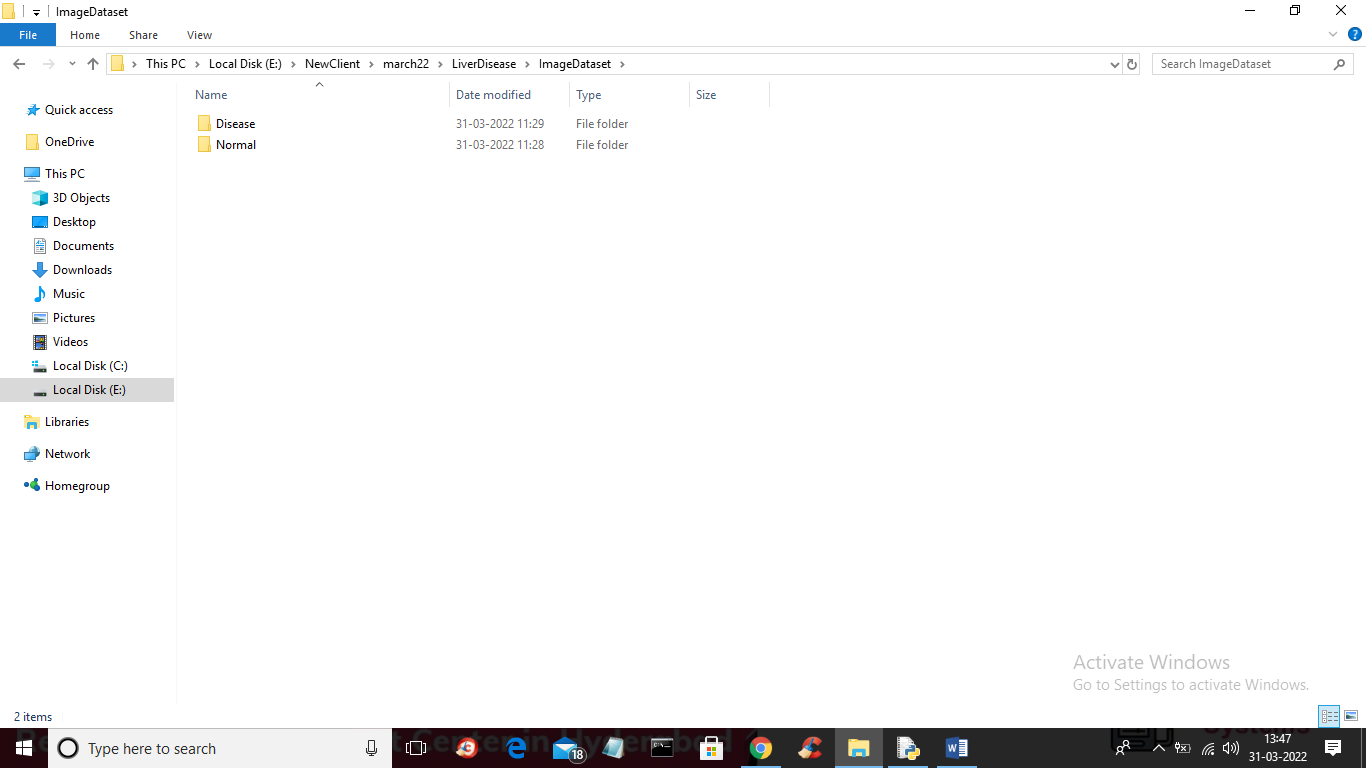
Diagnosis of Liver Diseases using Machine Learning

