

How to Cite:

Geetha, C., & Arunachalam, A. R. (2022). Liver tumor prediction by data mining and machine learning techniques in health care environment. *International Journal of Health Sciences*, 6(S4), 6722–6729. <https://doi.org/10.53730/ijhs.v6nS4.10398>

Liver tumor prediction by data mining and machine learning techniques in health care environment

C Geetha

Assistant Professor, Department of Computer Science and Engineering
Bharath Institute of Higher Education & Research, Tamilnadu, Chennai, India
Corresponding author email: geetha.cse@bharathuniv.ac.in

A R Arunachalam

Deputy Dean (Academics), Dr.M.G.R. Educational and Research Institute,
Tamilnadu, Chennai, India
Email: arunachalam.cse@drmgrdu.ac.in

Abstract---In recent medical field advancements, many medicines and high-end curable treatments has been discovered for tumor patients. The key purpose of this work is to center on prediction of tumor in liver of human body by data mining techniques and machine learning algorithms. The therapeutic administrations sector accumulates a massive amount of data that has not been appropriately mined and utilized for effective use. The disclosure of these cloaked real-time gatherings and links is routinely overlooked. Our work focuses on this aspect of medical discovery by translating knowledge about liver tumors to produce intelligent clinical decisions and emotionally supportive networks to assist clinicians. Here nine attributes of blood test values will be used as data set. The detection of tumor in early stage is most important factor to consider for patient's cure. Hence classification of tumors by classification models with the data mining predictive algorithms is performed. Our evaluation obtains better accuracy comparing with different classifier algorithms.

Keywords---data mining, liver tumor prediction, machine learning algorithms, healthcare.

Introduction

The health-care sector generates a great deal of sophisticated data on patients, clinic assets, infection detection, electronic patient records, clinical gadgets, and so on. The massive amount of data is a hidden component that must be prepared

and broken down for data extraction, which enables assistance for cost-investment funds and dynamic. Information from Healthcare Informatics, Medical Device Industry, Pharmaceutical Industry, Hospital Management, and Drug Industry is not restricted in the social insurance board envelope. as shown in Figure-1.

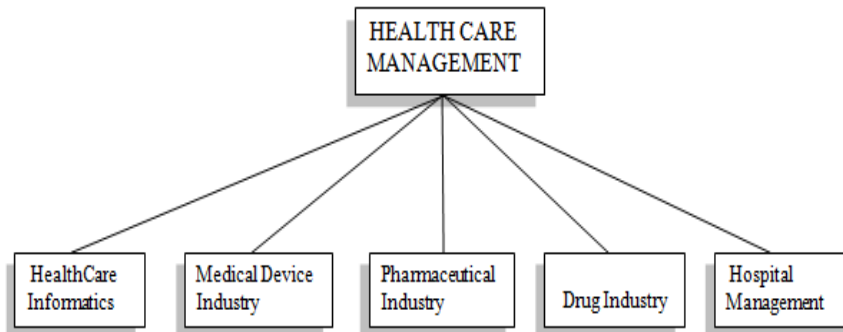


Figure 1 Healthcare Management

Assuming a key job in the digestion and serving a few basic capacities and its issue has gotten one of the huge issues of human sicknesses around the globe. Liver ailments are one of the most executioner illnesses by the most reason for Viral Hepatitis on the planet. Improved wellbeing examination might be cultivated finished programmed conclusion of patient record put away in wellbeing information, i.e., By gaining from past encounters [1]. Also, information mining is applied to this put away clinical record to recover data from the information. As indicated by data mining, methods arranged into Association, order, and grouping. Affiliation its discover the level of the co-event the thing in the information, order it's an information mining capacity that allocates the thing to the objective class, and it predicates the real class to the things however Clustering not at all like characterization it's not allotted the class, and it separates the dataset into little parcels dependent on its closeness [3]. Information mining brings a lot of devices and procedures that can be applied to this prepared information to find concealed examples that give medicinal services experts an extra wellspring of information for deciding [5]. The choice rests with social insurance experts [6].

