Abstract: Liver disease stands out as a prominent cause of mortality in India according to the World Health Organization, as well as on a global scale. The field of machine learning has emerged as a highly promising domain within the healthcare sector. Within the realm of medicinal industry, Computer Aided Diagnosis (CAD) represents a developing area of exploration. Extensive research has been conducted on the analysis of liver disease leveraging machine learning, highlighting advancements in the accuracy of disease detection and diagnosis. Through the utilization of machine learning, computers are equipped to assimilate knowledge and draw inferences from historical data. Consequently, computers can autonomously engage in self-learning processes, without the need for explicit programming by human developers. The present study offers an overview of machine learning methodologies employed in the context of liver disease, utilizing diverse datasets including liver function test data, Ultrasonic images (US), Computerized Tomography (CT) images, Magnetic Resonance Imaging (MRI).

Keywords: Deep learning, Liver Disease, Liver Function Test, Machine learning

I. INTRODUCTION

The contemporary era of computing has extended its scope to the extensive and effective utilization of Machine Learning Techniques (ML) in the field of bioinformatics. Due to the unpredictability present in medical datasets, extracting understandable insights poses a significant hurdle for medical practitioners. This obstacle has the potential to result in incorrect disease diagnoses, subsequently leading to inappropriate treatment regimens. It can be asserted that it would be advantageous for patients if healthcare professionals validate their evaluations utilizing decision-making systems. These systems are established by employing ML techniques, which adeptly analyze intricate and vague datasets.

Liver disease stands as a fatal ailment that has impacted one out of every five individuals in India. Projections suggest that by the year 2025, India might emerge as the "world capital" for cases of liver disease [1]. Various factors such as inherited conditions, contaminated food, viral, bacterial, or fungal infections in the liver cells, excessive lipid accumulation, and the overindulgence in alcohol or drugs commonly lead to liver disorders [2]. Detecting liver ailments early proves to be challenging as the organ can function normally even when partially impaired, intensifying the severity of the condition due to potential irreversible harm. Hence, prompt identification of liver disorders becomes imperative to facilitate timely treatment. The evaluation of intricate patient datasets during diagnosis prolongs the decision-making process for medical practitioners. To streamline this procedure and alleviate the burden, decision-making systems are devised using a multitude of intelligent methodologies.

This paper has contributed to the medical domain by showcasing an application of ML and Deep Learning (DL) techniques in the context of liver disorders, leveraging patients' demographic details, laboratory results, clinical information, as well as medical imagery encompassing US images, CT images, and MRI images. The rest of this paper is organized as follows: Section 2 presents the survey on various ML algorithm on Liver function test data and radiological data. Section 3 covers various DL approaches to liver diagnosis. In both Sections 2 and 3, information is listed in tabular form. Finally, conclusions are drawn in Section 4.

1 *Corresponding author: Research Scholar, Gujarat Technological University, Ahmedabad, Assistant Professor at Government Engineering College-Rajkot, India, Email: trupti.kodinariya@gmail.com
2 Professor and Head at Department of Information Technology, G H Patel College of Engineering and Technology, Vallabhdhiranagar, Gujarat, India

Copyright © JES 2024 on-line: journal.esrgroups.org
II. SURVEY BASED ON MACHINE LEARNING ALGORITHM

Machine learning algorithms play a crucial role in offering essential statistical information, real-time data, and sophisticated analytics. Numerous frameworks have been suggested by scholars for the purpose of diagnosing liver conditions. The majority of these frameworks rely on datasets comprising either patients’ demographic details, laboratory test results and clinical data, or images of the liver.

A. Classification of Liver Disease according to clinical, laboratory and patient data

Liver function tests are a category of blood tests that evaluate the health of the liver by measuring specific enzymes and proteins present in the blood, which can be utilized to determine liver functions and detect liver damage [3].