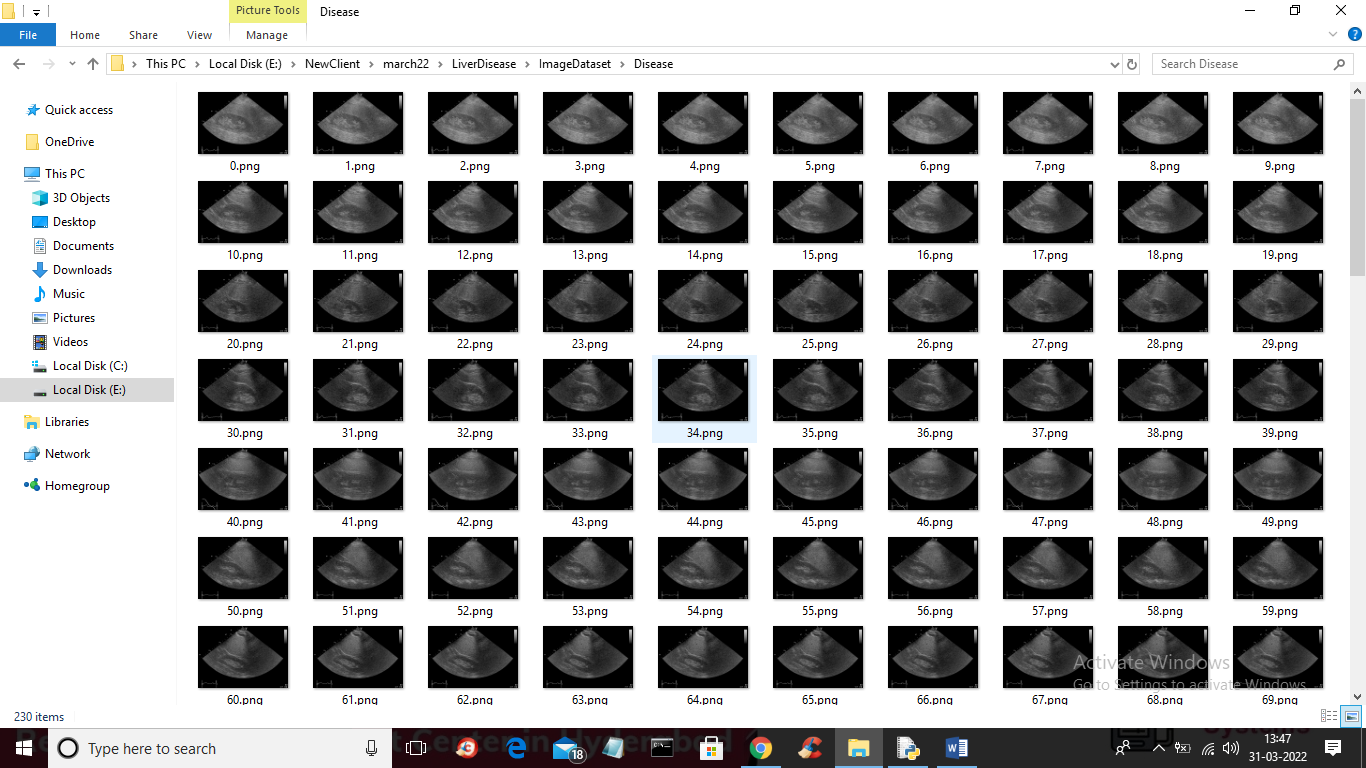
Due to liver diseases many peoples across the world lost their lives and its death rate can be reduced only by diagnosing disease on time but the main problem is LIVER will not show any symptoms for earlier damage. So author of this paper is applying two methods to predict liver disease.

Method1) in this method author is using INDIAN LIVER dataset to train various machine learning algorithms such as SVM, ANN and multilayer perceptron and this trained model will be applied on new patients TEST data to predict liver is normal or not but student ask us to implement Logistic Regression, Naïve Bayes and then compare its performance with SVM so we are using student suggested algorithms

Method2) in this method author is training ANN and CNN with gene MRNA images dataset and then training with CNN and ANN to predict whether liver disease inheriting in genes from ancestors. Student also asking to used liver images and then train with CNN and ANN but liver gene images are not available so we are using LIVER ULTRA SOUND SCAN IMAGES and below screen showing those images dataset

