OPEN ACCESS

Kashf Journal of Multidisciplinary Research
Vol: 01 - Issue 12 (2024)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

DETECTION OF HCV LIVER FIBROSIS APPLYING MACHINE LEARNING TECHNIQUE

Sadia Latif
Department of Computer Science Bahauddin Zakaria University, Multan, Pakistan.
Azhar Mehboob
Department of Computer Science, Islamia University of Bahawalpur, Pakistan
Assad Latif
School of management and engineering North China University of water resources and electric power Zhengzhou Henan, China
Salahuddin
Department of Computer Science, NFC Institute of Engineering and technology, Multan, Pakistan
Muhammad Ramzan
Department of Computer Science, Islamia University of Bahawalpur, Pakistan
Muhammad Ans Khalid
Department of Computer Science, Institute of southern Punjab Multan, Pakistan

Article Info

Abstract

The health department can be update using the latest technique that will increase the life expectation of the population. Cancer and HCV liver fibrosis are the most dangerous disease in the world due to these diseases death rate is increasing in the world. There are many tools and methodologies exist that predict the spread of HCV liver fibrosis and many performance algorithms applied on the given dataset but what's the big research gap is still no one used the attributes to find the main reason of HCV liver fibrosis. To overcome this research gap there is a need of best prediction of model by using different HCV liver fibrosis features to predict and overcome HCV liver fibrosis. We will analyze the dataset and will give the best HCV liver fibrosis disease prediction model. So, there is need to develop system which can handle all issues as well as produce quality result as compare to previous systems.

Keywords: *Deep learning, HCV Liver, Prediction, Convolution Neural Network*

Introduction

The liver is in the upper right of the stomach, below the diaphragm. This part of the body is also called an exocrine gland. It is in charge of several essential life processes, such as the production of bile, which is used in digestion; the cleansing of the blood; the management of blood toxicity; the elimination of bilirubin; the maintenance of an active metabolism; and the conversion of potentially harmful ammonia into urea. With that much body fat, fat builds up in the liver, making liver disease more likely. Fat accumulation in the liver can lead to a condition known as fatty liver disease, a situation quite prevalent in many regions of the world. For example, each year in India, there are reports of over 10 million instances of the disease. Because there are no adverse symptoms, testing is required before a diagnosis can be made. End-stage liver cirrhosis and hepatic fibrosis are becoming more common in every part of the world. Scarring from fibrosis can lead to cirrhosis, a condition that lasts for a long time. Scar tissue has formed bands, which have taken the place of the liver's natural structure. Many things can cause cirrhosis, but the most common ones are hepatitis B and C and drinking too much alcohol. A chronic HCV infection usually won't show any symptoms for a few years or even much longer. Some people with long-term illnesses don't have much liver damage, while others quickly develop liver cirrhosis and are at risk for hepatocellular carcinoma [1]. The development of chronic liver disease (CLD), which has a high prevalence rate worldwide, is significantly aided by the presence of chronic hepatitis B (CHB). CLD is an abbreviation that stands for chronic liver disease. The stage of liver fibrosis is the most precise indicator of the severity of the disease, and it accurately predicts the requirement of carrying out a variety of treatments. This is because there are many various approaches that can be used to treat liver fibrosis. At the moment, the most accurate method for diagnosing liver fibrosis is to do a liver biopsy. However, it has a number of drawbacks, including the fact that it is invasive, that it can cause complications, that it can make sampling errors (samples only represent about 1/50,000 of the whole liver), that it cannot be repeated in a short amount of time, that it has a complicated operating procedure, that it is susceptible to the observer's bias, and that it can be expensive. Research and development are being done on a number of potential alternatives right now. [2]. Because of these limits, the liver biopsy may lead to an incorrect diagnosis, and it is also possible that the patient will refuse to go through with the surgery. Because of this, there is a significant need for risk-free and non-invasive diagnostic procedures, especially for figuring out how lousy liver fibrosis is. The main reason for this demand is the need to find out how awful liver fibrosis is. Non-invasive diagnostic approaches generally incorporate serological diagnostics and imaging diagnosis. Serological models that use simple formulas and markers to measure liver fibrosis are becoming increasingly popular. [3]. Clinical decision support systems, or CDSSs, are being made to help doctors in a way that doesn't get in the form of treatment. Through the utilization of a wide range of ML algorithms and the facilitation of the correlation of a variety of patient test results, the Clinical Decision Support System (CDSS) assists medical professionals in selecting the most appropriate course of therapy at each stage of the process. As a consequence of this, it guarantees that the diagnosis will be prompt and cost-effective. The Hepatitis C virus (HCV), which is widely acknowledged to be a significant threat to human health, is a source of concern for around 200 million people across the globe. One of the most devasting effects of HCV infection is the development of life-threatening disorders such as fibrosis, cirrhosis, liver cancer, and hepatocellular carcinoma. These conditions can lead to death. This is due to the fact that the majority of infected individuals are unable to mount an effective defense against the virus and recover from its effects once it has taken hold in their body. The incidence of this disease in Egypt is a significant reason for concern because hepatitis C virus (HCV) is the leading cause of cirrhosis in that country. In addition, Egypt is accountable for approximately 13–15 percent of the total number of patients worldwide who are afflicted with the hepatitis C virus [4].

Around 200 million people on the globe are worried about (HCV), which is widely considered a substantial hazard to human health. This is a problem since HCV is a threat to human health. One of the most devasting effects of HCV infection is the development of life-threatening diseases such as fibrosis, cirrhosis, liver cancer, and hepatocellular carcinoma. These disorders can lead to death. This is because most infected individuals cannot mount an effective defense against the virus and recover from its effects once it has established a foothold in their bodies. Because the hepatitis C virus (HCV) is the most common cause of cirrhosis in Egypt, the prevalence of this disease in that country is a significant cause for concern. Cirrhosis is a condition in which the liver becomes scarred and inflamed. In addition, Egypt is responsible for roughly 13–15 percent of the total number of people worldwide infected with the hepatitis C virus. This percentage is based on the total number of patients worldwide who have the virus. Cirrhosis of the liver, chronic hepatitis, and hepatocellular carcinoma can all be traced back to the infectious condition that caused them. Chronic HCV, which has been found to have the highest prevalence rate in Egypt, has induced a significant amount of liver damage in the country. Egypt has suffered significantly from this condition. Depending on the circumstance, the percentage can range from 13 to 15 percent of the total. The hepatitis C virus is responsible for the inflammation in the liver, which can eventually lead to liver fibrosis (development of scar tissue in the liver). According to the Meta-analysis of Histological Data in Viral Hepatitis (METAVIR) system, the stages of liver fibrosis (F0-F4) can be broken down as follows: no fibrosis (F0), portal fibrosis (F1), few septa (F2), many septa (F3), and cirrhosis (F4), accordingly. The procedure that is still considered to be the gold standard for making the diagnosis of hepatic fibrosis is a biopsy of the liver. Liver biopsies are the method that is both the most common and the most reliable when it comes to diagnosing hepatic fibrosis. On the other hand, it is significantly more intrusive, time-consuming, and delicate, and the collection of residuals and historical assessment cause the patients to experience discomfort. In addition, there is a higher possibility of making a mistake. There has been a significant increase in the number of non-invasive diagnostic approaches made available in recent years to identify fibrosis and cirrhosis in HCV patients. In addition, the utilization of these procedures does not involve any hazards, is not complicated, can generate reliable results, and can be repeated. During the preliminary stages of this investigation, most of the focus was placed on the model founded on DT. The systematic research that used various classifiers was not included, and the study outcomes were not investigated, using several distinct assessment metrics. This body of work explores the effects of liver fibrosis by using a wide variety of feature selection and classification techniques to accomplish this goal. Because of this, the experiment results can be analyzed in greater detail, allowing us to determine which machine learning model is the most effective in recognizing HCV. [5].

1. Related work

The liver is responsible for digestion and for moving things along. It also acts as a promoter of enzymes and stores vitamins, glycogen, and minerals. Due to the nature of its work, it may be exposed to harmful pollutants. Diagnosis of liver disease might be subjective. Because there are so few signs, liver problems are frequently overlooked. The presence of hyperbilirubinemia almost always indicates liver disease, although not always. The amounts of enzymes can be learned from liver disease. The artificial intelligence anticipates liver problems. For the purpose of detecting and classifying liver illness, the suggested ensemble soft voting classifier makes use of decision trees, SVMs, and Naive Bayes classifiers. The variables gender, age, alanine, total bilirubin, aminotransferase, aspartate aminotransferase, direct bilirubin, albumin, alkaline phosphatase, and result are all included in the unbalanced dataset. Total bilirubin, aminotransferase, and aspartate aminotransferase are a few examples of additional variables. It is vital to take into

consideration both the accuracy of the algorithm's predictions as well as the error computations in order to reach the finest results that are possibly attainable.The goal of this study [1] is to look at the above classification algorithms on an unbalanced dataset and use parameters to judge how well they work. The suggested ensemble soft voting classifier uses three machine-learning techniques to diagnose and forecast liver disease. These methods are called Decision Tree, Support Vector Machine, and Naive Bayes classifiers. This puts everything into one of two categories. The proposed model's capacity for classification was significantly enhanced when it was trained on a collection of liver disorders that were not in any particular order. According to the study's findings, it would appear that the proposed model led to an improvement in the classification accuracy score. This conclusion was reached after taking into account the outcomes of the study. The implementation of the soft voting strategy resulted in a significant improvement in the efficiency of the classification procedure that was utilized in this investigation. Using this method, you may determine which algorithm works best with this specific dataset, the properties of which have already been determined in advance. The relative mean square error (RMSE), the mean square error (MSE), and accuracy are the three metrics that can be compared to one another. When it comes to accuracy, the predictions that the classification system for soft voting has produced have been the ones that have been the most spot on. The course of liver illness, which can be challenging to diagnose in its earlier stages, can be predicted with a high degree of accuracy with this model, which medical professionals use. The results of this study make it significantly more accessible for people who work in the medical field to make predictions. The possibility exists that in the not-too-distant future, this model will be used for far more extensive datasets that include an even more significant number of features, which will cause it to operate more efficiently [1].

In today's sedentary environment, a range of syndromes and undiagnosed diseases have emerged, leading to a surge in lifestyle disorders. These syndromes and conditions have increased the rate of lifestyle disorders. Physical examinations or medical diagnostics can detect illness in its early stages. This will prevent the infection early and require minimum medicine. Standard treatment includes a physical checkup and lab tests. Early liver disease diagnosis is difficult since symptoms appear late in the progression. This makes early liver disease detection challenging. Machine learning models would help detect the disease early and identify factors contributing to liver degeneration. Within the scope of this study, In [14] The researchers have come up with a method for simplifying features that revolves around the removal of recursive features. In addition, we make use of the Machine Learning Boosting Algorithms in order to improve the accuracy of our forecasts. After applying several different basic machine learning models to the dataset, the findings indicated that logistic regression and multi-layer perceptron had higher prediction accuracies with fewer characteristics. This was the case despite the fact that these two models had less characteristics. Following the application of the models to the dataset, this was found to be the case. The application of boosting algorithms such as Cat Boost, LGBM Classifier, XG Boost, and Gradient Boost helped to increase the accuracy of the dataset. In order to examine the data, several strategies were utilized. The research was conducted on the effect of feature reduction in relation to gradient-boosting machine learning algorithms to understand the potential

repercussions of this effect. Symptoms are seldom present in the early stages of most diseases that affect the liver. If the condition is recognized when it is still in the early phases of its growth, taking preventative measures and stopping its further spread will be much less complicated. The implementation of methods from machine learning makes early detection of the problem easier, and it also makes it possible to identify the key components that contribute to its development. In comparison to the results obtained by the other significant machine learning models, the Logistic Regression model and the Multi-Layer Perceptron model fared exceptionally well. The Recursive Feature Elimination technique was utilized to great effect in order to improve the prediction accuracy of the Gradient Boosting algorithms and bring them up to the level of precision required to reach an accuracy of 94%. It is not out of the realm of possibility that the creation of cutting-edge boosting algorithms for machine learning could result in an increase in the precision of forecasts.[14].

