



(Company No. 101067-P)

الجامعة الإسلامية العالمية ماليزيا
INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA
يُونِيسَيْتِي إِسْلَامُ، إِنْتَارَا بَغْسِيَا مَلَيْسِيَا

Garden of Knowledge and Virtue

CSCI 4340 MACHINE LEARNING COURSE

Dr. Amelia Ritahani Bt. Ismail

Assignment 2 Report:

HEPATITIS DISEASE DETECTION BY HYBRID MODEL MACHINE LEARNING

By

Mohamed Moubarak Mohamed Misbahou Mkoubi 1820705

Karim Fazlul 1822049

Mohammad Afif Muhajir 1816203

ABSTRACT

With the growth of technology and its rapid advancement in recent years many fields of activities have been taking advantage of said facts. Including the ever dire task of medical needs such as identifying diseases. Up to this moment, many kinds of diseases have been recorded, one prominent kind that most people have come to know is "Hepatitis".

In recent news, Hepatitis disease has been brought up because of the rise of cases especially in children, that are thought to be just a common cold, but with deeper investigation and study of cases they have high similarities and due to that accurate identification of this disease is of utmost importance. With that in mind, advanced computation has provided us with ease of trouble with the introduction of Machine Learning to help with identifying patterns and traits that can help overcome this trouble.

Keywords: Technology, Hepatitis, Disease Identification, Machine Learning

1. Introduction

The rise of technological advancement has been made known throughout these last few years, it can be noted by the number of discoveries made in those said years. According to a statement shown by the online World Economic Forum, "Technology has changed major sectors over the past 20 years, including media, climate action, and healthcare..." (*Here's How Technology Has Changed the World Since 2000*, 2020). With this in mind, many experts around the world, especially in the medical field, are eager to take full advantage of it. In the world of medical research, the growth of technology has been used to ease the process of disease identification through patterns and other factors through the means of machine learning.

As one report has stated,

Nowadays, humans face various diseases due to the current environmental conditions and their living habits. The identification and prediction of such diseases at their earlier stages are very important, to prevent the extremity of it. It is difficult for doctors to manually identify the diseases accurately most of the time. The goal of this paper is to identify and predict the patients with more common chronic illnesses. This could be achieved by using cutting-edge machine learning technology to ensure that this categorization reliably identifies persons with chronic diseases. (Alanazi, 2022, para. 1)

With the fact shown that many diseases are often too much to handle by manual

identification by limited human capabilities, with the help of automation provided by advanced computations, these kinds of hurdles are easier to overcome.

Amidst the numerous types of disease today each with its unique traits, some are even more prominent than the others, One prime example is Hepatitis disease. Hepatitis is usually the result of a viral infection or liver damage caused by alcohol, and symptoms include yellowing of the eyes and skin (jaundice), loss of appetite, a high temperature, dark urine, and feeling unusually tired all the time (*What Are the Symptoms of Hepatitis and Can It Be Treated?* 2022).

Reports have shown that Hepatitis cases have been on the rise, especially in children, such as one that is shown by the online news website Imperial College London, "An unexpected increase in hepatitis cases among young, previously healthy children was first reported in the United Kingdom in late March. As of 10 May, the World Health Organization (WHO) estimates that there are 348 cases globally." (Timmins & Head, 2022).

Some concerns also have been set forth about legal views regarding potential delays in identifying Hepatitis symptoms (Lovett, 2022). As with every disease, Hepatitis detection and identification still mostly relies on direct medical checkups, but by taking advantage of technological growth such as advanced computation in the form of machine learning which has been pointed out before, the process can be made easier. With that in mind, we want to set up a project of applying the method of disease identification through machine learning, especially deep learning by using hybrid algorithms as our model.

2. Technical Report

In this section of this report, the elaboration of computation methods, dataset and its preparation, algorithms, and model description will be pointed out.

A. Computation Methods

For this specific project, we were tasked with experimenting with a machine learning/deep learning branch of computation named the “hybrid algorithm” model.

A hybrid algorithm is one that mixes two or more different algorithms that solve the same issue, and is commonly used in programming languages such as Python, either selecting one (depending on the input) or switching between them during the method. This is usually done to combine the desirable qualities of each such that the total method outperforms the separate components.

