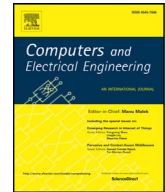




Contents lists available at ScienceDirect

## Computers and Electrical Engineering

journal homepage: [www.elsevier.com/locate/compeleceng](http://www.elsevier.com/locate/compeleceng)

## Novel classifiers for intelligent disease diagnosis with multi-objective parameter evolution<sup>☆</sup>

Nalluri MadhuSudana Rao<sup>a</sup>, Krithivasan Kannan<sup>a</sup>, Xiao-zhi Gao<sup>b</sup>,  
Diptendu Sinha Roy<sup>c,\*</sup>

<sup>a</sup> SASTRA Deemed to be University, Thanjavur, Tamilnadu, India

<sup>b</sup> Lappeenranta University of Technology, Finland

<sup>c</sup> National Institute of Technology, Meghalaya, India

### ARTICLE INFO

#### Article history:

Received 15 July 2017

Revised 26 January 2018

Accepted 29 January 2018

Available online xxx

#### Keywords:

Medical engineering data

Disease classifier

Data mining

Competitive collection, Sequential Minimal

Optimization (SMO)

Support Vector Machine (SVM)

Initial population generation

EHO

MOEA/D

NSGA-II

### ABSTRACT

In this research, intelligent classifiers for disease diagnosis are designed that use classifier parameters, such as cost, tolerance, gamma and epsilon, with multi-objective evolutionary algorithms. The multiple objective functions are prediction accuracy, sensitivity and specificity. This paper employs a Sequential Minimal Optimization (SMO), a variant of the classical Support Vector Machine (SVM), as the base classifier in conjunction with three popular evolutionary algorithms (EA), namely, Elephant Herding Optimization (EHO), Multi-objective Evolutionary Algorithm Based on Decomposition (MOEA/D) and Non-dominated Sorting Genetic Algorithm-II (NSGA-II), for parameter evolution. A new cuboids based initial population generation mechanism was also introduced to hybridize EHO, called CEHO. The performance of CEHO is compared with the other three EAs (EHO, MOEA/D and NSGA-II) over 17 medical engineering datasets, and pertinent statistical tests were conducted to substantiate their performances. The results demonstrate that the proposed CEHO exhibit better to competitive results across all datasets.

© 2018 Elsevier Ltd. All rights reserved.

## 1. Introduction

Machine learning, data mining and knowledge engineering techniques are being increasingly applied to assist healthcare professionals and doctors in analyzing medical data, which otherwise are challenging to handle [1]. It is well accepted in the healthcare industry that machine learning and data mining can support clinical decision making functions, including disease diagnoses, alternative treatment choices, and prognosis prediction, as well as administrative decisions such as staffing estimation and quality assurance [2].

Several machine learning and classification algorithms, such as the Naïve Bayes classifier, the SVM, decision trees (DT), fuzzy decision trees (FDT), neural networks, and the K-means clustering algorithm, have been applied to medical diagnoses, including various forms of cancer, cardio-vascular diseases, appendicitis, skin diseases, thyroid problems, liver problems, brain tumors, fertility problems, Parkinson's disease, and hepatitis. However, no single classifier can satisfactorily perform

<sup>☆</sup> Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. R. C. Poonia.

\* Corresponding author.

E-mail addresses: [madhu031083@gmail.com](mailto:madhu031083@gmail.com) (N. MadhuSudana Rao), [kannan@maths.sastra.edu](mailto:kannan@maths.sastra.edu) (K. Kannan), [xiao.z.gao@gmail.com](mailto:xiao.z.gao@gmail.com) (X.-z. Gao), [diptendu.sr@nitm.ac.in](mailto:diptendu.sr@nitm.ac.in) (D.S. Roy).

<https://doi.org/10.1016/j.compeleceng.2018.01.039>

0045-7906/© 2018 Elsevier Ltd. All rights reserved.

across different disease datasets, as concluded by most researchers in this field. This problem has led to a significant research focus on ensemble classifiers for improved prediction accuracy; such classifiers are grouped into several groups, and these groups are arranged in layers with sequential communication between them [3]. However, such ensemble designs are complicated and computationally heavy. Moreover, if the majority of the classifiers give the wrong prediction, the ensemble framework performs poorly. In this research, the problem of designing intelligent classifiers has been addressed by using a single base classifier and adjusting its parameters to attain maximum prediction accuracy, sensitivity and specificity using multi-objective evolutionary optimization. Such an attempt, when adequately performed, is computationally much more than are its ensemble counterparts. In this paper, SMO, a variant of the classical SVM [4], has been employed as the base classifier; popular and recent evolutionary and meta-heuristic algorithms, such as EHO, MOEA/D, NSGA-II and a novel cuboids based multi-objective optimization(MOO) using EHO (CEHO), have been used to optimize the SMO parameters using this evolutionary approach for medical diagnosis. More specifically, the initial population in CEHO is generated based on a competition. To validate the performance of the proposed classifiers, they were subjected to tests with as many as 17 medical datasets, and suitable statistical tests were performed to substantiate their performances. The experimental results presented in this paper demonstrate that the proposed CEHO outperforms state-of-the-art single and ensemble classifiers among the alternatives and shows better overall performance.

