ISSN 2277-2685
IJESR/April-June. 2025/ Vol-15/Issue-2s/351-358
Chillimuntha Hemant Kumar *et. al.*, / International Journal of Engineering & Science Research

Diagnosis of Liver Diseases using Machine Learning

Chillimuntha Hemant Kumar
PG scholar, Department of MCA, DNR College, Bhimavaram, Andhra Pradesh.
V.Sarala
(Assistant Professor), Master of Computer Applications, DNR college, Bhimavaram, Andhra Pradesh.

Abstract: ***Computer aided diagnosis on medical domain such as breast cancer detection, brain tumor, liver disease, etc plays a major rile to go for smart hospitality. Liver diseases are responsible for more than 2.4% of annual deaths in India. Early detection of liver conditions is challenging due to the mildness of initial symptoms, which often become noticeable only at advanced stages. This paper seeks to enhance liver disease diagnosis by investigating two methods for identifying patient parameters. Due to liver diseases many peoples across the world lost their lives and its death rate can be reduced only by diagnosing disease on time but the main problem is LIVER will not show any symptoms for earlier damage. So in this paper is applying two methods to predict liver disease. Method1) in this method author is using INDIAN LIVER dataset to train various machine learning algorithms such as SVM, ANN and multilayer perceptron and this trained model will be applied on new patients TEST data to predict liver is normal or not but student ask us to implement Logistic Regression, Naïve Bayes and then compare its performance with SVM so we are using student suggested algorithms. Method2) in this method author is training ANN and CNN with gene MRNA images dataset and then training with CNN and ANN to predict whether liver disease inheriting in genes from ancestors. Student also asking to used liver images and then train with CNN and ANN; we are using LIVER ULTRA SOUND SCAN IMAGES***

Keywords—Machine Learning, Bioinformatics, Artificial Neural Networks, CNN, SVM, LR, NB, LIVER ULTRA SOUND SCAN IMAGES.

I. INTRODUCTION

The aim of study to improve early diagnosis, thereby enabling timely intervention and treatment, ultimately reducing morbidity and mortality associated with liver diseases. Machine learning has the capability to identify early indicators of diseases, including subtle ones that may not be easily discernible even to experienced medical professionals.

Main objective of the project is to get standard dataset with both image and value dataset.

1. Apply ML techniques like SVM , KNN ,RF for Value dataset
2. Apply ANN and CNN for Image dataset
3. Prepare GUI for easy understanding of the application using python

The diagnosis of liver diseases presents a significant challenge due to the subtle and often asymptomatic nature of early-stage conditions. Traditional diagnostic methods, while effective, frequently detect liver diseases only at advanced stages, leading to delayed treatment and poorer patient outcomes. The advent of machine learning offers a promising avenue for improving the accuracy and timeliness of liver disease diagnosis.

Machine analysing vast amounts of patient data and identifying complex patterns, machine learning algorithms can assist in the early detection of liver diseases. This innovative approach not only enhances diagnostic precision but also facilitates personalised treatment strategies, ultimately aiming to reduce the burden of liver diseases on healthcare systems and improve patient prognosis. This paper explores the potential of machine learning in diagnosing liver diseases, focusing on various techniques and their efficacy in identifying key patient parameters.

Symptoms of Abnormal Liver

- Liver size changes:
- Fat deposits:
- Scarring or fibrosis:
- Lesions and growths:
- Bile duct changes:
- Inflammation:

Diagnosing liver disease is particularly challenging due to the subtle nature of its early symptoms. Often, liver issues go undetected until they have progressed significantly, as the liver can continue functioning even when partially damaged. Early detection is crucial and can be life-saving, significantly extending a patient's life expectancy.

While early symptoms may elude even experienced medical practitioners, they can be identified using advanced diagnostic methods.

The significance of this study lies in its potential impact on both computer science and medical fields. This paper evaluates two computer-aided diagnostic methods. The first method uses a symptomatic approach, training an Artificial Neural Network (ANN) to analyse various patient parameters, including alanine aminotransferase, alkaline phosphatase, bilirubin levels, age, and aspartate aminotransferase. The ANN classifies patients based on whether they suffer from chronic liver disease or are healthy.

