multiple-infectious DDM (disease diagnostic model) was developed for infectious disease classification and required seven different categories of infectious disease data. The deep learning (DL)-based model achieved better performance and had an accuracy of 97.31% for the mouth, 98.86% for the influenza virus, and 98.47% for the foot and hands. The proposed test also had an accuracy of 83.03% for syphilis, 87.30% for diarrhea, and 87.30% for measles. Berlin Richard et al., (2018) [12] proposed a medicine disease system with scalability beyond (SB) and boundary circumstances. The diseases and their phases were abstracted from medicine and were adaptable to modern science predictions, without a fixed hierarchy for the future. The proposal aimed to provide information about diseases such as types, boundaries, etc. and represented the disease explanation as an automatic categorical aspect of decomposition outside the lines of the threshold and boundaries. Ahirro Aditya et al., (2020) [13] discussed the increasing rates of disability and death caused by chronic diseases such

as those of the kidneys, lungs, and heart. These chronic diseases were responsible for the rise in death rates and the difficulty in finding a cure. Aditya et al. introduced a classification system based on machine learning (ML) to predict and detect various chronic diseases, such as lung, tumor, and heart disease, to expand the accuracy of Diagnosis and treatment in the early phases of disease classification through the analysis of medical data in healthcare and patient care services. The accuracy rate of diagnosis and reliable treatment were used as a parameter in the patient disease analysis, as incorrect diagnosis increases the risk of death in chronic diseases. Anil Dubey et al., (2021) [14] described the rapid evolution of biomedical data and the healthcare classification system. Machine learning (ML) techniques were employed in various analyses to classify the risk factors of chronic diseases. This paper presents a novel method intended for calculating multiple chronic infections expending deep learning (DL) methods. The general classification process consists of several phases, including data acquisition, optimum feature selection (FS), statistical feature extraction (FE), and classification. The health data of chronic diseases was obtained from various sources, and OFS was carried out on the existing set of features using hybrid methods such as the lion and butterfly methods. Alarsan et al. (2019) [15] suggested an electrocardiogram (ECG) classification technique using ML. The electrical signals of the heartbeat were measured using an ECG. The proposed method was implemented on the Apache Spark platform, using ML-lib and the Scala software design language. MLlib is an extensible ML library for Apache Spark. The most challenging aspect of ECG signal classification was dealing with anomalies in the waveforms, which were critical for determining the health diagnosis. Mustaqeem et al. (2017) [16] introduced a crossbred system for heart patients that provided sickness assessment and medication advice. The first step was to develop a statistical model that could recognize medical symptoms and categorize them into one of four class labels: silent hypoxia, non-cardiac chest pain, ischemia, and cardiovascular events. The research aimed to create an autonomous and dynamic cardiovascular disease decision support system. The results of the suggestion model were tested using a correlation coefficient, yielding an accuracy of 97.8%. Liu, et al., (2017) [17] discussed a system with two modules: a dimensionality reduction module based on the RFRS and a categorization module based on an ensemble classification. The system has three steps: information initialization, feature extraction with the ReliefF methodology, and dimensionality decrease

with the Rough Set approach. An ensemble classifier using C4.5 was proposed with a jack-knifing cross-validation strategy resulting in a maximum prediction performance of 92.59%. Ellaji et al., (2020) [18] analyzed the use of multiple smart wearable devices to improve medical care. The proposed system was a Narrow Band technology system that linked multiple devices. Sudharson, et al., (2021) [19] proposed a computer-assisted diagnosis (CAD) method for recognizing multi-class kidney disease using ultrasounds. The CAD system used a pre-trained ResNet-101 network and an SVM predictor. The deep RLN improved recognition accuracy compared to conventional techniques. The CAD system was found to be more effective than other classifiers such as K-nearest neighbor, tree, and discriminant. Muhammed et al., (2018) [20] proposed a network QoS utilizing Deep Learning and big data to detect network traffic or issues and to identify malicious traffic. Three datasets were used to calculate the ubeHealth system. Alzub et al., (2021) [21] analyzed that critical patient information was often abused as a result of cybercriminals. The authors proposed a blockchain technology secure data framework for biomedical sensor networks using LamportMerkle digital signatures, which was motivated and guided by these factors. At first, the LMDSG system was authenticated by smart electronics products by establishing leaves inside the tree-based structure to represent the hash algorithm of critical patient medical information. Sharma, et al., (2017) [22] studied the healthcare system and offers a minimal deep feature decision-making provision system for identifying healthcare renal disease using just 2nd order GLCM relevant properties based on the analyzed renal ultrasonography images to extract therapeutic renal disease. Bhaskar, et al., (2019) [23] suggested a new sensing approach. For the identification of sickness, the quantity of urea in the salivary is measured. For the measurement of nitrogen concentrations in the mouth swab, a novel sensing technique is used. Furthermore, we designed a 1D deep CNN technique with an SVM predictor to assess the input data acquired from the sensing. The model's accuracy of the system was improved by using a CNN-SVM interconnected system. The developed scheme correctly identified the specimens with 98.04 per cent accuracy. Pasyar, et al., (2021) [24] used a unique deep predictor built of pre-trained deep CNNs was proposed to categorize liver conditions. Various networks, including ResNeXt, ResNet34, ResNet18, ResNet50, and AlexNet, were used in conjunction with FCNs (fully connected networks). Deep extracted features by transfer learning could provide enough categorization data.

