A SYSTEMATIC REVIEW ON MACHINE LEARNING BASED TECHNIQUES IN MEDICAL DATA ANALYSIS

Author's Name: [1] C.Yosepu, [2]K.Ganapathi Babu

Affiliation: [1,2] *Assistant Professor of Computer Science and Engineering Colleges, Telangana, India*

E-Mail: cyosepu@gmail.com

Abstract

Machine Learning plays an important role in the fields of Artificial Intelligence (AI). In the present era, MachineLearning (ML) has been creating a centre of attention, and it has been successfully functional in many areas such as research, education, business, medical, forecasting etc., . In the present days the diagnosis of diseases is mainly from expert's decision. Diagnosis is a most important task in medical science as it is crucial in determining if a patient is having the disease or not. It plays vital role in the decision of relevant path of diagnosis for the treatment of the diseases. Applying machine learning techniques for disease diagnosis by using various intellectual algorithms has been a scorching research area of computer science. This paper focus on the systematic survey on the machine learning applications in the medical disease prognosis during the past decades.

Keywords: Artificial Intelligence, Diagnosis, Machine Learning, Pattern Recognition, Statistical Learning.

INTRODUCTION

It is well-known that, Medical diagnostic opinion is a prime factor of the clinical subject, considering that the diagnostic method likely to have comparatively significant ambiguities, which more advantageous the complication of medical prognosis and, consequently, gives importance for resolution help in clinical prognosis. Now-a-days in China, medical diagnostic solution is built-in with scientific devices for scan outcome of patients. Situated on the patient's signal of health problem, and the abilities got over the years via medical professionals having to do with the pathology area capabilities, examine symptoms, origins, for the period of making a diagnosis. The purpose of shrewd clinical analysis is that it makes it possible for physicians from the genuine medical signs; try to make the suitable scientific and special diagnosis of sufferers with the enhancement of computer science and artificial intelligence concept in reward many years. Within the scientific subject, computer-aided diagnosis techniques had been mostly used.

Artificial intelligence is a branch of computer science that makes an attempt to make PCs smarter. One of the imperative necessities for any intelligent behavior is to incorporate intelligence in an element of interest. A substantial part of the researchers in these days agree that there is no intelligence without studying. Along these strains, machine learning is certainly one of predominant branches of artificial intelligence and, definitely, it is likely one of the nice very swiftly developing subareas of AI study. Machine learning systems have been from the very dawn invented and used to evaluate scientific knowledge units [41]. On the present-day machine learning helps various fundamental means for intelligent knowledge evaluation. Machine learning technology is at present good suitable for evaluating clinical data, and mainly there is various study entire in medical analysis in small-scale distinctive diagnostic issues. Machine learning and statistical pattern recognition is the most significant discipline within the biomedical society because they advocate assurance for boosting the sensitivity &/ specificity of discovery and diagnosis of ailment, although even as establishing the objectivity of

choice-making mechanism. In the end, the preliminary assurance of those approaches has emerged in handiest finite medical utility, maybe the great eminent of which is the usage of such strategies for mammographic screening [42]. Machine learning study receives giant awareness within the additional development and analysis of linear ways for supervised and unsupervised function extraction and sample classification. Linear ways are drawing attention in decision methods alternatively than nonlinear classification and regression functions. This learn illustrates present progress in machine learning, targeting on supervised and unsupervised approaches and Bayesian inference, which have made robust have an effect on within the discovery and diagnosis of ailment in bio medicine.

RELATED WORK

Amit David and Boaz Lerner [1] carried out a support vector machine that categorizes actual cytogenetic signals from fluorescence in-situ hybridization (FISH) photographs in consideration of assessing genetic syndromes. The research work implements the SVM structural threat minimization concept in seeking the most appropriate structure of the classifier kernel and parameters. Outcomes show off correct performance of the SVM in classifying FISH signals in observation to other state-of the-art machine learning classifiers, demonstrating the capability of a SVM based genetic analysis framework.

Yan Zhang et al. [2] utilized SVM with RBF kernel function which depends on the statistical learning theory for the prognosis of heart disease. To pick the excellent parameters of kernel function and fine essential feature, Grid Search method of making improvements to criterions is utilized, to accomplish probably the most elevated classification accuracy.

