IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124209

PREDICTION OF CIRRHOSIS (LIVER FAILURE) USING ML TECHNIQUE

Mrs.R.Lavanya[1], S.Naveen[2], R.Romil[3], E.Vasim[4]

[1]Assistant Professor, Department of Computer Science and Engineering, DMI College of Engineering, Chennai, India

[2]Department of Computer Science and Engineering, DMI College of Engineering,Chennai, India

[3]Department of Computer Science and Engineering, DMI College of Engineering,Chennai, India

[4]Department of Computer Science and Engineering, DMI College of Engineering,Chennai, India

Abstract-The function of the liver is to filter blood that circulates through the body, converting nutrients and drugs absorbed from the digestive tract into ready-to-use chemicals. The liver performs many other important functions, such as removing toxins and other chemical waste products from the blood and readying them for excretion. Liver failure begins in the cells of your liver. Nowadays machine learning is applied to healthcare systems where there is a chance of predicting the disease early. The main necessity of Artificial intelligence is data. The past dataset is collected and that dataset is used to build a machine-learning model. The necessary pre-processing techniques are applied like univariate analysis and bivariate analysis are implemented. The data is visualized for a better understanding of the features and based on that a classification model is built by using a machine learning algorithm and a comparison of algorithms is done based on their performance metrics like accuracy, F1 score recall, etc.

I.INTRODUCTION

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data and apply knowledge and actionable insights from data across a broad range of application domains. The term "data science" has been traced back to 1974, when Peter Naur proposed it as an alternative name for computer science. In 1996, the International Federation of Classification Societies became the first conference to specifically feature data science as a topic. However, the definition was still in flux. The term “data science” was first coined in 2008 by D.J. Patil, and Jeff Hammerbacher, the pioneer leads of data and analytics efforts at LinkedIn and Facebook. In less than a decade, it has become one of the hottest and most trending professions in the market. Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data. Data science can be defined as a blend of mathematics, business acumen, tools, algorithms, and machine learning techniques, all of which help us in finding out the hidden insights or patterns from raw data which can be of major use in the formation of big business decisions.

Data Scientist: Data scientists examine which questions need answering and where to find the related data. They have business acumen and analytical skills as well as the ability to mine, clean, and present data. Businesses use data scientists to source, manage, and analyze large amounts of unstructured data.

A. Artificial intelligence

Artificial intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving. Artificial intelligence (AI) is intelligence demonstrated by machines, as opposed to the natural intelligence displayed by humans or animals. Leading AI textbooks define the field as the study of "intelligent agents" any system that perceives its environment and takes actions that maximize its chance of achieving its goals. Some popular accounts use the term "artificial intelligence" to describe machines that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem-solving", however, this definition is rejected by major AI researchers. Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems. Specific applications of AI include expert systems, natural language processing, speech recognition, and machine vision. AI applications include advanced web search engines, recommendation systems (used by Youtube, Amazon, and Netflix), Understanding human speech (such as Siri or Alexa), self-driving cars (e.g. Tesla), and competing at the highest level in strategic game systems (such as chess and Go), As machines become increasingly capable, tasks considered to require "intelligence" are often removed from the definition of AI, a phenomenon known as the AI effect. For instance, optical character recognition is frequently excluded from things considered to be AI, having become a routine technology. Artificial intelligence was founded as an academic discipline in 1956 and in the years since has experienced several waves of optimism, followed by disappointment and the loss of funding (known as an "AI winter"), followed by new approaches, success, and renewed funding. AI research has tried and discarded many

DOI: 10.17148/IJARCCE.2023.124209

different approaches during its lifetime, including simulating the brain, modeling human problem-solving, formal logic, large databases of knowledge, and imitating animal behavior. In the first decades of the 21st century, highly mathematical statistical machine learning has dominated the field, and this technique has proved highly successful, helping to solve many challenging problems throughout industry and academia. The various sub-fields of AI research are centered around particular goals and the use of particular tools. The traditional goals of AI research include reasoning, knowledge representation, planning, learning, natural language processing, perception, and the ability to move and manipulate objects. General intelligence (the ability to solve an arbitrary problem) is among the field's long-term goals. To solve these problems, AI researchers use versions of search and mathematical optimization, formal logic, artificial neural networks, and methods based on statistics, probability, and economics. This raises philosophical arguments about the mind and the ethics of creating artificial beings endowed with human-like intelligence. These issues have been explored by myth, fiction, and philosophy since antiquity. Science fiction and futurology have also suggested that, with its enormous potential and power, AI may become an existential risk to humanity. As the hype around AI has accelerated, vendors have been scrambling to promote how their products and services use AI. Often what they refer to as AI is simply one component of AI, such as machine learning. AI requires a foundation of specialized hardware and software for writing and training machine learning algorithms. No one programming language is synonymous with AI, but a few, including Python, R, and Java, are popular. In general, AI systems work by ingesting large amounts of labeled training data, analyzing the data for correlations and patterns, and using these patterns to make predictions about future states. In this way, a chatbot that is fed examples of text chats can learn to produce life-like exchanges with people, or an image recognition tool can learn to identify and describe objects in images by reviewing millions of examples.AI programming focuses on three cognitive skills: learning, reasoning, and self-correction.