The term "hybrid algorithm" does not just refer to integrating numerous algorithms to address a particular issue. Many algorithms may be thought of as combinations of simpler parts, but only when it comes to mixing algorithms that answer the same issue but differ in other ways, most notably in performance.

In this project specifically, we will be using a type of hybrid algorithm technique which is called “Blending Algorithm”.

B. Dataset

For this project, we will be experimenting with electronic health records on Hepatitis disease as our dataset. This dataset has 20 attributes and 142 instances of data. Details of the dataset can be seen inside the table below:

Attribute	Description
Class	Integer (1= Die, 0=Live)
AGE	Integer value
STEROID	Integer (1= Yes, 0=No)
ANTIVIRALS	Integer (1= Yes, 0=No)

FATIGUE	Integer (1= Yes, 0=No)
MALAISE	Integer (1= Yes, 0=No)
ANOREXIA	Integer (1= Yes, 0=No)
LIVER BIG	Integer (1= Yes, 0=No)
LIVER FIRM	Integer (1= Yes, 0=No)
SPLEEN PALPABLE	Integer (1= Yes, 0=No)
SPIDERS	Integer (1= Yes, 0=No)
ASCITES	Integer (1= Yes, 0=No)
VARICES	Integer (1= Yes, 0=No)
BILIRUBIN	Float, continuous value
ALK PHOSPHATE	Integer, real number
SGOT	Integer, real number
ALBUMIN	Float, continuous value
PROTIME	Integer, real number
HISTOLOGY	Integer (1= Yes, 2=No)

C. Algorithm Explanation

Blending Algorithm is a kind of ensemble machine learning algorithm. It uses a machine learning model to learn how to best combine the predictions from multiple contributing ensemble member models.

Blending is the same as “stacked generalization”, and as such has a similar

process and architecture, which involves two or more base models, often referred to as models, and a meta-model that combines the predictions of the base models referred to as a level-1 model. The meta-model is trained on the predictions made by base models on out-of-sample data.

Nevertheless, blending has specific connotations for how to construct a stacking ensemble model.

Blending may suggest developing a stacking ensemble where the base models are machine learning models of any type, and the meta-model is a linear model that “*blends*” the predictions of the base models. (Brownlee, 2020)

D. Data Preprocessing

Normalization: We have applied Min-Max normalization on numeric columns to reshape their distribution as close as possible to normal distribution because machine learning algorithms in general perform well on data that are naturally distributed.

Feature Engineering: In the data set age is given as a numeric value. We have converted it into categorical type to get better insight from this feature and then encode it using one-hot encoding so that our model can utilize this feature efficiently.

Several columns in the data set have discrete integer values which indicate that they are categorical values, we have encoded them using the pandas `get_dummies` method which converts categorical values into indicator variables.

The dataset is clean as there are no missing values in the dataset

level-0

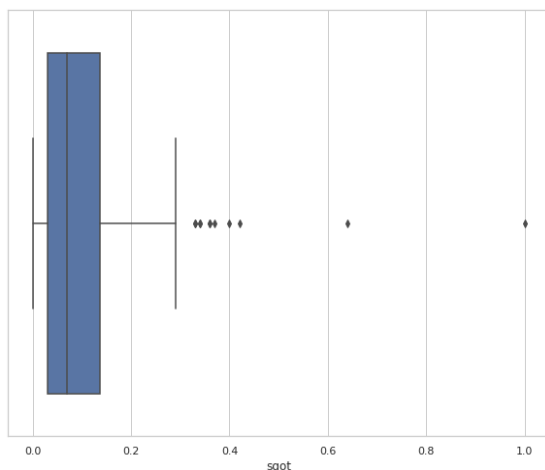
By the end of preprocessing we have created 14 new features and the final summary stat is given below

Experimental Setup: Our Experimental Setup includes

- Data Preparation
- Model Development: Dataset Splitting for Cross Validation-> Model Training -> Model Testing
- Evaluation: Our metric of evaluation will be by comparing our worked hybrid algorithm with a chosen available machine learning algorithm.