Chieh-Chen Wu et al. [4] employed four classification models namely Random Forest (RF), Naïve Bayes (NB), Artificial Neural Networks (ANN), and Logistic Regression (LR) to forecast the fatty liver disease. They utilized the Synthetic Minority Over-Sampling Technique (SMOTE) to create synthetic samples for the minority class. Moloud Abdar et al. [5] utilized rule-based classifiers Boosted C5.0 and CHi-squared Automatic Interaction Detection (CHAID) to classify patients with liver disease. Aman Singh et al. [6] employed correlation distance metric and a nearest rule-based K-Nearest Neighbor (KNN) approach to establish an efficient predictive model for liver disease. Mehrbakhsh Nilashi et al. [7] applied Non-linear Iterative Partial Least Squares for data dimensionality reduction, Self-Organizing Map (SOM) technique for clustering, Classification and Regression Trees (CART) for feature selection, and ensembles of Neuro-Fuzzy Inference System (ANFIS) for predicting hepatitis disease. C. Mahesh et al. [8] predicted hepatitis B disease by utilizing a Generalized Regression Neural Network (GRNN) based expert system. Sara Sweidan et al. [9] introduced a fuzzy fibrosis decision support (F2DS) system to classify different stages of fibrosis. Omar S. Soliman et al. [10] utilized Particle Swarm Optimization (PSO) algorithm and Least Squares Support Vector Machine (LS-SVM) to propose a hybrid classification model for HCV diagnosis, where Principle Component Analysis (PCA) algorithm was used for feature vector extraction. Modified-PSO Algorithm was employed to find the optimal values of LS-SVM parameters. Jagdeep Singh et al. [11] made predictions regarding liver disease based on a software engineering approach using various classification algorithms such as LR, Sequential minimal optimization (SMO), RF, NB, J48, and KNN on data from liver patients. El-Sappagh et al. [12] conducted a performance comparison between ANFIS and fuzzy analytical hierarchy process for fibrosis diagnosis.

The above works are summarized in Table 1.

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Methods</th>
<th>Purpose</th>
<th>Result</th>
</tr>
</thead>
</table>
Classifier: GRNN based expert system | Diagnosis Hepatitis B Disease | Accuracy: 98.04% |
Pre-processing: Dimension reduction using PCA.  
Classifier: Least Squares Support Vector with Modified-PSO Algorithm was used to search for the optimal values of LS-SVM parameters | Diagnosis Hepatitis C Disease | LS-SVM: 96.12%  
Proposed Method: 98.86% |
Classifier: KNN with correlation distance metric | Classification of liver disorders as sick and healthy | Accuracy: 96.74%  
Sensitivity: 95.81%  
Specificity: 97.12% |
<table>
<thead>
<tr>
<th>Name</th>
<th>Data set</th>
<th>Classifier</th>
<th>Classification of liver disorders as sick and healthy</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moloud Abdar et al [5] (2017)</td>
<td>Data set: ILPD Data Set</td>
<td>Booster C5.0 and CHAID</td>
<td>Classification of liver disorders as sick and healthy</td>
<td>Boosted C5.0: 92.33%</td>
<td>CHAID: 76.14%</td>
</tr>
<tr>
<td>El-Sappagh et al [12] (2018)</td>
<td>Data set: 119 patient’s Demographic, laboratory test and clinical data are collected from Liver Institute, Mansoura University, Egypt</td>
<td>Boosted C5.0 and CHAID</td>
<td>Prediction of liver fibrosis stages in patients</td>
<td>FAHP: 93.33%</td>
<td>ANFIS: 92.46%</td>
</tr>
<tr>
<td>Chieh-Chen Wu et al [4] (2019)</td>
<td>Data set: Demographic and clinical characteristics of overall 577 patient are collected from New Taipei City Municipal Hospital Banqiao Branch.</td>
<td>Boosted C5.0 and CHAID</td>
<td>Diagnosis fatty liver disease</td>
<td>RF: 87.48%</td>
<td>NB: 82.65%</td>
</tr>
<tr>
<td>Sara Sweidan et al [9] (2019)</td>
<td>Data set: 119 patient’s Demographic, laboratory test and clinical data are collected from Liver Institute, Mansoura University, Egypt</td>
<td>Boosted C5.0 and CHAID</td>
<td>Diagnosis fatty liver disease</td>
<td>ANFIS: 93.02%</td>
<td>SVM: 81.17%</td>
</tr>
<tr>
<td>Mehrbakhsh Nilashi et al. [7]</td>
<td>Data set: UCI Hepatitis data set</td>
<td>ANFIS ensemble</td>
<td>Diagnosis Hepatitis Disease</td>
<td>Accuracy: 93.02%</td>
<td></td>
</tr>
<tr>
<td>Jagdeep Singha et al. [11] (2020)</td>
<td>Data set: ILPD Data Set</td>
<td>Correlation-based Feature Selection Subset Evaluator was used as Feature evaluator and Greedy Stepwise used as search method</td>
<td>Classification of liver disorders as sick and healthy</td>
<td>LR: 74.36%</td>
<td>NB: 55.9%</td>
</tr>
</tbody>
</table>

