www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 5 May 2017, Page No. 21292-21296
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i5.16

PSO-Based Support Vector Machine Technique with Cuckoo Search For Lungs Cancer Diagnosis

Dr.K.Mahalakshmi[1], Ms.R.Keerthika[2], Ramyadevi[3]

[1] Professor, Department of Information Technology, Karpagam College of Engineering
2Asst. Prof, Department of Information Technology, Karpagam College of Engineering (Autonomous) Coimbatore -641 032, Tamil Nadu, India

[3]UG Scholar, Department of Information Technology, Karpagam College of Engineering (Autonomous) Coimbatore -641 032, Tamil Nadu, India

prof.dr.mlk@gmail.com[1]; keerthikait@gmail.com[2] ; pramyadevi1@gmail.com[3] ;

Abstract:

Disease diagnosis is conducted with a machine learning method. World Health Organization (WHO) reports that worldwide 7.6 million deaths are caused by cancer each year. Uncontrollable cell development in the tissues of the lung is called as lung cancer. These uncontrollable cells restrict the growth of healthy lung tissues. If not treated, this growth can spread beyond the lung in the nearby tissue called metastasis and, form tumours. In order to preserve the life of the people who are suffered by the lung cancer disease, it should be pre-diagnosed. So there is a need of pre diagnosis system for lung cancer disease which should provide enhanced result. We have proposed a novel machine learning method that hybridizes support vector machine (SVM), particle swarm optimization (PSO).The new method consists of two stages: initially, pso based approach for parameter optimization of SVM is developed to find the better initial parameters of kernel function, and then PSO is applied to continue SVM training and find the best parameters of SVM. Experimental results indicate that the proposed PSO-SVM model achieves better classification accuracy and F-measure than PSO-SVM and GASVM. Therefore, we can conclude that our proposed method is very efficient compared to the formerly reported algorithms.

Keywords: Cuckoo s e a r c h ; Support vector machines (SVMs); Model complexity minimization (MCM); Risk minimization (ERM)

Literature Review:

Data mining in healthcare is a very important field in diagnosis and in deeper understanding of medical data. Health data mining intends to solve real- world problems in diagnosing and treating diseases. One of the most important applications of data mining in the domain of machine learning is diagnosis, and this type of diagnosis of the disease asthma is a notable dispute due to the lack of sufficient knowledge of physicians concerning this disease and because of the complexity of asthma. The purpose of this research is the skilful diagnosis of asthma using efficient algorithms of machine learning. This study was conducted on a dataset consisting of 169 asthmatics and 85 non - asthmatics visiting the Imam Khomeini and Masseeh Daneshvari Hospital of Tehran. The algorithms of k – nearest neighbors, random forest, and support vector machine, together with pre – processing and efficient training.

1. Introduction:

Accurate diagnosis and effective treatment of disease are important issues in life science research and have a positive meaning for human health. Lung cancer is one of the [7] [8] commonest cancers in the industrialized world, and persons with this serious disease must deal not only with the physical effects but also with the psychosocial aspects [8]. Lung cancer [9] is a disease of abnormal cells multiply and growing into a tumour [10]. The overall 5-year survival rate for lung cancer patients increases from 14 to 49% if the disease is detected in time [11]. Among different types of cancer the lung cancer is the most aggressive and best practice to its accurate prognosis is the determination of the current stage of the disease. Three main factors in cancer staging are primary tumour, regional lymph nodes and metastasis [12]. One of the most important and difficult tasks a doctor has to carry out is the detection and diagnosis of cancerous lung nodules from x-ray image's result [13].

Given that lung cancer is one of the common cancers world-wide, the implications of focusing on excellence of life as well as survival require to be understood [14]. Mahalakshmi K et al. (2016) assisted a commitment in productive programming in a roundabout way expended less vitality by spending less equipment hardware to run. They discussed about building a two level green programming demonstration that converts the supportable life cycle of a product item and the product devices advancing greed n and earth practical programming [18,15].

Mahalakshmi et al. (2015) have also used a hybrid technique namely optimization and clustering in non-functional requirements based on the Support Vector Machine and Artificial Bee Colony as well as Differential Evolution. The proposed system also use the Support Vector Machine (SVM) for classify Non-Functional Requirements. Their proposed hybrid ABC- DE with ABC based feature selection shows high classification accuracy of 90.54% [16,17].Recently, medical experts pay more attention to early diagnosis of disease and suggest many new methods to deal with disease diagnosis problem. Using machine learning methods to diagnose disease is rapid development of a novel research branch

of machine learning. Researchers have applied artificial intelligence and computer technology to develop some medical diagnostic systems, which improve the efficiency of diagnosis and become practical tools.

