LITERATURE SURVEY

1, Paper Name: Liver Disease Prediction Using Machine Learning Algorithms: A Comparative Study

Author Name: Md. Fazle Rabbi, S.M. Mahedy Hasan, Arifa Islam Champa, Md. Asif Zaman, Md. Kamrul Hasan

link: https://ieeexplore.ieee.org/document/9333528/

Content :

In this paper, we have compared several ML methods, such as Logistic Regression (LR), Decision Tree (DT), Random Forest (RF) and Extra Trees (ET) for the prediction of liver disorders. Issues ignored by the previous researchers have been taken into consideration to improve the prediction accuracy. At the preprocessing steps, categorical values are encoded through label encoding. Then we have utilized the Pearson Correlation Coefficient to improve irrelevant features. After the removal of redundant features of, over sampling is used to mitigate the imbalanced class distribution problem, and feature scaling is used to handle the outliers. After the completion of data preprocessing, LR, DT, RF, and ET classifiers are used for classifying liver disorders patients. For further improvement, the AdaBoost classifiers are used to increase the performance of each classification algorithm.

2,Paper Name: A Survey on machine learning techniques for the diagnosis of liver disease

Author Name: Golmei Shaheamlung, Harshpreet Kaur, Mandeep Kaur

Link: https://ieeexplore.ieee.org/document/9160097/

Content:

This paper gives us the basic idea of past published paper of detection and diagnosis of liver disease based on different machine learning algorithm. With this survey and study it has clearly find and observed that some machine learning algorithm such as Decision tree, J48 and ANN provide better accuracy on detection and prediction of liver disease. And different algorithm has different performance based on different scenario but most importantly the dataset and feature selection is also very important to get better prediction results. And also the paper presents a survey on different types of machine learning techniques used by different authors and every machine learning techniques has some good and bad outcomes depend on the datasets and features selection etc. With this survey we found out that the accuracy and performance can be improve by using different combination or hybrid machine learning algorithm and in future we can also work on more parameter which help to get better performance than the existing technique.

3, Paper Name: Evaluation based Approaches for Liver Disease Prediction using Machine Learning Algorithms

Author Name: C. Geetha, AR. Arunachalam

Link: https://ieeexplore.ieee.org/document/9402463/

Content:

The scale of patient medical records increases day by day in the health care sector. Data mining is the method of using a computer-based information system (CBIS), using modern tactics, to uncover insights from data. The machine learning method is close to that of data mining. Algorithms in machine learning can be differentiated from either supervised or unsupervised methods of learning. For statistical modelling, supervised learning approaches are commonly used. Predictive modelling is a subset of the area of clinical and business intelligence that is used to identify health risks and also to forecast individuals' potential health status. In order to store large-scale information on patient outcomes, procedures, etc., electronic health records (EHR) are used. The data on the EHR can be organised or unstructured. Electronic health records are stored in a standardised data format using managed language to log patient knowledge as a written texts that is hyperlinked in existence. The EHR aims to streamline knowledge about the clinical workflow. Ensemble learning is a well-known method used for prediction by integrating multiple ensemble models of machine learning[1]. Aggregations of various classifiers are J48, C4.5 and Naive Bayes, etc. [2]. Ensembles search for better outcomes than all of the simple classifiers. The proposed work aims to enhance the predictive and classification quality of healthcare data by developing a hybrid predictive classifier model using the classifier ensemble [3][4].

Major issues deliberated on patients with liver disease are not readily detected at starting phase since that can usually operate even though it is partly impaired. An early detection of liver disorders will improve the survival rate of the patient. There is a high probability of liver failure among Indians. India is expected to become the World Capital of Liver Diseases by 2025. Because

of the deskbound lifestyle, increased alcohol intake and smoking, the pervasive prevalence of infection inside liver in India is contributing around 100 forms of liver infections are present. It would also be of great value in the medical field to build a computer that will increase the diagnosis of the disease. This systems can help doctors make correct treatment choices, and the patient queue will also be minimised by liver specialists such as endocrinologists assisted by Automated categorization Methods for Disorders in Liver part. In medical diagnosis and disease prediction, classification techniques are widely common. Michael J Sorich [5] described on chemical datasets, the classification (SVM) provides better prediction results. Paul R Harper [6] stated that an absolute greatest categorization method does not provide forecasted results. However, the unsurpassed achieving algorithm depends on distinct features of the dataset being evaluated.

4, Paper Name: Optimizing Liver disease prediction with Random Forest by various Data balancing Techniques

Author Name: Satessh Ambesange, Vijayalaxmi A, Rashmi Uppin, Shruthi Patil, Vilaskumar Patil

Link: https://ieeexplore.ieee.org/document/9500023/

Content:

The machine-learning model is capable of predicting diseases, based on a data set, which is built in combination of key health parameters of a person with diseases and without diseases. For building models, an effective data set is needed, with proper representation of disease classifications. In this work, Indian

Liver Patient Dataset hosted at ics.uci.edu is used. Several classification machine-learning algorithms are able to classify the liver diseases. Instead of selecting the algorithm, which gives better performance, the paper approaches how to tune the ML module for Random Forest algorithm in step-by-step ways. The Random Forest algorithm is to build a model, since it is trained on several samples of data obtained from splitting data so that the model is not tuned for very specific data. The main focus of the paper is to deeply analyze how models can be further tuned beyond one point of saturation due to an imbalanced data set. Various balancing techniques discussed and their impacts on performance are tabulated in later sections.

5, Paper Name: Liver Diseases Prediction using KNN with Hyper Parameter Tuning Techniques

Author Name: Sateesh Ambesange, Ranjana Nadagoudar, Rashmi Uppin, Vilaskumar Patil, Shruti

Link: https://ieeexplore.ieee.org/document/9297949/

Content:

The disease can be predicted based on health parameters, oral conditions - like alcohol, city pollution level, movement, body chemical compositions using advanced AI/ML techniques. ML is the branch of AI in which a machine learns from a dataset and its performance measures improve with real data over time. The different techniques of ML have been adopted for diagnosis and prophecy of various diseases in the field of medicine. Due to easy access to clinical data, ML algorithms play an important role

in medical decision making. Therefore to identify the disease and make a real-time effective decision the design and develop of a ML model will play a major role. Several ML classification algorithms exist to predict the Liver disease. Each algorithm has different ways of learning from the data set and can be refined / performance tuned. The paper focuses on KNN algorithms, steps to be performance to optimise the model, step by step developing several models. The reason for picking a KNN algorithm is, which looks for several nearby values to classify diseases. Which helps to analyse various more effectively, as increased K value looks at several nearest values, before classifying disease.

Instead of building models using US/Europe based data sets, the paper works on building ML models effectively using Indian dataset and paper discussed how to analyze and predict with more accuracy step by step - preprocessing of data, Univariate and by variate analysis, feature selection, Feature engineering then ML model is trained using this data. As part of preprocessing data is analyzed and checked each feature distribution, most importantly is the data set is balanced/imbalanced and then appropriate methods used to transform the data to normal distribution and imbalance data set is balanced using various methods.. Feature engineering of the dataset is performed to get the important features for prediction and remove the less contributing features so that computation time of the model can be reduced. Hyper parameter tuning methods, tune the parameter's values of KNN ML model to get high accuracy. The paper builds several models, using all these techniques, to indicate performance importance and finally achieving a final KNN ML model to predict liver disease effectively for Indian Dataset. The several models built to predict

liver disease and how performance improves is discussed in next sections.