IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693

Dual Threshold Based Classification Technique (DTBCT) For Assessing Liver Abnormalities from Medical Images

R. Malathi[1*], S. Ravichandran[2]

[1,2] PG and Research Department of Computer Science, H.H. The Rajah's College (A), Pudukkottai, Tamilnadu, India

Corresponding Author: shamalu2010@gmail.com, Tel.: +91-9791755884

DOI: https://doi.org/10.26438/ijcse/v7i5.14361439 | Available online at: www.ijcseonline.org

Accepted: 21/May/2019, Published: 31/May/2019

Abstract— Healthcare systems have been using data mining to predict disease in recent years. Early prediction of liver diseases is important to save human life, mainly to decrease mortality rates by taking appropriate disease control measures. This paper explores early predictions of liver disease through various classification techniques. The liver disease dataset selected for this study consists of 15 CT scan images of the liver. The images were segmented with GLCM features. The main purpose of this paper is to propose a hybrid classifier algorithm for predicting liver diseases involving multiple techniques. The proposed technique is also compared with existing classifiers like Naïve Bayes, K nearest neighbor and support vector machine (SVM) on the scales of sensitivity, specificity and classification accuracy. Experimental results of the proposed hybrid classifier algorithm were found to better in predicting liver diseases.

Keywords— Classification, GLCM, liver disease, prediction, standard deviation.

I. INTRODUCTION

Liver plays an important role in many human body functions like protein production, blood coagulation and maintenance of cholesterol, glucose (sugar) and iron metabolism levels in the body. It is also responsible in removing toxins from the human body, thus helping human survival. Any loss in these functions can cause serious damage to the body. When a liver is infected with a virus or damaged by chemicals or attacked by its own immune system, it gets severely damaged making survival difficult. Liver diseases caused by hepatotrophic viruses burden the human body and healthcare resources. Persistent infections from hepatitis B virus (HBV), hepatitis C virus, and hepatitis delta virus result in chronic liver diseases. The definition of acute liver disease stems from the duration of the disease with a medical history not greater than 6 months. Majority of acute liver diseases get formed from acute viral hepatitis [1]. Typical symptoms of liver disease are nausea, vomiting, upper right abdominal pain, fatigue and weakness. Symptoms in patients includes jaundice, abdominal pain, fatigue, nausea, vomiting, back pain, abdominal swelling, weight loss, fluid in abnormal cavity, general itching, pale stool, enlarged spleen and gallbladder. Though symptoms of liver diseases may vary, they basically include swelling of the abdomen and legs, easy staining, changes in stool and urine color, jaundice or yellowing of the skin and eyes. In some cases very limited symptoms exist in patients where imaging and liver function tests can help detect liver damages, thus helping in diagnostic procedures of liver diseases [2]. This paper proposes a novel technique called Dual threshold based classification technique (DTBCT) for assessing liver abnormalities from medical images.

II. RELATED WORK

S. Dangare et al. [3] analyzed predictive heart disease systems using more input attributes using data mining classification techniques like namely decision trees, naive Bayes and neural networks on the heart disease database. The performance of these techniques was compared based on accuracy. The author's showed that neural networks predicted heart disease with highest accuracy. Omar S. Soliman et al. [4] proposed a hybrid classification system for HCV diagnosis using an improved particle swarm optimization algorithm and a least squares support vector machine (LS-SVM). The technique extracted feature vectors using a principal component analysis algorithm. Since the LS-SVM algorithm is sensitive to changes in its parameter values, a Modified-PSO algorithm was used to find the optimal value of the LS-SVM parameters for lesser iterations. The proposed system was implemented and evaluated in the AVC reference data set from UCI repository of machine learning databases. The system was compared with another classification system using PCA and LS-SVM. Its experimental results achieved higher classification accuracy when compared to other systems. Karthik et.al [5] applied soft computing techniques for intelligent diagnosis of liver diseases. They implemented their classification and type detection in three phases. In the first phase, liver diseases were using an artificial neural network (ANN) classification algorithm. In the second phase, they uses the learning algorithm (LEM) to generate classification rules by approximation. In the third stage, fuzzy rules were applied to identify the type of liver disease. Dhamodharan et.al [6] predicted three important liver diseases like liver cancer, cirrhosis and hepatitis using different symptoms. They used NaïveBayes and FT Tree algorithms to predict the disease. The two algorithms were then compared based on their classification accuracy. Their experimental results concluded that NaïveBayes was a better algorithm for predicting diseases, since its classification accuracy was higher than other algorithms. Rosalina et al [7] predicted a disease with hepatitis prognosis using support vector machine (SVM) and packaging methods. The packaging methods eliminated noises before the classification process. SVM was used to choose features for achieving higher accuracy. A Functional selection process minimized noise or irrelevant data. The experimental results and observations showed higher accuracy of clinical laboratory testing with minimized costs and shorter execution times. They achieved their goals by combining Wrappers method with SVM technology.