An FCN was used to transform images into various diseases and disorders, such as liver hepatitis, normal liver, and cirrhosis. For liver scans that were categorized into three categories, the simulation results indicated an precision of 86.4 per cent utilizing ResNet50 through a crossbreed classifier. Singh, et al., (2020) [25] used a web development strategy to forecast liver problems and used a categorization and feature engineering methodology. The recommended implementation was based on ILPD (Irvine dataset's Indian liver patient dataset). On the liver patient dataset, several pre-trained models such as KNN, logistic regression (LR), Naive Bayes, SMO, and random forest (RF) algorithm were used to determine accuracy. Sharma, et al., (2018) [26] studied that FLD termed Fatty Liver Disease was among the most serious disorders which should be discovered and treated as soon as possible to reduce death. Ultrasonography scans had been frequently used mostly by psychiatrists to diagnose FLD. Researchers had difficulty recognizing FLD due to the poor quality of ultrasound images. Many scientists had built numerous CAD (Computer-aided diagnosis) techniques for the categorization of adipose and liver cell ultrasonography pictures to overcome the challenges. According to the simulation observations, the suggested CAD system could achieve 95.55 per cent accuracy and 97.77 per cent sensitivity using the 20 key qualities picked using the MI deep learning model. Baek, et al., (2020) [27] proposed H-scan, which used a channel estimation technique to differentiate the dispersion difference equation, as well as the Burr allocation. Quantifies speckled configurations and Burr allocation were two new investigations that were closely related to the thermodynamics of ultrasonic dispersion. Statistical investigations could yield at least 5 different characteristics that were strongly attributable to ultrasonography in cellular arithmetic. The various kinds of hepatic illness could be distinguished with 94.6 per cent accuracy using a multipara metric measuring technique and group identification. Table 1 displays various techniques, simulation tools, comparative methods, and performance metrics used in disease recommender and classification systems. In Table 2, the different existing methods for disease recommender and classification systems are outlined, including MIDDM, Multi-disease prediction, and others. The concluded and proposed methods are compared with different techniques, research datasets, and performance metrics, highlighting existing research problems and gaps, implemented tools, and metrics.

### 3. Problem statement

With the increase in chronic illness cases such as kidney, liver, and heart disease in the world, medical care support systems such as recommendation systems play an important role in disease prevention by offering consistent and precise risk factor treatment forecasts and acceptance of the cardiovascular disease, kidney disease, and liver disease status, allowing healthcare professionals to get a 24/7 mobile health monitoring system and clients to provide a 24/7 remote health monitoring framework[23]. The challenge in health data, which is characterized by unbalanced, large, unclear, and incomplete information, makes it challenging to provide high-accuracy predictions. The aim of the disease recommender system is to predict disease risk with high precision and reliability [25, 26]. The complexity of symptoms associated with various diseases complicates the job of surgeons [21, 27, 28]. Previous studies have attempted to automate the recognition and classification of multiple diseases using various techniques and machine learning algorithms. Despite these efforts, there are still challenges in accurately classifying heart, liver, and kidney diseases. In developing disease classification systems, it is crucial to maintain high accuracy. The main problems[16] identified with current systems include the accuracy of classification techniques, the error rate, and the processing time required for disease classification, among others.

### 4. Proposed methodology

The methodology involves several important steps, which are outlined in Fig. 1. It covers the process from selecting the dataset at the beginning to calculating the presentation parameters at the end of the simulation. Afterward, the research proceeded to implement the proposed work. The researchers then collected a clinical dataset from an online source and designed a graphical user interface. This pre-processing step helps to extract valuable content from the data and remove unwanted data. During the text pre-processing step, data cleaning, repetitive data, spelling exceptions, and improbable information are sensed. The missing information and outliers are preserved in the same way. The missing values in the information have been cleared. After pre-processing, the proposed feature extraction and enhanced classification algorithms extract the valuable feature sets and reduce the feature matrix of the uploaded dataset. The feature-based processing performs well when working with the optimized deep learning-

Table 1. Analysis of the different methods

Author Name	Year	Proposed methods	Comparative Methods	Dataset	Parameter
Ramasamy et al., [10]	2021	SHoT training method	HoT BPN	Heart Liver Breast cancer Drug consumption Thyroid disease	Accuracy Train time
Wang et al., [11]	2022	MIDDM	DT BM LR XGBoostr	Clinical infectious disease	Accuracy
Ahirrao et al., [13]	2020	Multi-disease prediction model	NA	Diabetics	Patient status
Dubey et al., [14]	2021	Hybrid lion-based BOA	NN DBN	Kaggle	Precision F1-score Accuracy SP SN

**Abbreviations:** SHoT (strassen’s half of threshold) train method; HoT (half of the threshold); BPN (back propagation neural network); MIDDM (multiple infectious disease diagnostic model); DT (Decision tree); BM (Bayesian method); LR(logistic regression); BOA (butterfly optimization algorithm); NN (neural network); NA (not applicable); DBN (deep belief network); SP(specificity); SN(Sensitivity).

