A Survey on Data mining Algorithms for Tumor Detection and Classification

T. R. Nisha Dayana [1] *, A. Lenin Fred [2]
[1]Research Scholar, Bharathiar University Tamil Nadu, India
[2]Mar Ephraem College of Engineering and Technology, Tamil Nadu, India
Email: leninfredip@gmail.com

Abstract

Data mining techniques plays vital role in engineering, medicine and management sectors for the analysis of huge data. Computer aided algorithms supports the clinicians for disease diagnosis and surgical preplanning. This research work highlights the widely used data mining algorithms for tumor diagnosis, segmentation and classification. The findings reveal that many improvements can be done in the data mining algorithms for improving the accuracy.

Keywords: segmentation, classification, preplanning, data mining

1. Introduction

The data mining has immense applications in image processing, statistics and robotics [1]. The role of data mining is inevitable in medical field for disease diagnosis and therapeutic planning. Huge amount of data is generated in medical field owing to the use of many instruments and equipment's for disease diagnosis and treatment planning [1]. The signal processing and image processing algorithms are coupled with the data mining algorithms for practical applications [2]. In [3], applications of data mining in medical field were analyzed. The data mining has proved its efficiency in the areas of predictive medicine, customer relationship management and management of health care. A detailed study was performed highlighting the applications of data mining in medical data classification[3].The data mining algorithms are broadly classified into two categories; predictive and descriptive. For cancer diagnosis, data mining algorithms on an average produces accuracy greater than 90% and for IVF treatment cases; an accuracy of 70% was produced [4]. The role of data mining algorithms in decision making and precision diagnosis in medical field was described in [5]. Section ii highlights the related works of data mining algorithms in medical field.

2. Related Works

The flow diagram represents the classification of data mining algorithms. In predictive mining, data is used for prediction and forecasting is done based on the patterns identified by known results, where as in descriptive mining, the properties of the data are analyzed and does not have predefined targets. Parveen Khan et al analyzes various data mining techniques for the analysis of MR brain data sets. The Decision Trees, Support Vector Machine, Artificial Neural Networks (ANN) and Fuzzy logic based techniques were analyzed and the results reveals that, it is difficult to judge a efficient single data mining algorithm [6]. The ANN based algorithms were found to be efficient for larger datasets and SVM approach was found to be efficient for smaller data sets. Data mining algorithms were found to be robust for cancer diagnosis based on the analysis of patient's health records [7]. The Decision tree classifier was found to be efficient for the brain tumor detection and classification, preprocessing was performed by median filter and association rules are applied for the extracted texture features [8]. An accuracy of 96% and sensitivity of 93% were obtained by the incorporation of association rules in decision tree classifier. The feed forward neural network based segmentation of white matter, grey matter and cerebro spinal fluid yields efficient results, when compared with the K-NN and Bayesian classifier [9]. The generalized eigenvalue proximal support vector machine (GEPSVM) with RBF kernel and feature extraction by stationary wavelet transform (SWT), optimization by principal component analysis (PCA) generates efficient classification results for MR brain images when compared with the classical techniques [10].

