Artificial Intelligence in the field of Medical Decision Support System

Riya Pandey

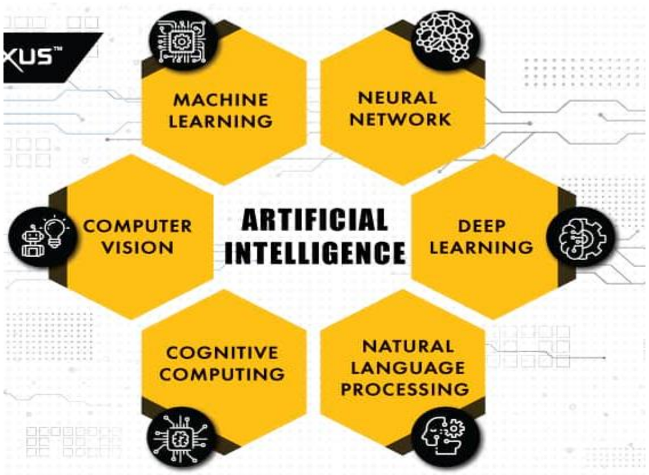
28th March 2022

ABSTRACT:

Management information system (MIS), decision support system (DSS) and executive support system (EES) are the inevitable constituents of the intelligent systems which are being integrated in the infrastructural and technological development of the organizations and to address non-routine decisions. The intelligent systems are incorporated with methodologies that support providing solutions to the unpredicted decisions by employing mathematical and statistical tools and incorporating software programs embedded with the cutting edge algorithms. We investigate the applicability of a number of algorithms in the healthcare domain and propose mechanisms of development of machine learning techniques in the area of artificial intelligence. Artificial intelligence (AI) encompasses integer linear programming (ILP) and machine learning (ML) that further motivates us to dig up the algorithms and learning techniques to find the best solution in the field of predictive analytics by examining a Liver disease stage classification problem and comparing the results for better diagnosis.

INTRODUCTION:

Machine learning (ML) has attracted scientists and researchers to explore its applicability in a multitude of socio-technical disciplines such as e-commerce, weather forecasting, financial predictions, healthcare, etc. One of the classic examples of ML could be recognizing handwritten text in the discipline of pattern recognition. Furthermore, identifying the DNA sequence matching with the DNA pattern of a patient from the data repository who is prone to developing diabetes falls in the pattern recognition ML category in the biomedical domain. Another example could be in the area of pharmaceuticals which involves investigating the biomarkers for detection of pancreatic cancer using the techniques of big data analytics. The machine learning algorithms is extensively studied in the biomedical domains such as robotics for medical device, signal processing, image processing and natural language for medical records in this proposition. Appropriate ML algorithms such as kNN, Logistic regression, random forest, decision tree etc. are used to make diagnostic predictions on the provided dataset. Here, we will be using the Support Vector Machines, Random Forest and decision trees algorithms and compare the results to find the most suitable method for classifying the dataset.



PROBLEM STATEMENT:

This Machine Learning Challenge requires participants to create predictive models to predict the stage of liver Cirrhosis using 18 clinical features. Cirrhosis damages the liver from a variety of causes leading to scarring and liver failure. Hepatitis and chronic alcohol abuse are frequent causes of the disease. Liver damage caused by cirrhosis can't be undone, but further damage can be limited. Treatments focus on the underlying cause. In advanced cases, a liver transplant may be required. Predicting the stage of cirrhosis and beginning the treatment before it's too late can prevent the fatal consequences of the disease.

Market / Customer/ Business Need Assessment:

There were 10.6 million (10.3–10.9) prevalent cases of decompensated cirrhosis and 112 million (107–119) prevalent cases of compensated cirrhosis globally in 2017. In liver medical imaging, physicians usually detect, characterize, and monitor diseases by assessing liver medical images visually. Sometimes, such visual assessment, which is based on expertise and experience, may be personal and inaccurate. AI can make a quantitative assessment by recognizing imaging information automatically instead of such qualitative reasoning. Therefore, AI can assist physicians to make more accurate and reproductive imaging diagnosis and greatly reduce the physicians' workload. There are two kinds of AI methods widely used in medical imaging currently, one is traditional machine learning algorithms, and the other one is deep learning algorithms. Studies showed an overall sensitivity to chronic liver disease of 65%-95%, with a positive predictive value of 98%.

In this article, I am going to emphasize about machine learning, a branch of artificial intelligence, helping physicians in prognosis and people suffering from liver cirrhosis. This is a system which takes in data, finds patterns, trains itself using the data and outputs an outcome. In this article we explore applications of artificial intelligence to provide business leaders with an understanding of current and emerging trends, presents representative examples of people facing different stages of liver cirrhosis and most importantly addressing the importance of its prognosis process in eradicating this disease. Here, I am going to classify the data into different stages of liver cirrhosis so that patients may get appropriate treatment on time.

Target Specifications and Characterization:

A. To change traditional liver cirrhosis diagnosis process to faster and accurate process.

B. Reducing frustration and death of patients due to delay in the prognosis process.

C. Predetermined dataset of patients suffering from different stages of liver cirrhosis is taken and based on that prediction is performed.

