IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Multi-Modal Prediction Of Liver Disease: Integrating Gene Expression And Ultrasound Imaging

[1]Nikkala Vasanthi, Student in Dept. Of Master of Computer Applications, at Miracle Educational Society Group of Institutions

[2]Dr. A. Arjuna Rao, Professor at Miracle Educational Society Group of Institutions

[3]Katuri Swamy, Associate Professor at Miracle Educational Society Group of Institutions

ABSTRACT

Liver disorders are one of the major health concerns on the globe, and the pesky issue is that they often aggravate without being diagnosed. In this proposal, a model which utilizes machine learning techniques is developed aimed at predicting liver health status. Following the Indian Liver Dataset, SVM, Logistic Regression, and Naïve Bayes algorithms were used for classification and an accuracy of 73% was obtained. For disease classification, convolutional and artificial neural network architectures were trained on liver ultrasound images, the CNN accuracy reached 97%. Data preprocessing features missing value replacement and imbalance resolution by oversampling. Hybridized mechanisms illustrated comparably higher performance for predicting images which can be employed as tools for diagnosis. Thus, in the case of liver disease, the integration of structured dataset with imaging should be stressed.

Keywords: CNN, ANN, SVM

INTRODUCTION:

While it is safe to state that Machine learning is one of the more exciting fields in computer science, it is fair to say that there are numerous applications of it in the real world most notably health care. Algorithms are trained through datasets or multi-dimensional data that enables the classification, prediction and even doing the task themselves. If a machine is not programmed to make a certain decision, that decision would be installed by the machine learning algorithms trained on that particular dataset. An aspect that relies heavily on classification, and even improves through it, is ML and AI which is extremely important in the medical world, one wrong definition can lead to severe harm. Classification of medical data sets also comes with its own set of challenges, as different datasets yield different models with different accuracies.

The focus on improving health is global and is dictated by the economic argument that ‘healthier populations are assets to society’. Amplifying human intervention through ML holds the potential of improved disease diagnostics, effective treatment plans, and the development of new drugs. ML spans from improving the clinical aspect of the drug development process to” smart” EHR systems, which together improve operational processes within healthcare systems. I am confident that with this advantage, we would be able to create an ecosystem of healthcare that is more efficient, reliable and more importantly more focused on patients.

GAP IDENTIFIED BASED ON LITERATURE SURVEY:

Until recently, ML was scarcely used for identifying and predicting liver diseases, and it focused on multi-modal datasets aimed at structuring clinical data and medical imaging. Existing research that leveraged the Indian Liver Dataset employed low-level ML techniques including SVM and ANN which led to subpar accuracy due to greatly skewed datasets as well as poorly executed feature engineering. In this sense, the majority of the techniques utilizing image-based diagnostics favored becoming steeped in the use of either CNNs or ANNs, however, seldom undertaking these contrasting approaches to benchmarks. There are few papers that consider combining these two types of data sets touching the issue of improving the reliability of the diagnostics.

Key gaps include:

1. There are no sophisticated algorithms for preparing the analysis on unbalanced datasets.
2. Little consideration on complex approaches to structured and image data.
3. A rather bare comparative discussion concerning the ML approaches with respect to metrics like precision and recall is presented.
4. There has been rather a little importance given to the aspect of explainability in the context of ML in regards to Health Care.

PROBLEM STATEMENT:

Lack of early damage symptoms makes it impossible to use the liver disease diagnosis model effectively. It is important to develop a robust non-invasive diagnostic such as combined structured clinical data and imaging.

Key Challenges:

• Data Imbalance: Clinical datasets often have a skewed distribution which affects the performance of the models.

• Feature Engineering: The process of selecting useful subsets of relevant features from structured and image datasets.

• Model accuracy: Must be consolidated to allow accurate predictions using various metrics.

• Data Availability: The problem of acquiring plausible annotated medical image datasets is a major bottleneck.

