*Research Journal of Computer Systems and Engineering*

*ISSN: 2230-8571, 2230-8563*

*Volume 02 Issue 02 - 2021 (July to December)*

*Page 17:21*

# Liver Tissue Based Disease Detection Using Pre-Processing and Feature Extraction Techniques

**Dr. Nitin Namdeo Patil**
*Computer Science & Communication Engineering*
*Associate Professor, R. C. Patel Institute of Technology, Shirpur*
*ORCHID ID -0000-0001-8006-0860*
*er.nitinpatil@gmail.com*

**Article History**

Received: 15 July 2021
Revised: 20 September 2021
Accepted: 22 November 2021

***Abstract***

Liver disorder diseases one of the major diseases in the world, Liver is one of huge solid organs in human body; and is also considered a gland because, among its many functions, it makes and secretes bile. Liver is the second largest organ of human body. The diseases caused in liver are not detected and predicted at the earlier stage. So, this research gives the novel predictive analysis method for earlier liver disease prediction as application of data mining techniques. In this paper we have given the preprocessing and feature extraction techniques of liver tissues at the earlier symptoms of alcoholic liver disease (ALD). The novel method used for the early detection of hepatitis C in liver tissue. Initially on basis on the amount of white blood cells (WBC) in blood the liver dataset has been preprocessed using adaptive histogram discretization. After the preprocessing, liver tissue has to be extracted using Kernelized independent component analysis (KICA) that has maximum amount of WBC.

Keywords: Liver disorder diseases, alcoholic liver disease(ALD), hepatitis C, preprocessing, feature extraction, adaptive histogram discretization



## 1. INTRODUCTION:

Recent developments in health-related research are focusing on disease risk prediction. Different classification algorithms are applied in the research of risk prediction from patient health records. Different classification techniques of data mining are used to extract the risk from patient health records. To repeatedly make decisions in novel situations, classification procedures are used [1].When a person contracts the hepatitis virus, the infection might attack his liver, causing swelling and redness. [2] A liver condition called hepatitis that affects the majority of people across all age groups. The diagnosis of hepatitis is a significant difficulty for many hospitals and public health care agencies.Many people can be saved with an accurate diagnosis and timely disease prognosis [3]. Hepatitis A, often known as infectious hepatitis, is a disease that frequently affects youngsters. The hepatitis A virus causes the hepatitis A disease type to manifest (HAV). The stools, faeces, or poop of infected people contain this hepatitis A.The liver swells as a result of hepatitis D. The HEV virus is what causes hepatitis E. Hepatitis E can be contracted by consuming contaminated water. The liver swells as a result, but there is no long-term harm. [4]

## 2. LITERATURE REVIEW

This section gives some related study for liver disease prediction using data mining techniques.work [5]implemented non-incremental ANFIS for learning the classification models. Author in [6] proposed the analysis of predicting the response for treatment in patient with hepatitis C virus. The Interferon Alfa (IFN) and ribavirin (RBV) combination is the standard treatment for chronic hepatitis C (CHC), but it is quite expensive and has many adverse effects, therefore it frequently fails.Work [7] proposes a new machine learning approach to detect HCC using 165 patients. To forecast liver illness at an early stage, the author in [8] used various decision tree algorithms, including J48, LMT, Random Forest, Random Tree, REPTree, Decision Stump, and Hoeffding Tree. The segmentation of liver tumours using computed tomography (CT) images is the subject of work [9]. Author [10] suggests a level set model that combines edge energy and likelihood energy.

## 3. SYSTEM MODEL:

This paper discuss about thenovel method used for the early detection of hepatitis C in liver tissue. Initially on basis on the amount of white blood cells (WBC) in blood the liver dataset has been preprocessed using adaptive histogram discretization. After the preprocessing, liver tissue has to be extracted using Kernelized independent component analysis (KICA)that has maximum amount of WBC. The architecture is shown in figure 1:

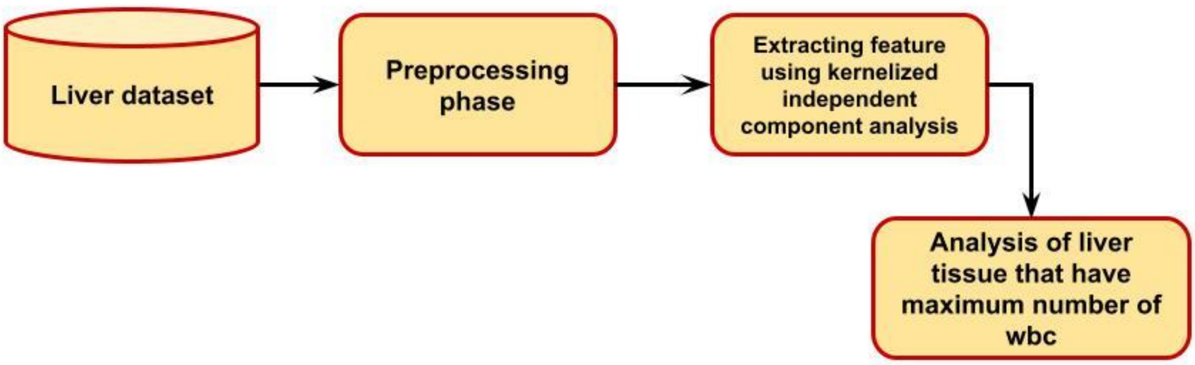


*Fig-1 proposed system architecture*

*3.1Preprocessing Phase:*

*3.1.1 Adaptive histogram discretization*

A typical strategy for enhancing the quality of natural and medical photographs is image enhancement. For human viewers, enhanced photos offer more information, and they also make it easier to conduct more image analysis. Furthermore, many image processing systems frequently demand images with contrast augmentation as input images. Spatial domain techniques and transform domain techniques are two broad categories for image enhancing methods.

*3.1.2 Histogram Analysis*

The histogram discretization approach is one of the most used improvement methods. Histogram discretization alters an image's dynamic range and contrast by reshaping the image's intensity histogram to a desired shape by eq. (1)- (3).

$$f_i(i_k) = \frac{n_k}{N} \tag{1}$$

$$F_k(i_k) = \sum_{j=0}^{k} f_i(i_j) \tag{2}$$

$$p_s(s)ds = p_r(r)dr \tag{3}$$

Kernelized independent component analysis (KICA)**:**

We define the +-correlation as the highest correlation between the random variables fi(x1) and fz, given an RKHS 3. (z2). where f and f2 fall between by eq. (4) (5)

$$\rho_{\mathcal{F}} = \max_{f_1,f_2\in\mathcal{F}} \operatorname{corr}\big(f_1(x_1), f_2(x_2)\big)$$
$$= \max_{f_1,f_2\in\mathcal{F}} \frac{\operatorname{cov}\big(f_1(x_1), f_2(x_2)\big)}{\big(\operatorname{var} f_1(x_1)\big)^{1/2}\big(\operatorname{var} f_2(x_2)\big)^{1/2}} \tag{4}$$

$$\langle f, C_{uv}(g)\rangle_{\mathcal{F}} = \mathbb{E}[f(u)g(v)] - \mathbb{E}[f(u)]\mathbb{E}[g(v)] \tag{5}$$

for all f ∈ F and g ∈ G. squared HS norm of covariance operator Cuv, denoted as HSIC, is then by eq. (6)

$$\|C_{uv}\|_{\mathrm{HS}}^2 = \mathbb{E}_{u,u',v,v'}\big[\psi(u,u')\hat{\psi}(v,v')\big] + \mathbb{E}_{u,u'}[\psi(u,u')]\mathbb{E}_{v,v'}\big[\hat{\psi}(v,v')\big] - 2\mathbb{E}_{u,v}\Big[\mathbb{E}_{u'}[\psi(u,u')]\mathbb{E}_{v'}\big[\hat{\psi}(v,v')\big]\Big]$$

$$\psi(a,b) = \hat{\psi}(a,b) := \phi(a-b) = \exp\left(-\frac{(a-b)^2}{2\lambda^2}\right) \tag{6}$$

contrast function over estimated signals Y ∈ R m×n is defined as eq. (7)

