

Post-Analysis: ML Interpretability Using Shapley Values for Trained ML Models - (for STREAMLINE)

This notebook allows user to run additional post-analysis methods for in-depth interpretability on pipeline results using Shapley Values through the SHAP framework (created by Scott M. Lundberg and Su-In Lee) to help explain model prediction outputs and feature importances.

When run, 1) trained models are unpickled 2) repsective training and testing datasets to CV partitions are loaded 3) create SHAP explainers for each model 4) compute SHAP values using the explainer & testing dataset 5) generate and save various forms of figures that provide different perspectives/interpretations of feature importances 5) create csv files of feature importances for each model per CV partition 6) save a master list of SHAP values (mean shapley values) for each model that stores all values from each CV partition. Additionally, running force plots is optional,user must specify 'True' in order to generate the plots (force plots for some models cannot be saved and can only be displayed in the Run cell).

During the run of this notebook, all figures and csv files containing feature impomrtance SHAP values are stored in a main folder called "shap_values" under "model_evaluation" of the user's experiment dataset used in STREAMLINE.

Some figures (such as force plot) will display '0' or '1' as the 'average' model prediction but these refer to the classification of 0 or 1. This simply demonstrates that SHAP values of features contributing have a certain magnitude on the final prediction, 0 or 1. User will most likely see this when running Decision Tree, Logistic Regression, Extreme Gradient Boosting, and other tree-based models.

NOTE: Creation of the notebook took into account how to apply explainers, SHAP value computation, and plots appropriately to its respective ML models and is still a work in progress due to the continuous changes in the SHAP packages.

Goals

- Laid out a rough outline of how SHAP would be computed, I thought I would give SHAP methods a try
- Earlier methods work and prove that the model is unpickled and can be used
- Be able to iterate through each trained model to it's respective CV dataset, create shap values, generate shap plots
- Be able to store each CV shap values for each model and store in csv file as a DataFrame

```
LR_shap_all_CVs.csv ==>
                        LR_0 --> CV0
                        LR_1 --> CV1
                        LR_2 --> CV2
```

Progress of Updates/Fixes:

Months of June & July

- Still need to figure out saving results into a file (pickle.dump()), create and save into designated folder **DONE**
- Figure out how to work TreeExplainer, expected_value function **DONE**
 - expected_value works for certain models/Explainers --> doesn't really work for Naive Bayes
- Find file with the feature names for corresponding dataset to load into program under 'Load Metadata" section **DONE**
- Figure out how to display other shap plots such as waterfall, force plot, etc
 - waterfall plots might only work for certain models like tree-based ones
 - some of the summary figures functions aren't working
- Most of the program is hardcoded to specifically load one of the trained models after running STREAMLINE **resolved**
- Was able to prove that the model can be unpickled and used for .predict() and .predictproba() **resolved**
- Was able to use model to create SHAP explainers, calculate shap_values for CV0 testing dataset, and display plots **resolved**
- However, still need to refine the SHAP methods as there were some issues for Decision Tree Classifier **resolved**
- Was able to display Decision Tree prediction using TreeExplainer or even Explainer....I might be doing something wrong **resolved**
- XGBOOST MODEL IS COMPATIBLE WITH ALL OF THE LISTED SHAP PLOTS **resolved**
- RF MODEL NEEDED IT'S OWN IF-STATEMENT FOR NOW BUT WILL CONDENSE FOR CLARITY ADN EFFICIENCY **resolved**
- STILL NEED TO WORK ON LIGHTGBM, CATBOOST **resolved**
- GO BACK TO FIX DECISION TREE **resolved**
- Go back to double check shap plot compatibility for global and local importance for linear models **resolved**
- Work through the DecisionTreeClassifier and compare to other codes out there (if possible) **resolved**
- Currently unsure if creating dataframe for each model's shap_values shuold be done in compute_shap_values() or within the nested for-loop in testing cell **resolved**
- Feature names when displaying shap plots
 - Beeswarm plots still aren't displaying feature names (everything else is fine)

7/29/22

- ALL given SHAP plots seems to work for NB() when not in a defined function block and if-statement **resolved**
- Bar, scatter, waterfall, and beeswarm plots don't work for LR(), other plots work fine on LinearExplainer() and shap_values = explainer.shap_values(data) **STILL NEEDS TO BE FIXED**

8/02/22

- Plots and shap_values for each trained model in each CV work
- Currently unsure if creating dataframe for each model's shap_values shuold be done in compute_shap_values() or within the nested for-loop in testing cell

8/04/22

- Can create DataFrames for each CV but feature names most likely are not matching actual values (double check it)
- Difficult looping through to merge Dataframes for all CVs features...tried temporary variable
- Must also consider that shap_values array are returned in order of features from test/train set it was passed from...not based on feature order in test/train set **FIXED on 8/05/22**
 - Consider mapping out and ordering the values to avoid shuffling of names and values **FIXED on 8/05/22**

8/05/22

- Saving feature importance scores for each cv
- Created two different runs, one for actual test (default) and another if the user chooses to run it on the training sets for comparison

8/08/22

- Iterating through multiclass shap values for Decision Tree poses issue?...ideally we'd want to get the shap absolute average for both classes 0 and 1...same might be for XGB and any other model that has multiclass output **FIXED on 8/08/22**
 - Figured out that when running the loop in shap_feature_ranking() for Decision Tree, both classes 0 and 1 are accounted for. The shap absolute averages are summed up automatically to get the overall CV feature importances for the model (i double checked this myself through creating a loop that would output two different csv files for each class it iterated through)
- **Current issue:** Figuring out how to save multiple figures for each model when calling shap_summary()...for now, I can only save each figure individually through each CV...if model NB has 2 plot function calls & iterate through 3 CVs --> total 6 shap plots for **ONE** model
.....
 - **POSSIBLE FIX** merge all images onto one pdf per model which would entail different shap summaries **OR** create the master list of feature impmortance of all CVs for each model and create shap summaries for those

08/09/22 - 08/12/22

- Goal is to save shap averages frm all CVs in each model in a final/master CSV file **(08/09/22)**
- **CURRENT ISSUE:** Struggling to create a method to save the mean shap values for each feature into a dataframe composed of all averages across all CVs **(08/10/22 - 08/11/22) --> FIXED**
 - Was able to iterate through CV0 for NB0 but still need to figure out how to loop through all CVs **(08/12/22)**
 - Maybe possible fix for this is either:
 - Implement this as a cell alone at the bottom of the notebook (my concern is redundancy in calling compute_shapValues() as well as other functions)
 - Leave as function and figure out how to append to an existing DataFrame after the firs iteration
 - Implement within shap_feature_ranking()

08/15/22

- Still trying to figure out best way on how to loop through all CVs for each model's mean absolute shap_values
 - **NOTE:** it's better to append each CV values to FI_all list before creatibg dataframe....it would be inefficient to continuously append a row to an existing dataframe

8/16/22

- Continuing on fixing csv issue for saving absolute mean shap_values for each model across all CVs
- Worked on (and fixed) saving summary plots to a new folder withing shap_values test run as well as saving figures for force plots if user decides to run it (aka user-specified run)
 - **NOTE:** SHAP force plots for DT, RF, LR, and XGB can only be displayed in the 'Run' cell of the notebook. These figures can be saved as png but it only shows blank space it works for NB

8/17/22

- Was able to loop through alls CVs for master list with for given models but values stored in csv file **do not correspond to the output in run cell**
 - **CURENT ISSUE** Seems like when concating new rows, previous row gets overwritten with different values than the original dataframe row **FIXED 08/18/22**

8/18/22

- Fixed the issue with saving each CV of mean shap values per model --> saves and creates master list of shap values for each model successfully
- Will be focused on tesing out the code with other ML algorithms and cleaning up code

8/29/22

- Was able to generate and save waterfall plots for a single preediction (only) for models LR, XGB, DT, and RF
 - * figure out why NB is labeled as 'Permutation' --> this is causing issues in displaying plots that require 'explainer.expected_value' parameter

Final Steps

- Saving shap figures per model in each cv
- Make sure you can loop through each pickled model, load it, create shap values and display plots
- Be able to load one model at a time, create shapley values for each CV train and test set, store shap scores in a dataframe
- Make sure to load original dataset features so that each csv file is the same length as the original dataset
 - This means when a CV dataset is missing a feature, we make sure to assign a shap score of 0
 - each new csv file for loading shap scores of each trained model must include all features

```
LR_shap_all_CVs.csv ==>
                        LR_0 --> CV0
                        LR_1 --> CV1
                        LR_2 --> CV2
```

- Save dataframe for each model in a csv file

Possible future changes to consider: changing shap_summary()

Instead of using the CV loop to create shap figures for each model per CV....

- Open & read masterList csv
- sum each feature over all CVs to get shap_values average of each feature for the model
 - CV0 Feature X shap_value + CV1 Feature X shap_value + CV2 Feature X shap_value = (SUM(shap_values) / # of CVs)
 - Feature shap_values average for Model A = (SUM(shap_values) / # of CVs)
- create shap figures using finalized shap_value average of features for each model

In []:

```
In [1]: # required packages
import os
import sys
import glob
import pickle
import warnings
warnings.filterwarnings('ignore')
import csv
import sklearn
import random
import shap
import numpy as np
import numpy.typing as npt
import pandas as pd
import scipy as sp
import matplotlib.pyplot as plt
from matplotlib.backends.backend_pdf import PdfPages
import itertools
from itertools import chain
from fpdf import FPDF
import collections
from termcolor import colored as cl #text customization

# Model packages
import xgboost
import lightgbm as lgb
from sklearn import *
from sklearn import tree
from shap.plots import waterfall

# Jupyter Notebook Hack: This code ensures that the results of multiple commands within a given cell are all displayed
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"

shap.initjs() # load JS visualization code to notebook. SHAP plots won't be displayed without this
```



Run Parameters

```
In [2]: dataset_path = "/Users/jessicakim/Desktop/STREAMLINE/DemoData"
experiment_path = "/Users/jessicakim/Desktop/STREAMLINE/DemoData/Output/hcc_demo"
```

Check for Analyzed Datasets and Remove Unecessary Files

```
In [3]: datasets = os.listdir(experiment_path)
experiment_name = experiment_path.split('/')[-1] #Name of experiment folder

datasets.remove('metadata.csv')
datasets.remove('metadata.pickle')
datasets.remove('algInfo.pickle')

try:
    datasets.remove('jobsCompleted')
except:
    pass
try:
    datasets.remove('UsefulNotebooks')
except:
    pass
try:
    datasets.remove('logs')
    datasets.remove('jobs')
except:
    pass
try:
    datasets.remove('DatasetComparisons') #If it has been run previously (overwrite)
except:
    pass
try:
    datasets.remove('KeyFileCopy') #If it has been run previously (overwrite)
except:
    pass
try:
    datasets.remove('.DS_Store') #If it has been run previously (overwrite)
except:
    pass
try:
    datasets.remove(experiment_name+'_ML_Pipeline_Report.pdf') #If it has been run previously (overwrite)
except:
    pass

datasets = sorted(datasets) #ensures consistent ordering of datasets
print("Analyzed Datasets: "+str(datasets))
```

Analyzed Datasets: ['hcc-data_example', 'hcc-data_example_no_covariates']

Load Metadata and Other Necessary Variables

```
In [4]: jupyterRun = 'True'
# Loading necessary variables specified earlier in the pipeline from metadatafor dataPrep()
file = open(experiment_path + '/' + "metadata.pickle", 'rb')
metadata = pickle.load(file)
# file.close()
# print(metadata)

class_label = metadata['Class Label']
instance_label = metadata['Instance Label']
cv_partitions = int(metadata['CV Partitions'])

# # # unpickle and load in feature_names from original dataset
original_headers = pd.read_csv(experiment_path+"/hcc-data_example/exploratory/OriginalFeatureNames.csv",sep=',').column
print(original_headers)

alg_file = open(experiment_path + '/' + "/algInfo.pickle", 'rb')
algInfo = pickle.load(alg_file)
alg_file.close()
algorithms = []

abbrev = {}
for key in algInfo: # pickling specific model while also checking for corresponding algInfo
    if algInfo[key][0]: # If that algorithm was used
        algorithms.append(key)
        abbrev[key] = (algInfo[key][1])

print(f'\nChecking for algorithms used in STREAMLINE...\n{algorithms}')
print(f'\nChecking for abbrev for algorithms used in STREAMLINE...\n{abbrev}')
```

```
['Gender', 'Symptoms ', 'Alcohol', 'Hepatitis B Surface Antigen', 'Hepatitis B e Antigen', 'Hepatitis B Core Antibod
y', 'Hepatitis C Virus Antibody', 'Cirrhosis', 'Endemic Countries', 'Smoking', 'Diabetes', 'Obesity', 'Hemochromatosi
s', 'Arterial Hypertension', 'Chronic Renal Insufficiency', 'Human Immunodeficiency Virus', 'Nonalcoholic Steatohepati
tis', 'Esophageal Varices', 'Splenomegaly', 'Portal Hypertension', 'Portal Vein Thrombosis', 'Liver Metastasis', 'Radi
ological Hallmark', 'Age at diagnosis', 'Grams of Alcohol per day', 'Packs of cigarets per year', 'Performance Status
*', 'Encephalopathy degree*', 'Ascites degree*', 'International Normalised Ratio*', 'Alpha-Fetoprotein (ng/mL)', 'Haem
oglobin (g/dL)', 'Mean Corpuscular Volume', 'Leukocytes(G/L)', 'Platelets', 'Albumin (mg/dL)', 'Total Bilirubin(mg/d
L)', 'Alanine transaminase (U/L)', 'Aspartate transaminase (U/L)', 'Gamma glutamyl transferase (U/L)', 'Alkaline phosp
hatase (U/L)', 'Total Proteins (g/dL)', 'Creatinine (mg/dL)', 'Number of Nodules', 'Major dimension of nodule (cm)',
'Direct Bilirubin (mg/dL)', 'Iron', 'Oxygen Saturation (%)', 'Ferritin (ng/mL)']
```

```
Checking for algorithms used in STREAMLINE...
['Naive Bayes', 'Logistic Regression', 'Decision Tree', 'Random Forest', 'Extreme Gradient Boosting']
```

```
Checking for abbrev for algorithms used in STREAMLINE...
{'Naive Bayes': 'NB', 'Logistic Regression': 'LR', 'Decision Tree': 'DT', 'Random Forest': 'RF', 'Extreme Gradient Boo
sting': 'XGB'}
```

dataPrep(): Loading Target CV Training & Testing Sets

```
In [5]: def dataPrep(train_file_path,instance_label,class_label, test_file_path):

        '''Loads target cv training dataset, separates class from features and removes instance labels'''

        train = pd.read_csv(train_file_path)
        if instance_label != 'None':
            train = train.drop(instance_label,axis=1)

        # get feature names from train dataset
        trainFeat = list(train.drop(class_label, axis=1).columns) #note: datatype --> list
        set(itertools.chain(*trainFeat))

        trainX = pd.DataFrame(train.drop(class_label, axis=1).values)
        trainY = pd.DataFrame(train[class_label].values)
        del train #memory cleanup

        test = pd.read_csv(test_file_path)
        if instance_label != 'None':
            test = test.drop(instance_label,axis=1)

        # get feature names from test dataset
        testFeat = list(test.drop(class_label, axis=1).columns)
        set(itertools.chain(*testFeat))

        testX = pd.DataFrame(test.drop(class_label, axis=1).values)
        testY = pd.DataFrame(test[class_label].values)
        del test #memory cleanup

        return trainX, trainY, testX, testY, trainFeat, testFeat
```

SHAP: get_explainer()

(NOTE: Have not included KNN, SVM, ANN yet)

- if algorithm name matches ['list model names'], create explainers
- return explainer based on given model from parameter

Types of SHAP Explainers

Source: <https://shap-lrjball.readthedocs.io/en/latest/api.html#core-explainers>

.Explainer()

- Uses Shapley values to explain any machine learning model or python function.
- This is the primary explainer interface for the SHAP library
- It takes any combination of a model and masker and returns a callable subclass object that implements the particular estimation algorithm that was chosen.

.TreeExplainer()

- Uses Tree SHAP algorithms to explain the output of ensemble tree models.
- Tree SHAP is a fast and exact method to estimate SHAP values for tree models and ensembles of trees, under several different possible assumptions about feature dependence.
- It depends on fast C++ implementations either inside an external model package or in the local compiled C extention.

.LinearExplainer()

- Computes SHAP values for a linear model, optionally accounting for inter-feature correlations.
- This computes the SHAP values for a linear model and can account for the correlations among the input features.
- Assuming features are independent leads to interventional SHAP values which for a linear model are `coef[i] * (x[i] - X.mean(0)[i])` for the *i*th feature.
- If instead we accountfor correlations then we prevent any problems arising from colinearity and share credit among correlated features.
- Accounting for correlations can be computationally challenging, but LinearExplainer uses sampling to estimate a transform that can then be applied to explain any prediction of the model.

.KernelExplainer()

- Uses the Kernel SHAP method to explain the output of any function.
- Kernel SHAP is a method that uses a special weighted linear regression to compute the importance of each feature.
- The computed importance values are Shapley values from game theory and also coefficents from a local linear regression.

```
In [6]: def get_explainer(model, abbrev, trainX):

    '''Pass loaded model and abbrev to match appropriate SHAP explainer'''

    '''Must always use training dataset as background data in order to
        evaluate SHAP values for either testing (usually)or training set'''

    explainer = None
    trained_model = model

    if abbrev in ["NB"]:
        explainer = shap.Explainer(trained_model.predict, trainX)

        # dont use model.predict for Linear Explainer (only for Explainer using Naive Bayes)
        # ^^ You get a class method error when creating shap plots and values
    if abbrev in ["LR"]:
        explainer = shap.LinearExplainer(trained_model, trainX)

    # not tested yet with LGB, CGB
    if abbrev in ['DT', 'RF', "XGB", "LGB","CGB"]:
        explainer = shap.TreeExplainer(trained_model)

    # not tested yet with KNN, SVM, ANN
    if abbrev in ['KNN','SVM','ANN']:
#         explainer = shap.KernelExplainer(trained_model.predict, trainX)
        explainer = shap.KernelExplainer(trained_model.predict_proba, trainX) # SVM seen with this

    return explainer
```

SHAP: compute_shapValues()

(NOTE: Have not included KNN, SVM, ANN yet)

NOTES

- Parameter 'X' in this context refers to whatever training or testing dataset that was passed in from the whole run from below
- Mentioned earlier, default run uses training dataset as background data and creates shap values using testing data
- The same follows for feature_names --> either train_feat or test_feat (default) will be passed

```
In [7]: def compute_shapValues(model, abbrev, explainer, X):

    '''This method will calculate shapley values and store these as a Pandas DataFrame for conversion to csv file
        This includes creating expected_values and shap_values --> returns shap_values (will be called by shap_summary)
    '''

    max_evals = max(500, (2 * len(X)) + 1)    # optional: declares number of permutations for shap.Explainer()
    shap_values = None

    if abbrev in ['NB']:
        shap_values= explainer(X)  # permutation object cannot use .expected_value function like LR
        print(shap_values)

    # not tested yet for KNN, SVM, ANN
    # SVM and KNN can use explainer.expected_value[0] when creating force plots
    if abbrev in ['LR', 'KNN', 'SVM', 'ANN']:
        shap_values = explainer.shap_values(X)
        print(shap_values)

    # i think shap_values() only works for TreeExplainer and LinearExplainer...Explainer for NB is considered a
    # permutation object
    # not tested yet for LGB, CGB
    if abbrev in ['DT', 'RF', "XGB", "LGB", "CGB"]:
```



```
shap_values = explainer.shap_values(X, approximate=False, check_additivity=False)
print(shap_values)

return shap_values
```

Able to confirm that shap_values are calculated in order based on data features passed in

shap_value feature importance method just orders features & values based on importance but feature:value remains the same

SHAP: shap_summary()

(NOTE: Have not included KNN, SVM, ANN yet)

Plot Types for SHAP v0.41.0

Waterfall

- Plots an explnation of a single prediction as a waterfall plot

Summary (type: violin & bar)

- Summary plots of SHAP values across a whole dataset

Dependence

- Plots the value of the feature on the x-axis and the SHAP value of the same feature on the y-axis
- This shows how the model depends on the given feature, and is like a richer extenstion of the classical parital dependence plots.
- Vertical dispersion of the data points represents interaction effects.
- Grey ticks along the y-axis are data points where the feature's value was NaN.

Force

- Visualize cumulative SHAP values with an additive force layout.

Beeswarm

- Summary plots of SHAP values across a whole dataset
- Designed to display an information-dense summary of how the top features in a dataset impact the model's output.

```
In [8]: def shap_summary(abbrev, feature_names, shap_values, explainer, X, cvCount, save_path, dataset):
        '''Retrieve shap_values from previous method;
            this method will return and display different types of shap plots

            Figures for each model CV is saved as a png which will be merged to a
            final summary report for each model
        '''

        # Make folder in shapFigures for all summary plots generated from this function
        # helps to differentiate summary plots from user-chosen force_plots
        # save_path is used as a parameter for saving shap figure for each model
        if not os.path.exists(save_path+'SummaryPlots'):
            os.mkdir(save_path+'SummaryPlots')

        # generates random number between range 0 and 54 (length of shap_values = 55)
        # as a random single prediction for force plots
        random_single_predict = random.randint(0, len(shap_values)-1)

        # checks algorithm in given list to execute shap summaries
        if abbrev in ["NB"]:

            print(f'Saving Summary Plot for SHAP Values in Class 0 & 1 in {dataset} Set...')
            shap.summary_plot(shap_values, X, feature_names, plot_type='violin', show=False)
            plt.savefig(f'{save_path}/SummaryPlots/{abbrev}_{str(cvCount)}_shapSummaryViolinPlot.png', bbox_inches='tight')
            plt.close()

            print(f'Saving Summary Plot for SHAP Values in Class 0 & 1 in {dataset} Set...')
            shap.summary_plot(shap_values, X, feature_names, plot_type='bar', show=False)
            plt.savefig(f'{save_path}/SummaryPlots/{abbrev}_{str(cvCount)}_shapSummaryBarPlot.png', bbox_inches='tight')
            plt.close()

            # print('SHAP Bar Plot for Summary Plot for SHAP Values in Class 0 & 1 in Test Set:\n')
            # shap.plots._bar.bar_legacy(shap_values, feature_names, show=True) # doesnt work but should for this...attrib

            print(f'Saving SHAP Beeswarm Plot for Top 5 SHAP Values in Class 0 & 1 in {dataset} Set...')
            shap.plots.beeswarm(shap_values, max_display=5, show=False) #max_display allows user to choose # of features
            plt.savefig(f'{save_path}/SummaryPlots/{abbrev}_{str(cvCount)}_beeswarmPlot.png', bbox_inches='tight')
            plt.close()

        # print('Waterfall Plot for SHAP Values in Class 0 in Test Set: \n')
```

```
#         shap.plots._waterfall.waterfall_legacy(explainer.expected_value[0], shap_values[0][0], testX.iloc[0], testFe

# scatter, bar, waterfall, beeswarm plots should work for this model
# waterfall plot also doesnt work...i get "AttributeError: 'numpy.ndarray' object has no attribute 'base_value
# Bar plot should work for this model if using .Explainer() and shap_values = explainer(data)-->
#         not explainer.shap_values
elif abbrev in ["LR", 'XGB']:

    expected_value = explainer.expected_value
    print(f'Expected value for {abbrev}: {expected_value}')

    print(f'Saving Summary Plot for SHAP Values in {dataset} Set...')
    shap.summary_plot(shap_values, X, feature_names, plot_type='violin', show=False)
    plt.savefig(f'{save_path}/SummaryPlots/{abbrev}_{str(cvCount)}_shapSummaryViolinPlot.png', bbox_inches='tight')
    plt.close()

    print(f'Saving SHAP Bar Plot for SHAP Values {dataset} Set...')
    shap.summary_plot(shap_values, X, feature_names, plot_type="bar", show=False)
    plt.savefig(f'{save_path}/SummaryPlots/{abbrev}_{str(cvCount)}_shapSummaryBarPlot.png', bbox_inches='tight')
    plt.close()

    print(f'Saving SHAP Decision Plot for SHAP Values in {dataset} Set...')
    shap.decision_plot(expected_value, shap_values, feature_names, show=False)
    plt.savefig(f'{save_path}/SummaryPlots/{abbrev}_{str(cvCount)}_shapDecisionPlot.png', bbox_inches='tight')
    plt.close()

    print(f'Saving SHAP Decision Plot for Single-Prediction in {dataset} Set...')
    shap.decision_plot(expected_value, shap_values[54], feature_names, show=False)
    plt.savefig(f'{save_path}/SummaryPlots/{abbrev}_{str(cvCount)}_shapDecisionPlot_singlePredict.png', bbox_inche
    plt.close()

    print(f'Saving Waterfall Plot for SHAP Values for a Single-Prediction in {dataset} Set...')
    shap.plots._waterfall.waterfall_legacy(explainer.expected_value, shap_values[random_single_predict], testX.ilo
    plt.savefig(f'{save_path}/SummaryPlots/{abbrev}_{str(cvCount)}_WaterfallPlot_singlePredict.png', bbox_inches='
    plt.close()

#         shap.plots._bar.bar_legacy(expected_value, shap_values[10], testX[10], feature_names, show=True)

#         # waterfall plot works for DT() if it uses .Explainer() and shap_vales = explainer(data)
#         # instead of using TreeExplainer but other plots listed here work
elif abbrev in ['DT', 'RF', 'LGB','CGB']:
    expected_value = explainer.expected_value
    print(f'Expected value for {abbrev}: {expected_value}')

    print(f'Saving Bar Summary Plot for SHAP Values in Class 0 & 1 in {dataset} Set...')
    #         #tree.tree_plot(testX)  ---> helps display Decision Tree
    shap.summary_plot(shap_values, X, feature_names, plot_type='bar', class_names=['0', '1'], show=False)
    plt.savefig(f'{save_path}/SummaryPlots/{abbrev}_{str(cvCount)}_shapSummaryBarPlot.png', bbox_inches='tight')
    plt.close()

    print(f'Saving Decision Plot for SHAP Values from Class 0 in {dataset} Set...')
    shap.decision_plot(expected_value[0], shap_values[0], feature_names=feature_names, show=False)
    plt.savefig(f'{save_path}/SummaryPlots/{abbrev}_{str(cvCount)}_DecisionPlot_Class0.png', bbox_inches='tight')
    plt.close()

    print(f'Saving Decision Plot for SHAP Values from Class 1 in {dataset} Set...')
    shap.decision_plot(expected_value[1], shap_values[1], feature_names=feature_names, show=False)
    plt.savefig(f'{save_path}/SummaryPlots/{abbrev}_{str(cvCount)}_DecisionPlot_Class1.png', bbox_inches='tight')
    plt.close()

    print(f'Saving Waterfall Plot for SHAP Values from Class 0 in {dataset} Set...')
    shap.plots._waterfall.waterfall_legacy(explainer.expected_value[0], shap_values[0][random_single_predict], tes
    plt.savefig(f'{save_path}/SummaryPlots/{abbrev}_{str(cvCount)}_WaterfallPlot_Class0_singlePredict.png', bbox_i
    plt.close()

    print(f'Saving Waterfall Plot for SHAP Values from Class 1 in {dataset} Set...')
    shap.plots._waterfall.waterfall_legacy(explainer.expected_value[1], shap_values[1][random_single_predict], tes
    plt.savefig(f'{save_path}/SummaryPlots/{abbrev}_{str(cvCount)}_WaterfallPlot_Class1_singlePredict.png', bbox_i
    plt.close()
```

SHAP: run_force_plots()

(NOTE: Have not included KNN, SVM, ANN yet)

OPTIONAL

- Only runs when user sets 'run_force=True' (default is run_force = False)


```
In [9]: def run_force_plots(abbrev, explainer, shap_values, X, feature_names, cvCount, save_path, dataset):
'''This method is optional but will save force plots for each given trained model per CV

By default, run_force_plots=False but can be run if user sets run_force_plots=True in
the 'Run' cell at the bottom of the notebook'''

# Make folder in shapFigures called ForcePlots to organize user-chosen figures
# save_path is used as a parameter for saving shap figures for each model

if not os.path.exists(save_path+'/ForcePlots'):
    os.mkdir(save_path+'/ForcePlots')

# generates random number between range 0 and 54 (length of shap__values = 55)
# as a random single prediction for force plots
random_single_predict = random.randint(0, len(shap_values)-1)

if abbrev in ['NB']:
    print(f'Saving Force Plot for {abbrev} SHAP Values in {dataset} Set...\n')
    shap.force_plot(shap_values[random_single_predict], X, feature_names=feature_names, matplotlib=True, show=False)
    plt.title(f'{abbrev}{cvCount} Force Plot for Single Prediction Classification 0 or 1')
    plt.savefig(f'{save_path}/ForcePlots/{abbrev}{str(cvCount)}_singlePredictFP.png', dpi=600, bbox_inches='tight')
    plt.close()

# can only return plot when running code, png can be saved but plots don't show up
# matplotlib doesn't support output for multiple samples just like DT and RF
elif abbrev in ['LR', 'XGB']: # does not allow .savefig save force plot for multiple samples
    print(f'\nDisplaying Force Plot for {abbrev} SHAP Values in Whole {dataset} Set...')
    return shap.force_plot(explainer.expected_value, shap_values, feature_names, show=True)
# plt.savefig(save_path+abbrev+'_'+str(cvCount)+'ForcePlot.png', bbox_inches='tight')
# plt.close()

# Decision Tree has multiclass output so needed to create two separate function calls
# Decision Tree doesn't work when just using shap_values as a parameter
# can only return plot when running code, png can be saved but plots don't show up
# matplotlib doesn't support output for multiple samples just like LR and XGB
else:
    print(f'\nDisplaying Force Plot for {abbrev} SHAP Values from Class 0 in {dataset} Set...')
    return shap.force_plot(explainer.expected_value[0], shap_values[0], feature_names, show=True)
# plt.savefig(f'{save_path}/ForcePlots/{abbrev}{str(cvCount)}_singlePredictFP.png', dpi=600, bbox_inches='tight')
# plt.close()

    print(f'\nDisplaying Force Plot for {abbrev} SHAP Values from Class 1 in {dataset} Set...')
    return shap.force_plot(explainer.expected_value[1], shap_values[1], feature_names, show=True)
# plt.savefig(save_path+abbrev+'_'+str(cvCount)+'singlePredictFP.png', bbox_inches='tight')
# plt.close()
```

SHAP: shap_feature_ranking()

(NOTE: Have not included KNN, SVM, ANN yet)

Calculating feature importance for each trained model per CV

```
In [10]: def shap_feature_ranking(abbrev, shap_values, X, feature_names): # 'X' and 'feature_names' argument is whichever test set was used

'''Calculate the average of the absolute SHAP values for each feature and use it to show
which features were the most important when making a prediction'''

if abbrev in ['NB']:
    feature_order = np.argsort(np.mean(np.abs(shap_values.values), axis=0))
    shap_means = (np.mean(np.abs(shap_values.values), axis=0))
    df = pd.DataFrame({"Features": [feature_names[i] for i in feature_order][::-1], "Importance": [ np.mean(np.abs(shap_values.values), axis=0)[i] for i in feature_order][::-1]})

#LR cant use shap_values.values
elif abbrev in ['LR', 'LGB', 'XGB', 'CGB']:
    feature_order = np.argsort(np.mean(np.abs(shap_values), axis=0))
    shap_means = (np.mean(np.abs(shap_values), axis=0))
    df = pd.DataFrame({"Features": [feature_names[i] for i in feature_order][::-1], "Importance": [ np.mean(np.abs(shap_values), axis=0)[i] for i in feature_order][::-1]})

else: # For multiclass models (can be used for NB)..Loops through Class 0 and Class 1
    # Sums up the shap average values form both classes to get the shap average for the whole CV for the model
    # The solution for the 'else' statement was found on StackOverflow

    c_idx = []
    columns = feature_names
    for column in range(0, (len(columns))):
        if isinstance(shap_values, list):
            c_idx.append(X.columns.get_loc(column))
            means = [np.abs(shap_values[class_][:, c_idx]).mean(axis=0) for class_ in range(len(shap_values))]
            shap_means = np.sum(np.column_stack(means), 1)
        else: # Else there is only one 2D array of shap values
            assert len(shap_values.shape) == 2, 'Expected two-dimensional shap values array.'
            shap_means = np.abs(shap_values).mean(axis=0)
    df = pd.DataFrame({"Features": feature_names, 'Importance': shap_means}).sort_values(by='Importance', ascending=False)
```

```
df.index += 1

return df, shap_means
```