Several tests are required to evaluate the severity of liver fibrosis brought on by chronic HBV infection. [3], created a technique that makes use of the assistance of computers. RFC, with its nine indications, has the potential to improve accuracy in detecting the severity of liver fibrosis compared to the 19 existing models and the other three machine learning algorithms. This is because RFC has more information to work with than the other models. This is especially relevant to stages S2 and S3, in that order. This holds especially true for phases S2 and S3 of the process. This is particularly important for the procedure's phases S2 and S3. This holds especially true for the process's phases S2 and S3, respectively. It has been established that the quality of the training samples has a very significant role in developing a classifier. It is important to conduct additional research based on large data sets that include information on serum producers and images in the physical layer to improve diagnosis accuracy and make its clinical application more straightforward. [3], research is also required to make it possible to increase its clinical use. In this investigation, we established models for determining the degree to which liver fibrosis was present by employing various machine learning strategies. These tactics consisted of DTC, RFC, LRC, and SVC. These 920 instances were all affected by a persistent HBV infection throughout their bodies. To compile the data, the Department of Infectious Diseases at Second Xiangya Hospital contacted each of them in a backwards fashion between April 2007 and December 2018 so that they could look at their medical histories. The entire data set consists of one hundred and fifty percent of the samples used for training and testing combined. It took the application of four distinct machine learning classifiers to sift through random combinations of 24 indicators, which included 67 108 760 group indicators, in order to identify the indicator combinations that yielded the most favorable outcomes. This was done to choose the indicator combinations that had the best results. This was done to determine which combinations of indicators gave the best outcomes and pick those combinations.

The concept of machine learning underpins this model [5] were able to collect the cases of liver fibrosis illness that Egyptian patients encountered by utilizing the machine learning repository located at UCI. The method of synthetic minority oversampling, which involves increasing the number of synthetic examples of patients, has been applied so that the instances of various categories can be brought into equilibrium. After that, we determined the significant aspects of the hepatitis C virus in this dataset using a range of techniques for feature

selection. These techniques included both manual and automated approaches. Patients have been placed into one of the following three categories using a variety of distinct classifiers: balanced primary, chosen feature, or primary HCV cases. KNN displays the highest accuracy (94.40 percent) after completing this analysis compared to other classifiers. The infectious disease caused by the hepatitis C virus has been analyzed, and decisions regarding the condition have been made with the assistance of this result. Various methods such as data balance, feature selection, and classification were applied during this inquiry of the HCV patient dataset. According to our research findings, the KNN algorithm yields the most accurate types when applied to HCV patients in varying stages. In addition, RF and SVM show likely results and indicate that they function more effectively when used for this analysis. During [5] investigation of HCV patient records, I encountered a few problems that need to be addressed. The raw dataset utilized for the experimental ECV does not include any additional instances or attributes that may be used for analysis. Additionally, this should be performed by employing more feature selection, transformation, and classification algorithms than was previously done. In the future, researchers will attempt to mitigate the effects of these limits by collecting additional cases and implementing Internet of Things (IoT)-based modules that are able to automatically detect HCV. This will be done in an effort to alleviate the effects of these constraints[5].

The goal of this work [19] is to select the most effective instrument for diagnosing and detecting hepatitis, as well as for calculating the life expectancy of hepatitis patients once treatment has been completed. Specifically, the goal is to determine which device has the best chance of succeeding in this endeavor. In addition, this research aims to estimate how long people with hepatitis can anticipate living. In order to accomplish the goal of this body of work, research into the similarities and differences between the several approaches to machine learning and neural networks was carried out. This inquiry was carried out to fulfil the objective of this body of work. The total mean square error as well as the proportion of questions that were answered correctly are taken into consideration by the performance metric. It was formerly believed that the classification and prediction approaches for diagnosing hepatitis disease were the machine learning algorithms such as support vector machines (SVM), kernel neural networks (KNN), and support vector regressions (SVR) (ANN). After it was established that there was a possibility for an improvement in the accuracy of disease diagnostic prediction, some early study on the algorithms was carried out. The programming environment known as MATLAB was applied throughout the process of putting machine learning into practice and verifying its accuracy. The work done during the Designing Phase will be put into action during the Implementation Phase, and the primary focus of this phase will be on putting that work into action so that we can see results that can be measured. Because the majority of the coding that was done for the Business Logic Lay comes into play during this stage, it is the stage that is considered to be the most fundamental component of the overall project. It is the developer's responsibility to finish this task at each step of the project. Additionally, it is the developer's responsibility to fine-tune the bug and module dependencies. However, it is only here that we will fix all of the runtime issues. The identification of hepatitis as the cause of this body of work was its primary objective. In order to accomplish this goal, many machine learning techniques and neural network approaches were utilized throughout the entirety of the project. In order to determine which

method of diagnosis is the most effective for hepatitis disease, a comparison of the accuracy achieved for a particular data set using a variety of ML and ANN algorithms was carried out. This comparison was carried out in order to determine which method is the most accurate. This was done in order to identify which method of diagnostic assessment is the most efficient. This comparison was carried out so that we could identify which approach was the most successful in diagnosing the condition.[19], produced an accurate forecast of the disease by utilizing a Support Vector Machine (SVM), an Artificial Neural Network (ANN), and a K Nearest Neighbor (KNN) technique. Based on the findings of this research, it is possible to conclude that the ANN model is the most accurate of all the models that were examined and its performance and that this is the case. The ANN model boasts an impressively high level of prediction accuracy, coming in at 96 percent and a low mean square error. In subsequent research projects, the same strategy, which uses RNN, will be utilized to produce forecasts regarding the occurrence of various diseases.

