International Journal of Advance Research in Science and Engineering
Vol. No.4, Special Issue (01), September 2015
www.ijarse.com
IJARSE
ISSN 2319 - 8354

MACHINE LEARNING TECHNIQUES FOR MEDICAL DIAGNOSIS: A REVIEW

Anju Jain

Asst. Professor, Deptt of CSE, G.J.U. S&T, Hisar, Haryana (India)

ABSTRACT

Machine learning algorithm can significantly help in solving the healthcare problems by developing classifier systems that can assist physicians in diagnosing and predicting diseases in early stages. However, extracting knowledge from medical data is challenging as this data may be heterogeneous, unorganized, and high dimensional and may contain noise and outliers. Most appropriate method can be chosen only after analyzing all the available machine learning techniques and validating their performances in terms of accuracy and comprehensibility. This literature has reviewed the use of machine learning algorithms like decision tree, support vector machine, random forest, evolutionary algorithms and swarm intelligence for accurate medical diagnosis. The dependence on medical images for diagnosing a disease is on rise. Since interpreting modern medical images is becoming increasingly complex, machine learning algorithms in medical imaging can provide significant assistance in medical diagnosis. Machine learning techniques could be used for large scale and complex biological data analysis as these techniques are efficient and inexpensive in solving bioinformatics problems.

***Keywords:* Decision Tree, Evolutionary Algorithms, Machine Learning, Medical Diagnosis, Protein Function Prediction, Random Forest, Medical Imaging, Swarm Intelligence, Support Vector Machine.**

I. INTRODUCTION

Machine learning, a sub discipline in the field of Artificial Intelligence, explores the study and design of algorithms that can learn from data [1]. Machine Learning provides methods/algorithms that make system computationally intelligent. Such algorithms build models based on input and then use these models to make predictions or decisions.

Machine Learning is mainly useful in cases where algorithmic/deterministic solutions are not available i.e. there is a lack of formal models or the knowledge about the application domain is scarce. The algorithms have been developed in diverse set of disciplines such as statistics, computer science, robotics, computer vision, physics, and applied mathematics. Advantages of machine learning over statistical models are accuracy, automation, speed, customizability and scalability.

As medicine plays a great role in human life, automated knowledge extraction from medical data sets has become an immense issue. Research on knowledge extraction from medical data is growing fast [2]. All activities in medicine can be divided into six tasks: screening, diagnosis, treatment, prognosis, monitoring and management. As the healthcare industry is becoming more and more reliant on computer technology, machine learning methods are required to assist the physicians in identifying and curing abnormalities at early stages.

Medical diagnosis is one of the important activities of medicine. The accuracy of the diagnosis contributes in deciding the right treatment and subsequently in cure of diseases. Machine Learning is extensively used in diagnosing several diseases such as cancer [3], [4], [5], diabetes [6], heart [7] and skin diseases [8]. Application of Machine learning algorithms improves the diagnostic speed, accuracy and reliability. Among various algorithms in data modelling, decision tree is known as the most popular due to its simplicity and interpretability [8], [9]. Recently, more efficient algorithms such as SVM and artificial neural networks have also become popular [2], [4], [10].

Further, medical imaging has also been one of the most successful techniques to diagnose diseases related to the internal human organs [11], [12], [5], [4], [13], [14]. Although the process of identifying any abnormalities in the captured images is completely dependent on the diagnosis given by the radiologist/physicians, yet the growth of the medical knowledge has made it difficult for radiologists or physicians to keep record of all the possible diagnosis of various diseases. Use of machine learning in medical imaging can assist less as well as highly experienced radiologists in diagnosing the complex cases.

It has been observed from literature review that research is also being done in application of machine learning algorithm in areas such as protein function prediction and gene expression [15], [16]. Unlike sequence and structure based methods for protein function prediction, machine learning methods do not require explicit knowledge of homology and homology-derived parameters for the purpose of function prediction. Therefore research on developing appropriate machine learning techniques for prediction of protein function for disease diagnosis is on rise.