For instance, some original studies in this area are:

- Cuckoo search via Levy flights [1].
- An efficient cuckoo search algorithm for numerical function optimization [2].
- Multimodal function optimisation [3].

Sartakhti proposed a novel machine learning method that hybridized support vector machine (SVM) and simulated annealing (SA) for hepatitis disease diagnosis. The obtained classification accuracy of SVM-SA method was 96.25% and it was very hopeful with regard to the other classification methods in the literature for this problem. The approach proposed by Ramirez was based on image parameter selection and support vector machine (SVM) classification.The proposed system yielded 90.38% accuracy in the early diagnosis of Alzheimer's disease and outperformed existing techniques including the voxel-asfeatures (VAF) approach. Abdi and Giveki developed a diagnosis model based on particle swarm optimization (PSO), support vector machines (SVMs), and association rules (ARs) to diagnose erythematosquamous diseases. The proposed model consists of two stages: first, AR is second-hand to select the optimal feature subset from the original feature set. Then, a PSO-based approach for parameter determination of SVM is developed to find the best in the SVM training system significantly influences the classification accuracy and PSO is a promising tool for global searching). Support vector machine is a machine learning algorithm based on statistical learning theory and has the strong predictive ability for nonlinear problems. However, SVM prediction performance is closely related to the quality of the selected parameters. Parameter optimization algorithms currently used are particle swarm optimization and hereditary algorithms, but these algorithms have their shortcomings and affect the accuracy of disease prediction. Firstly introduces related algorithms, such as support vector machine and cuckoo search, and then presents the novel models, CS-PSO-SVM. It gives results of different model in two real disease diagnoses datasets from University of California Irvine Machine Learning Repository. A mobile device and desktop computer varies from each other in many aspects. A few of the characteristics were the output that is, the size of thescreen, resolution features, and so on; input, that is, keypads, touchscreen and voice inputs; processing power, memory, application as well as their media types. The electronic and mobile learning platforms varies in compatibility as when the services of e-learning platforms are transferred into applications in m-learning platforms it could be observed that a few of the services should undergo certain changes to meet the restrictions of small mobile devices while a few of the services are not viable to be conveyed reliably [19].

2. EXISTING METHOD:

Cuckoo search (CS) is an existed swarm intelligent optimization algorithm. Preliminary study show that cuckoo search algorithm is simple and easy to implement and has less parameters. Cuckoo search algorithm is able to afford a new method for the SVM parameter optimization. This paper proposes a disease diagnosis model based on cuckoo search, particle swarm optimization (PSO), and support vector machine. Cuckoo search is an optimization algorithm developed by Xin-she Yang and Suash Deb [4, 5, 6]. It was inspired by the obligate brood parasitism of some cuckoo species by laying their eggs in the nests of other host birds (of other species). Some host birds can engage direct conflict with the intruding cuckoos. For example, if a host bird discover that the eggs are not their own, it will either throw these alien eggs away or simply abandon its nest and build a new nest somewhere else. Some cuckoo species such as the New World brood-parasitic *Tapera* have evolved in such a way that female parasitic cuckoos are often very specific in the mimicry in colours and pattern of the eggs of a few chosen host species.

2.1 Disadvantage:

- Simple algorithm
- Time consumption is high

3. PROPOSED METHOD:

Support vector machines(SVMs) are a set of related supervised learning methods used for classification and regression. A support vector machine constructs a hyper plane or set of hype planes in a high-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training data points of any class, since in general, the larger the margin, the lower the generalization error of the classifier. In order to extend the SVM methodology to handle data that is not completely linearly separable, we relax the constraints slightly to allow for misclassified points; the formulation is subsequent.

3.1. Parameter calculation:

```
load('ori_label.txt')
labels=ori_label;
cluster_value=clustered_values;
truelabels=cluster_value;

nrow = length(truelabels);
[R, truelabels] = ind2cluster(truelabels);
[truelabels, S] = sort(truelabels);
labels = labels(S);
R = ind2cluster(labels);
k = length(R);

S = ones(nrow,1);
high=[];
```

3.2 Accuracy:

```
acc=100-Rerror; % difference between the total percentange and
total error

acc1=100-Rerror1;
disp('Parameters for the PSO & Kmeans clustering');
Accuracy_pso=acc;
Accuracy_kmeans=acc1;
```