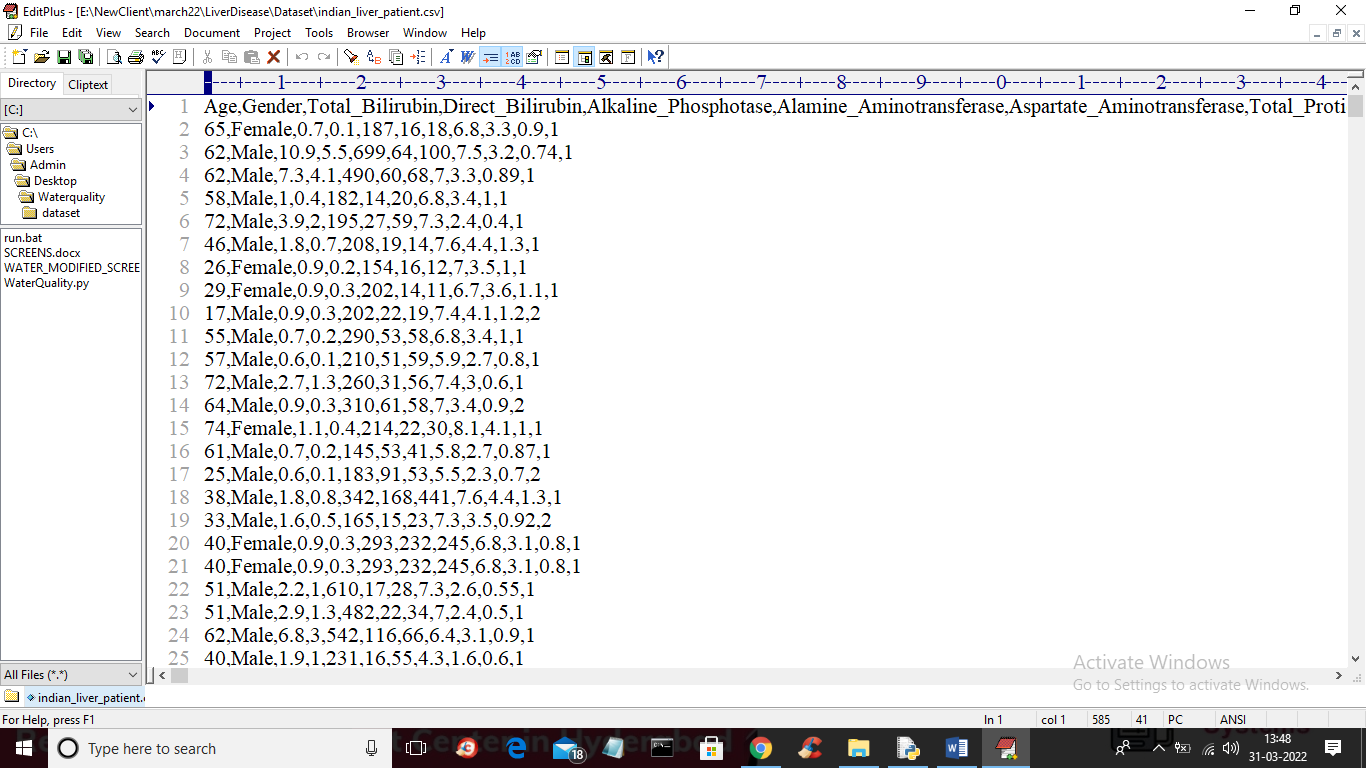


In above screen we are having two folders such as Normal and Disease and just go inside any folder to see those images



So we will use above images to train CNN and ANN and then evaluate their performance in terms of accuracy, precision and recall etc.