This paper looks into the machine learning techniques that have been utilized in building computer aided diagnosis. Section II briefly presents about the various classification algorithms used in medical domain. Section III reviews the literature covered in five major areas: Decision trees, Support vector machine, Random forest, Evolutionary algorithm and Swarm intelligence. The last section concludes and underlines the future work in this domain.

II. BACKGROUND DETAILS

Classification algorithms are widely used in various medical applications. Data classification is a two phase process in which first step is the training phase where the classifier algorithm builds classifier with the training set of tuples and the second phase is classification phase where the model is used for classification and its performance is analyzed with the testing set of tuples. Brief about the various classification algorithms in medical domain are:

2.1 Decision Tree Algorithm

The decision tree is one of the classification algorithms. The learning algorithm applies a divide and-conquer strategy to construct the tree [17]. The sets of instances are associated by a set of attributes. A decision tree comprises of nodes and leaves, where nodes represent a test on the values of an attribute and leaves represent the class of an instance that satisfies the conditions. The outcome is ‘true’ or ‘false’. Rules can be derived from the path starting from the root node to the leaf and utilizing the nodes along the way as preconditions for the rule, to predict the class at the leaf. The tree pruning has to be carried out to remove unnecessary preconditions and duplications.

2.2 Support Vector Machine

SVM algorithms are based on the learning system which uses the statistical learning methodology and they are popularly used for classification. In SVM technique, the optimal boundary, known as hyperplane, of two sets in a vector space is obtained independently on the probabilistic distribution of training vectors in the set. This hyperplane locates the boundary that is most distant from the vectors nearest to the boundary in both sets. The vectors that are placed near the hyperplane are called supporting vectors. If the space is not linearly separable there may be no separating hyperplane. The kernel function is used to solve the problem. The kernel function analyses the relationship among the data and it creates a complex divisions in the space.

2.3 Random Forests

Random forest algorithm is one of the best among classification algorithms and is able to classify large amounts of data with high accuracy. It is an ensemble learning method building models that constructs a number of decision trees at training time and outputs the modal class out of the classes predicted by individual trees. It is a combination of tree predictors where each tree depends on the values of a random vector sampled independently with the same distribution for all the trees in the forest. The basic principle is that a group of “weak learners” can come together to form a “strong learner”.

2.4 Evolutionary Algorithms

A Genetic Algorithm (GA) is an evolutionary and stochastic method for finding optimal solutions in large and complex search spaces. A GA is inspired by natural evolution: a population of encoded candidate solutions (called “chromosomes”) is evolved through generations using genetic-like operations such as crossover and mutation. At each generation, solutions are probabilistically selected based on their fitness, in order to generate offspring and create the next generation. The initial population is randomly generated, and at each generation, every candidate solution is evaluated against an objective function in order to gain a fitness score. In a learning system, the objective function is typically the measure of the accuracy of a candidate over a training set of instances.

2.5 Swarm Intelligence

Swarm intelligence (SI) is a computational intelligence technique to solve complex real-world problems. It involves the study of collective behaviour of individuals in a population who interact locally with one another and with their environment in a decentralized control system. The inspiration often comes from nature, especially biological systems. The agents follow very simple rules, and although there is no centralized control structure dictating how individual agents should behave, local and to a certain degree random interactions between the agents lead to an "intelligent" global behaviour which is unknown to the individual agents. Some of the popular SI algorithms are Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO) and Artificial Bee Colony (ABC).

III. REVIEW OF LITERATURE

Classification of medical data is a complex optimization problem. The aim is not just to find optimal solution but to provide an accurate diagnosis. Many researchers are applying different kinds of machine learning algorithms for solving this problem. This section reviews the literature covered in five major areas: Decision trees, Support vector machine, Random forest, Evolutionary algorithm and Swarm intelligence.

3.1 Decision Trees

Research on using decision trees for medical classification or disease diagnosis purpose has caught a lot of attention in recent past [9], [18]– [21]. Among various algorithms in data classification, decision tree is known as one of the most popular due to its simplicity and interpretability [2]. It has been used in the diagnosis of diseases such as breast cancer, liver cancer, brain tumour and dermatologic diseases.