SHAP: save_shap()

(NOTE: Have not included KNN, SVM, ANN yet)

Function that saves shap values using shap_means taken from shap_feature_ranking() and saves all values from all CVs into one master list per trained model

```
In [11]: def save_shap(abbrev, shap_values, shap_means, original_headers, cvCount, dataset, each): # 'df' parameter is the data
        '''Create a new dataframe that stores the model's absolute mean SHAP feature importance values over each CV
            and combines with features from original dataset later on

            temp_list[] is returned as an array so that each CV iterated here can be stored into FI_all in the Run
            cell of the Notebook'''

        temp_list = []
        shap_vals = []

        headers = pd.read_csv(f'{experiment_path}/{each}/CVDatasets/{each}_CV_{cvCount}_{dataset}.csv').columns.values.tolist()

        if instance_label != 'None':
            headers.remove(instance_label)
            headers.remove(class_label)

        shap_vals = np.array(shap_means)

        for name in original_headers:
            if name in headers:
                index = headers.index(name)
                print(f'Checking for matches...{name} is {name}')
                print(f'{name} value is {shap_vals[index]}\n')
                temp_list.append(shap_vals[index])
            else:
                temp_list.append(0.0)

        return np.asarray(temp_list)
```

Unit Testing Cell

```
In [12]: # # # # # ^^^^
# dr = pd.DataFrame(columns=original_headers)
# dr.head()
# dr2 = pd.DataFrame()
# dr2.head()
# if not os.path.exists(full_path+'/model_evaluation/shap_values/testResults/shapFigures'):
#     os.mkdir(full_path+'/model_evaluation/shap_values/testResults/shapFigures')
# save_path = full_path + '/model_evaluation/shap_values/testResults/shapFigures'

# FI_all = []
# shap_vals = []
# temp_df = pd.DataFrame()

# for cvCount in range(0, cv_partitions):

# result_file = experiment_path+ '/hcc-data_example/models/pickledModels/NB_0.pickle'
# file = open(result_file, 'rb')
# model = pickle.load(file)
# file.close()
# print('\nChecking if correct model is loaded...\n', model)

# train_path = experiment_path + '/hcc-data_example/CVDatasets/hcc-data_example_CV_0_Train.csv'
# test_path = experiment_path + '/hcc-data_example/CVDatasets/hcc-data_example_CV_0_Test.csv'
# trainX, trainY, testX, testY, trainFeat, testFeat = dataPrep(train_path,instance_label,class_label, test_path)

# explainer = get_explainer(model, 'NB', trainX)

# shap_values = compute_shapValues(model, 'NB', explainer, testX)
# shap-fi_df, shap_means = shap_feature_ranking('LR', shap_values, testX, testFeat)
# # display(shap-fi_df)
# shapFI_path = (f'{filepath}{abbrev[algorithm]}_{str(cvCount)}_shapFIValues_Test.csv')
# shap-fi_df.to_csv(shapFI_path, header=True, index=True)
```

```
# headers = pd.read_csv(f'{experiment_path}/hcc-data_example/CVDatasets/hcc-data_example_CV_{str(cvCount)}_Test.csv')
# if instance_label != 'None':
#     headers.remove(instance_label)
# headers.remove(class_label)

# temp_list = []

# vals = np.mean(np.abs(shap_values), axis=0)
# shap_vals = np.array(vals)
# for name in original_headers:
#     if name in headers:
#         index = headers.index(name)
#         print(f'Checking for matches...{name} is {name}')
#         print(f'Index is {index}')
#         print(f'Shap value is {shap_vals[index]}\n')
#         print(f'{name} value is {shap_vals[index]}')
#         temp_list.append(shap_vals[index])

#     else:
#         temp_list.append(0.0)
# FI_all.append(temp_list)

# del temp_list

# print(np.asarray(FI_all))
# df = pd.DataFrame(np.asarray(FI_all), columns=[original_headers])
# display(df)
#
# # print(np.mean(np.abs(shap_values.values)))
# shap_summary('NB', testFeat, shap_values, explainer, testX, 0, save_path, 'Test')
# run_force_plots('DT', explainer, shap_values, testX, testFeat, cvCount, save_path)
```

In []:

Run SHAP for Testing Datasets

Loop through each hcc_demo dataset to unpickle and load trained models to create Shapley values and plots Default run

- The default setting runs explainer and shap values for the TESTING datasets for each model and CV
- User has the option below to run the loop for training sets as well

In [13]:

```
# testing all methods
run_force = True # parameter in run_force_plot(); set to True if user wants to display force plots for trained models
run_test = True

if run_test == True:
    for each in datasets:
        print("-----")
        print(each)
        print("-----")
        full_path = experiment_path+'/' + each
        filepath = f'/{full_path}/model_evaluation/shap_values/testResults/' #path to save SHAP FI value results

        #Make folder in experiment folder/datafolder to store all shap_values per algorithm/CV combination
        if not os.path.exists(full_path+'/model_evaluation/shap_values/testResults'):
            os.mkdir(full_path+'/model_evaluation/shap_values/testResults')

        if not os.path.exists(full_path+'/model_evaluation/shap_values/testResults/shapFigures'):
            os.mkdir(full_path+'/model_evaluation/shap_values/testResults/shapFigures')
        save_path = full_path + '/model_evaluation/shap_values/testResults/shapFigures'

        for algorithm in algorithms: #loop through algorithms
            print("-----")
            print(abbrev[algorithm])
            print("-----")

            if not os.path.exists(f'{experiment_path}/hcc-data_example/model_evaluation/shap_values/testResults/'):
                os.mkdir(f'{experiment_path}/hcc-data_example/model_evaluation/shap_values/testResults/')
            file_path = (f'{experiment_path}/hcc-data_example/model_evaluation/shap_values/testResults/{abbrev[algorithm]}')

            FI_all = [] # list to store feature importanes to create shap values master list

            for cvCount in range(0,cv_partitions): #loop through cv's
                print(f"{abbrev[algorithm]}{cvCount} In CV{cvCount}...")

                # unpickle and load model
                result_file = f'/{full_path}/models/pickledModels/{abbrev[algorithm]}_{str(cvCount)}.pickle'
                file = open(result_file, 'rb')
                model = pickle.load(file)
                file.close()
                print(f'\nChecking if correct model is loaded...\n{model}')

                # Load CV datasets, paths to datasets updates with each iteration
                train_path = f'/{experiment_path}/{each}/CVDatasets/{each}_CV_{str(cvCount)}_Train.csv"
```

```

test_path =f"/{experiment_path}/{each}/CVDatasets/{each}_CV_{str(cvCount)}_Test.csv"
trainX, trainY, testX, testY, trainFeat, testFeat = dataPrep(train_path, instance_label, class_label, tes
#
    print(trainX)

# shap computation and plots
# Sanity check: print explainer to check if explainer exists
explainer = get_explainer(model, abbrev[algorithm], trainX) #explainer must always use training set
print(f"\nChecking explainer for {abbrev[algorithm]}{cvCount}...\n{explainer}\n")

print(f"Checking shap values for {abbrev[algorithm]}{cvCount}...\n")
shap_values = compute_shapValues(model, abbrev[algorithm], explainer, testX)

print(f"\nGenerating SHAP plots for {abbrev[algorithm]}{cvCount}...\n")
shap_summary(abbrev[algorithm], testFeat, shap_values, explainer, testX, cvCount, save_path, 'Test')

#save SHAP FI results for each model per CV
print('Saving feature importance ranking for {}{}\n'.format(abbrev[algorithm], cvCount))
shap_fi_df, shap_means = shap_feature_ranking(abbrev[algorithm], shap_values, testX, testFeat) # can e
print(f'shap means for {abbrev[algorithm]} in CV{cvCount}...\n{shap_means}\n')
shapFI_path = (f'{filepath}{abbrev[algorithm]}_{str(cvCount)}_shapFIValues_Test.csv")
shap_fi_df.to_csv(shapFI_path, header=True, index=True)

# OPTIONAL: set to 'TRUE' to save force plot figures
if run_force == True:
    run_force_plots(abbrev[algorithm], explainer, shap_values, testX, testFeat, cvCount, save_path, 'T
else:
    continue

# create masterList of mean SHAP Values taken from shap_feature_ranking() for each model
save = save_shap(abbrev[algorithm], shap_values, shap_means, original_headers, cvCount, 'Test', each)
FI_all.append(save)
print("-----")
temp = FI_all
del FI_all # free up space

# create master list for each model after looping through all CVs
df = pd.DataFrame(np.asarray(temp), columns=original_headers)
display(df)

path = (f'{filepath}{abbrev[algorithm]}_shapMasterList.csv')
df.to_csv(path, index=False)

```

```
-----
hcc-data_example
-----
-----
NB
-----

NB0 In CV0...

Checking if correct model is loaded...
GaussianNB()

Checking explainer for NB0...
shap.explainers.Permutation()

Checking shap values for NB0...

.values =
array([[ 0.00583333, -0.03166667, -0.0225      , ..., -0.00666667,
         0.00833333, -0.02166667],
       [ 0.00166667,  0.03333333,  0.04666667, ...,  0.00166667,
         0.0125      , -0.01583333],
       [ 0.005      ,  0.          ,  0.02666667, ...,  0.00666667,
         0.0575      , -0.00666667],
       ...,
       [ 0.          , -0.00583333, -0.00833333, ...,  0.00083333,
         0.00333333, -0.0275      ],
       [-0.00666667, -0.00416667, -0.0125      , ...,  0.00166667,
        -0.03666667, -0.02916667],
       [ 0.01083333,  0.04666667, -0.00833333, ...,  0.00916667,
        -0.09333333,  0.44          ]])

.base_values =
array([0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33,
       0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33,
       0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33,
       0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33])

.data =
array([[ 0.0368995,  1.0551286, -0.5852099, ..., -1.2247449,  0.6264224,
        -0.4383745],
       [-0.3323658, -0.2719725,  3.8874108, ...,  0.8164966,  0.6264224,
        -0.3900995],
       [-0.0973788,  0.3178502,  0.1974987, ...,  0.8164966,  0.6264224,
        -0.4866495],
       ...,
       [-0.6680615,  0.6127615, -0.708207 , ...,  0.8164966,  0.6264224,
        -0.3740079],
       [-0.9198333,  0.3178502, -0.484576 , ...,  0.8164966, -1.5963668,
        -0.2613663],
       [ 0.1376082, -0.3457004, -0.3615789, ...,  0.8164966, -1.5963668,
        1.1546993]])

Generating SHAP plots for NB0...

Saving Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving SHAP Beeswarm Plot for Top 5 SHAP Values in Class 0 & 1 in Test Set...
Saving feature importance ranking for NB0...

shap means for NB in CV0...
[0.00534848 0.02095455 0.0150303  0.01983333 0.00298485 0.02857576
 0.03245455 0.02789394 0.00921212 0.07512121 0.01466667 0.03872727
 0.05901515 0.01948485 0.02295455 0.00872727 0.01277273 0.01519697
 0.03604545 0.01331818 0.          0.01207576 0.02374242 0.02318182
 0.01474242 0.03533333 0.02319697 0.00521212 0.00421212 0.02495455
 0.01715152 0.02792424 0.0369697  0.00225758 0.03387879 0.00212121
 0.00245455 0.025      0.04078788]

Saving Force Plot for NB SHAP Values in Test Set...

Checking for matches...Symptoms  is Symptoms
Symptoms  value is 0.025

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.013318181818181814

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.012075757575757576

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.00921212121212121

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.019484848484848484

Checking for matches...Smoking is Smoking
Smoking value is 0.002121212121212122

Checking for matches...Diabetes is Diabetes
```


Diabetes value is 0.014666666666666665

Checking for matches...Obesity is Obesity
Obesity value is 0.024954545454545455

Checking for matches...Hemochromatosis is Hemochromatosis
Hemochromatosis value is 0.03604545454545454

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.0029848484848484843

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.027893939393939398

Checking for matches...Esophageal Varices is Esophageal Varices
Esophageal Varices value is 0.022954545454545453

Checking for matches...Splenomegaly is Splenomegaly
Splenomegaly value is 0.0024545454545454536

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.002257575757575758

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.033878787878787876

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.035333333333333333

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.027924242424242428

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.03696969696969696

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.059015151515151534

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.028575757575757574

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.02374242424242423

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.019833333333333334

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.015196969696969695

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.005212121212121215

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.01474242424242424

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.020954545454545448

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.04078787878787878

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.00534848484848485

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.03245454545454545

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.012772727272727267

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.015030303030303028

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.07512121212121209

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.004212121212121211

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.0231969696969697

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.03872727272727273

Checking for matches...Iron is Iron
Iron value is 0.023181818181818192

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.017151515151515157

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)

Ferritin (ng/mL) value is 0.0087272727272726

NB1 In CV1...

Checking if correct model is loaded...
GaussianNB()

Checking explainer for NB1...
shap.explainers.Permutation()

Checking shap values for NB1...

```
.values =  
array([[ -0.015      ,  0.02333333, -0.0725      , ...,  0.12416667,  
        -0.0025      , -0.01          ],  
       [ 0.0325      , -0.005         , -0.01          , ..., -0.20833333,  
        -0.03666667, -0.01583333],  
       [ 0.01166667,  0.015          ,  0.02083333, ..., -0.0575      ,  
        -0.0075      , -0.00666667],  
       ...,  
       [ 0.0075      , -0.015         ,  0.09583333, ..., -0.09166667,  
        -0.01          , -0.01166667],  
       [-0.00416667, -0.005          ,  0.01          , ...,  0.          ,  
        -0.0075      , -0.00333333],  
       [ 0.04333333,  0.01416667, -0.0375      , ...,  0.08583333,  
        -0.0125      , -0.01          ]])  
  
.base_values =  
array([0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49,  
       0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49,  
       0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49,  
       0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49,  
       0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49, 0.49])
```

```
.data =  
array([[ -0.7178553, -0.3317865,  1.9018142, ...,  0.6831301, -0.4188214,  
        -0.0949777],  
       [ 0.9544964, -0.9389338,  0.0448948, ..., -1.4638501, -0.4505005,  
        -0.2135745],  
       [ 0.5909417, -0.2169208, -0.558604 , ..., -1.4638501, -0.1812288,  
        -0.176513 ],  
       ...,  
       [-0.4270115,  0.83328  ,  0.1996381, ..., -1.4638501, -0.3713029,  
        -0.2358114],  
       [-0.4270115,  1.1450584, -0.1098485, ...,  0.6831301, -0.2445869,  
        -0.2358114],  
       [ 1.0272073, -0.6763836,  1.1280978, ...,  0.6831301, -0.3871424,  
        -0.2580483]])
```

Generating SHAP plots for NB1...

Saving Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving SHAP Beeswarm Plot for Top 5 SHAP Values in Class 0 & 1 in Test Set...
Saving feature importance ranking for NB1...

```
shap means for NB in CV1...  
[0.01960606 0.01687879 0.04354545 0.00671212 0.04760606 0.02828788  
 0.04548485 0.0089697  0.          0.00337879 0.00860606 0.02086364  
 0.09939394 0.00290909 0.01383333 0.00506061 0.0269697  0.03965152  
 0.0055303  0.01465152 0.00263636 0.00798485 0.03839394 0.0265  
 0.00028788 0.02809091 0.01269697 0.01477273 0.06919697 0.00365152  
 0.0099697  0.02204545 0.01448485 0.0605303  0.01965152 0.04143939  
 0.02804545 0.00533333 0.06159091 0.02986364 0.01510606]
```

Saving Force Plot for NB SHAP Values in Test Set...

Checking for matches...Symptoms is Symptoms
Symptoms value is 0.06159090909090909

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.0067121212121211

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.0146515151515151

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.00263636363636355

Checking for matches...Hepatitis B Core Antibody is Hepatitis B Core Antibody
Hepatitis B Core Antibody value is 0.005530303030303031

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.007984848484848489

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0993939393939394

Checking for matches...Smoking is Smoking

Smoking value is 0.005333333333333334

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.00860606060606061

Checking for matches...Obesity is Obesity
Obesity value is 0.009969696969696972

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0691969696969697

Checking for matches...Esophageal Varices is Esophageal Varices
Esophageal Varices value is 0.0029090909090909085

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.041439393939393936

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.028045454545454537

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.028090909090909097

Checking for matches...Age at diagnosis is Age at diagnosis
Age at diagnosis value is 0.01960606060606061

Checking for matches...Grams of Alcohol per day is Grams of Alcohol per day
Grams of Alcohol per day value is 0.026969696969696966

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.014484848484848486

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.06053030303030303

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.045484848484848475

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.038393939393939404

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.028287878787878785

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.03965151515151516

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.014772727272727276

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.0002878787878787878

Checking for matches...Platelets is Platelets
Platelets value is 0.019651515151515146

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.04354545454545455

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.029863636363636367

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.016878787878787882

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.008969696969696968

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.0050606060606060614

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.04760606060606061

Checking for matches...Total Proteins (g/dL) is Total Proteins (g/dL)
Total Proteins (g/dL) value is 0.015106060606060612

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.0033787878787878804

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.0036515151515151526

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.012696969696969696

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.020863636363636362

Checking for matches...Iron is Iron
Iron value is 0.026499999999999996

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)

Oxygen Saturation (%) value is 0.022045454545454535

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.013833333333333335

NB2 In CV2...

Checking if correct model is loaded...
GaussianNB()

Checking explainer for NB2...
shap.explainers.Permutation()

Checking shap values for NB2...

```
.values =  
array([[ 0.          ,  0.          ,  0.          , ...,  0.          ,  
        -0.00083333,  0.          ],  
       [ 0.0025      ,  0.          ,  0.          , ...,  0.00166667,  
        -0.00166667,  0.          ],  
       [-0.00333333,  0.01333333, -0.015        , ...,  0.          ,  
        -0.0125      ,  0.00666667],  
       ...,  
       [ 0.00083333,  0.01916667,  0.00333333, ..., -0.00333333,  
        -0.00583333,  0.0025      ],  
       [-0.01916667, -0.06166667,  0.00666667, ..., -0.00916667,  
        -0.03916667,  0.01          ],  
       [ 0.          ,  0.          ,  0.          , ...,  0.          ,  
        0.          ,  0.          ]])  
  
.base_values =  
array([0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05,  
       0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05,  
       0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05,  
       0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05,  
       0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05])
```

```
.data =  
array([[ -0.8437697, -0.4184704,  0.5703518, ...,  0.6264224, -0.5282705,  
        -1.3228757],  
       [ -3.7262129, -0.4184704,  0.5703518, ..., -1.5963668, -0.5282705,  
        0.7559289],  
       [ -1.0908363,  0.0205566, -1.7533038, ...,  0.6264224, -0.5282705,  
        0.7559289],  
       ...,  
       [ -1.0908363, -1.8818937,  0.5703518, ...,  0.6264224, -0.5282705,  
        0.7559289],  
       [ -1.2555473,  1.0449529,  0.5703518, ...,  0.6264224, -0.5282705,  
        0.7559289],  
       [ -0.1025701, -2.028236 ,  0.5703518, ...,  0.6264224, -0.5282705,  
        0.7559289]])
```

Generating SHAP plots for NB2...

Saving Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving SHAP Beeswarm Plot for Top 5 SHAP Values in Class 0 & 1 in Test Set...
Saving feature importance ranking for NB2...

```
shap means for NB in CV2...  
[0.00121212 0.00242424 0.00071212 0.00228788 0.00066667 0.00113636  
 0.004        0.00334848 0.00066667 0.00030303 0.00092424 0.05527273  
 0.00133333 0.00022727 0.01795455 0.0010303  0.00245455 0.00354545  
 0.00912121 0.00098485 0.05312121 0.01395455 0.00069697 0.00334848  
 0.00606061 0.00154545 0.00156061 0.00201515 0.00575758 0.00021212  
 0.0179697   0.00362121 0.00154545 0.00040909 0.00421212 0.00060606]
```

Saving Force Plot for NB SHAP Values in Test Set...

Checking for matches...Symptoms is Symptoms
Symptoms value is 0.0006060606060606061

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.000712121212121212

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.009121212121212125

Checking for matches...Hepatitis B Core Antibody is Hepatitis B Core Antibody
Hepatitis B Core Antibody value is 0.0035454545454545456

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0009848484848484845

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0002272727272727274

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.0009242424242424244

Checking for matches...Obesity is Obesity
Obesity value is 0.005757575757575757

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.001136363636363636

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.00066666666666666666

Checking for matches...Human Immunodeficiency Virus is Human Immunodeficiency Virus
Human Immunodeficiency Virus value is 0.053121212121212125

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.0004090909090909088

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.004212121212121212

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.0060606060606060615

Checking for matches...Age at diagnosis is Age at diagnosis
Age at diagnosis value is 0.00121212121212121

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.017969696969696972

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.003621212121212121

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.0013333333333333333

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.004000000000000001

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.013954545454545459

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.0006666666666666668

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.002454545454545454

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0015606060606060603

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.0033484848484848476

Checking for matches...Platelets is Platelets
Platelets value is 0.0015454545454545456

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.0024242424242424242

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.003348484848484848

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.0010303030303030305

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.0022878787878787875

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.00030303030303030276

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.002015151515151515

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.0015454545454545454

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.05527272727272727

Checking for matches...Iron is Iron
Iron value is 0.0006969696969696964

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.00021212121212121218

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.017954545454545456

	Gender	Symptoms	Alcohol	Hepatitis B Surface Antigen	Hepatitis B e Antigen	Hepatitis B Core Antibody	Hepatitis C Virus Antibody	Cirrhosis	Endemic Countries	Smoking	...	Gamma glutamyl transferase (U/L)	Alkaline phosphatase (U/L)	Pr
0	0.0	0.025000	0.000000	0.013318	0.000000	0.000000	0.012076	0.009212	0.019485	0.002121	...	0.012773	0.015030	0.0
1	0.0	0.061591	0.006712	0.014652	0.002636	0.005530	0.007985	0.000000	0.099394	0.005333	...	0.005061	0.047606	0.0
2	0.0	0.000606	0.000712	0.000000	0.009121	0.003545	0.000985	0.000000	0.000227	0.000000	...	0.001030	0.002288	0.0

3 rows x 49 columns

```
-----
LR
-----
LR0 In CV0...

Checking if correct model is loaded...
LogisticRegression(C=0.006606805070193189, dual=True,
                    max_iter=193.8544995971634, random_state=42,
                    solver='liblinear')

Checking explainer for LR0...
<shap.explainers._linear.Linear object at 0x7fecb3c3a490>

Checking shap values for LR0...

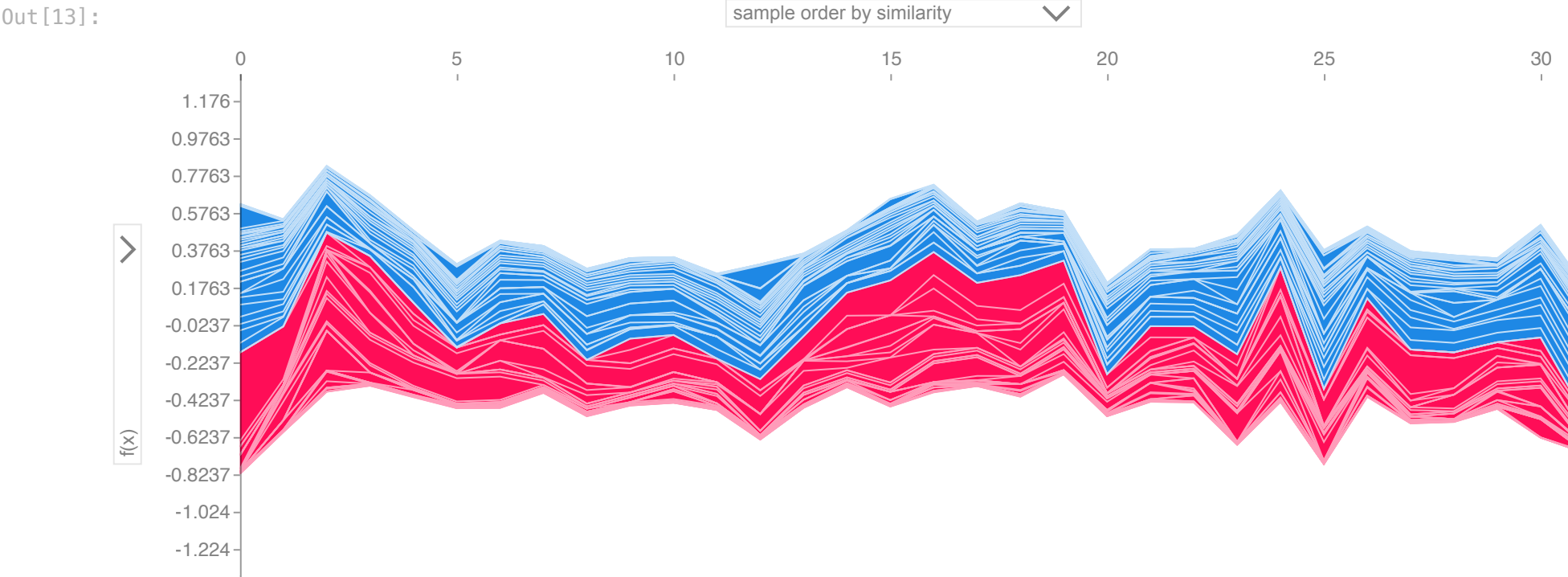
[[-1.54247265e-04 -5.20008470e-02 -4.40485958e-02 ... -1.47969848e-03
  2.31157863e-02 -2.02239904e-02]
 [ 6.85778818e-04  1.94894821e-02  2.13041189e-01 ...  1.21066239e-03
  2.31157863e-02 -1.80842415e-02]
 [ 1.51216786e-04 -1.22839969e-02  9.42115197e-04 ...  1.21066239e-03
  2.31157863e-02 -2.23637392e-02]
 ...
 [ 1.44943883e-03 -2.81707337e-02 -5.11185666e-02 ...  1.21066239e-03
  2.31157863e-02 -1.73709948e-02]
 [ 2.02218390e-03 -1.22839969e-02 -3.82640794e-02 ...  1.21066239e-03
 -6.24982370e-02 -1.23782505e-02]
 [-3.83345247e-04  2.34611703e-02 -3.11941086e-02 ...  1.21066239e-03
 -6.24982370e-02  5.03876684e-02]]

Generating SHAP plots for LR0...

Expected value for LR: -0.023696555525940875
Saving Summary Plot for SHAP Values in Test Set...
Saving SHAP Bar Plot for SHAP Values Test Set...
Saving SHAP Decision Plot for SHAP Values in Test Set...
Saving SHAP Decision Plot for Single-Prediction in Test Set...
Saving Waterfall Plot for SHAP Values for a Single-Prediction in Test Set...
Saving feature importance ranking for LR0...

shap means for LR in CV0...
[0.00140633 0.04251623 0.03865761 0.00609194 0.01046446 0.04334771
 0.04264836 0.02153193 0.00967007 0.03609579 0.04411194 0.01784339
 0.01750026 0.0158771  0.04361159 0.03592322 0.02784188 0.03343238
 0.00743959 0.01902675 0.          0.02683842 0.01723709 0.00580197
 0.02052612 0.05890593 0.0426101  0.00347406 0.00263454 0.018688
 0.00247347 0.01348655 0.06691778 0.02108598 0.05053329 0.00215841
 0.00131828 0.03886877 0.01698003]
```

Displaying Force Plot for LR SHAP Values in Whole Test Set...