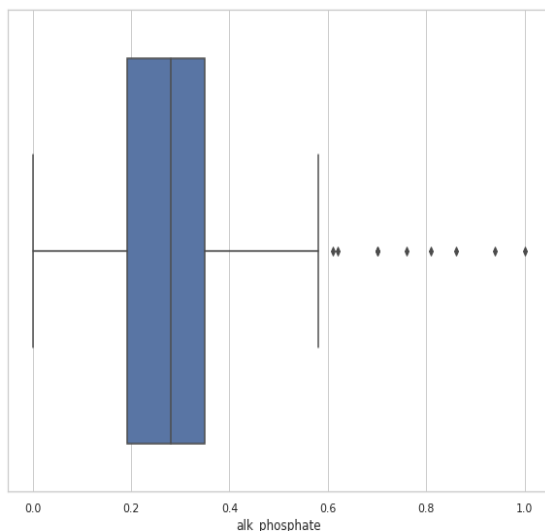
3. Implementation and Processes

The process of how our program model handles data will be clarified in this section of the report. This will include the process of creating the model coding, training, and testing the model before analyzing the result.



Data Normalization: Distribution of data has a significant effect on the ML model as most of the machine learning model assumes that data are in a normal distribution, data without normal distribution can cause the model to learn the wrong feature because of the variance difference of such feature. We have used standard normalization to normalize our data and it is mathematically expressed as

$$Z = (x - \mu) / \sigma$$



Outlier Removal: Outliers are data points that are 3 standard deviations apart from the mean. These values can have a significant influence on our model performance which generally causes bad performance, as these values are not common in natural events, and most of the time they are caused by mistakes while data is being prepared.

We have used a numpy built-in function to create a mask of data points that lies 3 standard deviations away from the mean then filtered out the data point using that mask to remove them.