Decision tree has been used as classifier for breast cancer diagnosis [18], [19], [20]-[22]. It has been compared with different classification algorithms in each paper such as ANN, logistic regression, Bayesian network, KNN and case based reasoning. Case based fuzzy decision tree outperforms other approaches with an average accuracy rate of 99.5% in breast cancer diagnosis [21]. Azar and El-Metwally [9] proposed a decision support tool for the detection of breast cancer based on three types of classifiers. They are single decision tree SDT, boosted decision tree (BDT) and decision tree forest and reported that BDT performed better than SDT with 98.83% and 97.07% accuracy respectively.

Yeh et al. [23] presented decision tree model as the optimum model for cerebrovascular disease with accuracy of 99.59% in comparison to Bayesian classifier and back propagation neural network.

Luk et al. [24] proposed classification and regression tree (CART) model that provides discrimination between Hepatocellular Carcinoma (HCC) and non-malignant liver tissue. HCC is known as the most dangerous cancer due to not being diagnosed until advanced tumour stages. Decision tree algorithms were successfully applied for building classification model based on hidden pattern in the problem dataset. Further, Fan et al. [21] proposed hybrid model case-based reasoning and fuzzy decision tree (CBFDT) for liver disease whose accuracy is highest among various other models with accuracy of 81.6%.

Major limitations of decision tree in medical data are imbalancement and cost sensitivity problem. Further they are sensitive to inconsistent data.

3.2 Support Vector Machines

SVM algorithms have been proposed as an effective statistical learning method for classification because of their high generalization performance. Intuitively, given a set of points which belong to either one of the two classes, a SVM can find a hyperplane having the largest possible fraction of points of the same class on the same plane. This hyperplane, called the optimal separating hyperplane (OSH), can minimize the risk of misclassifying examples of a test set.

There has been a lot of research on medical diagnosis of breast cancer using SVM with Wisconsin breast cancer diagnosis (WBCD) data in literature and most of them reported high classification accuracies [25]–[30]. In Polat and Gunes [26], least square SVM was used and an accuracy of 98.53% was obtained. Further, SVM model with grid search and feature selection was proposed [27], [28] for breast cancer diagnosis. When using SVM, two problems are confronted: how to choose the optimal input feature subset for SVM and how to set the best kernel parameters. Feature selection limits the number of input features in a classifier in order to have a good predictive and less computationally intensive model. In addition to the feature selection, proper model parameters setting can improve the SVM classification accuracy. The parameters that should be optimized include penalty parameter C and the kernel function parameters such as the gamma (c) for the radial basis function (RBF) kernel. F-score is adapted to find the important features, and the grid search approach is used to search the optimal SVM parameters. Proposed model showed 99.51% accuracy for 80-20% training-test partition. Hassanien and Kim [31] diagnosed breast cancer using hybrid approach (ANN + SVM + Fuzzy). The proposed approach utilized type-II fuzzy algorithm for improving the quality of MRI image. Then, segmentation was done using pulse coupled neural networks in order to extract regions of interest. Wavelet features were extracted from these regions and finally, SVM was used for the actual diagnosis and discrimination of different regions of interest to determine whether they represent cancer or not. Results showed that the accuracy offered by proposed hybrid model using SVM was high compared to other machine learning algorithms such as decision tree, neural network etc.

SVM has been largely and successfully used in hybrid approach for medical diagnosis of various diseases such as Genetic + Fuzzy + SVM for the diagnosis of diabetes, liver and heart diseases [10], ANN + SVM for prostate cancer diagnosis [32], ANFIS + SVM for pain identification [33] and nonnegative matrix factorization + SVM for Alzheimer's disease [14]. However, practical obstacle of the SVM-based classification model is its black-box nature. A possible solution for this issue is the use of SVM rule extraction techniques or the use of hybrid-SVM model combined with other more interpretable models.