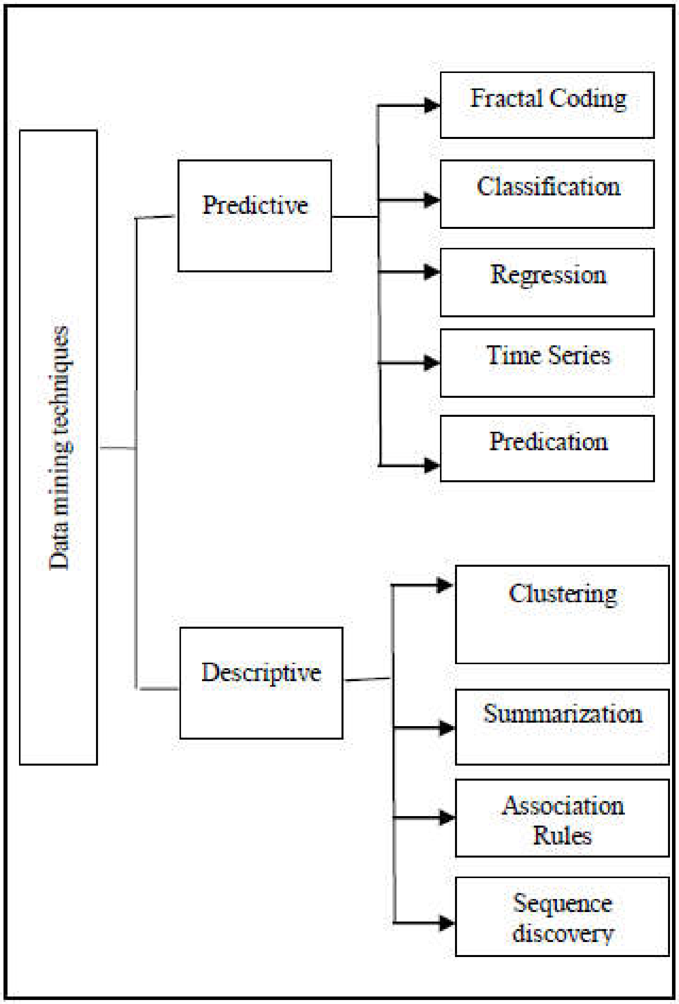

Figure 1: Classification of data mining algorithms

The least square support vector machine classifier with multilayer perceptron efficiently classifies the brain tumor stages in MR brain images [11]. The preprocessing was performed by median filtering and feature extraction was performed by gray level co-occurrence matrix. The generalized eigenvalue proximal SVM with feature extraction by weighted-type fractional Fourier transform (WFRFT) and optimization by principal component analysis (PCA) yields efficient classification results for MR brain images [12]. The FCM along with SVM performs the automatic detection of brain tumor in MR brain images; feature extraction was performed by Grey level run length matrix (GLRLM) [13]. The genetic algorithm was coupled with SVM for brain tumor classification; feature extraction was performed by 2D Wavelet Transform and Spatial Gray Level Dependence Matrix [14]. A predictive linear regression model was proposed for the prediction of brain tumor, the training data was balanced by SMOTE oversampling technique [15]. The cross correlated K-nearest neighbor with feature extraction by gray level co-occurrence matrix efficiently performs the classification of breast MR images [16].

The Nonnegative Matrix Factorization was found to be effective for reducing the dimensionality of data for brain tumor detection [17].In [18], a detailed analysis has been performed on the data mining algorithms for brain tumor detection in MR images of children. The MR data set comprising of 190 images were used; decision tree algorithm has best classification rate of 96.16%, while in terms of computation time, KNN is best having an execution time of 0.03 seconds. A novel data mining approach combining feature selection with ensemble-based classification was used for the brain tumor detection in patients' health records [19]. The globally optimized Artificial Neural Network Input Gain Measurement Approximation (GANNIGMA) based hybrid feature selection that is Coupled d with an ensemble classification was used, ensemble classification improves the accuracy. The statistical analysis of the result was performed by regression technique. The SVM and ANN gives promising results for the detection of carcinogenic tissues in hyper spectral database of human beings [20]. The sensitivity and specificity of SVM classification was 92.61% and 92.44%, whereas for ANN based system, the sensitivity and specificity was 92.44% and 92.77%.