III. METHODOLOGY

This research work identifies liver diseases by using the liver image structure, liver features and high order statistics. A batch of 15 scanned segmented liver CT images were used for the development this prototype. The procedure includes pre-processing procedures like filtering, dual threshold, erosion and dilation. All the selected features were quantified and analyzed on the statistical tool, MATLAB version 2011a [8].

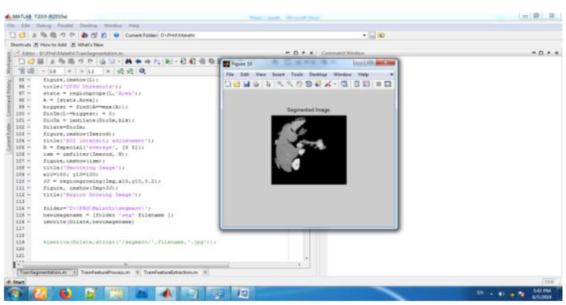

Figure 1. Segmented CT scan liver query image

Figure 1 depicts the image of a segmented liver CT scan image from the group of 15 training images. The batch of images is processed filtered using techniques discussed above. The resulting images are a noise reduced and enhanced image. Gaussian median filtering is used for removing noises in the images. Figure 2 depicts a noise removed image.

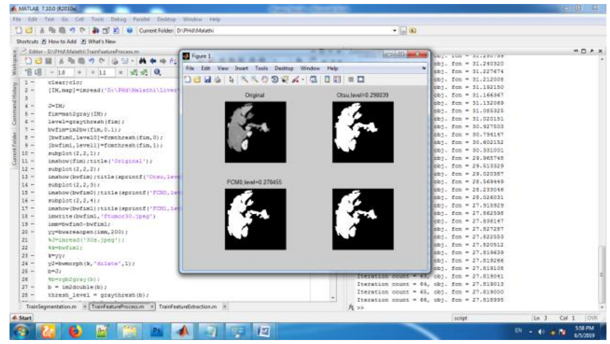

Figure 2. Filtered CT scan liver image

The filtered images then are passed for further processing called dual threshold. The inbound and outbound edges are detected using this technique which helps in drawing a perfect image with accurate boundaries. Figure 3 shows the dual threshold image with a clear boundary.

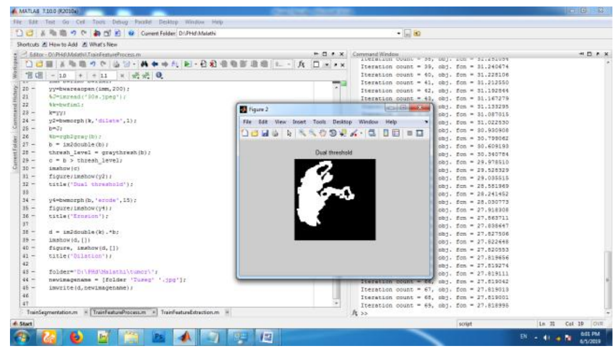

Figure 3. Dual threshold processed liver image

The inner and outer boundaries of the image are corrected in the next step. DTBCT uses erosion and dilation methods for correcting the boundaries. Finally a segmented liver image that is the best output in all respects results for feature extractions. An example liver image after passing through erosion and dilation techniques is depicted in Figures 4 and 5.