8/31/22, 2:58 PM

SHAP_final

Checking for matches...Symptoms is Symptoms
Symptoms value is 0.03886876657626088

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.01902674623276478

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.026838415026258885

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.009670065204705305

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.015877099981977692

Checking for matches...Smoking is Smoking
Smoking value is 0.0021584128121270303

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.04411193743084139

Checking for matches...Obesity is Obesity
Obesity value is 0.018688002101128375

Checking for matches...Hemochromatosis is Hemochromatosis
Hemochromatosis value is 0.007439589240777946

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.010464459888218829

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.021531933553371088

Checking for matches...Esophageal Varices is Esophageal Varices
Esophageal Varices value is 0.04361158968680013

Checking for matches...Splenomegaly is Splenomegaly
Splenomegaly value is 0.0013182768276836729

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.021085978608251846

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.05053329221562375

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.058905931019859296

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.013486554186244934

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.06691778037192776

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.017500259021430268

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.04334770889735201

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.017237085335443568

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.0060919367120545135

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.03343237999610053

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0034740606021462683

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.020526117890812018

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.04251623086938566

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.016980033669758456

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.0014063259378561783

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.042648362852658314

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.02784187597044018

```
Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.03865761449100018

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.03609579477129794

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.0026345418487333004

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.04261009919109683

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.01784338626207558

Checking for matches...Iron is Iron
Iron value is 0.005801966424714463

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.0024734668502215075

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.03592322253205616

-----
LR1 In CV1...

Checking if correct model is loaded...
LogisticRegression(C=0.06359900885943309, max_iter=48.076782938152924,
                    random_state=42, solver='sag')

Checking explainer for LR1...
<shap.explainers._linear.Linear object at 0x7fecb30277f0>

Checking shap values for LR1...

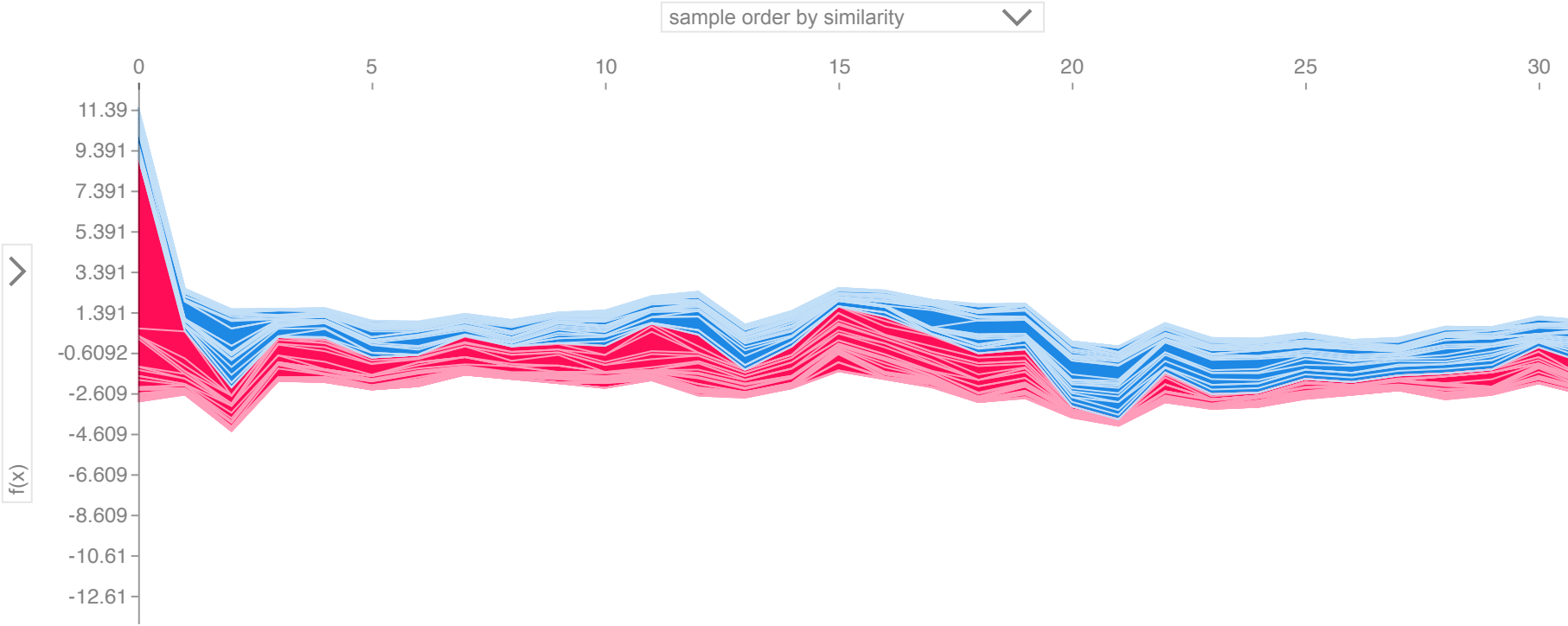
[[-0.19905492  0.03850256 -0.30345684 ...  0.14567922 -0.09989041
  -0.01025168]
 [ 0.18762272  0.10977384 -0.00291614 ... -0.39387344 -0.10644632
  -0.02397595]
 [ 0.10356236  0.0250188   0.09475959 ... -0.39387344 -0.05072119
  -0.01968712]
 ...
 [-0.13180663 -0.09826126 -0.0279612  ... -0.39387344 -0.09005657
  -0.02654925]
 [-0.13180663 -0.13486004  0.02212892 ...  0.14567922 -0.063833
  -0.02654925]
 [ 0.20443478  0.07895383 -0.17823155 ...  0.14567922 -0.09333451
  -0.02912255]]

Generating SHAP plots for LR1...

Expected value for LR: -0.6091565598361125
Saving Summary Plot for SHAP Values in Test Set...
Saving SHAP Bar Plot for SHAP Values Test Set...
Saving SHAP Decision Plot for SHAP Values in Test Set...
Saving SHAP Decision Plot for Single-Prediction in Test Set...
Saving Waterfall Plot for SHAP Values for a Single-Prediction in Test Set...
Saving feature importance ranking for LR1...

shap means for LR in CV1...
[0.16654343 0.07266444 0.14603831 0.00492052 0.22157166 0.21656434
 0.24265308 0.09538771 0.01955599 0.04794785 0.15645302 0.07717627
 0.12170538 0.03470505 0.14257453 0.05444321 0.01157082 0.22367285
 0.02900849 0.01100653 0.01148055 0.13612008 0.20269228 0.15509792
 0.07512847 0.00666055 0.10291185 0.11299246 0.09496667 0.01784108
 0.07432644 0.00910401 0.05617509 0.19525583 0.02225843 0.07756351
 0.09654913 0.10977749 0.22690642 0.08378449 0.02584071]
```

Out [13]:



Checking for matches...Symptoms is Symptoms
Symptoms value is 0.2269064171534283

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.004920516143671562

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.011006526845122698

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0114805528046631

Checking for matches...Hepatitis B Core Antibody is Hepatitis B Core Antibody
Hepatitis B Core Antibody value is 0.029008490739009542

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.13612007680249777

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.019555993342348636

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.12170538451100983

Checking for matches...Smoking is Smoking
Smoking value is 0.10977748998232592

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.1564530216006723

Checking for matches...Obesity is Obesity
Obesity value is 0.07432644045275326

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.09496667033560677

Checking for matches...Esophageal Varices is Esophageal Varices
Esophageal Varices value is 0.034705053111518044

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.0775635063691718

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.09654913169476316

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.006660545435984733

Checking for matches...Age at diagnosis is Age at diagnosis
Age at diagnosis value is 0.1665434287207862

Checking for matches...Grams of Alcohol per day is Grams of Alcohol per day
Grams of Alcohol per day value is 0.011570822629152567

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.05617509323066895

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.19525582905924707

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.24265307616145004

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.20269228347766075

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.2165643372235713

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.22367284993229583

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.1129924604280148

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.07512846977281235

Checking for matches...Platelets is Platelets
Platelets value is 0.022258432750088153

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.14603830828422823

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.08378449223791448

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.07266443613666665

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.09538770841862036

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.05444321320404681

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.22157165571529303

Checking for matches...Total Proteins (g/dL) is Total Proteins (g/dL)
Total Proteins (g/dL) value is 0.025840706896034186

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.04794784856260298

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.017841083480363045

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.1029118464267115

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.07717626605291422

Checking for matches...Iron is Iron
Iron value is 0.15509792157625088

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.009104009767384343

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.14257453003812987

LR2 In CV2...

Checking if correct model is loaded...
LogisticRegression(C=0.0006580360277501316, class_weight='balanced', dual=True, max_iter=112.07606211860569, random_state=42, solver='liblinear')

Checking explainer for LR2...
<shap.explainers._linear.Linear object at 0x7fecb2ff92b0>

Checking shap values for LR2...

```
[[-0.00209083  0.00438031  0.00118267 ... -0.00203892 -0.00439287
  -0.00869176]
 [-0.00982644  0.00438031  0.00118267 ...  0.00524293 -0.00439287
   0.00355015]
 [-0.00275388  0.00044561 -0.00473069 ... -0.00203892 -0.00439287
   0.00355015]
 ...
 [-0.00275388  0.01749597  0.00118267 ... -0.00203892 -0.00439287
   0.00355015]
 [-0.00319591 -0.00873536  0.00118267 ... -0.00203892 -0.00439287
   0.00355015]
 [-0.00010167  0.01880754  0.00118267 ... -0.00203892 -0.00439287
   0.00355015]]
```

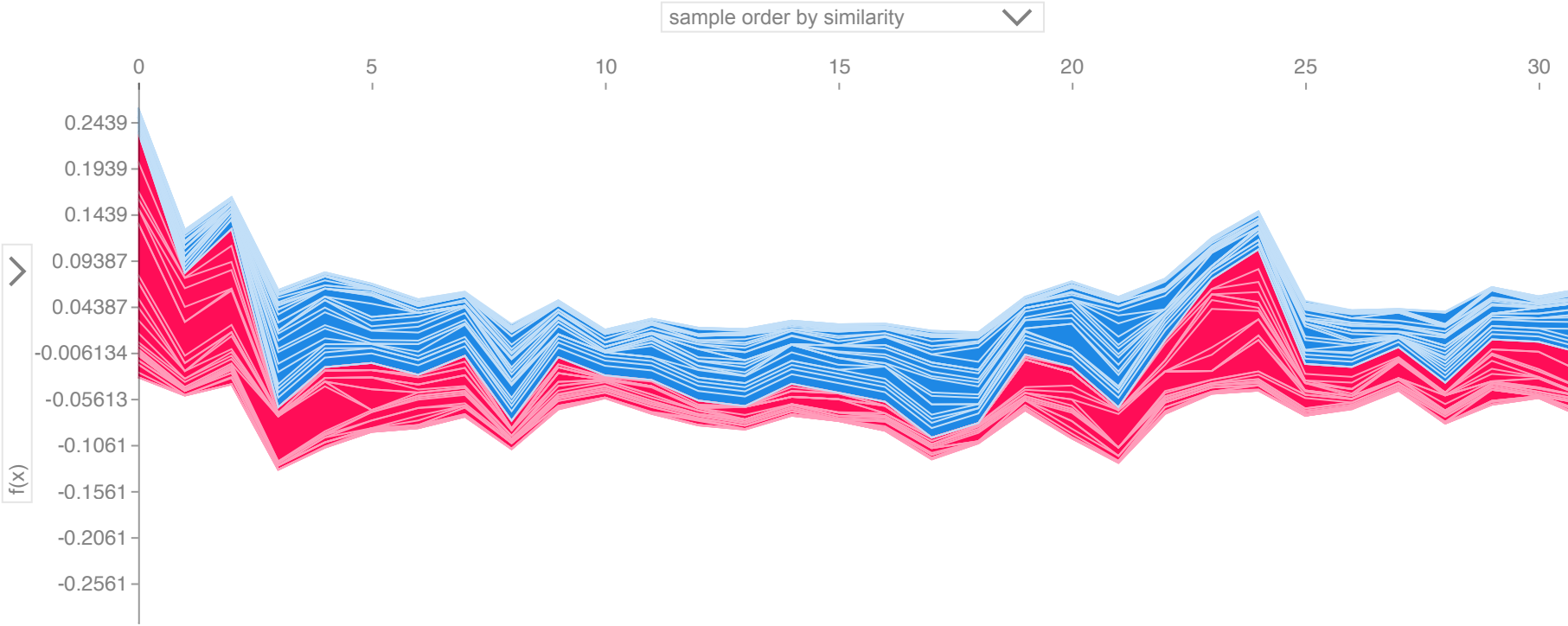
Generating SHAP plots for LR2...

Expected value for LR: -0.006133751932115765
Saving Summary Plot for SHAP Values in Test Set...
Saving SHAP Bar Plot for SHAP Values Test Set...
Saving SHAP Decision Plot for SHAP Values in Test Set...
Saving SHAP Decision Plot for Single-Prediction in Test Set...
Saving Waterfall Plot for SHAP Values for a Single-Prediction in Test Set...
Saving feature importance ranking for LR2...

shap means for LR in CV2...
[2.52578912e-03 7.22348091e-03 2.21482139e-03 7.76975487e-03
5.94797251e-05 3.91915048e-03 6.47924621e-03 4.43775359e-03
1.64743809e-03 9.91359698e-04 5.27881775e-03 4.72258480e-03
7.57378172e-04 3.30706823e-04 7.53856654e-03 3.77134111e-03
6.63804149e-03 3.80908259e-03 9.57013968e-04 1.91522175e-03
2.73061318e-03 3.56031389e-03 5.37089763e-03 6.26003349e-03
8.86535277e-03 5.64753789e-03 2.09495109e-03 5.68981824e-03
3.75910628e-03 4.18704626e-04 4.80513725e-04 1.07792792e-02
4.05709949e-03 2.79623086e-03 7.26821080e-03 4.76544111e-03]

Displaying Force Plot for LR SHAP Values in Whole Test Set...

Out [13]:



Checking for matches...Symptoms is Symptoms
Symptoms value is 0.0047654411054461375

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.0022148213857691858

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0009570139678777011

Checking for matches...Hepatitis B Core Antibody is Hepatitis B Core Antibody
Hepatitis B Core Antibody value is 0.0038090825853725127

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0019152217497143459

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0003307068230837304

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.005278817754709944

Checking for matches...Obesity is Obesity
Obesity value is 0.003759106278018618

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.003919150479863269

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.0016474380902237297

Checking for matches...Human Immunodeficiency Virus is Human Immunodeficiency Virus
Human Immunodeficiency Virus value is 0.0027306131765694056

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.0027962308607005497

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.0072682108030158316

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.008865352768042291

Checking for matches...Age at diagnosis is Age at diagnosis
Age at diagnosis value is 0.0025257891209047784

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.0004805137253571042

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.010779279186138606

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.0007573781723251926

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.0064792462052921185

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.0035603138897862584

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 5.947972507110492e-05

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.006638041492021804

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0020949510893413683

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.006260033491679873

Checking for matches...Platelets is Platelets
Platelets value is 0.004057099489974041

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.007223480911784905

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.004437753586863367

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.003771341108453564

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.007769754870870507

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.000991359698411762

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.005689818237105034

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.005647537886930736

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.0047225847988951256

Checking for matches...Iron is Iron
Iron value is 0.005370897631439668

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.0004187046261779091

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.00753856653534736

	Gender	Symptoms	Alcohol	Hepatitis B Surface Antigen	Hepatitis B e Antigen	Hepatitis B Core Antibody	Hepatitis C Virus Antibody	Cirrhosis	Endemic Countries	Smoking	...	Gamma glutamyl transferase (U/L)	Alkaline phosphatase (U/L)	Pr
0	0.0	0.038869	0.000000	0.019027	0.000000	0.000000	0.026838	0.009670	0.015877	0.002158	...	0.027842	0.038658	0.0
1	0.0	0.226906	0.004921	0.011007	0.011481	0.029008	0.136120	0.019556	0.121705	0.109777	...	0.054443	0.221572	0.0
2	0.0	0.004765	0.002215	0.000000	0.000957	0.003809	0.001915	0.000000	0.000331	0.000000	...	0.003771	0.007770	0.0

3 rows x 49 columns

DT

DT0 In CV0...

Checking if correct model is loaded...
DecisionTreeClassifier(max_depth=17, min_samples_leaf=35, min_samples_split=45,
 random_state=42)

Checking explainer for DT0...
<shap.explainers._tree.Tree object at 0x7fecc756c190>

Checking shap values for DT0...

```
[array([[0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        ...,
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.])), array([[0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        ...,
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.]])]
```

Generating SHAP plots for DT0...

Expected value for DT: [0.57272727 0.42727273]
Saving Bar Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Decision Plot for SHAP Values from Class 0 in Test Set...
Saving Decision Plot for SHAP Values from Class 1 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 0 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 1 in Test Set...
Saving feature importance ranking for DT0...

shap means for DT in CV0...

0.	0.	0.	0.45031405	0.	0.
0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.
0.	0.	0.]		

Displaying Force Plot for DT SHAP Values from Class 0 in Test Set...

Out [13]:



Checking for matches...Symptoms is Symptoms
Symptoms value is 0.0

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.0

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Smoking is Smoking
Smoking value is 0.0

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.0

Checking for matches...Obesity is Obesity
Obesity value is 0.0

Checking for matches...Hemochromatosis is Hemochromatosis
Hemochromatosis value is 0.0

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.0

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.0

Checking for matches...Esophageal Varices is Esophageal Varices
Esophageal Varices value is 0.0

Checking for matches...Splenomegaly is Splenomegaly
Splenomegaly value is 0.0

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.0

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.0

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.0

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.0

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.0

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.0

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.0

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.0

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.4503140495867769

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.0

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.0

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.0

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.0

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.0

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.0

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.0

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.0

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.0

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.0

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.0

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.0

Checking for matches...Iron is Iron
Iron value is 0.0

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.0

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.0

```
-----
DT1 In CV1...

Checking if correct model is loaded...
DecisionTreeClassifier(criterion='entropy', max_depth=21, min_samples_leaf=3,
                      min_samples_split=23, random_state=42,
                      splitter='random')

Checking explainer for DT1...
<shap.explainers._tree.Tree object at 0x7fecb2ba4190>

Checking shap values for DT1...

[array([[ 0.          ,  0.          ,  0.          , ..., -0.02795699,
         0.          ,  0.          ],
       [ 0.          ,  0.          ,  0.          , ...,  0.07004662,
         0.          ,  0.          ],
       [ 0.          ,  0.          ,  0.          , ...,  0.07004662,
         0.          ,  0.          ],
       ...,
       [ 0.          ,  0.          ,  0.          , ...,  0.08666667,
         0.          ,  0.          ],
       [ 0.          ,  0.          ,  0.          , ..., -0.11345397,
         0.          ,  0.          ],
       [ 0.          ,  0.          ,  0.          , ..., -0.02795699,
         0.          ,  0.          ]]), array([[ 0.          ,  0.          ,  0.          , ...,  0.02795699,
         0.          ,  0.          ],
       [ 0.          ,  0.          ,  0.          , ..., -0.07004662,
         0.          ,  0.          ],
       [ 0.          ,  0.          ,  0.          , ..., -0.07004662,
         0.          ,  0.          ],
       ...,
       [ 0.          ,  0.          ,  0.          , ..., -0.08666667,
         0.          ,  0.          ],
       [ 0.          ,  0.          ,  0.          , ...,  0.11345397,
         0.          ,  0.          ],
       [ 0.          ,  0.          ,  0.          , ...,  0.02795699,
         0.          ,  0.          ]]])]

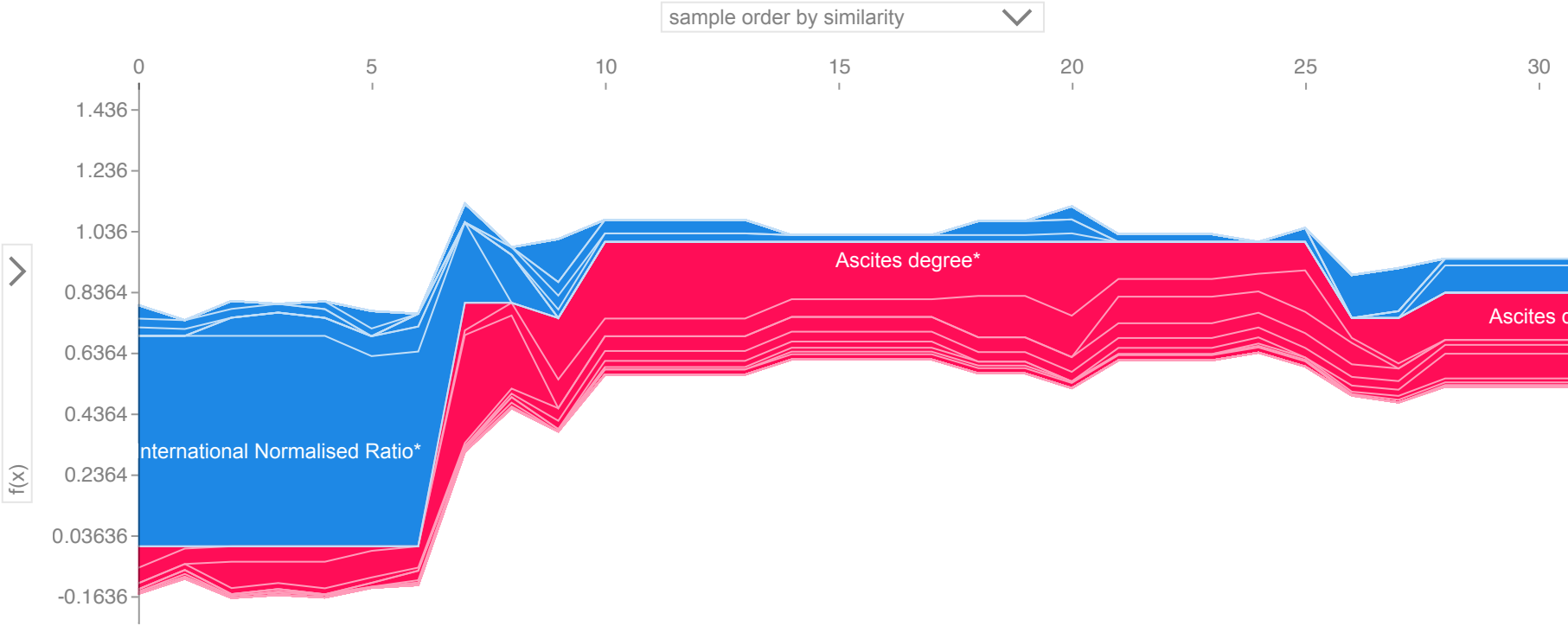
Generating SHAP plots for DT1...

Expected value for DT: [0.63636364 0.36363636]
Saving Bar Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Decision Plot for SHAP Values from Class 0 in Test Set...
Saving Decision Plot for SHAP Values from Class 1 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 0 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 1 in Test Set...
Saving feature importance ranking for DT1...

shap means for DT in CV1...
[0.          0.          0.          0.13587539 0.          0.04034871
 0.30717224 0.          0.          0.          0.          0.
 0.          0.          0.          0.04657883 0.          0.08993169
 0.          0.          0.          0.          0.25005612 0.
 0.          0.          0.          0.05505981 0.          0.
 0.          0.          0.          0.15649071 0.          0.
 0.          0.          0.12766554 0.          0.          ]

Displaying Force Plot for DT SHAP Values from Class 0 in Test Set...
```

Out [13]:



Checking for matches...Symptoms is Symptoms
Symptoms value is 0.12766554142993597

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.13587538667643773

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.0

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis B Core Antibody is Hepatitis B Core Antibody
Hepatitis B Core Antibody value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Smoking is Smoking
Smoking value is 0.0

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.0

Checking for matches...Obesity is Obesity
Obesity value is 0.0

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0

Checking for matches...Esophageal Varices is Esophageal Varices
Esophageal Varices value is 0.0

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.0

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.0

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.0

Checking for matches...Age at diagnosis is Age at diagnosis
Age at diagnosis value is 0.0

Checking for matches...Grams of Alcohol per day is Grams of Alcohol per day
Grams of Alcohol per day value is 0.0

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.0

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.15649070978746688

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.30717224088870315

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.2500561240512013

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.040348707689985115

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.08993168740204315

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.055059813609845784

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.0

Checking for matches...Platelets is Platelets
Platelets value is 0.0

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.0

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.0

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.0

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.0

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.04657883451217262

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.0

Checking for matches...Total Proteins (g/dL) is Total Proteins (g/dL)
Total Proteins (g/dL) value is 0.0

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.0

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.0

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.0

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.0

Checking for matches...Iron is Iron
Iron value is 0.0

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.0

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.0

DT2 In CV2...

Checking if correct model is loaded...
DecisionTreeClassifier(class_weight='balanced', max_depth=29,
min_samples_leaf=30, min_samples_split=45,
random_state=42)

Checking explainer for DT2...
<shap.explainers._tree.Tree object at 0x7feca04407f0>

Checking shap values for DT2...

```
[array([[ 0.          , -0.14385676,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          , -0.0462963 ,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          , -0.0462963 ,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       ...,
       [ 0.          , -0.0462963 ,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          ,  0.05769231,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          , -0.14385676,  0.          , ...,  0.          ,
         0.          ,  0.          ]]), array([[ 0.          ,  0.14385676,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          ,  0.0462963 ,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          ,  0.0462963 ,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       ...,
       [ 0.          ,  0.0462963 ,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          , -0.05769231,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          ,  0.14385676,  0.          , ...,  0.          ,
         0.          ,  0.          ]]])
```

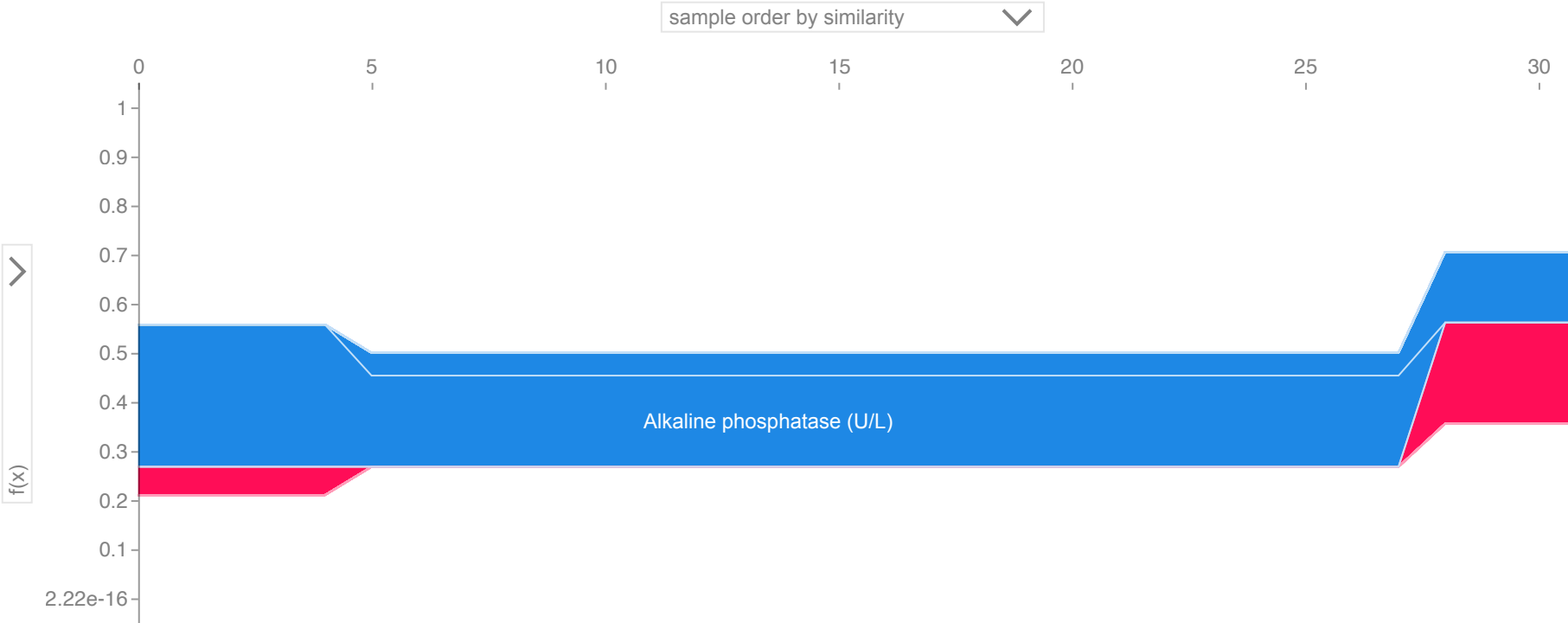
Generating SHAP plots for DT2...

Expected value for DT: [0.5 0.5]
Saving Bar Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Decision Plot for SHAP Values from Class 0 in Test Set...
Saving Decision Plot for SHAP Values from Class 1 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 0 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 1 in Test Set...
Saving feature importance ranking for DT2...

shap means for DT in CV2...
[0. 0.20461559 0. 0.45585129 0. 0.
 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0.]

Displaying Force Plot for DT SHAP Values from Class 0 in Test Set...

Out [13]:



Checking for matches...Symptoms is Symptoms
Symptoms value is 0.0

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.0

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis B Core Antibody is Hepatitis B Core Antibody
Hepatitis B Core Antibody value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.0

Checking for matches...Obesity is Obesity
Obesity value is 0.0

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.0

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.0

Checking for matches...Human Immunodeficiency Virus is Human Immunodeficiency Virus
Human Immunodeficiency Virus value is 0.0

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.0

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.0

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.0

Checking for matches...Age at diagnosis is Age at diagnosis
Age at diagnosis value is 0.0

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.0

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.0

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.0

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.0

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.0

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.0

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.0

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.0

Checking for matches...Platelets is Platelets
Platelets value is 0.0

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.20461558566581414

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.0

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.0

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.4558512856187384

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.0

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.0

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.0

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.0

Checking for matches...Iron is Iron
Iron value is 0.0

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.0

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.0

	Gender	Symptoms	Alcohol	Hepatitis B Surface Antigen	Hepatitis B e Antigen	Hepatitis B Core Antibody	Hepatitis C Virus Antibody	Cirrhosis	Endemic Countries	Smoking	...	Gamma glutamyl transferase (U/L)	Alkaline phosphatase (U/L)	Prothrombin Time (s)
0	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.000000	
1	0.0	0.127666	0.135875	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.046579	0.000000	
2	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.455851	

3 rows x 49 columns

RF

RF0 In CV0...

Checking if correct model is loaded...
RandomForestClassifier(criterion='entropy', max_depth=1, max_features=None,
min_samples_leaf=17, min_samples_split=41,
n_estimators=960, random_state=42)

Checking explainer for RF0...
<shap.explainers._tree.Tree object at 0x7fecb3397910>

Checking shap values for RF0...

```
[array([[ 0.          ,  0.04181896,  0.00365806, ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          , -0.01898534, -0.01999966, ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          , -0.01898534, -0.01840236, ...,  0.          ,
         0.          ,  0.          ],
       ...,
       [ 0.          ,  0.04012049,  0.02614831, ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          , -0.01898534, -0.00631506, ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          , -0.01898534, -0.01061794, ...,  0.          ,
         0.          ,  0.          ]]), array([[ 0.          , -0.04181896, -0.00365806, ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          ,  0.01898534,  0.01999966, ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          ,  0.01898534,  0.01840236, ...,  0.          ,
         0.          ,  0.          ],
       ...,
       [ 0.          , -0.04012049, -0.02614831, ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          ,  0.01898534,  0.00631506, ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          ,  0.01898534,  0.01061794, ...,  0.          ,
         0.          ,  0.          ]]])
```

Generating SHAP plots for RF0...

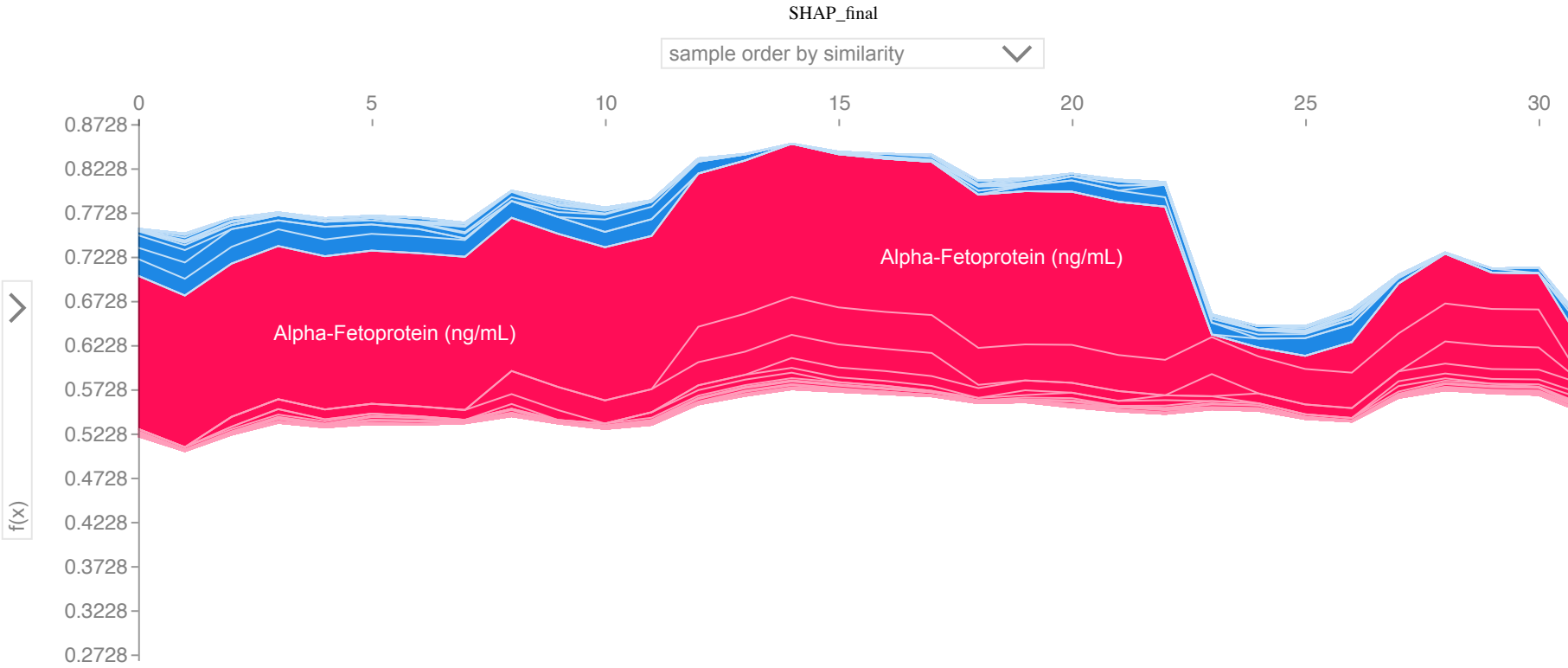
Expected value for RF: [0.5728125 0.4271875]
Saving Bar Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Decision Plot for SHAP Values from Class 0 in Test Set...
Saving Decision Plot for SHAP Values from Class 1 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 0 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 1 in Test Set...
Saving feature importance ranking for RF0...

shap means for RF in CV0...

```
[0.          0.05778136 0.03348325 0.26381148 0.          0.
 0.00738881 0.          0.          0.00362092 0.          0.00285239
 0.          0.          0.          0.00571113 0.00338564 0.00389609
 0.          0.          0.          0.          0.          0.02296876
 0.00103436 0.00339649 0.00240992 0.          0.          0.
 0.          0.          0.01077766 0.          0.          0.
 0.          0.          0.          ]
```

Displaying Force Plot for RF SHAP Values from Class 0 in Test Set...