4. PSO-SVM:

DOI: 10.18535/ijecs/v6i5.16

Particle swarm optimization is an evolutionary computation technique proposed by Kennedy and Eberhart. It is a population-based stochastic search process, modeled after the social behavior of a bird flock. It is similar in spirit to birds migrating in a flock toward some destination, where the intelligence and competence lie in the cooperation of an entire flock. PSO algorithms make use of particles moving in an n-dimensional space to search for solutions for n-variable function optimization problem. All particles have fitness values which are evaluated by the fitness function to be optimized and have velocity which direct the flying of the particles. The particles fly through the problem space by following the particles with the best solutions so far. PSO is initialized with a group of random particles (solutions) and then searches for optima by updating each generation. SVM also has a drawback that limits the use of SVM on academic and industrial platform: there are free parameters (SVM hyper parameters and SVM kernel parameters) that need to be defined by the user. Since the quality of SVM regression models depends on a proper setting of these parameters, the foremost issue for practitioners trying to apply SVM is how to set these parameter values (to ensure good generalization performance) for a given training dataset. SVM based on PSO optimizes two important hyperparameters C and ε using PSO.

5. Evaluating the output:

Sensitivity is the relation of correct positive diagnoses that are correctly revealed by a specific test. Based on this definition, one hundred percent sensitivity implies correct diagnoses of the disease in all patients

$$\text{SENSITIVITY} = \frac{\text{the number of correct positive diagnoses}}{\text{The number of false negative diagnoses+the number of correct positive diagnoses}}$$

$$\text{ACCURACY} = \frac{\text{the number of correct positive diagnoses}}{\text{the number of correct positive diagnoses+The number of false negative diagnoses+the number of correct positive diagnoses+ the number of correct positive diagnoses}}$$

Table. Performance Of Our Proposed Lung Cancer Pre-Diagnosis Technique And Other Optimization Technique Such As FFBNN-GA, FFBNN-PSO And FFBNN

Measures	Proposed FFBNN-ABC	FFBNN-GA	FFBNN-PSO	FFBNN
Accuracy	90	60	80	50
Sensitivity	88	77	82	0
Specificity	100	29	67	22
FPR	0.0	71.4	33.3	77.8
PPV	100	67	93	53
NPV	60	40	40	40
FDR	0	33	7	47
MCC	66.4	6.8	39.1	0.0

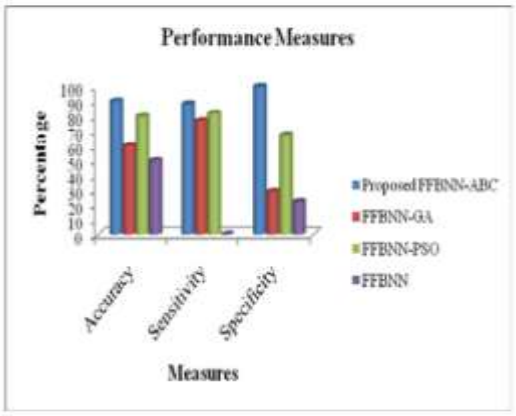
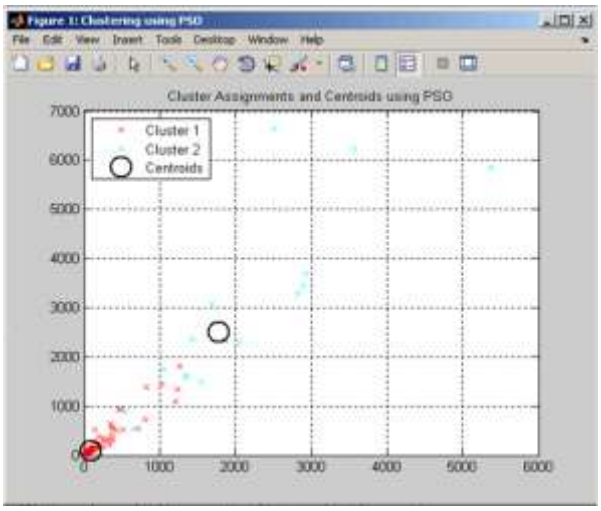
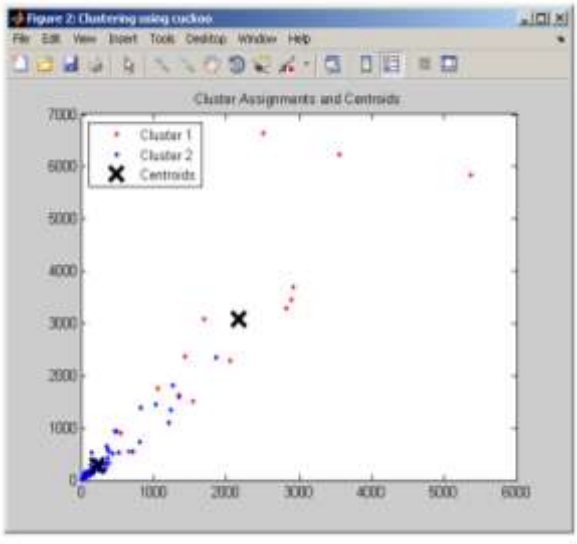