Figure- Outlier Plot

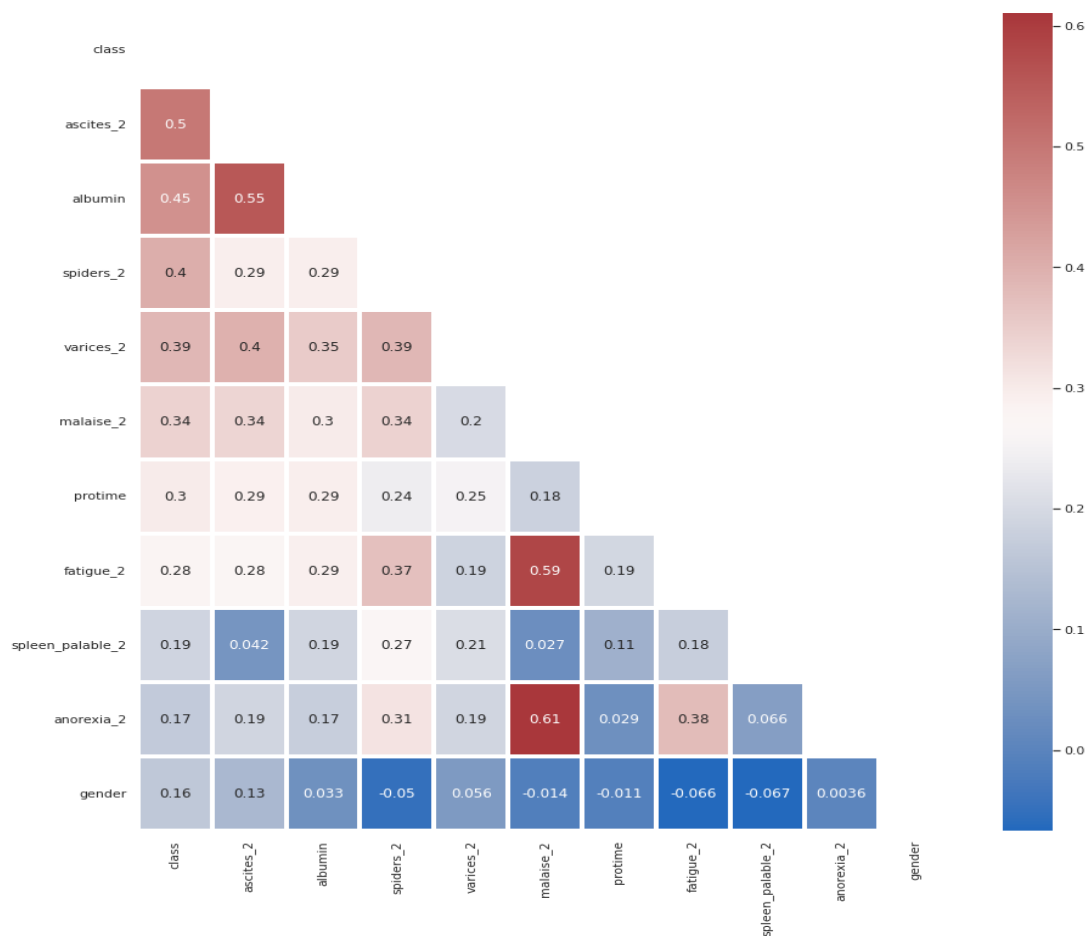


Figure- Correlation Plot

Feature Reduction by Correlation Value

The above figure represents the distribution of correlation of attributes with another attribute in a pair. The values shown in the boxes represent the level of dependency, in other words, correlations between two attributes(i.e. correlation value of attribute spleen_palpable_2 and albumin is shown on the box with coordinates (3,3); which means three boxes from the bottom and three boxes from the left that shows the value of 0.19).

From this figure, it can be concluded that Ascites have the highest correlation while gender has the lowest correlation with the predicted label. Ascites and albumin are highly correlated with each other so we can eliminate one of them for feature reduction.