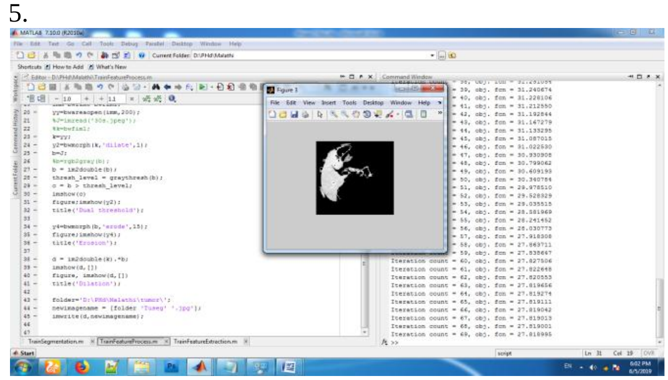

Figure 4. Eroded liver image

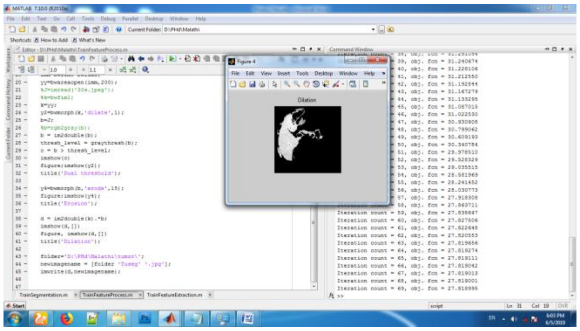

Figure 5. Dilated liver image

IV. EXPERIMENTAL RESULTS

Methods suggested for accurate diagnosis of liver disease using image processing methods are voluminous. These techniques extract features from the image and are based on image characteristics. However, the use of statistical methods in feature extractions for image analysis from medical images have been less proposed [9]. In DTBCT, all the 15 images features were extracted by adopting higher order statistical calculations as mentioned previously in materials and methods. A total of 24 features such as auto correlation, contracts, correlation, dissimilarity, energy, entropy, homogeneity, cluster prominence, cluster shade, sum average, sum of squares, area, eccentricity, orientation, convex area, filled area, Euler number, equivalent diameter, solidity, major axis length, minor axis length and perimeter were determined. These quantified units were processed for obtaining mean and standard deviation values (STDEV). Sample GLCM feature values are furnished in Table 1, while Table 2 lists the STDEV values of the image batch.

Table 1. Extracted feature set 1 (GLCM features)

Img id	auto correlation	Contracts	Correlation	Dissimilarity	energy	entropy
1	3.170644685	0.187131	0.897843	0.08563	0.586065	0.890249
2	6.170583169	0.689838	0.832082	0.304134	0.381904	1.513508
3	6.131028543	0.305549	0.908533	0.186823	0.326543	1.526393
4	5.701279528	0.321973	0.925952	0.160187	0.501184	1.239122
5	6.091535433	0.792815	0.821177	0.30844	0.419222	1.428142
6	2.704662894	0.289247	0.830318	0.14813	0.644858	0.935741
7	5.018147146	0.515563	0.818049	0.263103	0.349165	1.529005
8	7.434793307	0.85876	0.835201	0.371678	0.343536	1.674971
9	3.170644685	0.187131	0.897843	0.08563	0.586065	0.890249
10	2.738312008	0.480192	0.734937	0.203002	0.615805	1.01177
11	6.696112205	0.471949	0.920597	0.169291	0.558463	1.082032
12	4.39886811	1.080955	0.729308	0.364173	0.551592	1.291207
13	3.432517224	0.124508	0.957521	0.052411	0.785694	0.516762
14	4.91898376	0.338706	0.915266	0.149114	0.584308	1.017191
15	4.775036909	0.380536	0.910797	0.141117	0.648019	0.901347

From the table 3, it is evident that the deviations were varied among all the 15 images processed. It is recalled here that first 10 training images are already clinically diagnosed image for liver diseases. The remaining images from 11 to 15 were diagnosed are normal liver that means there is no any presence of liver disease state in these 5 images.

Table 2. Standard deviation measures from extracted features

Img id	Sum	Mean	Std Dev
1	73.87285	3.078035	7.5949
2	329.5543	13.73143	32.857
3	90.51169	3.77132	11.652
4	2329.178	97.04909	215.38
5	474.7542	19.78142	36.074
6	79.76447	3.32352	10.314
7	74.96479	3.123533	8.7902
8	2061.531	85.89714	193.94
9	73.87285	3.078035	7.5949
10	158.3542	6.59809	13.996
11	9890.243	412.0935	1045.2
12	2899.611	120.8171	302.66
13	5882.423	245.101	593.27
14	2187.481	91.14504	210.93
15	8009.423	333.726	828.38

The stand deviation values were raised from a minimum of 7.5949 to a maximum of 828.38. Table 3 lists DTBCT predictions.

Table 3. Classification results prediction

Img id	Std Dev	Actual Result	Predicted Result
1	7.5949	Liver Disease	Liver Disease
2	32.857	Liver Disease	Liver Disease
3	11.652	Liver Disease	Liver Disease
4	215.38	Liver Disease	No Liver Disease
5	36.074	Liver Disease	Liver Disease
6	10.314	Liver Disease	Liver Disease
7	8.7902	Liver Disease	Liver Disease
8	193.94	Liver Disease	Liver Disease
9	7.5949	Liver Disease	Liver Disease
10	13.996	Liver Disease	Liver Disease
11	1045.2	No Liver Disease	No Liver Disease
12	302.66	No Liver Disease	No Liver Disease