Liver Tumor Prediction Process Diagram

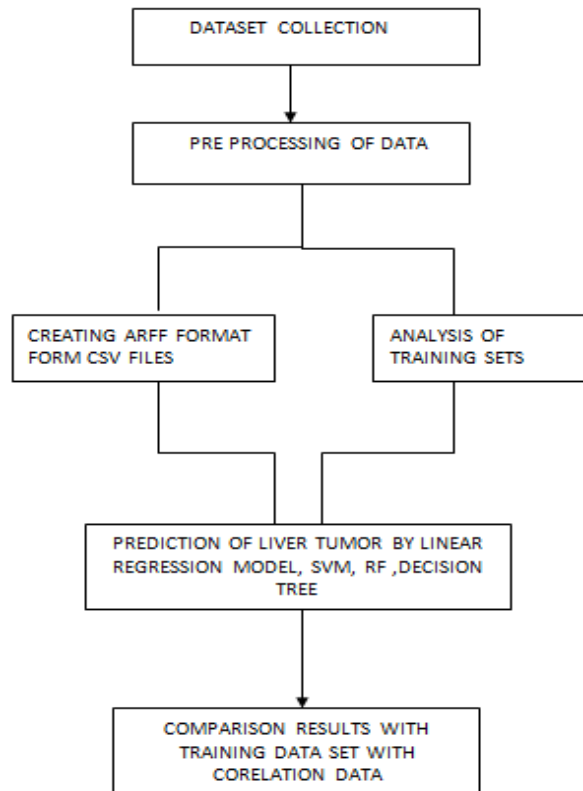


Figure 2 Training and Testing Process

Figure 2 depicts the general organization of the preparation and testing procedure. Pre-processing, regional liver identification, fix extraction, surface element extraction, surface element measurement reduction, classifier preparation and order, and treatment relapse model preparation and forecast are among the tactics [4]. With a high discriminatory particle, we intend to isolate a large number of surface features that give hopeful therapy with the most crucial risk of perseverance. In addition, we use a clear yet appropriate way to deal with region of liver and liver tumors on CT images of any other notable AI or AI algorithms. Figure 3 below shows the original picture, the region of the liver and the region of the tumor. [2].

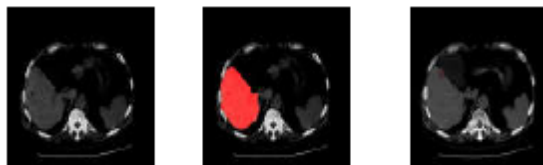


Figure 3 Example of Liver tumor image

Data Mining and ML Techniques

There are many data mining and machine learning techniques presented in this section. These techniques are used for efficient prediction of liver disease in healthcare environment. Some of the important techniques are listed below.

1. Decision Tree: The decision tree algorithm follows standard order model, with consistency of good accuracy rate, simple implied grouping and lower blunder intercession with frequency. Selection tree analysis is a commonly used inductive rationale estimation.
2. Support Vector Machine: Throughout each survey, specialists have made reliable forecasts of the risk of liver malignant development, high or healthy. That is, every snippet of accumulated information is considered with the details. SVM is a targeted different classification learning model that can recognize the limit (hydroplane) with the widest range of information, making the calculation suitable for emulating current data.
3. Logistic Regression: Logistic regression is a Organizational remission is a linear relapse test. Scientists can fabricate linear capabilities and use AI to obtain categorization or anticipate expectations.
4. Algorithm for neighbor selection: The algorithm centers around the Nearest neighbor dimension that has reduced datasets in addition to distinguish between the ideal recurrence section and the forecast collection. This is called as KNN algorithm
5. Collection Learning: algorithm related to ensemble Based learning is used toward test the actual statistical model and develop a latest set of information with the same size as the original sample. Actual set of data T, such fresh datasets are used to educate numerous classifications. The conclusive outcomes of the characterization will be resolved based on the votes of the individual aftereffects of those groupings. The investigations have been performed by collection Based ensemble erudition for the classification which is built with respect to decision algorithm.
6. Algorithm on neural pertained Networks: This algorithm is known as ANN that recreates set of ranges with respect to determinant events. It is a pattern recognition scheme depending upon framework together with utility of the neuron-based network. This has been used by the majority of researchers for classification and prediction.