$$H: O(m) \to \mathbb{R}, H(X) := \sum_{1\le i<j\le m}^{m} \mathbb{E}_{k,l}\big[\phi\big(x_i^\top \bar{w}_{kl}\big)\phi\big(x_j^\top \bar{w}_{kl}\big)\big] + \mathbb{E}_{k,l}\big[\phi\big(x_i^\top \bar{w}_{kl}\big)\big]\mathbb{E}_{k,l}\big[\phi\big(x_j^\top \bar{w}_{kl}\big)\big] - 2\mathbb{E}_k\Big[\mathbb{E}_l\big[\phi\big(x_i^\top \bar{w}_{kl}\big)\big]\mathbb{E}_l\big[\phi\big(x_j^\top \bar{w}_{kl}\big)\big]\Big] \tag{7}$$

The difference between the kth and lth samples of the whitened observations is indicated in equation (7c), where X:= [x1,...,xm] O(m), wkl = wkwl R m, and the empirical expectation over all k are all represented (m). We define the set of all m m skew-symmetric matrices as so(m):= R mm| = , and we regard the orthogonal group O(m) as a m(m 1)/2 dimensional embedded submanifold of R mm. It should be noted that so(m) = R m(m1)/2 indicates that so(m) is isomorphic to R m(m1)/2. At point X O(m), the tangent space TXO(m) of O(m) is gi by eq. (8).

$$T_X O(m) := \{\Xi \in \mathbb{R}^{m\times m} \mid \Xi = X\Omega, \Omega \in \mathfrak{so}(m)\} \tag{8}$$

Endowing the manifold O is a common strategy for creating a Newton-type method for optimising a smooth function H: O(m) R. (m).

## 4. PERFORMANCE ANALYSIS:

The complete implementation of the suggested method is carried out using the Python tool, and the following specifications were taken into account for the experiment: a PC running Ubuntu, 4GB of RAM, and an Intel i3 processor.

| Parameters | RBV | LMT | LTDD_FET |
|---|---|---|---|
| **Accuracy** | 91 | 95 | 97 |
| **Precision** | 71 | 77 | 79 |
| **Recall** | 81 | 83 | 85 |
| **F1_Score** | 65 | 68 | 72 |
| **RMSE** | 55 | 58 | 59 |
| **MAP** | 41 | 43 | 45 |
| **AUC** | 32 | 33 | 36 |