The second method employs a genetic approach, applying Artificial Neural Networks and Multi-Layer Perceptron's to microarray analysis for diagnosing liver diseases. This comparative study aims to determine the effectiveness of these two methods in improving early diagnosis and patient outcomes.

Diagnosing liver diseases presents a unique challenge due to the subtlety of symptoms in the early stages. Traditional diagnostic methods often fail to detect these diseases until they have advanced significantly, making early intervention difficult. However, machine learning offers a promising solution to this problem by providing tools that can analyse complex medical data and identify patterns indicative of liver diseases at an early stage.

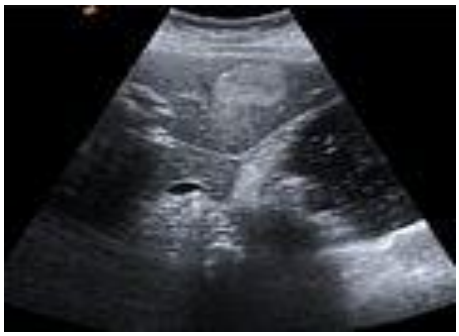


Fig.1.1 Liver Ultrasound Images

- A liver ultrasound can detect fat deposits (steatotic liver disease), inflammation and swelling (hepatitis), and the presence of scar tissue (fibrosis or cirrhosis).
- These conditions represent the three primary stages of chronic liver disease. Additionally, the scan can identify liver lesions, including unusual spots or growths on the liver.

Drawbacks of some pervious work are,

- Time complexity
- Low accuracy
- It works on less number of images

Machine Learning

Machine learning algorithms can process vast amounts of data, including patient demographics, laboratory test results, imaging studies, and genetic information. By training models on this data, machine learning can identify correlations and patterns that might not be evident to human practitioners. These models can then predict the likelihood of liver disease in new patients, facilitating earlier diagnosis and treatment.

Table 1.1 Difference between supervised learning and unsupervised learning.

Factors	Supervised learning	Unsupervised learning
Input	Known and labeled data	Unknown data
Complexity	Very complex	Less Complex
Number of classes	Known	Unknown
Accuracy	Accurate and reliable	Moderately Accurate and reliable

II. LITERATURE SURVEY

Liver disease, a leading cause of mortality in Taiwan, is challenging to detect early. Early diagnosis is crucial for effective treatment, making the development of accurate diagnostic models essential. This study introduces an intelligent diagnosis model using Classification and Case-Based Reasoning (CBR) and Regression Tree (CART) techniques to improve liver disease diagnosis accuracy. Data from 510 outpatients at a Taiwanese medical centre were used, with 340 cases for model development and 170 for comparative analysis. The model's first phase

employs CART to determine the presence of liver disease, achieving a 92.94% accuracy rate. The second phase uses CBR to identify specific liver disease types, with a 90.00% accuracy rate. The integration of CART and CBR provides a robust decision-support system for liver disease diagnosis and treatment, enhancing diagnostic accuracy and reducing errors. [1]

In modern cancer treatment, traditional methods of tumour classification based on microscopic examination can be subjective and inconsistent, even among experienced clinicians. Recently, DNA microarray technology has enabled the development of molecular cancer classifiers using gene expression data. Previous studies have focused on pairwise cancer classification, but extending this to multiple tumour types presents new computational and methodological complexities. This article describes a way to use computers and class-specific binary support vector machines (SVMs) to make predictions about 14 common types of cancer in adults. The methodology achieves an overall classification accuracy of 78%, significantly surpassing random classification expectations. In a substantial subset of samples (80%), the algorithm achieves an accuracy of 90%. This study demonstrates the feasibility of accurate multiclass cancer diagnosis using gene expression data and underscores the analytical challenges of applying such methods to biomedical research.[2]