Out [13]:



8/31/22, 2:58 PM

SHAP_final

Checking for matches...Symptoms is Symptoms
Symptoms value is 0.0

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.0

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Smoking is Smoking
Smoking value is 0.0

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.0

Checking for matches...Obesity is Obesity
Obesity value is 0.0

Checking for matches...Hemochromatosis is Hemochromatosis
Hemochromatosis value is 0.0

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.0

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.0

Checking for matches...Esophageal Varices is Esophageal Varices
Esophageal Varices value is 0.0

Checking for matches...Splenomegaly is Splenomegaly
Splenomegaly value is 0.0

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.0

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.0

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.003396490003583018

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.0

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.01077766356614318

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.0

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.0

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.0

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.26381148486457795

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.0038960872846985255

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.001034357372425554

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.05778135700342246

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.0

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.0

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.007388808063013539

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.003385640346541343

localhost:8888/nbconvert/html/streamline/SHAP_final.ipynb?download=false

38/98

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.03348325011888675

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.003620916529437587

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.0

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.0024099224231534586

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.002852385183600379

Checking for matches...Iron is Iron
Iron value is 0.022968761629997517

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.0

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.005711131335828448

RF1 In CV1...

Checking if correct model is loaded...
RandomForestClassifier(class_weight='balanced', criterion='entropy',
max_depth=2, max_features='log2', min_samples_leaf=9,
min_samples_split=31, n_estimators=207, random_state=42)

Checking explainer for RF1...
<shap.explainers._tree.Tree object at 0x7fecb339b430>

Checking shap values for RF1...

```
[array([[ 3.95998286e-03,  8.78166228e-05,  5.42231605e-02, ...,
        -4.99207302e-03,  7.28955367e-03,  2.80322166e-03],
       [-3.19678857e-03,  8.79064012e-04,  1.60776918e-03, ...,
         1.80742612e-02,  5.37725417e-03, -1.63881455e-03],
       [-2.91886671e-03, -4.37616095e-04, -2.66173097e-02, ...,
         2.10142415e-02,  3.35299303e-03,  5.93880632e-04],
       ...,
       [ 2.62833254e-03, -1.78372567e-04,  1.24812736e-02, ...,
         1.73079030e-02,  4.94635866e-03, -1.53940799e-03],
       [ 2.33387141e-03, -4.42763000e-04, -2.58700195e-02, ...,
        -6.86380765e-03,  4.07652440e-03, -1.52713434e-03],
       [-5.01472471e-03,  1.13411715e-03,  5.63221558e-02, ...,
        -5.51702530e-03,  5.22884231e-03, -4.18441541e-04]])], array([[ -3.95998286e-03, -8.78166228e-05, -5.42231605e-0
2, ...,
         4.99207302e-03, -7.28955367e-03, -2.80322166e-03],
       [ 3.19678857e-03, -8.79064012e-04, -1.60776918e-03, ...,
        -1.80742612e-02, -5.37725417e-03,  1.63881455e-03],
       [ 2.91886671e-03,  4.37616095e-04,  2.66173097e-02, ...,
        -2.10142415e-02, -3.35299303e-03, -5.93880632e-04],
       ...,
       [-2.62833254e-03,  1.78372567e-04, -1.24812736e-02, ...,
        -1.73079030e-02, -4.94635866e-03,  1.53940799e-03],
       [-2.33387141e-03,  4.42763000e-04,  2.58700195e-02, ...,
         6.86380765e-03, -4.07652440e-03,  1.52713434e-03],
       [ 5.01472471e-03, -1.13411715e-03, -5.63221558e-02, ...,
         5.51702530e-03, -5.22884231e-03,  4.18441541e-04]])])]
```

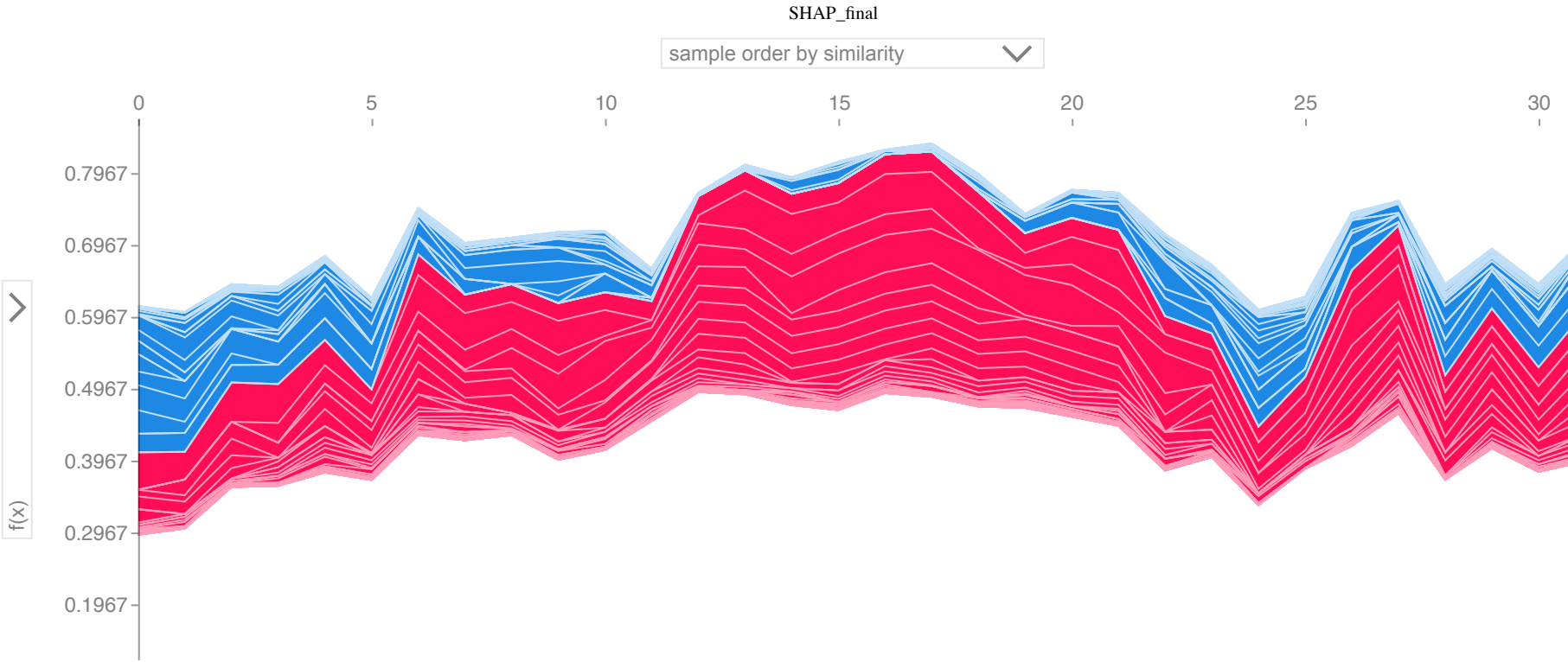
Generating SHAP plots for RF1...

Expected value for RF: [0.49673858 0.50326142]
Saving Bar Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Decision Plot for SHAP Values from Class 0 in Test Set...
Saving Decision Plot for SHAP Values from Class 1 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 0 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 1 in Test Set...
Saving feature importance ranking for RF1...

shap means for RF in CV1...
[0.00616708 0.00111862 0.06046394 0.00077522 0.03486096 0.0340225
0.05590628 0.00749152 0. 0.00562259 0.00123675 0.04072415
0. 0. 0.01997052 0.00509053 0.00389264 0.06545126
0.00088551 0. 0. 0.000764 0.02460811 0.0519972
0.00218131 0. 0.01086884 0.01422362 0. 0.00199138
0. 0.04454033 0.00793646 0.0415923 0.00357104 0.00107033
0.00023718 0.00150827 0.01973532 0.01162116 0.00338441]

Displaying Force Plot for RF SHAP Values from Class 0 in Test Set...

Out [13]:



Checking for matches...Symptoms is Symptoms
Symptoms value is 0.019735316941878446

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.0007752169838459056

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.0

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis B Core Antibody is Hepatitis B Core Antibody
Hepatitis B Core Antibody value is 0.0008855053507077019

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0007639992732988261

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Smoking is Smoking
Smoking value is 0.0015082743786287906

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.0012367462089529642

Checking for matches...Obesity is Obesity
Obesity value is 0.0

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0

Checking for matches...Esophageal Varices is Esophageal Varices
Esophageal Varices value is 0.0

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.0010703293843331851

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.00023718135882364253

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.0

Checking for matches...Age at diagnosis is Age at diagnosis
Age at diagnosis value is 0.006167076926786898

Checking for matches...Grams of Alcohol per day is Grams of Alcohol per day
Grams of Alcohol per day value is 0.003892638996650175

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.007936460247694677

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.04159230337320716

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.055906281750399786

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.02460810937430933

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.03402249946647537

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.0654512610453232

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.01422361902188939

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.002181312845692389

Checking for matches...Platelets is Platelets
Platelets value is 0.0035710406264135093

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.06046394211156472

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.011621157721183936

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.0011186161036797612

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.007491518036034972

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.005090525665644779

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.03486095715017262

Checking for matches...Total Proteins (g/dL) is Total Proteins (g/dL)
Total Proteins (g/dL) value is 0.0033844115849051715

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.00562258648460939

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.001991382127581023

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.010868835835614063

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.04072414767347222

Checking for matches...Iron is Iron
Iron value is 0.051997196164826134

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.044540329249476256

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.01997052406201156

RF2 In CV2...

Checking if correct model is loaded...
RandomForestClassifier(max_depth=11, max_features=None, min_samples_leaf=14,
min_samples_split=27, n_estimators=10, random_state=42)

Checking explainer for RF2...
<shap.explainers._tree.Tree object at 0x7fecc5ddd670>

Checking shap values for RF2...

```
[array([[ 0.01794027, -0.01378879,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.00711062, -0.00448347,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.00711062,  0.00318182,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       ...,
       [ 0.00711062, -0.00448347,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.01210694,  0.00318182,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.01794027, -0.01378879,  0.          , ...,  0.          ,
         0.          ,  0.          ]]), array([[ -0.01794027,  0.01378879,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [-0.00711062,  0.00448347,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [-0.00711062, -0.00318182,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       ...,
       [-0.00711062,  0.00448347,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [-0.01210694, -0.00318182,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [-0.01794027,  0.01378879,  0.          , ...,  0.          ,
         0.          ,  0.          ]]])]
```

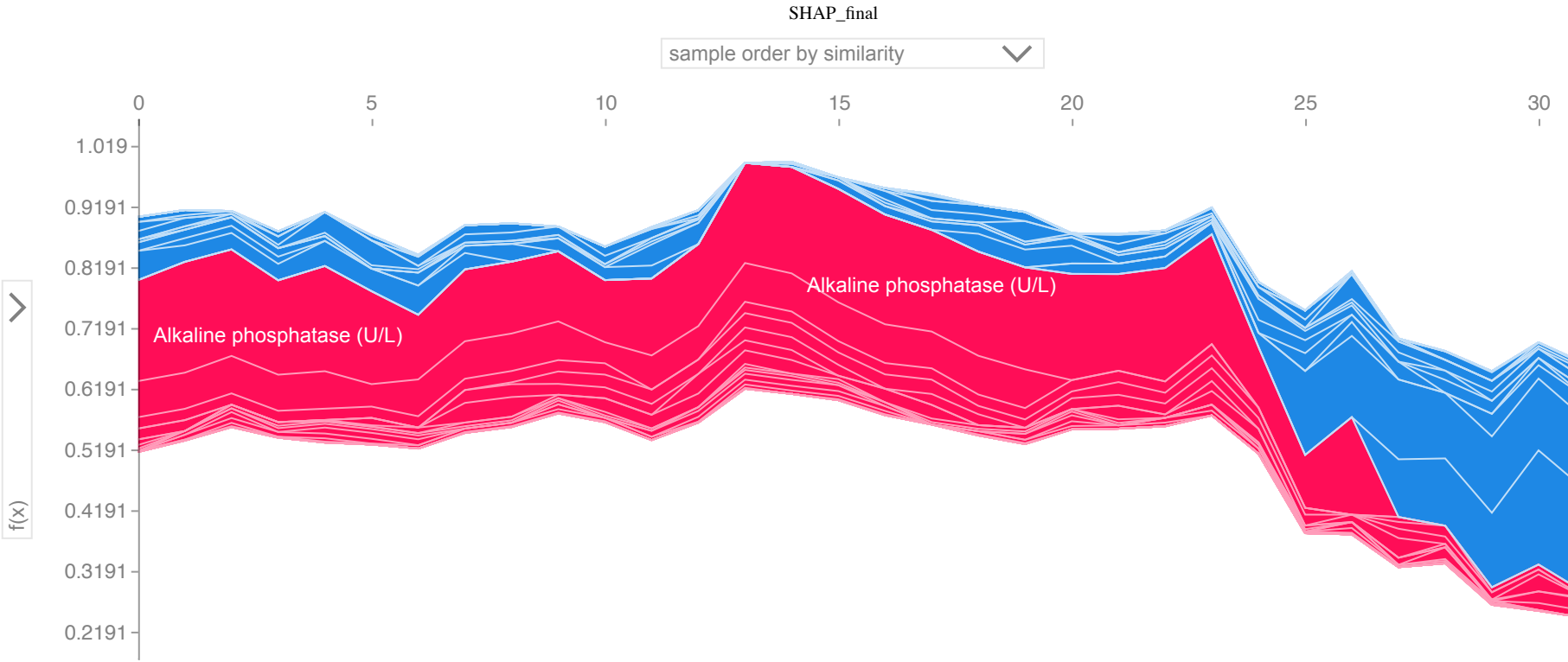
Generating SHAP plots for RF2...

Expected value for RF: [0.61909091 0.38090909]
Saving Bar Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Decision Plot for SHAP Values from Class 0 in Test Set...
Saving Decision Plot for SHAP Values from Class 1 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 0 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 1 in Test Set...
Saving feature importance ranking for RF2...

shap means for RF in CV2...
[0.02955806 0.01578466 0. 0.34348013 0.03598729 0. 0.02417711 0.02192333 0. 0. 0. 0.022355 0. 0. 0. 0.11953585 0.04343473 0. 0. 0.04015099 0.0171707 0.03920637 0. 0. 0.00814669 0.01507267 0. 0. 0. 0.]

Displaying Force Plot for RF SHAP Values from Class 0 in Test Set...

Out [13]:



Checking for matches...Symptoms is Symptoms
Symptoms value is 0.0

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.0

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis B Core Antibody is Hepatitis B Core Antibody
Hepatitis B Core Antibody value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.0

Checking for matches...Obesity is Obesity
Obesity value is 0.0

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.0

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.0

Checking for matches...Human Immunodeficiency Virus is Human Immunodeficiency Virus
Human Immunodeficiency Virus value is 0.0

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.0

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.0

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.0

Checking for matches...Age at diagnosis is Age at diagnosis
Age at diagnosis value is 0.029558056161740173

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.008146688615273838

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.015072670276432032

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.0

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.02417711379529561

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.0

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.0359872854418309

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.0

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.017170701398205646

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.0

Checking for matches...Platelets is Platelets
Platelets value is 0.0

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.01578465616716046

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.021923325686171547

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.0022355033143750366

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.34348012918289555

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.0

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.03920637228510218

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.0401509898135691

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.0

Checking for matches...Iron is Iron
Iron value is 0.04343473360863735

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.0

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.11953585318415587

	Gender	Symptoms	Alcohol	Hepatitis B Surface Antigen	Hepatitis B e Antigen	Hepatitis B Core Antibody	Hepatitis C Virus Antibody	Cirrhosis	Endemic Countries	Smoking	...	Gamma glutamyl transferase (U/L)	Alkaline phosphatase (U/L)	Prothrombin Time (sec)
0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	...	0.003386	0.033483	0.000000
1	0.0	0.019735	0.000775	0.0	0.0	0.000886	0.000764	0.0	0.0	0.001508	...	0.005091	0.034861	0.000000
2	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	...	0.002236	0.343480	0.000000

3 rows x 49 columns

XGB

XGB0 In CV0...

Checking if correct model is loaded...
XGBClassifier(alpha=0.0002575842389979265, base_score=0.5, booster='gbtree', callbacks=None, colsample_bylevel=1, colsample_bynode=1, colsample_bytree=0.9181376162919086, early_stopping_rounds=None, enable_categorical=False, eta=5.623331491160975e-07, eval_metric=None, gamma=0.0002786718840103683, gpu_id=-1, grow_policy='lossguide', importance_type=None, interaction_constraints='', learning_rate=5.62333128e-07, max_bin=256, max_cat_to_onehot=4, max_delta_step=0, max_depth=27, max_leaves=0, min_child_weight=0.20525460238584922, min_samples_leaf=27, min_samples_split=37, missing=nan, monotone_constraints=('', n_estimators=164, n_jobs=1, nthread=1, ...)

Checking explainer for XGB0...
<shap.explainers._tree.Tree object at 0x7fecc7492490>

Checking shap values for XGB0...

[[-3.0952360e-06 -3.7619997e-05 -2.3696571e-06 ... -1.2713954e-07
1.3318062e-07 -2.2155659e-06]
[-1.1003718e-05 1.4350392e-05 2.1088497e-06 ... 6.4105130e-08
4.7997673e-07 2.5433717e-06]
[-3.2040905e-06 1.4657915e-05 1.6045622e-05 ... 2.1734811e-08
1.6514244e-07 -8.6456166e-06]
...
[1.4764141e-06 -2.2817851e-05 -1.9707142e-05 ... 4.7425072e-08
1.5751922e-07 6.8362704e-07]
[1.7882336e-05 9.4924808e-06 9.1785603e-07 ... 2.1734811e-08
-1.7996480e-06 7.2638748e-09]
[-3.0676049e-06 1.4631669e-05 5.5639280e-06 ... 2.1734811e-08
-1.6272652e-06 2.0639069e-07]]

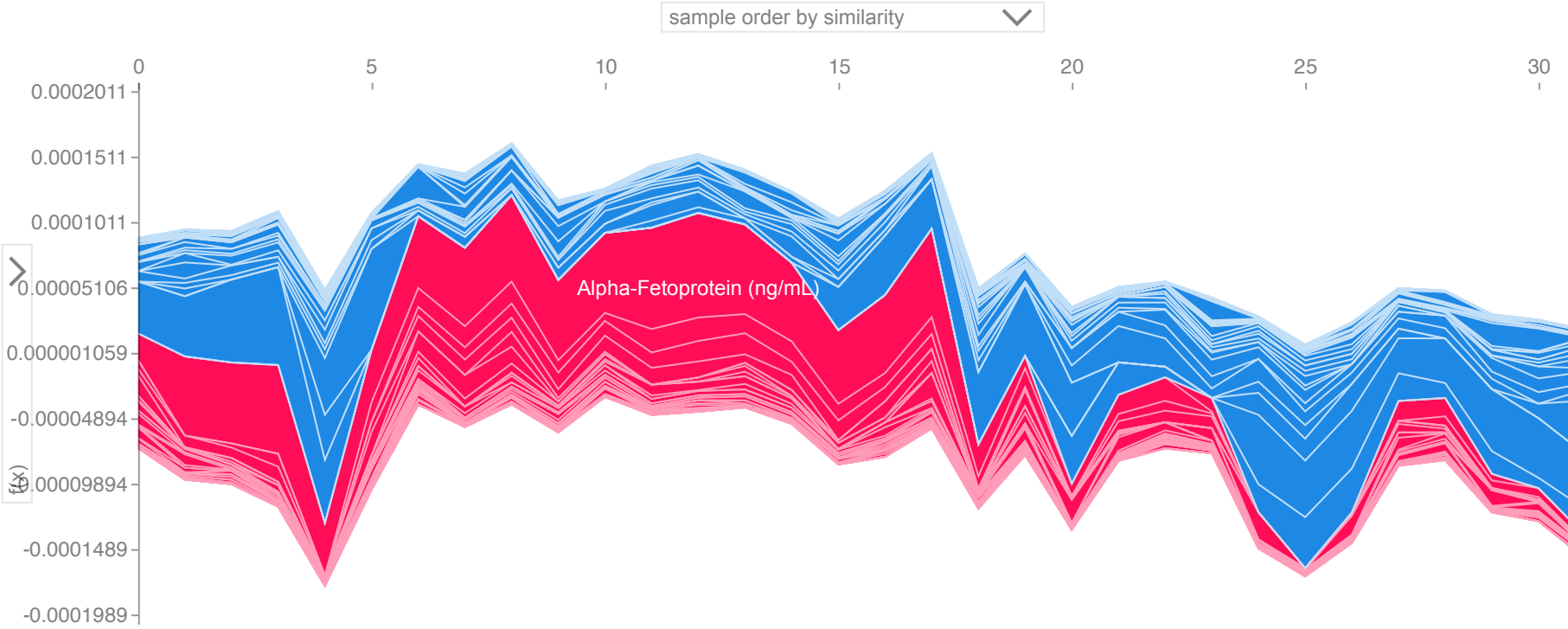
Generating SHAP plots for XGB0...

Expected value for XGB: 1.0591810450932826e-06
Saving Summary Plot for SHAP Values in Test Set...
Saving SHAP Bar Plot for SHAP Values Test Set...
Saving SHAP Decision Plot for SHAP Values in Test Set...
Saving SHAP Decision Plot for Single-Prediction in Test Set...
Saving Waterfall Plot for SHAP Values for a Single-Prediction in Test Set...
Saving feature importance ranking for XGB0...

shap means for XGB in CV0...
[5.0704757e-06 2.2996033e-05 1.3565857e-05 5.5260436e-05 3.6212236e-07
1.6286810e-06 1.4313478e-05 5.8636260e-07 1.2952528e-06 4.3385403e-06
5.9431727e-07 3.7154391e-06 8.8449019e-07 1.5119959e-06 1.2578693e-06
7.5676094e-06 3.1827817e-06 6.2529321e-06 6.7650809e-08 2.4962540e-06
0.0000000e+00 4.5599251e-07 9.7675377e-07 1.1370757e-05 2.0100431e-06
4.2798874e-06 4.0618061e-06 7.1399029e-07 9.4845092e-07 0.0000000e+00
1.4046256e-06 1.1307814e-06 3.3824342e-06 6.6143878e-07 4.8722768e-07
0.0000000e+00 8.0868695e-08 5.8277635e-07 1.1251765e-06]

Displaying Force Plot for XGB SHAP Values in Whole Test Set...

Out [13]:



8/31/22, 2:58 PM

SHAP_final

Checking for matches...Symptoms is Symptoms
Symptoms value is 5.827763516208506e-07

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 2.4962539555417607e-06

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 4.559925059766101e-07

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 1.2952527868037578e-06

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 1.5119959471121547e-06

Checking for matches...Smoking is Smoking
Smoking value is 0.0

Checking for matches...Diabetes is Diabetes
Diabetes value is 5.943172709521605e-07

Checking for matches...Obesity is Obesity
Obesity value is 0.0

Checking for matches...Hemochromatosis is Hemochromatosis
Hemochromatosis value is 6.765080939885593e-08

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 3.621223640948301e-07

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 5.86362602916779e-07

Checking for matches...Esophageal Varices is Esophageal Varices
Esophageal Varices value is 1.2578692576425965e-06

Checking for matches...Splenomegaly is Splenomegaly
Splenomegaly value is 8.08686948516879e-08

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 6.61438775750866e-07

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 4.872276804235298e-07

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 4.279887434677221e-06

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 1.1307813565508695e-06

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 3.3824342153820908e-06

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 8.844901913107606e-07

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 1.6286810478050029e-06

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 9.767537676452775e-07

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 5.526043605641462e-05

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 6.252932053030236e-06

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 7.139902891140082e-07

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 2.010043090194813e-06

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 2.2996033294475637e-05

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 1.1251764817643561e-06

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 5.070475708635058e-06

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 1.4313477549876552e-05

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 3.182781711075222e-06

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 1.3565857443609275e-05

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 4.338540293247206e-06

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 9.484509178037115e-07

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 4.061806066602003e-06

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 3.7154391065996606e-06

Checking for matches...Iron is Iron
Iron value is 1.1370757420081645e-05

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 1.4046255500943516e-06

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 7.567609372927109e-06

XGB1 In CV1...

Checking if correct model is loaded...
XGBClassifier(alpha=0.00029260435288728723, base_score=0.5, booster='gbtree',
 callbacks=None, colsample_bylevel=1, colsample_bynode=1,
 colsample_bytree=0.5441411005619007, early_stopping_rounds=None,
 enable_categorical=False, eta=0.05120369776687421,
 eval_metric=None, gamma=0.4526660690706259, gpu_id=-1,
 grow_policy='depthwise', importance_type=None,
 interaction_constraints='', learning_rate=0.0512036979,
 max_bin=256, max_cat_to_onehot=4, max_delta_step=0, max_depth=18,
 max_leaves=0, min_child_weight=0.12415100550271539,
 min_samples_leaf=9, min_samples_split=27, missing=nan,
 monotone_constraints=()), n_estimators=464, n_jobs=1, nthread=1, ...)

Checking explainer for XGB1...
<shap.explainers._tree.Tree object at 0x7fecc7044370>

Checking shap values for XGB1...

```
[[-0.45429307 -0.06374221 -0.94646686 ... 0.2016137 -0.11681356
  -0.14071724]
 [ 0.5519699  0.13245122 -0.10298917 ... -0.83697766 -0.3408652
  -0.07272914]
 [ 0.21548487 -0.07396804  0.14945313 ... -0.72553927  0.11201834
  0.11979318]
 ...
 [-0.237244 -0.04841679 -0.3906867 ... -0.8205372 -0.11913119
  -0.15298116]
 [-0.33601356 -0.06487641  0.14479543 ... 0.26082134  0.02685894
  -0.13012125]
 [ 0.47213364 -0.15212956 -0.5969934 ... 0.20478216 -0.17158583
  -0.18285887]]
```

Generating SHAP plots for XGB1...

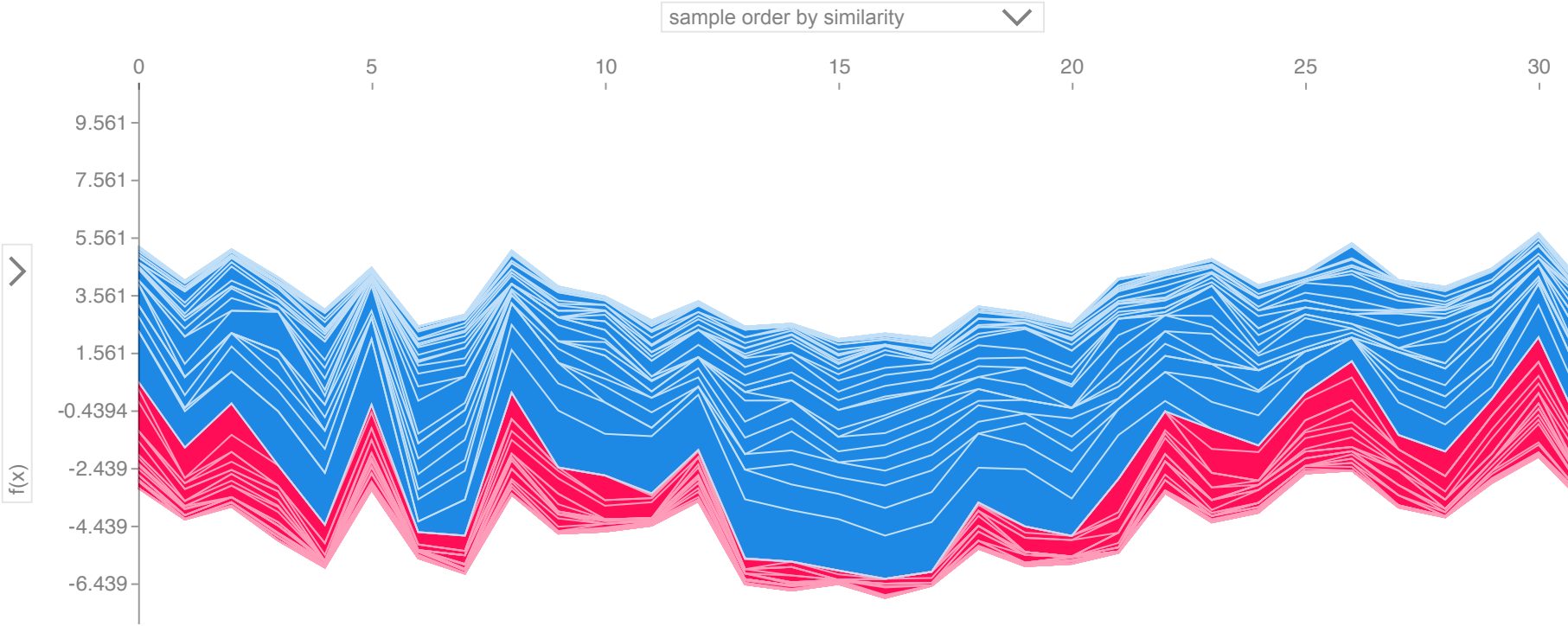
Expected value for XGB: 1.5605792999267578
Saving Summary Plot for SHAP Values in Test Set...
Saving SHAP Bar Plot for SHAP Values Test Set...
Saving SHAP Decision Plot for SHAP Values in Test Set...
Saving SHAP Decision Plot for Single-Prediction in Test Set...
Saving Waterfall Plot for SHAP Values for a Single-Prediction in Test Set...
Saving feature importance ranking for XGB1...

shap means for XGB in CV1...

```
[0.3576165 0.11124935 0.35397023 0.07787172 0.47773322 0.5834394
 0.17404112 0.22541404 0.00938301 0.09362442 0.00415343 0.4011129
 0.04800781 0.          0.36292264 0.06283091 0.14283706 0.94141096
 0.00840189 0.          0.          0.02011156 0.44907653 0.28699368
 0.29294854 0.01342451 0.17626652 0.3927353  0.          0.04733162
 0.          0.76446444 0.10945793 0.30105153 0.11617845 0.01967243
 0.          0.          0.39740062 0.1771822  0.13916534]
```

Displaying Force Plot for XGB SHAP Values in Whole Test Set...