Saeed Shariati et al. [3] utilized Fuzzy NN with SVM and ANN to discover and prognosis hepatitis and thyroid illnesses. Further, analysis of disease, they identified the variety and the stage of ailment which incorporate 6 classes for hepatitis ailment, i.e. Hep B (two stages) Hep C (two stages), hepatitis and non- hepatitis and for thyroid disease 5 classes were labeled, specifically: Hypothyroid, thyrotoxicosis, Subclinical hypothyroid, Subclinical thyrotoxicosis and non-presence of thyroid. For hepatitis disease the excellent accuracies bounds to 98% and for thyroid ailment to 99%.

Hanaa Ismail Elshazly et al. [4] proposed a Genetic algorithm headquartered support vector machine classifier for lymph ailments analysis. In the foremost stage, the dimensions of lymph ailments dataset have 18 features and they are diminished to six aspects by means of utilizing GA. A support vector machine with more than a few kernel functions including linear, Quadratic and Gaussian were used as a classifier in the 2nd stage. The SVM classifier with every kernel function is used to assess the performance by making use of efficiency indices similar to accuracy, sensitivity, specificity, AUC/ROC, Matthews Correlation Coefficient and F-Measure. Linear kernel function bought a most improved effect which validates the competency of GA-linear system.

D. Tomar et al. [5] presented an effective Parkinson ailment analysis system using LST-SVM and Particle Swarm Optimization-PSO. PSO is utilized for characteristic selection and parameter optimization. The efficiency of proposed procedure is compared with different distinctive current procedures in terms of accuracy, sensitivity and specificity. Empirical outcomes justify the adequacy of the proposed Parkinson ailment evaluation method over different existing procedures with 98% accuracy.

SNO	AUTHORS	YEAR	METHODOLOGY	DATASET
1	Tomar et al.[6]	2014	Feature Selection Based Least Square Twin SVM	Heart disease
2	S.Ghosh et al.[7]	2014	MLP BPN &Support Vector Machine	Breast Cancer
3	Avik Basu et al.[8]	2015	Support Vector Machine with linear kernel	Erythemato-Squamous Disease
4	Seyede Zahra Paylakhi et al.[9]	2015	GA-SVM	DNA microarray gene expression data of Alzheimer's disease
5	Mohammed El Amine LAZOUNI et al.[10]	2014	SVM with different kernel functions	Pre-anesthetic Examination
6	Tipawan Silwattananusarn et al.[11]	2016	Information Gain feature selection and Correlation- based feature selection with SVM ensembles	Cardiotocography
7	Austeclino Magalhaes Barros Junior et al.[12]	2010	ANN and Bayesian Network	Asymptomatic Malaria Infection
8	Luis Javier Herrera et al.[14]	2013	Wavelet Feature Extraction, Support Vector Machines	Alzheimer's disease
9	Carlos Arizmendi et al.[13]	2011	Artificial Neural Networks	Brain Tumour
10	Mrudula Gudadhe et al.[15]	2010	Support Vector Machine and Artificial Neural Network	Heart Disease
11	Suchita Saha et al.[16]	2015	Support Vector Machine ,Extreme Learning Machine	Cardiac Pathology
12	R. Suganya, S. Rajaram[17]	2013	Support Vector Machine	Ultrasound Liver Images
13	Muhammad Fathurachman et al.[18]	2014	ANN, Extreme Learning Machine	Heart Disease
14	Paweł Marciniak et al.[19]	2014	Artificial Neural Networks, Bayesian Networks, Decision Trees	Cardiovascular Diseases
15	M. Lopez et al.[20]	2009	Kernel PCA and Support Vector Machines	Neurological Image Classification for the Alzheimer's Disease
16	Joarder Kamruzzamanand et al.[21]	2006	Linear Discriminant Analysis, Multilayer Perceptron, SVM	Cerebral Palsy Gait
17	Hung T. Nguyen et al.[22]	2007	Bayesian neural network	Hypoglycemic Episodes in Children with Type 1Diabetes
18	F. D. S. Farias et al.[23]	2012	Artificial Neural Intelligence, Naïve Bayes, Bayesian Networks and Decision Tree	Lymphotropic Virus
19	Muhammad Shoaib. B.Sehga1 et al.[24]	2004	Generalized Regression Neural Network ,Support Vector Machine and Probabilistic Neural Network	Genetic Mutations in Ovarian Cancer
20	Bo Pang et al.[25]	2004	Digital Image Processing techniques and Bayesian networks	Tongue Abnormal Appearances and Diseases

Fatemeh Saiti et al. [26] recommended SVM and Probabilistic NN for the classification of two thyroid diseases Hypothyroidism and Hyperthyroidism from the thyroid disorder database. These algorithms rely regularly likely on effective classification algorithms to handle excessive and unimportant aspects. Genetic Algorithm tested a handy and strong framework for settling on fair subsets of aspects that result in fortifies prognosis rates.