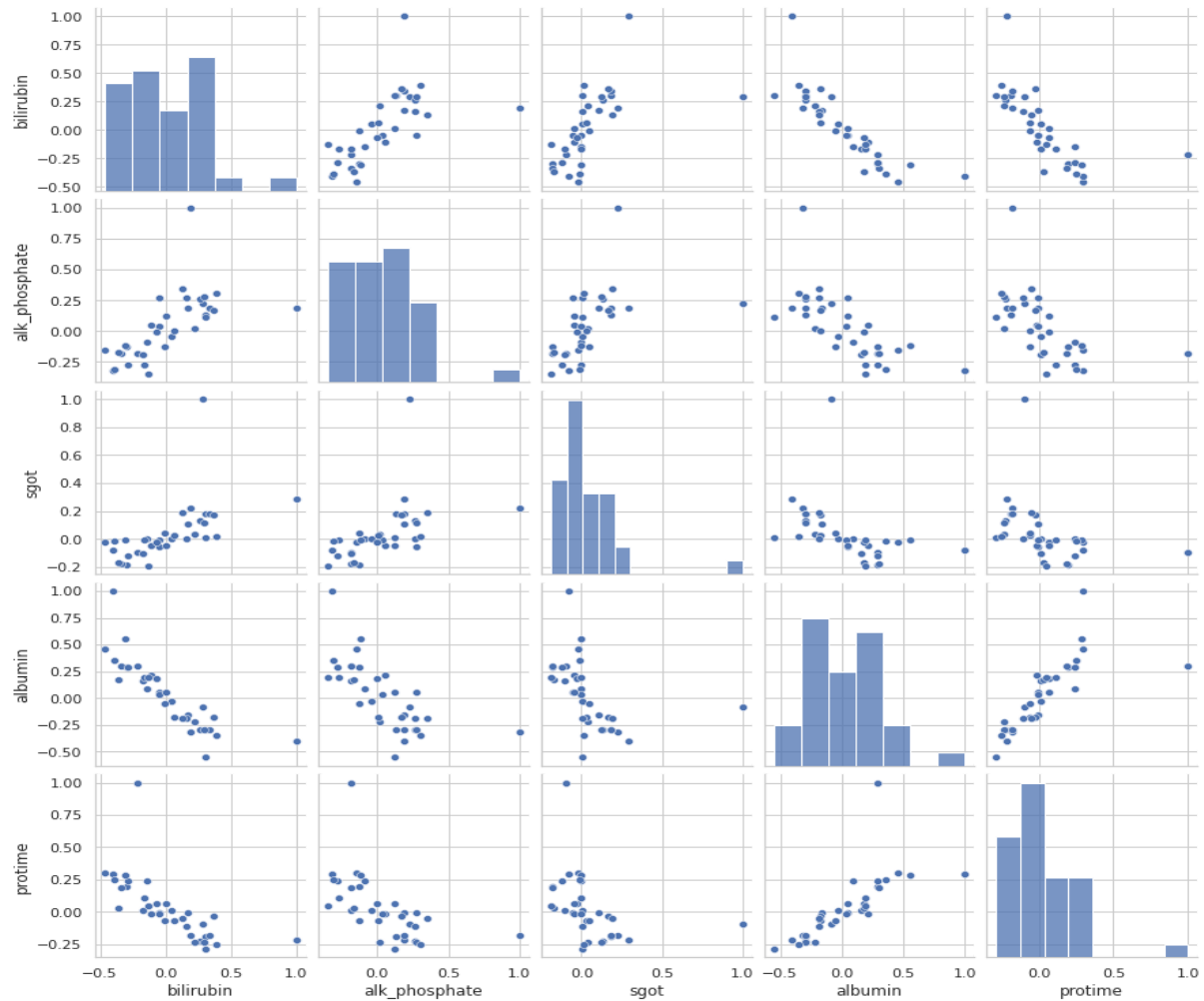
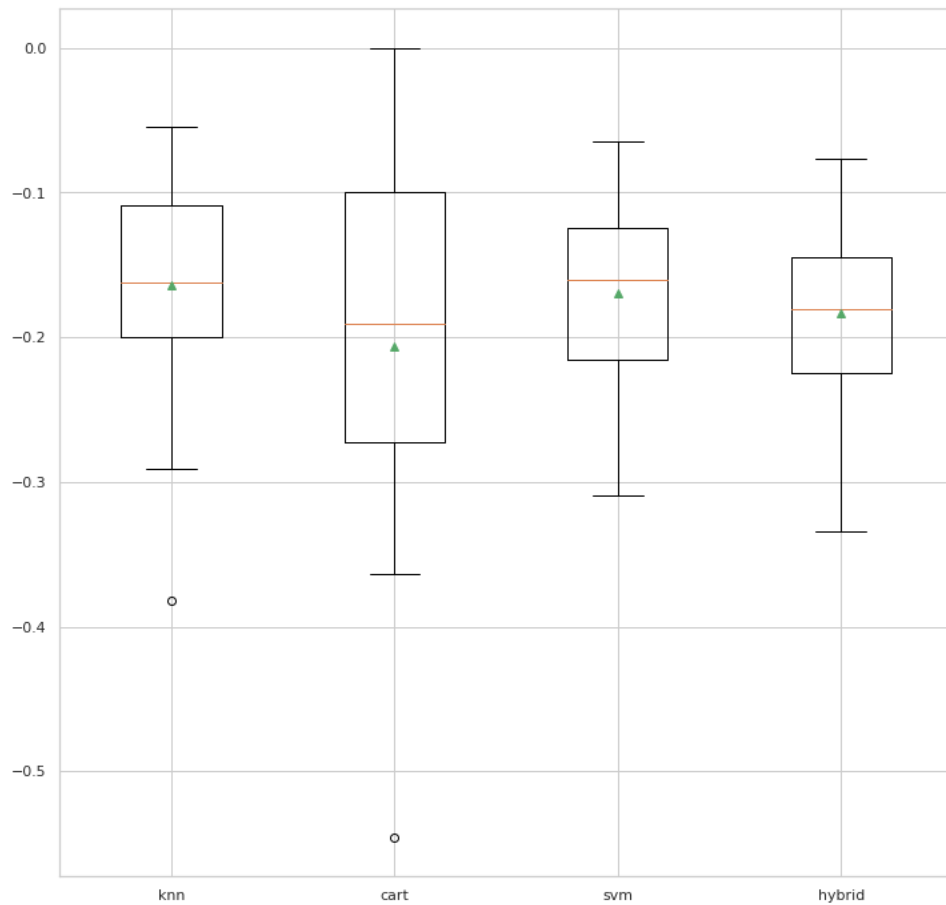