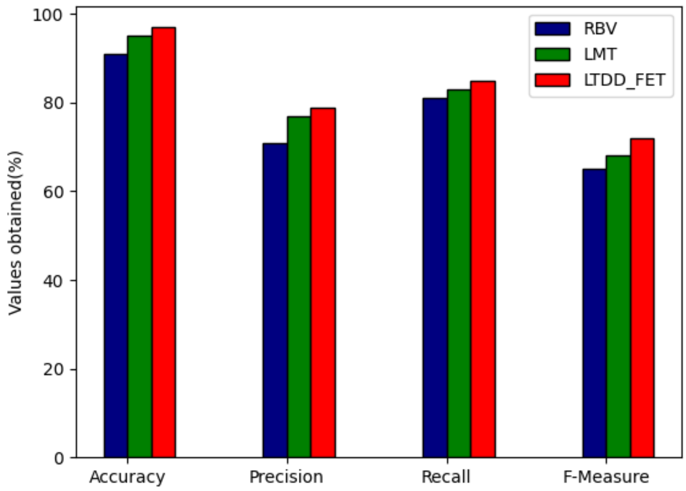


*Figure 2: Comparison of various parameters*

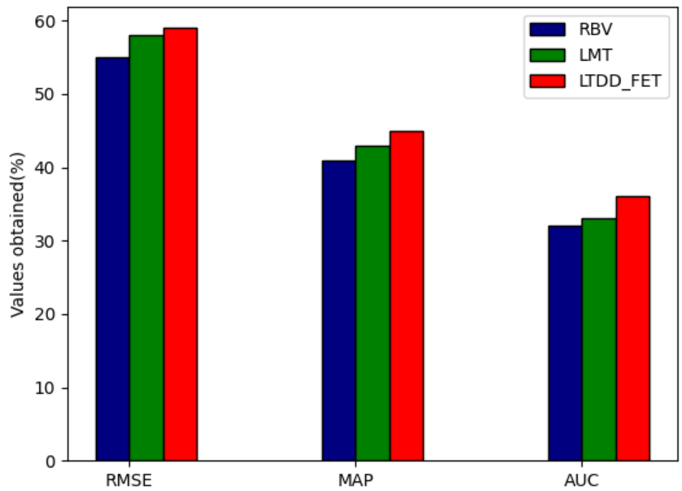


*Figure 3: Comparison*

The models are evaluated with accuracy, precision, recall, F-1 score, RMSE, MAP and AUC metrics as shown in table-2. This enables us to evaluate the performance of the models inside a dataset and the potency of the suggested methods for detecting liver illness. Figures 4 and 5 depict the confusion matrix produced by the four models utilising CT scans for the various numbers of epochs.Figs. 4 and 5 depict the accuracy and loss during the training and validation phases for various epoch counts. We can sum up the results of the many models that have already been researched in relation to the CT images based on the confusion matrices. Proposed technique attained accuracy of 97%, precision of 79%, recall of 85%, F-1 score of 72%, RMSE of 59%, MAP of 45% and AUC of 36% for Kaggle repository.

## 5. CONCLUSION:

Given that liver disease's symptoms are modest, diagnosing it can be particularly challenging. Nearly 38,170 of the 2,62,6,418 fatalities reported in the United States in 2014 were due to chronic liver disease. The value of computer-based prediction will only increase. Further, a lot of work is being done using the liver dataset has been preprocessed using adaptive histogram discretization. After the preprocessing, liver tissue has to be extracted using Kernelized independent component analysis (KICA) that has maximum amount of WBC. The majority of the PSO-based multi-objective feature selection algorithms now in use use binary tournament selection to choose the best mutation and gbest. If improved selection and mutation processes are utilised, there is potential to further limit the search space for better liver categorization accuracy. The proposed method in this paper can be used in the future with the heart dataset and disease classification to identify heart illnesses.

## REFERENCES:


[1] Rivenson, Y., Wang, H., Wei, Z., de Haan, K., Zhang, Y., Wu, Y., ...&Ozcan, A. (2019). Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nature biomedical engineering*, *3*(6), 466-477.

[2] Trivizakis, E., Manikis, G. C., Nikiforaki, K., Drevelegas, K., Constantinides, M., Drevelegas, A., & Marias, K. (2018). Extending 2-D convolutional neural networks to 3-D for advancing deep learning cancer classification with application to MRI liver tumor differentiation. *IEEE journal of biomedical and health informatics*, *23*(3), 923-930.

[3] MehrbakhshNilashia, HosseinAhmadi, "A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique", Journal of Infection and Public Health, Elsevier, 2017.

[4] Das, A., Acharya, U. R., Panda, S. S., &Sabut, S. (2019). Deep learning based liver cancer detection using watershed transform and Gaussian mixture model techniques. *Cognitive Systems Research*, *54*, 165-175.

[5] WojciechKsiazek, MoloudAbdar, "A novel machine learning approach for early detection of hepatocellular carcinoma patients", 2018 Elsevier.

[6] Nazmun Nahar1 and Ferdous Ara, "LIVER DISEASE PREDICTION BY USING DIFFERENT DECISION TREE TECHNIQUES", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.8, No.2, March 2018.

[7] Biswas, M., Kuppili, V., Edla, D. R., Suri, H. S., Saba, L., Marinhoe, R. T., ...&Suri, J. S. (2018). Symtosis: A liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm. *Computer methods and programs in biomedicine*, *155*, 165-177.

[8] Nancy.P, Sudha.V, "Analysis of feature Selection and Classification algorithms on Hepatitis Data", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 6, Issue 1, January 2017.

[9] Hsiao-Hsien Rau Anis Fuad Chien-Yeh Hsu Li-Ming, "Development of a Web-Based Liver Cancer Prediction Model for Type II Diabetes Patients by Using an Artificial Neural Network", Computer Methods and Programs in Biomedicine, 2015.

[10] SrinivasVadali, G. V. S. R. Deekshitulu, "Analysis of Liver Cancer Using Data Mining SVM Algorithm in MATLAB", Advances in Intelligent Systems and Computing 816, https://doi.org/10.1007/978-981-13-1592-3_12, 2018.