**Notes:**
- **PPV:** Positive Predictive Value
- **NPV:** Negative Predictive Value
- **FAHP:** Fuzzy AHP
- **ANFIS:** Adaptive Neuro-Fuzzy Inference System
- **RF:** Random Forest
- **NB:** Naive Bayes
- **ANN:** Artificial Neural Network
- **LR:** Logistic Regression
B. Classification of Liver Diseases Using Radiography Images and Machine Learning

Computer Aided Diagnosis has emerged as an expansive domain within the realm of medical research in recent years and has assumed a pivotal role in the field of medical diagnosis [13][14].

Andreia Andrade et al. [15] introduce a semi-automatic method for classifying steatotic (fatty) liver tissues utilizing B-scan ultrasound images, ANN and SVM. Alivar A. et al. [16] proposed a CAD system for categorizing normal, fatty, and cirrhotic liver ultrasound images by employing spatial domain and transform domain characteristics with Regions of Interest (ROI) manually delineated. Owjimehr M [17] suggested an automated ROI detection system for the characterization of normal, fatty, and heterogeneous liver tissues through textural analysis of liver ultrasound images. Aggarwal, K et al. [18] detected cirrhosis of the liver in ultrasound images by utilizing modified local binary pattern and OTSU methods, with SVM and KNN as classifiers. Xu S. S. D et al. [19] presented a CAD system designed to differentiate between hepatocellular carcinoma (HCC) and liver abscess based on ultrasound image texture features, employing a SVM classifier with ROIs manually extracted by a radiologist. H. Alahmera et al. [20] developed an automated CAD system for the classification of liver lesions as Benign or Malignant. Wang et al. [21] categorized three types of hepatic tissue (Normal, HCC, and Hemangioma) using multiclass SVMs constructed through one-against-all and one-against-one strategies. Kayaatli O et al. [22] determined fibrosis stage utilizing texture properties from computed tomography images of the liver with manual ROI extraction. K. Mala et al. [23] classified liver tumors as hepatocellular carcinoma, cholangiocarcinoma, hepatocellular adenoma, and hemangioma using Probabilistic Neural Networks (PNN) as a classifier, along with adaptive threshold decision-making based on intensity information and morphological processing for liver segmentation, with tumor region delineation accomplished through Fuzzy C Means Clustering. Liu H et al. [24] categorized cirrhosis into normal, early, middle, and advanced stages through the utilization of multi-sequence MRIs with manual ROI extraction and feature extraction via RF, applying Nearest Neighbors for classification of cirrhosis stages. Jansen MJA et al. [25] distinguished liver lesions as Adenoma, Cyst, Hemangioma, HCC, and Metastasis using MRI images and a Randomized Tree Classifier. The above works are summarized in table 2.

Table 2: Performance of liver disease diagnosis based on machine learning using radiology images.