Above, mentioned targets can be achieved by analyzing:

1. What the patient look for.
2. How are present diagnosis processes being performed?
3. Problems faced by people suffering from liver cirrhosis.
4. How to identify and provide treatment in initial stage accurately.
5. How efficiently are the physicians able to detect the severity of the disease?
6. When and where a patient likes to trust and spend on?
7. Analyzing the needs of the patients suffering from liver cirrhosis.
8. To help patient fight liver cirrhosis early stage.
9. To send results to the patient within minutes and prescribing the further steps to be taken by the patient depending on the stage he/she has been diagnosed with.
10. To remind the patient about the latest changes in the diagnosis process.

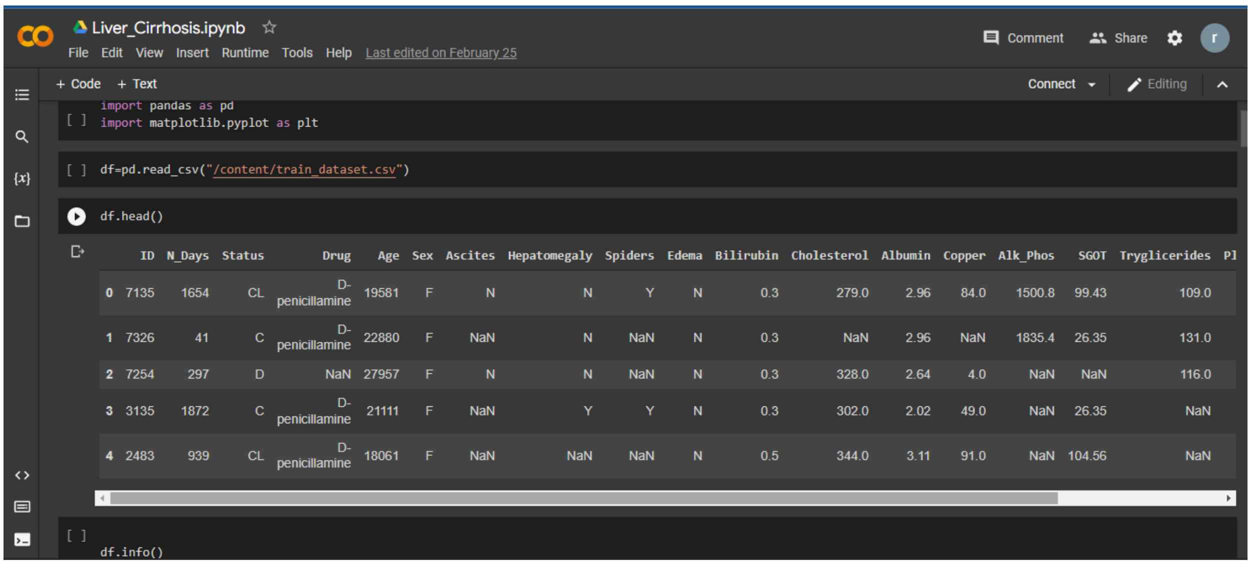
Machine learning-based prediction of survival prognosis:

Survival prediction after first diagnosis is important for both disease specialist and patients or their family members. First, as the survival ability of the liver cirrhosis patients largely depends on the stage of the disease, accurately forecasting the prognosis would be helpful for estimating the severity and the time point of disease progression. On the other hand, patients and the families can set appropriate goals base on the accurate survival prediction. As the result, the timely prevention and treatment would be made and the worse treatment decision, such as over-treatment or late palliative care, would be effectively avoided.

DATASET:

Cirrhosis is a late stage of scarring (fibrosis) of the liver caused by many forms of liver diseases and conditions, such as hepatitis and chronic alcoholism. The following data contains the information collected from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984. A description of the clinical background for the trial and the covariates recorded here is in Chapter 0, especially Section 0.2 of Fleming and Harrington, Counting Processes and Survival Analysis, Wiley, 1991. A more extended discussion can be found in Dickson, et al., Hepatology 10:1-7 (1989) and in Markus, et al., N Eng J of Med 320:1709-13 (1989).

A total of 424 PBC patients, referred to Mayo Clinic during that ten-year interval, met eligibility criteria for the randomized placebo-controlled trial of the drug D-penicillamine. The first 312 cases in the dataset participated in the randomized trial and contain largely complete data. The additional 112 cases did not participate in the clinical trial but consented to have basic measurements recorded and to be followed for survival. Six of those cases were lost to follow-up shortly after diagnosis, so the data here are on an additional 106 cases as well as the 312 randomized participants.