Figure- Linear Relationship between top five features in the data set

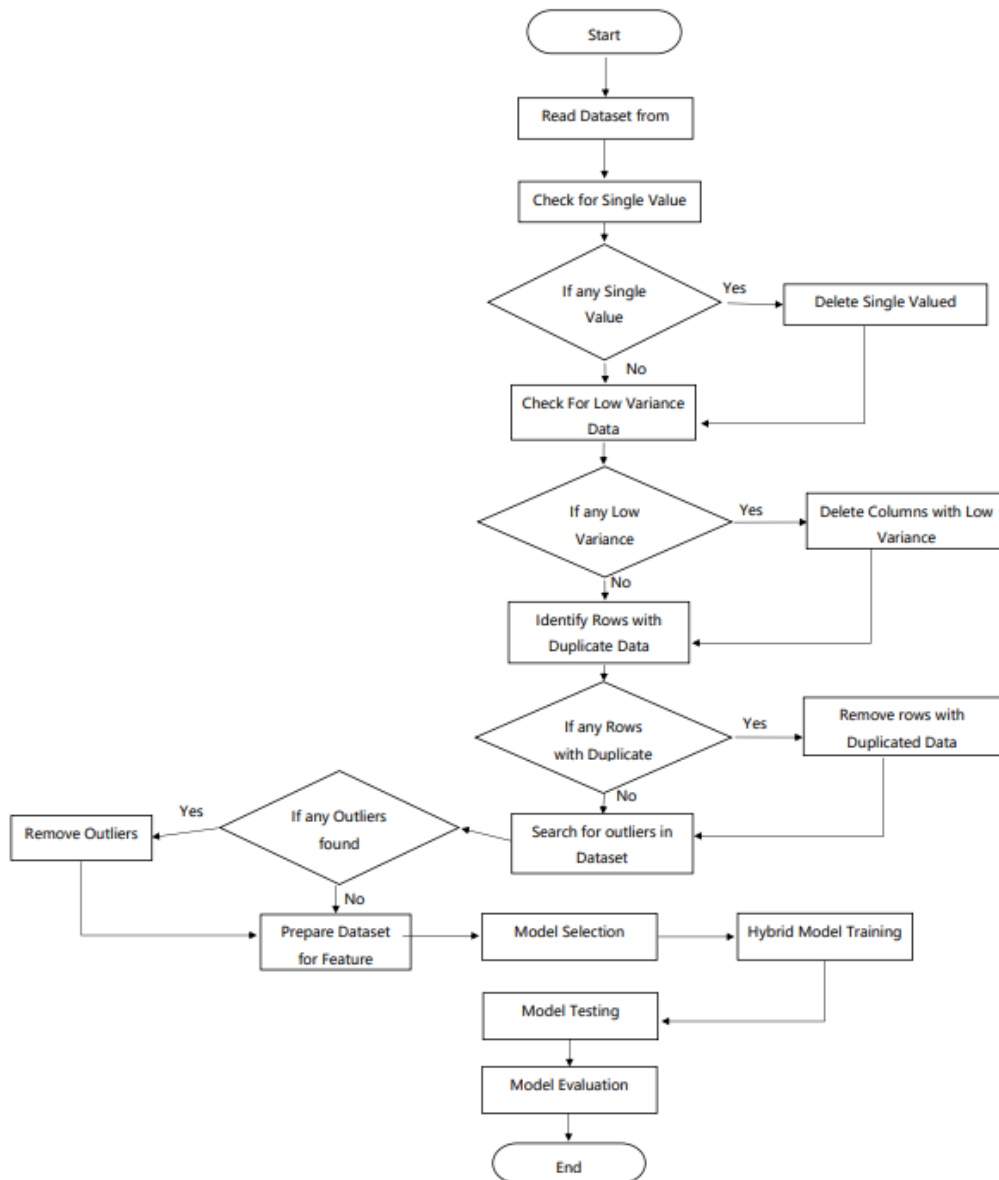
This graph represents the correlation of the top 5 features (based on statistical analysis) of the data set. The diagonal components of the figure represent the relation with the feature itself, that is why the shape of the graph appears to be in natural form. Protamine has a negative correlation with bilirubin while a strong correlation with albumin. The relationship between sgot and protamine appears to be neutral.