Fig. Proposed, FFBNN- GA, FFBNN-PSO And FFBNN Techniques Performance Outcomes In Terms Of Accuracy, Sensitivity And Specificity

5.1.CLUSTERING USING PSO

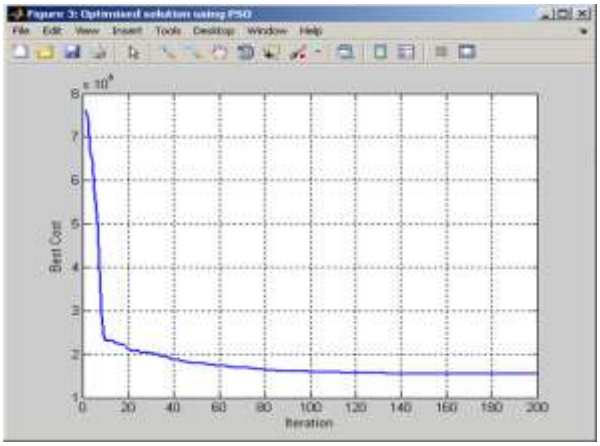


5.2.CLUSTERING USING CUCKOO

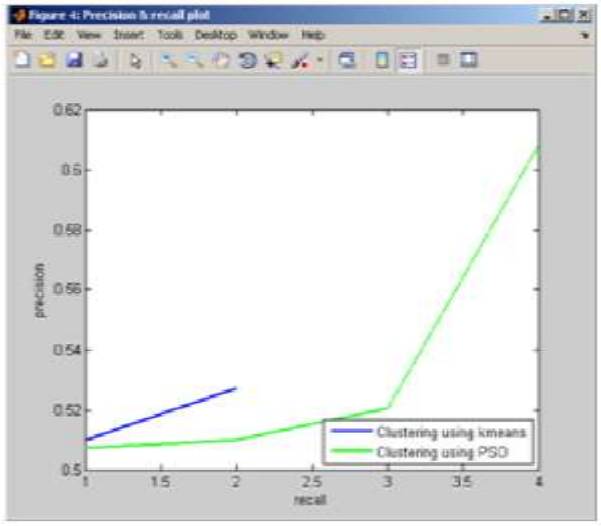


DOI: 10.18535/ijecs/v6i5.16

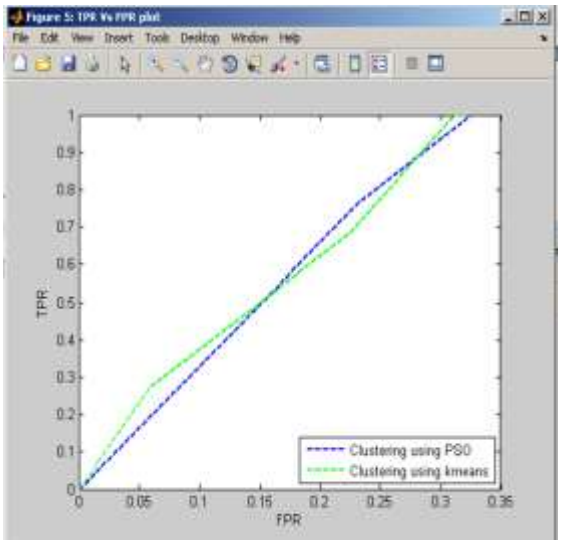
5.3. OPTIMIZED SOLUTION USING PSO



5.3.PRECISION AND RECALL PLOT



5.5. TPR VS FPR



5.6. ERROR PREDICTION:

5.7. Average prediction:

6. CONCLUSION:

In the last few decades, several disease diagnosis models have been developed for the disease prediction. The objective of disease diagnosis models is to make a definite diagnosis from patients' laboratory sheet early as soon as possible and initiate timely treatment. Accurate diagnosis has an important sense for human health. In this paper, we design a new disease diagnosis model, PSO-SVM. The experimental research results show that the novel algorithm is better than GA-SVM and PSO-SVM.