3.3 Random Forest

Research has been done using Random forest as a classifier and feature selection algorithm for medical diagnosis. Ozcift [34] used best first search random forest algorithm to select optimal features for four medical datasets: colon cancer, leukemia cancer, breast cancer and lung cancer. The proposed model with extracted features accuracy was compared with 15 widely used classifiers trained with all features and showed improved classification accuracy. Nguyen et al. [35] also used random forest classifier combined with feature selection for breast cancer diagnosis and reported 99.82% classification accuracy. Taking into account sensitivity, specificity and overall classification accuracy, random forest is ranked first among all the classifiers tested in prediction of dementia using several neuropsychological tests [36]. Random forest has been successfully used as a classifier in the diagnosis of diseases such as Abdominal lymphadenopathy [37], Alzheimer’s disease and cardiovascular risk.

3.4 Evolutionary Algorithms

Genetic algorithms (GA) are used in developing the decision support system for the diagnosis of several diseases such as breast cancer [38], prostate cancer [39], brain tumour [40], colon cancer[41] and heart diseases [42]. Evolutionary algorithms (EAs) are generally combined with other classification algorithms to form hybrid systems to develop computer aided diagnosis for certain organs [41]–[44]. Chaochang et al. [43] built a diagnostic model for hypertension by a hybrid of GA, apriori and decision tree with high accuracy. [41], [42] diagnosed heart diseases, colon, lymphoma and leukaemia cancer by hybrid fuzzy + GA. Asymptomatic carotid stenosis is considered as an important factor of stroke. It has several risk factors such as smoking, hypertension, diabetes, cardiac diseases and physical inactivity. Bilge et al. [44] discovered rules for these risk factors and asymptomatic carotid stenosis by hybrid approach of GA and regression. EAs are usually applied in medical data mining as a parameter finder. Evolutionary techniques search for the parameter values of the knowledge representation set up by the designer so that the mined data are optimally interpreted (presence or absence of disease). A GA efficiently searches the significant boundary features of the brain tumour region and feed them to ANFIS [40]. A GA searches for optimal structure and training parameters of neural network for a better prediction of lung sounds and hence reducing the processing load and time [45]. A system has been developed to analyze digital mammograms using novel neuro-genetic algorithm [46]. The system starts first by extracting features, then the GA selects the most significant features and uses them as input to a neural network. This system has achieved a very satisfactory performance.Genetic programming along with other classification algorithms are used to develop model for disease diagnosis such as ANN + GP to diagnose thalassemia [47], decision tree + GP for chest pain diagnosis [48] and GP with image processing techniques to detect lung abnormalities at an early stage [49].

3.5 Swarm Intelligence

Swarm intelligence (SI) algorithms such as PSO, ACO are generally used as ideal pre-processing tools to help optimize feature selection in classifier systems for medical diagnosis. This increases the classification accuracy and keeps the computational resources needed to a minimum [50]. Improved binary particle swarm optimization (IBPSO) is used to implement a feature (gene) selection, and a K-nearest neighbor (K-NN) serves as an evaluator of IBPSO for gene expression data classification problems [51]. The classification accuracy obtained by the proposed method was the highest in nine out of the 11 gene expression data test problems. An ACO has been used to select appropriate wavelet coefficients from mass spectral data as feature selection method for ovarian cancer diagnosis and has showed high classification accuracy [52]. Hybrid classification models are developed using PSO/ACO for the diagnosis of several diseases. Case based reasoning + PSO (CBRPSO) and SVM +PSO hybrid models for diagnosis of breast cancer have also been proposed [53], [54]. The CBRPSO has been found to outperform the other approaches with an average accuracy rate of around 97.4% for breast cancer. The CBRPSO model is also used for liver disorders with an average accuracy of 76.8%. Hybrid model using PSO with other classification algorithms for diagnosis of coronary artery disease [55], leukaemia [56] and MR brain image classifier [13] have been proposed with high classification accuracy. Thus SI algorithms have been used as optimization techniques in many areas such as function optimization, ANN training, fuzzy system control and medical diagnosis.