The text and data mining has immense applications in genomics, pharmacogenomics and other medical fields [21]. The Multiple Kernel Learning with Adaptive Neuro-Fuzzy Inference System (MKL with ANFIS) based deep learning was applied for the diagnosis of heart diseases [22]. The MKL-ANFIS algorithm generates efficient results when compared with the Least Squarewith Support Vector Machine (LS with SVM), General Discriminant Analysis and Least Square Support Vector Machine (GDA with LS-SVM), Principal Component Analysis with Adaptive Neuro-Fuzzy Inference System (PCA with ANFIS) and Latent Dirichlet Allocation with Adaptive Neuro-Fuzzy Inference System (LDA with ANFIS algorithms. The MKL-ANFIS algorithm has sensitivity (98%), specificity (99%) and Mean Square Error (0.01) for the for the KEGG Metabolic Reaction Network dataset. In [23], various classification algorithms like Decision Table, OneR and DTNB are used for the analysis of liver disorder analysis. The DTNB data mining algorithm yields efficient results than other techniques; classification accuracy of 86.48% was produced, kappa coefficient was 0.85 and error rate was 0.2315.

The brain disorder analysis was performed by Naive Bayes, Support Vector Machine, Random Tree and C4.5 data mining classifiers [24]. The SVM was found to generate efficient results when compared with the other algorithms. For optimum feature selection, principal component analysis (PCA), Runs filtering, Fisher filtering and Relief feature selectors are used. The SVM was found to generate an accuracy of 73.33% with Relief feature selector; MLC brain data set was used for analysis. In [25], neural network, KNN and .Bayesian classifier were used for the tissue segmentation and disorder classification in MR brain images. The neural network based system was found to be proficient, when compared with the other techniques. The 20 MR brain data sets were used, an accuracy of 83% was generated by neural network based classification, while KNN and Bayesian classifier has 67% accuracy.

3. Conclusion

This research work analyzes widely used data mining algorithms for the disease diagnosis and treatment planning. The hybrid approaches were found to generate robust results, when compared with the classical methods. The outcome of this work will be an aid for researchers developing novel algorithms in data mining for medical applications.

References

[1] Suresh R, Dhivya P, Bhuvana N. "Analysis on image mining techniques". Innovations in Information, Embedded and Communication Systems (ICIIECS), IEEE (2017);(1-5).

[2] Acharya UR, Yu W,"Data mining techniques in medical informatics", The open medical informatics journal, **(2010)** ;4:21.

[3] Lashari SA, Ibrahim R, Senan N, Taujuddin NS," Application of Data Mining Techniques for Medical Data Classification: A Review", InMATEC Web of Conferences; Vol. 150, **(2018)** EDP Sciences.

[4] Tomar D, Agarwal S,"A survey on Data Mining approaches for Healthcare", International Journal of Bio-Science and Bio-Technology, **(2013)** Oct;5(5):241-66.

[5] Lashari SA, Ibrahim R, Senan N, Taujuddin NS,"Application of Data Mining Techniques for Medical Data Classification: A Review", InMATEC Web of Conferences Vol. 150, **(2018)** EDP Sciences.

[6] Khan P, Singh A, Maheshwari S,"Automated brain tumor detection in medical brain images and clinical parameters using data mining techniques: a review", International Journal of Computer Applications. **(2014)** Jan 1;98(21).

[7] Murphy DR, Laxmisan A, Reis BA, Thomas EJ, Esquivel A, Forjuoh SN, Parikh R, Khan MM, Singh H,"Electronic health record-based triggers to detect potential delays in cancer diagnosis", BMJ Qual Saf. **(2014)** Jan 1;23(1):8-16.

[8] Naik J, Patel S," Tumor detection and classification using decision tree in brain MRI", International Journal of Computer Science and Network Security (IJCSNS), **(2014)** Jun 1;14(6):87

[9] Damodharan S, Raghavan D," Combining Tissue Segmentation and Neural Network for Brain Tumor Detection", International Arab Journal of Information Technology (IAJIT), **(2015)** Jan 1;12(1).

[10] Zhang Y, Dong Z, Liu A, Wang S, Ji G, Zhang Z, Yang J,"Magnetic resonance brain image classification via stationary wavelet transform and generalized eigenvalue proximal support vector machine", Journal of Medical Imaging and Health Informatics. **(2015)** Dec 1;5(7):1395-403.