Out [13]:



Checking for matches...Symptoms is Symptoms
Symptoms value is 0.3974006175994873

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.07787172496318817

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.0

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis B Core Antibody is Hepatitis B Core Antibody
Hepatitis B Core Antibody value is 0.008401886560022831

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.020111560821533203

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.009383011609315872

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.04800781235098839

Checking for matches...Smoking is Smoking
Smoking value is 0.0

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.00415342953056097

Checking for matches...Obesity is Obesity
Obesity value is 0.0

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0

Checking for matches...Esophageal Varices is Esophageal Varices
Esophageal Varices value is 0.0

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.019672434777021408

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.0

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.013424506410956383

Checking for matches...Age at diagnosis is Age at diagnosis
Age at diagnosis value is 0.35761651396751404

Checking for matches...Grams of Alcohol per day is Grams of Alcohol per day
Grams of Alcohol per day value is 0.1428370624780655

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.10945793241262436

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.301051527261734

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.17404112219810486

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.4490765333175659

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.5834394097328186

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.9414109587669373

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.3927353024482727

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.29294854402542114

Checking for matches...Platelets is Platelets
Platelets value is 0.11617845296859741

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.3539702296257019

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.17718219757080078

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.11124935001134872

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.22541403770446777

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.06283091008663177

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.47773322463035583

Checking for matches...Total Proteins (g/dL) is Total Proteins (g/dL)
Total Proteins (g/dL) value is 0.13916534185409546

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.09362442046403885

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.04733162000775337

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.17626652121543884

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.4011129140853882

Checking for matches...Iron is Iron
Iron value is 0.28699368238449097

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.7644644379615784

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.36292263865470886

XGB2 In CV2...

Checking if correct model is loaded...
XGBClassifier(alpha=5.77534955247629e-07, base_score=0.5, booster='gbtree',
 callbacks=None, colsample_bylevel=1, colsample_bynode=1,
 colsample_bytree=0.41771820514444086, early_stopping_rounds=None,
 enable_categorical=False, eta=8.67291826605322e-06,
 eval_metric=None, gamma=0.07212410933578818, gpu_id=-1,
 grow_policy='lossguide', importance_type=None,
 interaction_constraints='', learning_rate=8.67291874e-06,
 max_bin=256, max_cat_to_onehot=4, max_delta_step=0, max_depth=22,
 max_leaves=0, min_child_weight=6.66045104839759,
 min_samples_leaf=11, min_samples_split=39, missing=nan,
 monotone_constraints='()', n_estimators=884, n_jobs=1, nthread=1, ...)

Checking explainer for XGB2...
<shap.explainers._tree.Tree object at 0x7fecb3220c10>

Checking shap values for XGB2...

```
[[[-1.8894803e-04  2.3255567e-04  0.0000000e+00 ...  0.0000000e+00
   -8.9294117e-06 -7.9064384e-06]
 [-1.9001741e-04  1.9959988e-04  0.0000000e+00 ...  0.0000000e+00
   -1.0370755e-05  3.5941350e-06]
 [-1.6012945e-04  1.6149672e-04  0.0000000e+00 ...  0.0000000e+00
   -1.5008896e-05  3.5941350e-06]
 ...
 [-1.7405280e-04  2.5253580e-04  0.0000000e+00 ...  0.0000000e+00
   -1.0370755e-05  3.5941350e-06]
 [-1.8333328e-04 -4.4896262e-04  0.0000000e+00 ...  0.0000000e+00
   -1.5008896e-05  3.5941350e-06]
 [-1.6480772e-04  2.5592663e-04  0.0000000e+00 ...  0.0000000e+00
   -1.3567553e-05  3.5941350e-06]]
```

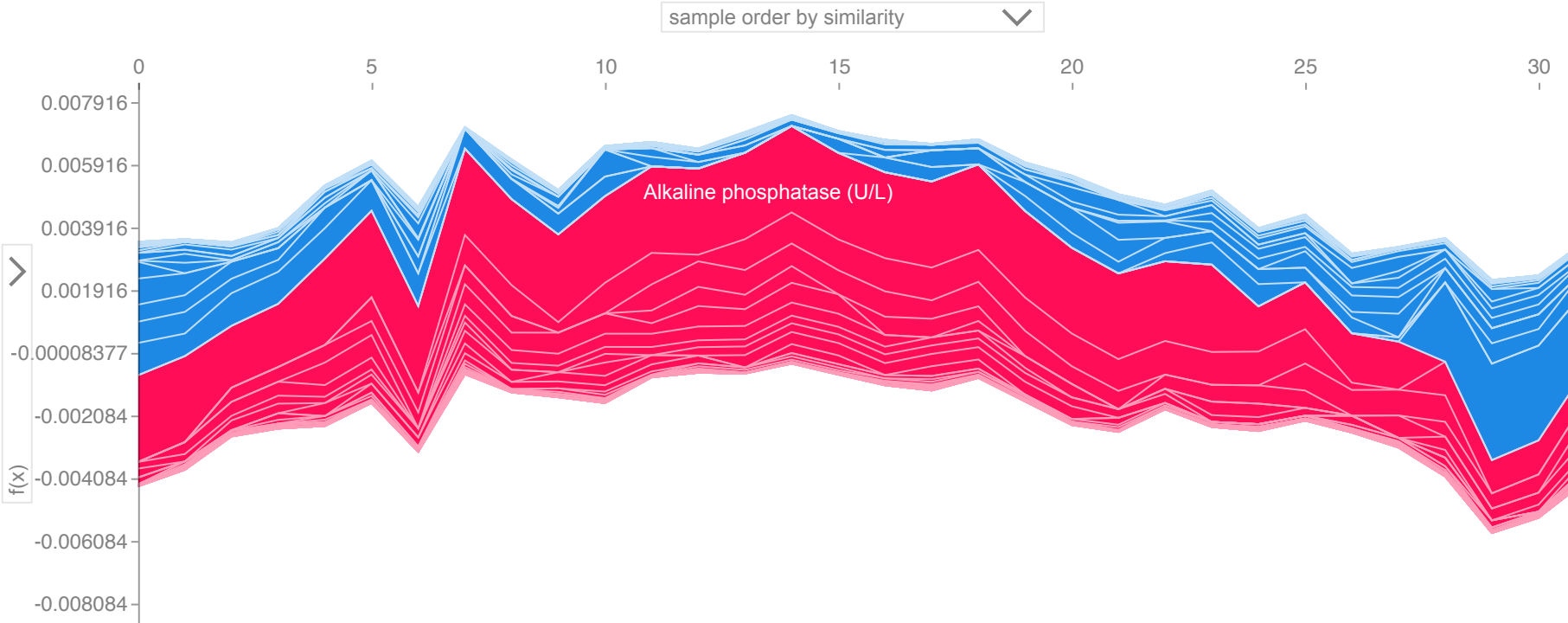
Generating SHAP plots for XGB2...

Expected value for XGB: -8.377160702366382e-05
Saving Summary Plot for SHAP Values in Test Set...
Saving SHAP Bar Plot for SHAP Values Test Set...
Saving SHAP Decision Plot for SHAP Values in Test Set...
Saving SHAP Decision Plot for Single-Prediction in Test Set...
Saving Waterfall Plot for SHAP Values for a Single-Prediction in Test Set...
Saving feature importance ranking for XGB2...

shap means for XGB in CV2...
[1.8230948e-04 2.8346531e-04 0.0000000e+00 2.6645351e-03 9.0496865e-04
0.0000000e+00 0.0000000e+00 6.8523477e-05 0.0000000e+00 2.5999354e-05
1.5617621e-05 1.3878002e-05 0.0000000e+00 0.0000000e+00 6.4939231e-04
2.9202370e-04 5.7312089e-04 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00 2.1182994e-05 5.4415216e-04 2.8262383e-04 1.1534676e-05
2.2458608e-04 3.4793742e-05 3.9722818e-05 0.0000000e+00 8.0097008e-05
3.5846941e-05 4.6863221e-04 1.0432031e-04 0.0000000e+00 1.3717378e-05
5.4281068e-06]

Displaying Force Plot for XGB SHAP Values in Whole Test Set...

Out [13]:



Checking for matches...Symptoms is Symptoms
Symptoms value is 5.428106760518858e-06

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.0

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis B Core Antibody is Hepatitis B Core Antibody
Hepatitis B Core Antibody value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Diabetes is Diabetes
Diabetes value is 1.561762110213749e-05

Checking for matches...Obesity is Obesity
Obesity value is 0.0

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.0

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.0

Checking for matches...Human Immunodeficiency Virus is Human Immunodeficiency Virus
Human Immunodeficiency Virus value is 0.0

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.0

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 1.3717378351429943e-05

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 1.153467565018218e-05

Checking for matches...Age at diagnosis is Age at diagnosis
Age at diagnosis value is 0.00018230947898700833

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 3.584694059100002e-05

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.00046863220632076263

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.0

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.0

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 2.118299380526878e-05

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.0009049686486832798

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.0005731208948418498

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 3.479374208836816e-05

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.00028262383420951664

Checking for matches...Platelets is Platelets
Platelets value is 0.00010432030830997974

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.0002834653132595122

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 6.85234772390686e-05

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.0002920237020589411

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.0026645350735634565

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 2.5999353965744376e-05

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 3.972281774622388e-05

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.00022458608145825565

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 1.3878002391720656e-05

Checking for matches...Iron is Iron
Iron value is 0.0005441521643660963

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 8.009700832189992e-05

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.0006493923137895763

	Gender	Symptoms	Alcohol	Hepatitis B Surface Antigen	Hepatitis B e Antigen	Hepatitis B Core Antibody	Hepatitis C Virus Antibody	Cirrhosis	Endemic Countries	Smoking	...	Gamma glutamyl transferase (U/L)	Alkaline phosphatase (U/L)
0	0.0	5.827764e-07	0.000000	0.000002	0.0	0.000000	4.559925e-07	0.000001	0.000002	0.0	...	0.000003	0.000014
1	0.0	3.974006e-01	0.077872	0.000000	0.0	0.008402	2.011156e-02	0.009383	0.048008	0.0	...	0.062831	0.477733
2	0.0	5.428107e-06	0.000000	0.000000	0.0	0.000000	0.000000e+00	0.000000	0.000000	0.0	...	0.000292	0.002665

3 rows x 49 columns

```
-----  
hcc-data_example_no_covariates  
-----  
-----  
NB  
-----  
NB0 In CV0...
```

```
Checking if correct model is loaded...  
GaussianNB()
```

```
Checking explainer for NB0...  
shap.explainers.Permutation()
```

```
Checking shap values for NB0...
```

```
.values =  
array([[ 0.00333333, -0.02      ,  0.      , ...,  0.005      ,  
        -0.03083333, -0.02916667],  
       [ 0.005      ,  0.00583333,  0.00083333, ...,  0.01416667,  
        -0.03166667, -0.02333333],  
       [ 0.00583333,  0.03583333,  0.      , ...,  0.08083333,  
        -0.005      , -0.02      ],  
       ...,  
       [-0.00083333, -0.01166667,  0.      , ...,  0.005      ,  
        -0.03833333, -0.03      ],  
       [ 0.      , -0.00416667,  0.      , ..., -0.01916667,  
        -0.02      , -0.03416667],  
       [ 0.0075     ,  0.01666667,  0.      , ..., -0.03333333,  
        0.50166667, -0.02083333]])  
  
.base_values =  
array([0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32,  
       0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32,  
       0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32,  
       0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32,  
       0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32, 0.32])
```

```
.data =  
array([[ 0.0376161,  1.0547612, -1.7113069, ...,  0.6264224, -0.4385377,  
        -0.1829094],  
       [-0.331668 , -0.2723382, -1.7113069, ...,  0.6264224, -0.3902633,  
        -0.3086845],  
       [-0.096669 ,  0.3174837, -1.7113069, ...,  0.6264224, -0.486812 ,  
        -0.1829094],  
       ...,  
       [-0.667381 ,  0.6123947,  0.5843487, ...,  0.6264224, -0.3741719,  
        -0.1290058],  
       [-0.9191656,  0.3174837,  0.5843487, ..., -1.5963668, -0.2615318,  
        -0.1379898],  
       [ 0.13833  , -0.346066 ,  0.5843487, ..., -1.5963668,  1.154515 ,  
        -0.1379898]])
```

```
Generating SHAP plots for NB0...
```

```
Saving Summary Plot for SHAP Values in Class 0 & 1 in Test Set...  
Saving Summary Plot for SHAP Values in Class 0 & 1 in Test Set...  
Saving SHAP Beeswarm Plot for Top 5 SHAP Values in Class 0 & 1 in Test Set...  
Saving feature importance ranking for NB0...
```

```
shap means for NB in CV0...  
[0.00557576 0.0175      0.00022727 0.01109091 0.01821212 0.02536364  
 0.02621212 0.02556061 0.00739394 0.0705303  0.01254545 0.03837879  
 0.06295455 0.01445455 0.0185      0.00837879 0.00975758 0.01377273  
 0.00560606 0.01034848 0.01030303 0.02404545 0.01722727 0.01187879  
 0.03071212 0.01921212 0.00433333 0.00762121 0.00292424 0.01389394  
 0.02289394 0.03425758 0.03477273 0.00206061 0.01924242 0.0425303  
 0.05490909]
```

```
Saving Force Plot for NB SHAP Values in Test Set...
```

```
Checking for matches...Symptoms is Symptoms  
Symptoms value is 0.019242424242424252
```

```
Checking for matches...Alcohol is Alcohol  
Alcohol value is 0.0002272727272727274
```

```
Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen  
Hepatitis B Surface Antigen value is 0.010348484848484846
```

```
Checking for matches...Hepatitis B Core Antibody is Hepatitis B Core Antibody  
Hepatitis B Core Antibody value is 0.005606060606060607
```

```
Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody  
Hepatitis C Virus Antibody value is 0.010303030303030305
```

```
Checking for matches...Cirrhosis is Cirrhosis  
Cirrhosis value is 0.007393939393939392
```

```
Checking for matches...Endemic Countries is Endemic Countries  
Endemic Countries value is 0.014454545454545458
```

```
Checking for matches...Diabetes is Diabetes
```

Diabetes value is 0.01254545454545454

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.025560606060606065

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.00762121212121215

Checking for matches...Esophageal Varices is Esophageal Varices
Esophageal Varices value is 0.018499999999999996

Checking for matches...Splenomegaly is Splenomegaly
Splenomegaly value is 0.00206060606060606

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.03477272727272727

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.030712121212121208

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.0228939393939394

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.03425757575757576

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.06295454545454544

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.025363636363636363

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.024045454545454537

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.018212121212121207

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.013772727272727273

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.004333333333333336

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.011878787878787874

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.0175

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.04253030303030302

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.005575757575757576

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.02621212121212121

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.009757575757575757

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.011090909090909092

Checking for matches...Total Proteins (g/dL) is Total Proteins (g/dL)
Total Proteins (g/dL) value is 0.05490909090909089

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.07053030303030303

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.00292424242424243

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.01921212121212121

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.03837878787878788

Checking for matches...Iron is Iron
Iron value is 0.017227272727272723

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.013893939393939396

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.008378787878787878

NB1 In CV1...

Checking if correct model is loaded...


```
GaussianNB()

Checking explainer for NB1...
shap.explainers.Permutation()

Checking shap values for NB1...

.values =
array([[ 0.015      , -0.07166667,  0.34416667, ..., -0.01083333,
         0.115      , -0.00583333],
       [-0.00083333, -0.01166667, -0.04833333, ..., -0.0725     ,
        -0.21416667, -0.01916667],
       [ 0.00666667,  0.02666667,  0.02083333, ..., -0.00666667,
        -0.095      , -0.0025     ],
       ...,
       [-0.16666667, -0.06833333,  0.14583333, ..., -0.04       ,
        -0.1875     , -0.04916667],
       [-0.00083333,  0.01916667,  0.0075     , ...,  0.02       ,
         0.015      , -0.00333333],
       [ 0.02416667, -0.0275     , -0.00166667, ..., -0.02       ,
         0.1075     , -0.01583333]])

.base_values =
array([0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48,
       0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48,
       0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48,
       0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48, 0.48])

.data =
array([[ -0.3307946,  1.9018224,  2.7240451, ..., -0.5141671,  0.6831301,
        -0.4175834],
       [-0.9380037,  0.0449382, -0.5189802, ..., -0.5141671, -1.4638501,
        -0.449261 ],
       [-0.2159173, -0.5585492,  0.9323539, ..., -0.5141671, -1.4638501,
        -0.1800009],
       ...,
       [ 0.8343903,  0.1996785,  1.465366 , ..., -0.5141671, -1.4638501,
        -0.3700669],
       [ 1.1462003, -0.1098022,  1.3626167, ...,  1.944893 ,  0.6831301,
        -0.2433562],
       [-0.6754268,  1.1281206,  0.1103594, ..., -0.5141671,  0.6831301,
        -0.3859057]])

Generating SHAP plots for NB1...

Saving Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving SHAP Beeswarm Plot for Top 5 SHAP Values in Class 0 & 1 in Test Set...
Saving feature importance ranking for NB1...

shap means for NB in CV1...
[0.01628788 0.04440909 0.04562121 0.02784848 0.00575758 0.05356061
 0.00787879 0.          0.00298485 0.00728788 0.0235303  0.02672727
 0.0954697  0.01274242 0.00427273 0.03601515 0.00593939 0.00278788
 0.00928788 0.03927273 0.02130303 0.00051515 0.02809091 0.01386364
 0.01428788 0.06469697 0.00287879 0.00668182 0.01821212 0.01581818
 0.05901515 0.01725758 0.03760606 0.02651515 0.06425758 0.02471212]

Saving Force Plot for NB SHAP Values in Test Set...

Checking for matches...Symptoms is Symptoms
Symptoms value is 0.06425757575757576

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.002787878787878788

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.009287878787878788

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.09546969696969697

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.007287878787878787

Checking for matches...Obesity is Obesity
Obesity value is 0.006681818181818183

Checking for matches...Hemochromatosis is Hemochromatosis
Hemochromatosis value is 0.005939393939393941

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.005757575757575757

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0646969696969697

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.037606060606060606
```

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.02651515151515152

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.028090909090909083

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.015818181818181815

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.059015151515151534

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.026727272727272714

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.05356060606060604

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.03927272727272726

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.027848484848484848

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.036015151515151514

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.01428787878787879

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.00051515151515151

Checking for matches...Platelets is Platelets
Platelets value is 0.017257575757575753

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.044409090909090905

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.024712121212121217

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.016287878787878785

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.007878787878787883

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.004272727272727272

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.04562121212121212

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.0029848484848484856

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.00287878787878787

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.013863636363636364

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.023530303030303023

Checking for matches...Iron is Iron
Iron value is 0.021303030303030306

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.018212121212121218

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.012742424242424244

```
-----  
NB2 In CV2...  
  
Checking if correct model is loaded...  
GaussianNB()  
  
Checking explainer for NB2...  
shap.explainers.Permutation()  
  
Checking shap values for NB2...  
  
.values =  
array([[ 0.          ,  0.          ,  0.          , ...,  0.          ,  
        0.          ,  0.          ],  
       [ 0.          ,  0.          ,  0.01666667, ...,  0.          ,  
        0.          ,  0.          ],  
       [ 0.00083333, -0.02583333,  0.00416667, ..., -0.01          ,
```

```
-0.02      ,  0.00916667],
...,
[ 0.02666667,  0.0025      ,  0.02166667, ..., -0.00166667,
-0.01083333, -0.00083333],
[-0.03916667,  0.005      , -0.0125      , ..., -0.00583333,
-0.025      , -0.01      ],
[ 0.      ,  0.      ,  0.      , ...,  0.      ,
0.      ,  0.      ]])

.base_values =
array([0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05,
0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05,
0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05,
0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05])

.data =
array([[ -0.4184767,  0.5703518, -0.3573003, ...,  0.6264224, -0.5282705,
-0.7859052],
[ -0.4184767,  0.5703518,  4.7719674, ..., -1.5963668, -0.5282705,
1.272418 ],
[ 0.0205491, -1.7533038,  0.3559594, ...,  0.6264224, -0.5282705,
-0.7859052],
...,
[ -1.8818959,  0.5703518,  1.5964111, ...,  0.6264224, -0.5282705,
1.272418 ],
[ 1.0449426,  0.5703518, -0.1898393, ...,  0.6264224, -0.5282705,
1.272418 ],
[ -2.0282378,  0.5703518, -0.5867838, ...,  0.6264224, -0.5282705,
-0.7859052]])
```

Generating SHAP plots for NB2...

Saving Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving SHAP Beeswarm Plot for Top 5 SHAP Values in Class 0 & 1 in Test Set...
Saving feature importance ranking for NB2...

```
shap means for NB in CV2...
[0.00178788 0.00072727 0.00142424 0.00039394 0.0004697  0.00260606
0.00281818 0.00066667 0.00043939 0.00048485 0.00066667 0.06024242
0.00016667 0.01372727 0.00040909 0.00143939 0.00134848 0.00183333
0.01024242 0.00092424 0.05330303 0.01442424 0.00065152 0.00259091
0.00631818 0.0009697  0.00090909 0.00266667 0.00154545 0.00030303
0.01784848 0.00240909 0.00075758 0.00036364 0.00156061 0.0005      ]
```

Saving Force Plot for NB SHAP Values in Test Set...

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.0007272727272727274

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.0018333333333333333

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.010242424242424249

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.00092424242424242424

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0004393939393939394

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0001666666666666667

Checking for matches...Smoking is Smoking
Smoking value is 0.0005000000000000002

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.0006666666666666666

Checking for matches...Hemochromatosis is Hemochromatosis
Hemochromatosis value is 0.0013484848484848478

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.0004696969696969697

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.0006666666666666668

Checking for matches...Human Immunodeficiency Virus is Human Immunodeficiency Virus
Human Immunodeficiency Virus value is 0.05330303030303031

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0026666666666666666

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.0003636363636363637

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.0015606060606060605

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.0063181818181817

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.01784848484848485

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.0024090909090909093

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.0026060606060606057

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.014424242424242424

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.00039393939393939407

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.0014393939393939393

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0009090909090909089

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.0025909090909090908

Checking for matches...Platelets is Platelets
Platelets value is 0.0007575757575757574

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.001787878787878788

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.002818181818181818

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.000409090909090909

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.001424242424242424

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.0004848484848484849

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.0015454545454545454

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.0009696969696969691

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.060242424242424236

Checking for matches...Iron is Iron
Iron value is 0.0006515151515151514

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.0003030303030303031

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.013727272727272729

	Gender	Symptoms	Alcohol	Hepatitis B Surface Antigen	Hepatitis B e Antigen	Hepatitis B Core Antibody	Hepatitis C Virus Antibody	Cirrhosis	Endemic Countries	Smoking	...	Gamma glutamyl transferase (U/L)	Alkaline phosphatase (U/L)	Pr...
0	0.0	0.019242	0.000227	0.010348	0.000000	0.005606	0.010303	0.007394	0.014455	0.0000	...	0.009758	0.011091	0.0...
1	0.0	0.064258	0.000000	0.000000	0.002788	0.000000	0.009288	0.000000	0.095470	0.0000	...	0.004273	0.045621	0.0...
2	0.0	0.000000	0.000727	0.001833	0.010242	0.000000	0.000924	0.000439	0.000167	0.0005	...	0.000409	0.001424	0.0...

3 rows x 49 columns

```
-----
LR
-----
LR0 In CV0...

Checking if correct model is loaded...
LogisticRegression(C=0.0076324520136090606, dual=True,
                    max_iter=383.683139958808, random_state=42,
                    solver='liblinear')

Checking explainer for LR0...
<shap.explainers._linear.Linear object at 0x7fecc7005ca0>

Checking shap values for LR0...

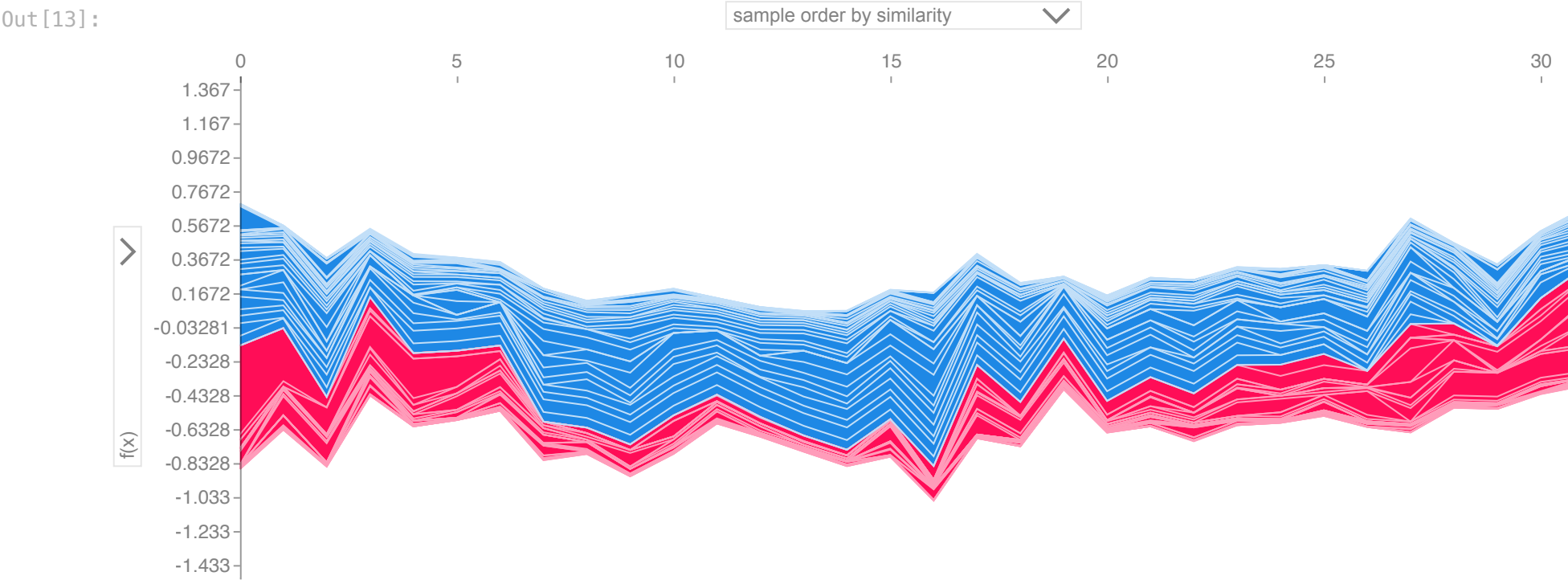
[[-0.00015674 -0.05535253  0.01520368 ...  0.02453187 -0.02176526
  -0.01030959]
 [ 0.00068073  0.02080485  0.01520368 ...  0.02453187 -0.01946545
  -0.01559663]
 [ 0.0001478  -0.01304287  0.01520368 ...  0.02453187 -0.02406506
  -0.01030959]
 ...
 [ 0.00144207 -0.02996674 -0.00683064 ...  0.02453187 -0.01869885
  -0.00804372]
 [ 0.00201307 -0.01304287 -0.00683064 ... -0.0663269  -0.01333263
  -0.00842137]
 [-0.00038514  0.02503582 -0.00683064 ... -0.0663269   0.05412839
  -0.00842137]]

Generating SHAP plots for LR0...

Expected value for LR: -0.032810633342135964
Saving Summary Plot for SHAP Values in Test Set...
Saving SHAP Bar Plot for SHAP Values Test Set...
Saving SHAP Decision Plot for SHAP Values in Test Set...
Saving SHAP Decision Plot for Single-Prediction in Test Set...
Saving Waterfall Plot for SHAP Values for a Single-Prediction in Test Set...
Saving feature importance ranking for LR0...
```

```
shap means for LR in CV0...
[0.00140116 0.04528556 0.00911419 0.04269167 0.00676139 0.04691372
 0.04760978 0.02417621 0.01084964 0.03987523 0.04972073 0.01931754
 0.01954522 0.01863871 0.04679377 0.04076409 0.0305404  0.03517929
 0.01120396 0.0219318  0.02888885 0.01847507 0.00279009 0.02309836
 0.06607258 0.04874631 0.00372936 0.00598961 0.00118051 0.00295957
 0.01536827 0.07389625 0.05683262 0.00165448 0.04124988 0.01827661
 0.01735989]
```

Displaying Force Plot for LR SHAP Values in Whole Test Set...



Checking for matches...Symptoms is Symptoms
Symptoms value is 0.04124988333284681

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.009114194957419365

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.021931800198248246

Checking for matches...Hepatitis B Core Antibody is Hepatitis B Core Antibody
Hepatitis B Core Antibody value is 0.011203964193253292

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.028888854722906872

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.010849640129879672

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.01863871226140451

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.049720725126536205

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.024176207685650656

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0059896102049319675

Checking for matches...Esophageal Varices is Esophageal Varices
Esophageal Varices value is 0.04679376824390308

Checking for matches...Splenomegaly is Splenomegaly
Splenomegaly value is 0.001654476497957493

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.056832615018323526

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.06607258151962792

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.015368266419514635

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.07389624848357253

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.019545221175570546

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.04691372270672035

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.018475069395498186

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.006761393898885898

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.035179291380850006

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0037293579510649677

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.023098362493688785

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.04528556133231627

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.018276606287655997

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.0014011642070957946

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.0476097788016792

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.030540403355380764

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.042691667541604264

Checking for matches...Total Proteins (g/dL) is Total Proteins (g/dL)
Total Proteins (g/dL) value is 0.017359885564995438

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.03987522970487768

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.0011805113171685208

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.048746311825010664

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.019317543618077915

Checking for matches...Iron is Iron
Iron value is 0.0027900938299610955

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.0029595698369759994

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.04076409160970465

LR1 In CV1...

Checking if correct model is loaded...
LogisticRegression(C=0.0003630322466779861, class_weight='balanced',
max_iter=159.30522616241012, random_state=42)

Checking explainer for LR1...
<shap.explainers._linear.Linear object at 0x7fecc6fea5b0>

Checking shap values for LR1...

```
[[ 0.00034788 -0.01330814  0.01706337 ... -0.00222993  0.00364155
 -0.00301374]
 [ 0.00099548 -0.00012806 -0.00335686 ... -0.00222993 -0.00984567
 -0.00321245]
 [ 0.00022536  0.00415547  0.0057817  ... -0.00222993 -0.00984567
 -0.00152343]
 ...
 [-0.00089483 -0.0012264   0.00913789 ... -0.00222993 -0.00984567
 -0.00271568]
 [-0.00122738  0.00097028  0.00849092 ...  0.00602908  0.00364155
 -0.00192084]
 [ 0.00071544 -0.00781644  0.00060588 ... -0.00222993  0.00364155
 -0.00281503]]
```

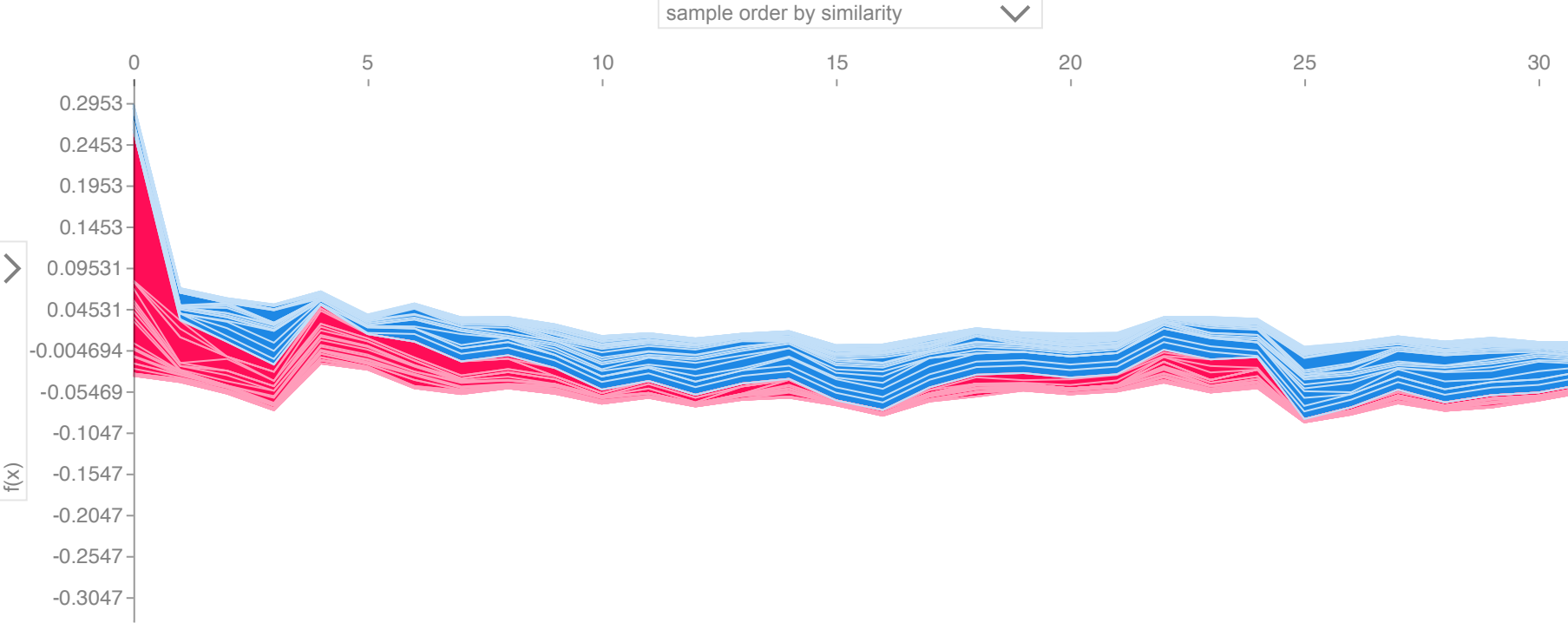
Generating SHAP plots for LR1...