2. Methods and materials

2.1. Dataset Collections

We have collected the dataset from the UCI Dataset Repository; the name of the **Dataset 2** is **HCV Data (Comprehensive)**[1] [1]. 14 hepatitis attributes make up the hepatitis illness dataset, which are as follows: X (Patient ID/No), Category, Age, Sex, ALB, ALP, ALT, AST, BIL, CHE, CHOL, CREA, GGT, and PROT. This dataset contains a total of **615** samples from people suffering from hepatitis **Disease**. The specifics of a dataset are displayed in Figure 01.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 589 entries, 0 to 588
Data columns (total 13 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Category  589 non-null    int64
 1   Age       589 non-null    int64
 2   Sex       589 non-null    int64
 3   ALB       589 non-null    float64
 4   ALP       589 non-null    float64
 5   ALT       589 non-null    float64
 6   AST       589 non-null    float64
 7   BIL       589 non-null    float64
 8   CHE       589 non-null    float64
 9   CHOL      589 non-null    float64
 10  CREA      589 non-null    float64
 11  GGT       589 non-null    float64
 12  PROT      589 non-null    float64
dtypes: float64(10), int64(3)
memory usage: 59.9 KB
```

Figure 1 Data Set Details

3.1.2 Data Set Describe

This type of Meta Data Description is a clean, well understanding set of records. Anyhow the significance of some of the attributes is not much clear. Let's see the meaning of;

[1] https://archive.ics.uci.edu/ml/datasets/HCV+data

- Age: How much person's old now in terms of years
- Sex: Differentiate between Male and Female in term of (1 = Male, 0 = Female)
- Category: The category has five different values (Blood Donor is related to Value 0, Suspect Blood Donor is related to Value 0s, Hepatitis is related to Value 1, Fibrosis is related to Value 2, and Cirrhosis value is 3)
- ALB: The ALB is related to the albumin level in patients
- ALP: The ALP (**Alkaline phosphatase**) is related to the amount of enzyme in the liver.
- GGT: This test tells us the total amount of GGT (GAMMA-Glutamyl Transferase)
- PORT: This is the attribute, which post-transfusion hepatitis (Yes is 1, No is 0)

Our research use Description – **Meta Data 01** with the complete details of this attribute to understand it.

	Category	Age	Sex	ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	
count	615.000000	615.000000	615.000000	614.000000	597.000000	614.000000	615.000000	615.000000	615.000000	605.000000	615.000000	615.000000	614.0
mean	0.299187	47.408130	0.613008	41.620195	68.283920	28.450814	34.786341	11.396748	8.196634	5.368099	81.287805	39.533171	72.0
std	0.841657	10.055105	0.487458	5.780629	26.028315	25.469689	33.090690	19.673150	2.205657	1.132728	49.756166	54.661071	5.4
min	0.000000	19.000000	0.000000	14.900000	11.300000	0.900000	10.600000	0.800000	1.420000	1.430000	8.000000	4.500000	44.8
25%	0.000000	39.000000	0.000000	38.800000	52.500000	16.400000	21.600000	5.300000	6.935000	4.610000	67.000000	15.700000	69.3
50%	0.000000	47.000000	1.000000	41.950000	66.200000	23.000000	25.900000	7.300000	8.260000	5.300000	77.000000	23.300000	72.2
75%	0.000000	54.000000	1.000000	45.200000	80.100000	33.075000	32.900000	11.200000	9.590000	6.060000	88.000000	40.200000	75.4
max	4.000000	77.000000	1.000000	82.200000	416.600000	325.300000	324.000000	254.000000	16.410000	9.670000	1079.100000	650.900000	90.0

Figure 2 Data set overview

Figure 02 is showing the details of our Dataset in terms of the total number of records (**615**) Dataset 2, Find out the mean, Standard deviation, Min value, Max value**, 25%, 50%, and 75%** Percentage of the Dataset.

3.1.3 Data Set Visualization

3.1.3.1 HCV Category

In our dataset 2, there is four different category of class lab in it. In data set 2, 0 is related to number of patients, which can work as Blood Donor, 1 is relating to patient with Cirrhosi-suspect, 2 is relating to Blood Donor Hepatitis, and 3 is belonging to patient with Fibrosis issues. The figure 3 is showing amount of people in all four categories.

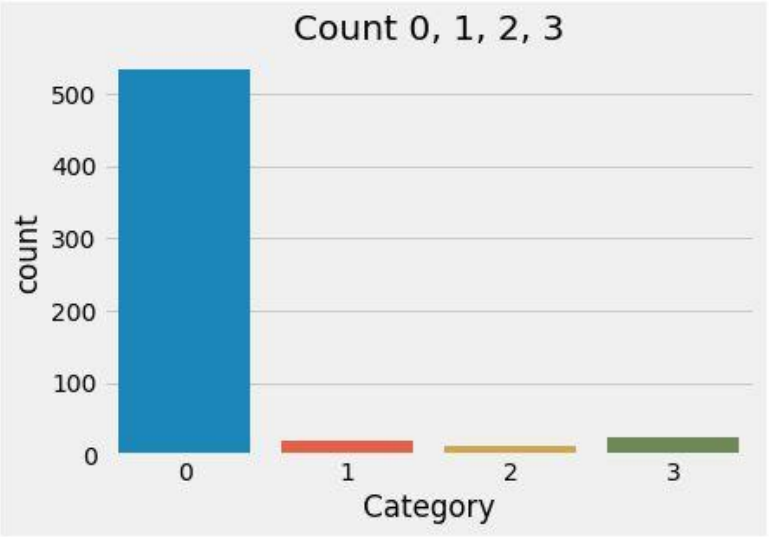

Figure 3 Amount of people in all four categories

3.1.3.2 Correlation Between Features

The cluster Map pertaining to Target is displayed down below in Figure 4. The Cluster Heatmap makes it simple to distinguish the characteristics of the dataset that are most closely connected to the target characteristic. In order to plot the associated characteristics of the heatmap, we made use of the seaborn library. As shown in Figure 4, a positive association exists between the categories of Sex, CHE, CREA, and ALB and the Category attribute.