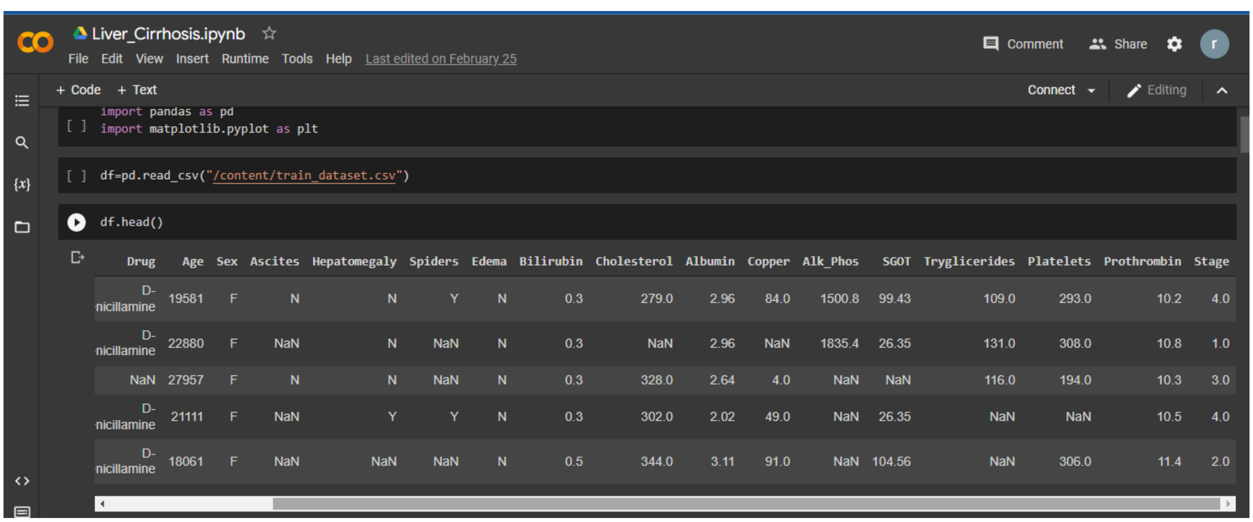

```
import pandas as pd
import matplotlib.pyplot as plt
```

```
df=pd.read_csv("/content/train_dataset.csv")
```

```
df.head()
```

	ID	N_Days	Status	Drug	Age	Sex	Ascites	Hepatomegaly	Spiders	Edema	Bilirubin	Cholesterol	Albumin	Copper	Alk_Phos	SGOT	Tryglicerides	Pl
0	7135	1654	CL	D-penicillamine	19581	F	N	N	Y	N	0.3	279.0	2.96	84.0	1500.8	99.43	109.0	
1	7326	41	C	D-penicillamine	22880	F	NaN	N	NaN	N	0.3	NaN	2.96	NaN	1835.4	26.35	131.0	
2	7254	297	D	NaN	27957	F	N	N	NaN	N	0.3	328.0	2.64	4.0	NaN	NaN	116.0	
3	3135	1872	C	D-penicillamine	21111	F	NaN	Y	Y	N	0.3	302.0	2.02	49.0	NaN	26.35	NaN	
4	2483	939	CL	D-penicillamine	18061	F	NaN	NaN	NaN	N	0.5	344.0	3.11	91.0	NaN	104.56	NaN	

```
df.info()
```




```
import pandas as pd
import matplotlib.pyplot as plt
```

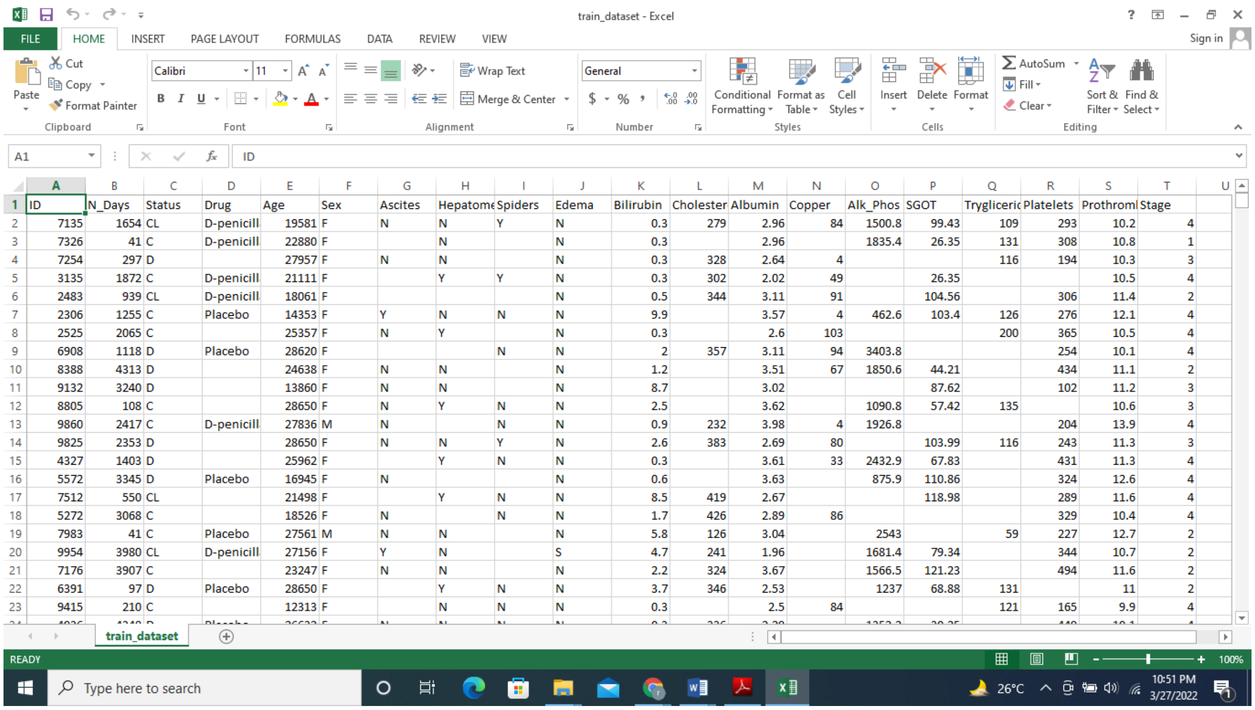
```
df=pd.read_csv("/content/train_dataset.csv")
```

```
df.head()
```

