



Project Work :

Final ISA(Review 4) / ESA 2020

Project Title : Prediction Of Hepatitis and Liver Damage.
Project ID : PW20MG02
Project Guide : Prof. Mahitha G
Project Team : 01FB16ECS183 - M Pradeep Kumar
01FB16ECS185 - M Sumukha
01FB16ECS395 - Sreerama Priyanka





Problem Statement

- Chronic viral hepatitis is now the second biggest killer after tuberculosis. More than 3600 people die every day in case of the viral hepatitis disease (WHO, 2017). Therefore, this study aims to develop model that can improve the compressiveness of advice provided for the people especially people in remote areas.
- Hepatitis B:
prediction of survival of the patient based on clinical report.
- Hepatitis C:
predicting the stages of fibrosis.
- Liver Damage:
Predicting the liver damage so, they can go for further treatment.



To confirm the Hepatitis disease accurately person must undergo liver Biopsy which is invasive procedure, costly, needs expertise and might lead to complications.

- Doctor can take actions according to the model result.
- Disease can be confirmed without Biopsy.
- Dentists are more prone to Hepatitis even they can use this.



Literature Survey

List the papers/references studied

Research Paper:

- Title: Adaptive Learning Expert System for Diagnosis and Management of Viral Hepatitis
- Link: https://www.researchgate.net/publication/332294787_Adaptive_Learning_Expert_System_for_Diagnosis_and_Management_of_Viral_Hepatitis
- Authors: Henok Yared Agizew
- Dataset - <http://archive.ics.uci.edu/ml/datasets/Hepatitis>
- Methodology: This study aims to develop an intelligent expert systems. It is modeled by using decision tree to simulate procedures involved in diagnosis and management of viral hepatitis.



List the papers/references studied

Research Paper:

- Title: Prediction of Liver Fibrosis stages by Machine Learning model.
- Link: Authors: Heba Ayeldeen, Olfat Shaker, Ghada Ayeldeen, Khaled.
- Dataset: <https://archive.ics.uci.edu/ml/datasets/Hepatitis+C+Virus+%28HCV%29>
- Methodolgy: Data attributes are classified considering the models SVM. We have used PCA for feature selection.



List the papers/references studied

Research Paper:

- Title: Liver Disease prediction using different Decision tree techniques.
- Link: <https://aircconline.com/ijdkp/V8N2/8218ijdkp01.pdf>
- Authors: Nazmun Nahar and Ferdous Ara
- Dataset - <https://www.kaggle.com/uciml/indian-liver-patient-records>
- Methodolgy: The study used the dataset from (ILPD) with 583 samples with 10 independent variables and 1 dependent variable and on them classification algorithms are applied.



Proposed Solution

1. Viral Hepatitis B:

1. Dataset- UCI- Number of samples: 155
2. Data Pre-processing - replacing text to number and imputer is used for missing values
3. Classification Algorithm: Decision Tree, Gradient Boosting and Random forest.

2. Hepatitis C:

1. Dataset: UCI-Number Of samples: 1485
2. Data Pre-processing:
 - > Data attributes are classified by considering the clinical ranges for an attribute.
 - > SMOTE Over Sampling Technique is used.
3. Feature Selection:
 - > Dataset: number of features - 30.
 - > PCA feature selection and medical research for selection.



Proposed Solution

4. Classification Algorithm

-> Models - Decision Tree, logistic regression, Random forest and SVM.

3. Liver Damage:

1. dataset- ILPD- number of samples:583

independent Variables:10, dependent:1

2. Data Pre-processing

->Imputer is used for missing values

-> SMOTE Over Sampling Technique is used

3. Classification Algorithm: SVM classifier and Logistic Regression.



Why Your Solution is Better?

- For pre-processing we have not only used machine learning algorithms we have also done medical research for feature selections.



Technologies / Methodologies

- Jupyter Notebook: Analyze Code
- Language : Python
- HTML/CSS and jQuery: UI visualization
- Flask: to run the server in the backend.

Software Engineering Approach : Agile

Why? : Create goals and distribute work.

Update each team member on the progress and discuss implementation.