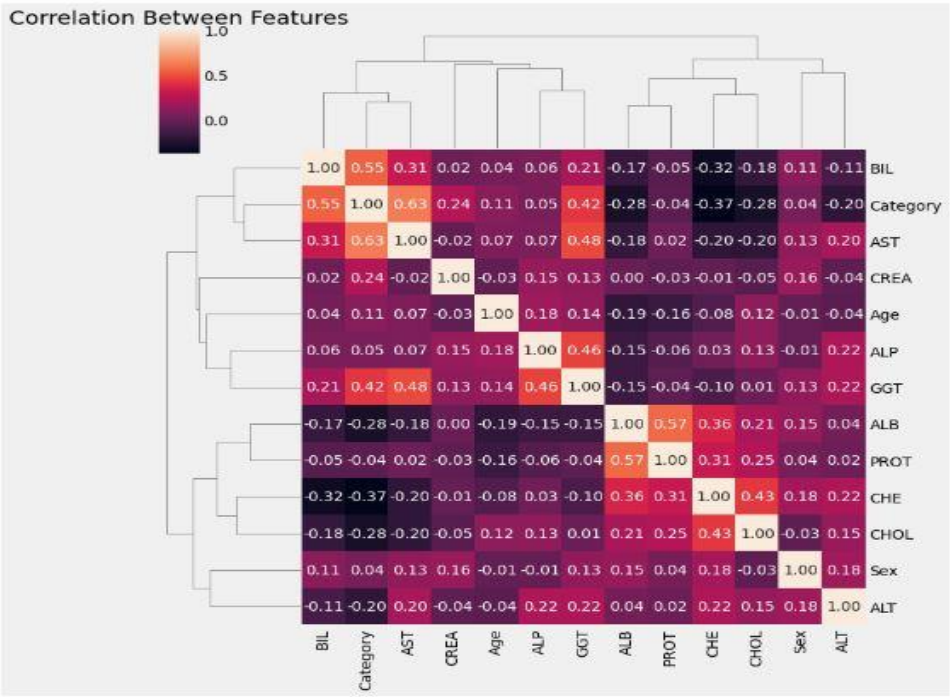

Figure 4 Correlation Between Features

3.1.3.3 Features with Category

Blood Donors, Cirrhosis, Suspected Blood Donors, Hepatitis, and Fibrosis are all shown in Figure 5, together with their Age, Sex, ALB, ALP, ALT, AST, BIL, CHE, CHOL, CREA, GGT, and PROT levels, to illustrate the relationship between these factors and the **hepatitis patient disease**. There is an outlier in our data set (CREA, GGT, and ALP), and it can be identified by the fact that the median value of these features is significantly higher than the median value of its other features.

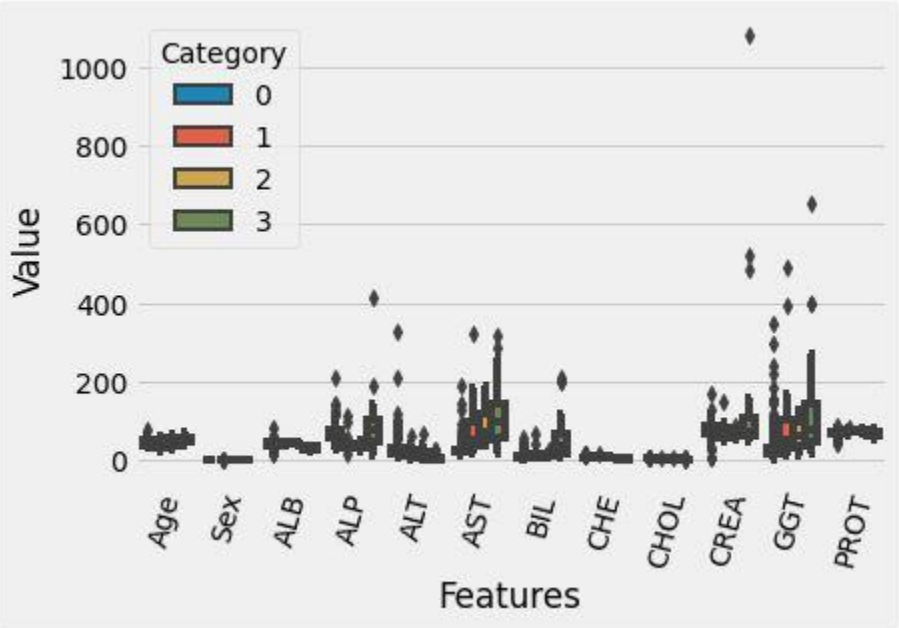

Figure 5 Features with Category

3.2. Analysis of Case 0 (Blood Donor)

Figure 6 depicts an analysis of the Blood Donor category taking into account all of its qualities. The majority of the records in our dataset 2 pertain to patients who donate their blood. People whose ages fall between 30 and 50 years old have fewer hepatitis symptoms in their bodies, according to the age attribute. When compared to the female gender, the male gender is more commonly employed in the blood donation industry. The ALB, ALP, ALT, AST, and BIL value range from **31 - 56, 30- 168, 0.5 – 531, 0 – 43.2, 0.0 – 16** respectively.

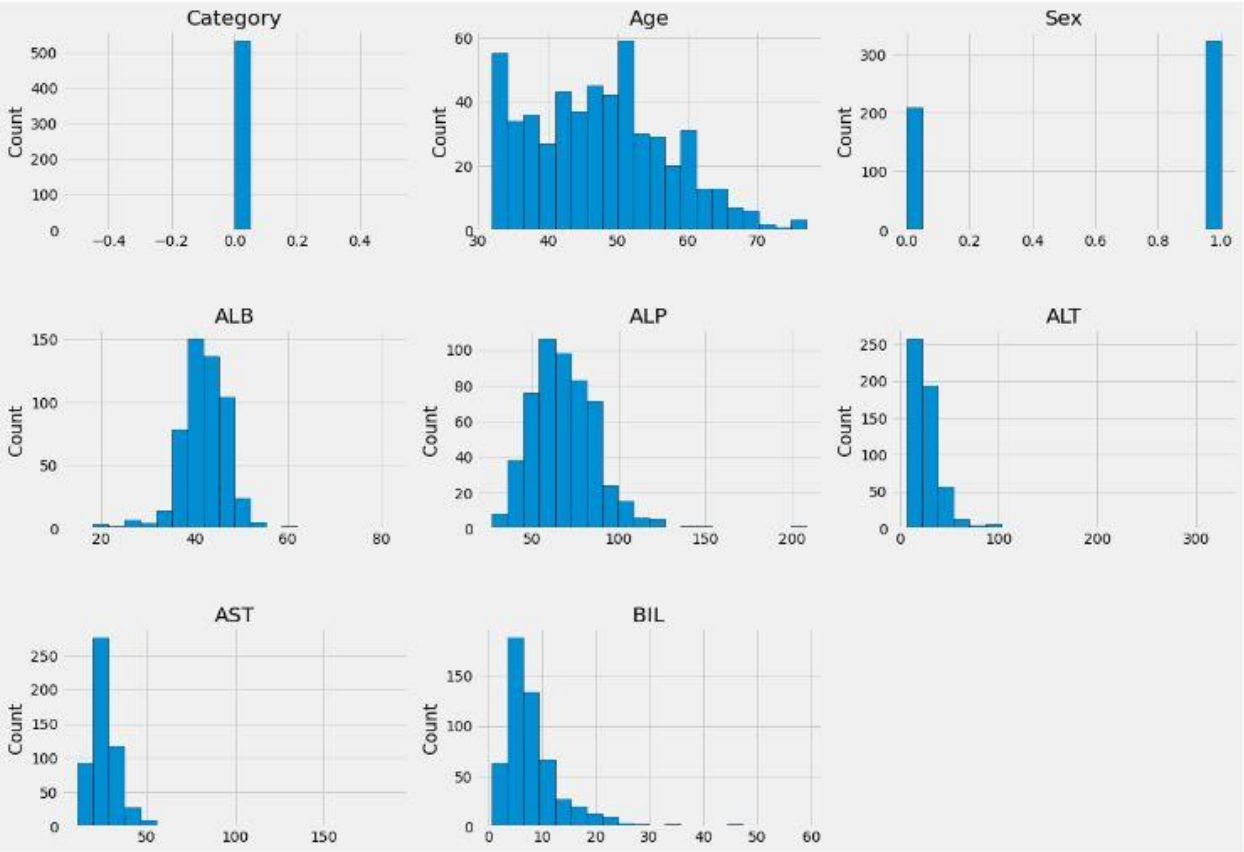

Figure 6 Analysis of Case 0 (Blood Donor)

3.3. Analysis of Case 1 (Cirrhosi-suspect)

The study of the **Cirrhosi**-suspect category may be seen in Figure 7, together with all of its properties. In our dataset, cirrhosis-suspect patients account for two-thirds of the records, making them the second most common category. People who are **50** years old or older are starting to show signs of cirrhosis in their bodies, according to the age characteristic. The masculine gender is significantly more likely to be suspected than the female gender. The values for **ALB, ALP, ALT,** and **AST** can range anywhere from **40.0** to **47.5, 20** to **49, 0.0** to **38, 0.0** to **59.5,** and **0.0 to 26** correspondingly.

Figure 7 Analysis of Case 1 (Cirrhosi-suspect)

3.4. Analysis of Case 2 (Hepatitis Patient)

The study of the **Hepatitis Patient** category, including all of its properties, can be shown in Figure 8. People who are **30** years old or older have a higher risk of having **hepatitis patient** symptoms in their bodies, as shown by the age attribute. The masculine gender is significantly more likely to be suspected than the female gender. The ALB, ALP, ALT, AST, and BIL value range from (**38-41, 44- 48), 25-50, (0 – 20, 44, 65), (75, 125,155, 176), (10, 19)** respectively.

Figure 8 Analysis of Case 2 (Hepatitis Patient)

3.5. Analysis of Case 3 (Fibrosis Patient)

The study of the **Fibrosis Patient** category, including all of its properties, can be seen in Figure 9. People who are **40** years old or older have an increased risk of developing **fibrosis** in their bodies, as shown by the age attribute. The masculine gender is significantly more likely to be suspected than the female gender. The ALB, ALP, ALT, AST, and BIL value range from (**25, 30-35), 0-200, (2.3 – 12.11, 25, 20, 30), (0, 100,150, 200), (0-50, 200)** respectively.

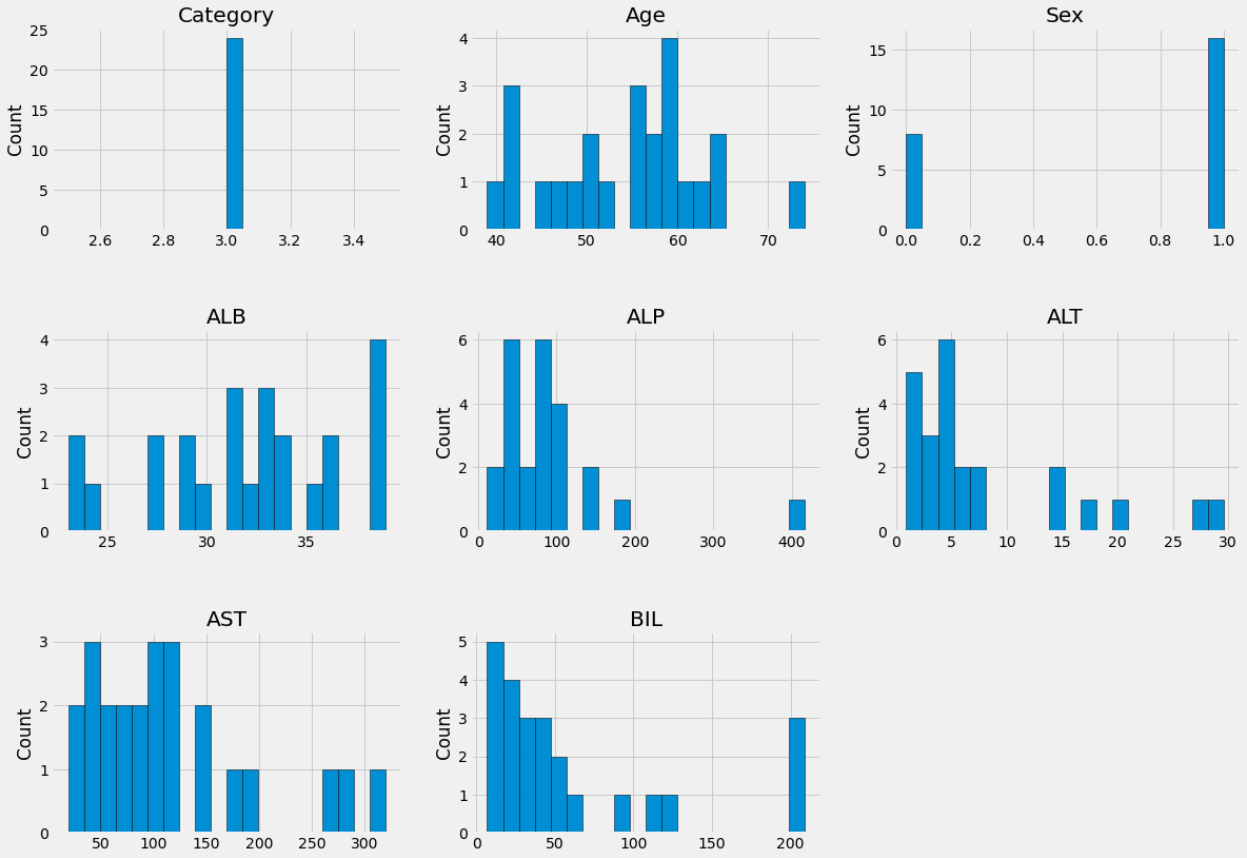

Figure 9 Analysis of Case 3 (Fibrosis Patient)

3.6 Visualization of Features

It will be seen in figure 10 that our dataset-2 contains two different kinds of outlier's altogether. In comparison to outliers, outlier skreas has a lower value (mean discontinue linear pattern). Both outliers can be found in every category (blood donor, **cirrhosis** suspect, hepatitis patient, and **fibrosis patient**, respectively). Therefore, in the subsequent stage, we will need to drop these outliers in order to proceed with the subsequent operation.

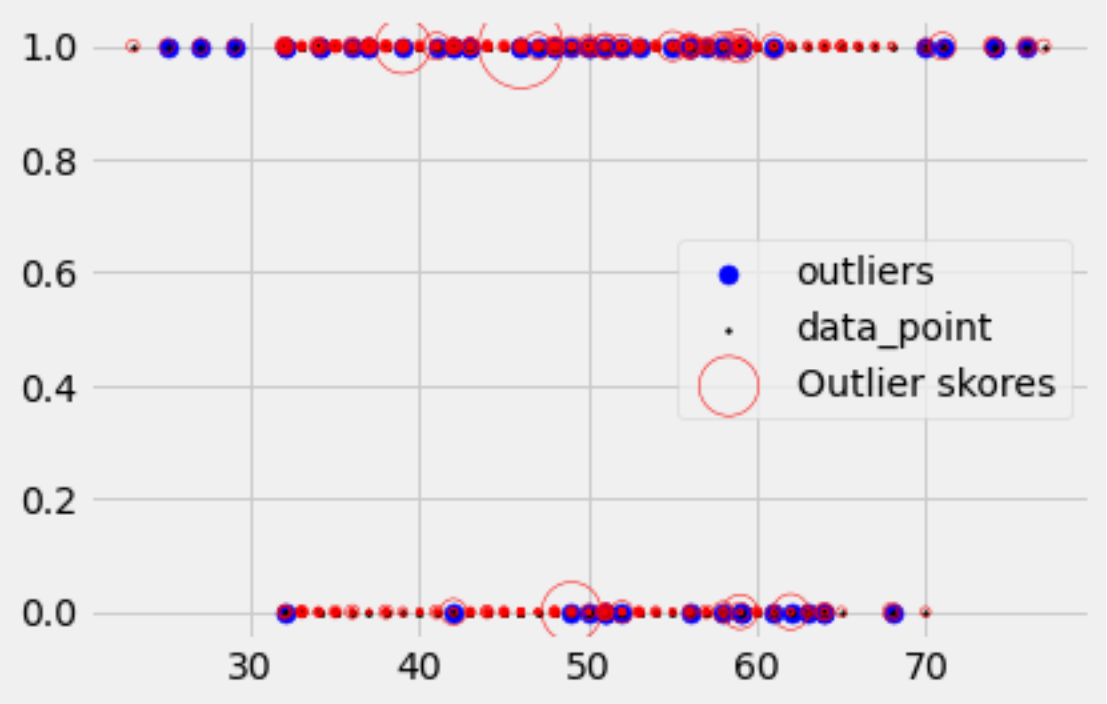

Figure 10 Outlier in Dataset 2

3.6.1 Data Set 1

HCV-Egy-Data is the name of the dataset that we have obtained from the UCI Dataset Repository; the dataset is referred to as **Dataset** 1. The hepatitis illness dataset contains a total of **29** different hepatitis features (Age, Gender, BMI, Fever, Nausea/Vomiting, Headache, Diarrhea, Fatigue & widespread bone ache, Jaundice, Epigastric Pain, WBC, RBC, HGB, Plat, AST 1, ALT 1, ALT4, ALT 12, ALT 24, ALT 36, ALT 48, ALT after 24 weeks, this dataset has a total of **1384** samples of hepatitis **Disease** patients**.** Figure 11 shows the details of a dataset.