A cross-validation is a popular approach for evaluating a machine learning model's generalization performance. In this graph, the average of the support vector machines(SVM) has a high cross-validation performance compared to our hybrid model. The last one has almost the same result as our hybrid model. Therefore, the amount of the data in this dataset is not enough for a complex model, and that leads to our hybrid being underfitting.

Flowchart

FLOWCHART FOR Hepatitis Detection Deep Learning



References

Brownlee, J. (2020, November 30).

Blending Ensemble Machine

Learning With Python. Machine

Learning Mastery. Retrieved June

11, 2022, from

[https://machinelearningmastery.co](https://machinelearningmastery.com/blending-ensemble-machine-learning-with-python/)

[m/blending-ensemble-machine-lea](https://machinelearningmastery.com/blending-ensemble-machine-learning-with-python/)

[rning-with-python/](https://machinelearningmastery.com/blending-ensemble-machine-learning-with-python/)

Here's how technology has changed the

world since 2000. (2020,

November 18). The World

Economic Forum. Retrieved June

4, 2022, from

[https://www.weforum.org/agenda/2](https://www.weforum.org/agenda/2020/11/heres-how-technology-has-changed-and-changed-us-over-the-past-20-years)

[020/11/heres-how-technology-has-](https://www.weforum.org/agenda/2020/11/heres-how-technology-has-changed-and-changed-us-over-the-past-20-years)

[changed-and-changed-us-over-the](https://www.weforum.org/agenda/2020/11/heres-how-technology-has-changed-and-changed-us-over-the-past-20-years)

[-past-20-years](https://www.weforum.org/agenda/2020/11/heres-how-technology-has-changed-and-changed-us-over-the-past-20-years)

Lovett, S. (2022, May 10). *Hepatitis in*

children: Any delays in diagnosis

could lead to negligence claims,

say lawyers. Hepatitis in children:

Any delays in diagnosis could lead

to negligence claims, say lawyers.

[https://www.independent.co.uk/ne](https://www.independent.co.uk/news/health/hepatitis-children-covid-a-denovirus-b2074816.html)

[ws/health/hepatitis-children-covid-a](https://www.independent.co.uk/news/health/hepatitis-children-covid-a-denovirus-b2074816.html)

[denovirus-b2074816.html](https://www.independent.co.uk/news/health/hepatitis-children-covid-a-denovirus-b2074816.html)

Timmins, G., & Head, E. (2022, May 13).

Hepatitis outbreak in children:

What do we know so far? | Imperial

News. Imperial College London.

Retrieved June 4, 2022, from

[https://www.imperial.ac.uk/news/23](https://www.imperial.ac.uk/news/236414/hepatitis-outbreak-children-what-know/)

[6414/hepatitis-outbreak-children-w](https://www.imperial.ac.uk/news/236414/hepatitis-outbreak-children-what-know/)

[hat-know/](https://www.imperial.ac.uk/news/236414/hepatitis-outbreak-children-what-know/)

What are the symptoms of hepatitis and

can it be treated? (2022, April 26).

The Independent. Retrieved June

4, 2022, from

[https://www.independent.co.uk/life-](https://www.independent.co.uk/life-style/health-and-families/hepatitis-symptoms-treatments-sick-signs-b2065451.html)

[style/health-and-families/hepatitis-s](https://www.independent.co.uk/life-style/health-and-families/hepatitis-symptoms-treatments-sick-signs-b2065451.html)

[ymptoms-treatments-sick-signs-b2](https://www.independent.co.uk/life-style/health-and-families/hepatitis-symptoms-treatments-sick-signs-b2065451.html)

[065451.html](https://www.independent.co.uk/life-style/health-and-families/hepatitis-symptoms-treatments-sick-signs-b2065451.html)