Drug	Age	Sex	Ascites	Hepatomegaly	Spiders	Edema	Bilirubin	Cholesterol	Albumin	Copper	Alk_Phos	SGOT	Tryglicerides	Platelets	Prothrombin	Stage
D-nicillamine	19581	F	N	N	Y	N	0.3	279.0	2.96	84.0	1500.8	99.43	109.0	293.0	10.2	4.0
D-nicillamine	22880	F	NaN	N	NaN	N	0.3	NaN	2.96	NaN	1835.4	26.35	131.0	308.0	10.8	1.0
NaN	27957	F	N	N	NaN	N	0.3	328.0	2.64	4.0	NaN	NaN	116.0	194.0	10.3	3.0
D-nicillamine	21111	F	NaN	Y	Y	N	0.3	302.0	2.02	49.0	NaN	26.35	NaN	NaN	10.5	4.0
D-nicillamine	18061	F	NaN	NaN	NaN	N	0.5	344.0	3.11	91.0	NaN	104.56	NaN	306.0	11.4	2.0

Attribute Information-

1) ID: Unique Identifier
2) N_Days: number of days between registration and the earlier of death, transplantation, or study analysis time.
3) Status: status of the patient C (censored), CL (censored due to liver tx), or D (death)
4) Drug: type of the drug. D-penicillamine or placebo
5) Age: age in [days]
6) Sex: M (male) or F (female)
7) Ascites: presence of ascites N (No) or Y (Yes)
8) Hepatomegaly: the presence of hepatomegaly N (No) or Y (Yes)
9) Spiders: the presence of spiders N (No) or Y (Yes)
10) Edema: the presence of edema N (no edema and no diuretic therapy for edema), S (edema present without diuretics, or edema resolved by diuretics), or Y (edema despite diuretic therapy)
11) Bilirubin: serum bilirubin in [mg/dl]
12) Cholesterol: serum cholesterol in [mg/dl]
13) Albumin: albumin in [gm/dl]
14) Copper: urine copper in [ug/day]
15) Alk_Phos: alkaline phosphatase in [U/liter]
16) SGOT: SGOT in [U/ml]
17) Triglycerides: triglicerides in [mg/dl]
18) Platelets: platelets per cubic [ml/1000]
19) Prothrombin: prothrombin time in seconds [s]
20) Stage: histologic stage of disease (1, 2, 3, or 4)



ID	N_Days	Status	Drug	Age	Sex	Ascites	Hepatome	Spiders	Edema	Bilirubin	Cholester	Albumin	Copper	Alk_Phos	SGOT	Tryglicerid	Platelets	Prothrom	Stage
7135	1654	CL	D-penicill	19581	F	N	N	Y	N	0.3	279	2.96	84	1500.8	99.43	109	293	10.2	4
7326	41	C	D-penicill	22880	F		N		N	0.3		2.96		1835.4	26.35	131	308	10.8	1
7254	297	D		27957	F	N	N		N	0.3	328	2.64	4			116	194	10.3	3
3135	1872	C	D-penicill	21111	F		Y	Y	N	0.3	302	2.02	49		26.35			10.5	4
2483	939	CL	D-penicill	18061	F				N	0.5	344	3.11	91		104.56		306	11.4	2
2306	1255	C	Placebo	14353	F	Y	N	N	N	9.9		3.57	4	462.6	103.4	126	276	12.1	4
2525	2065	C		25357	F	N	Y		N	0.3		2.6	103			200	365	10.5	4
6908	1118	D	Placebo	28620	F			N	N	2	357	3.11	94	3403.8			254	10.1	4
8388	4313	D		24638	F	N	N		N	1.2		3.51	67	1850.6	44.21		434	11.1	2
9132	3240	D		13860	F	N	N		N	8.7		3.02			87.62		102	11.2	3
8805	108	C		28650	F	N	Y	N	N	2.5		3.62		1090.8	57.42	135		10.6	3
9860	2417	C	D-penicill	27836	M	N		N	N	0.9	232	3.98	4	1926.8			204	13.9	4
9825	2353	D		28650	F	N	N	Y	N	2.6	383	2.69	80		103.99	116	243	11.3	3
4327	1403	D		25962	F		Y	N	N	0.3		3.61	33	2432.9	67.83		431	11.3	4
5572	3345	D	Placebo	16945	F	N			N	0.6		3.63		875.9	110.86		324	12.6	4
7512	550	CL		21498	F		Y	N	N	8.5	419	2.67			118.98		289	11.6	4
5272	3068	C		18526	F	N		N	N	1.7	426	2.89	86				329	10.4	4
7983	41	C	Placebo	27561	M	N	N		N	5.8	126	3.04		2543		59	227	12.7	2
9954	3980	CL	D-penicill	27156	F	Y	N		S	4.7	241	1.96		1681.4	79.34		344	10.7	2
7176	3907	C		23247	F	N	N		N	2.2	324	3.67		1566.5	121.23		494	11.6	2
6391	97	D	Placebo	28650	F		Y	N	N	3.7	346	2.53		1237	68.88	131		11	2
9415	210	C		12313	F		N	N	N	0.3		2.5	84			121	165	9.9	4

Machine learning as the future of liver disease stage prediction:

AI is set to change the medical industry in the coming decades — it wouldn't make sense for pathology to not be disrupted too. Currently, ML models are still in the testing and experimentation phase for liver cirrhosis prognosis. As datasets are getting larger and of higher quality, researchers are building increasingly accurate models. Machine Learning is the next step forward for us to overcome this hurdle and create a high accuracy medical decision support system. It is of great significance to indicate that AI is different from human intelligence in numerous ways. Although various forms of AI have exceeded human performance, they lacked higher-level background knowledge and failed to establish associations like the human brain.