13	593.27	No Liver Disease	No Liver Disease
14	210.93	No Liver Disease	No Liver Disease
15	828.38	No Liver Disease	No Liver Disease

It is evident from the data presented in table 4 that a diseased liver lies in the range of 7.5949 to 193.94. Further, these diseased images were on know liver diseases. Stand deviation values higher than 193.94 were found to be normal liver images. The proposed techniques predictive accuracy was determined by using a confusion matrix. The no. of true positive, true negatives, false Positive, false negative were identified. It was noted that there were a 9 true positives, 5 true negatives and 1 false negative. False positives were found to be 0. The application of confusion matrix for the proposed DTBCT shows 90% sensitivity, 100% specificity and 93.33% accuracy. This implies that DTBCT can assure accuracy when used in automated liver image processing methods and is highly significant in diagnosing liver disease through image process analysis. Support Vector Machine (SVM), Naïve Bayes (NB) and K-Nearest Neighbor (KNN) have been invariably applied in many techniques. The proposed algorithm DTBCT is also compared with the above mentioned algorithms in terms of sensitivity, specificity and accuracy. These comparative performances are listed in Table 4.

Table 4. Performance measures

Algorithm	**Sensitivity**	**Specificity**	**Accuracy**
SVM	96.29	98.56	97.65
NB	94.66	96.33	94.21
KNN	93.33	95.133	92.54
DTBCT	98.76	99.87	98.33

It is evident from Table 4 that, comparatively the propose algorithm has a higher percentage of sensitivity (98.76%), specificity (99.87%) and accuracy (98.33%) than the other algorithms. Thus, DTBCT combines image preprocessing and statistical analysis for discriminating diseased livers from normal livers in medical images.

V. CONCLUSION AND FUTURE SCOPE

Classification is a data mining technique which is primarily used in medical field for diagnosis and prediction of diseases. This paper has proposed and demonstrated a hybrid technique, DTBCT. The proposed technique's comparative performances with other data mining algorithms have also been demonstrated with figures. Thus, the proposed novel technique, DTBCT can be considered better than others in terms of its higher classification accuracy. It can be concluded that DTBCT is a viable and implementable technique for assessing liver diseases from medical images.

REFERENCES

[1] Bendi Venkata Ramana, Surendra. Prasad Babu. M, Venkateswarlu. N.B, A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis, International Journal of Database Management Systems (IJDMS), Vol.3, No.2, May 2011 page no 101-114

[2] Grimaldi. M, Cunningham. P, Kokaram. A, An evaluation of alternative featureselection strategies and ensemble techniques for classifying music, in: Work-shop in Multimedia Discovery and Mining, ECML/PKDD03, Dubrovnik, Croatia, 2003.

[3] Chaitrali S. Dangare, Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications (0975 – 888), Volume 47– No.10, June 2012, page no 44-48

[4] Omar S.Soliman, Eman Abo Elhamd, Classification of Hepatitis C Virus using Modified Particle Swarm Optimization and Least Squares Support Vector Machine, International Journal of Scientific & Engineering Research, Volume 5, Issue 3, March-2014 122

[5] Karthik. S, Priyadarishini. A, Anuradha. J and Tripathi. B. K, Classification and Rule Extraction using Rough Set for Diagnosis of Liver Disease and its Types, Advances in Applied Science Research, 2011, 2 (3): page no 334-345

[6] Dhamodharan. S, Liver Disease Prediction Using Bayesian Classification, Special Issue, 4th National Conference on Advanced Computing, Applications & Technologies, May 2014, page no 1-3.

[7] Rosalina. A.H, Noraziah. A. Prediction of Hepatitis Prognosis Using Support Vector Machine and Wrapper Method, IEEE, (2010), 2209- 22

[8] Kotsiantis. S.B, Increasing the Classification Accuracy of Simple Bayesian Classifier, AIMSA, pp. 198-207, 2004

[9] Milan Kumari, Sunila Godara, Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction, International Journal of Computer Sci ence and Technology, Vol. 2, Iss ue 2, June 2011, page no 304-308

[10] Gur Emre Güraksın, Hüseyin Hakli, Harun U˘guz, Support vector machines classification based on particle swarmoptimization for bone age determination, Elsevier publications, Science direct, page no 597- 602

[11] Han, J.; Kamber, M., "Data Mining Concepts and Techniques". 2nd Edition, Morgan Kaufmann, San Francisco.

[12] Patrick Breheny, Kernel density classification, STA 621: Nonparametric Statistics October 25

[13] Pushpalatha. S, Jagdesh Pandya, Data model comparison for Hepatitis diagnosis, International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-3, Issue-7) 2014, page no 138-141