Table 2. Existing methods of disease recommender and classification system

Author’s Name	Year	Proposed method	Problem/Gaps	Simulation Tool/ Dataset	Performance Metrics
Alarsan, et al., [15]	2019	ML Classification Procedures (Random Forest DT Gradient-Boosted Trees )	Need to increase and enhance the accuracy of the system	MATLAB tool/Spark-Scala tool/ MIT-BIH Arrhythmia Dataset	Time (sec) Accuracy (%) SP (Specificity %) SN (Sensitivity %) CC (Correct Classification %)
Mustaqeem, et al.,[16]	2017	Attribute evaluator method Ranker Search Algorithm SVM MLP RF Statistical analysis based heart disease recommender model	Multi-class classification issues and Computational time is high	P.O.F hospital patients dataset/Decision-Making tools	Accuracy (%) Kappa Statistics RMSE (Root Mean Square Error) rate Entropy (%) Sensitivity (%)
Liu, et al., [17]	2017	C4 .5 classifier RelifF and Rough Set-based hybrid method	The value of the threshold is not stable	MATLAB tool/ UCI and Statlog dataset	Accuracy TPR (True Positive Rate)
Ellaji, C., et al., [18]	2020	Narrow Band IoT within elegant HCS (heath care system)	Difficult to improve and deployment of the system is challenging	NA	NA
Sudharson, et al., [19]	2021	ResNet101 SVM RLN (Residual learning network)	Noisy feature sets are degrading the classification results	Pre-trained Kidney Image Dataset	Accuracy SN (Sensitivity) SE (Selectivity)

Muhammed, A., et al., [20]	2018	MLP(multi-layer perceptron) DNN (deep neural network) RNN(recurrent neural network)	Performance problems like maximum latency and bandwidth limitation. Enhanced the security, privacy, and scalability	ISPDSL-II and Waikato-VIII Dataset/ soft flowed tool	Accuracy (%) Kappa (%)
Alzubi, et al., [21]	2021	LMDS (Lamport Merkle Digital Signatures) based technique	High computational time and cost	Authentication Tool	Authentication accuracy (%) CT (computation Times) CO (computation overhead)
Sharma, et al., [22]	2017	GLCM Efficient decision support system SVM KNN PNN (probabilistic neural network)	Need to enhance ROI extraction method	Siemens ACUSON X300 machine Dataset	Accuracy (%) OCA (overall classification accuracy) ICA (individual classification accuracy)
Bhaskar, et al., [23]	2019	CNN-SVM (Convolution neural network-support vector machine)	Consume more time for computational, error rate, and efficiency	Diagnostic tool/ Random sample collection dataset	Accuracy Sensitivity Specificity F1-score Precision MCC FPR (false positive rate) FNR (false negative rate) CT (Computation Time)
Pasyar, et al., [24]	2021	CNN (convolutional neural network) ResNet 50 FCN (fully connected network) CDNN (Convolution deep neural network)	Limited or Large dataset	Two class datasets of 216 images	Accuracy Specificity Sensitivity

based technique. The intelligent models use the feature extraction process to reduce the error probability.

A brief series of phases used in the heart[28], liver[29], and kidney[30] disease BFCNN classification model is defined in Fig. 2.

The main steps are as follows:

**(i) Upload and pre-process the dataset:** This section enables users to upload datasets for multiple diseases, with labeled formats for training and unlabeled formats for testing. The collected data is pre-processed, and unwanted data entries are selected

and rejected. Once the iterations are complete, the data is passed on to the feature extraction module.

**(ii) Feature extraction using KPCA:** The feature-based processing yields high performance when used with deep learning-based techniques. The intelligent models reduce the error probability by using the feature extraction process. This proposal employs the KPCA algorithm to extract the feature vectors. The KPCA method is a linear particle component analysis that provides linear measurement reduction. KPCA extends PCA into nonlinear and parametric forms,

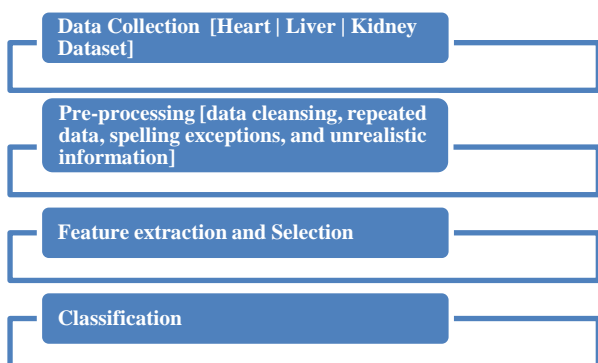


Figure. 1 General flow of the research work

fitting the required dataset. KPCA is a form of PCA that displays nonlinear mappings in high-dimensional feature space [31].

**Steps of the KPCA Method:**

1. Create the kernel matrix  $K1$  from the training dataset that defines in Eq. (1)

$$K1_{x,y} = K(m_x, m_y) \tag{1}$$

2. Calculate the Gram matrix  $K2$  using Eq. (2);

$$K2 = K - 1n K - K1n + 1nK1n \tag{2}$$

3. Resolve the vectors  $V1$  by using Eq. (3);

$$K_{V1k} = \lambda k P1_k \tag{3}$$

4. Calculate the KPCA Eq. (4);

$$M_k(m) = \phi(m)1Vi = \sum_{x=1}^k 1k K(m_x, m_y) \tag{4}$$

**Pseudo Code: Feature Extraction Using KPCA method**

- Process KPCA(m) [32]
- Assumed input:  $K1m * N \leftarrow [m1, m2, \dots, mN]$ .
- Collect  $K1c \leftarrow K1m * N$
- Kernel Matrix=  $Kim * N \quad K1m * M, \quad K1_{x,y} = K(m_x, m_y)$ .
- Concentration on the feature matrix
- $K1 = K1 - mK1/m - K1m/m + mK1m/m2$
- procedure  $K - PCA(X) 2$ : Given Input:  $XN \times M \leftarrow [x1, x2, \dots, xM]$ .
- Removing eigenvectors as;
- $K_{V1k} = \lambda k P1_k$
- Normalization as
- 