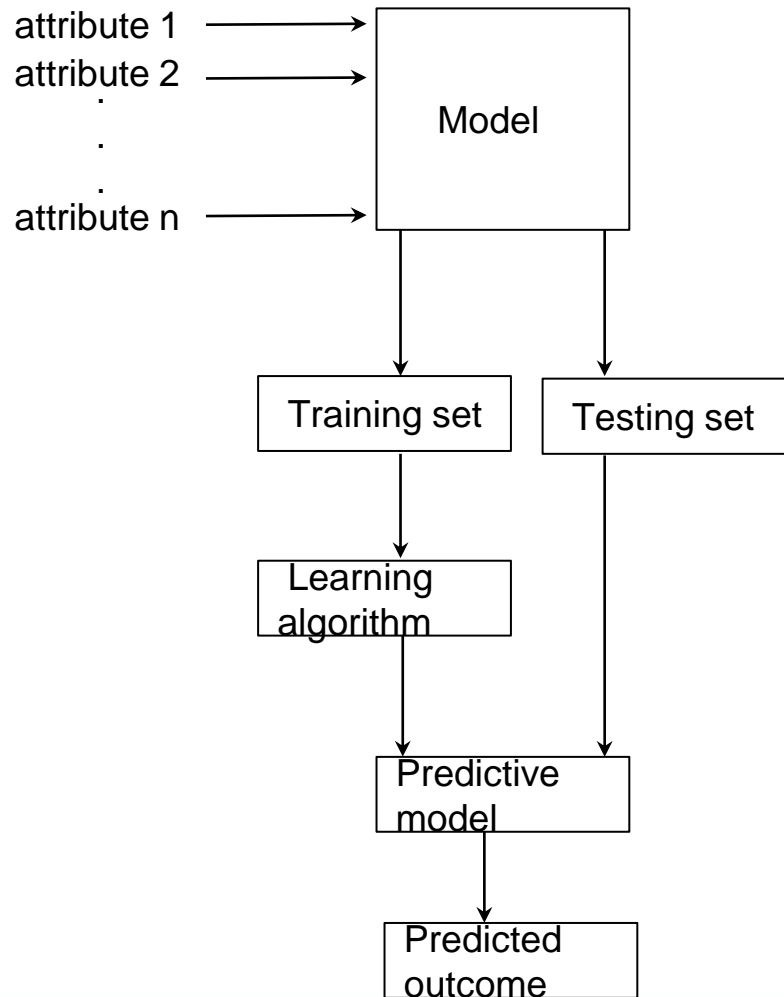


Dependencies and Risks

- Patient must consult Doctor to know the Diagnosis and further treatment plan.
- Patients must undergo blood tests.
- Since our model is probabilistic predictor, doesn't have 100% accuracy.