[11] Praveen GB, Agrawal A," Hybrid approach for brain tumor detection and classification in magnetic resonance images", InCommunication, Control and Intelligent Systems (CCIS), **(2015)** Nov 7 pp. 162-166. IEEE.

[12] Zhang YD, Chen S, Wang SH, Yang JF, Phillips P," Magnetic resonance brain image classification based on weighted-type fractional Fourier transform and nonparallel support vector machine", International Journal of Imaging Systems and Technology, **(2015)** Dec;25(4):317-27

[13] Singh A," Detection of brain tumor in MRI images, using combination of fuzzy c-means and SVM", InSignal Processing and Integrated Networks (SPIN), 2015 2nd International Conference on 2015 Feb 19, **(2015)** pp. 98-102. IEEE.

[14] Kharrat A, Halima MB, Ayed MB," MRI brain tumor classification using Support Vector Machines and meta-heuristic method", InISDA **(2015)** Dec 14 pp. 446-451.

[15] Zhou M, Hall LO, Goldgof DB, Gillies RJ, Gatenby RA, "Imbalanced learning for clinical survival group prediction of brain tumor patients", InMedical Imaging 2015: Computer-Aided Diagnosis, International Society for Optics and Photonics, Vol. 9414 **,(2015)** Mar 20, p. 94142K.

[16] Waugh SA, Purdie CA, Jordan LB, Vinnicombe S, Lerski RA, Martin P, Thompson AM," Magnetic resonance imaging texture analysis classification of primary breast cancer",European radiology, **(2016)** Feb 1; 26(2):322-30.

[17] Prajapati SJ, Jadhav KR,"Brain tumor detection by various image segmentation techniques with introduction to non-negative matrix factorization. Brain", **(2015)** Mar; 4(3):600-3.

[18] Ali EM, Seddik AF, Haggag MH," Using Data Mining Techniques for Children Brain Tumors Classification based on Magnetic Resonance Imaging", International Journal of Computer Applications. **(2015)**; 131(2).

[19] Huda S, Yearwood J, Jelinek HF, Hassan MM, Fortino G, Buckland M,"A hybrid feature selection with ensemble classification for imbalanced healthcare data: A case study for brain tumor diagnosis", IEEE access. **(2016)** Jan 1; 4:9145-54

[20] Ortega S, Callicó GM, Plaza ML, Camacho R, Fabelo H, Sarmiento R," Hyperspectral database of pathological in-vitro human brain samples to detect carcinogenic tissues", InBiomedical Imaging (ISBI), IEEE 13th International Symposium on 2016, **(2016)** Apr 13 (pp. 369-372). IEEE.

[21] Gonzalez GH, Tahsin T, Goodale BC, Greene AC, Greene CS,"Recent advances and emerging applications in text and data mining for biomedical discovery", Briefings in bioinformatics**. (2015)** Sep 29; 17(1):33-42.

[22] Manogaran G, Varatharajan R, Priyan MK,"Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system", Multimedia tools and applications. **(2018)** Feb 1; 77(4):4379-99.

[23] Mishra S, Tripathy HK, Mishra B, Sahoo S,"Implementation of Classification Rule Mining to minimize Liver Disorder risks", International Journal of Control Theory and Applications. **(2017)**; 10(18):17-24.

[24] Ramani RG, Sivaselvi K,"Classification of Pathological Magnetic Resonance Images of Brain Using Data Mining Techniques", InRecent Trends and Challenges in Computational Models (ICRTCCM), Second International Conference on 2017, **(2017)** Feb 3 (pp. 77-82). IEEE

[25] Damodharan S, Raghavan D,"Combining Tissue Segmentation and Neural Network for Brain Tumor Detection.", International Arab Journal of Information Technology (IAJIT),**(2015)** Jan 1;12