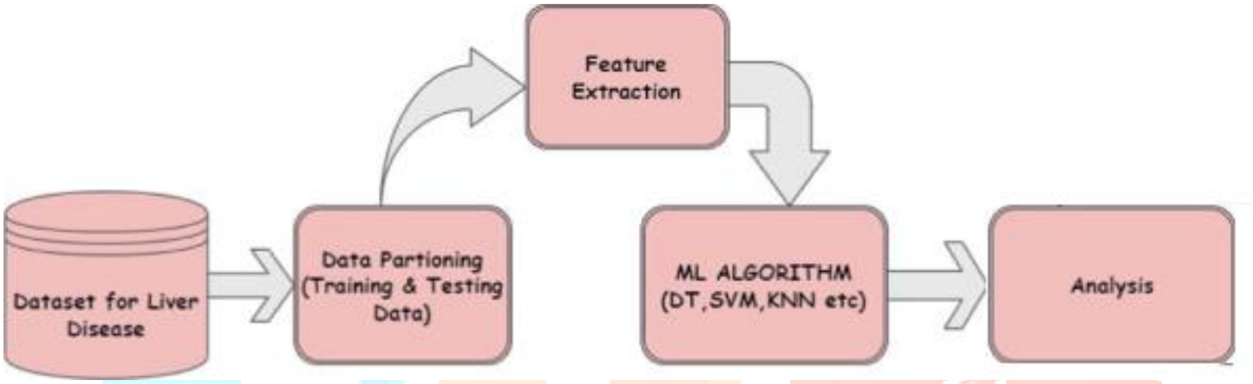
• Integration: Efficiently combining structured predictions and those from image-based models.

PROPOSED METHOD:

I would like to draw your attention to the proposed Modern Machine Learning Framework for the task of detecting liver diseases. We have utilized over sampling and interpolation techniques to pre-process the Indian Liver Dataset. It was identified that there were some missing values in the dataset. Oversampling was carried out to remedy this problem, and in this way, a balance was created in the data. In terms of structured data, classification is performed based on performance metrics comparison with Logistic Regression, SVM and

Naïve Bayes. Convolutional Neural Network as well as Artificial Neural Network models are employed for analysis of ultrasound images of the liver in order to detect the disease. One of the basic problems to be addressed in elaborating medical imaging related systems is the choice of set parameters like image size, augmentation, normalization and so on that would ensure maximum accuracy for differentiating normal imaging features from diseased ones. The evaluation of the models is performed on pre-defined precision, accuracy, recall and F1 measures. Other than this, a relative comparison between the performance of structured models and image-based models was also of our interest. The reason why the hybrid model was proposed is to combine the strengths of those individual models so that a single useful and reliable prediction tool was achieved, which could be used to supplement diagnosis.

ARCHITECTURE:


Feature Extraction
Dataset for Liver Disease
Data Partioning (Training & Testing Data)
ML ALGORITHM (DT,SVM,KNN etc)
Analysis

DATASET:

The Indian Liver Dataset consists of age, gender and some biochemical markers. The dataset's 583 records with a skewed distribution of cases consist of 167 normal observations and 450 observations with some disease pathologies. It was also noted that some of these values were missing, which were filled in, while oversampling was done to balance the records and prepare them for effective model training. The data set contains images of a patient that has ultrasound scans, which has images categorised as normal and diseased ones. Now the images are prepared for usage, They have been formatted for CNN and ANN training. This dataset aids the visual characteristic learning of liver diseases.

METHODOLOGY:

Indian Liver Dataset to be loaded.

Fill the missing values to zero.

Implement the oversampling algorithms to normalize the dataset.

Pre-processing of clinical data

Assign numeric values for categorical features like gender

Standardization of the data for consistent feature scaling

Divide dataset into 80% working and 20% testing datasets

ML model training on clinical data:

Deploy Models of Logistic Regression, Naive Bayes and SVM.

Assess models on accuracy, precision, recall and F1 measure.

Perform analysis of diagrams to evaluate the results.

Image dataset preparation

Organize folders of liver ultrasound pictures that contain images of normal and diseased livers.

Alter images to have the equivalent size in order to develop a CNN and ANN.

Apply some image rotation, shifting, warping etc. in order to increase the size of the dataset.

Model training on images:

Build CNN that consists of convolutional and pooling layers for feature extraction.

Form ANN with fully connected layers to the images which have been processed.

Assess image models using confusion matrices and accuracies.

Hybrid integration model:

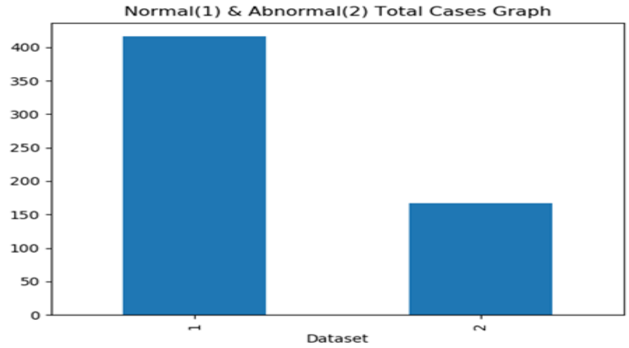
Integrate structured data and prediction of the image model.