IV. CONCLUSION

Successful implementation of machine learning algorithms in medical diagnosis can help the integration of computer based systems in the healthcare environment. Especially in developing and highly populated country like India where mortality ratio is high and there is only one doctor for every 1700 persons, machine learning techniques in medical diagnosis can assist physicians to diagnose and cure diseases at early stage. Technology can no way replace a doctor's experience and expertise, but it can take care of relatively straightforward yet time consuming diagnostic tasks and doctors can take up clinically more demanding procedure.

The dependence on medical images for diagnosing a disease is on rise. Since interpreting modern medical images is becoming increasingly complex, machine learning algorithms in medical imaging can provide significant assistance in medical diagnosis. They can help interns or less experienced physicians to reliably evaluate medical images and thus improve their diagnostic accuracy, sensitivity and specificity.

Protein function prediction is another important area where machine learning techniques have a vital role to play. Machine learning techniques could be used for large scale and complex biological data analysis as these techniques are efficient and inexpensive in solving bioinformatics problems. The research in this area will not only be beneficial for physicians in terms of diagnosing diseases, it may also help health planners for diagnosing and preventing diseases at a large scale.

REFERENCES

[1] H. Brink and J. W. Richards, Real-World Machine Learning, 7th ed. Manning Pulications, 2013.

[2] N. Esfandiari, M. R. Babavalian, A.-M. E. Moghadam, and V. K. Tabar, "Knowledge discovery in medicine: Current issue and future trend," Expert Syst. Appl., vol. 41, no. 9, pp. 4434–4463, Jul. 2014.

[3] M. Çınar, M. Engin, E. Z. Engin, and Y. Ziya Ateşçi, "Early prostate cancer diagnosis by using artificial neural networks and support vector machines," Expert Syst. Appl., vol. 36, no. 3, Part 2, pp. 6357–6361, Apr. 2009.

[4] A. E. Hassanien and T. Kim, "Breast cancer MRI diagnosis approach using support vector machine and pulse coupled neural networks," J. Appl. Log., vol. 10, no. 4, pp. 277–284, Dec. 2012.

[5] J. Nahar, T. Imam, K. S. Tickle, A. B. M. Shawkat Ali, and Y.-P. P. Chen, "Computational intelligence for microarray data and biomedical image analysis for the early diagnosis of breast cancer," Expert Syst. Appl., vol. 39, no. 16, pp. 12371–12377, Nov. 2012.

[6] H. Temurtas, N. Yumusak, and F. Temurtas, "A comparative study on diabetes disease diagnosis using neural networks," Expert Syst. Appl., vol. 36, no. 4, pp. 8610–8615, May 2009.

[7] S. Shilaskar and A. Ghatol, "Feature selection for medical diagnosis : Evaluation for cardiovascular diseases," Expert Syst. Appl., vol. 40, no. 10, pp. 4146–4153, Aug. 2013.

[8] C.-L. Chang and C.-H. Chen, "Applying decision tree and neural network to increase quality of dermatologic diagnosis," Expert Syst. Appl., vol. 36, no. 2, Part 2, pp. 4035–4041, Mar. 2009.

[9] A. T. Azar and S. M. El-Metwally, "Decision tree classifiers for automated medical diagnosis," Neural Comput. Appl., vol. 23, no. 7–8, pp. 2387–2403, Dec. 2013.

[10] Y. C. T. Bo Jin, "Support vector machines with genetic fuzzy feature transformation for biomedical data classification.," Inf Sci, vol. 177, no. 2, pp. 476–489, 2007.

[11] K. Suzuki, Ed., Machine Learning in Computer-Aided Diagnosis: Medical Imaging Intelligence and Analysis. IGI Global, 2012.

[12] S. N. Deepa and B. A. Devi, "A Survey on Artificial Intelligence Approaches for Medical Image Classification," Indian J. Sci. Technol., vol. 4, no. 11, pp. 1583–1595, Nov. 2011.