$$\alpha \leftarrow \frac{\alpha}{\text{mod}(\alpha) \sqrt{\lambda k P1_k}}$$

- Loop  $x \leftarrow 1: P1_k$  as
- $P1_k = \sum_{x=1}^k 1k K(m_x, m_y)$

**(iii) Feature selection using ABC method:** It feature selection using ABC method: The ABC method provides a smaller, but more specific subset of the original data by selecting important features over a group of features and removing the unimportant ones. The result of this process is a reduction in data dimensionality and an improvement in classification accuracy, both of which lead to faster processing time.

The ABC method operates as a population-based optimization method and takes inspiration from the smart foraging behavior of honeybee swarms. It involves three steps: employed, onlooker, and scout bees[33].

- **Employed Bee:** The employed bee creates a slight change in location through its innate learning and the development of a new source. This bee evaluates the fitness of the new source by comparing it to the nectar value of the previous source and chooses the one with the higher value. If the novel location is more than the old is forgotten. If the new location is superior, the previous one is forgotten. The location of the previous source is then retained in its memory. All the employed bees in the process share information about their location and the nectar source with the onlooker bees through the dance region.
- **Onlooker Bee:** The onlooker bee evaluates the information from all the employed bees and chooses a nectar source based on its potential value. The probability value of P1 is computed as;

$$P1 = \frac{Fit_x}{\sum_{m=1}^S Fit_m} \tag{1}$$

The Eq. (1) computes the probability value P1, where  $S_m$  is the food source and the number of employed and  $Fit_x$  solution is assumed in Eq. (2);

$$Fit_x = \frac{1}{1+f_1} \tag{2}$$

The  $f_1$  is assumed as the cost function of the clustering issue.

$$V_n^m = Z_{mn} + \phi_{mn} (Z_{mn} - Z_{pn}) \tag{3}$$

in equation (iii),  $p = 1, 2, 3, \dots, S_m$ ,  $\phi_{mn} =$  random number.

The Pseudo code of the ABC method represents. **(iv) Training and testing:** The training phase enables the system to train models and store them in



**Pseudo code: Feature selection using ABC method**

- Load preparing samples
- Produce the primary population  $x_m, m = 1, \dots, S_m$
- Estimate the  $f_1$  of the population and regular the cycle to initial **repeat**
- **For** each employed bee, Create a novel solution with eq(iii).
- Compute the value  $f_1$  Apply greedy collection development.
- Compute the P1 with eq (ii), **For** each onlooker bee P1 based novel solution  $V_{mn}$  Analyse the value  $f_1$ .
- If an available solution for scout, then swap it for a novel solution.
- Memorize the best solution  $C = C + 1$  until cycle = M.

a directory for future use. It combines the feature array and labeled vectors and trains them with a set number of iterations. The testing stage is used to evaluate the presentation of the trained model. It takes samples from the user and compares them to the training set to regulate the classification grades. The system processes the results to determine the performance and display the outcomes.

**(v) BFCNN model for disease classification**

The research model is separated into three phases: (i) data pre-processing, (ii) feature extraction (FE) and selection (FS), and (iii) disease classification, which was discussed in the previous section. A novel approach for detecting multiple diseases in medical datasets using a BEE-feathered CNN model. The BEE-feathered CNN model is a modified version of the traditional CNN model that incorporates the behavior of bee swarms to enhance the accuracy of the model. The bee-inspired algorithm used in the model involves the use of multiple artificial bees that simulate the behavior of real bees to search for optimal solutions. The proposed model was trained and tested using a dataset consisting of a medical dataset of patients with different diseases. The images were collected from various sources and labeled with the corresponding disease type. The dataset was split into training and validation sets, with 80% of the datasets used for training and 20% for validation. The proposed BEE-feathered CNN model is a promising approach for detecting multiple diseases in medical datasets. The model's high accuracy and ability to detect multiple diseases simultaneously can aid healthcare professionals in the accurate and timely diagnosis, ultimately improving patient outcomes. Fig. 2 illustrates the various steps involved in the research classification architecture. The feature sets for diseases are input into the planning model, which then arranges the mapped feature sets for disease prediction. This process assigns class labels and

determines the disease category of the patient. Convolutional neural network (CNN) [24] is used in this research method to classify the disease of each patient. The CNN model is created from a variety of different layers, with a specific stack of layers chosen for the system engineering. Fig. 3 illustrates the various layers of the CNN. The layers of the CNN include the convolutional layer (CL), rectified linear unit (ReLU), and pooling layer (PL), with options for average pooling and max pooling, and a fully connected layer (FCL). Algorithm 1 outlines the detailed steps of the research model.

The reason for the BFCNN model using multiple disease classification is an advanced ML method that gained immense popularity in the domain of medical analysis. The proposed model uses a multi-task learning method to classify multiple diseases simultaneously. It means that the model is trained to classify multiple diseases from a single dataset. The BFCNN model using multiple disease classification is a main enhancement over existing techniques for medical datasets. It is more accurate, efficient, and reliable as compared to the existing one.

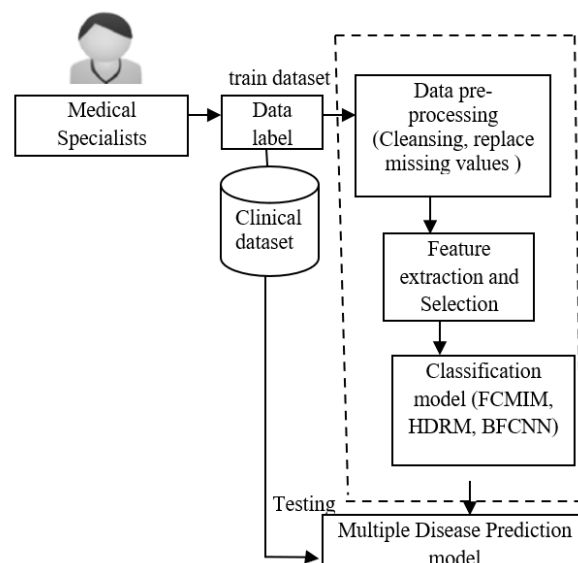


Figure. 2 Multiple disease diagnosis model

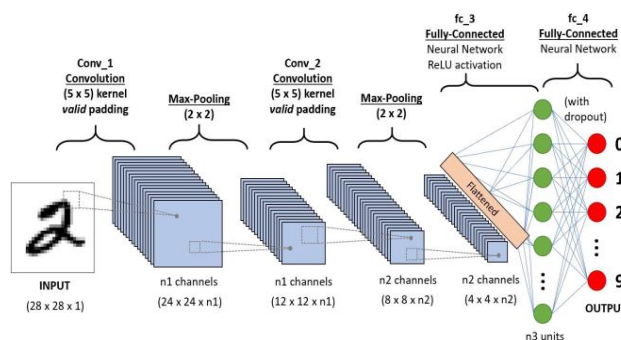


Figure. 3 Structural of the CNN model [24]



Algorithm 1: Multiple disease classification

```

INPUT: Datasetdb
OUTCOME: Category of diseases (heart, liver, and kidney).
1. Read db >= {db} || db → database
2. {db} train_set >={db} [0:G]
3. {db} test_set >={db} [G++:n]
4. CNN model = init_cnn(CNN,size ([hidden, weight]))
5. Set parameters (filter, padding, stride)
6. Define CNN including CONV LAYER (CNN, filter, padding, stride, sig (G))
7. Feature set = [[n+2p-k]/stride+1]
8. Defined PL (CNN, sub-sampling kind, sub-sampling rate) as labels \\ label (0,1,2)
9. Set classifier
For (init i in 0..labels) do
  If (i== labels.size) then
    Classifier = (CNN, labels)
  End if
End for

Category of diseases = (classifier, {db} test_set).
    
```

### 5. Result and discussions

This section outlines the simulation setup for predicting the disease classification model. The patients are identified based on the datasets saved for different diseases, such as liver, heart, and kidney. The research consists of three steps for disease identification, including data pre-processing, feature extraction and selection, and classification. The research methodology includes various techniques applied in each of these phases.

#### 5.1 Dataset utilized

The research involves datasets with various categories of diseases, including liver, kidney, and heart. Each disease category has its unique components, so no dataset contains similar components for all categories. The datasets were obtained from the UCI Machine repository[28, 29, 30]. Patient 1032 was selected and constructed in the specified dataset based on the medical factors and circumstances listed in Table 4. The dataset comprises 43 attributes and 3 output groups for each patient, and the data from the dataset is used for both preparation and analysis purposes.

The heart is capable of pumping blood throughout the body, and several factors can affect its functioning. The research focuses on metrics such as saturated fat and blood pressure (BP) aimed at this syndrome. Table 5 lists the various stages of cholesterol and BP that can disrupt the human body.

Table 4. Attributes related to health: dataset

Attributes	Heart[28]	Kidney[29]	Liver[30]
Age	Chest pain category	Sugar	TB
Hemo	Chol (mg/dl)	PC	Alkphos
PCV	FBS	PCC	Sgpt alamine
WBC	Resting ECG	Bacteria	Sgot aspartate
RBC	Max. heart rate	BGR	TP
Hyper tension	Exang	BU	Albumin
CAD	Slope	Sod	Sg

Table 5. HD: Level of cholesterol and BP metric [28]

Cholesterol	
Range (mg/dL)	Heart disease – Risk
160-189	High
190 and above	Very_high
BP	
Range of BP (mm hg)	Disease_risk
120-80	Normal means no infection
140-190	High means disease present

Table 6. Performance metrics

Metrics	Formula's
Accuracy	$\frac{T_P+T_N}{T_P+F_P+F_N+T_N}$
SP	$\frac{T_N}{T_N+F_P}$
SN	$\frac{T_P}{T_P+F_N}$

#### 5.2 Result analysis

The research employs an optimized CNN method to classify the disease category of a given patient. To estimate the presence of the suggested model, specificity (SP), Sensitivity (SN), and accuracy are used (as shown in Table 6). The accuracy of classification is calculated and constructed on the percentage of accurately predicted cases using four metrics: true positive (T\_P), true negative (T\_N), false positive (F\_P), and false negative (F\_N). Specificity measures the measurement of patients without the disease who are accurately identified, while sensitivity represents the percentage of patients with the infection who are perfectly detected.

**Abbreviations:** T\_P: true positive, T\_P: true negative, F\_P: false positive, F\_N: false negative.

The dataset gathered from heart, liver, and kidney diseases provides different properties, some of which are crucial in identifying the disease. To choose the most reliable information, the ABC (artificial bee

Table 7. Outcomes of multiple classification methods

Classification methods	Parameters		
	Accuracy	SP	SN
BFCNN	96.7	99	97
SAdHDRM	94.2	97.4	96.8
FCMIM	84.3	93.3	90.9

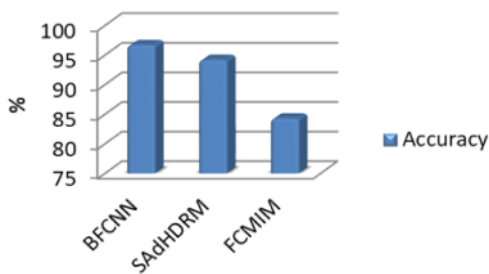


Figure. 4 Comparison analysis with BFCNN, SAdHDRM and FCMIM: Accuracy (%)

colony optimizer) is used to select the features, and the BFCNN model is used for prediction. The classification method with various metrics has been applied to multiple datasets, and the outcomes of the expectation technique in terms of accuracy, specificity (SP), and Sensitivity (SN), with and without features, are shown in Table 7.

This section compares the data to show that the implemented model with its developed feature selection and classification techniques performs expressively better than the original classification model. The ABC is recycled to select the most reliable information by evaluating the best score and fitness values, leading to improved outcomes for the classification methods and accurate disease analysis. The experimental analysis demonstrates the BFCNN model, a supervised-based model that works best with the convolutional neural network (CNN) method, achieving a maximum accuracy rate of 96.7%, as shown in Fig. 4.

This model manages missing and clean data for both the training and testing steps, with missing information typically being managed in the data pre-processing step. The BFCNN classification outperforms other techniques because of its deep learning classification method, reducing the error probability and increasing the accuracy rate, specificity (SP), and sensitivity (SN). The outcomes of the research BFCNN model are presented in this section. The classification procedure is illustrated in Fig. 4.

The outcomes of the forecast models are assessed using various parameters such as accuracy, specificity (SP), and sensitivity (SN) rates, which are shown in Figs. 4, 5, and 6. Table 5 presents the outcomes of the actual prediction classes as labeled

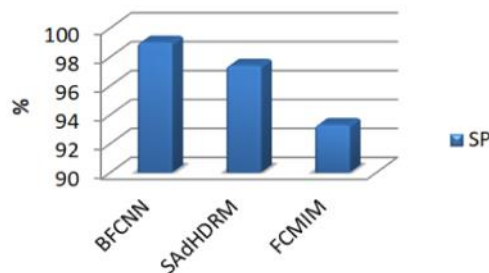


Figure. 5 Comparison analysis with BFCNN, SAdHDRM and FCMIM: Specificity (%)

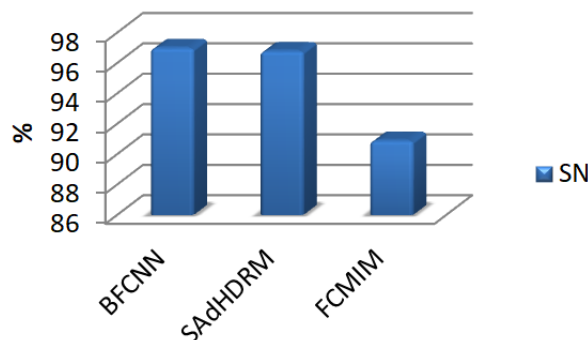


Figure. 6 Comparison analysis with BFCNN, SAdHDRM and FCMIM: Sensitivity (%)

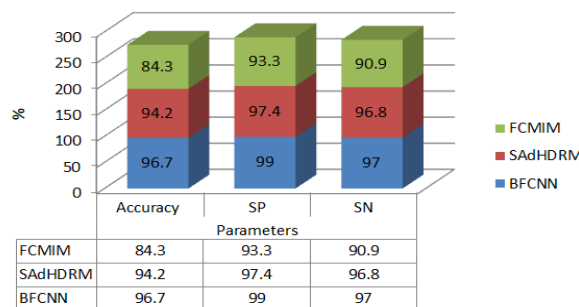


Figure. 7 Consilidated Comparison analysis with BFCNN, SAdHDRM and FCMIM

by the healthcare experts. The presentation metrics of the proposed model are calculated. The grades demonstrate that the research model achieved an accuracy of 96.7%, a specificity of 99%, and a sensitivity of 97%.

Fig. 7 shows the comparative analysis with different performance metrics and methods such as accuracy, specificity, and sensitivity rate. The proposed BFCNN has achieved a high accuracy rate as compared with the existing models such as SAdHDRIM [16], and FCMIM [28].

### 5.3 Discussions

The research model proposed in this study is aimed at classifying or predicting various diseases such as heart, liver, and kidney diseases for early-stage detection. The proposed model achieved a

maximum accuracy rate of 96.7 percent, surpassing other FCMIM [28] and SAdHDRM [16] models, making it more reliable and accurate due to its use of medical processes and specialist input. We have compared our proposed BFCNN model with existing models such as FCMIM [28] and SAdHDRM [16]. These models had lower accuracy rates and higher error probabilities, but our model has overcome these issues. It has achieved the lowest error rate and the highest accuracy rate, specificity, and Sensitivity.

## 6. Conclusion and future scope

In conclusion, this article has presented an optimized deep learning method for healthcare classification systems (HCCS). The model was tested using three different diseases (heart, liver, and kidney), and the outcomes were analyzed and constructed on metrics such as specificity, sensitivity, and accuracy. The proposed BEE-featured CNN model is a promising approach for detecting multiple diseases in medical datasets. The model's high accuracy and ability to detect multiple diseases simultaneously can aid healthcare professionals in the accurate and timely diagnosis, ultimately improving patient outcomes. The results were found to be highly capable and accurate, ranging between 91 percent and 99 percent, with an accuracy rate of 96.7 percent being achieved after training and testing the data. The proposed model achieved a maximum accuracy rate of 96.7 percent, surpassing other FCMIM [28] and SAdHDRM [16] models, making it more reliable and accurate due to its use of medical processes and specialist input. We have compared our proposed BFCNN model with existing models such as FCMIM [28] and SAdHDRM [16]. These models had lower accuracy rates and higher error probabilities, but our model has overcome these issues. It has achieved the lowest error rate and the highest accuracy rate, specificity, and sensitivity.

Future work on this research model will involve designing it as a deep-learning recommendation system for heart, liver, and kidney diseases. This will help identify risk factors and aid in early-stage disease detection.

## Conflicts of interest

There are no conflicts of interest declared by any of the authors.

## Author contributions

Authors proposed a Multi-Disease Classification Framework on Medical Datasets Using BFCNN Model to overcome the problem of existing models

and tested data sets on Matlab and results obtained are expected. The paper background work, conceptualization, methodology, dataset collection, implementation, result analysis and comparison, preparing and editing draft, visualization have been done by first author. The supervision, review of work and project administration, have been done by second author.

## References

- [1] M. R. Cowie, J. Bax, N. Bruining, J. G. Cleland, F. Koehler, and M. Malik, "e-Health: a position statement of the European Society of Cardiology", *European Heart Journal*, Vol. 37, No. 1, p. 63, 2016.
- [2] M. Lan, L. Samy, N. Alshurafa, M. K. Suh, H. Ghasemzadeh, and A. M. O. Connell, "Wanda: An end-to-end remote health monitoring and analytics system for heart failure patients", In: *Proc. of the Conference on Wireless Health*, Vol. 6, No. 8, pp. 1-8, 2012.
- [3] W. G. Members, D. L. Jones, R. J. Adams, T. M. Brown, M. Carnethon, S. Dai, and J. W. Rosett, "Heart disease and stroke statistics—2010 update: a report from the American Heart Association", *Circulation*, Vol. 121, No. 7, pp. e46-e215, 2010.
- [4] S. Chen, X. Zhu, S. Zhang, and J. Wang, "A framework for massive data transmission in a remote real-time health monitoring system", In: *Proc. of 18th International Conference on Automation and Computing (ICAC)*, Vol. 6, No. 9, pp. 1-5, 2012.
- [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 Research*, Vol. 2, No. 1, pp. 2-11, 2015.
- [6] R. Bellazzi, F. Ferrazzi, and L. Sacchi, "Predictive data mining in clinical medicine: a focus on selected methods and applications", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 1, No. 5, pp. 416-430, 2011.
- [7] T. Zheng, W. Xie, L. Xu, X. He, Y. Zhang, M. You, and Y. Chen, "A machine learning-based framework to identify type 2 diabetes through electronic health records", *International Journal of Medical Informatics*, Vol. 97, pp. 120-127, 2017.
- [8] T. S. Polonsky, R. L. McClelland, N. W. Jorgensen, D. E. Bild, G. L. Burke, A. D. Guerci, and P. Greenland, "Coronary artery calcium score and risk classification for coronary heart

- disease prediction”, *Jama*, Vol. 303, No. 16, pp. 1610-1616, 2010.
- [9] H. C. Koh, and G. Tan, “Data mining applications in healthcare”, *Journal of Healthcare Information Management*, Vol. 19, No. 2, p. 65, 2011.
- [10] M. D. Ramasamy, K. Periasamy, L. Krishnasamy, R. K. Dhanaraj, S. Kadry, and Y. Nam”, Multi-Disease Classification Model using Strassen’s Half of Threshold (SHoT) Training Algorithm in Healthcare Sector”, *IEEE Access*, Vol. 9, pp. 112624-112636, 2021.
- [11] M. Wang, Z. Wei, M. Jia, L. Chen, and H. Ji, “Deep learning model for multi-classification of infectious diseases from unstructured electronic medical records”, *BMC Medical Informatics and Decision Making*, Vol. 22, No. 1, pp. 1-13, 2022.
- [12] R. Berlin, R. Gruen, and J. Best, “Systems medicine disease: disease classification and scalability beyond networks and boundary conditions”, *Frontiers in Bioengineering and Biotechnology*, Vol. 6, p. 112, 2018.
- [13] A. Ahirrao, A. Bhagwat, P. Desai, S. Kaneri, and A. Shaikh, “Multi Disease Detection and Predictions Based On Machine Learning”, *International Journal of Emerging Technologies and Innovative Research*, Vol. 7, No. 2, pp. 950-953, 2020.
- [14] A. K. Dubey, “Optimized hybrid learning for multi disease prediction enabled by lion with butterfly optimization algorithm”, *Sādhanā*, Vol. 46, No. 2, pp. 1-27, 2021.
- [15] F. I. Alarsan, and M. Younes, “Analysis and classification of heart diseases using heartbeat features and machine learning algorithms”, *Journal of Big Data*, Vol. 6, No. 1, pp. 1-15, 2019.
- [16] A. Mustaqeem, S. M. Anwar, A. R. Khan, and M. Majid, “A statistical analysis based recommender model for heart disease patients”, *International Journal of Medical Informatics*, Vol. 108, pp. 134-145, 2017.
- [17] X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu, Q. Wang, and Q. Wang, “A hybrid classification system for heart disease diagnosis based on the RFRS method”, *Computational and Mathematical Methods in Medicine*, Vol. 3, No. 1, pp. 1-12, 2017.
- [18] C. Ellaji, G. Sreehitha, and B. L. Devi, “Efficient health care systems using intelligent things using NB-IoT”, *Materials Today: Proceedings*, Vol. 3, No. 4, pp. 1243-1255, 2020.
- [19] S. Sudharson, and P. Kokil, “Computer-aided diagnosis system for the classification of multi-class kidney abnormalities in the noisy ultrasound images”, *Computer Methods and Programs in Biomedicine*, Vol. 205, No. 1, p. 106071, 2021.
- [20] T. Muhammed, R. Mehmood, A. Albeshri, and I. Katib, “UbeHealth: a personalized ubiquitous cloud and edge-enabled networked healthcare system for smart cities”, *IEEE Access*, Vol. 6, pp. 32258-32285, 2018.
- [21] J. A. Alzubi, “Blockchain-based LamportMerkle Digital Signature: Authentication tool in IoT healthcare”, *Computer Communications*, Vol. 170, No. 2, pp. 200-208, 2021.
- [22] K. Sharma, and J. Virmani, “A decision support system for classification of normal and medical renal disease using ultrasound images: a decision support system for medical renal diseases”, *International Journal of Ambient Computing and Intelligence (IJACI)*, Vol. 8, No. 2, pp. 52-69, 2021.
- [23] N. Bhaskar, and S. Manikandan, “A deep-learning-based system for automated sensing of chronic kidney disease”, *IEEE Sensors Letters*, Vol. 3, No. 10, pp. 1-4, 2019.
- [24] P. Pasyar, T. Mahmoudi, S. Z. M. Kouzehkanan, S. Z. M. Ahmadian, A. Arabalibeik, H. Soltanian, N. A. R. Radmard, “Hybrid classification of diffuse liver diseases in ultrasound images using deep convolutional neural networks”, *Informatics in Medicine Unlocked*, Vol. 22, No. 3, p. 100496, 2021.
- [25] I. Mohit, K. S. Kumar, U. A. K. Reddy, B. S. Kumar, “An Approach to detect multiple diseases using machine learning algorithm”, In *Journal of Physics: Conference Series*, IOP Publishing, Vol. 2089, No. 1, p. 012009, 2021.
- [26] A. S. Hussein, W. M. Omar, X. Li, and M. Ati, “Accurate and reliable recommender system for chronic disease diagnosis”, *Global Health*, Vol. 3, No. 2, pp. 113-118, 2012.
- [27] T. N. T. Tran, A. Felfernig, C. Trattner, and A. Holzinger “Recommender systems in the healthcare domain: state-of-the-art and research issues”, *Journal of Intelligent Information Systems*, Vol. 3, No. 4, pp. 1-31, 2020.
- [28] J. P. Li, A. U. Haq, S. U. Din, J. J. Khan, A. Khan, A. Saboor, “Heart disease identification method using machine learning classification in e-healthcare”, *IEEE Access*, Vol. 8, No. 3, pp. 107562-107582, 2020.
- [29] *UCI Machine Learning Repository: ILPD (Indian liver patient dataset) data set*. Available at: [https://archive.ics.uci.edu/ml/datasets/ILPD+\(In](https://archive.ics.uci.edu/ml/datasets/ILPD+(In)

- dian+Liver+Patient+Dataset) (Accessed: January 24, 2023).
- [30] *UCI Machine Learning Repository: Chronic\_kidney\_disease Data set*. Available at: [http://archive.ics.uci.edu/ml/datasets/chronic\\_kidney\\_disease](http://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease) (Accessed: January 24, 2023).
- [31] M. M. Ahmadinejad, and E. Sherly, "A comparative study on pca and kpca methods for face recognition", *International Journal of Science and Research (IJSR)*, Vol. 5, No. 6, pp. 1-7, 2016.
- [32] *Nonlinear component analysis as a kernel eigenvalue ... - github pages* (no date). Available at: <https://namanuiuc.github.io/assets/projects/KPCA/report.pdf> (Accessed: January 25, 2023).
- [33] M. S. Uzer, N. Yilmaz, and O. Inan, "Feature selection method based on artificial bee colony algorithm and support vector machines for medical datasets classification", *The Scientific World Journal*, Vol. 2, No. 7, pp. 1-8, 2013.