Below screen showing Indian Liver Dataset

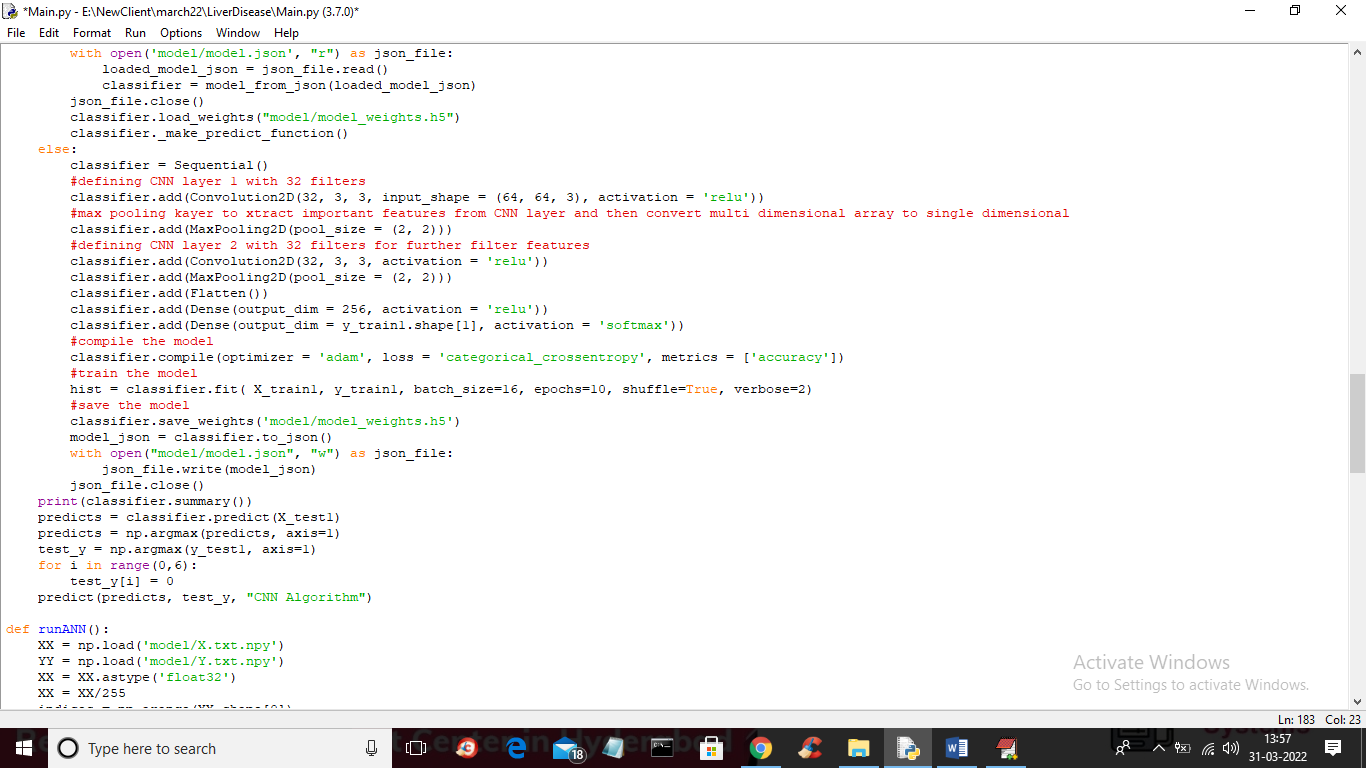


In above screen first row represents dataset column names and y-axis represents dataset values and in last column we have class label as 1 or 2 where 1 means normal and 2 means disease and we will use above dataset to train SVM, Logistic Regression and Naïve Bayes and evaluate their performance in terms of accuracy and other metrics.

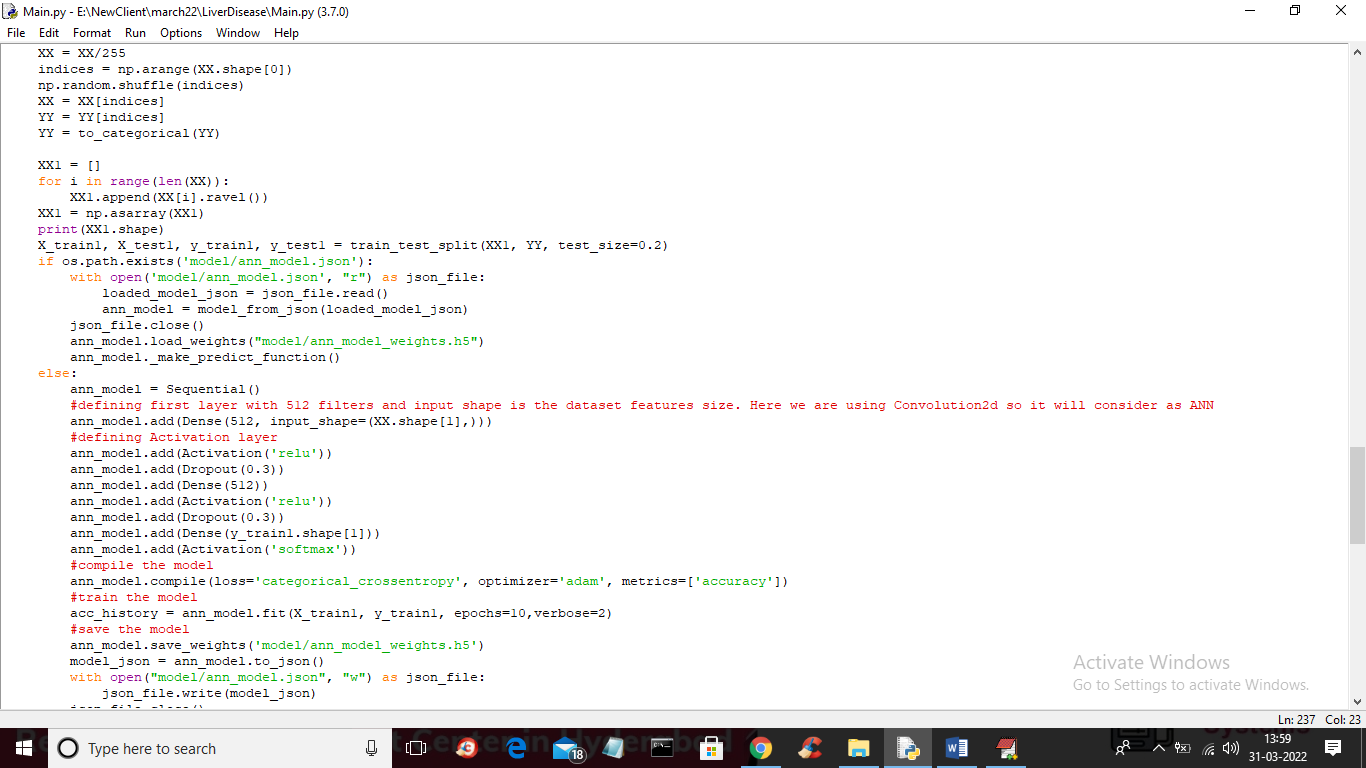
Modules Information

1. Upload Indian Liver Dataset: using this module we will upload Indian Liver dataset to application
2. Preprocess Dataset: using this module we will read dataset and then replace missing values with 0 and this dataset is highly imbalance as normal records are 167 and disease records are 450 so we are applying over and under sampling algorithm to equal both records and then split dataset into training and testing where application used 80% dataset for training and 20% for testing
3. Run SVM Algorithm: using this module we will train SVM with above dataset and then calculate its performance
4. Run Logistic Regression Algorithms: using this module we will train Logistic Regression and calculate its performance
5. Run Naïve Bayes Algorithms: using this module we will train Naïve Bayes and calculate its performance
6. Run CNN Images Algorithm: using this module we will read all normal and disease images and then train with CNN and this trained model will be applied on test data to calculate its prediction accuracy
7. Run ANN Images Algorithm: using this module we will read all normal and disease images and then train with ANN and this trained model will be applied on test data to calculate its prediction accuracy
8. Comparison Graph: using this module we will plot comparison graph of all algorithms.

Below screen showing code for CNN and ANN



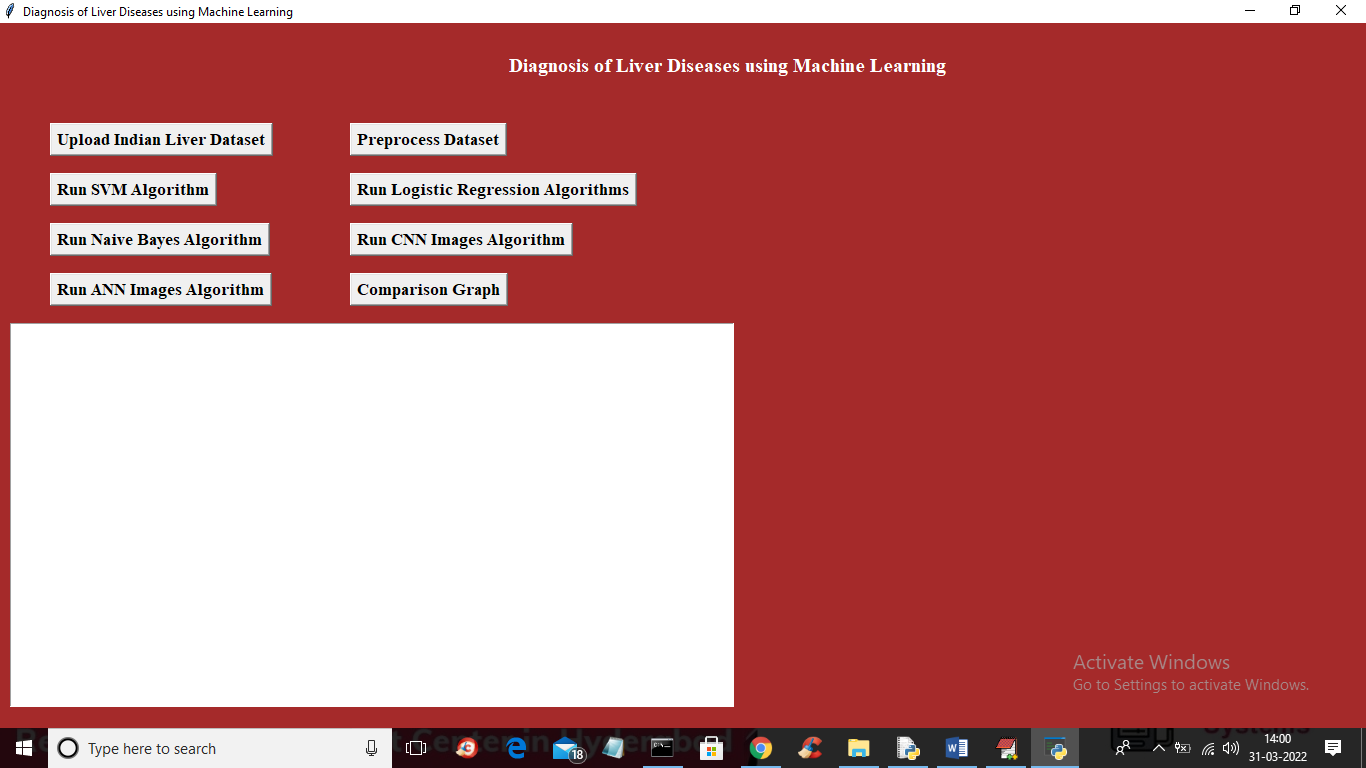
In above screen read red colour comments to know about CNN and below is the code for ANN



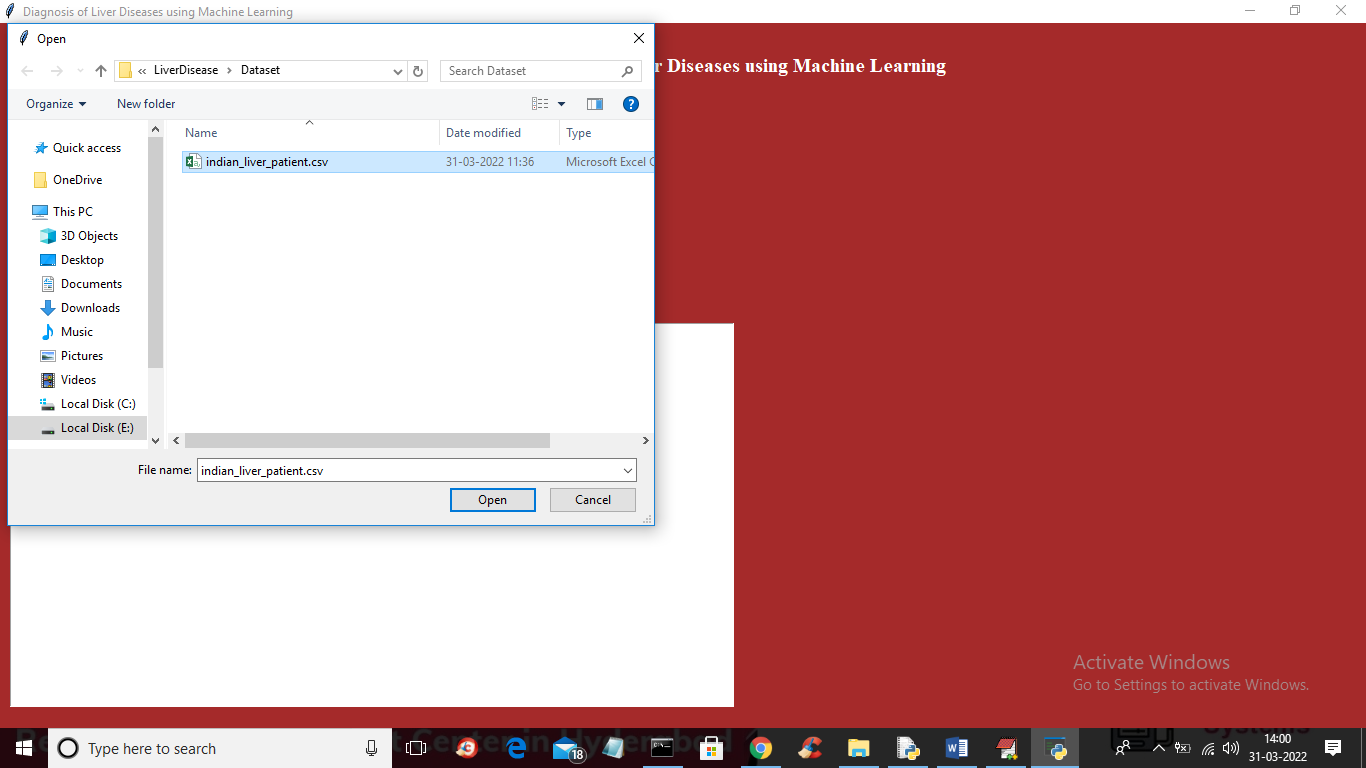
In above screen read red colour comments to know about ANN.

SCREEN SHOTS

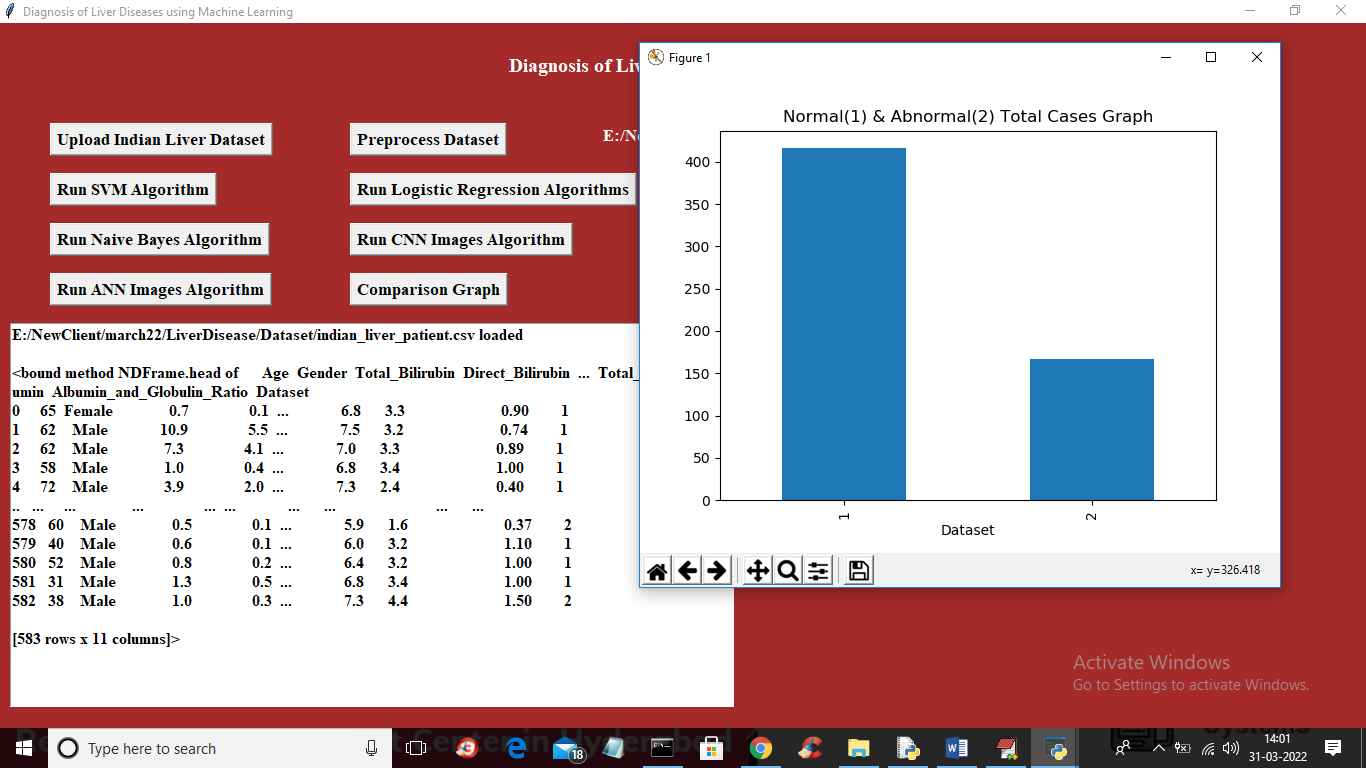
To run project double click on ‘run.bat’ file to get below screen