Enhancing the accuracy of machine learning algorithms is crucial for developing effective computer-aided diagnosis (CADx) systems. Previous studies indicate that ensemble classification techniques can significantly improve base classifier performance. This study is mostly about putting together Rotation Forest (RF) ensembles with 30 different machine learning algorithms to see how well they classify datasets from existing literature about Parkinson's disease, diabetes, and heart disease. The study employs correlation-based feature selection (CFS) to reduce feature dimensions initially. We then individually assess each of the 30 algorithms' performance on the three datasets. We then build RF ensemble classifiers to improve the performance of these base classifiers on disease data. We conduct evaluation using leave-one-out validation, measuring classification accuracy (ACC), kappa error (KE), and area under the receiver operating characteristic curve (AUC). Results show that individual base classifiers achieve average accuracy of 72.15%, 77.52%, and 84.43% for the diabetes, heart disease, and Parkinson's disease datasets, respectively. In contrast, RF ensemble classifiers achieve average accuracies of 74.47%, 80.49%, and 87.13% for the same diseases. This study underscores the potential of RF as a promising ensemble algorithm to elevate the accuracy of diverse machine learning methods, paving the way for advanced CADx systems in medical diagnosis. [3]

This study focuses on applying Rotation Forest, a novel multiple classifier system (MCS), to classify microarray datasets in cancer research, marking its inaugural use in this domain. In the Rotation Forest framework, you need a linear transformation method to move data into new feature spaces for each classifier. This makes the ensemble more diverse and improves the accuracy of each classifier. We have used previous approaches such as Random Projections (RP), Non-Parametric Discriminant Analysis (NDA), and Principal Component Analysis (PCA), but this paper introduces Independent Component Analysis (ICA) due to its suitability in capturing microarray data properties. We demonstrate Rotation Forest's superior performance over other MCSs, such as bagging and boosting, using breast cancer and prostate datasets. Experimental findings also highlight ICA's ability to further enhance Rotation Forest's classification accuracy compared to traditional transformation methods. [4]

The prevalence of liver disease is on the rise, attributed to factors like alcohol abuse, exposure to harmful substances, consumption of contaminated food, pickles, and medications. To alleviate the strain on healthcare providers, automated classification tools are proposed. This study assesses several classification algorithms—Naïve Bayes, C4.5 decision tree, Back propagation Neural Network, and Support Vector Machines—using datasets of liver patient records. Evaluation criteria include Accuracy, Precision, Sensitivity, and Specificity, aiming to identify the most

effective algorithm for diagnosing liver conditions reliably. [5]

In India, liver diseases cause more than 2.4% of fatalities annually. Because of its modest symptoms, liver disease can be challenging to diagnose in its early stages. The symptoms frequently show up when it's too late. The goal of this work is to enhance the diagnosis of liver illnesses through the investigation of two identification techniques: patient parameters and genome expression. The study outlines drawbacks and goes over the computational methods that can be used to the previously outlined methodology. It suggests ways to increase these algorithms' effectiveness. [6]

Liver diseases are a significant global health concern and the 11th leading cause of death. Diagnosing these diseases through traditional methods like liver biopsies and MRI scans is labor-intensive and time-consuming. Early detection is crucial as symptoms often appear too late. This study evaluates the effectiveness of various machine learning algorithms—SVM, LR, NB, CNN and ANN in detecting liver disease. The algorithms were assessed using metrics such as accuracy, F1 score, precision, and recall. Results showed that CNN had the highest accuracy at 97.5%, indicating its potential for reliable early diagnosis using clinical data and imaging scans.[7]