Learning processes. This aspect of AI programming focuses on acquiring data and creating rules for how to turn the data into actionable information. The rules, which are called algorithms, provide computing devices with step-by-step instructions for how to complete a specific task.
Reasoning processes. This aspect of AI programming focuses on choosing the right algorithm to reach a desired outcome.
Self-correction processes**.** This aspect of AI programming is designed to continually fine-tune algorithms and ensure they provide the most accurate results possible.AI is important because it can give enterprises insights into their operations that they may not have been aware of previously and because, in some cases, AI can perform tasks better than humans. Particularly when it comes to repetitive, detail-oriented tasks like analyzing large numbers of legal documents to ensure relevant fields are filled in properly, AI tools often complete jobs quickly and with relatively few errors.Artificial neural networks and deep learning artificial intelligence technologies are quickly evolving, primarily because AI processes large amounts of data much faster and makes predictions more accurately than humanly possible.

B. Natural Language Processing (NLP)

Natural language processing (NLP) allows machines to read and understand human language. A sufficiently powerful natural language processing system would enable natural-language user interfaces and the acquisition of knowledge directly from human-written sources, such as newswire texts. Some straightforward applications of natural language processing include information retrieval, text mining, question answering, and machine translation. Many current approaches use word co-occurrence frequencies to construct syntactic representations of text. "Keyword spotting" strategies for search are popular and scalable but dumb; a search query for "dog" might only match documents with the literal word "dog" and miss a document with the word "poodle". "Lexical affinity" strategies use the occurrence of words such as "accident" to assess the sentiment of a document. Modern statistical NLP approaches can combine all these strategies as well as others, and often achieve acceptable accuracy at the page or paragraph level. Beyond semantic NLP, the ultimate goal of "narrative" NLP is to embody a full understanding of commonsense reasoning. By 2019, transformer-based deep learning architectures could generate coherent text.

C. Machine learning

Machine learning is to predict the future from past data. Machine learning (ML) is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data and the basics of Machine Learning, implementation of a simple machine learning algorithm using Python. The process of training and prediction involves the use of specialized algorithms. It feeds the training data to an algorithm, and the algorithm uses this training data to give predictions on new test data. Machine learning can be roughly separated into three categories. There are supervised learning, unsupervised learning, and reinforcement learning. A supervised learning program is both given the input data and the corresponding labeling to learn data that has to be labeled by a human being beforehand. Unsupervised learning is no labels. It provided the learning algorithm. This algorithm has to figure out the clustering of the input data. Finally, Reinforcement learning dynamically interacts with its environment and it receives positive or negative feedback to improve its performance. Data scientists use many different kinds of machine learning algorithms to discover patterns in

DOI: 10.17148/IJARCCE.2023.124209

Python that lead to actionable insights. At a high level, these different algorithms can be classified into two groups based on the way they "learn" about data to make predictions: supervised and unsupervised learning. Classification is the process of predicting the class of given data points. Classes are sometimes called targets/ labels or categories.

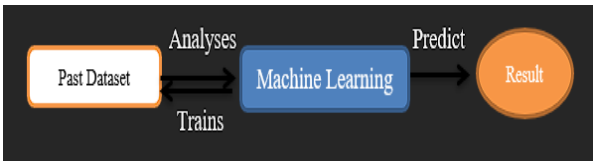


Fig1.1: Process of Machine Learning

Supervised Machine Learning **is the** majority of practical machine learning uses supervised learning. Supervised learning is where input variables (X) and an output variable (y) and use an algorithm to learn the mapping function from the input to the output **is y = f(X).** The goal is to approximate the mapping function so well that when you have new input data (X) that you can predict the output variables (y) for that data. Techniques of Supervised Machine Learning algorithms include **logistic regression, multi-class classification, Decision Trees** and **support vector machines, etc**. Supervised learning requires that the data used to train the algorithm is already labeled with correct answers. Supervised learning problems can be further grouped into **Classification** problems. This problem has as its goal the construction of a succinct model that can predict the value of the dependent attribute from the attribute variables. The difference between the two tasks is the fact that the dependent attribute is numerical for categorical classification. A classification model attempts to draw some conclusions from observed values. Given one or more inputs a classification model will try to predict the value of one or more outcomes. A classification problem is when the output variable is a category, such as "red" or "blue".