Expected value for LR: -0.004693505131985552
Saving Summary Plot for SHAP Values in Test Set...
Saving SHAP Bar Plot for SHAP Values Test Set...
Saving SHAP Decision Plot for SHAP Values in Test Set...
Saving SHAP Decision Plot for Single-Prediction in Test Set...
Saving Waterfall Plot for SHAP Values for a Single-Prediction in Test Set...
Saving feature importance ranking for LR1...

shap means for LR in CV1...
[6.59582008e-04 6.40443545e-03 5.19879014e-03 4.66436210e-03
1.10540294e-03 7.15460519e-03 1.45982163e-03 2.36738955e-05
4.88314797e-04 1.38603130e-03 2.13837462e-03 2.88647226e-03
2.15906676e-03 2.74347731e-03 1.00568752e-03 6.42732235e-03
1.51142641e-04 5.20122378e-04 1.22688122e-03 5.78614231e-03
3.65582895e-03 1.66643432e-04 2.59137018e-03 2.48640295e-03
1.05485235e-03 2.05926953e-03 6.75858829e-04 4.42724624e-04
1.26668493e-03 4.01168683e-04 7.80247284e-03 1.92864839e-03
3.20187976e-03 3.12791521e-03 5.67198922e-03 2.52836113e-03]

Displaying Force Plot for LR SHAP Values in Whole Test Set...

Out[13]:



Checking for matches...Symptoms is Symptoms
Symptoms value is 0.005671989215052704

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0005201223783531668

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.001226881222520814

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 2.3673895500685356e-05

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.002159066761354565

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.0013860313033

Checking for matches...Obesity is Obesity
Obesity value is 0.00044272462420167925

Checking for matches...Hemochromatosis is Hemochromatosis
Hemochromatosis value is 0.00015114264072037324

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.0011054029444667148

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.002059269525580652

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.003201879763155339

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.003127915213817235

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.002591370179934829

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.00040116868337315156

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.007802472836625624

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.00288647225535381

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.007154605186744367

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.005786142312824131

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.004664362095345179

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.006427322350691086

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0010548523452812277

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.0001666434324191902

Checking for matches...Platelets is Platelets
Platelets value is 0.001928648391141849

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.00640443544504057

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.002528361125426319

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.0006595820079446869

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.0014598216340550093

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.0010056875205088232

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.0051987901388542085

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.0004883147965576142

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.0006758588294186788

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.0024864029457391038

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.002138374617276843

Checking for matches...Iron is Iron
Iron value is 0.0036558289538485738

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.001266684934688209

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.0027434773085111695

LR2 In CV2...

Checking if correct model is loaded...
LogisticRegression(C=1.1355889848345609e-05, dual=True,
max_iter=287.08753481954665, random_state=42,
solver='liblinear')

Checking explainer for LR2...
<shap.explainers._linear.Linear object at 0x7feca047a640>

Checking shap values for LR2...

```
[[ 7.32083460e-05  1.92822706e-05 -4.36797008e-05 ... -3.15759647e-05
 -7.30478160e-05  3.36532768e-05]
 [ 7.32083460e-05  1.92822706e-05  8.74173101e-04 ...  8.11953378e-05
 -7.30478160e-05 -5.73015253e-05]
 [ 7.45330978e-06 -7.71290826e-05  8.39539991e-05 ... -3.15759647e-05
 -7.30478160e-05  3.36532768e-05]
 ...
 [ 2.92391780e-04  1.92822706e-05  3.05925660e-04 ... -3.15759647e-05
 -7.30478160e-05 -5.73015253e-05]
 [-1.45975103e-04  1.92822706e-05 -1.37135230e-05 ... -3.15759647e-05
 -7.30478160e-05 -5.73015253e-05]
 [ 3.14310121e-04  1.92822706e-05 -8.47444465e-05 ... -3.15759647e-05
 -7.30478160e-05  3.36532768e-05]]
```

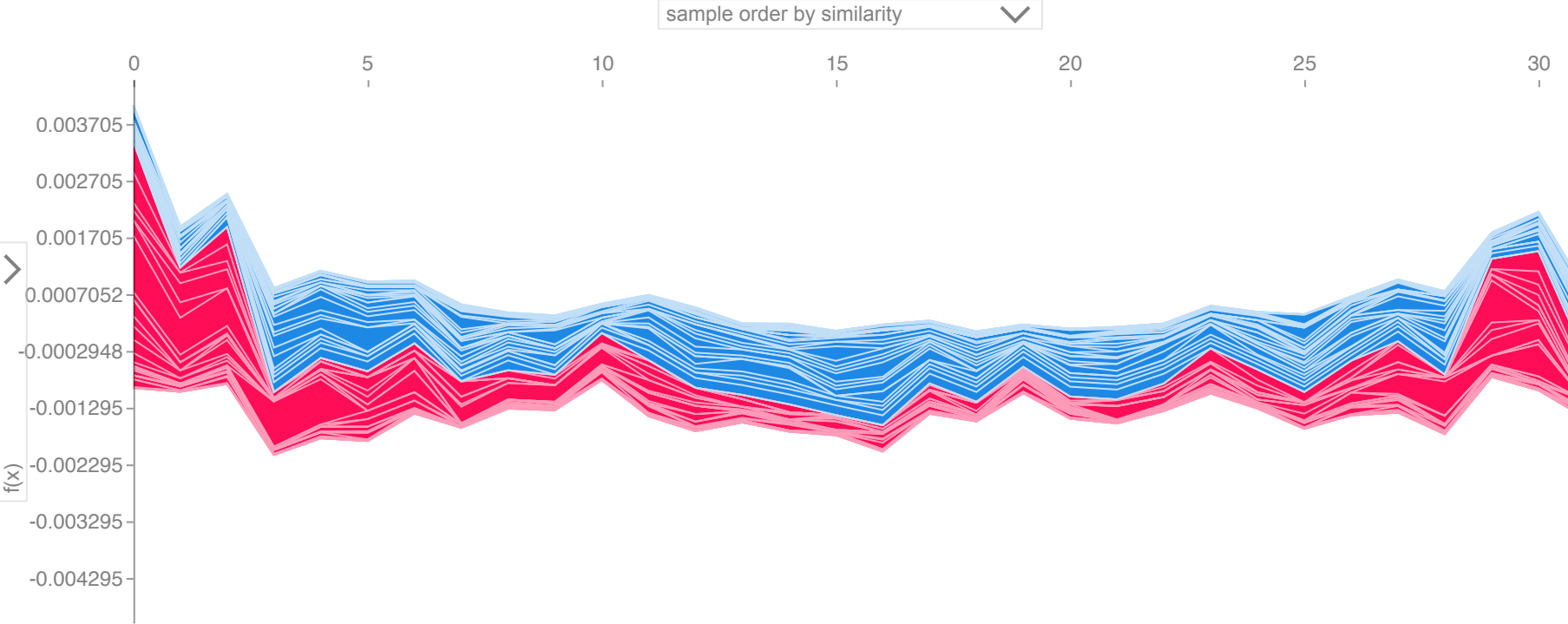
Generating SHAP plots for LR2...

Expected value for LR: -0.0002947812564940089
Saving Summary Plot for SHAP Values in Test Set...
Saving SHAP Bar Plot for SHAP Values Test Set...
Saving SHAP Decision Plot for SHAP Values in Test Set...
Saving SHAP Decision Plot for Single-Prediction in Test Set...
Saving Waterfall Plot for SHAP Values for a Single-Prediction in Test Set...
Saving feature importance ranking for LR2...

shap means for LR in CV2...
[1.20703617e-04 3.61104341e-05 1.30361444e-04 4.21001643e-07
6.21079433e-05 1.07603971e-04 7.28317332e-05 2.64396863e-05
2.10054599e-05 1.56830750e-05 8.57463121e-05 8.38978365e-05
3.50432110e-06 1.24865360e-04 6.42190902e-05 1.10112994e-04
1.58137203e-05 3.74835449e-05 1.62587155e-05 2.91571716e-05
4.46537494e-05 5.93378478e-05 8.84658729e-05 1.04116231e-04
1.47733584e-04 9.41730202e-05 3.43163849e-05 8.75614518e-06
9.53497753e-05 3.88489713e-06 6.60462180e-06 1.80050043e-04
7.03508691e-05 4.33041801e-05 1.20860932e-04 4.26826080e-05]

Displaying Force Plot for LR SHAP Values in Whole Test Set...

Out[13]:



Checking for matches...Alcohol is Alcohol
Alcohol value is 3.611043410527636e-05

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 3.7483544912258247e-05

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 1.6258715545930826e-05

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 2.91571715918726e-05

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 2.1005459891650037e-05

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 3.504321104586795e-06

Checking for matches...Smoking is Smoking
Smoking value is 4.26826080104935e-05

Checking for matches...Diabetes is Diabetes
Diabetes value is 8.574631207165098e-05

Checking for matches...Hemochromatosis is Hemochromatosis
Hemochromatosis value is 1.581372030006737e-05

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 6.210794327084585e-05

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 2.643968628389975e-05

Checking for matches...Human Immunodeficiency Virus is Human Immunodeficiency Virus
Human Immunodeficiency Virus value is 4.465374941547495e-05

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 8.756145175607693e-06

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 4.330418013553576e-05

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.00012086093191716493

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.00014773358416846413

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 6.6046218043750655e-06

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.00018005004255871305

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.00010760397111963915

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 5.9337847786120445e-05

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 4.210016429335878e-07

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.00011011299356774825

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 3.4316384882787026e-05

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.00010411623051167003

Checking for matches...Platelets is Platelets
Platelets value is 7.035086907195455e-05

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.00012070361674913643

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 7.283173323778625e-05

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 6.421909021732613e-05

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.0001303614442638027

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 1.5683074962712685e-05

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 9.534977530274743e-05

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 9.417302016310787e-05

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 8.38978365248278e-05

Checking for matches...Iron is Iron
Iron value is 8.846587285094806e-05

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 3.884897127619816e-06

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.00012486536030582927

	Gender	Symptoms	Alcohol	Hepatitis B Surface Antigen	Hepatitis B e Antigen	Hepatitis B Core Antibody	Hepatitis C Virus Antibody	Cirrhosis	Endemic Countries	Smoking	...	Gamma glutamyl transferase (U/L)	Alkaline phosphatase (U/L)	Pro (
0	0.0	0.041250	0.009114	0.021932	0.000000	0.011204	0.028889	0.010850	0.018639	0.000000	...	0.030540	0.042692	0.0
1	0.0	0.005672	0.000000	0.000000	0.000520	0.000000	0.001227	0.000024	0.002159	0.000000	...	0.001006	0.005199	0.0
2	0.0	0.000000	0.000036	0.000037	0.000016	0.000000	0.000029	0.000021	0.000004	0.000043	...	0.000064	0.000130	0.0

3 rows x 49 columns

DT

DT0 In CV0...

Checking if correct model is loaded...
DecisionTreeClassifier(max_depth=17, min_samples_leaf=35, min_samples_split=45, random_state=42)

Checking explainer for DT0...
<shap.explainers._tree.Tree object at 0x7fecc720ae50>

Checking shap values for DT0...

```
[array([[0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        ...,
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.])), array([[0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        ...,
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.],
        [0., 0., 0., ..., 0., 0., 0.]])]
```

Generating SHAP plots for DT0...

Expected value for DT: [0.57272727 0.42727273]
Saving Bar Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Decision Plot for SHAP Values from Class 0 in Test Set...
Saving Decision Plot for SHAP Values from Class 1 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 0 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 1 in Test Set...
Saving feature importance ranking for DT0...

shap means for DT in CV0...

[0.	0.	0.	0.	0.45031405	0.
0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.
0.]				

Displaying Force Plot for DT SHAP Values from Class 0 in Test Set...

Out [13]:



Checking for matches...Symptoms is Symptoms
Symptoms value is 0.0

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.0

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.0

Checking for matches...Hepatitis B Core Antibody is Hepatitis B Core Antibody
Hepatitis B Core Antibody value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.0

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.0

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0

Checking for matches...Esophageal Varices is Esophageal Varices
Esophageal Varices value is 0.0

Checking for matches...Splenomegaly is Splenomegaly
Splenomegaly value is 0.0

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.0

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.0

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.0

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.0

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.0

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.0

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.0

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.4503140495867769

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.0

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.0

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.0

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.0

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.0

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.0

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.0

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.0

Checking for matches...Total Proteins (g/dL) is Total Proteins (g/dL)
Total Proteins (g/dL) value is 0.0

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.0

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.0

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.0

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.0

Checking for matches...Iron is Iron
Iron value is 0.0

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.0

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.0

DT1 In CV1...

Checking if correct model is loaded...
DecisionTreeClassifier(criterion='entropy', max_depth=30, min_samples_leaf=20,
min_samples_split=5, random_state=42, splitter='random')

Checking explainer for DT1...
<shap.explainers._tree.Tree object at 0x7fecc719de20>

Checking shap values for DT1...

```
[array([[ 0.          ,  0.08222611,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          , -0.02219251,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          , -0.02219251,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       ...,
       [ 0.          , -0.02219251,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          , -0.06287879,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          ,  0.02902098,  0.          , ...,  0.          ,
         0.          ,  0.          ]]), array([[ 0.          , -0.08222611,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          ,  0.02219251,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          ,  0.02219251,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       ...,
       [ 0.          ,  0.02219251,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          ,  0.06287879,  0.          , ...,  0.          ,
         0.          ,  0.          ],
       [ 0.          , -0.02902098,  0.          , ...,  0.          ,
         0.          ,  0.          ]]])]
```

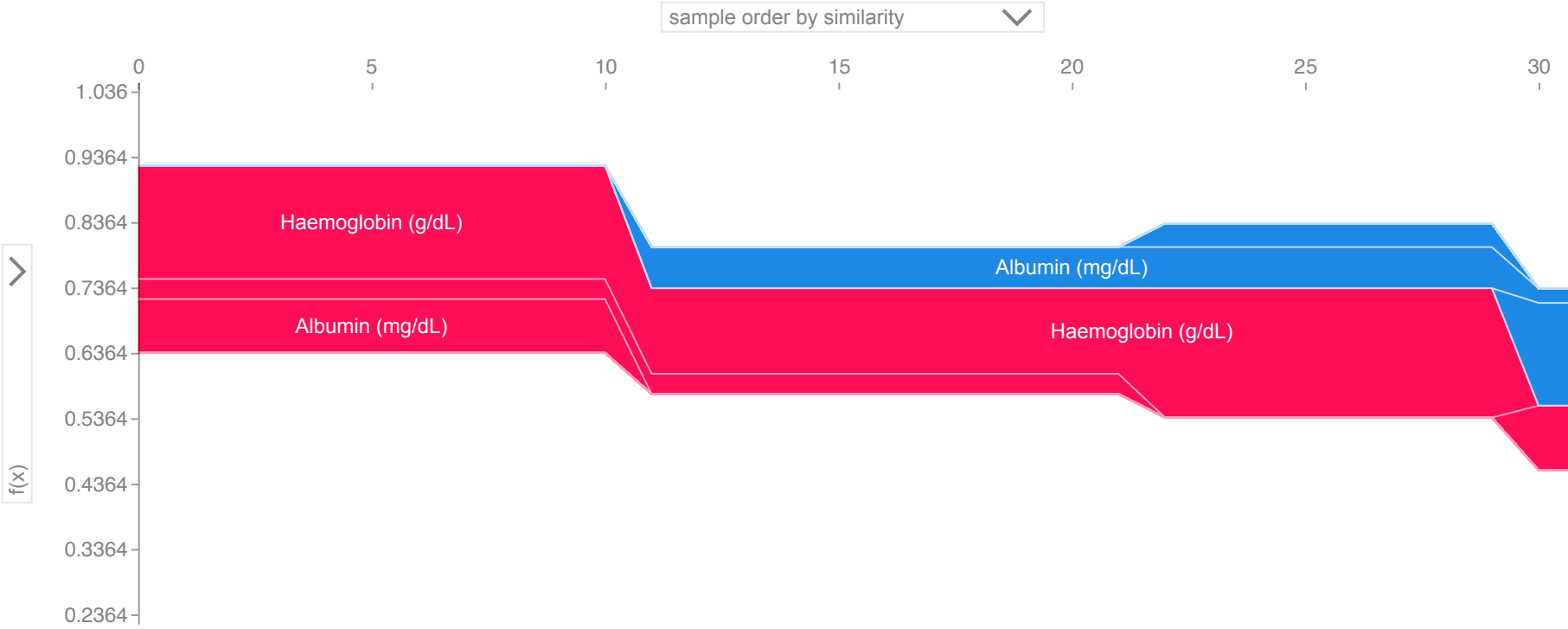
Generating SHAP plots for DT1...

Expected value for DT: [0.63636364 0.36363636]
Saving Bar Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Decision Plot for SHAP Values from Class 0 in Test Set...
Saving Decision Plot for SHAP Values from Class 1 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 0 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 1 in Test Set...
Saving feature importance ranking for DT1...

shap means for DT in CV1...
[0. 0.0970056 0. 0. 0. 0.12790833
 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0.34071246 0. 0.
 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0.]

Displaying Force Plot for DT SHAP Values from Class 0 in Test Set...

Out [13]:



Checking for matches...Symptoms is Symptoms
Symptoms value is 0.0

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.0

Checking for matches...Obesity is Obesity
Obesity value is 0.0

Checking for matches...Hemochromatosis is Hemochromatosis
Hemochromatosis value is 0.0

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.0

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.0

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.0

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.0

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.0

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.0

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.0

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.12790833233520985

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.0

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.0

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.3407124560577246

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.0

Checking for matches...Platelets is Platelets
Platelets value is 0.0

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.09700559689864502

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.0

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.0

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.0

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.0

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.0

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.0

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.0

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.0

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.0

Checking for matches...Iron is Iron
Iron value is 0.0

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.0

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.0

DT2 In CV2...

Checking if correct model is loaded...
DecisionTreeClassifier(class_weight='balanced', max_depth=29,
min_samples_leaf=30, min_samples_split=45,
random_state=42)

Checking explainer for DT2...
<shap.explainers._tree.Tree object at 0x7fecc7195790>

Checking shap values for DT2...

```
[array([[ -0.14385676,  0.          ,  0.20558515, ...,  0.          ,
         0.          ,  0.          ],
       [ -0.0462963 ,  0.          , -0.18566247, ...,  0.          ,
         0.          ,  0.          ],
       [ -0.0462963 ,  0.          , -0.18566247, ...,  0.          ,
         0.          ,  0.          ],
       ...,
       [ -0.0462963 ,  0.          , -0.18566247, ...,  0.          ,
         0.          ,  0.          ],
       [  0.05769231,  0.          , -0.28965107, ...,  0.          ,
         0.          ,  0.          ],
       [ -0.14385676,  0.          ,  0.20558515, ...,  0.          ,
         0.          ,  0.          ]]), array([[ 0.14385676,  0.          , -0.20558515, ...,  0.          ,
         0.          ,  0.          ],
       [  0.0462963 ,  0.          ,  0.18566247, ...,  0.          ,
         0.          ,  0.          ],
       [  0.0462963 ,  0.          ,  0.18566247, ...,  0.          ,
         0.          ,  0.          ],
       ...,
       [  0.0462963 ,  0.          ,  0.18566247, ...,  0.          ,
         0.          ,  0.          ],
       [-0.05769231,  0.          ,  0.28965107, ...,  0.          ,
         0.          ,  0.          ],
       [  0.14385676,  0.          , -0.20558515, ...,  0.          ,
         0.          ,  0.          ]]])]
```

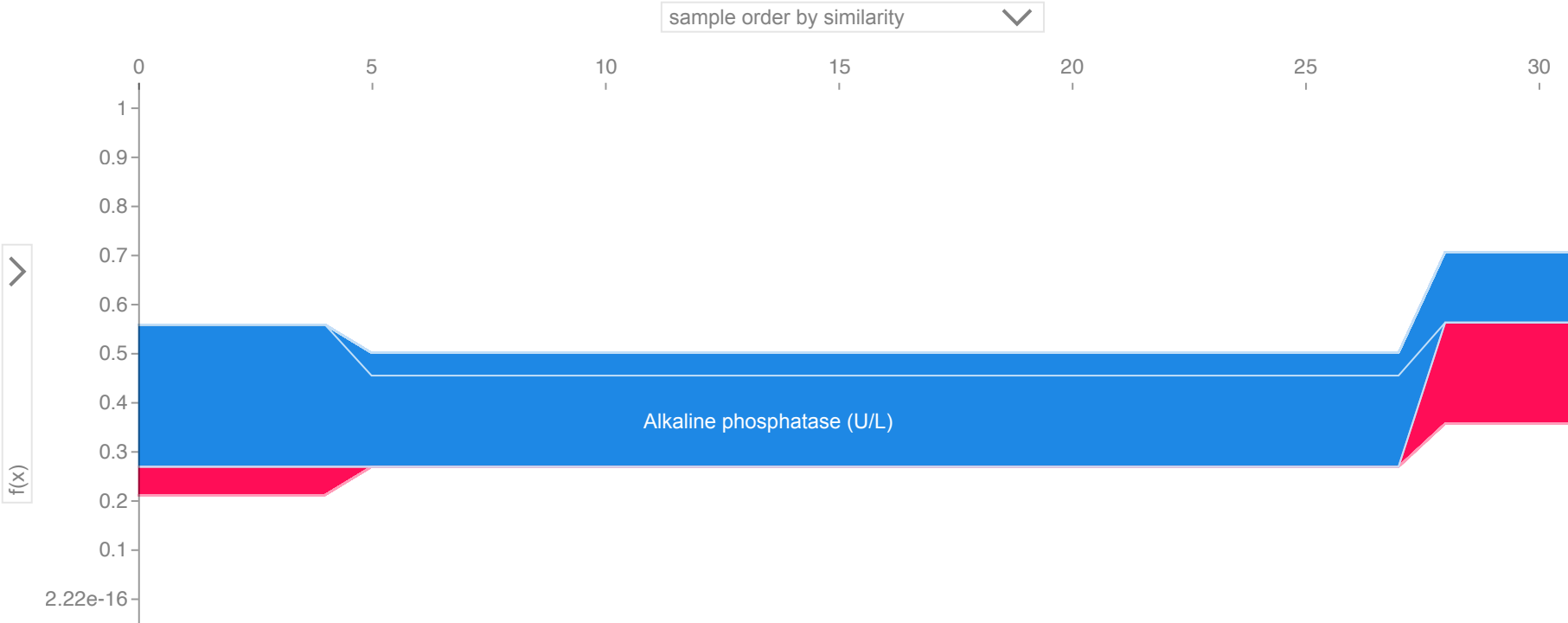
Generating SHAP plots for DT2...

Expected value for DT: [0.5 0.5]
Saving Bar Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Decision Plot for SHAP Values from Class 0 in Test Set...
Saving Decision Plot for SHAP Values from Class 1 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 0 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 1 in Test Set...
Saving feature importance ranking for DT2...

shap means for DT in CV2...
[0.20461559 0. 0.45585129 0. 0. 0.]
0. 0. 0. 0. 0. 0.]
0. 0. 0. 0. 0. 0.]
0. 0. 0. 0. 0. 0.]
0. 0. 0. 0. 0. 0.]
0. 0. 0. 0. 0. 0.]

Displaying Force Plot for DT SHAP Values from Class 0 in Test Set...

Out [13]:



8/31/22, 2:58 PM

SHAP_final

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.0

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.0

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Smoking is Smoking
Smoking value is 0.0

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.0

Checking for matches...Hemochromatosis is Hemochromatosis
Hemochromatosis value is 0.0

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.0

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.0

Checking for matches...Human Immunodeficiency Virus is Human Immunodeficiency Virus
Human Immunodeficiency Virus value is 0.0

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.0

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.0

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.0

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.0

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.0

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.0

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.0

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.0

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.0

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.0

Checking for matches...Platelets is Platelets
Platelets value is 0.0

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.20461558566581417

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.0

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.0

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.45585128561873833

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.0

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.0

localhost:8888/nbconvert/html/streamline/SHAP_final.ipynb?download=false

75/98

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.0

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.0

Checking for matches...Iron is Iron
Iron value is 0.0

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.0

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.0

	Gender	Symptoms	Alcohol	Hepatitis B Surface Antigen	Hepatitis B e Antigen	Hepatitis B Core Antibody	Hepatitis C Virus Antibody	Cirrhosis	Endemic Countries	Smoking	...	Gamma glutamyl transferase (U/L)	Alkaline phosphatase (U/L)	Tc Prote (g/dl)
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.000000	
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.455851	

3 rows x 49 columns

```
-----
RF
-----
RF0 In CV0...

Checking if correct model is loaded...
RandomForestClassifier(max_depth=9, max_features=None, min_samples_leaf=9,
                        min_samples_split=24, n_estimators=935, random_state=42)

Checking explainer for RF0...
<shap.explainers._tree.Tree object at 0x7fecb32f7e50>

Checking shap values for RF0...
```

```
[array([[ -8.04892952e-04,  3.49123556e-02, -7.39028100e-04, ...,
        -2.55865722e-05, -4.80373900e-04, -5.51427280e-03],
       [ 8.98581287e-04, -2.36897659e-02, -7.01881775e-05, ...,
        -4.69087156e-05, -1.53604128e-03, -9.93491258e-03],
       [ 1.09297656e-04, -2.10001659e-02,  2.39694450e-04, ...,
        -4.18242045e-05, -3.61466526e-04, -6.98520090e-03],
       ...,
       [ 3.18715672e-04,  2.90540264e-02,  2.67651465e-04, ...,
        -4.18242045e-05,  7.09998438e-05,  7.60365635e-03],
       [-1.87966055e-03, -1.58333798e-02,  2.67651465e-04, ...,
        3.18309142e-04,  1.93293842e-03, -7.31627298e-04],
       [-1.82586614e-03, -2.32531905e-02,  1.25596859e-04, ...,
        3.18309142e-04, -1.88973059e-03, -1.80372814e-03]])], array([[ 8.04892952e-04, -3.49123556e-02,  7.39028100e-0
4, ...,
        2.55865722e-05,  4.80373900e-04,  5.51427280e-03],
       [-8.98581287e-04,  2.36897659e-02,  7.01881775e-05, ...,
        4.69087156e-05,  1.53604128e-03,  9.93491258e-03],
       [-1.09297656e-04,  2.10001659e-02, -2.39694450e-04, ...,
        4.18242045e-05,  3.61466526e-04,  6.98520090e-03],
       ...,
       [-3.18715672e-04, -2.90540264e-02, -2.67651465e-04, ...,
        4.18242045e-05, -7.09998438e-05, -7.60365635e-03],
       [ 1.87966055e-03,  1.58333798e-02, -2.67651465e-04, ...,
        -3.18309142e-04, -1.93293842e-03,  7.31627298e-04],
       [ 1.82586614e-03,  2.32531905e-02, -1.25596859e-04, ...,
        -3.18309142e-04,  1.88973059e-03,  1.80372814e-03]])])

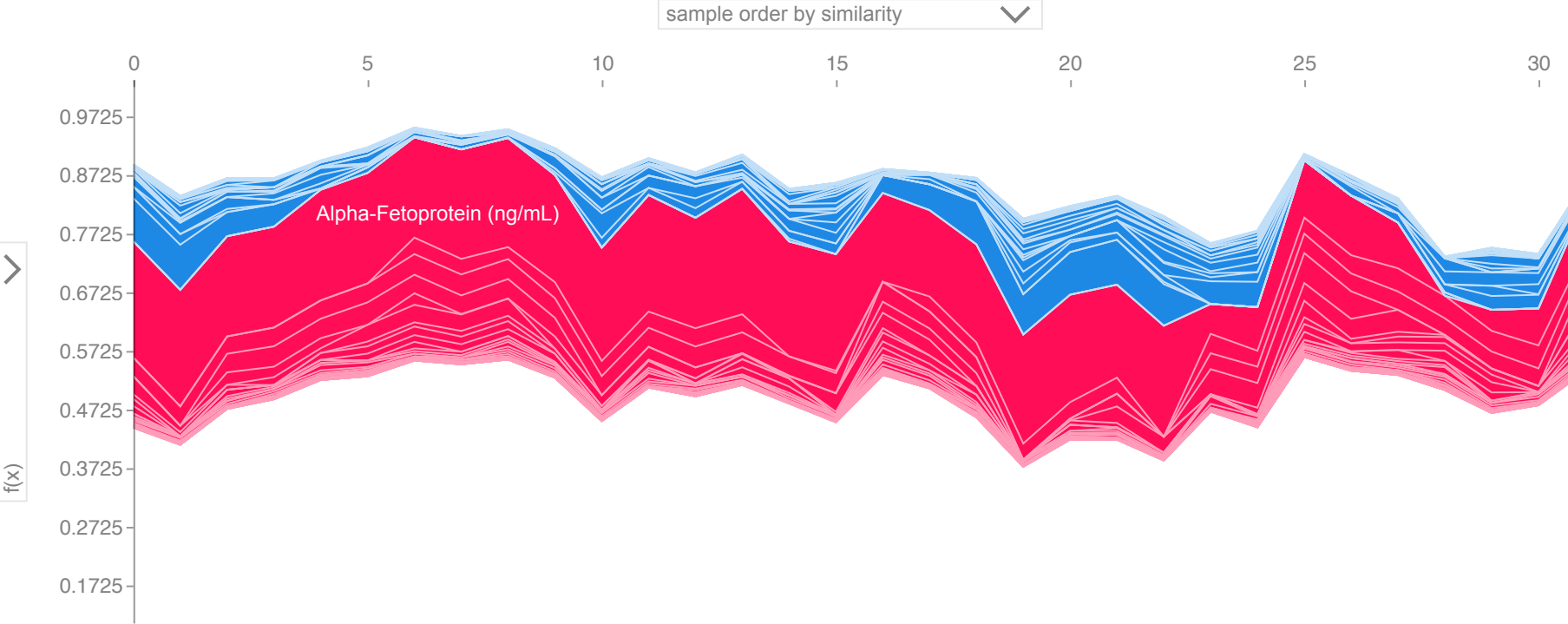
Generating SHAP plots for RF0...
```

```
Expected value for RF: [0.57252309 0.42747691]
Saving Bar Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Decision Plot for SHAP Values from Class 0 in Test Set...
Saving Decision Plot for SHAP Values from Class 1 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 0 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 1 in Test Set...
Saving feature importance ranking for RF0...
```

```
shap means for RF in CV0...
[3.19942297e-03 5.67437669e-02 4.84800100e-04 3.69305780e-02
 2.98213726e-01 2.06980669e-03 3.76885936e-02 1.95240541e-04
 0.00000000e+00 2.22415035e-02 2.95175702e-03 1.38474826e-02
 0.00000000e+00 0.00000000e+00 1.88034845e-03 6.30126715e-02
 7.04483607e-03 2.14466544e-02 9.22727603e-04 0.00000000e+00
 3.89407922e-04 1.04031699e-02 6.60506693e-02 6.24117169e-03
 8.50247591e-03 2.44389504e-02 4.76767927e-03 0.00000000e+00
 1.18700021e-02 8.76713050e-03 4.19877143e-03 2.73102701e-02
 5.51265543e-04 1.72420103e-04 2.43806037e-04 2.04004893e-03
 1.37081846e-02]
```

```
Displaying Force Plot for RF SHAP Values from Class 0 in Test Set...
```

Out[13]:



Checking for matches...Symptoms is Symptoms
Symptoms value is 0.00024380603660597525

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.0004848000995061814

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.0

Checking for matches...Hepatitis B Core Antibody is Hepatitis B Core Antibody
Hepatitis B Core Antibody value is 0.0009227276031035558

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0003894079218432087

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.002951757024125973

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.0001952405412753242

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0

Checking for matches...Esophageal Varices is Esophageal Varices
Esophageal Varices value is 0.0018803484504917316

Checking for matches...Splenomegaly is Splenomegaly
Splenomegaly value is 0.00017242010265236288

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.0005512655427968731

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.008502475906368403

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.00419877142554893

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.02731027009637193

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.0

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.002069806690269925

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.010403169863992902

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.2982137257678314

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.02144665440006814

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0047676792694041105

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.006241171687849642

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.05674376690998474

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.0020400489295341615

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.0031994229675763068

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.037688593636958326

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.0070448360739481875

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.03693057797462684

Checking for matches...Total Proteins (g/dL) is Total Proteins (g/dL)
Total Proteins (g/dL) value is 0.013708184622317671

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.022241503485770874

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.011870002063751826

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.02443895039805761

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.013847482608104967

Checking for matches...Iron is Iron
Iron value is 0.06605066925780469

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.008767130495977389

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.06301267151123399

RF1 In CV1...