III. PROPOSED METHOD

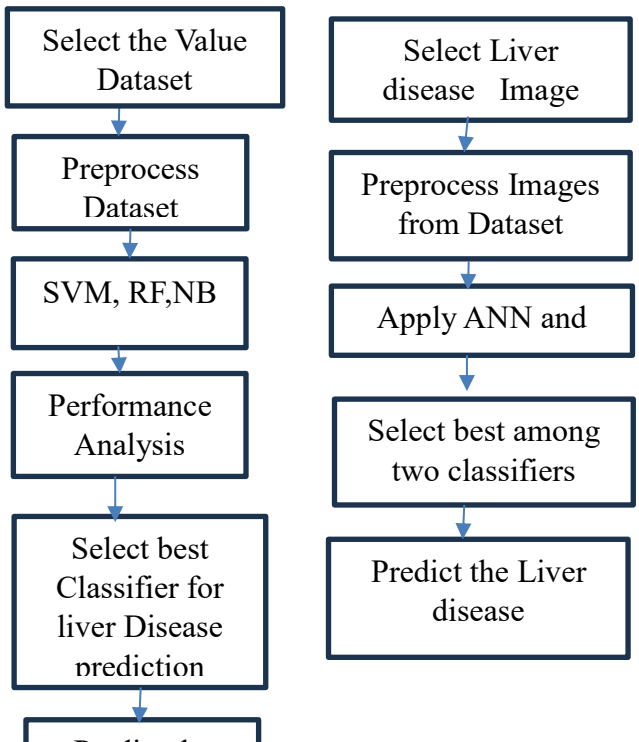


Fig.3.1 Proposed Work Block Diagram

Method 1. For Value Dataset:

In this method author is using INDIAN LIVER dataset to train various machine learning algorithms such as SVM, ANN and multilayer perceptron and this trained model will be applied on new patients TEST data to predict liver is normal or not but student ask us to implement Logistic Regression, Naïve Bayes and then compare its performance with SVM so we are using student suggested algorithms.

Methodology

1. **Data Pre-processing**:
   - **Feature Selection**: Relevant features from the dataset are selected.
   - **Normalization**: Data is normalized to improve model performance.
2. **Training the Models**:
   - **Support Vector Machine (SVM)**: Trained to classify liver function status.
   - **Artificial Neural Network (ANN)**: A neural network with multiple layers to learn complex patterns in the data.
   - **Multilayer Perceptron (MLP)**: A type of ANN specifically designed for supervised learning.
   - **Logistic Regression**: A simple yet effective algorithm for binary classification problems.
   - **Naïve Bayes**: A probabilistic classifier based on Bayes' theorem, assuming feature independence.
3. **Model Evaluation**:

- Each model is trained using the training portion of the Indian Liver dataset.
- The trained models are then tested on new patient data to predict liver functionality.

4. **Performance Metrics**:
   - The performance of each algorithm is evaluated using metrics such as Accuracy, Precision, Sensitivity (Recall), and Specificity.
   - Comparative analysis is conducted to determine which model performs best in predicting liver health.

Method 2. For Image Dataset:

In this method author is training ANN and CNN with gene MRNA images dataset and then training with CNN and ANN to predict whether liver disease inheriting in genes from ancestors. Student also asking to used liver images and then train with CNN and ANN but liver gene images are not available so we are using LIVER ULTRA SOUND SCAN IMAGES and below screen showing those images dataset.

Methodology:

1. **Data Acquisition**:

- **Gene mRNA Images**: Collected and used to train models for predicting hereditary liver disease.
- **Liver Ultrasound Scan Images**: Since gene images are unavailable, these are used to develop an alternative diagnostic model.

2. **Data Pre-processing**:

- **Gene mRNA Images**:
  - Image normalization to standardize the input data.
  - Augmentation techniques to increase the diversity of training data.
- **Liver Ultrasound Images**:
  - Gray scale conversion to simplify the image data.
  - Image resizing to a standard dimension suitable for CNN input.
  - Augmentation (e.g., rotation, scaling) to enhance model robustness.

3. **Model Training**:

- **Artificial Neural Network (ANN)**:
  - Designed with multiple layers to capture complex patterns in the gene mRNA images.
  - Trained to predict whether liver disease is inherited.
- **Convolutional Neural Network (CNN)**:
  - Architected to efficiently handle and extract features from image data.
  - Trained on both gene mRNA images and liver ultrasound images to predict liver disease status.