```
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Age                             1385 non-null   int64
 1   Gender                          1385 non-null   int64
 2   BMI                             1385 non-null   int64
 3   Fever                           1385 non-null   int64
 4   Nausea/Vomting                  1385 non-null   int64
 5   Headache                        1385 non-null   int64
 6   Diarrhea                        1385 non-null   int64
 7   Fatigue & generalized bone ache 1385 non-null   int64
 8   Jaundice                        1385 non-null   int64
 9   Epigastric pain                 1385 non-null   int64
 10  WBC                             1385 non-null   int64
 11  RBC                             1385 non-null   int64
 12  HGB                             1385 non-null   int64
 13  Plat                            1385 non-null   int64
 14  AST 1                           1385 non-null   int64
 15  ALT 1                           1385 non-null   int64
 16  ALT4                            1385 non-null   int64
 17  ALT 12                          1385 non-null   int64
 18  ALT 24                          1385 non-null   int64
 19  ALT 36                          1385 non-null   int64
 20  ALT 48                          1385 non-null   int64
 21  ALT after 24 w                  1385 non-null   int64
 22  RNA Base                        1385 non-null   int64
 23  RNA 4                           1385 non-null   int64
 24  RNA 12                          1385 non-null   int64
 25  RNA EOT                         1385 non-null   int64
 26  RNA EF                          1385 non-null   int64
 27  Baseline histological Grading   1385 non-null   int64
 28  Outcome                         1385 non-null   int64
dtypes: int64(29)
memory usage: 313.9 KB
None
```

Figure 11 Data Set Details

3.6.2 Data Set Describe

This particular kind of Meta Data Description refers to a tidy and easily digestible collection of records. In any case, the importance of a few of the characteristics is not quite obvious. Let's look at what this phrase means;

- Age: How much person's old now in terms of years
- Sex: Differentiate between Male and Female in term of (1 = Male, 0 = Female)
- Category: The category has five different values (Blood Donor is related to Value 0, Suspect Blood Donor is related to Value 0s, Hepatitis is related to Value 1, Fibrosis is related to Value 2, and Cirrhosis value is 3)
- ALB: The ALB is related to the albumin level in patients
- ALP: The ALP (**Alkaline phosphatase**) is related to the amount of enzyme in the liver.

Our research use Description – **Meta Data 01** with the complete details of this attribute to understand it.

	count	mean	std	min	25%	50%	75%	max
Age	1385.0	4.631913e+01	8.781508	32.0	39.0	46.0	54.0	61.0
Gender	1385.0	1.489531e+00	0.500071	1.0	1.0	1.0	2.0	2.0
BMI	1385.0	2.860866e+01	4.076215	22.0	25.0	29.0	32.0	35.0
Fever	1385.0	1.515523e+00	0.499939	1.0	1.0	2.0	2.0	2.0
Nausea/Vomting	1385.0	1.502527e+00	0.500174	1.0	1.0	2.0	2.0	2.0
Headache	1385.0	1.496029e+00	0.500165	1.0	1.0	1.0	2.0	2.0
Diarrhea	1385.0	1.502527e+00	0.500174	1.0	1.0	2.0	2.0	2.0
Fatigue & generalized bone ache	1385.0	1.498917e+00	0.500179	1.0	1.0	1.0	2.0	2.0
Jaundice	1385.0	1.501083e+00	0.500179	1.0	1.0	2.0	2.0	2.0
Epigastric pain	1385.0	1.503971e+00	0.500165	1.0	1.0	2.0	2.0	2.0
WBC	1385.0	7.533386e+03	2668.220333	2991.0	5219.0	7498.0	9902.0	12101.0
RBC	1385.0	4.422130e+06	346357.711599	3816422.0	4121374.0	4436465.0	4721279.0	5018451.0
HGB	1385.0	1.258773e+01	1.713511	10.0	11.0	13.0	14.0	15.0
Plat	1385.0	1.583481e+05	38794.785550	93013.0	124479.0	157916.0	190314.0	226484.0
AST 1	1385.0	8.277473e+01	25.993242	39.0	60.0	83.0	105.0	128.0
ALT 1	1385.0	8.391625e+01	25.922800	39.0	62.0	83.0	106.0	128.0
ALT4	1385.0	8.340578e+01	26.529730	39.0	61.0	82.0	107.0	128.0
ALT 12	1385.0	8.351047e+01	26.064478	39.0	60.0	84.0	106.0	128.0
ALT 24	1385.0	8.370903e+01	26.205994	39.0	61.0	83.0	107.0	128.0
ALT 36	1385.0	8.311769e+01	26.399031	5.0	61.0	84.0	106.0	128.0
ALT 48	1385.0	8.362960e+01	26.223955	5.0	61.0	83.0	106.0	128.0
ALT after 24 w	1385.0	3.343627e+01	7.073569	5.0	28.0	34.0	40.0	45.0
RNA Base	1385.0	5.909512e+05	353935.357602	11.0	269253.0	593103.0	886791.0	1201086.0
RNA 4	1385.0	6.008956e+05	362315.132786	5.0	270893.0	597869.0	909093.0	1201715.0
RNA 12	1385.0	2.887536e+05	285350.674511	5.0	5.0	234359.0	524819.0	3731527.0
RNA EOT	1385.0	2.876603e+05	264559.525070	5.0	5.0	251376.0	517806.0	808450.0
RNA EF	1385.0	2.913783e+05	267700.691713	5.0	5.0	244049.0	527864.0	810333.0
Baseline histological Grading	1385.0	9.761733e+00	4.023896	3.0	6.0	10.0	13.0	16.0
Outcome	1385.0	2.536462e+00	1.121392	1.0	2.0	3.0	4.0	4.0

Figure 112 Data set overview

Figure 12 presents the information regarding our dataset in terms of the total number of records (**1384**) Dataset 1, Determine the mean, the standard deviation, the minimum value, the maximum value, and the percentages that represent **25%, 50%, and 75%** of the dataset.

3.6.3 Data Set Visualization

3.6.3.1 HCV Category

In our dataset 1, there is four different category of class lab in it. In data set 1, 1 is related to number of patients, which can work as Blood Donor, 2 is relating to patient with Cirrhosi-suspect, 3 is relating to Blood Donor Hepatitis, and 4 is belonging to patient with Fibrosis issues. The figure 13 is showing amount of people in all four categories.

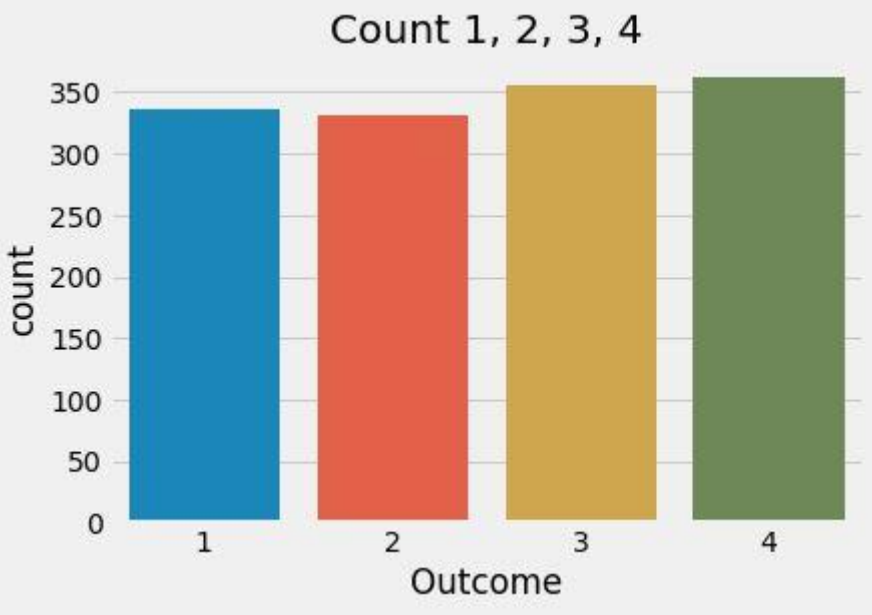

Figure 113 *Amount of people in all four categories*

3.6.3.2 Correlation Between Features

The cluster Map relating to Target is displayed down below in Figure 14. The Cluster Heatmap makes it simple to distinguish the characteristics of the dataset that are most closely connected to the target characteristic. In order to plot the associated characteristics of the heatmap, we made use of the seaborn library. According to Figure 14, there is a positive link between the attributes of Sex, RNA 4, ALT 12, and RAN Base and the Category property.

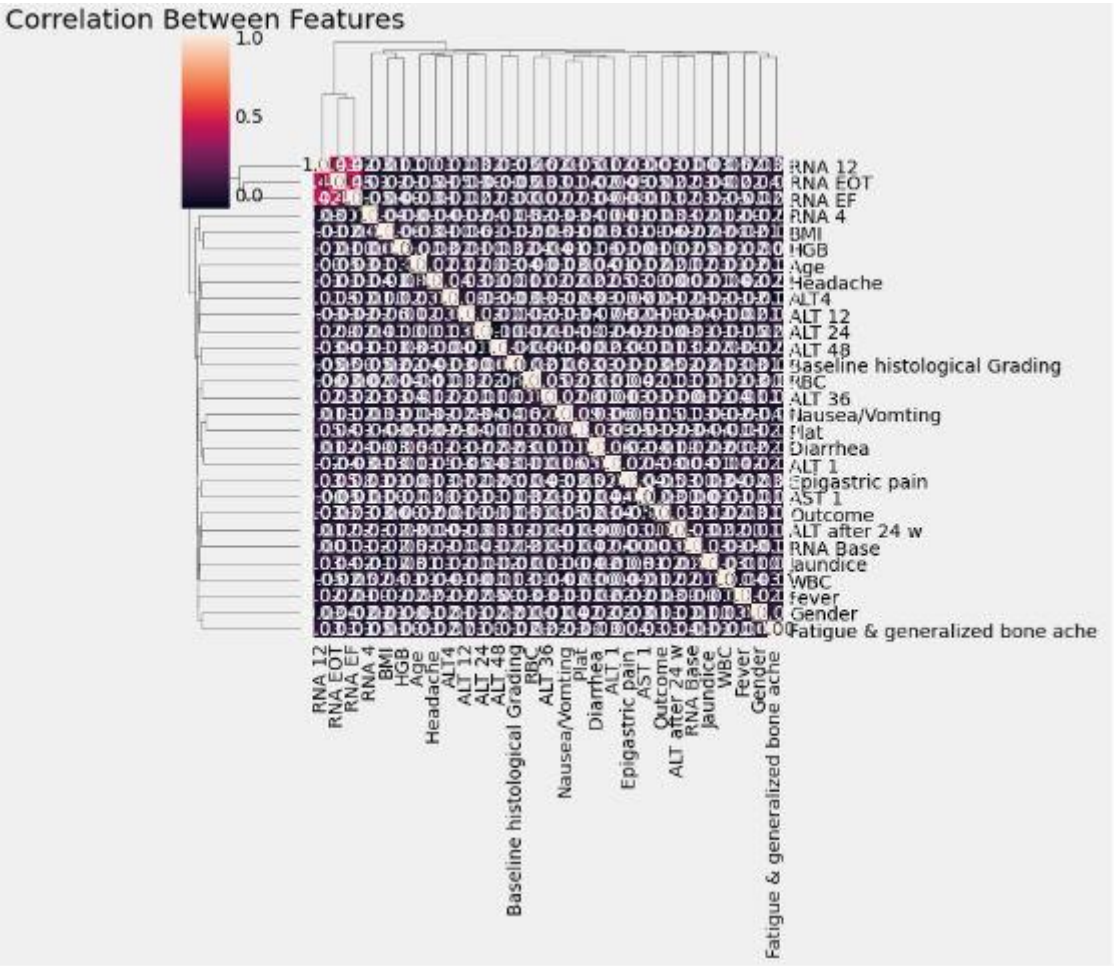

Figure 114 Correlation Between Features

3.6.3.3 Features with Category

Figure 15 is showing a number of Blood Donors, Cirrhosis, suspected Blood Donors, **Hepatitis**, and **Fibrosis** in relation to Age, Gender, BMI, Fever, Nausea/Vomiting, Headache, Diarrhea, Fatigue & generalized bone ache, Jaundice, Epigastric Pain, WBC, RBC, HGB, Plat, AST 1, ALT 1, ALT4, ALT 12, ALT 24, ALT 36 There is at least one outlier in our dataset, which consists of RBC, HGB, and PLAT. This is due to the fact that the median of these features is significantly higher than the median of its other features.

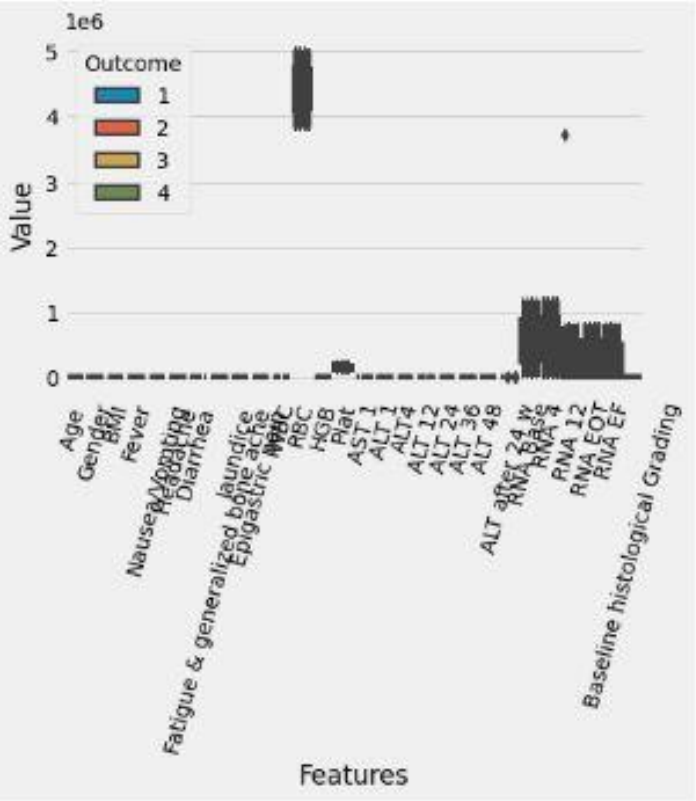

Figure 15 Features with Category

3.7. Analysis of Case 1 (Blood Donor)

Figure 16 is showing the analysis of Blood Donor category with all features. In our dataset 1, most of records are relating to patient, which works as blood donor. Age attribute is showing people that in range of **30** to **60** are less hepatitis symptom in their body. Male gender is mostly work as blood donor as compare to female gender. The BMI, Fever, Nausea/Vomiting, Headache, Diarrhea and Fatigue & Generalized bone ache value range from **22.5** – **25.0, Max 160, 152 – 161, 160 – 165, 150 – 170** respectively.