System Model

The Body Mass Index (BMI) is a worldwide benchmark for measure the extent of body fat distribution and slimness, as well as whether or not it is safe. Equation (1) shows how to define it:

$$BMI = wg/n^2 \quad (1)$$

Where wg is weight which is measured in kilogram and n is height which is measured in meters as shown in eqn 2. Here SI is represented as index value of smoking people.

$$SI = X_1 * X_2 \quad (2)$$

Where X_1 is cigarettes smoked and X_2 is the year. The model of proposed model is shown in equation 3.

$$f_{\omega, \gamma}(h) = \omega^R u + \gamma \quad (3)$$

$f_{\omega, \gamma}(h)y_i > 0$ implies the outcome values that are compatible with the authentic categorization. The groups has been segregated into two which are the classified as marks affirmative (+1), pessimistic (-1). The f value point towards prediction outcome value is dependable with the authentic categorization predicament. Equation 4 depicts the knowing the values of and for which $f_{\omega, \gamma}(h)y_i > 0$ is a positive value:

$$(\omega^R h_i + \gamma)y_i > 0 \quad (4)$$

Experimental Study and Analysis

The weka tool [8] has developed algorithms for classification of data. The informational set Linked to be posted and perpetual arrangement of information ought to be in format with “. arff” extension. When the required data is given as an input, essential estimation is applied in the Classifiers tab and the going with boundary can be picked. In the event that the beginning button is clicked, it gives the favored yield with the plans, for example, root mean square esteem, absolute error esteem and so forth. Our Evaluation has been processed with SVM, Linear regression and Random Forest algorithm. Our dataset consists of 583 patient instances with 9 attributes such as Age, Gender, Total bilirubin, Direct bilirubin, Alkaline phosphatase Aspartate Aminotransferase ,Total proteins, Albumin, Globulin. Figure 6 shows an output screen of classifier tab

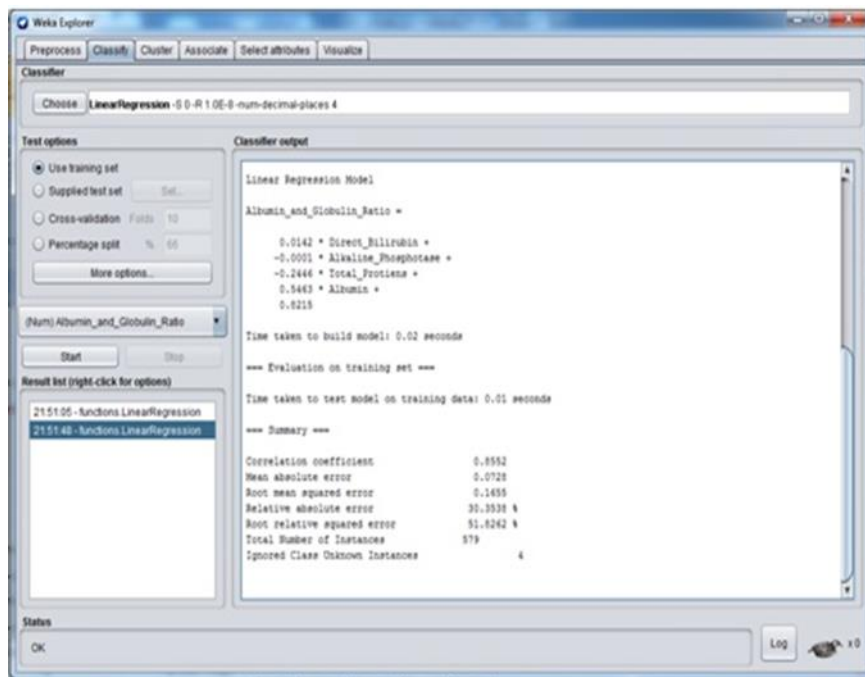


Figure 4 Output Window of Model Processing

Table 1 Comparative Estimate Models from Training Data

S.No	Algorithm/Model	Correlation coefficient	Mean absolute error	Root mean squared error	Accuracy(After Cross Validation)
1	Linear Regression Model	0.8567	0.0724	0.1647	93%
2	Random Forest	0.9794	0.0408	0.0764	90%
3	Decision tree	1	0.0006	0.0024	92%
4.	SVM	0.8467	0.0003	0.0022	92%

Table 1 shows the comparison results of prediction of liver cancer from the training datasets. The table also shows the accuracy of prediction where linear regression and decision tree is better as compared with SVM and Random forest. The accuracy is calculated using the linear equation $y=mx+c$.