Alaa Elsayad et al. [27] evaluated the affectivity of Bayesian classifier in diagnosing the hazard of cardiovascular ailment. Two Bayesian network classifiers; Tree Augmented Naïve Bayes and

the Markov Blanket Estimation are implemented and their prognosis certainties are the reference points in action to the Support Vector Machine. The exploratory outcomes exhibit that Bayesian networks with MBE have the classification precision of 97%, as long as TAN and SVM units have 88 and 71 percentages.

S Hongzong et al. [28] keen on the application of SVM on the coronary heart ailment and non-coronary heart disorder classification. Linear discriminant evaluation and SVM with a radial basis function kernel are compared. The prediction precisions of training and evaluation units of SVM were 96% and 78% subsequently, and for LDA it was 90 and 72 percentages subsequently. The cross-validated precision of SVM and LDA were 92 and 85 percentages.

Emre Çomak et al. [29] presented a DSS that categorizes the Doppler signals of heart valve to 2 lessons (traditional and abnormal) to help the cardiologist. LS-SVM and ANN with Back Propagation are implemented to classify the drawn out elements. Further, ROC curve is used to analyze sensitivities & specificities of those classifiers and estimate the AUC. Ultimately, two classifiers are evaluated in all facets. Javad Salimi Sartakhti et al [30] suggested an innovative machine learning system that infuses support vector machine and simulated annealing for the analysis of hepatitis disorder. The accuracy of the classification is received by using 10-fold cross validation. The assessed classification precision of the proposed process was 96.25 percentages.

D. Vassis et al. [31] made a complete overview regarding the usage of neural networks in automated medical prognosis, with a designated precedence in Support Vector Machines (SVMs), which are exact varieties of neural functions. Over the evaluation, in abounding cases, symptoms and ailments may also be well predicted by way of neural programs, while SVMs are step by step used in clinical prognosis because of their special classification elements.

Vidyullatha .P et al. [32] gave an idea about decision making for the heart patient. For diagnosing the heart disease, the right decision and treatment recommended by the physicians is so important so far. This Rough Set Theory is used to make a reliable decision for heart data patients. The rough set method generates the rules which will provide clarity with higher accuracy so that the cardiologist can make the decision whether the patient suffers heart problem or not.

Paweł Szewczyk et al. [33] discussed a LS-SVM to the fundamental analysis of patients. K-fold Cross Validation, Grid-Search, and PSO methods had been used to search out some most suitable parameters. The classification results had been confirmed by using some class labels determined through a reliable trained. The process developed has been proved on the unreal knowledge and the data gathered from the true data base. Konstantina Kourou et al. [34] reviewed recent Machine Learning procedures which can be employed within the modeling of cancer progression. The items mentioned right here for prediction is related to more than a few supervised Machine Learning procedures moreover to exceptional input points and data samples.

M. Naresh Kumar [35] developed a vigorous and persuasive decision tree headquartered method for predicting dengue sickness. A decision tree is trained and in comparison its efficiency with C4.5 algorithm for dengue sickness analysis. The dataset with sixty five patient

documents are diagnosed and from which fifty three individuals were rooted to have dengue fever. A decision tree centered algorithm has been equipped to discriminate the dengue fever utilizing analytical and laboratory information with no. of perfectly classified examples as 89%, F-measure of 0.86 & ROC of 0.826 in comparison with C4.5.

Shweta Kharya [36] mentioned quite a lot of data mining techniques which had been used for breast malignancy analysis and prognosis. More than a few survey and scientific articles on breast cancer analysis and prognosis are summarized as well as directed on research accomplished utilizing the data mining techniques to reinforce the breast malignancy analysis and diagnosis.