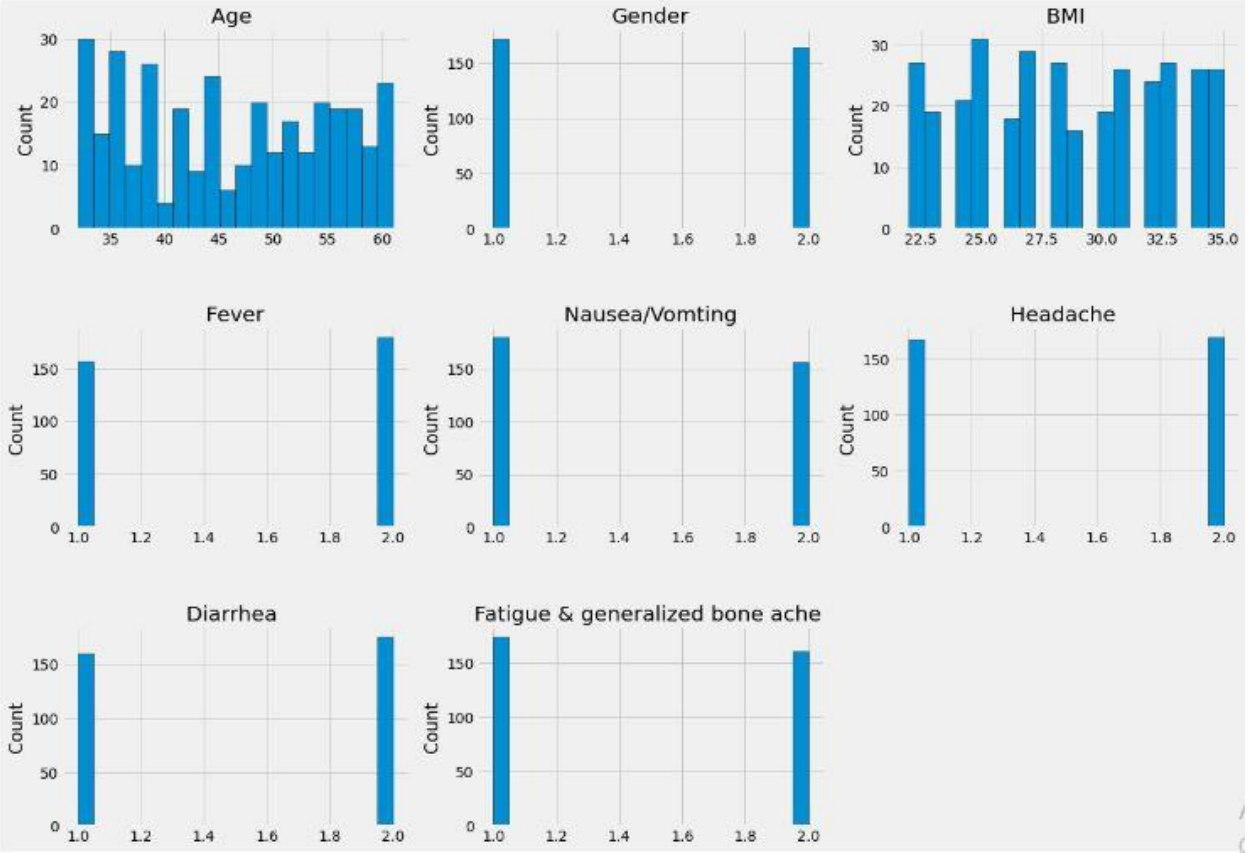

Figure 116 Analysis of Case 1 (Blood Donor)

3.8. Analysis of Case 2 (Cirrhosi-suspect)

Figure 17 is showing the analysis of **Cirrhosi-suspect** category with all features. In our dataset 1, other most of records are relating to patient, which works as **Cirrhosi-suspect**. Age attribute is showing people that **60** plus are **Cirrhosi-suspect** symptom in their body. Male gender is mostly common suspected as compare to female gender. The BMI, Fever, Nausea/Vomiting, Headache, Diarrhea and Fatigue & Generalized bone ache value range from **10 - 33, 153-161, 150 – 160, 220 – 300, 140 – 180** respectively.

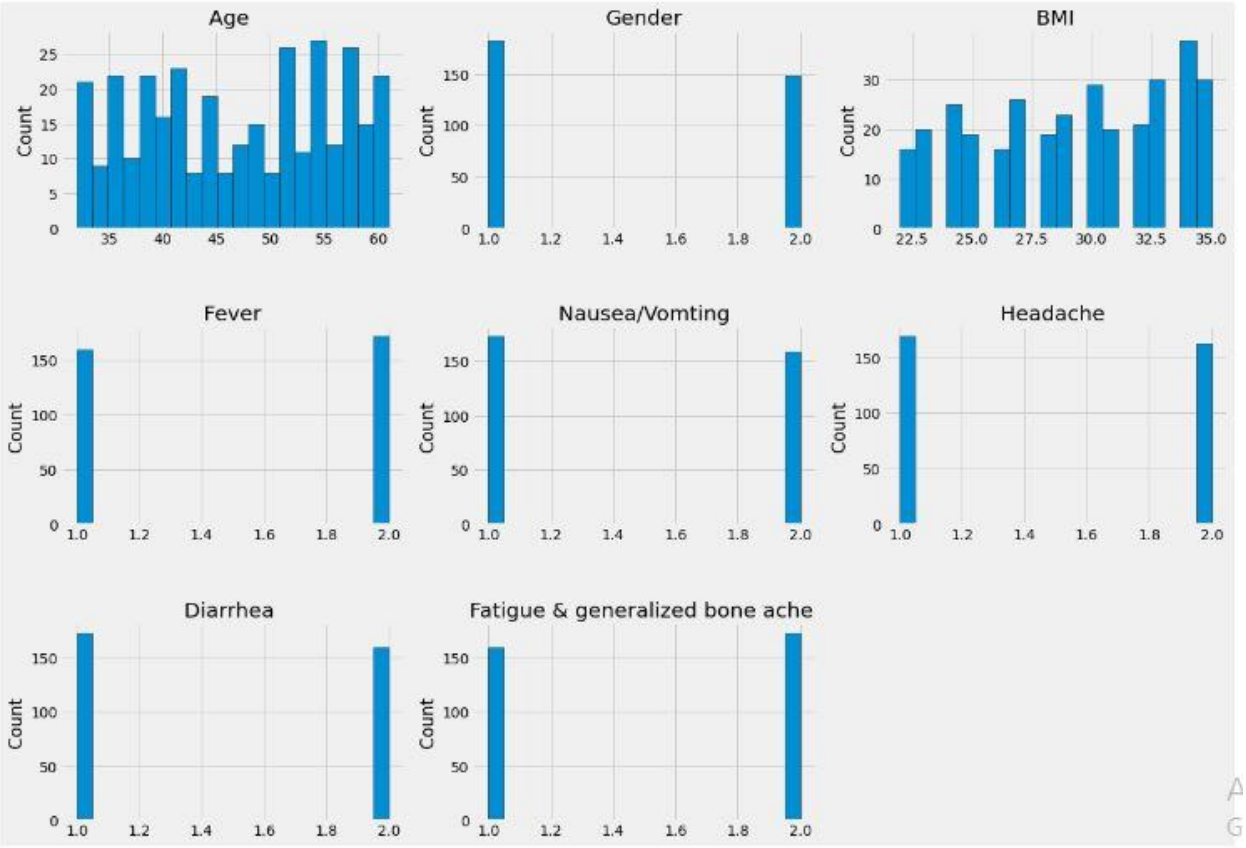

Figure 117 Analysis of Case 2 (Cirrhosi-suspect)

3.9. Analysis of Case 3 (Hepatitis Patient)

Figure 18 is showing the analysis of **Hepatitis Patient** category with all features. Age attribute is showing people that **20** plus group of people are **Hepatitis Patient** symptom in their body. Male gender is mostly common suspected as compare to female gender. The BMI, Fever, Nausea/Vomiting, Headache, Diarrhea and Fatigue & Generalized bone ache value range from (**16-35, 150- 169, 150-170, 220 - 250, 155-160, 157-160** respectively.

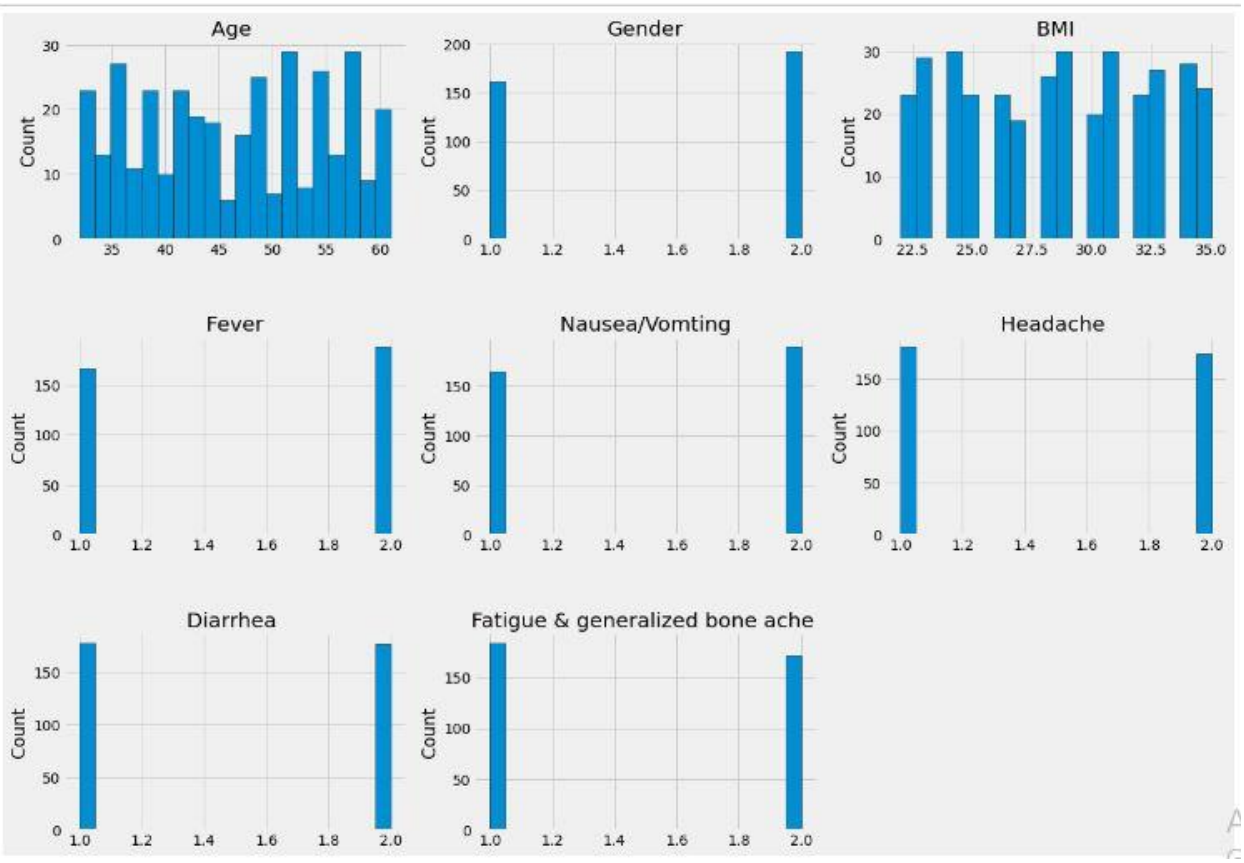

Figure 18 Analysis of Case 3 (Hepatitis Patient)

3.10. Analysis of Case 4 (Fibrosis Patient)

Figure 19 is showing the analysis of **Fibrosis Patient** category with all features. Age attribute is showing people that **21** plus group of people are **Fibrosis Patient** symptom in their body. Male gender is mostly common suspected as compare to female gender. The BMI, Fever, Nausea/Vomiting, Headache, Diarrhea and Fatigue & Generalized bone ache value range from **17-35, 152- 159, 153-160, 210-230, 152-160, and 162-154** respectively.

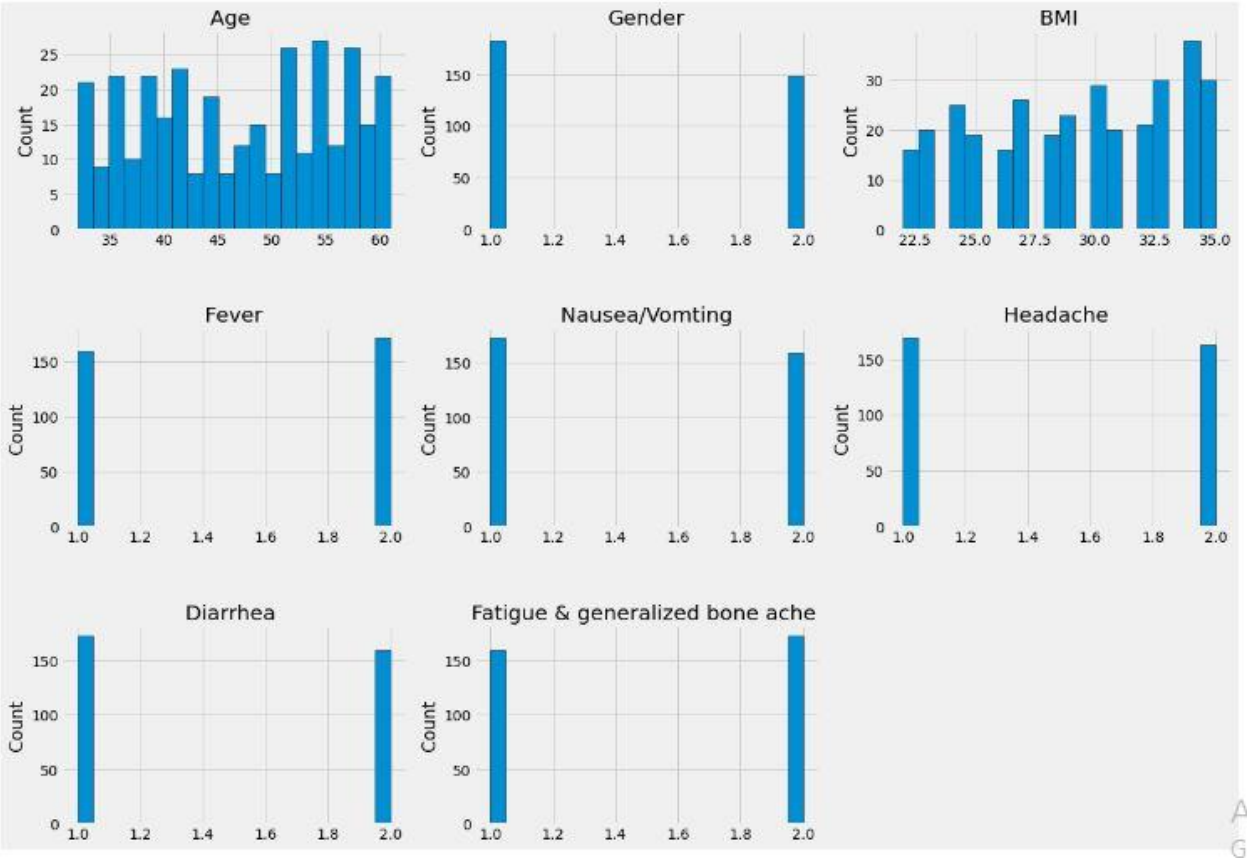

Figure 119 Analysis of Case 3 (Fibrosis Patient)

3.11. Visualization of Features

In figure 20, it will be observing that our dataset-1 has two type of outlier in it. Outlier skores, and outliers (mean discontinue linear pattern). All category (Blood Donor, **Cirrhosi-suspect**, Hepatitis Patient, **Fibrosis Patient) has** both outliers in it. So in upcoming step, we will need to Drop these outliers for further upcoming procedure.

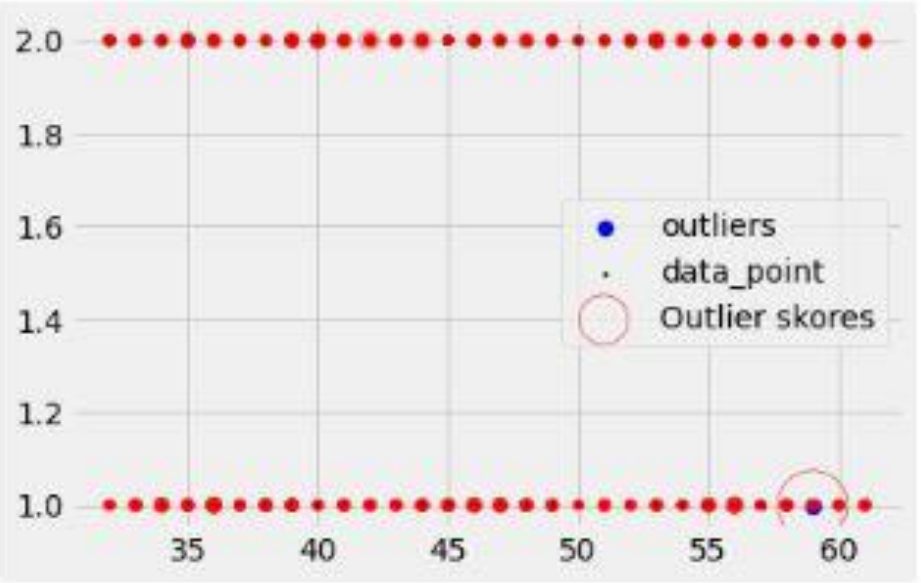

Figure 20 Outlier in Dataset 1

Previous studies stated that large amount of HCV liver fibrosis data collected from different sources and on different websites and different techniques applied on it. This thesis proposes a new machine learning approach for HCV liver fibrosis diagnosis using multiple real data which is taken from online sources. The dataset contains 1000 plus instances. This method is developed using machine learning techniques to use different algorithms such as Extreme Machine Learning (EML), using Naïve Bayes, Random Forest, Decision Tree, SVM algorithm etc. for that purpose python language is used. Machine learning algorithms applied on the given dataset for the detection of HCV liver fibrosis.

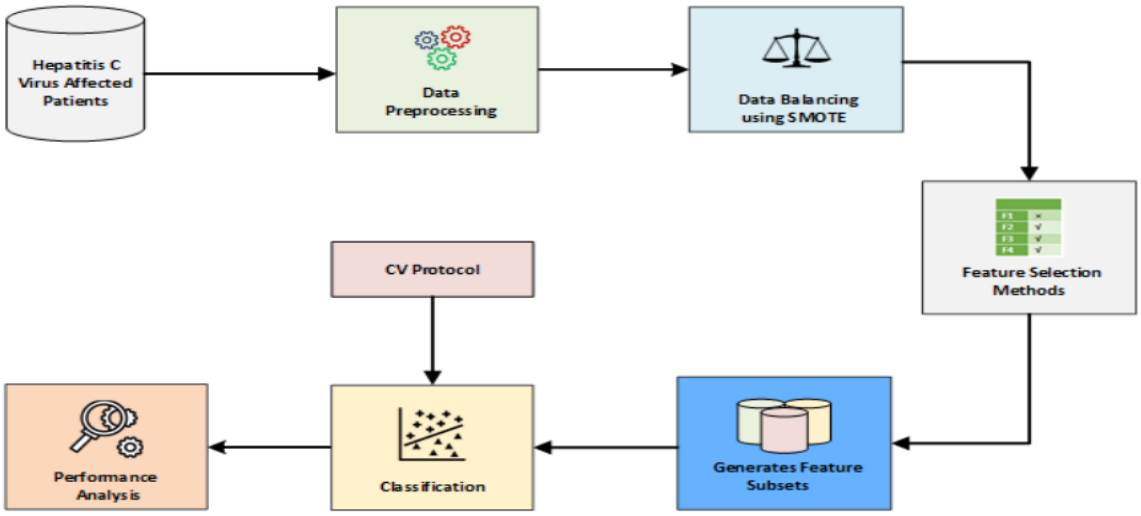

Figure 21 Proposed methodology which is further explained one by one

➢ **Collection of data**

We will use Multiple HCV datasets which we download from UCI machine learning repository. **Dataset 1** Contain 1000 plus records and 34 attributes and **Dataset 2** contains 615 records and 14 attributes. These datasets will be used for our system and on these datasets we will apply different classification algorithms.

- **Data pre-processing**

In this step, we take the HCV liver fibrosis datasets from the internet in these datasets some irrelevant data are present and some missing values are present. We remove all these irrelevant data and use useful information for future work. ata will be standardized and normalized to achieve better results.

- **Data Balancing**

These datasets have a little imbalance, and they do not have a sufficient number of examples, which means that we should anticipate to see a low level of accuracy. The Synthetic Minority Oversampling Techniques, often known as **SMOTE**, is a method for oversampling data that is used to construct similar instances by making use of raw records. In order to normalize (**dataset-2, dataset-1**), remove any outliers, and divide the data set into a test group and a train group with a ratio of **80:20**.

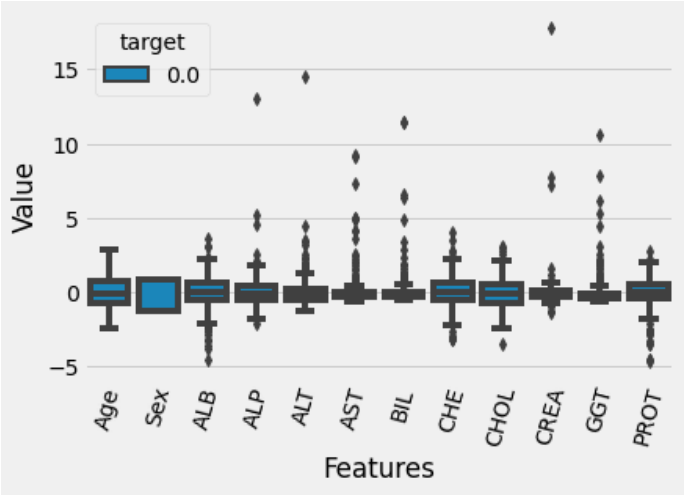

Figure 22 Normalize Dataset -2 (Standardization)

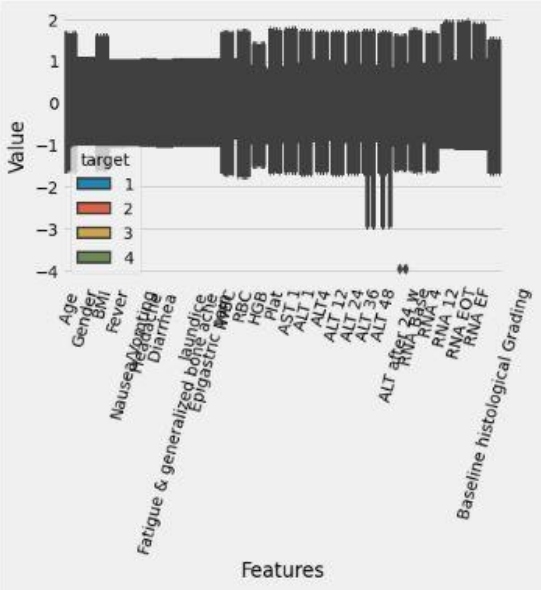

Figure 23 Normalize Dataset -1 (Standardization)

3: Feature Selection and Feature subsets

A machine learning model can be helped to discover acceptable features with the assistance of feature selection. This enables the model to reduce the likelihood of overfitting, save time and money, and deliver more accurate results. Chi-Square Attribute Evaluation (CSAE), Gain Ratio Attribute Evaluation (GRAE), Info Gain Attribute Evaluation (IGAE), and Relief (RFAE) are some of the filter-based feature selection approaches that will be implemented in order to rank and identify characteristics from both HCV datasets.

3. Results and implementation

In this section, classification will be implemented into HCV datasets using the programming language python. Additionally, several different classifiers, including Random Forest (RF), K-nearest neighbour (KNN), Decision Tree (DT), Naive Bayes (NB), Logistic Regression (LR), Xg boost classifier (XGB), and Support Vector Machine (SVM), will be used to conduct an analysis of HCV datasets.

Logistical Regression

- **Dataset-1:** We used the Scikit library for logistic regression, and with the assistance of the Logistic Regression header file, we were able to achieve an accuracy score using logistic regression that is as follows: average accuracies **0.2601801801801802**, standard deviation accuracies **0.04031923314334112**, and test accuracy **0.259927797833935**.
- **Dataset-2**: For Logistic Regression, we have used the Scikit library and with the help of the Logistic Regression header file, we have achieved an accuracy score using Logistic Regression is average accuracies **95.97%,** standard deviation accuracies **0.02403**, and test accuracy **92.37%**.

K-nearest neighbors (KNN)

- **Dataset-1**: For K-Nearest Neighbors, we have used the Scikit library and with the help of the Gaussian NB header file, we have achieved an accuracy score using KNN is score **0.367660343,** best training accuracy **0.28182637**, and test accuracy **0.205776173**.
- **Dataset-2**: For K-Nearest Neighbors, we have used the Scikit library and with the help of the Gaussian NB header file, we have achieved an accuracy score using KNN is score **91.52%,** best training accuracy **95.54%**, and test accuracy **94.06%**.