Figure 5 Prediction Accuracy

In figure 5, LR denotes Linear Regression, RT denotes Random Forest, DT stands for Decision Tree and SVM stands for Support vector machine model. The comparative results show that after cross validation with 11 attributes of the dataset, the prediction level has been increased efficiently.

Conclusion and Future Work

In this work, data mining techniques have been addressed for prediction of liver tumor in healthcare environment. Our results have been identified in a comparative manner with respect to linear regression, Decision Tree, SVM and Random Forest classifiers. The prediction accuracy has been calculated from the parameters like coefficient of correlation data, error data of accumulated, Mean absolute values and error values of Root mean square. The results showed an efficient improvement strategy of the classifier models. Our future work will be focused with CT scan images and patient datasets with alcoholic and non-alcoholic characteristics.

References

1. Cheng-Mei Chen, Chien-Yeh Hsu, 'Prediction of Survival in Patients with Liver Cancer using Artificial Neural Networks and Classification and Regression Trees', Seventh IEEE International Conference on Natural Computation', 26-28 July 2011, pp 811-815, China.
2. Intouch Kunakorntrum, Woranich Hinthong, Sumet Amonyongchareon, Phond Phunchongharn, 'Liver Cancer Prediction Using Synthetic Minority based on Probabilistic Distribution (SyMProD) Oversampling Technique', IEEE 10th International Conference on Awareness Science and Technology (iCAST), 23-25 Oct. 2019, Japan, pp 1-6.
3. L. Ali, A. Hussain, J. Li, U. Zakir, X. Yan, Mufti Mahmud, M. Rajak, 'Intelligent Image Processing Techniques for Cancer Progression Detection, Recognition and Prediction in the Human Liver', IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE), 9-12 Dec. 2014, USA

4. P R Anisha ,C Kishor Kumar Reddy ,L V Narasimha Prasad, 'International Conference on Signal Processing and Communication Engineering Systems-IEEE xplore '2-3 Jan. 2015,pp 352-357,India.
5. Reetu, Narender Kumar, 'Medical Diagnosis for Liver Cancer using Classification Techniques', International Journal of Science and Research, Volume 4 Issue 5, ,pp 2058-2061, 2015.
6. Shambel Kefelegn, Pooja Kamat, 'Prediction and Analysis of Liver Disorder Diseases by using Data Mining Technique: Survey', International Journal of Pure and Applied Mathematics, Volume 118 issue 9, pp 765-770, 2018.
7. Tapas Ranjan Baitharu, Subhendu Kumar Pani, 'Analysis of Data Mining Techniques For Healthcare Decision Support System Using Liver Disorder Dataset', Procedia Computer Science Vol 85,pp 862 – 870, 2016.
8. <https://www.cs.waikato.ac.nz/ml/weka/>
9. Jani, J. R., Bajamal, A. H., Utomo, S. A., Parenrengi, M. A., Fauzi, A. A., Utomo, B., & Dwihapsari, Y. (2021). Correlation between magnetic resonance imaging (MRI) and dynamic mechanical analysis (DMA) in assessing consistency of brain tumor. *International Journal of Health & Medical Sciences*, 4(2), 260-266. <https://doi.org/10.31295/ijhms.v4n2.1737>
10. Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2021). Get vaccinated when it is your turn and follow the local guidelines. *International Journal of Health Sciences*, 5(3), x-xv. <https://doi.org/10.53730/ijhs.v5n3.2938>
11. Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2021). Health and treatment of diabetes mellitus. *International Journal of Health Sciences*, 5(1), i-v. <https://doi.org/10.53730/ijhs.v5n1.2864>