[13] Y. Zhang, S. Wang, G. Ji, and Z. Dong, “An MR brain images classifier system via particle swarm optimization and kernel support vector machine,” Sci. World J., vol. 2013, 2013.

[14] P. Padilla, M. López, J. M. Górriz, J. Ramírez, D. Salas-González, I. Álvarez, and Alzheimer’s Disease Neuroimaging Initiative, “NMF-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer’s disease,” IEEE Trans. Med. Imaging, vol. 31, no. 2, pp. 207–216, Feb. 2012.

[15] A. A. Freitas, D. C. Wieser, and R. Apweiler, “On the importance of comprehensible classification models for protein function prediction,” IEEEACM Trans. Comput. Biol. Bioinforma. IEEE ACM, vol. 7, no. 1, pp. 172–182, Mar. 2010.

[16] R. C. Barros, M. P. Basgalupp, A. A. Freitas, and A. C. P. L. F. de Carvalho, “Evolutionary Design of Decision-Tree Algorithms Tailored to Microarray Gene Expression Data Sets,” IEEE Trans. Evol. Comput., vol. 18, no. 6, pp. 873–892, Dec. 2014.

[17] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd Revised ed. Burlington, MA: Morgan Kaufmann Publishers In, 2011.

[18] D. Delen, G. Walker, and A. Kadam, “Predicting Breast Cancer Survivability: A Comparison of Three Data Mining Methods,” Artif Intell Med, vol. 34, no. 2, pp. 113–127, Jun. 2005.

[19] D. M. F. bin Othman and T. M. S. Yau, “Comparison of Different Classification Techniques Using WEKA for Breast Cancer,” in 3rd Kuala Lumpur International Conference on Biomedical Engineering 2006, F. Ibrahim, N. A. A. Osman, J. Usman, and N. A. Kadri, Eds. Springer Berlin Heidelberg, 2007, pp. 520–523.

[20] N. Cruz-Ramírez, H.-G. Acosta-Mesa, H. Carrillo-Calvet, and R.-E. Barrientos-Martínez, “Discovering interobserver variability in the cytodiagnosis of breast cancer using decision trees and Bayesian networks,” Appl. Soft Comput., vol. 9, no. 4, pp. 1331–1342, Sep. 2009.

[21] C.-Y. Fan, P.-C. Chang, J.-J. Lin, and J. C. Hsieh, “A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification,” Appl. Soft Comput., vol. 11, no. 1, pp. 632–644, Jan. 2011.

[22] N. Cruz-Ramírez, H. G. Acosta-Mesa, H. Carrillo-Calvet, L. Alonso Nava-Fernández, and R. E. Barrientos-Martínez, “Diagnosis of breast cancer using Bayesian networks: A case study,” Comput. Biol. Med., vol. 37, no. 11, pp. 1553–1564, Nov. 2007.

[23] D.-Y. Yeh, C.-H. Cheng, and Y.-W. Chen, “A predictive model for cerebrovascular disease using data mining,” Expert Syst. Appl., vol. 38, no. 7, pp. 8970–8977, Jul. 2011.

[24] J. M. Luk, B. Y. Lam, N. P. Y. Lee, D. W. Ho, P. C. Sham, L. Chen, J. Peng, X. Leng, P. J. Day, and S.-T. Fan, “Artificial neural networks and decision tree model analysis of liver cancer proteomes,” Biochem. Biophys. Res. Commun., vol. 361, no. 1, pp. 68–73, Sep. 2007.

[25] H. X. Liu, R. S. Zhang, F. Luan, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan, “Diagnosing breast cancer based on support vector machines,” J. Chem. Inf. Comput. Sci., vol. 43, no. 3, pp. 900–907, Jun. 2003.

[26] K. Polat and S. Güneş, “Breast cancer diagnosis using least square support vector machine,” Digit. Signal Process., vol. 17, no. 4, pp. 694–701, Jul. 2007.

[27] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," Expert Syst. Appl., vol. 36, no. 2, Part 2, pp. 3240–3247, Mar. 2009.