II.LITERATURE SURVEY

A. General

A literature review is a body of text that aims to review the critical points of current knowledge on and/or methodological approaches to a particular topic. It is a secondary source that discusses published information in a particular subject area and sometimes information in a particular subject area within a certain period. Its ultimate goal is to bring the reader up to date with current literature on a topic and forms the basis for another goal, such as future research that may be needed in the area and precedes a research proposal and maybe just a simple summary of sources. Usually, it has an organizational pattern and combines both summary and synthesis. A summary is a recap of important information about the source, but a synthesis is a re-organization or reshuffling of information. It might give a new interpretation of old material or combine new with old interpretations or it might trace the intellectual progression of the field, including major debates. Depending on the situation, the literature review may evaluate the sources and advise the reader on the most pertinent of them. Loan default trends have been long studied from a socio-economic standpoint. Most economics surveys believe in empirical modeling of these complex systems to be able to predict the loan default rate for a particular individual. The use of machine learning for such tasks is a trend that it is observing now. Some of the surveys are to understand the past and present perspectives of loan approval or not.

B. Related Work

The Diagnosis of Chronic Liver Disease using Machine Learning Techniques Golmei Shaheamlung, Harpreet Kaur (26-March-2021). In the 21st century, the issue of liver disease has been increasing all over the world. As per the latest survey report, the liver disease death toll has been rising by approximately 2 million per year worldwide. The overall percentage of death by liver disease is 3.5% worldwide. Chronic Liver disease is also considered to be one of the deadly diseases, so early detection and treatment can recover the disease easily. The hidden knowledge of liver disease is recognized and extracted using a historical liver disease database. The complex queries are responded to to diagnose liver disease The proposed model was improved by applying a combination of three classifiers, Logistic regression, Random forest, and KNN algorithm. Python is employed for the implementation of the suggested model and the result proved accuracy that is achieved at 77.58 percent. In the future, the execution of the clustering algorithm is performed with the hybrid classifier technique for the division of data. Prediction and Analysis of Liver Disorder Diseases by Using Data Mining Technique Shambel Kefelegn, Pooja Kamat (2018) Liver disorder diseases one of the major diseases in the world, Liver is one of the huge solid organs in the human body; and is also considered a gland because, among its many functions, it makes and secretes bile. The liver theatres a vital role in many physical functions from protein manufacture and blood clotting to fat, sugar, and iron metabolism. Liver disorder diseases are any trouble of the liver purpose that reason for sickness.The study paper is to predict and analyze liver disorder diseases to produce better performance accuracy by comparing various data mining classification algorithms and the performance of the accuracy is measured by confusion matrices. Decision Tree considered for performance evaluation in liver disorder diseases prediction Future work we can use the Hybrid approach to get better performance accuracy for liver disorder diseases prediction with their suitable data sets. Liver Disease