Checking if correct model is loaded...
RandomForestClassifier(class_weight='balanced', max_depth=25,
 max_features='log2', min_samples_leaf=16,
 min_samples_split=45, n_estimators=194, oob_score=True,
 random_state=42)

Checking explainer for RF1...
<shap.explainers._tree.Tree object at 0x7fecb27dc970>

Checking shap values for RF1...

```
[array([[ 6.89282193e-04,  5.16384335e-02, -1.76947858e-02, ...,
         4.15186605e-04, -5.22220942e-03,  1.52801529e-02],
       [-2.86516996e-03, -1.75170521e-03,  1.53660207e-02, ...,
         1.57204645e-04,  1.14384010e-02,  1.36293862e-02],
       [ 9.58343029e-04, -3.08481544e-02, -1.63920514e-02, ...,
         1.57204645e-04,  1.17329978e-02, -6.96607661e-03],
       ...,
       [-4.62392460e-05,  2.51854494e-02, -1.70193937e-02, ...,
         1.57204645e-04,  1.17917953e-02,  1.21731827e-02],
       [ 8.00804127e-04, -3.08295224e-02, -1.58989779e-02, ...,
        -1.04418188e-03, -5.37882055e-03, -6.99630666e-04],
       [-1.58770860e-03,  5.12094387e-02, -9.49257495e-03, ...,
         1.57204645e-04, -5.09113830e-03,  1.43028838e-02]])], array([[ -6.89282193e-04, -5.16384335e-02,  1.76947858e-0
2, ...,
        -4.15186605e-04,  5.22220942e-03, -1.52801529e-02],
       [ 2.86516996e-03,  1.75170521e-03, -1.53660207e-02, ...,
        -1.57204645e-04, -1.14384010e-02, -1.36293862e-02],
       [-9.58343029e-04,  3.08481544e-02,  1.63920514e-02, ...,
        -1.57204645e-04, -1.17329978e-02,  6.96607661e-03],
       ...,
       [ 4.62392460e-05, -2.51854494e-02,  1.70193937e-02, ...,
        -1.57204645e-04, -1.17917953e-02, -1.21731827e-02],
       [-8.00804127e-04,  3.08295224e-02,  1.58989779e-02, ...,
         1.04418188e-03,  5.37882055e-03,  6.99630666e-04],
       [ 1.58770860e-03, -5.12094387e-02,  9.49257495e-03, ...,
        -1.57204645e-04,  5.09113830e-03, -1.43028838e-02]])])
```

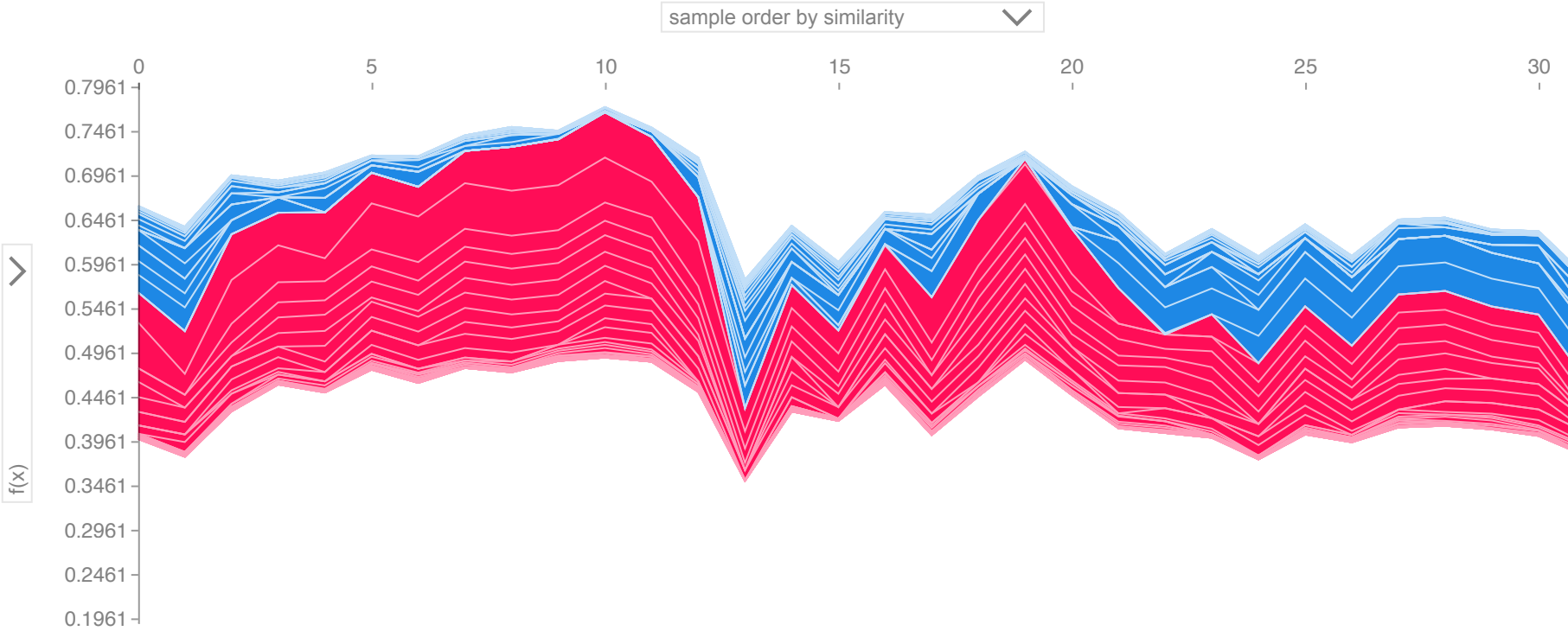
Generating SHAP plots for RF1...

Expected value for RF: [0.49611628 0.50388372]
Saving Bar Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Decision Plot for SHAP Values from Class 0 in Test Set...
Saving Decision Plot for SHAP Values from Class 1 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 0 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 1 in Test Set...
Saving feature importance ranking for RF1...

shap means for RF in CV1...
[0.00264805 0.06822466 0.02612097 0.03263829 0. 0.03433725
 0.00539887 0. 0.00051941 0.00117829 0.02959989 0.
 0. 0.00622004 0.00443315 0.06544777 0. 0.
 0. 0.01709888 0.04313898 0.00236821 0. 0.01021446
 0.00709165 0. 0.00395138 0. 0.02701932 0.00185567
 0.03220521 0.00290079 0.00128005 0.00072348 0.01524718 0.02758143]

Displaying Force Plot for RF SHAP Values from Class 0 in Test Set...

Out [13]:



8/31/22, 2:58 PM

SHAP_final

Checking for matches...Symptoms is Symptoms
Symptoms value is 0.015247176712858846

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.001178286135152906

Checking for matches...Obesity is Obesity
Obesity value is 0.0

Checking for matches...Hemochromatosis is Hemochromatosis
Hemochromatosis value is 0.0

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.0

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.001280045196774621

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.000723483837280238

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.0

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.0018556665211397474

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.032205205095110115

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.0

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.03433725081841752

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.017098883790089066

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.03263828816439576

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.06544777013559176

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.007091652485888949

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.0023682100400009556

Checking for matches...Platelets is Platelets
Platelets value is 0.002900789989556397

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.06822466197111779

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.027581431989709027

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.002648053115764084

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.005398873640950709

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.004433145676385586

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.026120973930810663

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.0005194082445076238

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.003951375540850356

localhost:8888/nbconvert/html/streamline/SHAP_final.ipynb?download=false

81/98

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.010214462581339564

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.02959988900568732

Checking for matches...Iron is Iron
Iron value is 0.04313898091479226

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.02701931763240256

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.0062200362640664995

RF2 In CV2...

Checking if correct model is loaded...
RandomForestClassifier(class_weight='balanced', max_depth=16, max_features=None,
min_samples_leaf=20, min_samples_split=40,
n_estimators=299, oob_score=True, random_state=42)

Checking explainer for RF2...
<shap.explainers._tree.Tree object at 0x7fecc7e52af0>

Checking shap values for RF2...

```
[array([[ -1.64930049e-02,  0.00000000e+00,  9.65987236e-02, ...,
         0.00000000e+00,  0.00000000e+00, -2.92612124e-04],
        [ -1.10170786e-02,  0.00000000e+00, -1.61874431e-01, ...,
         0.00000000e+00,  0.00000000e+00,  4.35547992e-04],
        [ -1.17523110e-02,  0.00000000e+00, -1.57953243e-01, ...,
         0.00000000e+00,  0.00000000e+00, -2.92612124e-04],
        ...,
        [ -1.24810033e-02,  0.00000000e+00, -1.55756997e-01, ...,
         0.00000000e+00,  0.00000000e+00,  4.35547992e-04],
        [  2.57852085e-02,  0.00000000e+00, -8.08842985e-02, ...,
         0.00000000e+00,  0.00000000e+00,  4.35547992e-04],
        [ -1.64930049e-02,  0.00000000e+00,  1.95650937e-01, ...,
         0.00000000e+00,  0.00000000e+00, -1.15877556e-04]]), array([[ 1.64930049e-02,  0.00000000e+00, -9.65987236e-0
2, ...,
         0.00000000e+00,  0.00000000e+00,  2.92612124e-04],
        [  1.10170786e-02,  0.00000000e+00,  1.61874431e-01, ...,
         0.00000000e+00,  0.00000000e+00, -4.35547992e-04],
        [  1.17523110e-02,  0.00000000e+00,  1.57953243e-01, ...,
         0.00000000e+00,  0.00000000e+00,  2.92612124e-04],
        ...,
        [  1.24810033e-02,  0.00000000e+00,  1.55756997e-01, ...,
         0.00000000e+00,  0.00000000e+00, -4.35547992e-04],
        [ -2.57852085e-02,  0.00000000e+00,  8.08842985e-02, ...,
         0.00000000e+00,  0.00000000e+00, -4.35547992e-04],
        [  1.64930049e-02,  0.00000000e+00, -1.95650937e-01, ...,
         0.00000000e+00,  0.00000000e+00,  1.15877556e-04]])]
```

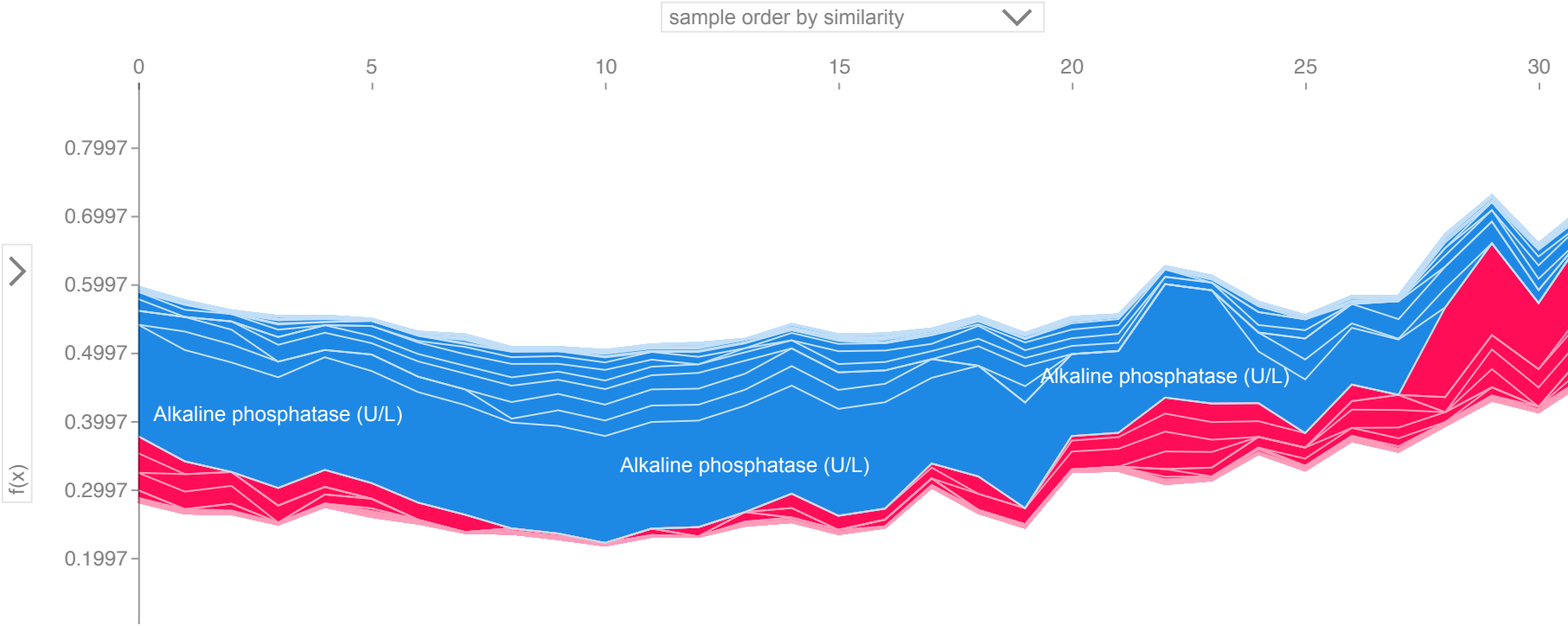
Generating SHAP plots for RF2...

Expected value for RF: [0.49969365 0.50030635]
Saving Bar Summary Plot for SHAP Values in Class 0 & 1 in Test Set...
Saving Decision Plot for SHAP Values from Class 0 in Test Set...
Saving Decision Plot for SHAP Values from Class 1 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 0 in Test Set...
Saving Waterfall Plot for SHAP Values from Class 1 in Test Set...
Saving feature importance ranking for RF2...

shap means for RF in CV2...
[0.03408604 0. 0.31589118 0.04625691 0.00087878 0.
0.00323506 0. 0. 0.00274943 0.00068531 0.00081468
0. 0.04803709 0.00389811 0.03569443 0. 0.
0. 0. 0. 0. 0.04638195 0.02383206
0. 0.00562718 0.00082995 0. 0.00143203 0.00309383
0.00178158 0.01819939 0.00481963 0. 0. 0.00051831]

Displaying Force Plot for RF SHAP Values from Class 0 in Test Set...

Out [13]:



Checking for matches...Alcohol is Alcohol
Alcohol value is 0.0

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.0

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Smoking is Smoking
Smoking value is 0.0005183068903087402

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.0006853094103879076

Checking for matches...Hemochromatosis is Hemochromatosis
Hemochromatosis value is 0.0

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.0008787778330564218

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.0

Checking for matches...Human Immunodeficiency Virus is Human Immunodeficiency Virus
Human Immunodeficiency Virus value is 0.0

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.0

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.0

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.0

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.0017815839983008862

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.018199389907456946

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.0

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.0

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.04625690721396444

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.03569442705626348

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0008299485092251701

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.023832055334701036

Checking for matches...Platelets is Platelets
Platelets value is 0.004819628863925804

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.03408603703114281

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.0032350607592149666

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.003898109011038946

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.3158911806696007

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.0027494323399706814

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.0014320261570804037

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.005627180856920699

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.0008146839117934954

Checking for matches...Iron is Iron
Iron value is 0.04638194744554026

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.0030938281755690323

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.048037094761025974

	Gender	Symptoms	Alcohol	Hepatitis B Surface Antigen	Hepatitis B e Antigen	Hepatitis B Core Antibody	Hepatitis C Virus Antibody	Cirrhosis	Endemic Countries	Smoking	...	Gamma glutamyl transferase (U/L)	Alkaline phosphatase (U/L)	Prothrombin
0	0.0	0.000244	0.000485	0.0	0.0	0.000923	0.000389	0.0	0.0	0.000000	...	0.007045	0.036931	0.000000
1	0.0	0.015247	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	...	0.004433	0.026121	0.000000
2	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000518	...	0.003898	0.315891	0.000000

3 rows x 49 columns

XGB

XGB0 In CV0...

Checking if correct model is loaded...
XGBClassifier(alpha=0.0003085901759707382, base_score=0.5, booster='gbtree',
 callbacks=None, colsample_bylevel=1, colsample_bynode=1,
 colsample_bytree=0.31595586732894876, early_stopping_rounds=None,
 enable_categorical=False, eta=0.0016131413768891527,
 eval_metric=None, gamma=1.086786493948363e-07, gpu_id=-1,
 grow_policy='lossguide', importance_type=None,
 interaction_constraints='', learning_rate=0.00161314139,
 max_bin=256, max_cat_to_onehot=4, max_delta_step=0, max_depth=3,
 max_leaves=0, min_child_weight=9.912142174935715,
 min_samples_leaf=32, min_samples_split=43, missing=nan,
 monotone_constraints=()), n_estimators=305, n_jobs=1, nthread=1, ...)

Checking explainer for XGB0...
<shap.explainers._tree.Tree object at 0x7fecb32f8820>

Checking shap values for XGB0...

```
[[ 0.00033414 -0.00198181  0.          ...  0.          0.
  0.00147086]
 [-0.0001525  0.00104749  0.          ...  0.          0.
  0.00147086]
 [ 0.00033414 -0.00198181  0.          ...  0.          0.
  0.00174883]
 ...
 [ 0.00013163 -0.00198181  0.          ...  0.          0.
 -0.00234208]
 [-0.0001525  -0.00198181  0.          ...  0.          0.
  0.00110088]
 [ 0.0001405   0.00104749  0.          ...  0.          0.
  0.00082496]]
```

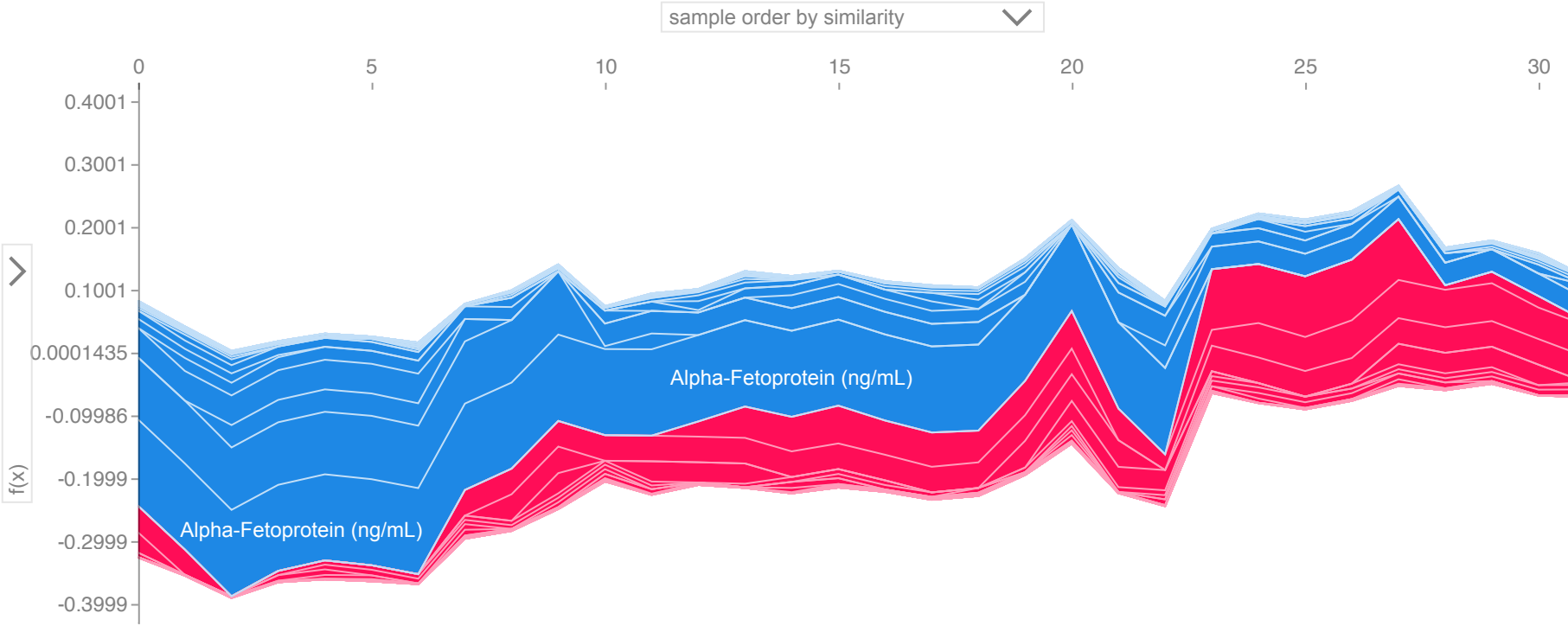
Generating SHAP plots for XGB0...

Expected value for XGB: 0.00014352845028042793
Saving Summary Plot for SHAP Values in Test Set...
Saving SHAP Bar Plot for SHAP Values Test Set...
Saving SHAP Decision Plot for SHAP Values in Test Set...
Saving SHAP Decision Plot for Single-Prediction in Test Set...
Saving Waterfall Plot for SHAP Values for a Single-Prediction in Test Set...
Saving feature importance ranking for XGB0...

shap means for XGB in CV0...
[0.00017262 0.00154013 0. 0.06297818 0.09290499 0.
 0.01105899 0. 0. 0.00785514 0. 0.0015779
 0. 0. 0. 0.00434002 0.00566131 0.01006154
 0. 0. 0. 0.00033577 0.03775872 0.0008829
 0. 0.0248299 0. 0. 0. 0.00037678
 0.00134643 0.04380364 0. 0. 0. 0.
 0.00179555]

Displaying Force Plot for XGB SHAP Values in Whole Test Set...

Out [13]:



8/31/22, 2:58 PM

SHAP_final

Checking for matches...Symptoms is Symptoms
Symptoms value is 0.0

Checking for matches...Alcohol is Alcohol
Alcohol value is 0.0

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.0

Checking for matches...Hepatitis B Core Antibody is Hepatitis B Core Antibody
Hepatitis B Core Antibody value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.0

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.0

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0

Checking for matches...Esophageal Varices is Esophageal Varices
Esophageal Varices value is 0.0

Checking for matches...Splenomegaly is Splenomegaly
Splenomegaly value is 0.0

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.0

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.0

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.0013464291114360094

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.04380363970994949

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.0

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.0

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.0003357660898473114

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.09290499240159988

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.010061542503535748

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.0008829032303765416

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.0015401346608996391

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.0

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.0001726230839267373

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.011058987118303776

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.00566130643710494

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.06297817826271057

Checking for matches...Total Proteins (g/dL) is Total Proteins (g/dL)
Total Proteins (g/dL) value is 0.0017955484800040722

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.007855135947465897

localhost:8888/nbconvert/html/streamline/SHAP_final.ipynb?download=false

87/98

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.0

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.024829896166920662

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.001577903749421239

Checking for matches...Iron is Iron
Iron value is 0.0377587229013443

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.0003767825255636126

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.004340018145740032

XGB1 In CV1...

Checking if correct model is loaded...
XGBClassifier(alpha=0.2551087479356069, base_score=0.5, booster='gbtree',
 callbacks=None, colsample_bylevel=1, colsample_bynode=1,
 colsample_bytree=0.7752104293797693, early_stopping_rounds=None,
 enable_categorical=False, eta=0.01045236594154273,
 eval_metric=None, gamma=1.2055493072793457e-05, gpu_id=-1,
 grow_policy='depthwise', importance_type=None,
 interaction_constraints='', learning_rate=0.0104523655,
 max_bin=256, max_cat_to_onehot=4, max_delta_step=0, max_depth=11,
 max_leaves=0, min_child_weight=1.1403861951454617,
 min_samples_leaf=23, min_samples_split=24, missing=nan,
 monotone_constraints='()', n_estimators=372, n_jobs=1, nthread=1, ...)

Checking explainer for XGB1...
<shap.explainers._tree.Tree object at 0x7fecb331d9a0>

Checking shap values for XGB1...

```
[[[-4.1213792e-02 -3.3411807e-01  3.1376600e-01 ... -3.7167885e-03
   5.5054866e-02 -5.6627635e-02]
 [ 8.7244794e-02  3.1003144e-03 -1.0110403e-01 ... -8.8498723e-03
 -2.7989864e-01 -7.0158280e-02]
 [-4.6925444e-02  1.0683852e-01  3.3140031e-01 ... -7.7774939e-03
 -2.4780509e-01 -1.7681159e-04]
 ...
 [-1.6688924e-02 -1.4458889e-01  2.3780176e-01 ... -4.2031193e-03
 -2.1513927e-01 -1.9822428e-02]
 [-1.9945810e-02  1.5445979e-01  2.7486852e-01 ...  8.2303239e-03
  7.4652217e-02 -1.7326010e-02]
 [-3.5156224e-02 -2.9176781e-01  7.5178228e-02 ... -5.8821207e-03
  5.9577417e-02 -8.8170484e-02]]]
```

Generating SHAP plots for XGB1...

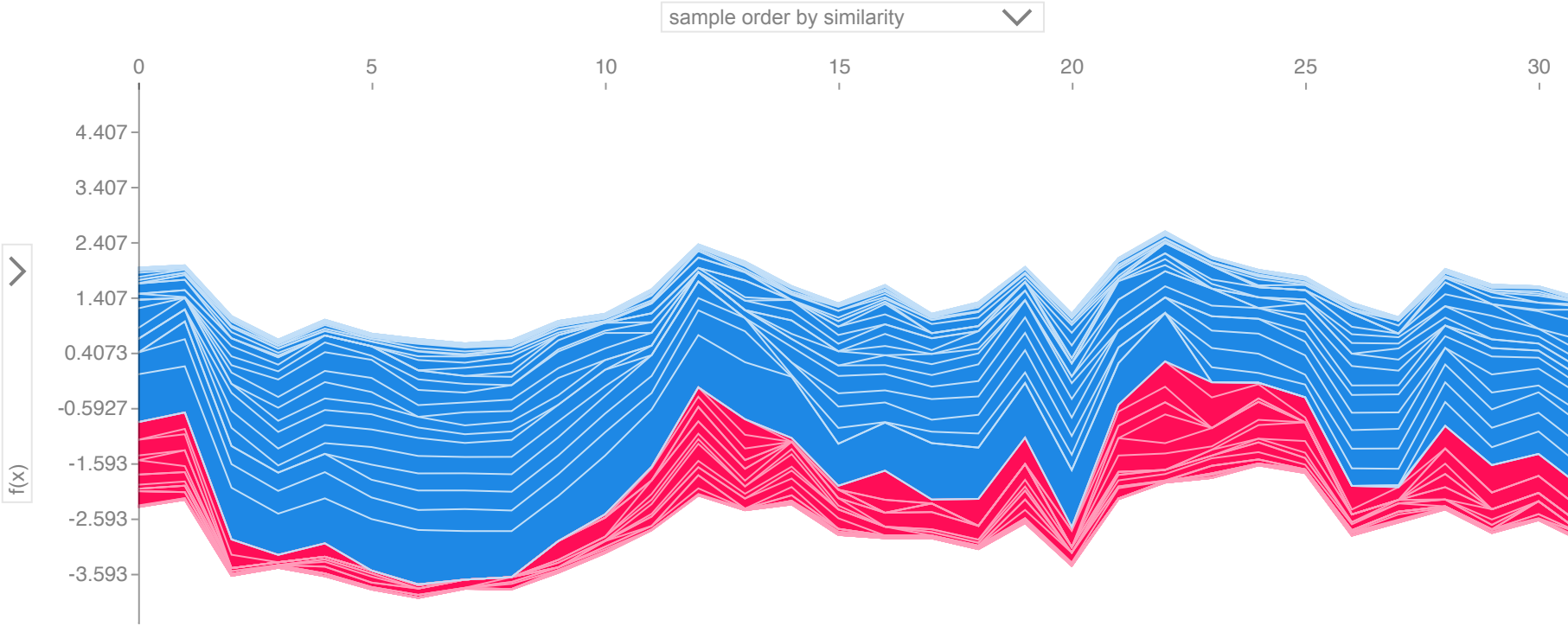
Expected value for XGB: 0.40734654664993286
Saving Summary Plot for SHAP Values in Test Set...
Saving SHAP Bar Plot for SHAP Values Test Set...
Saving SHAP Decision Plot for SHAP Values in Test Set...
Saving SHAP Decision Plot for Single-Prediction in Test Set...
Saving Waterfall Plot for SHAP Values for a Single-Prediction in Test Set...
Saving feature importance ranking for XGB1...

shap means for XGB in CV1...

```
[0.04532614 0.18305469 0.19840291 0.32052973 0.00730014 0.18359938
 0.1174619  0.          0.03076757 0.02206666 0.25751933 0.01147493
 0.          0.16940679 0.03576979 0.5790477  0.          0.
 0.02004449 0.22447011 0.2718499  0.12752493 0.0020189  0.0934651
 0.22346753 0.          0.03206018 0.00267068 0.30437937 0.04777434
 0.3019611  0.12592266 0.00799157 0.00673484 0.13446881 0.0617146 ]
```

Displaying Force Plot for XGB SHAP Values in Whole Test Set...