The remainder of this paper is organized as follows: existing research is presented in Section 2, specifically focusing on several machine learning algorithms employed for processing medical datasets. Because this paper discusses a novel mechanism for initial population generation, the review also presents several attempts for initial population generation that helps with performance improvements. The problem formulation of the proposed MOO for the classifying problem is presented in Section 3. Section 4 elaborately explains the materials and methods used in this work. The novel mechanism used to generate the initial population based on the cuboid-based competition is given in Section 4. A very rudimentary introduction to SVM and its improved SMO classifiers is given in Section 4. The performance of the proposed CEHO is tested over 17 benchmark medical datasets, and Section 5 provides a brief account of the experimental setup, the experiments and summarizes the results. Section 5 also presents statistical analyses of the obtained results to validate the credibility of the results. The research conclusions are presented in Section 6.

## 2. Review of related work

Since this paper has two primary goals for designing intelligent diagnosis systems, namely employing machine learning and data mining mechanisms for disease diagnosis and a novel initial population generation mechanism, this section has been organized in two parts where related works are presented. Additionally, a summary of a number of recently published meta-heuristic algorithms is given at the end of this section.

Disease diagnosis with a variety of data mining and machine learning techniques has been studied in the literature over the years. Some of the literature employs specially modified single classifiers for accurately predicting a certain disease or multiple diseases, but others employ specially designed evolutionary-based ensemble classifiers with defined single or multiple objectives.

A significant amount of literature has focused on designing a suitable classifier for diagnosing a specific disease. For instance, in [5], decision trees, MLP and SVM were used to predict seminal quality and demonstrated that MLP provides higher sensitivity, whereas SVM achieves higher specificity values. Another attempt to develop a disease diagnosis system was made in [6] for patients who had undergone thoracic surgery; a novel associative classifier was deliberated using a modified Particle Swarm Optimization (PSO) for rule mining. In [7], the problem of huge dimensionality of medical datasets was addressed. It employed five different classifiers, including ridge regression (RR), online sequential RR, kernel RR, SVM and random forest, with six performance evaluation metrics. The results concluded that kernel RR outperforms other methods. The fruit fly algorithm was used in [8] to evolve the parameters of several variations of SVM.

To improve the prediction accuracy and other quality metrics of prediction, classifiers have often been grouped together with some arrangements to form ensemble classifiers. In [3], an ensemble framework of classifiers comprising 7 classifiers was developed with a multi-layer weighted bagging prediction to classify as many as 14 medical engineering datasets for superior diagnosis performance. Another ensemble classifier was developed in [9] with seven classifiers architected in three phases (data acquisition and cleaning, classifier training, and prediction and evolution).

The remainder of this section discusses the initial population generation mechanisms that have been employed in recently published literature. For mining interesting association rules, a multiple seeds based genetic algorithm (GA) is developed in [10], where the entire solution space is uniformly divided into chunks. From each chunk, a fixed number of chromosomes was generated randomly, and from each chunk, a fixed number of chromosomes was selected based on their fitness value. In [11], an efficient initial population was generated based on the lower and upper limits. This mechanism was tested for 3 benchmark problems, and the results were found to be more reliable.

In the literature, several novel and variations of existing evolutionary algorithms (EA) have been proposed to solve global optimization problems. These EAs include several variations of the Krill Herding algorithm [12], and EHO [13].

However, from such studies, it can be concluded that no individual classifier achieves acceptably high prediction accuracy for a wide variety of disease datasets, which can be addressed by ensemble classifiers. However, such ensembles compromise the simplicity that an individual classifier offers. Additionally, it is more promising to employ efficient initial population generating mechanisms so the solutions can converge to a global optimum in fewer iterations. To this end, this

Download English Version:

<https://daneshyari.com/en/article/6883468>

Download Persian Version:

<https://daneshyari.com/article/6883468>

[Daneshyari.com](https://daneshyari.com)