4. **Model Evaluation**:

  - Models are evaluated using a separate test dataset to ensure unbiased performance metrics.
  - Performance metrics include Accuracy, Precision, Sensitivity (Recall), and Specificity.

5. **Comparative Analysis**:

- **Gene mRNA Images**:
  - Performance of ANN and CNN on gene mRNA images is compared to determine which model better predicts hereditary liver disease.
- **Liver Ultrasound Images**:
  - Models trained on ultrasound images are compared to those trained on gene mRNA images to assess the effectiveness of using ultrasound data for liver disease diagnosis.

IV. RESULT

1. For Value dataset

The Indian Liver dataset contains records of patients with various liver conditions. Each

record includes several attributes, such as age, gender, and various biochemical markers, along with a label indicating whether the liver is functioning normally or abnormally. The study analyses two approaches for predicting chronic liver disease. Because liver disease symptoms are so mild, diagnosing it can be very challenging. Nearly 38,170 fatalities from chronic liver disease were reported in the United States in 2014 out of 2,626,418 total deaths.

Normal and Abnormal

To understand the distribution of liver conditions in the dataset, we categorize the records into "Normal" (healthy liver) and "Abnormal" (liver disease) based on the provided labels.

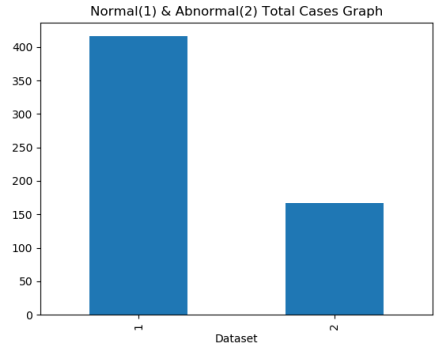


Fig.4.1 normal and abnormal total class graph

In above graph we can see Normal and Abnormal Count in INDIAN LIVER dataset

The dataset is loaded and cleaned to ensure that all entries are valid and relevant for analysis,

Fig.4.2 Dataset Record

In below screen, we can see the dataset record. For diagnosis of liver we are using various machines learning algorithm like SVM, LR and NB.