Out [13]:



Checking for matches...Symptoms is Symptoms
Symptoms value is 0.13446880877017975

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.020044488832354546

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.022066660225391388

Checking for matches...Obesity is Obesity
Obesity value is 0.0026706834323704243

Checking for matches...Hemochromatosis is Hemochromatosis
Hemochromatosis value is 0.0

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.007300143130123615

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.007991569116711617

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.00673484243452549

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.0020188980270177126

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.047774337232112885

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.30196109414100647

Checking for matches...Encephalopathy degree* is Encephalopathy degree*
Encephalopathy degree* value is 0.011474931612610817

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.18359938263893127

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.2244701087474823

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.3205297291278839

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.579047679901123

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.22346752882003784

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.12752492725849152

Checking for matches...Platelets is Platelets
Platelets value is 0.12592266499996185

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.18305468559265137

Checking for matches...Total Bilirubin(mg/dL) is Total Bilirubin(mg/dL)
Total Bilirubin(mg/dL) value is 0.06171460077166557

Checking for matches...Alanine transaminase (U/L) is Alanine transaminase (U/L)
Alanine transaminase (U/L) value is 0.04532614350318909

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.1174619048833847

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.03576979041099548

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.19840291142463684

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.030767574906349182

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.032060179859399796

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.09346509724855423

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.25751933455467224

Checking for matches...Iron is Iron
Iron value is 0.2718499004840851

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.3043793737888336

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.16940678656101227

XGB2 In CV2...

Checking if correct model is loaded...
XGBClassifier(alpha=2.925626174679448e-07, base_score=0.5, booster='gbtree',
 callbacks=None, colsample_bylevel=1, colsample_bynode=1,
 colsample_bytree=0.10192679583709846, early_stopping_rounds=None,
 enable_categorical=False, eta=0.0002501390367621765,
 eval_metric=None, gamma=4.155123430773271e-05, gpu_id=-1,
 grow_policy='depthwise', importance_type=None,
 interaction_constraints='', learning_rate=0.000250139041,
 max_bin=256, max_cat_to_onehot=4, max_delta_step=0, max_depth=2,
 max_leaves=0, min_child_weight=6.224435947630516,
 min_samples_leaf=10, min_samples_split=45, missing=nan,
 monotone_constraints='()', n_estimators=956, n_jobs=1, nthread=1, ...)

Checking explainer for XGB2...
<shap.explainers._tree.Tree object at 0x7fecc7450700>

Checking shap values for XGB2...

```
[[ 5.08398050e-03  1.09896610e-05 -1.44367898e-02 ... -2.88566196e-04
 -5.45803457e-04  5.12634986e-04]
 [ 4.83505754e-03  2.37361382e-06  1.88300461e-02 ...  5.14981046e-04
 -4.53948451e-04 -9.65057523e-04]
 [ 4.78631770e-03 -3.62699211e-05  1.88307650e-02 ... -2.88566196e-04
 -5.20780042e-04  5.26774325e-04]
 ...
 [ 4.53065475e-03  2.37361382e-06  1.87011994e-02 ... -2.88566196e-04
 -4.90862702e-04 -9.11067938e-04]
 [-1.01444125e-02  2.37361382e-06  9.76637285e-03 ... -2.88566196e-04
 -4.72357147e-04 -8.52823956e-04]
 [ 4.44364175e-03  1.09896610e-05 -2.03168858e-02 ... -2.76739825e-04
 -5.45803457e-04  5.12634986e-04]]
```

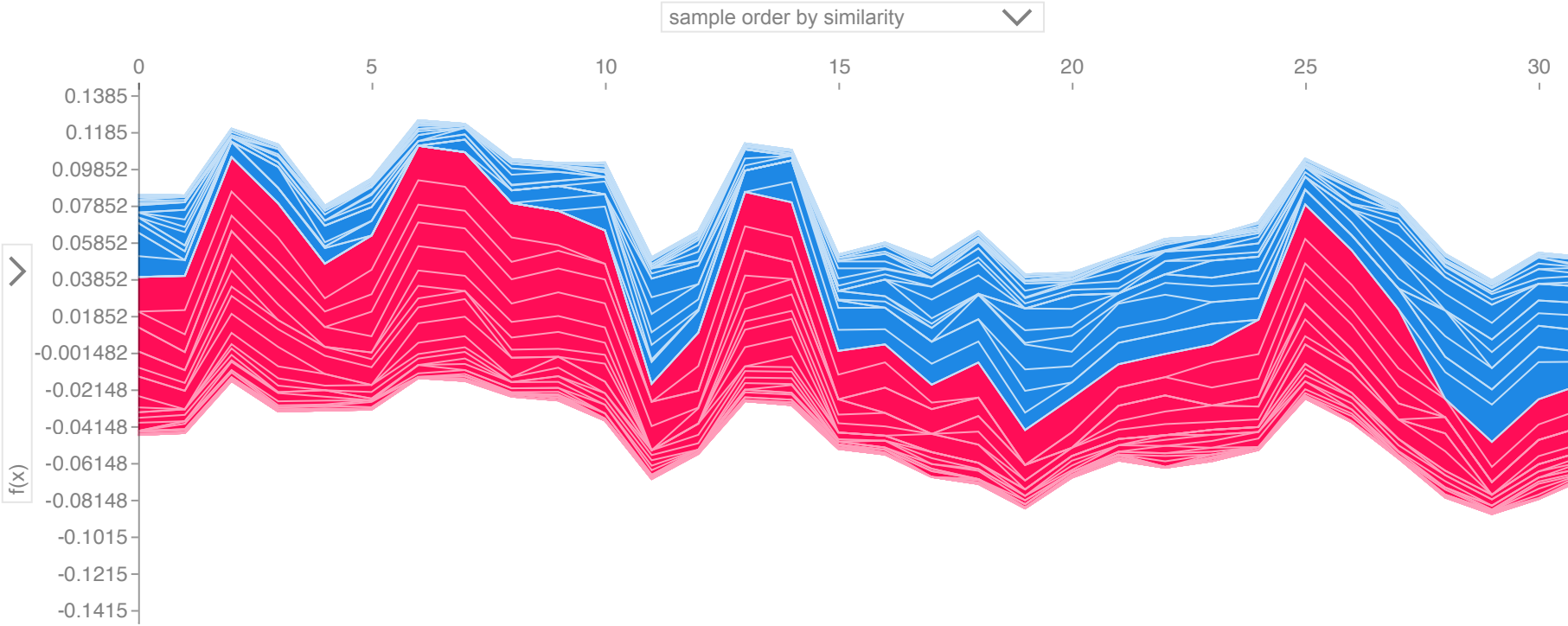
Generating SHAP plots for XGB2...

Expected value for XGB: -0.0014823254896327853
Saving Summary Plot for SHAP Values in Test Set...
Saving SHAP Bar Plot for SHAP Values Test Set...
Saving SHAP Decision Plot for SHAP Values in Test Set...
Saving SHAP Decision Plot for Single-Prediction in Test Set...
Saving Waterfall Plot for SHAP Values for a Single-Prediction in Test Set...
Saving feature importance ranking for XGB2...

shap means for XGB in CV2...
[6.1399941e-03 1.3443283e-05 1.7751092e-02 1.1837184e-02 8.3396683e-04
1.5874461e-03 4.5042816e-03 0.0000000e+00 0.0000000e+00 1.2435629e-03
2.8041655e-03 2.5228446e-03 0.0000000e+00 1.0483522e-02 7.2023137e-03
8.1350375e-03 0.0000000e+00 0.0000000e+00 0.0000000e+00 0.0000000e+00
0.0000000e+00 1.1928915e-03 1.0485627e-02 8.1559177e-03 1.5648296e-03
7.1601393e-03 3.2767998e-03 0.0000000e+00 3.1696835e-03 4.1859788e-03
1.1703122e-03 1.0534652e-02 3.0733834e-03 3.2062203e-04 6.3230167e-04
6.3157361e-04]

Displaying Force Plot for XGB SHAP Values in Whole Test Set...

Out [13]:



8/31/22, 2:58 PM

SHAP_final

Checking for matches...Alcohol is Alcohol
Alcohol value is 1.3443283023661934e-05

Checking for matches...Hepatitis B Surface Antigen is Hepatitis B Surface Antigen
Hepatitis B Surface Antigen value is 0.0

Checking for matches...Hepatitis B e Antigen is Hepatitis B e Antigen
Hepatitis B e Antigen value is 0.0

Checking for matches...Hepatitis C Virus Antibody is Hepatitis C Virus Antibody
Hepatitis C Virus Antibody value is 0.0

Checking for matches...Cirrhosis is Cirrhosis
Cirrhosis value is 0.0

Checking for matches...Endemic Countries is Endemic Countries
Endemic Countries value is 0.0

Checking for matches...Smoking is Smoking
Smoking value is 0.0006315736100077629

Checking for matches...Diabetes is Diabetes
Diabetes value is 0.0028041654732078314

Checking for matches...Hemochromatosis is Hemochromatosis
Hemochromatosis value is 0.0

Checking for matches...Arterial Hypertension is Arterial Hypertension
Arterial Hypertension value is 0.0008339668274857104

Checking for matches...Chronic Renal Insufficiency is Chronic Renal Insufficiency
Chronic Renal Insufficiency value is 0.0

Checking for matches...Human Immunodeficiency Virus is Human Immunodeficiency Virus
Human Immunodeficiency Virus value is 0.0

Checking for matches...Nonalcoholic Steatohepatitis is Nonalcoholic Steatohepatitis
Nonalcoholic Steatohepatitis value is 0.0

Checking for matches...Portal Hypertension is Portal Hypertension
Portal Hypertension value is 0.0003206220280844718

Checking for matches...Portal Vein Thrombosis is Portal Vein Thrombosis
Portal Vein Thrombosis value is 0.0006323016714304686

Checking for matches...Liver Metastasis is Liver Metastasis
Liver Metastasis value is 0.0015648296102881432

Checking for matches...Packs of cigarets per year is Packs of cigarets per year
Packs of cigarets per year value is 0.0011703121708706021

Checking for matches...Performance Status* is Performance Status*
Performance Status* value is 0.010534651577472687

Checking for matches...Ascites degree* is Ascites degree*
Ascites degree* value is 0.001587446080520749

Checking for matches...International Normalised Ratio* is International Normalised Ratio*
International Normalised Ratio* value is 0.0011928915046155453

Checking for matches...Alpha-Fetoprotein (ng/mL) is Alpha-Fetoprotein (ng/mL)
Alpha-Fetoprotein (ng/mL) value is 0.011837184429168701

Checking for matches...Haemoglobin (g/dL) is Haemoglobin (g/dL)
Haemoglobin (g/dL) value is 0.008135037496685982

Checking for matches...Mean Corpuscular Volume is Mean Corpuscular Volume
Mean Corpuscular Volume value is 0.0032767998054623604

Checking for matches...Leukocytes(G/L) is Leukocytes(G/L)
Leukocytes(G/L) value is 0.00815591774880886

Checking for matches...Platelets is Platelets
Platelets value is 0.003073383355513215

Checking for matches...Albumin (mg/dL) is Albumin (mg/dL)
Albumin (mg/dL) value is 0.006139994133263826

Checking for matches...Aspartate transaminase (U/L) is Aspartate transaminase (U/L)
Aspartate transaminase (U/L) value is 0.004504281561821699

Checking for matches...Gamma glutamyl transferase (U/L) is Gamma glutamyl transferase (U/L)
Gamma glutamyl transferase (U/L) value is 0.007202313747256994

Checking for matches...Alkaline phosphatase (U/L) is Alkaline phosphatase (U/L)
Alkaline phosphatase (U/L) value is 0.017751092091202736

Checking for matches...Creatinine (mg/dL) is Creatinine (mg/dL)
Creatinine (mg/dL) value is 0.0012435629032552242

Checking for matches...Number of Nodules is Number of Nodules
Number of Nodules value is 0.0031696835067123175

localhost:8888/nbconvert/html/streamline/SHAP_final.ipynb?download=false

93/98

Checking for matches...Major dimension of nodule (cm) is Major dimension of nodule (cm)
Major dimension of nodule (cm) value is 0.00716013927012682

Checking for matches...Direct Bilirubin (mg/dL) is Direct Bilirubin (mg/dL)
Direct Bilirubin (mg/dL) value is 0.0025228445883840322

Checking for matches...Iron is Iron
Iron value is 0.010485626757144928

Checking for matches...Oxygen Saturation (%) is Oxygen Saturation (%)
Oxygen Saturation (%) value is 0.004185978788882494

Checking for matches...Ferritin (ng/mL) is Ferritin (ng/mL)
Ferritin (ng/mL) value is 0.010483521968126297

	Gender	Symptoms	Alcohol	Hepatitis B Surface Antigen	Hepatitis B e Antigen	Hepatitis B Core Antibody	Hepatitis C Virus Antibody	Cirrhosis	Endemic Countries	Smoking	...	Gamma glutamyl transferase (U/L)	Alkaline phosphatase (U/L)	Prothrombin Time (sec)
0	0.0	0.000000	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	...	0.005661	0.062978	0.000000
1	0.0	0.134469	0.000000	0.0	0.0	0.0	0.020044	0.0	0.0	0.000000	...	0.035770	0.198403	0.000000
2	0.0	0.000000	0.000013	0.0	0.0	0.0	0.000000	0.0	0.0	0.000632	...	0.007202	0.017751	0.000000

3 rows x 49 columns

Run SHAP for Training Sets

Optional

- This runs on training CV Datasets that were partiioned during STREAMLINE
- User can set run_train to 'True' for comparison between training and testing sets

In [14]:

```
# testing all methods
run_force = False # parameter in run_force_plot(); set to True if user wants to display force plots for trained models
run_train = False

if run_train == True:
    for each in datasets:
        print("-----")
        print(each)
        print("-----")
        full_path = experiment_path+'/' + each
        filepath = f"/{full_path}/model_evaluation/shap_values/trainResults/" #path to save SHAP FI value results

        #Make folder in experiment folder/datafolder to store all shap_values per algorithm/CV combination
        if not os.path.exists(full_path+'/model_evaluation/shap_values/trainResults'):
            os.mkdir(full_path+'/model_evaluation/shap_values/trainResults')

        if not os.path.exists(full_path+'/model_evaluation/shap_values/trainResults/shapFigures'):
            os.mkdir(full_path+'/model_evaluation/shap_values/trainResults/shapFigures')
        save_path = full_path + '/model_evaluation/shap_values/trainResults/shapFigures'

        for algorithm in algorithms: #loop through algorithms
            print("-----")
            print(abbrev[algorithm])
            print("-----")

            if not os.path.exists(f'{experiment_path}/hcc-data_example/model_evaluation/shap_values/trainResults/'):
                os.mkdir(f'{experiment_path}/hcc-data_example/model_evaluation/shap_values/trainResults/')
            file_path = (f'{experiment_path}/hcc-data_example/model_evaluation/shap_values/trainResults/{abbrev[algorithm]}')

            FI_all = [] # list to store feature importanes to create shap values master list

            for cvCount in range(0,cv_partitions): #loop through cv's
                print(f"{abbrev[algorithm]}{cvCount} In CV{cvCount}...")

                # unpickle and load model
                result_file = f"/{full_path}/models/pickledModels/{abbrev[algorithm]}_{str(cvCount)}.pickle"
                file = open(result_file, 'rb')
                model = pickle.load(file)
                file.close()
                print(f'\nChecking if correct model is loaded...\n{model}')

                # Load CV datasets, paths to datasets updates with each iteration
                train_path = f"/{experiment_path}/{each}/CVDatasets/{each}_CV_{str(cvCount)}_Train.csv"
                test_path =f"/{experiment_path}/{each}/CVDatasets/{each}_CV_{str(cvCount)}_Test.csv"
                trainX, trainY,testX, testY, trainFeat, testFeat = dataPrep(train_path,instance_label,class_label, test_path)
                print(trainX)

                # shap computation and plots
                # Sanity check: print explainer to check if explainer exists
```

```
explainer = get_explainer(model, abbrev[algorithm], trainX) #explainer must always use training set
print(f"\nChecking explainer for {abbrev[algorithm]}{cvCount}...\n{explainer}\n")

print(f"Checking shap values for {abbrev[algorithm]}{cvCount}...\n")
shap_values = compute_shapValues(model, abbrev[algorithm], explainer, trainX)

print(f"\nGenerating SHAP plots for {abbrev[algorithm]}{cvCount}...\n")
shap_summary(abbrev[algorithm], trainFeat, shap_values, explainer, trainX, cvCount, save_path, 'Train')

#save SHAP FI results for each model per CV
print('Saving feature importance ranking for {}{}\n'.format(abbrev[algorithm], cvCount))
shap_fi_df, shap_means = shap_feature_ranking(abbrev[algorithm], shap_values, trainX, trainFeat) # can
print(f'shap means for {abbrev[algorithm]} in CV{cvCount}...\n{shap_means}')
shapFI_path = (f'{filepath}{abbrev[algorithm]}_{str(cvCount)}_shapFIValues_Test.csv')
shap_fi_df.to_csv(shapFI_path, header=True, index=True)

# OPTIONAL: set to 'TRUE' to save force plot figures
if run_force == True:
    run_force_plots(abbrev[algorithm], explainer, shap_values, trainX, trainFeat, cvCount, save_path,
else:
    continue

# create masterList of mean SHAP Values taken from shap_feature_ranking() for each model
save = save_shap(abbrev[algorithm],shap_values, shap_means, original_headers, cvCount, 'Train', each)
FI_all.append(save)
print("-----")
temp = FI_all
del FI_all # free up space

# create master list for each model after looping through all CVs
df = pd.DataFrame(np.asarray(temp), columns=original_headers)
display(df)

path = (f'{filepath}{abbrev[algorithm]}_shapMasterList.csv')
df.to_csv(path, index=True)
```

Read in Model SHAP Value Master List to Create SHAP Figures for the Final Model Over All CVs

NOT ENTIRELY SURE IF THIS IS ACTUALLY DISPLAYING SHAP VALUE AVERAGES OF EACH FEATURE (OVER ALL CVs)FOR ONE MODEL

Areas of concern/ISSUES:

- Not necessarily correct if I'm using just the model from different partitions to display overall SHAP value averages using the respective CV0-CV2 datasets
 - Unless I'm comparing the figures using different Explanation objects to show how different trained models are 'explaining' the prediction

```
In [16]: '''Read in master list file for each trained model to create average SHAP value for each
feature over all CVs

Will be used to create summative figures of the model rather than summative figures of each CV per model'''

abbrev='LR'
read_df = pd.read_csv(f'{experiment_path}/hcc-data_example/model_evaluation/shap_values/testResults/LR_shapMasterList.
# print(len(read_df.columns))

final_mean = []
for i in range(0, len(read_df.columns)):
#     print(read_df.iloc[:, i].values)
    final_mean.append(np.mean(read_df.iloc[:, i].values))
final_mean = np.asarray(final_mean)

result_file = experiment_path+ '/hcc-data_example/models/pickledModels/LR_0.pickle'
file = open(result_file, 'rb')
model = pickle.load(file)
file.close()
# print('\nChecking if correct model is loaded...\n', model)

vals = []
vals = np.asarray(read_df.values)
print(vals)

# original_dataset_vals = pd.read_csv(dataset_path+"/hcc-data_example.csv")
# print(original_dataset_vals)

train_path = experiment_path + '/hcc-data_example/CVDatasets/hcc-data_example_CV_0_Train.csv'
```



```
test_path = experiment_path +  '/hcc-data_example/CVDatasets/hcc-data_example_CV_0_Test.csv'
trainX, trainY,testX, testY, trainFeat, testFeat = dataPrep(train_path,instance_label,class_label, test_path)

explainer = get_explainer(model, abbrev, trainX)

'''Might not be correct since the dataset being used is testing CV0 unless you open the original
    hcc-data_example dataset csv file

    Another concern might be that this wouldn't be the correct usage of Explainer objects & creating
    figures from it.....in this case it seems I'm using trained models from each CV (& its respective CV dataset)
    to explain the SHAP value average of feature taken over all CVs (from the master list)'''

shap.summary_plot(vals, original_headers, plot_type='bar')
shap.force_plot(explainer.expected_value, vals, original_headers, show=True)
shap.decision_plot(explainer.expected_value, vals, original_headers, show=True)

print(f'Single Prediction of {abbrev} SHAP Values Master List\n')
shap.plots._waterfall.waterfall_legacy(explainer.expected_value, vals[0], testX.iloc[0], original_headers, show=True)
shap.force_plot(explainer.expected_value, vals[0], original_headers, show=True)

# explainer = get_explainer(model, abbrev, trainX)

# # shap_values = compute_shapValues(model, 'NB', explainer, testX)
# shap.summary_plot(np.mean(read_df.iloc[:, :].values), testX, original_headers)
# print(final_mean)
```

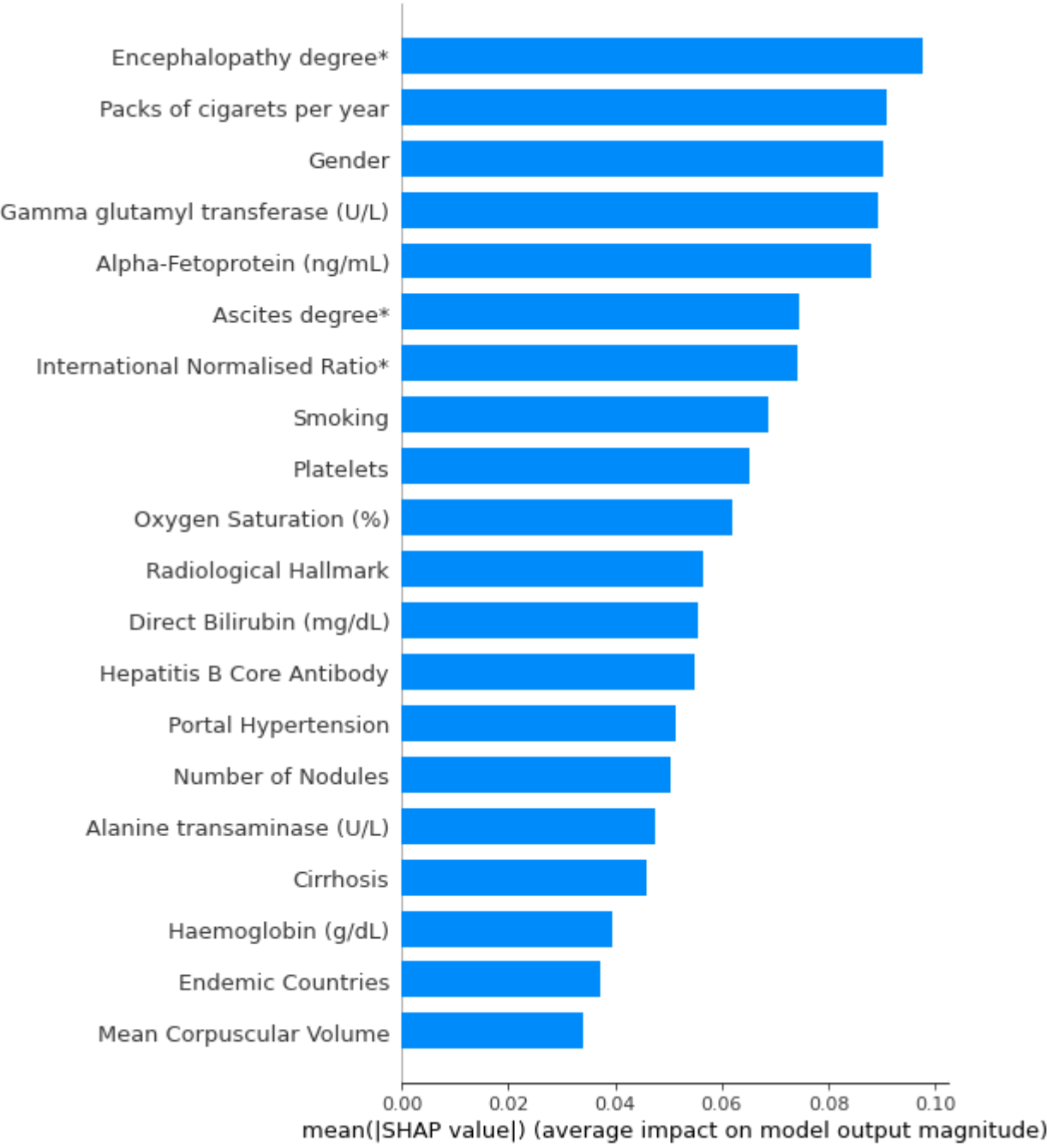
Out[16]:

'Read in master list file for each trained model to create average SHAP value for each \n feature over all CVs\n\nWill be used to create summative figures of the model rather than summative figures of each CV per model'

[[3.88687666e-02 0.00000000e+00 1.90267462e-02 0.00000000e+00
0.00000000e+00 2.68384150e-02 9.67006520e-03 1.58771000e-02
2.15841281e-03 4.41119374e-02 1.86880021e-02 7.43958924e-03
1.04644599e-02 2.15319336e-02 0.00000000e+00 0.00000000e+00
4.36115897e-02 1.31827683e-03 2.10859786e-02 5.05332922e-02
5.89059310e-02 0.00000000e+00 0.00000000e+00 0.00000000e+00
1.34865542e-02 6.69177804e-02 1.75002590e-02 4.33477089e-02
1.72370853e-02 6.09193671e-03 3.34323800e-02 3.47406060e-03
2.05261179e-02 0.00000000e+00 4.25162309e-02 1.69800337e-02
1.40632594e-03 4.26483629e-02 2.78418760e-02 3.86576145e-02
0.00000000e+00 3.60957948e-02 2.63454185e-03 4.26100992e-02
1.78433863e-02 5.80196642e-03 2.47346685e-03 3.59232225e-02]
[2.26906417e-01 4.92051614e-03 1.10065268e-02 1.14805528e-02
2.90084907e-02 1.36120077e-01 1.95559933e-02 1.21705385e-01
1.09777490e-01 1.56453022e-01 7.43264405e-02 0.00000000e+00
0.00000000e+00 0.00000000e+00 0.00000000e+00 9.49666703e-02
3.47050531e-02 0.00000000e+00 7.75635064e-02 9.65491317e-02
6.66054544e-03 0.00000000e+00 1.66543429e-01 1.15708226e-02
5.61750932e-02 1.95255829e-01 0.00000000e+00 2.42653076e-01
2.02692283e-01 2.16564337e-01 2.23672850e-01 1.12992460e-01
7.51284698e-02 2.22584328e-02 1.46038308e-01 8.37844922e-02
7.26644361e-02 9.53877084e-02 5.44432132e-02 2.21571656e-01
2.58407069e-02 4.79478486e-02 1.78410835e-02 1.02911846e-01
7.71762661e-02 1.55097922e-01 9.10400977e-03 1.42574530e-01]
[4.76544111e-03 2.21482139e-03 0.00000000e+00 9.57013968e-04
3.80908259e-03 1.91522175e-03 0.00000000e+00 3.30706823e-04
0.00000000e+00 5.27881775e-03 3.75910628e-03 0.00000000e+00
3.91915048e-03 1.64743809e-03 2.73061318e-03 0.00000000e+00
0.00000000e+00 0.00000000e+00 2.79623086e-03 7.26821080e-03
8.86535277e-03 0.00000000e+00 2.52578912e-03 0.00000000e+00
4.80513725e-04 1.07792792e-02 7.57378172e-04 6.47924621e-03
3.56031389e-03 5.94797251e-05 6.63804149e-03 2.09495109e-03
6.26003349e-03 4.05709949e-03 7.22348091e-03 0.00000000e+00
0.00000000e+00 4.43775359e-03 3.77134111e-03 7.76975487e-03
0.00000000e+00 9.91359698e-04 5.68981824e-03 5.64753789e-03
4.72258480e-03 5.37089763e-03 4.18704626e-04 7.53856654e-03]]

Out[16]:

"Might not be correct since the dataset being used is testing CV0 unless you open the original \n hcc-data_example
dataset csv file\n \n Another concern might be that this wouldn't be the correct usage of Explainer objects & cr
eating \n figures from it.....in this case it seems I'm using trained models from each CV (& its respective CV
dataset)\n to explain the SHAP value average of feature taken over all CVs (from the master list)"



```

-----
IndexError                                Traceback (most recent call last)
Input In [16], in <cell line: 49>()
    41 '''Might not be correct since the dataset being used is testing CV0 unless you open the original
    42     hcc-data_example dataset csv file
    43
    44     Another concern might be that this wouldn't be the correct usage of Explainer objects & creating
    45     figures from it.....in this case it seems I'm using trained models from each CV (& its respective CV d
ataset)
    46     to explain the SHAP value average of feature taken over all CVs (from the master list)'''
    48 shap.summary_plot(vals, original_headers, plot_type='bar')
--> 49 shap.force_plot(explainer.expected_value, vals, original_headers, show=True)
    50 shap.decision_plot(explainer.expected_value, vals, original_headers, show=True)
    52 print(f'Single Prediction of {abbrev} SHAP Values Master List\n')

File ~/opt/anaconda3/lib/python3.9/site-packages/shap/plots/_force.py:201, in force(base_value, shap_values, features,
feature_names, out_names, link, plot_cmap, matplotlib, show, figsize, ordering_keys, ordering_keys_time_format, text_r
otation, contribution_threshold)
    189     e = AdditiveExplanation(
    190         base_value,
    191         np.sum(shap_values[k, :]) + base_value,
    (...)
    197         DenseData(np.ones((1, len(feature_names))), list(feature_names))
    198     )
    199     exps.append(e)
--> 201 return visualize(
    202     exps,
    203     plot_cmap=plot_cmap,
    204     ordering_keys=ordering_keys,
    205     ordering_keys_time_format=ordering_keys_time_format,
    206     text_rotation=text_rotation
    207 )

File ~/opt/anaconda3/lib/python3.9/site-packages/shap/plots/_force.py:349, in visualize(e, plot_cmap, matplotlib, figs
ize, show, ordering_keys, ordering_keys_time_format, text_rotation, min_perc)
    347     assert False, "Matplotlib plot is only supported for additive explanations"
    348     else:
--> 349     return AdditiveForceArrayVisualizer(e, plot_cmap=plot_cmap, ordering_keys=ordering_keys, ordering_keys
_time_format=ordering_keys_time_format)
    350 else:
    351     assert False, "visualize() can only display Explanation objects (or arrays of them)!"

File ~/opt/anaconda3/lib/python3.9/site-packages/shap/plots/_force.py:471, in AdditiveForceArrayVisualizer.__init__(se
lf, arr, plot_cmap, ordering_keys, ordering_keys_time_format)
    465 for (ind,e) in enumerate(arr):
    466     self.data["explanations"].append({
    467         "outValue": ensure_not_numpy(e.out_value),
    468         "simIndex": ensure_not_numpy(clustOrder[ind])+1,
    469         "features": {}
    470     })
--> 471 for i in filter(lambda j: e.effects[j] != 0 or e.instance.x[0,j] != 0, range(len(e.data.group_names))):
    472     self.data["explanations"][-1]["features"][i] = {
    473         "effect": ensure_not_numpy(e.effects[i]),
    474         "value": ensure_not_numpy(e.instance.group_display_values[i])
    475     }

File ~/opt/anaconda3/lib/python3.9/site-packages/shap/plots/_force.py:471, in AdditiveForceArrayVisualizer.__init__.<1
ocals>.<lambda>(j)
    465 for (ind,e) in enumerate(arr):
    466     self.data["explanations"].append({
    467         "outValue": ensure_not_numpy(e.out_value),
    468         "simIndex": ensure_not_numpy(clustOrder[ind])+1,
    469         "features": {}
    470     })
--> 471 for i in filter(lambda j: e.effects[j] != 0 or e.instance.x[0,j] != 0, range(len(e.data.group_names))):
    472     self.data["explanations"][-1]["features"][i] = {
    473         "effect": ensure_not_numpy(e.effects[i]),
    474         "value": ensure_not_numpy(e.instance.group_display_values[i])
    475     }

IndexError: index 48 is out of bounds for axis 0 with size 48

```

In []: