

# Scaling Of Medical Disease Data Classification Based On A Hybrid Model Using Feature Optimization

Jayesh Mohanrao Sarwade<sup>1</sup>, Harsh Mathur<sup>2</sup>

<sup>1,2</sup>Department of Computer Science & Engineering Madhyanchal Professional University,  
Bhopal, India.

---

## Abstract

The classification of medical disease data leverages the healthcare informatics system. The variant of medical disease data is very complex such as document, numerical and constant. Therefore, the classification algorithms face a problem of data transformation and are compromised with the accuracy and sensitivity of classification algorithms. The increased classification rate of medical disease data promotes telemedicine's automatic diagnosis of critical illness. This paper proposed a feature optimization-based hybrid model for medical disease data classification. The feature optimization of disease data applies particle swarm optimization, and optimal features constraints increase the classification rate of proposed algorithms. The hybrid model of the classification algorithm is the process of clustering and classification. The proposed algorithm uses K-means and support vector machines (KSVM). The KSVM algorithm boosts the process of classification of medical disease data. The hybrid algorithm has been simulated in MATLAB environments and tested with a series of reputed datasets such as diabetics, cancer, hepatitis, lymphography and other three datasets. The performance of the proposed algorithm compares SVM and CNN. The analysis of evaluation results suggests that the proposed algorithm increase 2-3% of accuracy and sensitivity of data classification.

**Keywords:** - Data classification, Healthcare, Machine learning, CNN, PSO hybrid Model

## Introduction

The classification of medical disease data plays a vital role in the healthcare system. The versatile nature of human diseases such as heart disease, cancer, hepatitis and viral infection such as covid19. The prediction of symptoms and treatment required accurate information. The accuracy of information help to handle and control the spread of disease. The classification algorithm analyzed the diseases data and set some standard evaluation matrix worldwide [1,2]. The conventional classification algorithms such as data mining and neural network face a problem of classification accuracy. The recent trends of machine learning and swarm intelligence improve the classification efficiency and trustworthiness of medical disease data

classification [3,4,5]. Machine learning can reduce the cost of healthcare management while also facilitating the development of better doctor-patient relationships. However, there are various clinical difficulties, such as quick, precise, and reliable decision models. For accurate disease diagnosis, this issue must be addressed. Because of the large volume of unstructured data in healthcare, they categorise and quantify a conversation between a clinician and a patient [6,7,8]. Furthermore, because the bulk of medical datasets contains incomplete, redundant, irrelevant, and noisy information, the performance of classification models worsens [9,10]. Approximately one out of every seven adults in the United States have a mental illness. The machine learning algorithms support vector machine, decision tree, naïve Bayes MLP, CNN, and many more methods enhance medical disease data classification performance. The major issue related to the classification of medical disease data is mixed categories of data such as structured, instructed and mixed categories of data. The process of transformation and normalization handles the data related issues and improves the classification rate of a machine learning algorithm. Another issue of the classification algorithm is data balancing scenarios such as underfitting and overfitting. The balancing factors decrease the classification rate of the machine learning algorithm. The techniques and algorithms help build models and draw essential conclusions from the dataset. Some of the basic features are Age, Sex, Cholesterol, FBP, Rest ECG (Electrocardiogram) [14]. ECG (heart electrical activity test), number of major fluoroscopic vessels, thrust blood pressure, serum cholesterol, ST depression, pain location, hypertension, eating habits, smoking, weight, height and obesity [24] used in critical disease diagnosis. The features of medical disease data are very ambiguous in some cases; due to this performance of the classification algorithm is compromised. The ambiguity of features is minimized through feature optimization and feature selection approach. Most authors reported swarm-based optimization algorithms applied to select features and optimization. This paper uses the particle swarm optimization (PSO) algorithm for the feature optimization of disease data. Particle swarm optimization is a dynamic population-based heuristic algorithm. The concept of particle swarm optimization is birds of a fork. The applied swarm optimization algorithm enhances the performance of the proposed classification algorithm. This paper proposed a hybrid model for the classification of medical disease data—the hybrid model based on clustering and classification algorithms. The clustering algorithm is a better way to the group of data. The classification algorithm finds the categories of data—the support vector machine (SVM) algorithm combined with the K-means algorithm and proposed KSVM. The rest of paper follow order of section II. Related Work, in section III describes the proposed algorithm, in section IV describes the experimental analysis and in section V conclusion & future work.

## **II. Related work**

The medical data classification area is most prominent area for the society. The accurate classification rate improves the service and reliability of digital diagnosis of critical illness. The machine learning algorithm play a vital role in classification algorithms, the contribution of authors in recent years describes here.

In this [1] author present an FS method based on the MFO and a k-NN classifier. MFO is a new meta-heuristic algorithm that has demonstrated its ability to solve a variety of complex problems in an acceptable amount of time. The updated MFO method was assessed using

sixteen benchmark medical data sets, and the outcomes reveal that it performs well in terms of the assessment measures used. In this [2] author applied method for improving classification performance with a collection of meta-classifiers by combining wrapper subset evaluation with a meta-heuristic algorithm in a two-phase procedure. The feature domain is filtered in the first step using an attribute assessment method that uses the information gain ratio. The outcome reveal that the proposed technique improves Accuracy performance significantly when compared to other classifiers. In this [3] author utilize Swarm intelligence techniques to classify and pick features from various microarray datasets. In this study, they give a brief overview of microarray data and its various kinds. Furthermore, their research provides an overview of the most common swarm intelligence algorithms. In this [4] author proposed the Feature Selection methods and classification approaches They discovered that, when compared to Nave Bayes and K-nearest neighbour algorithms, the C4.5 method Decision tree produced the greatest outcomes for feature classification techniques. In this [5] author using a best-fit feature subset combination Simulated Annealing-Genetic Algorithms approach (SA-GA). the outcomes reveal that SAGA outperforms them in terms of accuracy and processing time. With a segmentation accuracy rate of 97.82 percent 0.74 for glioma tumor and 95.12 percent 3.21 for pituitary tumor, their method outperforms state-of-the-art technologies. In this [6] author choose features in illness classification, offer an ensemble-genetic learning strategy that uses wrapper feature reduction. The findings revealed that the Thallium Scan and vascular occlusion were the most essential aspects in the diagnosis of heart disease, with 97.57 percent accuracy in distinguishing between sick and healthy patients. In this [7] author propose an ensemble-genetic learning strategy for selecting features in disease classification utilizing wrapper feature reduction. The findings revealed that the Thallium Scan and vascular occlusion were the most essential aspects in the diagnosis of heart disease, with 97.57 percent accuracy in distinguishing between sick and healthy patients. In this [8] author propose The GREO algorithm is a meta-heuristic FS approach that uses a hybrid of GRO and EO algorithms. The XG Boost classifier receives the model's optimally picked features. LPC and LPCC based features are used as inputs, and the GREO method is used to optimize them. On two standard datasets, SAVEE and EMODB, they achieved excellent recognition accuracies of 97.31 percent and 98.46 percent, respectively. In this [9] author propose a process of determining the most informative genes as well as removing redundant and unnecessary genes is known as microarray data analysis. Feature selection is one of the most successful strategies for overcoming these obstacles. They want to evaluate and review the most current hybrid approaches that use bio-inspired evolutionary methods as the wrapper method in this research. In this [10] author proposed an intrusion detection system wrapper feature selection algorithm to handle the chosen method, this algorithm employs the Dynamic differential annealed optimizer (DDAO). The discussed method has outperformed such methods; several cases of excellent performance have been demonstrated. In this [11] author proposed software defect prediction, a hybrid heterogeneous ensemble approach. Heterogeneous ensembles are made up of a collection of classifiers from various learning base methods, each with its own set of strengths and weaknesses. The evaluation outcomes show that the ensemble version outperforms other well-known basic and ensemble classifiers. In this [12] author proposed the IPIO method is used to optimize the ELM initial weights and thresholds, and the modified ELM method is used to forecast the subseries.

In addition, multidimensional scaling and K-means clustering methods are used to divide forecasting outcomes into high frequency, medium-high frequency, medium-low frequency, and low frequency subseries. In this [13] author used the diagnosis of type II diabetes, an expert system-based ensemble model. The research revealed that the ensemble technique ensured an accuracy of 98.60 percent, which combines the predictive performance of multiple AI-based algorithms and is superior to all other individual counterparts. In this [14] author used the meta-heuristic optimized neural network to analyses patient health in order to predict various diseases. The efficiency of the system is evaluated using simulation outcomes, which ensure a maximum accuracy of 98.9 percent -ELT-COPD and 98 percent -NIH clinical dataset. In this [15] author proposed the vocal characteristics of people with Parkinson's disease are analyzed using sophisticated computational models. The outcomes show that the discussed algorithm outperformed the benchmarked models, demonstrating its efficacy in differentiating between affected and healthy PD samples.

In this [16] author proposed a less costly computational model for automatic COVID-19 detection from chest X-ray and CT-scan images Their work has a dual impact. The model was tested on three publicly available datasets, namely the COVID CT-dataset, the SARS-Cov-2 dataset, and the Chest X-Ray dataset, and achieved state-of-the-art accuracies of 98.32%, 95.35%, and 97.36%. In this [17] author proposed a hybrid method (MPG) for selecting gene expression is mentioned, which combines the Mutual information method with PSO and the GA. The discussed approach is appropriate for gene data set analysis using various classifier techniques and demonstrates the higher efficiency and accuracy of discussed data sets when compared to traditional selection mechanisms. In this [18] author examines the entire body of literature on meta-heuristic-based text clustering applications, including numerous variants such as basic, modified, hybridized, and multi-objective methods. The main procedures for text clustering and critical discussions are also provided. In this [19] author use of fifteen real-world project data sets obtained from the PROMISE repository. There are three different classifiers used k-NN, DT, and LDA. The reported outcomes show that the discussed EBMFO improves the overall performance of classifiers and outperforms the outcomes in the literature, demonstrating the importance of TF for feature selection algorithms. In this [20] author proposed Ring Theory based Harmony Search is a new hybrid meta-heuristic FS model based on a well-known meta-heuristic HS algorithm and a recently discussed Ring Theory based Evolutionary Algorithm. The research findings show that RTHS outperforms the state-of-the-art methods compared here. In this [21] author distinguish MCI patients based on RS-FMRI, a new RS-FMRI data analysis approach based on CBGWO and ANFIS is presented. This study suggests that the discussed CBGWO-ANFIS approach on RS-FMRI could be a useful tool for the early detection of MCI. In this [22] author used optimize feature selection, the discussed HBBO, a combination of ABC's onlooker bee and the migration operator of biogeography-based optimization. According to the outcomes, the overall accuracy and average accuracy of HBBO SVM are 95.21 percent and 94.21 percent, respectively, which is the best outcomes when compared to other outcomes. In this [23] author propose an ensemble classification approach for imbalanced classification based on feature space partitioning A hybrid metaheuristic called GACE is used to optimize the various parameters related to feature space partitioning. According to the outcomes of the analysis, its performance is comparable to that

of reference techniques in the literature. In this [24] approach, users generate training sets by clustering an initial set of time series with respect to pollution values, and they propose a novel Clustering-based Ensemble model for air pollution prediction based on evolving Spiking Neural Networks, with each ESNN network trained on a different set of time series. In terms of forecasting accuracy, the experimental outcomes show that the discussed ensemble model outperforms the other three models. In this [25] author propose a strong meta-heuristic feature selection method, as well as a hybrid NB and SWRF classification method The features are first extracted from the pre-processed image and then chosen using the multi-objective enhanced firefly algorithm. The image is segmented using an ensemble of NB and SWRF classifiers after the feature selection process. The NB classifier is trained and used to estimate the weights of training samples in this case. In this [26] author used Scopus and Web of Science databases for bibliometric analysis This bibliometric analysis takes into account important keywords, datasets, and the significance of the research study under consideration. In some ways, this research work can be beneficial to researchers interested in contributing to the field of feature selection and optimization in healthcare. In this [27] author proposed the Levy flight and evolutionary selection operators into the MFO algorithm, an effective wrapper approach the main purpose is to solve the FS problem in medical applications. However, FS is a difficult NP-hard problem because the search space grows exponentially as the number of features increases. In this [28] author The MFO optimizer has been used as a search strategy within a wrapper approach to take advantage of the moths' swarming behaviour. The Levy flight operator is discussed in order to improve the MFO's exploratory behaviour and alleviate stagnation in local minima. The proposed methods are validated using 23 medical data sets sourced from reputable data repositories. In this [29] author proposed a new ensemble clustering framework based on cluster-level weighting. The level of certainty that a given ensemble has about a cluster is considered the cluster's reliability. The final ensemble is created by selecting the best clusters and weighting each chosen cluster based on its reliability. In this [30] author propose a feature selection framework based on a tri-stage wrapper-filter for medical report-based disease detection in the first stage, four filter methods—Mutual Information, Relief F, Chi Square, and Variance—were used to create an ensemble, and then each feature from the union set was evaluated by three classification algorithms—support vector machine, naive Bayes, and k-nearest neighbours—and an average accuracy was calculated.

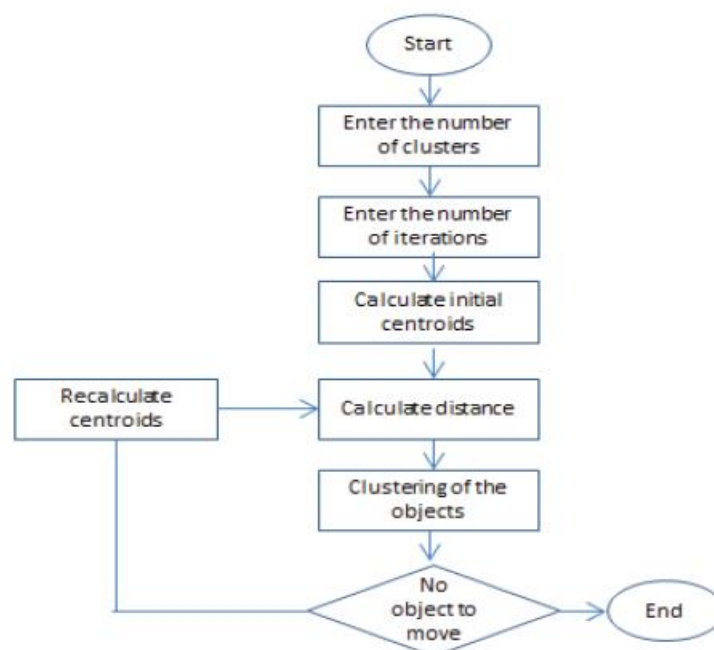
### **III. Proposed Algorithm**

The proposed algorithm of medical disease data classification encapsulates K-means algorithm, particle swarm optimization and support vector machine. The k-means algorithm is simple and fast clustering algorithm for the grouping of data. the processing of K-means algorithm work as feature reduction of disease data. particle swarm optimization algorithm work as feature selection process along with k-means algorithm. The population of PSO algorithm is medical disease data. the process of algorithm describes in three sections. In 1<sup>st</sup> section describes K-means clustering algorithm, in 2<sup>nd</sup> section describes the particle swarm optimization (PSO) and support vector machine and in 3<sup>rd</sup> section describes the proposed algorithm.

## 1<sup>st</sup> section

### K-means Algorithm

K-means clustering algorithm is subgroup of unsupervised learning. The process of data grouping is no need of data label. The K-means clustering algorithm work as function of similarity of data objects and formed the group, all task of clustering done by the iteration function. Uncontrolled tasks such as clustering, which would analyse and characterise the data set before beginning a controlled learning aim, are frequently used. Because clustering is done without the use of a class name, some sense of similarity must be derived from object properties. The clustering algorithm is used to describe commonalities and the technique in which dots are grouped. The K-means algorithm is a straightforward iterative clustering technique that separates a set of data into the number of clusters,  $k$ , that the user specifies. The method is simple to construct and run very quickly, as well as versatile and widely used. It is one of the most important data mining algorithms in past[14,25,28]. The processing of K-means algorithm describes in figure 1.



**Figure 1** process of K-means clustering algorithm.

## 2<sup>nd</sup> section

### Particle Swarm optimization (PSO)

PSO is a random optimization method inspired by the foraging behaviour of birds or fish. In the PSO, each particle represents a candidate of an optimized problem, and all the particles form a set of candidate solutions. Each particle adjusts its position by learning two best positions, i.e., the personal best position (or the personal leader) that the particle found so far itself, and let the global best position (or the global leader) found by neighbours of this particle so far. Considering a swarm including  $N_p$  particles[20,21,22],

let the position of the  $i$ -th particle be  $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ , its velocity be  $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})$ , its personal and global leaders be  $P_{bi} = (P_{bi1}, P_{bi2}, \dots, P_{biD})$  and  $G_b = (G_{b1}, G_{b2}, \dots, G_{bD})$ , then the particle is updated as the following:

$$V_{ij}^{t+1} = W V_{ij}^t + c_1 r_1 (P_{bij}^t - X_{ij}^t) + c_2 r_2 (G_{bj}^t - X_{ij}^t) \dots \dots \dots (1)$$

$$X_{ij}^{t+1} = X_{ij}^t + v_{ij}^{t+1}, j = 1, 2, \dots, D \dots \dots \dots (2)$$

where  $t$  is the iteration,  $D$  is the size of decision variables,  $w$  is the inertia coefficient,  $c_1$  and  $c_2$  are two acceleration factors,  $r_1$  and  $r_2$  are two random numbers within  $[0,1]$ .

### Support Vector machine (SVM)

SVM (Support vector machine) is machine learning algorithm derived by Vapnik in 1990. The support vector machine applied in various filed of image classification and pattern recognition. The nature of support vector machine is linear, non-linear and sigmoid. The non-linear support vector machine mapping the feature data with respect to one plane to another plan. The separation of data plan is non-linear and decision factor correlate with margin function of support vector. The hyperplane of equation is derived as

$$W \cdot x_i + b \geq 1 \text{ if } y_i = 1 \tag{3}$$

$$W \cdot x_i + b \leq -1 \text{ if } y_i = -1$$

Here  $W$  is weight vector,  $x$  is input vector  $y_i$  label o class and  $b$  is bias.

The minimization formulation of support vector

$$\begin{aligned} &\text{Minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \epsilon_i, i = 1, 2, \dots, n \\ &\text{subject to } y_i (w^T D \cdot x_i + b) \geq 1 - \epsilon_i \\ &\epsilon_i \geq 0 \quad i = 1, 2, \dots, n \dots \dots \dots (4) \end{aligned}$$

Here  $C$  is constant,  $n$  is number of observation and  $\epsilon_i$  is slack variable.

The rule of decision function is

$$f(x) = \sum_{i=1}^n y_i \alpha_i K(x_i, x_j) + b \dots \dots \dots (5)$$

### 3<sup>rd</sup> section

#### Proposed Algorithm

The proposed algorithm has been two algorithms one is algorithm for feature selection and reduction with K-means and particle swarm optimization (PSO) and another algorithm 2 describes the integration of support vector machine with optimal data of clustering algorithm.

#### Algorithm-1

Input:

D- disease data

F number of features of D

K number of cluster

C seeds of cluster

P population of PSO

for each  $i, j \in \{f_1, f_2, f_3, \dots, f_n, |c|\}$  and  $i \neq j$  do

Set iter = 0,  $P_D^c(i, j) = 100$

Measure similarity of data points

$$X = \sqrt{X_i - X_j}$$

Processing of data as

$\{C_1, C_2, C_3, \dots, C_n\}$

while iter  $\leq$  f do

Merge clusters  $C_1, C_2, \dots, C_n$

Formed intermediate clusters

Update new population of features  $\{f_1, f_2, \dots, f_n\}$

Iter=iter+1

Output:

Pair of clusters of indices i and j such that  $k_1, k_2, \dots, k_n$  for the processing of algorithm-2

#### Algorithm-2

The optimal cluster  $K_1, k_2$  is reduces dataset of D,  $\{k_1, k_2, k_3, \dots, k_n\} \in D$

$X_i$  vector of features  $\langle f_1, f_2, f_3, \dots, f_n \rangle$

$Y_i$  class variable with  $Y_i \in [0, 1]$

Put set of clusters(  $k_1, k_2, \dots, k_n$ )

Define Voting(  $C_0, C_1$ )



Set iter=0

While iter< maxiter do

For each class

train Cluester on I iteration

Set test data of class D1,D2,.....,Dn

Update

$$K_{class}^{iter+1} < - - k_{class}^{iter} U\{ X_j, \arg \max P(Y = class|X_i) \forall j \in K\}$$

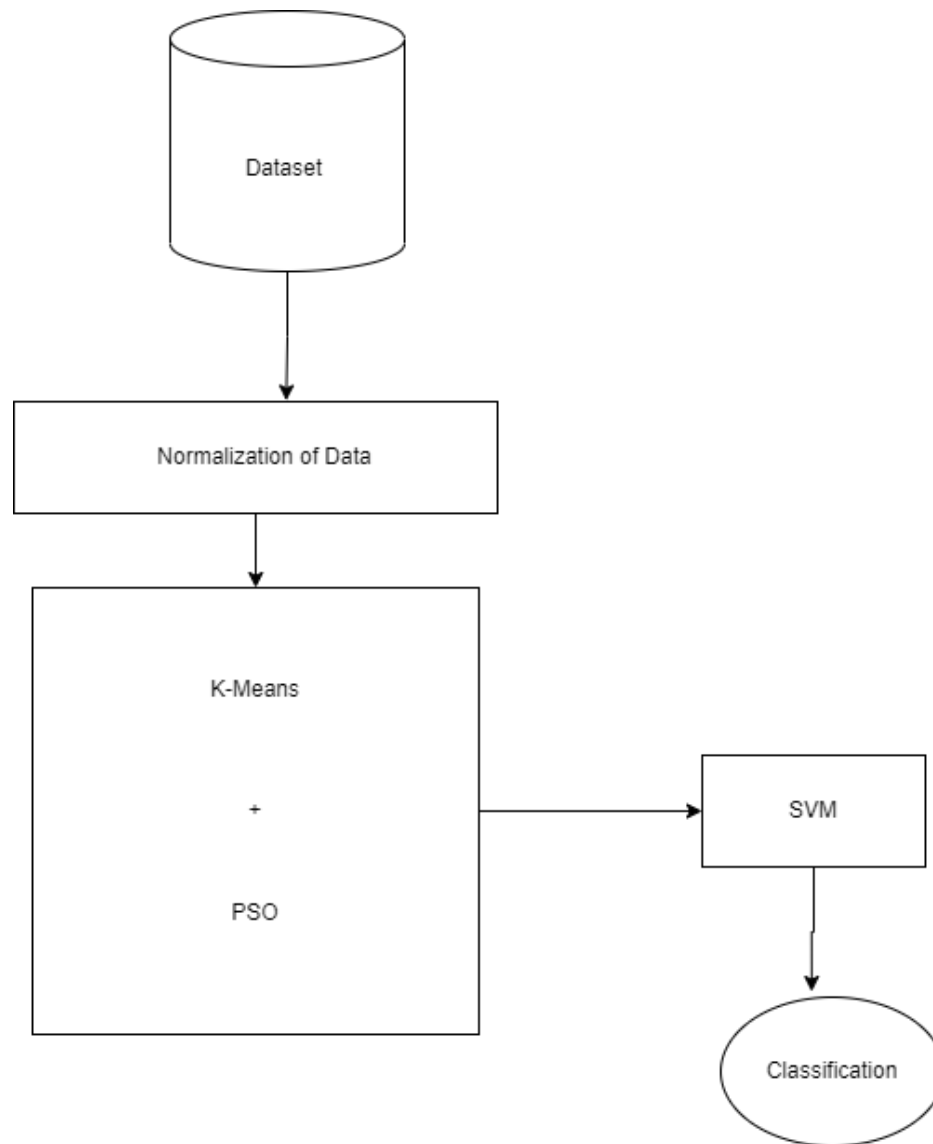
Iter=iter+1

Output

Predict class label of test data

Estimate confusion matrix

Estimate accuracy, sensitivity and specificity



**Figure 2** process block diagram of hybrid model of K-means and support vector machine.

#### IV. Experimental Analysis

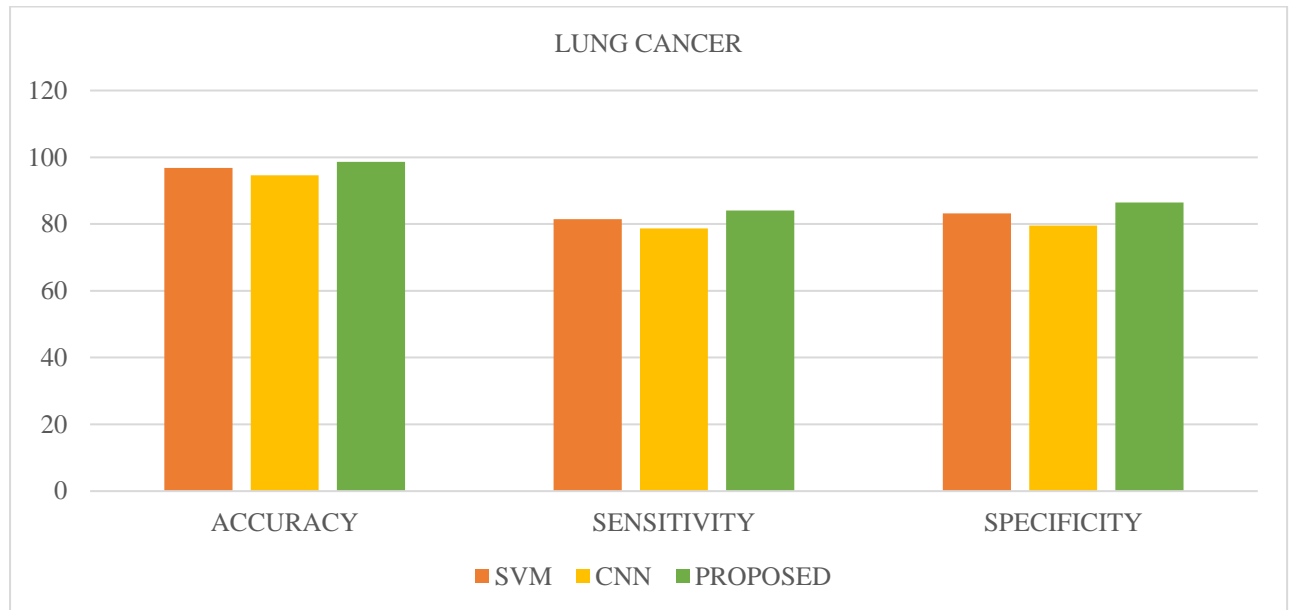
To validate the proposed algorithm, implement in MATLAB software. The version of software is R2014a, and the configuration of system is I7 processor, 16GB RAM and windows10 operating system. The MATLAB provides the basic support library file of SVM and K-means. But the other function of classifier defines and programmed with function file and compile with library file. For the process of detection applied UCI machine learning dataset. The process of sample of data applied 10 cross folds for the processing of prediction and measurement of parameters such as accuracy and precision. The name of datasets mentions in table. Table also mention the feature selection approach for the classification [28,29,30]. the standard parameters

$$\text{Accuracy} = \frac{\text{Total No. of Correctly Classified Instances}}{\text{Total No. of Instances}} \times 100$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100$$

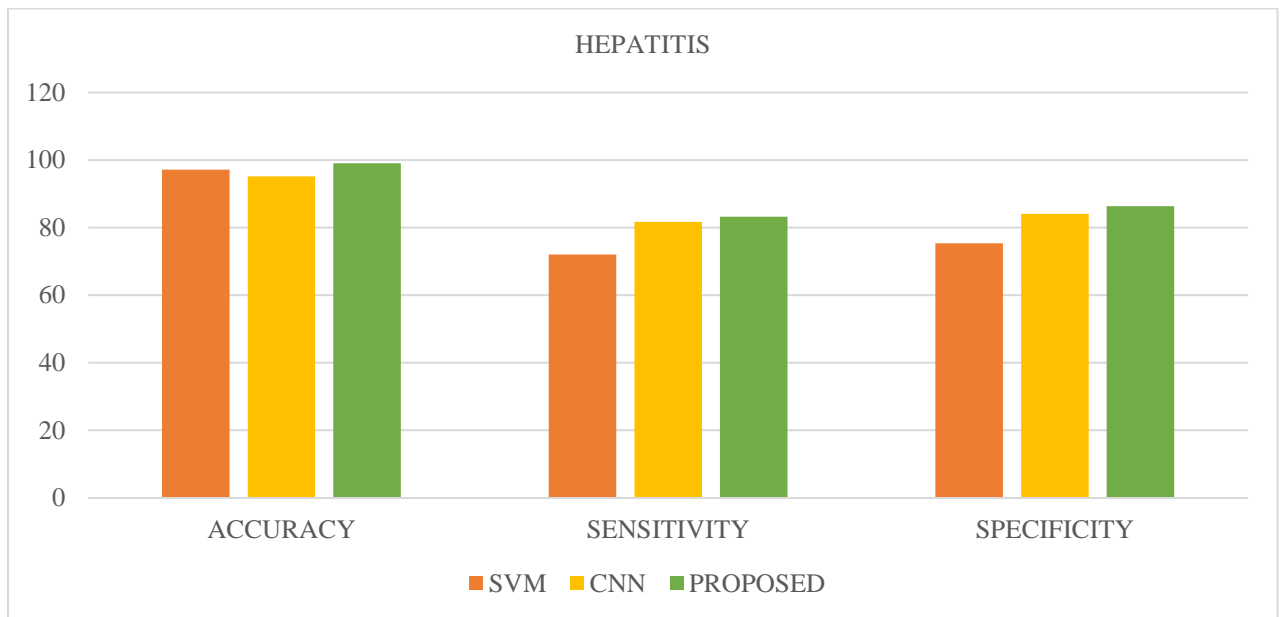
$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100$$

**Table 1: Estimated results of three algorithms SVM, CNN and proposed as accuracy, sensitivity and specificity**



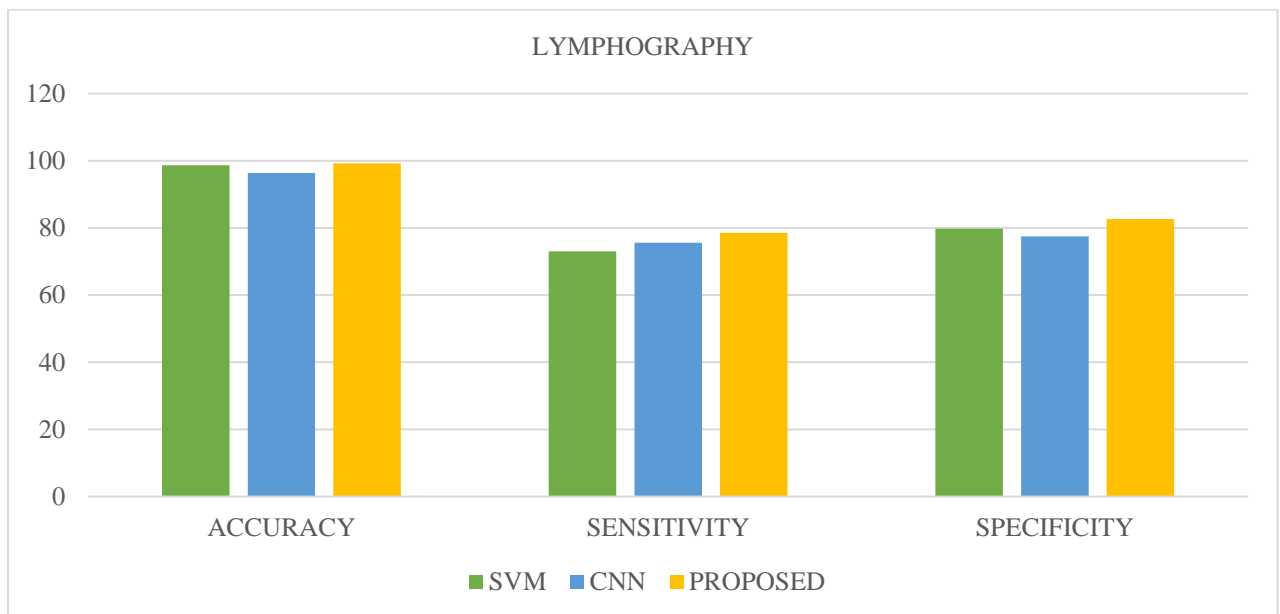
**Figure 3: represents the value of accuracy, sensitivity and specificity of lung cancer.**

The proposed algorithm improves the performance of lung cancer data in terms of accuracy, sensitivity and specificity the comparative analysis of proposed algorithm with CNN and SVM. The difference of results improves the accuracy 98.63, sensitivity 84.10 and specificity 86.41.



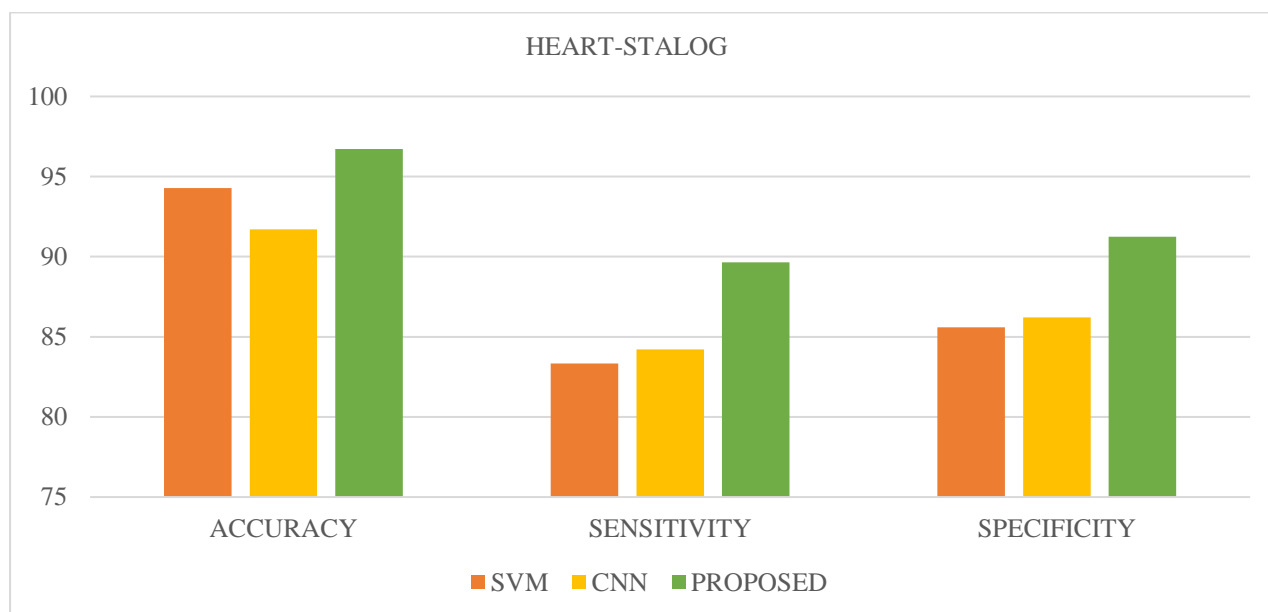
**Figure 4:** represents the value of accuracy, sensitivity and specificity of Hepatitis.

The proposed algorithm improves the performance of Hepatitis data in terms of accuracy, sensitivity and specificity the comparative analysis of proposed algorithm with CNN and SVM. The difference of results improves the accuracy 99.02, sensitivity 83.21 and specificity 86.32.



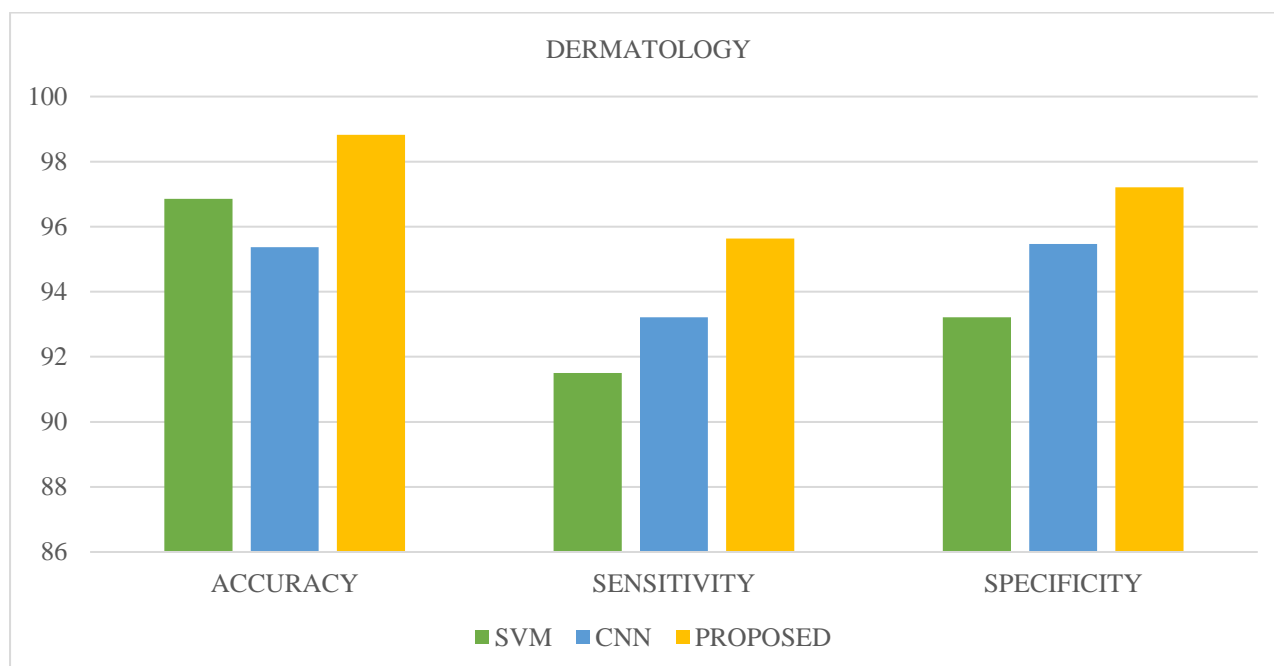
**Figure 5:** represents the value of accuracy, sensitivity and specificity of Lymphography.

The proposed algorithm improves the performance of Lymphography data in terms of accuracy, sensitivity and specificity the comparative analysis of proposed algorithm with CNN and SVM. The difference of results improves the accuracy 99.21, sensitivity 78.56 and specificity 82.61.



**Figure 6:** represents the value of accuracy, sensitivity and specificity of Heart-Stalog.

The proposed algorithm improves the performance of Heart-Stalog data in terms of accuracy, sensitivity and specificity the comparative analysis of proposed algorithm with CNN and SVM. The difference of results improves the accuracy 96.71, sensitivity 89.65 and specificity 91.24.



**Figure 7:** represents the value of accuracy, sensitivity and specificity of dermatology.

The proposed algorithm improves the performance of dermatology data in terms of accuracy, sensitivity and specificity the comparative analysis of proposed algorithm with CNN and SVM. The difference of results improves the accuracy 98.82, sensitivity 95.64 and specificity 97.21.

## V. Conclusion & Future Work

The hybrid algorithm for the classification of medical disease data is very efficient in the manner of accuracy and other measuring parameters. The process of feature optimization using particle swarm optimization reduces the ambiguous and irrelevant features of data and improves the validation of the clustering algorithm. The experimental results validate the importance of the proposed algorithm. We have compared the performance of the proposed algorithm with the two algorithms, SVM and CNN, on the five publicly available medical disease datasets. The experimental results showed that the proposed algorithm could achieve the best classification accuracy with relatively more minor running time on most of the datasets, indicating that it has strong competitiveness for classifying extensive medical data. Despite the promising results, there are still some limitations in work. First, the feature clustering method plays a vital role in medical disease. More sophisticated clustering methods that do not need to set any threshold manually need to be designed. It may be a feasible way to automatically adjust the threshold based on the algorithm's feedback result.

## Reference

- [1].Khurma, Ruba Abu, Ibrahim Aljarah, and Ahmad Sharieh. "Rank based moth flame optimisation for feature selection in the medical application." In 2020 IEEE congress on evolutionary computation (CEC), pp. 1-8. IEEE, 2020.
- [2].Almayyan, Waheeda. "Improved Discriminatory Ability using Hybrid Feature Selection via Approach Inspired by Grey Wolf Search and Ensemble Classifier for Medical Datasets." *International Journal of Computer Science and Information Security (IJCSIS)* 19, no. 3 (2021).
- [3].Jahwar, Alan, and Nawzat Ahmed. "Swarm intelligence algorithms in gene selection profile based on classification of microarray data: A review." *Journal of Applied Science and Technology Trends* 2, no. 01 (2021): 01-09.
- [4].Rajabhushanam, C. "A Survey on Feature Selection to Perform Classification Using Meta Heuristic Algorithms in Data Mining Domain." *Journal of Critical Reviews* 7, no. 10 (2020): 2031-2042.
- [5].Kharrat, Ahmed, and N. E. J. I. Mahmoud. "Feature selection based on hybrid optimization for magnetic resonance imaging brain tumor classification and segmentation." *Applied Medical Informatics*. 41, no. 1 (2019): 9-23.
- [6].Yılmaz, Asuman Günay, Ugur Turhal, and Vasif V. NabiyeV. "Effect of feature selection with meta-heuristic optimization methods on face spoofing detection." *J Mod Technol Eng* 5, no. 1 (2020): 48-59.
- [7].Abdollahi, Jafar, and Babak Nouri-Moghaddam. "Feature selection for medical diagnosis: Evaluation for using a hybrid Stacked-Genetic approach in the diagnosis of heart disease." *Ar Xiv preprint arXiv:2103.08175* (2021).
- [8].Dey, Arijit, Soham Chattopadhyay, Pawan Kumar Singh, Ali Ahmadian, Massimiliano Ferrara, and Ram Sarkar. "A Hybrid Meta-Heuristic Feature Selection Method Using Golden Ratio and Equilibrium Optimization Algorithms for Speech Emotion Recognition." *IEEE Access* 8 (2020): 200953-200970.

- [9]. Almugren, Nada, and Hala Alshamlan. "A survey on hybrid feature selection methods in microarray gene expression data for cancer classification." *IEEE Access* 7 (2019): 78533-78548.
- [10]. Wilson, Allan J., and S. Giriprasad. "A Feature Selection Algorithm for Intrusion Detection System Based on New Meta-Heuristic Optimization." *J. Soft Comput Eng. Appl* 1, no. 1 (2020).
- [11]. Alsawalqah, Hamad, Neveen Hijazi, Mohammed Eshtay, Hossam Faris, Ahmed Al Radaideh, Ibrahim Aljarah, and Yazan Alshamaileh. "Software defect prediction using heterogeneous ensemble classification based on segmented patterns." *Applied Sciences* 10, no. 5 (2020): 1745.
- [12]. Jiang, Feng, Jiaqi He, and Tianhai Tian. "A clustering-based ensemble approach with improved pigeon-inspired optimization and extreme learning machine for air quality prediction." *Applied Soft Computing* 85 (2019): 105827.
- [13]. Sarwar, Abid, Mehbob Ali, Jatinder Manhas, and Vinod Sharma. "Diagnosis of diabetes type-II using hybrid machine learning based ensemble model." *International Journal of Information Technology* 12, no. 2 (2020): 419-428.
- [14]. Abugabah, Ahed, Ahmad Ali AlZubi, Feras Al-Obeidat, Abdulaziz Alarifi, and Ayed Alwadain. "Data mining techniques for analyzing healthcare conditions of urban space-person lung using meta-heuristic optimized neural networks." *Cluster Computing* 23 (2020): 1781-1794.
- [15]. Xiong, Yanhao, and Yaohua Lu. "Deep feature extraction from the vocal vectors using sparse autoencoders for Parkinson's classification." *IEEE Access* 8 (2020): 27821-27830.
- [16]. Chattopadhyay, Soham, Arijit Dey, Pawan Kumar Singh, Zong Woo Geem, and Ram Sarkar. "COVID-19 detection by optimizing deep residual features with improved clustering-based golden ratio optimizer." *Diagnostics* 11, no. 2 (2021): 315.
- [17]. Singh, Rachhpal. "A gene expression data classification and selection method using hybrid meta-heuristic technique." *EAI Endorsed Transactions on Scalable Information Systems* 7, no. 25 (2020).
- [18]. Abualigah, Laith, Amir H. Gandomi, Mohamed Abd Elaziz, Husam Al Hamad, Mahmoud Omari, Mohammad Alshinwan, and Ahmad M. Khasawneh. "Advances in meta-heuristic optimization algorithms in big data text clustering." *Electronics* 10, no. 2 (2021): 101.
- [19]. Tumar, Iyad, Yousef Hassouneh, Hamza Turabieh, and Thaeer Thaeer. "Enhanced binary moth flame optimization as a feature selection algorithm to predict software fault prediction." *IEEE Access* 8 (2020): 8041-8055.
- [20]. Ahmed, Shameem, Kushal Kanti Ghosh, Pawan Kumar Singh, Zong Woo Geem, and Ram Sarkar. "Hybrid of harmony search algorithm and ring theory-based evolutionary algorithm for feature selection." *IEEE Access* 8 (2020): 102629-102645.
- [21]. Anter, Ahmed M., Yichen Wei, Jiahui Su, Yueming Yuan, Beiying Lei, Gaoxiong Duan, Wei Mai et al. "A robust swarm intelligence-based feature selection model for neuro-fuzzy recognition of mild cognitive impairment from resting-state fMRI." *Information Sciences* 503 (2019): 670-687.

- [22]. Rostami, Omid, and Mehrdad Kaveh. "Optimal feature selection for SAR image classification using biogeography-based optimization (BBO), artificial bee colony (ABC) and support vector machine (SVM): a combined approach of optimization and machine learning." *Computational Geosciences* 25, no. 3 (2021): 911-930.
- [23]. Lopez-Garcia, Pedro, Antonio D. Masegosa, Eneko Osaba, Enrique Onieva, and Asier Perallos. "Ensemble classification for imbalanced data based on feature space partitioning and hybrid metaheuristics." *Applied Intelligence* 49, no. 8 (2019): 2807-2822.
- [24]. Maciag, Piotr S., Nikola Kasabov, Marzena Kryszkiewicz, and Robert Bembenik. "Air pollution prediction with clustering-based ensemble of evolving spiking neural networks and a case study for London area." *Environmental Modelling & Software* 118 (2019): 262-280.
- [25]. Babu, Melingi Sunil, and V. Vijayalakshmi. "An effective approach for sub-acute Ischemic stroke lesion segmentation by adopting meta-heuristics feature selection technique along with hybrid Naive Bayes and sample-weighted random forest classification." *Sensing and Imaging* 20, no. 1 (2019): 1-24.
- [26]. Joshi, Rahul, Harita Gadikta, Saneeka Kharat, Soumi Mandal, Kalyani Kadam, and Siddhant Pandit. "Bibliometric of Feature Selection Using Optimization Techniques in Healthcare using Scopus and Web of Science Databases." *Library Philosophy and Practice* (2021): 1-38.
- [27]. Khurma, Ruba Abu, Ibrahim Aljarah, and Ahmad Sharieh. "A Simultaneous Moth Flame Optimizer Feature Selection Approach Based on Levy Flight and Selection Operators for Medical Diagnosis." *Arabian Journal for Science and Engineering* (2021): 1-26.
- [28]. Khurma, Ruba Abu, Ibrahim Aljarah, and Ahmad Sharieh. "A Simultaneous Moth Flame Optimizer Feature Selection Approach Based on Levy Flight and Selection Operators for Medical Diagnosis." *Arabian Journal for Science and Engineering* (2021): 1-26.
- [29]. Nazari, Ahmad, Ayob Dehghan, Samad Nejatian, Vahideh Rezaie, and Hamid Parvin. "A comprehensive study of clustering ensemble weighting based on cluster quality and diversity." *Pattern Analysis and Applications* 22, no. 1 (2019): 133-145.
- [30]. Tan, Teck Yan, Li Zhang, and Chee Peng Lim. "Adaptive melanoma diagnosis using evolving clustering, ensemble and deep neural networks." *Knowledge-Based Systems* 187 (2020): 104807.