BENCHMARKING ALTERNATE PRODUCTS:

There is a long history in the treatment of liver diseases of using traditional Chinese medicine which is considered a complementary or an alternative medical system in most Western countries. Forms of herbal drug include ointment, pellet, ball, powder, and fluid decoction. More standardized treatments are extracts from single herbs and formula injections. However, both medical professionals and the patients should be more concerned about drug safety when herbal injection is used. Some patients who used herbal injection suffered adverse reactions, including death.

Overall, the patients who chose alternative treatments were more likely to be younger, female, healthier, and have higher incomes and education levels. Some of these traits, such as overall better health, would normally improve the odds of survival after a liver cirrhosis diagnosis.

Applicable Regulations (Government and Environmental):

a. Patents on ML algorithms developed
b. Laws related to privacy for collecting data from users
c. Protection/ownership regulations
d. Creating an e-mail service to mail the report to the patient and doctor.
e. Being responsible by design.
f. Ensuring open-source, academic and research community for an audit of Algorithms.
h. Review of existing work authority regulations.

Applicable Constraints:

Expertise:
A. Requires a lot of research to obtain universal dataset of liver disease patients in-order to provide more sophisticated and accurate results.

B. Establishing e-mail service in the product which have to send the report after the machine learning model is deployed in any server.

C. Confidential health data to be obtained to train the model.

D. Thorough understanding of dataset and verification of the results must be performed by the pathologist from the machine learning model to provide a great health prescription and service to the user.

Business Opportunity:

Physicians are pretty good in diagnosing liver cirrhosis by looking at the images obtained by various medical imaging techniques such as computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, positron emission tomography (PET), mammography, and X-ray. In clinical work, the interpretation and analysis of medical images are mainly done by human experts. Recently, medical doctors have begun to benefit from the help of computer-aided diagnosis. Artificial intelligence (AI) is intelligence applied by machines, in contrast to the natural intelligence displayed by humans. In liver medical imaging, physicians usually detect, characterize, and monitor diseases by assessing liver medical images visually. Sometimes, such visual assessment, which is based on expertise and experience, may be personal and inaccurate. AI can make a quantitative assessment by recognizing imaging information automatically instead of such qualitative reasoning. Our main objective is to use Machine Learning, which not only gives faster results but also demonstrates higher accuracy in the Liver Cirrhosis prediction process.

Concept Generation:

Cirrhosis is a complication of liver disease that involves loss of liver cells and irreversible scarring of the liver. Alcohol and viral hepatitis B and C are common causes of cirrhosis, although there are many other causes. Cirrhosis in itself is already a late stage of liver damage. In the early stages of liver disease, there will be inflammation of the liver. If this inflammation is not treated it can lead to scarring (fibrosis). At this stage, it is still possible for the liver to heal with treatment.

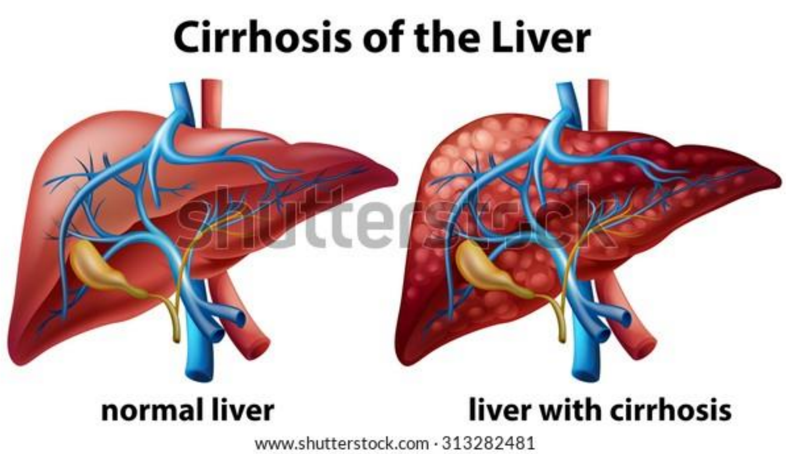
If fibrosis of the liver is not treated, it can result in cirrhosis. At this stage, the scar tissue cannot heal, but the progression of the scarring may be prevented or slowed. People with cirrhosis who have signs of complications may develop the end-stage liver disease (ESLD) and the only treatment at this stage is liver transplantation.

- **Stage 1 cirrhosis** involves some scarring of the liver, but few symptoms. This stage is considered compensated cirrhosis, where there are no complications.
- **Stage 2 cirrhosis** includes worsening portal hypertension and the development of varices.
- **Stage 3 cirrhosis** involves the development of swelling in the abdomen and advanced liver scarring. This stage marks decompensated cirrhosis, with serious complications and possible liver failure.
- **Stage 4 cirrhosis** can be life-threatening and people have developed the end-stage liver disease (ESLD), which is fatal without a transplant.