[28] C.-L. Huang, H.-C. Liao, and M.-C. Chen, "Prediction model building and feature selection with support vector machines in breast cancer diagnosis," Expert Syst. Appl., vol. 34, no. 1, pp. 578–587, Jan. 2008.

[29] C.-S. Lo and C.-M. Wang, "Support vector machine for breast MR image classification," Comput. Math. Appl., vol. 64, no. 5, pp. 1153–1162, Sep. 2012.

[30] L. Li, H. Tang, Z. Wu, J. Gong, M. Gruidl, J. Zou, M. Tockman, and R. A. Clark, "Data mining techniques for cancer detection using serum proteomic profiling," Artif. Intell. Med., vol. 32, no. 2, pp. 71–83, Oct. 2004.

[31] A. E. Hassanien and T. Kim, "Breast cancer MRI diagnosis approach using support vector machine and pulse coupled neural networks," J. Appl. Log., vol. 10, no. 4, pp. 277–284, Dec. 2012.

[32] M. Çınar, M. Engin, E. Z. Engin, and Y. Ziya Ateşçi, "Early prostate cancer diagnosis by using artificial neural networks and support vector machines," Expert Syst. Appl., vol. 36, no. 3, Part 2, pp. 6357–6361, Apr. 2009.

[33] M. Vatankhah, V. Asadpour, and R. Fazel-Rezai, "Perceptual pain classification using ANFIS adapted RBF kernel support vector machine for therapeutic usage," Appl. Soft Comput., vol. 13, no. 5, pp. 2537–2546, May 2013.

[34] A. Ozcift, "Enhanced Cancer Recognition System Based on Random Forests Feature Elimination Algorithm," J. Med. Syst., vol. 36, no. 4, pp. 2577–2585, May 2011.

[35] C. Nguyen, Y. Wang, and H. N. Nguyen, "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic," J. Biomed. Sci. Eng., vol. 06, no. 05, pp. 551–560, 2013.

[36] J. Maroco, D. Silva, A. Rodrigues, M. Guerreiro, I. Santana, and A. de Mendonça, "Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests," BMC Res. Notes, vol. 4, no. 1, pp. 299, Aug. 2011.

[37] K. M. Cherry, S. Wang, E. B. Turkbey, and R. M. Summers, "Abdominal lymphadenopathy detection using random forest," Proc. SPIE 9035, vol. 9035, pp. 90351G-90351G-6 ,March 2014

[38] T.-C. Chen and T.-C. Hsu, "A GAs based approach for mining breast cancer pattern," Expert Syst. Appl., vol. 30, no. 4, pp. 674–681, May 2006.

[39] W. Froelich, E. I. Papageorgiou, M. Samarinas, and K. Skriapas, "Application of evolutionary fuzzy cognitive maps to the long-term prediction of prostate cancer," Appl. Soft Comput., vol. 12, no. 12, pp. 3810–3817, Dec. 2012.

[40] A. Das and M. Bhattacharya, "A Study on Prognosis of Brain Tumors Using Fuzzy Logic and Genetic Algorithm Based Techniques," in International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing, 2009. IJCBS '09, 2009, pp. 348–351.

[41] G. Schaefer and T. Nakashima, "Data Mining of Gene Expression Data by Fuzzy and Hybrid Fuzzy Methods," IEEE Trans. Inf. Technol. Biomed., vol. 14, no. 1, pp. 23–29, Jan. 2010.

[42] A. Lahsasna, R. N. Ainon, R. Zainuddin, and A. Bulgiba, "Design of a fuzzy-based decision support system for coronary heart disease diagnosis," J. Med. Syst., vol. 36, no. 5, pp. 3293–3306, Oct. 2012.

[43] C. Chiu, K.-H. Hsu, P.-L. Hsu, C.-I. Hsu, P.-C. Lee, W.-K. Chiou, T.-H. Liu, Y.-C. Chuang, and C.-J. Hwang, “Mining Three-Dimensional Anthropometric Body Surface Scanning Data for Hypertension Detection,” IEEE Trans. Inf. Technol. Biomed., vol. 11, no. 3, pp. 264–273, May 2007.