Use ensemble techniques to make the final prediction.

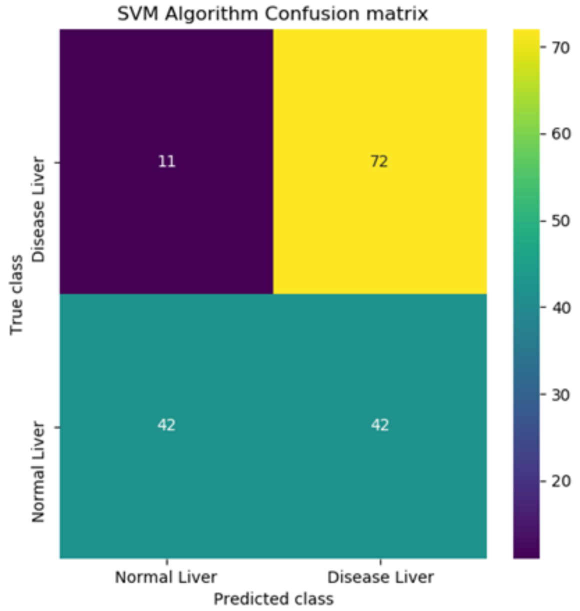
Visualization and comparison:

Draw comparison graphs for all algorithms across the different metrics.

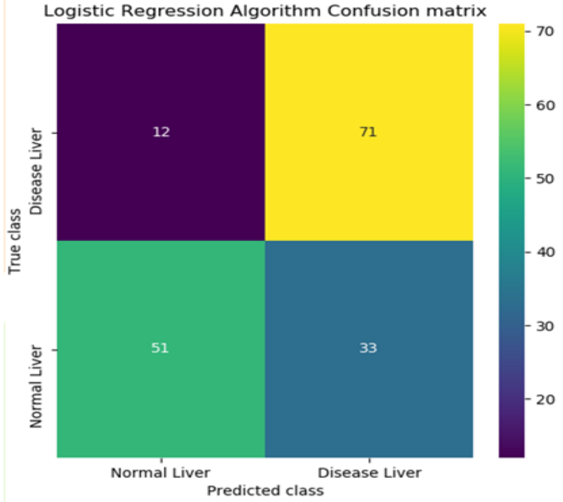
RESULTS:



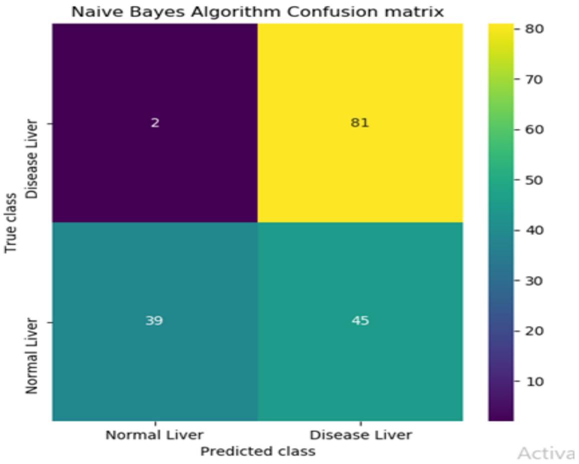
Number of records for normal as 1 and disease as 2



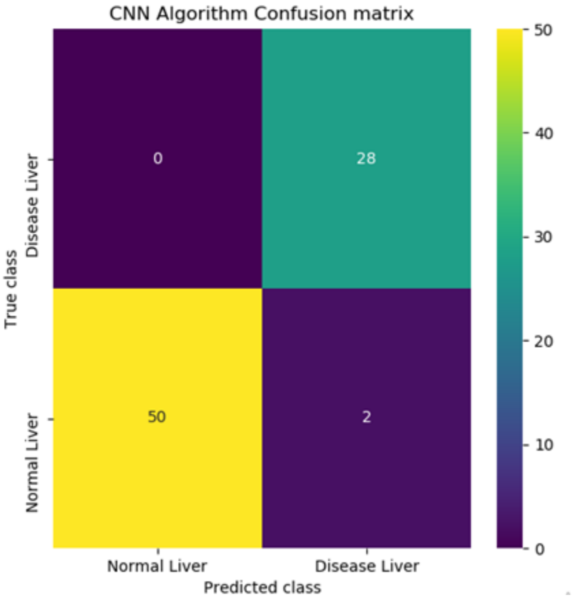
SVM is trained and we got its accuracy as 68%