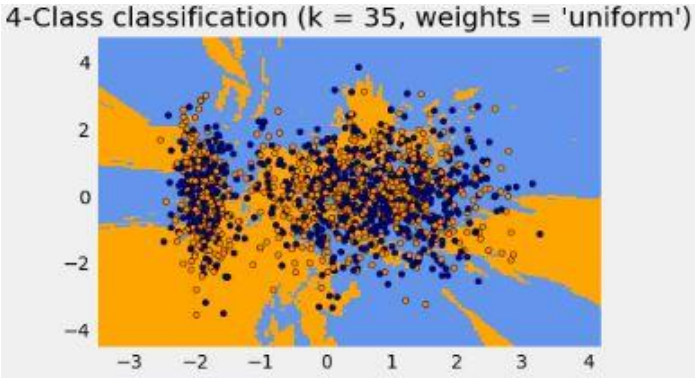

Figure 24 KNN 4-Classification (K=35, Weights = Uniform) –Dataset -1

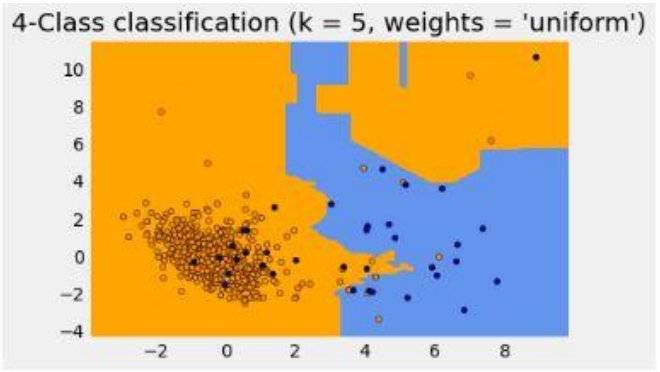

Figure 25 KNN 4-Classification (K=35, Weights = Uniform) –Dataset -2

Support Vector Machine (SVM)

- **Dataset-1:** For Support Vector Machine, we have used the Scikit library and with the help of the SVM header file, we have achieved an accuracy score using Support Vector Machine is average accuracies **0.27006644653** standard deviation accuracies **0.018097802**, and test accuracy **0.2563176895**.
- **Dataset-2:** For Support Vector Machine, we have used the Scikit library and with the help of the SVM header file, we have achieved an accuracy score using Support Vector Machine is average accuracies **94.47%,** standard deviation accuracies **0.010448**, and test accuracy **93.22%**.

Naïve Bayes (NB)

- **Dataset-1:** For Naïve Bayes (NB), we have used the Scikit library and with the help of the Gaussian NB header file, we have achieved an accuracy score using Naïve Bayes (NB) is average accuracies **0.254665898** standard deviation accuracies **0.060669878**, and test accuracy **0.263537906**.
- **Dataset-2:** For Naïve Bayes (NB), we have used the Scikit library and with the help of the Gaussian NB header file, we have achieved an accuracy score using Naïve Bayes (NB) is average accuracies **92.44%,** standard deviation accuracies **0.09500**, and test accuracy **87.28%**.

Decision Tree

- **Dataset-1:** For Decision Tree, we have used the Scikit library and with the help

of the Decision Tree Classifier header file, we have achieved an accuracy score using Decision Tree is average accuracies **0.26051948,** standard deviation accuracies **0.0934707211**, and test accuracy **0.25270758122**.

- **Dataset-2:** For Decision Tree, we have used the Scikit library and with the help of the Decision Tree Classifier header file, we have achieved an accuracy score using Decision Tree is average accuracies **94.61%,** standard deviation accuracies **0.0708080**, and test accuracy **89.83%**.

Random Forest

- **Dataset-1:** For Random Forest, we have used the Scikit library and with the help of the Random Forest Classifier header file, we have achieved the accuracy score using Random Forest average accuracies **0.250229320,** standard deviation accuracies **0.0343223700**, and test accuracy **0.205776173%.**
- **Dataset-2:** For Random Forest, we have used the Scikit library and with the help of the Random Forest Classifier header file, we have achieved the accuracy score using Random Forest average accuracies **95.33%,** standard deviation accuracies **0.028149**, and test accuracy **89.83%.**

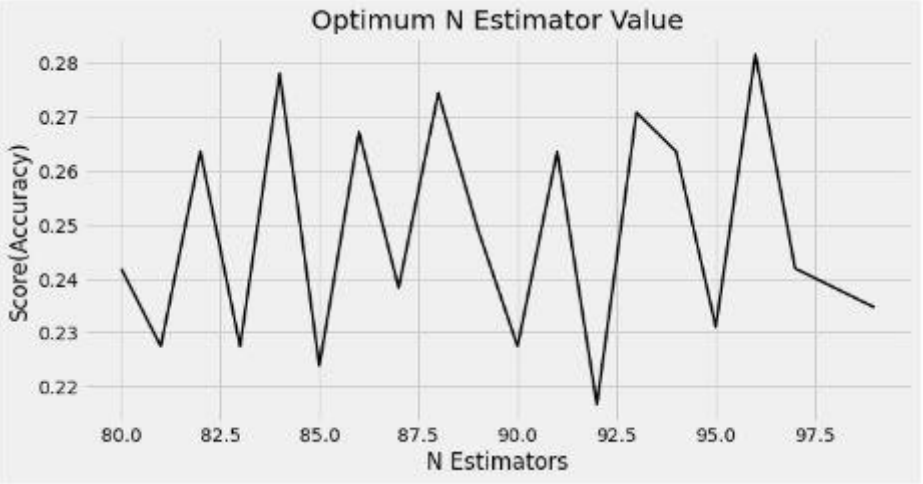

Figure 26 Random Forest Optimum N Estimator Data Set – 1

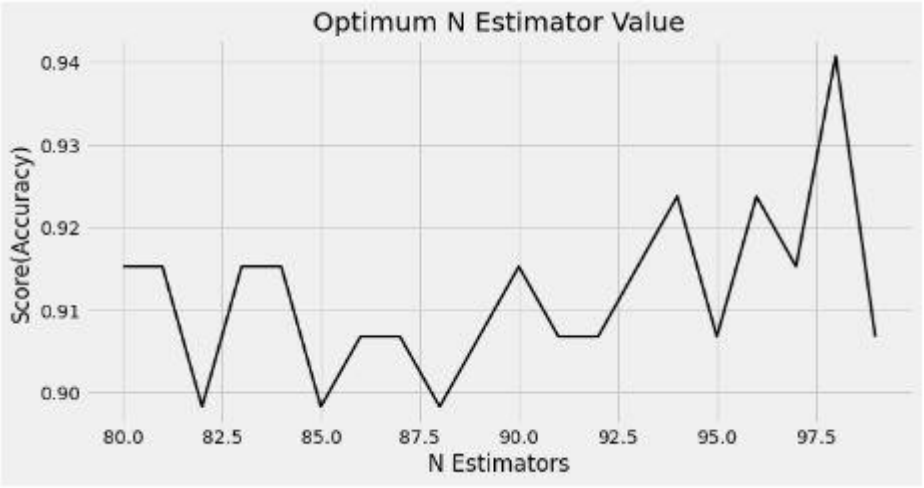

Figure 27 Random Forest Optimum N Estimator Data Set – 2

Artificial Neural Network

- **Dataset-1:** For ANN, we have used the Scikit library and with the help of the Keras Classifier header file, we have achieved the accuracy score using ANN accuracies **98.94%.** which is one of the highest accuracy as compare to all other machine learning algorithms.
- **Dataset-2:** For ANN, we have used the Scikit library and with the help of the Keras Classifier header file, we have achieved the accuracy score using ANN accuracies **98.94%.** which is one of the highest accuracy as compare to all other machine learning algorithms.

Gradient Boosting Machine (GBM)

- **Dataset-1:** For GBN, we have used the Scikit library and with the help of the Gradient Boosting Classifier header file, we have achieved the accuracy score using ANN accuracies **83.33%.** which is one of the highest accuracy as compare to all other machine learning algorithms.
- **Dataset-2:** For GBN, we have used the Scikit library and with the help of the Gradient Boosting Classifier header file, we have achieved the accuracy score using ANN accuracies **83.33%.** which is one of the highest accuracy as compare to all other machine learning algorithms.

Table 1 Accuracy Comparison Data set - 1

Algorithms	Avg Accuracy	SD Accuracy	Test Accuracy	F1-score
SVM	0.27%	0.01%	0.25%	0.84%
Decision Tree	0.26%	0.09%	0.25%	0.72%
Logistic Regression	0.26%	0.04%	0.26%	0.85%
KNN	0.36%	0.03%	0.20%	0.74%
Naïve Bayes	0.25%	0.06%	0.26%	0.84%
Random Forest	0.27%	0.03%	0.20%	0.87%

The above-mentioned table 2 illustrates that Accuracy comparison & performance evaluation of ML algorithms and our proposed model. The Random Forest model achieved accuracy of **87%**. This value is more efficient as compared to individual machine learning algorithms.

Table 2 : Accuracy Comparison Data set - 2

Algorithms	Avg Accuracy	SD Accuracy	Test Accuracy	F1-score
SVM	0.94%	0.01%	0.93%	0.84%
Decision Tree	0.94%	0.07%	0.89%	0.72%
Logistic Regression	0.95%	0.02%	0.92%	0.86%
KNN	0.94%	0.09%	0.87%	0.75%
Naïve Bayes	0.92%	0.09%	0.87%	0.84%
Random Forest	0.95%	0.02%	0.90%	0.88%

The above-mentioned table 3 illustrates that Accuracy comparison & performance evaluation of ML algorithms and our proposed model. The Random Forest model and Logistic Regression achieved same accuracy of **95%**. This value is more efficient as compared to individual machine learning algorithms.

4. Conclusion

In our study, four machine learning algorithms are applied for the Classification of **HCV LIVER FIBROSIS** patients. The dataset (dataset-1, dataset-2) that we have used in our study is publicly available on UCI. Our study evaluates the classification algorithms performance on **HCV LIVER FIBROSIS** patients by using Python language and improves the accuracy. It discovers that individual model is providing accuracy up to **95%** in dataset -2 and **87%** in dataset-1. This Highest accuracy has been achieved by Random Forest in both dataset (dataset-1, dataset-2).

Reference

1. A. Choudhury and D. Gupta, *A Survey on Medical Diagnosis of Diabetes Using Machine Learning*. Springer Singapore. doi: 10.1007/978-981-13-1280-9.
2. Khan, S. R., Raza, A., Shahzad, I., & Ijaz, H. M. (2024). Deep transfer CNNs models performance evaluation using unbalanced histopathological breast cancer dataset. Lahore Garrison University Research Journal of Computer Science and Information Technology, 8(1).
3. Bilal, Omair, Asif Raza, and Ghazanfar Ali. "A Contemporary Secure Microservices Discovery Architecture with Service Tags for Smart City Infrastructures." VFAST Transactions on Software Engineering 12, no. 1 (2024): 79-92.
4. S. Abhari, S. R. N. Kalhori, M. Ebrahimi, and H. Hasannejadasl, "Artificial Intelligence Applications in Type 2 Diabetes Mellitus Care : Focus on Machine Learning Methods," vol. 25, no. 4, pp. 248–261, 2019.
5. H. Kaur and V. Kumari, "Predictive Modelling and Analytics for Diabetes using a Machine Learning Approach," *Appl. Comput. Informatics*, no. December, 2018, doi: 10.1016/j.aci.2018.12.004.
6. I. M. Ibrahim and A. M. Abdulazeez, "The Role of Machine Learning Algorithms for Diagnosing Diseases," vol. 02, no. 01, pp. 10–19, 2021, doi: 10.38094/jastt20179.