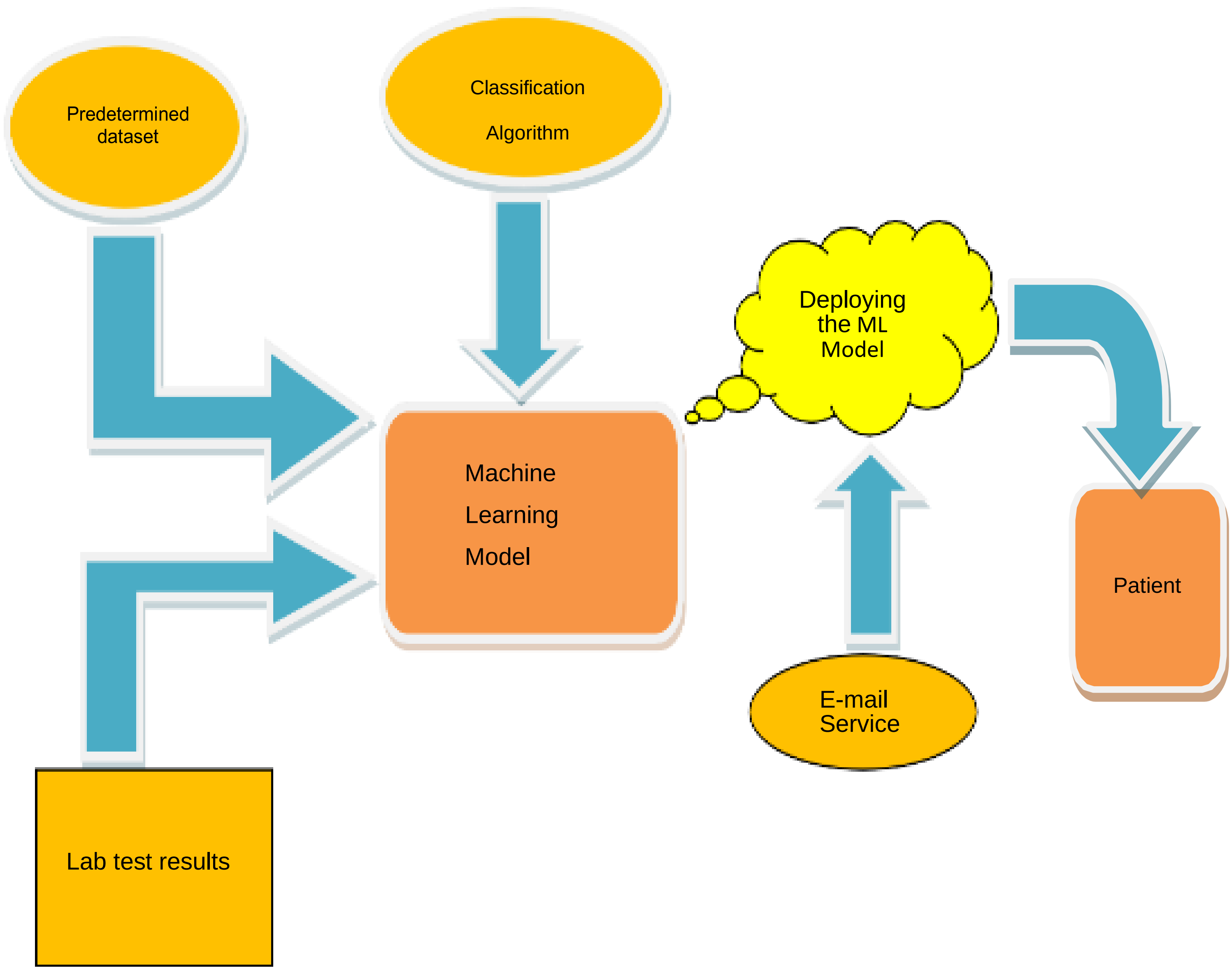
So in order to generate the model based on the problem stated above, we need to use Machine learning. Machine learning (ML) is the study of computer algorithms that improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. The accuracy of the model is determined by evaluation of the trained model on “testing dataset”. There are many IDEs for building the model using Python programming language. I have used the Google colaboratory for the same.

Firstly the dataset is prepared, in which data is extracted from the results of the patients who have undergone certain tests concerning lliver cirrhosis. This dataset includes results of the features mentioned earlier from the test and stored in the form of comma separated variable (csv) format. This dataset is imported into the google colaboratory environment. Python contains some in-built libraries such as pandas, scikit learn, Seaborn which is exclusively used for training machine learning algorithms.

Clinical tests are performed either at a clinic or at home. Data is inputted into a pathological ML system. A few minutes later, an email which is programmed in the deployment software with a detailed report that has an accurate prediction about the development of the disease is sent to the patient. Machine learning not only predicts cirrhosis prognosis results faster but also gives higher accuracy which is around 70 to 80 % which is greater when compared with the pathologists. So machine learning is a very useful technique which can be applied in replacing present prognosis processes.