DOI: 10.17148/IJARCCE.2023.124209

Prediction by Using Different Decision Tree Techniques Nazmun Nahar and Ferdous Ara (2018) Early prediction of liver disease is very important to save human life and taking proper steps to control the disease. Decision Tree algorithms have been successfully applied in various fields, especially in medical science. This research work explores the early prediction of liver disease using various decision tree techniques. The liver disease dataset which is selected for this study is consisting of attributes like total bilirubin, direct bilirubin, age, gender, total proteins, albumin, and globulin ratio. The main purpose of this work is to calculate the performance of various decision tree techniques and compare their performance. The study employed some decision tree algorithms such as J48, LMT, Random Forest, Random tree, REPTree, Decision Stump, and Hoeffding Tree to predict liver disease at an earlier stage. This algorithm gives various results based on Accuracy, Mean Absolute Error, Precision, Recall, Kappa statistics, and Runtime. These techniques were evaluated and their performance was compared. From the analysis. The results of this study will encourage us to continue developing other advanced decision trees such as CART. A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms A.K.M Sazzadur Rahman, F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Joy Roy, Syed Akhter Hossain : (11, NOVEMBER 2019) Chronic Liver Disease is the leading cause of global death that impacts the massive quantity of humans around the world. This disease is caused by an assortment of elements that harm the liver. For example, obesity, undiagnosed hepatitis infection, and alcohol misuse. Which is responsible for abnormal nerve function, coughing up or vomiting blood, kidney failure, liver failure, jaundice, liver encephalopathy and there are many more. This disease diagnosis is very costly and complicated. Therefore, the goal of this work is to evaluate the performance of different Machine Learning algorithms to reduce the high cost of chronic liver disease diagnosis by prediction. In this work, we used six algorithms of Logistic Regression. We just explored some popular supervised machine learning algorithms, more algorithms can be picked to assemble an increasingly precise model of liver disease prediction and performance can be progressively improved. Additionally, this work is likewise ready to assume a significant role in health care research and just as restorative focuses to anticipate liver infection. Performance Evolution of Different Machine Learning Algorithms for Prediction of Liver Disease Muktevi Srivenkatesh (December 2019) Liver malady is an overall medical issue that is related to different inconveniences and high mortality. It is of basic significance that illness be recognized before such huge numbers of these lives can be spared. The phases of liver ailment are a significant viewpoint for focused treatment. It is a troublesome undertaking for therapeutic analysts to foresee the disease inside the beginning times on account of sensitive manifestations. Generally, the side effects become evident once it's past the point of no return. To beat this issue, we have a liver infection forecast. Liver sickness might be distinguished with incalculable order systems, and these have been classified the utilization forecast of several highlights and classifier blends. In the end, the use of information-digging systems for prescient examination is significant in the well-being field since it enables us to confront ailments prior and accordingly spare individuals' lives through the expectation of a fix. In this work, we utilized a few learning calculations K-Nearest Neighbour, Support Vector Machines, Logistic Regression, Navi Bayes, and Random Forest to foresee patients with constant liver disappointment infection and patients who are not experiencing this illness. Re-enactment results demonstrated that the Logistic regression classifier demonstrated its exhibition in foreseeing with best outcomes regarding precision and least execution time.

C. Summary

A summary is a recap of important information about the source, but a synthesis is a re-organization or reshuffling of information. It might give a new interpretation of old material or combine new with old interpretations or it might trace the intellectual progression of the field, including major debates. Depending on the situation, the literature review may evaluate the sources and advise the reader on the most pertinent of them. Loan default trends have been long studied from a socio-economic standpoint. Most economics surveys believe in empirical modeling of these complex systems to be able to predict the loan default rate for a particular individual. The use of machine learning for such tasks is a trend that it is observing now. Some of the surveys to understand the past and present perspectives of loan approval or not.

III.PROPOSED SYSTEM

Liver failure is a condition or disease that happens when normal cells in the liver become abnormal in appearance and behavior. The failure cells can then become destructive to adjacent normal tissues and can spread both to other areas of the liver and to organs outside the liver. The proposed method is a machine learning model based on the past data of liver failure the features and target column were identified first using our domain knowledge related to health care. Then dataset is viewed for a better understanding of features and then the dataset is split into two parts normally in a 7:3 ratio where the data is used for training and testing. The algorithm is applied to the trained data to get a better understanding of the features and a classification model is built based on their learning different algorithms are compared and performance is measured and compared using their performance metrics.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 4, April 2023

DOI: 10.17148/IJARCCE.2023.124209

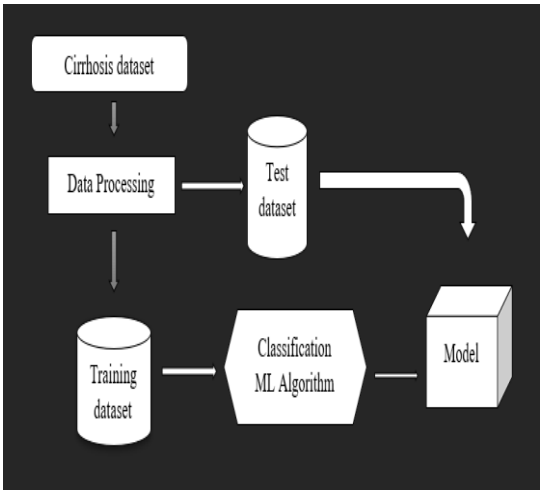
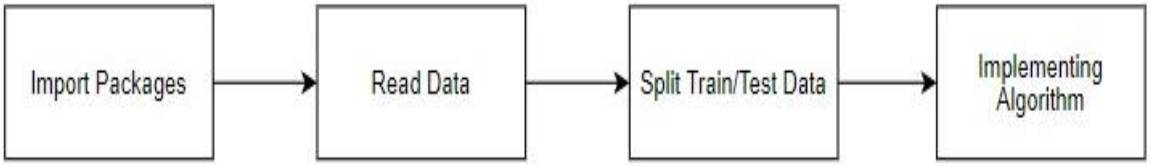


Fig: 4.1 Proposed system diagram

IV.DATA PRE-PROCESSING

Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered as close to the true error rate of the dataset. If the data volume is large enough to be representative of the population, you may not need the validation techniques. However, in real-world scenarios, to work with samples of data that may not be a true representative of the population of a given dataset. To find the missing value, duplicate value, and description of data type whether it is float variable or integer. The sample of data is used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. The evaluation becomes more biased as a skill on the validation dataset is incorporated into the model configuration. The validation set is used to evaluate a given model, but this is for frequent evaluation. It as machine learning engineers use this data to fine-tune the model hyperparameters. Data collection, data analysis, and the process of addressing data content, quality, and structure can add up to a time-consuming to-do list. During the process of data identification, it helps to understand your data and its properties; this knowledge will help you choose which algorithm to use to build your model.



Module Diagram

V.CONCLUSION AND FUTURE WORK

The analytical process started with data cleaning and processing, missing value, exploratory analysis, and finally model building and evaluation. The best accuracy on a public test set is a higher accuracy score is will find out. This project can help to find the Liver Failure stage based on the patient's health.

ACKNOWLEDGEMENT

Acknowledgment is a section where the writer thanks and expresses gratitude to everyone who assisted with the project. In research project writing, acknowledgment is often included to recognize and appreciate everyone who participated in the research.

REFERENCES

[1]A.Smith et al., "Multi-wavelength spectrophotometric analysis for detection of xanthochromia in cerebrospinal fluid and accuracy for the diagnosis of subarachnoid hemorrhage," Clin. Chim Acta, vol. 424, pp. 231-236, Sep.2013.

[2] C.C. Zabetta et al., "Bilistick: A low-cost point-of-care system to measurement total plasma bilirubin." Neonatology, vol. 103, no. 3, pp. 177-181, 2013.

[3]D.Jayaweera et al., "Chloroform ingestion causing severe gastrointestinal injury, hepatotoxicity and dermatitis

DOI: 10.17148/IJARCCE.2023.124209

confirmed with plasma chloroform concentrations," Clin. Toxicol., vol. 55, no. 2, pp. 147-150, Oct. 2017.
[4]F. Belfiore and H.J. Zimmerman, "Cytotoxicity of chlorinated hydrocarbons in vitro observations on chloroform-induced hemolysis," in Proc.Soc. Exp. Biol. Med., vol. 134, no. 1, pp. 61-66, May 1970.
[5]G.M. Hirschfield and G.J. Alexander, "Gilbert's syndrome: An overview for clinical biochemists," Ann. Clin. Biochem., vol. 43, no. Pt 5, pp.340-343, Sep. 2006.
[6]H.J> Kim and H.W. Lee, "Important predictor of mortality in patients with end-stage liver disease," Clin.Mol. Hepatol., vol.19,no.2, pp. 105-115, Jun. 2013.
[7]H.P. Sennels et al., "Diurnal variation of hematology parameters in healthy young males: The bispebjerg study of diurnal variations," Scand.J. Clin.Lab. Investigatigation,vol.71,no.7, pp.532- 541, Oct.2011.
[8]I.Yamanouchi et al.," Transcutaneous bilirubinometry: Preliminary studies of noninvasive transcutaneous bilirubin meter in the Okayama National hospital," Pediatrics, vol.65, no.2, pp.195- 202, Feb.1980.
[9]J.A. Lopez-Velazquez et al.," Bilirubin alone as a biomarker for short-term mortality in acute-on-chronic liver failure: An important prognostic indicator," Ann. Hepatol., vol.13, no.1,pp.98- 104,Jan./Feb.2014.
[10]J.Yao and L.V.Wang, Photoacoustic microscopy," Laser Photon.Rev., vol.7, no.5, pp. 758-778, Jan.2013.