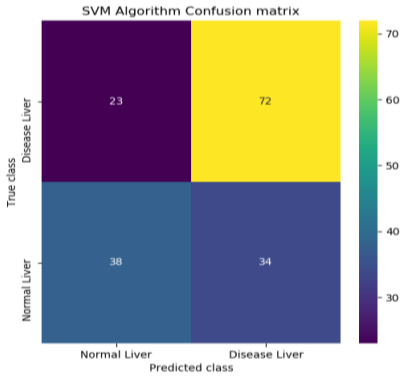


Fig.4.3 Confusion matrix for SVM Algorithm

In above screen we can see the Confusion matrix for SVM Algorithm,

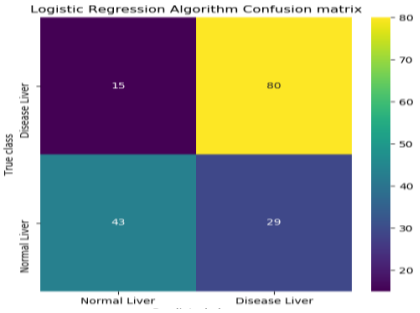


Fig.4.4 Confusion matrix for LR Algorithm

In above screen we can see the Confusion matrix for LR Algorithm,

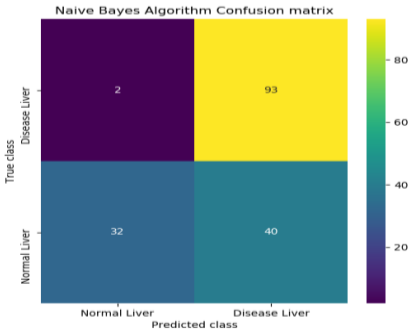


Fig.4.5 Confusion matrix for NB Algorithm

In above screen we can see the Confusion matrix for NB Algorithm,

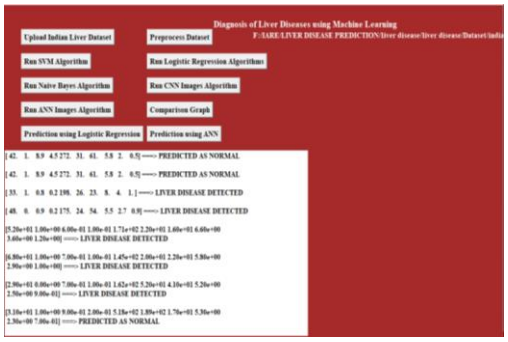


Fig.4.6 Prediction of image

In the above screen, we can see liver disease predicted as normal or abnormal.

2. For Image dataset

The dataset is loaded into the analysis environment, ensuring all images are correctly labelled and formatted for processing.

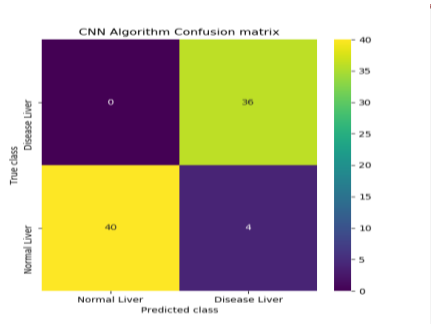


Fig.4.7 Confusion matrix for CNN Algorithm

On the above screen, we can see the confusion matrix for the CNN algorithm**,**

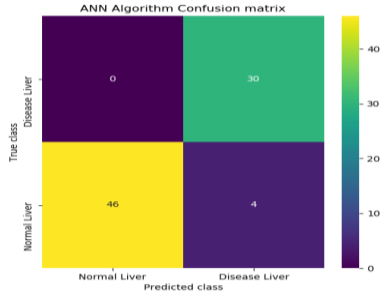


Fig.4.8 Confusion matrix for ANN Algorithm

On the above screen, we can see the confusion matrix for the ANN algorithm**.**
In the below screen, the liver disease image is predicted to be normal,

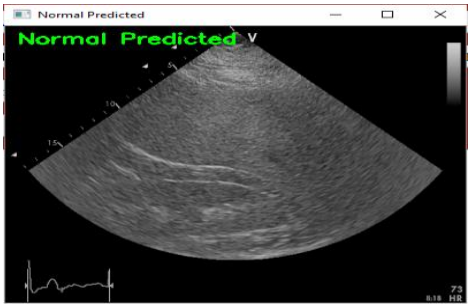


Fig.4.8 Disease predicted as a normal

In the below screen, the liver disease image is predicted to be abnormal,

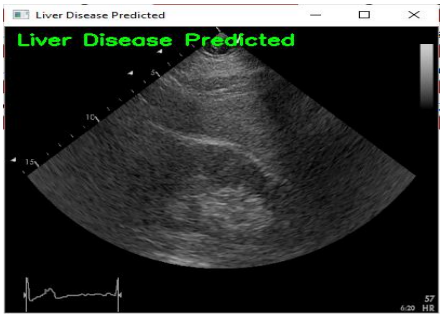
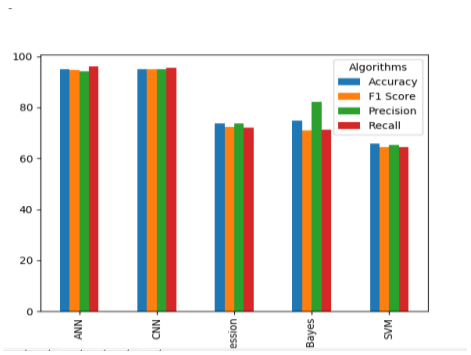


Fig.4.9 Liver Disease predicted



Fg.4.10 Performance Graph

We are using various machine algorithms like ANN, CNN, LR, NB, and SVM for the diagnosis of liver disease. In the above screen, we can see the performance graph for all algorithms**.**

V. CONCLUSION

The paper studies two approaches for predicting chronic liver disease, a condition challenging to diagnose due to its subtle symptoms. In 2014, chronic liver disease was responsible for approximately 38,170 out of 2,626,418 deaths in the United States. As computer-based prediction

methods become increasingly vital, this study examines two machine learning models that have the potential to enhance the accuracy and effectiveness of liver disease diagnosis. By using both image and value dataset we have analysed proposed model successfully. We are using various machine algorithms like ANN, CNN, LR, NB, and SVM for the diagnosis of liver disease. From these, we got better accuracy for CNN.