Final product prototype:



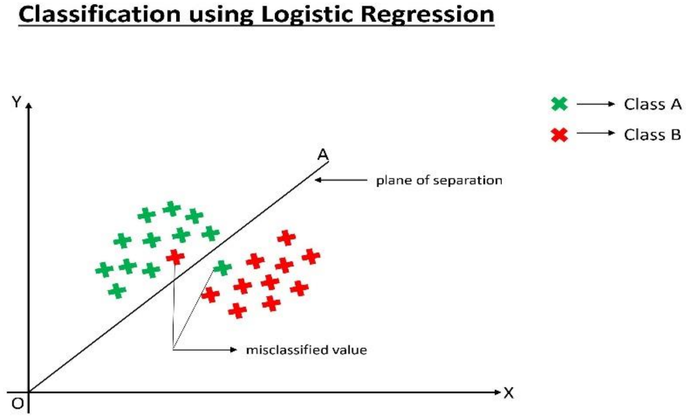
PRODUCT DETAILS:

It's a system which takes in data, finds patterns, trains itself using the data and displays the output. As discussed earlier, ML models have various advantages over the manual clinical results. Not only they are fast but also can repeat themselves thousands of times without getting exhausted. After every iteration, the machine repeats the process to do it better.

Machines have greater accuracy. With the advent of the Internet of Things technology, there is so much data out in the world that humans can't possibly go through it all. That's where machines help us. They can do work faster than us and make accurate computations and find patterns in data.

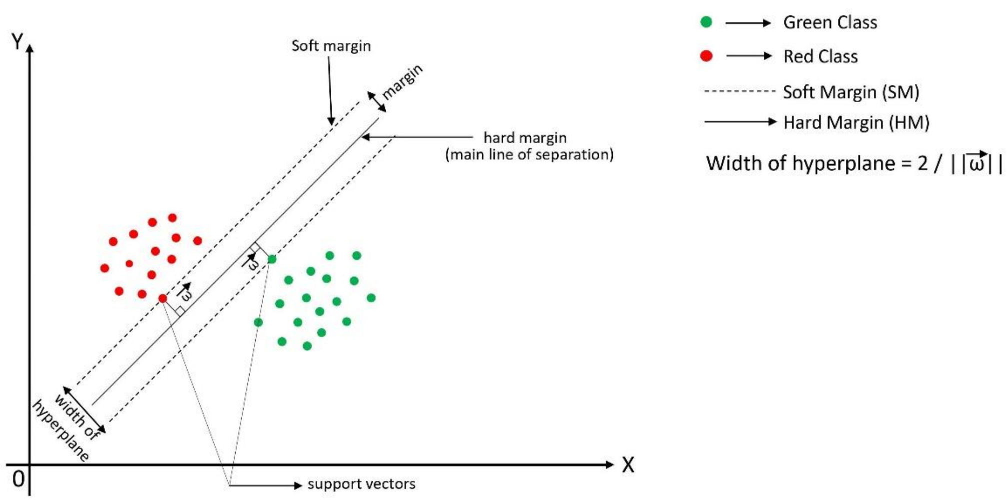
Various ML algorithms are used to classify the datasets into respective categories. Some of the classification algorithms used are:

1. **Logistic Regression:** Logistic regression is a classification algorithm which is used to predict a binary outcome based on a set of independent variables. A binary outcome is one where there are only two possible scenarios—either the event happens (1) or it does not happen (0). Independent variables are those variables or factors which may influence the outcome (or dependent variable). So: Logistic regression is the correct type of analysis to use when you're working with binary data. You know you're dealing with binary data when the output or dependent variable is dichotomous or categorical in nature; in other words, if it fits into one of two categories (such as "yes" or "no", "pass" or "fail", and so on).