[44] U. Bilge, S. Bozkurt, and S. Durmaz, “Application of data mining techniques for detecting asymptomatic carotid artery stenosis,” Comput. Electr. Eng., vol. 39, no. 5, pp. 1499–1505, Jul. 2013.

[45] İ. Güler, H. Polat, and U. Ergün, “Combining Neural Network and Genetic Algorithm for Prediction of Lung Sounds,” J. Med. Syst., vol. 29, no. 3, pp. 217–231, Jun. 2005.

[46] B. Verma and P. Zhang, “A novel neural-genetic algorithm to find the most significant combination of features in digital mammograms,” Appl. Soft Comput., vol. 7, no. 2, pp. 612–625, Mar. 2007.

[47] W. Wongseree, N. Chaiyaratana, K. Vichittumaros, P. Winichagoon, and S. Fucharoen, “Thalassaemia classification by neural networks and genetic programming,” Inf. Sci., vol. 177, no. 3, pp. 771–786, Feb. 2007.

[48] C. C. Bojarczuk, H. S. Lopes, and A. A. Freitas, “Genetic programming for knowledge discovery in chest-pain diagnosis,” IEEE Eng. Med. Biol. Mag., vol. 19, no. 4, pp. 38–44, Jul. 2000.

[49] W.-J. Choi and T.-S. Choi, “Genetic programming-based feature transform and classification for the automatic detection of pulmonary nodules on computed tomography images,” Inf. Sci., vol. 212, pp. 57–78, Dec. 2012.

[50] H. H. Inbarani, A. T. Azar, and G. Jothi, “Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis,” Comput. Methods Programs Biomed., vol. 113, no. 1, pp. 175–185, Jan. 2014.

[51] L.-Y. Chuang, H.-W. Chang, C.-J. Tu, and C.-H. Yang, “Improved binary PSO for feature selection using gene expression data,” Comput. Biol. Chem., vol. 32, no. 1, pp. 29–38, Feb. 2008.

[52] W. Zhao and C. E. Davis, “Swarm intelligence based wavelet coefficient feature selection for mass spectral classification: An application to proteomics data,” Anal. Chim. Acta, vol. 651, no. 1, pp. 15–23, Sep. 2009.

[53] P.-C. Chang, J.-J. Lin, and C.-H. Liu, “An attribute weight assignment and particle swarm optimization algorithm for medical database classifications,” Comput. Methods Programs Biomed., vol. 107, no. 3, pp. 382–392, Sep. 2012.

[54] H.-L. Chen, B. Yang, G. Wang, S.-J. Wang, J. Liu, and D.-Y. Liu, “Support Vector Machine Based Diagnostic System for Breast Cancer Using Swarm Intelligence,” J. Med. Syst., vol. 36, no. 4, pp. 2505–2519, May 2011.

[55] S. Muthukaruppan and M. J. Er, “A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease,” Expert Syst. Appl., vol. 39, no. 14, pp. 11657–11665, Oct. 2012.

[56] H. J. Escalante, M. Montes-y-Gómez, J. A. González, P. Gómez-Gil, L. Altamirano, C. A. Reyes, C. Reta, and A. Rosales, “Acute leukemia classification by ensemble particle swarm model selection,” Artif. Intell. Med., vol. 55, no. 3, pp. 163–175, Jul. 2012.

IJARSE
ISSN 2319 - 8354

Anju Jain received the MCA degree from Guru Jambheshwar University of Science & Technology, Hisar. She is topper and gold medallist in MCA year 2009. She is net qualified in computer science (June 2012) and gate qualified (2013). She also received the M.TECH degree in Computer Sc. & Engg from GJUS&T, Hisar. She is topper in M.Tech year 2013. Presently she is a Phd research scholar in GJU S&T Hisar. Her research area is Machine Learning Techniques for Medical Diagnosis. Presently she is working as the Assistant Professor in the CSE dept. of GJU S&T Hisar.