7. REFERENCES:

[1] Yang, X.-S., Deb, S.: Cuckoo search via lévy flights. In: World Congress on Nature & Bio- logically Inspired Computing, 2009. NaBIC 2009, pp. 210–214. IEEE (2009).
[2] Ong, P., Zainuddin, Z.: An efficient cuckoo search algorithm for numerical function optimiza- tion. In: AIP Conference Proceedings, vol. 1522, p. 1378 (2013).
[3] Jamil, M., Zepernick, H.: Multimodal function optimisation with cuckoo search algorithm. Int. J. Bio-Inspired Comput. **5**(2), 73–83 (2013).
[4] Yang,X.-S.:Metaheuristic algorithms for self-organizing systems: a tutorial. In: IEEE Sixth Interational Conference on Self-Adaptive and Self-Organizing Systems (SASO 2012), pp. 249–250. IEEE Conference Publications (2012).

[5] Yang, X.-S.: Bat algorithm and cuckoo search: a tutorial. In: Artificial Intelligence, Evolu- tionary Computing and Metaheuristics, pp. 421–434. Springer (2013).
[6] Vigneshwaran, A. An optimal job scheduling in grid using cuckoo algorithm. Int. J. Comput. Sci. Telecomm, 3(2), 65-69 (2012).
[7] Liu Lu and Liu Wanyu, "A Method of Pulmonary Nodules Detection with Support Vector Machines", In IEEE Proceedings of the Eighth International Conference on Intelligent Systems Design and Applications, Kaohsiung, pp. 32-35, 2008.
[8] Kathrine Carlsen, Anders Bonde Jensen, Erik Jacobsen, Mark Krasnik and Christoffer Johansen, "Psychosocial aspects of lung cancer Review Article", Lung Cancer, Vol. 47, No. 3, pp. 293-300, 2005.
[9] Ying-Wooi Wan and Guo, "Constructing gene- expression based survival prediction model for Non-Small Cell Lung Cancer (NSCLC) in all stages and early stages", In IEEE Proceedings of International Conference on Bioinformatics and Biomedicine Workshop, Washington, pp.338, 2009.
[10] Mokhled S. Al-Tarawneh, "Lung Cancer Detection Using Image Processing Techniques", Leonardo Electronic Journal of Practices and Technologies, Vol. 11, No. 20, pp. 147-158,2012.
[11] Chaudhary and Singh , "Lung Cancer Detection on CT Images by Using Image Processing", In IEEE Proceedings of International Conference on Computing Sciences, Phagwara, pp.142-146,2012.
[12] Almasslawi and Kabir , "Using non-rigid image registration and Thin-Plate Spline warping for lung cancer progression assessment", In IEEE proceedings of International Conference on Computer Science and Automation Engineering, Shanghai, pp. 506-510, 2011.
[13] Abdullah and Mohamaddiah, "Development of cellular neural network algorithm for detecting lung cancer symptoms", In IEEE Proceedings of International Conference on Biomedical Engineering and Sciences, Kuala Lumpur, pp.138-143, 2010.
[14] Ali Montazeri, Robert Milroy, David Hole, James McEwen and Charles R. Gillis, "Quality of life in lung cancer patients: As an important prognostic factor Original Research Article", Lung Cancer, Vol. 31, No. 2–3, pp. 233-240,2001.
[15] Mahalakshmi, K., D. R. Prabhakar, and D. V. Balakrishnan. "Kernel Optimization For Improved Nonfunctional Requirements Classification." Journal of Theoretical and Applied Information Technology 60.1: 64-72.
[16] Mahalakshmi, K., and R. Prabhakar. "Hybrid Optimization of SVM for Improved Non-Functional Requirements Classification." International Journal of Applied Engineering Research 10.20 (2015): 2015.
[17] Mahalakshmi, K., R. Prabhakar, and V.Balakrishnan. "Optimizing Support Vector Machine for Classifying Non Functional Requirements." Research Journal of Applied Sciences, Engineering and Technology 7.17 (2014): 3643-3648.
[18] K.Mahalakshmi, S.Manikandan, S.Nithyanantham, "Unsupervised Learning Technique Using Hybrid Optimization for Non-Functional Requirements Classification" Asian Journal of Information Technology, 15.9, (2016), 1457 – 1467.
[19] An Optimized Support Vector Machine for Classifying Opinions in M Learning Systems Applied To Biotechnology Domain", K.Mahalakshmi, MathiVananP, MohanaPriya D, R Keerthika, A.Nagajothi Research Journal of Biotechnology, World Research Journals, 2017/1special issue 168-176