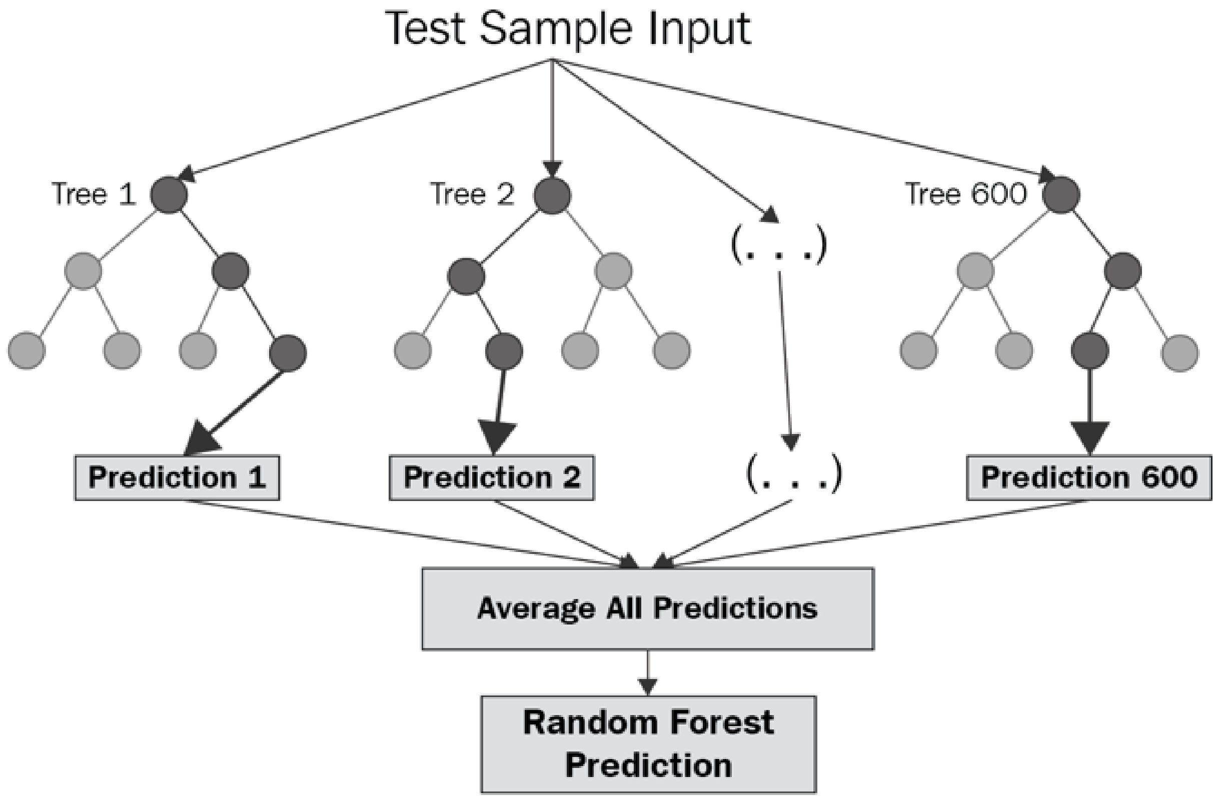


2. **Support Vector Machine:** The goal of an SVM algorithm is to classify data by creating a boundary with the widest possible margin between itself and the data. The advantages of support vector machines are: Effective in high dimensional spaces. Still effective in cases where number of dimensions is greater than the number of samples. Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient Versatile: different Kernel functions can be specified for the decision function.

Classification using Support Vector Machine



3. **Random Forest:** The fundamental concept behind random forest is a simple but powerful one — the wisdom of crowds. In data science speak, the reason that the random forest model works so well is: A large number of 16 relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models. The low correlation between models is the key. Just like how investments with low correlations (like stocks and bonds) come together to form a portfolio that is greater than the sum of its parts, uncorrelated models can produce ensemble predictions that are more accurate than any of the individual predictions. The reason for this wonderful effect is that the trees protect each other from their individual errors. While some trees may be wrong, many other trees will be right, so as a group the trees are able to move in the correct direction.



Python-libraries Used:

• Pandas: Pandas is a Python library used for working with data sets. It is used for analyzing, cleaning, exploring, and manipulating data. It is used for reading the csv file containing the dataset and converting it into a dataframe.

Syntax: import pandas as pd
df = pd.read_csv('File_name’)

• Scikit Learn: Scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library based on python programming language. It contains various classification, regression and clustering algorithms including (SVM) support vector machines, random forests, gradient boosting, k-means and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

Syntax for Logistic Regression: from sklearn.linear_model import LogisticRegression
Syntax for SVM: from sklearn.svm import SVC
Syntax for Random Forest: from sklearn.ensemble import RandomForestClassifier

• Seaborn: Seaborn is a library for making statistical graphics in Python. It is built on matplotlib and closely integrated with pandas data structures. This is a visualization tool used to demonstrate the datas.

Syntax: import seaborn as sns

Team required:

Algorithm with highest accuracy among classification algorithms is chosen as the best algorithm for Liver cirrhosis stage prediction. The manpower domain includes:

1. Machine learning engineering
2. Business analyst
3. Software developer
4. Cloud engineer
5. Data Researcher

Conclusion:

AI is set to change the medical industry in the coming decades — it wouldn't make sense for pathology to not be disrupted too. Currently, ML models are still in the testing and experimentation phase. As datasets are getting larger and of higher quality, researchers are building increasingly accurate models. AI is rapidly becoming an extremely promising aid in liver image tasks, leading to improved performance in detecting and evaluating liver lesions, facilitating liver clinical therapy, and predicting liver treatment response. In the future, the development of AI is inseparable from physicians and the work of physicians will be closely linked with AI. Machine-assisted medical services will be the optimal solution for future liver medical care. We need to determine which specific radiology tasks are most likely to benefit from the deep learning algorithm, taking into account the strengths and limitations of these algorithms. In the context of the rapid development of AI technology, physicians must keep pace with the times and apply technology rigorously in order to become a technology driver and better serve patients.

External Searches/ References:

[1] Poole D, Mackworth A, Goebel R. Computational Intelligence: A Logical Approach. New York: Oxford University Press, 1998

[2] J. Lata, “Chronic liver diseases as liver tumor precursors,” *Digestive Diseases*, vol. 28, no. 4-5, pp. 596–599, 2010.

[3] M. Basaranoglu, G. Basaranoglu, T. Sabuncu, and H. Sentürk, “Fructose as a key player in the development of fatty liver disease,” *World Journal of Gastroenterology*, vol. 19, no. 8, pp. 1166–1172, 2013.

[4] LOGISTIC REGRESSION: BINARY AND MULTINOMIA, 2014 Edition, by G. David Garson and Statistical Associates Publishing

[5] Han, J., Kamber, M., & Pei, J. (2012). Data Mining Concepts and Techniques (3rd ed). USA: Elsevier Inc.

[6] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6378542/

[7] https://www.shutterstock.com/search/cirrhosis

[8] https://www.hindawi.com/journals/ecam/2013/321234/

[9] https://dphi.tech/challenges/192/overview/about

GITHUB LINK: Project1_Code_file