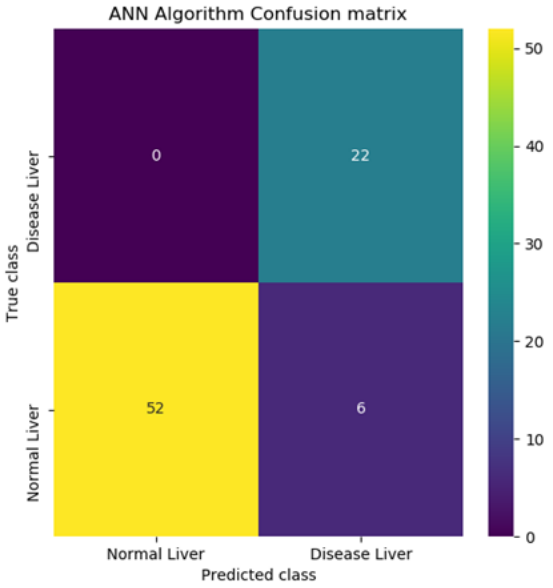
Logistic Regression we got accuracy as 73%



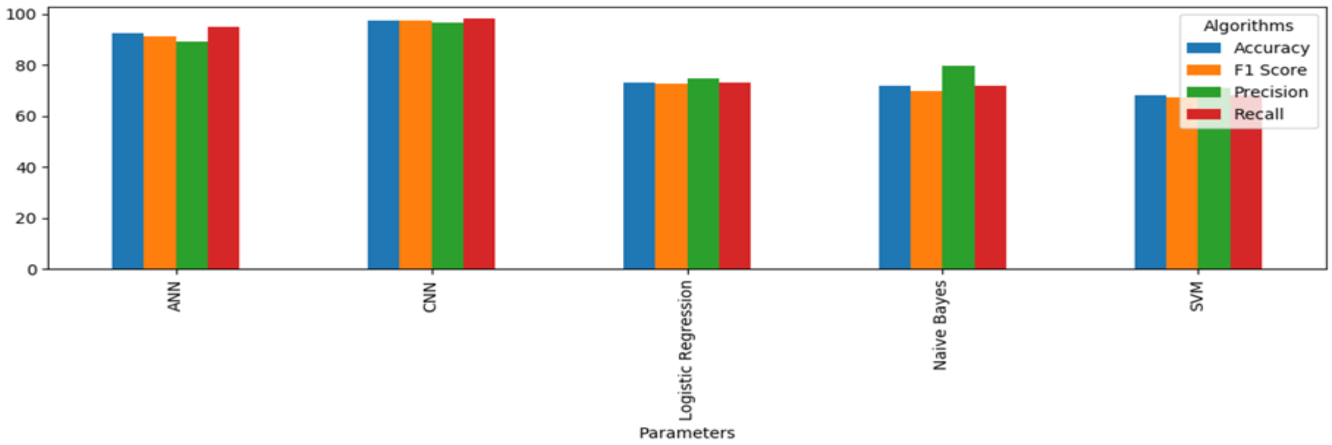
Naïve Bayes we got 71% accuracy



CNN we got 97% accuracy



ANN we got 92% accuracy



Graph x-axis represents algorithm names and y-axis represents accuracy

CONCLUSION

This project confirms how advisable integration of both structured clinical data and liver ultrasound images is in liver disease diagnosis. Logistic Regression structured data best achieving 73 % accuracy while CNN had the advantage on image data with 97% accuracy. The proposed hybrid architecture is helpful to alleviate the concerns about dataset imbalance and increases the reliability of the model. This study suggests perspective of combining various data in reaching to early diagnosis of advanced liver cancer as an alternative approach. The findings justify the application of higher ML paradigms in health care which opens up room for more research on multi-modal data aggregation.

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] V.N. Vapnik, "Statistical Learning Theory", Wiley Publications, 1998

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Delving Deep into Rectifiers", Microsoft Research, 2009

[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