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Methods</th>
<th>Purpose</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>K. mala et al. [23] (2008)</td>
<td>Data set: Private CT image Data set ROI: Automatically extracted using adaptive threshold technique, morphological processing and Fuzzy C-means clustering Feature extraction: Biorthogonal wavelet based texture features Feature selection: Sequential backward selection Classifier: Probabilistic Neural Network (PNN), Learning Vector Quantization (LVQ)</td>
<td>Classify liver tumor as hepatocellular carcinoma (HCC), cholangiocarcinoma (CC), hemangioma (H), hepatoadenoma (HA)</td>
<td>PNN: 90.2% LVQ: 83.5%</td>
</tr>
<tr>
<td>Wang et al. [21] (2009)</td>
<td>Data set: Private CT image Data set ROI-Manually extracted with 60 x 60 pixels from each image Feature Extraction: FOS, Spatial Gray Level Dependence Matrix (SGLDM), GLRLM, and Gray Level Difference Matrix (GLDM) Classifier: Multiclass SVM</td>
<td>Classify hepatic tissue as Normal, HCC, and Hemangioma</td>
<td>SVM+OAA: 94.44% SVM+OAO: 97.78%</td>
</tr>
<tr>
<td>Authors</td>
<td>Data set:</td>
<td>ROI:</td>
<td>Feature extraction:</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------------------</td>
<td>-----------------------------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>Kayaalti O et al. [22] (2012)</td>
<td>Private CT image Data set</td>
<td>Manually extracted with 32 x 32 pixels</td>
<td>GLCM, DWT and DFT</td>
</tr>
<tr>
<td>Andrade A, et al. [15] (2012)</td>
<td>Private US B-Mode Data set collected from the Coimbra University Hospital (177 Images)</td>
<td>Manually extracted with 50 x 50 pixels from each image.</td>
<td>Nine First order features: Mean, Variance, Standard deviation, Skewness, Kurtosis, Median, Entropy, Mode and Range, GLRLM, GLCM, law’s texture energy and fractal dimension</td>
</tr>
<tr>
<td>Alaleh Alivar et al [16] (2014)</td>
<td>Private US B-Mode Data set</td>
<td>Extracted Manually by radiologists (64 x 64)</td>
<td>GLCM, energy and energy deviation of 2-D WPT and 2-D Gabor filter banks sub-images, these feature are combined and normalized</td>
</tr>
</tbody>
</table>
*Date set: Private CT images*
*ROI: Automatically FCM clustering algorithm on histogram of CT image is applied which split image into three parts*
*Feature extraction: Shape feature extracted from entire segmented ROI; Intensity and texture feature are extracted from multiple ROI*
*Classifier: SVM*

Classify liver tumor as Benign and Malignant

Proposed Method: 98.3%
Single ROI: 95.1%

*Data set: Private MRI Image dataset*
*ROI: Extracted manually*
*Feature extraction: Contrast curve features, Texture features are extracted*
*Feature selection: ANOVA F-Score*
*Classifier: Randomized tree classifier*

Classify liver lesion as Adenoma, Cyst, Hemangioma, HCC, Metastasis

Accuracy: 77%
Sensitivity: 79%
Specificity: 77%

### III. SURVEY BASED ON DEEP LEARNING ALGORITHM

The primary distinguishing factor of deep learning methodologies is their emphasis on feature acquisition: the autonomous acquisition of data representations. This constitutes the fundamental contrast between deep learning strategies and more conventional machine learning techniques. The identification of features and the execution of a task become intertwined into a singular issue, thereby enhancing both aspects concurrently throughout the training procedure.

Hassan T. M. et al. [26] introduced a feature depiction utilizing a stacked sparse auto-encoder to classify various focal liver ailments by selecting the most probable outcomes for each class. Das A. et al. [27] proposed a CAD system to differentiate between liver cancer types such as HEM, HCC, and metastatic carcinoma (MET) by employing a deep neural network (DNN) as the classifier. Liver segmentation was achieved through a marker-controlled watershed segmentation technique, followed by the segmentation of cancer-affected lesion using the Gaussian mixture model algorithm. D Santhosh Reddy et al. [28] suggested a CAD framework utilizing convolutional neural networks (CNN). Frid-Adar M et al. [29] outlined approaches for producing synthetic medical images through DL Generative Adversarial Networks (GANs) derived from CT scans. The GANs' output is utilized as input for the data augmentation phase of the CNN to categorize liver lesions. Meng, D et al. [30] classified liver fibrosis stages utilizing FCNet as the classifier and Ultrasonic images. They conducted pre-training on VGGnet to extract features that are subsequently fed as input to FCNet. Choi K. J et al. [31] proposed a Deep Learning System (DLS) for liver fibrosis staging utilizing CT images. They implemented CNN for image segmentation and categorization. Ben-Cohen et al. [32] introduced a fully convolutional network for recognizing liver and lesions. Ghoniem, R. M. [33] suggested a bio-inspired DL strategy to enhance the predictive outcomes of liver cancer. Liver and lesion segmentation is accomplished using SegNet and UNet respectively in this methodology. The optimization of deep learning network hyper-parameters is performed through artificial bee colony optimization (ABC).

The above works are summarized in table 3.