7. A. Mujumdar and V. Vaidehi, "ScienceDirect ScienceDirect ScienceDirect ScienceDirect Diabetes Prediction using Machine Learning Aishwarya Mujumdar Diabetes Prediction using Machine Learning Aishwarya Mujumdar Aishwarya," *Procedia Comput. Sci.*, vol. 165, pp. 292–299, 2019, doi: 10.1016/j.procs.2020.01.047.
8. S. Brian and R. R. B. Pharmd, "Prediction of Nephropathy in Type 2 Diabetes: An Analysis of the ACCORD Trial applying Machine Learning Techniques," no. 317, doi: 10.1111/cts.12647.
9. HUSSAIN, S., RAZA, A., MEERAN, M. T., IJAZ, H. M., & JAMALI, S. (2020). Domain Ontology Based Similarity and Analysis in Higher Education. IEEEP New Horizons Journal, 102(1), 11-16.
10. Asif, S., Wenhui, Y., ur-Rehman, S., ul-ain, Q., Amjad, K., Yueyang, Y., ... & Awais, M. (2024). Advancements and Prospects of Machine Learning in Medical Diagnostics: Unveiling the Future of Diagnostic Precision. Archives of Computational Methods in Engineering, 1-31.
11. Asif, S., Zhao, M., Li, Y., Tang, F., Ur Rehman Khan, S., & Zhu, Y. (2024). AI-Based Approaches for the Diagnosis of Mpox: Challenges and Future Prospects. Archives of Computational Methods in Engineering, 1-33.
12. J. Li *et al.*, "International Journal of Medical Informatics Establishment of noninvasive diabetes risk prediction model based on tongue features and machine learning techniques," *Int. J. Med. Inform.*, vol. 149, no. August 2020, p. 104429, 2021, doi: 10.1016/j.ijmedinf.2021.104429.
13. S. G. Azevedo *et al.*, "System-Independent Characterization of Materials Using Dual-Energy Computed Tomography," *IEEE Trans. Nucl. Sci.*, vol. 63, no. 2, pp. 341–350, 2016, doi: 10.1109/TNS.2016.2514364.
14. B. Pranto, S. M. Mehnaz, E. B. Mahid, and I. M. Sadman, "Evaluating Machine Learning Methods for Predicting Diabetes among Female Patients in Bangladesh," 2020, doi: 10.3390/info11080374.
15. Shahzad, I., Khan, S. U. R., Waseem, A., Abideen, Z. U., & Liu, J. (2024). Enhancing ASD classification through hybrid attention-based learning of facial features. Signal, Image and Video Processing, 1-14.
16. A. T. Nagi, N. Ayesha, M. J. Awan, and R. Javed, "A Comparison of Two-Stage Classifier Algorithm with Ensemble Techniques On Detection of Diabetic Retinopathy," pp. 9–12, 2021.
17. H. Lu, S. Uddin, F. Hajati, M. Ali, and M. Matloob, "A patient network-based machine learning model for disease prediction : The case of type 2 diabetes mellitus," 2021.
18. Y. Sun and D. Zhang, "Machine Learning

Techniques for Screening and Diagnosis of Diabetes : a Survey," vol. 3651, pp. 872–880, 2019.
19. Meeran, M. T., Raza, A., & Din, M. (2018). Advancement in GSM Network to Access Cloud Services. Pakistan Journal of Engineering, Technology & Science [ISSN: 2224-2333], 7(1).
20. B. G. Choi, S. Rha, S. W. Kim, J. H. Kang, J. Y. Park, and Y. Noh, "Machine Learning for the Prediction of New-Onset Diabetes Mellitus during 5-Year Follow-up in Non-Diabetic Patients with Cardiovascular Risks," vol. 60, no. 2, pp. 191–199, 2019.
21. Khan, S.U.R.; Raza, A.;Waqas, M.; Zia, M.A.R. Efficient and Accurate Image Classification Via Spatial Pyramid Matching and SURF Sparse Coding. Lahore Garrison Univ. Res. J. Comput. Sci. Inf. Technol. 2023, 7, 10–23.
22. Farooq, M.U.; Beg, M.O. Bigdata analysis of stack overflow for energy consumption of android framework. In Proceedings of the 2019 International Conference on Innovative Computing (ICIC), Lahore, Pakistan, 1–2 November 2019; pp. 1–9.
23. F. Farrokhi *et al.*, "Investigating risk factors and predicting complications in deep brain stimulation surgery with machine learning algorithms," *World Neurosurg.*, 2019, doi: 10.1016/j.wneu.2019.10.063.
24. Raza, A., & Shahzad, I. (2024). Residual Learning Model-Based Classification of COVID-19 Using Chest Radiographs. Spectrum of engineering sciences, 2(3).
25. Farooq, M. U., Khan, S. U. R., & Beg, M. O. (2019, November). Melta: A method level energy estimation technique for android development. In 2019 International Conference on Innovative Computing (ICIC) (pp. 1-10). IEEE.
26. Khan, S. U. R., & Asif, S. (2024). Oral cancer detection using feature-level fusion and novel self-attention mechanisms. Biomedical Signal Processing and Control, 95, 106437.
27. Raza, A.; Meeran, M.T.; Bilhaj, U. Enhancing Breast Cancer Detection through Thermal Imaging and Customized 2D CNN Classifiers. VFAST Trans. Softw. Eng. 2023, 11, 80–92.
28. Dai, Q., Ishfaque, M., Khan, S. U. R., Luo, Y. L., Lei, Y., Zhang, B., & Zhou, W. (2024). Image classification for sub-surface crack identification in concrete dam based on borehole CCTV images using deep dense hybrid model. Stochastic Environmental Research and Risk Assessment, 1-18.
29. Khan, S.U.R.; Asif, S.; Bilal, O.; Ali, S. Deep hybrid model for Mpox disease diagnosis from skin lesion images. Int. J. Imaging Syst. Technol. 2024, 34, e23044.
30. Khan, S.U.R.; Zhao, M.; Asif, S.; Chen, X.; Zhu, Y. GLNET: Global–local CNN's-based informed model for detection of breast cancer categories from histopathological slides. J. Supercomput. 2023, 80, 7316–7348.
31. Khan, S.U.R.; Zhao, M.; Asif, S.; Chen, X. Hybrid-NET: A fusion of DenseNet169 and advanced machine learning classifiers for enhanced brain tumor diagnosis. Int. J. Imaging Syst. Technol. 2024, 34, e22975.
32. Mahmood, F., Abbas, K., Raza, A., Khan,M.A., & Khan, P.W. (2019). Three Dimensional Agricultural Land Modeling using Unmanned Aerial System (UAS). International Journal of Advanced Computer Science and Applications (IJACSA) [p-ISSN : 2158-107X, e-ISSN : 2156-5570], 10(1).
33. IoannisKavakiotis, Olga Tsave, Athanasios Salifoglou, and NicosMaglaveras, "Machine Learning and Data Mining Methods in Diabetes Research", Computational and Structural Biotechnology Journal, vol. 15, pp. 104– 116, 2017
34. Khan, U. S., & Khan, S. U. R. (2024). Boost diagnostic performance in retinal disease classification utilizing deep ensemble classifiers based on OCT. Multimedia Tools and Applications, 1-21.
35. Khan, M. A., Khan, S. U. R., Haider, S. Z. Q., Khan, S. A., & Bilal, O. (2024). Evolving knowledge representation learning with the dynamic asymmetric embedding model. Evolving Systems, 1-16.

36. Raza, A., & Meeran, M. T. (2019). Routine of encryption in cognitive radio network. Mehran University Research Journal of Engineering & Technology, 38(3), 609-618.
37. Al-Khasawneh, M. A., Raza, A., Khan, S. U. R., & Khan, Z. (2024). Stock Market Trend Prediction Using Deep Learning Approach. Computational Economics, 1-32.
38. Khan, U. S., Ishfaque, M., Khan, S. U. R., Xu, F., Chen, L., & Lei, Y. (2024). Comparative analysis of twelve transfer learning models for the prediction and crack detection in concrete dams, based on borehole images. Frontiers of Structural and Civil Engineering, 1-17.
39. Hekmat, A., Zhang, Z., Ur Rehman Khan, S., Shad, I., & Bilal, O. (2025). An attention-fused architecture for brain tumor diagnosis. Biomedical Signal Processing and Control, 101, 107221. https://doi.org/https://doi.org/10.1016/j.bspc.2024.107221