FUTURE SCOPE:

- Liver Disease typical treatment suggested by authorized doctors can be suggested or stages can be suggested.
- Android Application can be designed for same

REFERENCES

[1] Rong-Ho Lin, "An Intelligent Model for Liver Disease Diagnosis," Artificial Intelligence in Medicine, 2009"

[2] Ryan Rifkin, Sridhar Ramaswamy, Pablo Tamayo, Sayan Mukherjee, Chen-Hsiang Yeang, Micheal Angelo, Christine Ladd, Micheal Reich, Eva Latulippe, Jill P Merisov, Tomaso Poggio, William Gerald, Massimo Loda, Eric S Lander, Todd R Golub, "An Analytical Method For Multi-Class Molecular Cancer Classification ", 2003

[3] Akin Ozcivit and Arif Gulten "Classifier Ensemble Construction With Rotation Forest To Improve Medical Diagnosis Performance Of Machine Learning Algorithms",2011

[4] Kun-Hong Liu and De-Shuang Huang. "Cancer classification using Rotation forest", Computers in Biology and Medicine, 2008

[5] BendiVenkataRamana, Prof. M.Surendra Prasad Babu and Prof. N. B. Venkateswarlu, "A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis". International Journal of Engineering Reasearch and Development, 2012

[6] S. Sontakke, J. Lohokare and R. Dani, "Diagnosis of liver diseases using machine learning," *2017 International Conference on Emerging Trends & Innovation in ICT (ICEI)*, Pune, India, 2017, pp. 129-133, doi: 10.1109/ETIICT.2017.7977023.

[7] P. S. Harshini, K. Naresh, S. R. Pamulapati and A. Lavanya, "Diagnosis of Liver Diseases Using Machine Learning Algorithms and their Prediction Using Logistic Regression and ANN," *2023 3rd International Conference on Intelligent Technologies (CONIT)*, Hubli, India, 2023, pp. 1-6, doi: 10.1109/CONIT59222.2023.10205819.

[8] Beilharz TH, Preiss T: Translational profiling: the genome-wide measure of the nascent proteome. Brief Funct Genomic Proteomic, 2009.

[9] Gros F: From the messenger RNA saga to the transcriptome era. C R Biol. 2003, 326: 893-900.

[10] Shackel NA, Gorrell MD, McCaughan GW: Gene array analysis and the liver. Hepatology. 2002, 36: 1313-1325. 10.1053/jhep.2002.36950.

[11] Yano N, Habib NA, Fadden KJ, Yamashita H, Mitry R, Jauregui H, Kane A, Endoh M, Rifai A: Profiling the adult human liver transcriptome: analysis by cDNA array hybridization. J Hepatol. 2001, 35: 178-186. 10.1016/S0168-8278(01)00104-0.

[12] Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P, Nieselt_Struwe K, Muchmore E, Varki A, Ravid R, Doxiadis GM, Bontrop RE, Paabo S: Intra- and interspecific variation in primate gene expression patterns. Science. 2002, 296: 340-343. 10.1126/science.1068996.

[13] Nicholas A Shackel, Devanshi Seth, Paul S Haber, Mark D Gorrell and Geoffrey W McCaughan, "The Hepatic Transcriptome in human Liver Disease". 10.1186/1476-5926-5-6, BioMedCentral, 2006

[14] World Health Rankings, www.worldlifeexpectancy.com

[15] UCI Machine Learning Repository http://archive.ics.uci.edu/ml/datasets/ILPD+%28Indian+Liver+Patient+Dat aset%29