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Methods</th>
<th>Purpose</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
<td>Feature extraction:</td>
<td>Classifier:</td>
<td>Proposed Method:</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------</td>
<td>-------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Meng, D et al. [30] (2017)</td>
<td>stacked sparse auto-encode</td>
<td>a softmax layer</td>
<td>fine-tuning: 97.2%</td>
</tr>
<tr>
<td>D Santhosh Reddy et al. [28] (2018)</td>
<td>Pre-trained VGG16 with transfer learning</td>
<td>manually extracted</td>
<td>CNN: 84.3%</td>
</tr>
<tr>
<td>Ben-Cohen et al. [32] (2016)</td>
<td>Data augmentation</td>
<td>manually segmented using FCN-8s and FCN-4s receptivity</td>
<td>Liver Segmentation and Lesions Detection</td>
</tr>
<tr>
<td>Choi K. J et al. [31] (2018)</td>
<td>Data augmentation</td>
<td>automatically segmented using CNN</td>
<td></td>
</tr>
<tr>
<td>Frid-Adar M et al. [29] (2018)</td>
<td>Synthesis of liver lesions from CT images using Generative Adversarial networks (GANs) variants DCGAN and ACGAN</td>
<td>manually segmented</td>
<td>Without augmentation: 57%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
IV. CONCLUSIONS

The Computerized Diagnosis approach relies on cost-effective standard blood tests, patient physical assessments, and demographic information, primarily utilized for distinguishing between healthy and ill patients, diagnosis of Hepatitis and fatty liver Disease. Embedded methods like Random Forest have shown superior outcomes compared to other classifiers within this approach. Studies have revealed that age, sex, TG, ALT, GOT, GPT, AST/ALT ratio, total bilirubin, platelet count (PLT) are key variables impacting liver disease diagnosis, with ALT, AST, and PLT being the most frequently referenced in medical literature, particularly concerning liver ailments. The rise in levels of biomarkers such as serum glutamic oxaloacetic transaminase (SGOT), serum glutamic pyruvic transaminase (SGPT) in the bloodstream is predominantly linked to liver damage. However, crucial predictors for Non-Alcoholic Fatty Liver Disease (NAFLD) include BMI, WHR, triglycerides, glucose, systolic blood pressure (SBP), and alanine aminotransferase.

In CAD systems for liver imaging, there exist four primary stages: Data pre-processing, segmentation of the liver and/or lesions, feature extraction, and classification. Various methods are available for segmenting the liver region, most of which are manual while a few are semi-automatic or automatic. When segmenting liver lesions, utilizing multiple Regions of Interest (ROI) (inside, boundary, and surrounding lesion) yields better outcomes compared to a single ROI approach due to the distinct characteristics of benign and malignant lesions across different regions.

Numerous techniques have been proposed for feature extraction from liver tissue. The liver surface can be viewed as a texture, prompting the selection of various texture models for analyzing liver images. Key subjective parameters considered include echogenicity (brightness) in US images, contrast, homogeneity, granularity, and smoothness of the liver surface. Common features employed for liver disease diagnosis encompass first-order and second-order statistical features derived from texture analysis of images. GLCM, GLRLM, law’s texture energy, fractal dimension, and WPT are prevalent feature extraction techniques in the literature. Ultrasonic imaging is predominantly utilized for diagnosing fatty liver. Various medical imaging modalities used for early detection and diagnosis of liver tumors include US, CT scan, MRI Imaging, and Angiography. Although Ultrasonography is
non-invasive and radiation-free, its results are operator-dependent and lack specificity in distinguishing benign and malignant tumors based on echogenicity. MRI Imaging provides high tissue contrast and multiplanar capabilities for accurate tumor detection and differentiation, albeit being costly compared to CT. Researchers have shown that using GLCM texture features for classifying cirrhosis stages based on MRI images does not yield satisfactory results, suggesting the use of compressed patch vectors as alternative texture features.

Researchers have proposed various classifiers in CAD systems for liver disease diagnosis. In cases of nonlinear relationships between input and output features, ANN and SVM demonstrate superior performance compared to other methods. The automation of feature extraction in deep learning algorithms has recently emerged as a prominent research area. Nevertheless, a significant challenge in deep learning pertains to the substantial computational requirements involved. This is primarily due to the pixel-level operations in image processing, which necessitate high-speed and high-capacity computing resources. Some studies have utilized pre-trained models as a strategy to address this issue. To mitigate the problem of limited data availability, techniques such as Data Augmentation (e.g., flipping, rotating, and shearing) can be implemented as a preprocessing step. Furthermore, research efforts have integrated bio-inspired methodologies like PSO and ABC for optimizing hyper-parameters in DL networks, aiming to enhance overall performance.

REFERENCES

1844

1835–1844