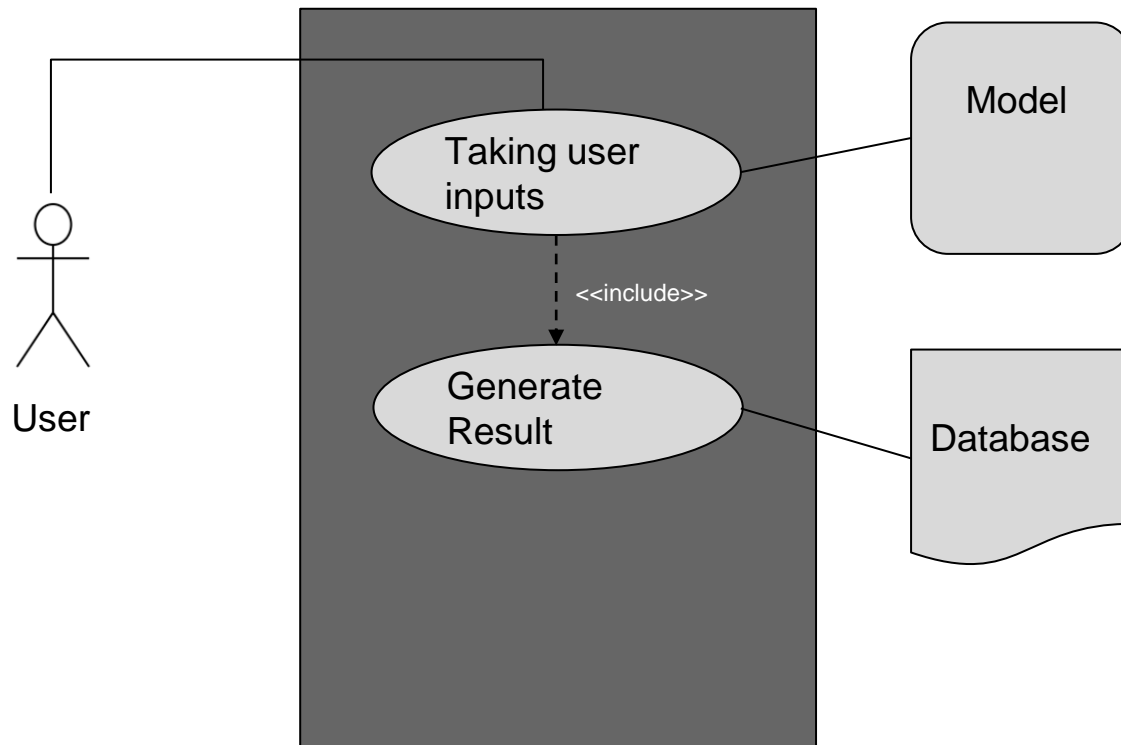


System Architecture





UI/ Use Case





- User: User has the lab reports which are sent to the Liver damage predictor.
- Database: 583 data samples are trained.
- Model: model uses logistic regression and SVM classifier to generate results.



Implementation Details

- Data preprocessing: Model cannot understand textual data so replacing missing values using Imputer.
- Feature selection: We have applied PCA, SVD, Recursive Feature elimination.
- Applying Machine learning models:
 - Random Forest
 - Decision Tree
 - Gradient Boosting
 - Logistic Regression



Design Approach

What is the design approach followed?

We have followed the Experimental research design approach

Why?

Experimental research design is a powerful design for testing hypotheses of causal relationship among variables

Benefits of the approach :

The controlled environment in which the study is conducted can yield a greater degree of purity in observation.

Drawbacks of the approach :

It is also difficult to get cooperation from the participants in the study.





Project Completion

Project Report Status -Ready
Project Demo - Fully done

Future work: Currently model predicts based on laboratory test dataset, In future preparing a model which takes both image (Ultrasound scanning of liver) and laboratory test dataset to enhance prediction.



Project Results

LIVER DAMAGE PREDICTION

Method	Accuracy
Logistic Regression	71.59
SVM	73.14
Random Forest	70.29
Smote: Logistic Regression	73.2
Smote: Radom Forest	80.4



HEPATITIS B LIVE OR DIE PREDICTION

Method	Accuracy
Random Forest Classifier	74.19
PCA- 15 attributes, Random Forest	83.87
Gradient Boosting Classifier	68
Decision Tree	61.29

HEPATITIS C – Fibrosis Stages

Method	Accuracy
SVM	27.8
Random Forest	32.49
Logistic Regression	22.18
Decision Tree	24.18





HEPATITIS C – Fibrosis Stages after Pre-processing

Method	Accuracy
SVM	27.44
Random Forest	20.58
Logistic Regression	22.18

HEPATITIS C – Enzyme and virus count

Method	Accuracy
Logistic Regression (ALT)	75.45
Logistic Regression (RNA)	65.34
Decision Tree (ALT)	66.42
Decision Tree (RNA)	79.6





Planned Effort Vs Actual Effort

Planned Effort

Topic	Time
Literature Survey and Data Collection	Review 1
Pre processing and feature selection	Review 2
Model development	Review 3
Final Model	Review 4

Actual Effort

Topic	Time
Literature Survey and Data Collection	Review 1
Literature Survey and Pre processing	Review 2
feature selection	Review 3
Final Model Development	Review 4



Lessons Learnt

Learnt things:

- > Working as a team.
- > understood importance of domain knowledge to solve a problem.

Challenges faced:

- > Difficulty in acquiring dataset in earlier stage.
- > Algorithms like RFE were time consuming.



Thank You