Erdem Alkım et al. [38] developed a new, rapid and adaptive disease prognosis system founded on LVQ- RM artificial neural networks. The proposed system handles data with absent values to justify that this procedure does no longer offer a specified solution for a distinctive ailment and a right analysis percentage after replacing missing values utilizing median approach over 99%.

Samaa Farhan et al. [39] developed a DSS for the prognosis of three ENT ailments using Artificial Neural Network (ANN). Those three diseases are: chronic infection rhinosinusitis, Otitis external and Pharyngitis. A MLP neural network is implemented in developing the system. The input layer consists of thirty eight variables and the output layer includes one neuron which predicts one type of the three ENT diseases. Different activation functions such as, Linear Activation Function (LF), Hyperbolic Tangent Sigmoid Activation Function (TANH) and Log-Sigmoid Activation Function are used for modeling this system. The results obtained for classification accuracy in our system is 95.41%.

CONCLUSION

Hence, a comprehensive survey has been carried out on previous works covered on the usage of automated and semi-automated medical diagnosis used in the application of specific diseases like cancer, brain tumor, ENT ailment, coronary heart disease, thyroid, liver disorders etc; and classification in specific. The use of various neural networks, KNN, fuzzy systems and several classifiers similar to support vector machines, decision trees, and Bayesian networks has been focused in this survey. Lastly, Data Mining techniques, Machine Learning approaches and several other algorithms has proven their efficiency in the area of medical data analysis and in the prediction of diseases.

REFERENCES

1. Amit David and Boaz Lerner," PATTERN CLASSIFICATION USING A SUPPORT VECTOR MACHINE FOR GENETIC DISEASE DIAGNOSIS" Proceedings. 2004
2. 23rd IEEE Convention of Electrical and Electronics Engineers in Israel, pp 289~292.
3. Yan Zhang, Fugui Liu, Zhigang Zhao, Dandan Li, Xiaoyan Zhou, Jingyuan Wang," Studies on application of Support Vector Machine in diagnose of coronary heart disease" Sixth International Conference on Electromagnetic Field Problems and Applications (ICEF), 2012,pp 1~4.
4. Saeed Shariati, Mahdi Motavalli Haghighi ," Comparison of ANFIS Neural Network with Several Other ANNs and Support Vector Machine for Diagnosing Hepatitis and Thyroid Diseases", International Conference on Computer Information Systems and Industrial Management Applications (CISIM), 2010,pp 596~599.

5. Hanaa Ismail Elshazly, Abeer Mohamed Elkorany, Aboul Ella Hassanien," Lymph diseases diagnosis approach based on support vector machines with different kernel functions", 9th International Conference on Computer Engineering & Systems (ICCES), 2014,pp 198~203.

6. D. Tomar; B. R. Prasad; S. Agarwal," An efficient Parkinson disease diagnosis system based on Least Squares Twin Support Vector Machine and Particle Swarm Optimization" 9th International Conference on Industrial and Information Systems (ICIIS), 2014,pp 1~6.

7. Tomar, Divya, and Sonali Agarwal. "Feature Selection based Least Square Twin Support Vector Machine for Diagnosis

8. Avik Basu, Sanjiban Sekhar Roy, Ajith Abraham," A Novel Diagnostic Approach Based on Support Vector Machine with Linear Kernel for classifying the erythemato- squamous disease", International Conference on Computing Communication Control and Automation (ICCUBEA), 2015 ,pp 343 – 347.

9. Seyede Zahra Paylakhi, Sadjaad Ozgoli, Seyed Hassan Paylakhi, "A novel gene selection method using GA/SVM and Fisher criteria in Alzheimer's disease", IEEE 23rd Iranian Conference on Electrical Engineering (ICEE)2015.

10. Mohammed El Amine LAZOUNI, Nesma SETTOUTI, Mostafa El HABIB DAHO, Mostafa El HABIB DAHO," An SVM Intelligent System for Pre-anesthetic Examination", IEEE Second World Conference on Complex Systems (WCCS), 2014,pp 73 – 78

11. Tipawan Silwattananusarn, Wanida Kanarkard, Kulthida Tuamsuk," Enhanced Classification Accuracy for Cardiotocogram Data with Ensemble Feature Selection and Classifier Ensemble", Journal of Computer and Communications, 2016, 4, 20-35.

12. Austeclino Magalhaes Barros Junior, Angelo Amancio Duarte, Manoel Barral Netto/Bruno Bezerril Andrade," ARTIFICIAL NEURAL NETWORKS AND BAYESIAN NETWORKS AS SUPPORTTING TOOLS FOR DIAGNOSIS OF ASYMPTOMATIC MALARIA" 12th IEEE International Conference on e-Health Networking Applications and Services (Healthcom), 2010, pp 106 – 111.

13. Carlos Arizmendi, Daniel A. Sierra, Alfredo Vellido, Enrique Romero," Brain Tumour Classification Using Gaussian Decomposition and Neural Networks", 33rd Annual International Conference of the IEEE EMBS Boston, Massachusetts USA, August 30 – September 3, 2011,pp 5645-5648.

14. Luis Javier Herrera, Ignacio Rojas, H. Pomares, A. Guillén, O. Valenzuela, O. Baños," Classification of MRI images for Alzheimer 's disease detection", SocialCom/PASSAT/BigData/EconCom/BioMedCom 2013,pp 846-851.

15. Mrudula Gudadhe, Kapil Wankhade, Snehlata Dongre, "Decision Support System for Heart Disease based on Support Vector Machine and Artificial Neural Network", IEEE International Conference on Computer & Communication Technology(ICCCT'10),pp 741-745.

16. Suchita Saha, Santanu Ghorai," Effect of Feature Fusion for Discrimination of Cardiac Pathology", Third International Conference on Computer, Communication, Control and Information Technology (C3IT), 2015, pp 1-6.

17. R.Suganya, S.Rajaram," Feature Extraction and Classification of Ultrasound Liver Images using Haralick Texture-Primitive Features: Application of SVM Classifier", 2013 International Conference on Recent Trends in Information Technology (ICRTIT), pp 596 – 602.

18. Muhammad Fathurachman, Umi Kalsum, Noviyanti Safitri, Chandra Prasetyo Utomo," Heart Disease Diagnosis using Extreme Learning Based Neural Networks", 2014

19. International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA),pp 23-27.

20. Paweł Marciniak, Rafał Kotas, Marek Kamiski, Zygmunt Ciota," Implementation of Artificial Intelligence Methods on Example of Cardiovascular Diseases Risk Stratification", MIXDES 2014, 21st

21. International Conference "Mixed Design of Integrated Circuits and Systems", June 19-21, 2014, Lublin, Poland, pp 503-507.

22. M. Lopez, J. Ramırez, J. M. Gorriz, D. Salas-Gonzalez, I. Alvarez, F. Segovia and R. Chaves," Neurological Image Classification for the Alzheimer's Disease Diagnosis Using Kernel PCA and Support Vector Machines", 2009 IEEE Nuclear Science Symposium Conference Record, pp 2486-2489.

23. Joarder Kamruzzamanand, Rezaul K. Begg," Support Vector Machines and Other Pattern Recognition Approaches to the Diagnosis of Cerebral Palsy Gait", IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 53, NO. 12, DECEMBER 2006, pp 2479-2490.

24. Hung T. Nguyen, Nejhdeh Ghevondian, Son T. Nguyen, Timothy W. Jones," Detection of Hypoglycemic Episodes in Children with Type 1Diabetes using an Optimal Bayesian Neural Network Algorithm", Proceedings of the 29th Annual International Conference of the IEEE EMBS Cité Internationale, Lyon, France August 23-26, 2007,pp 3140-3143.

25. F. D. S. Farias, L. V. D. Souza, R. C. M. Sousa, C. A. M. Caldas, and L. F. Gomes, J. C. W.Costa," Data Mining Applied to Diagnose Diseases Caused by Lymphotropic Virus: a Performance Analysis", IEEE LATIN AMERICA TRANSACTIONS, VOL. 10, NO. 1, January. 2012, pp 1319-1323.

26. Muhammad Shoaib. B.Sehga1, Iqbal Gondal, Laurence Dooley," Statistical Neural Networks and Support Vector Machine for the Classification of Genetic Mutations in Ovarian Cancer" Proceedings of the 2004 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, 2004. CIBCB '04, pp 140 – 146.

27. Bo Pang, David Zhang, Naimin Li, and Kuanquan Wang," Computerized Tongue Diagnosis Based on Bayesian Networks", IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, NO.10, VOL. 51, OCTOBER 2004, pp 1803-1810.