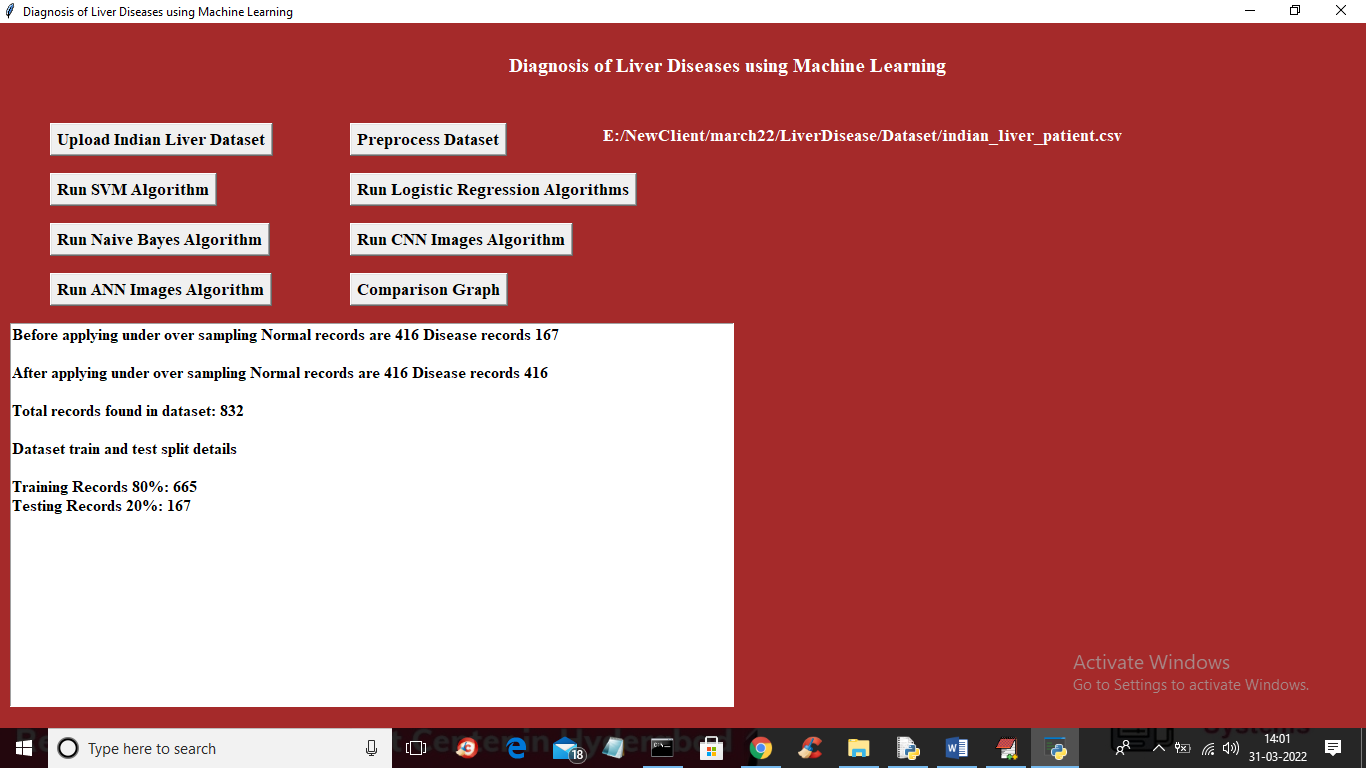
In above screen click on ‘Upload Indian Liver Dataset’ button to upload dataset and to get below screen



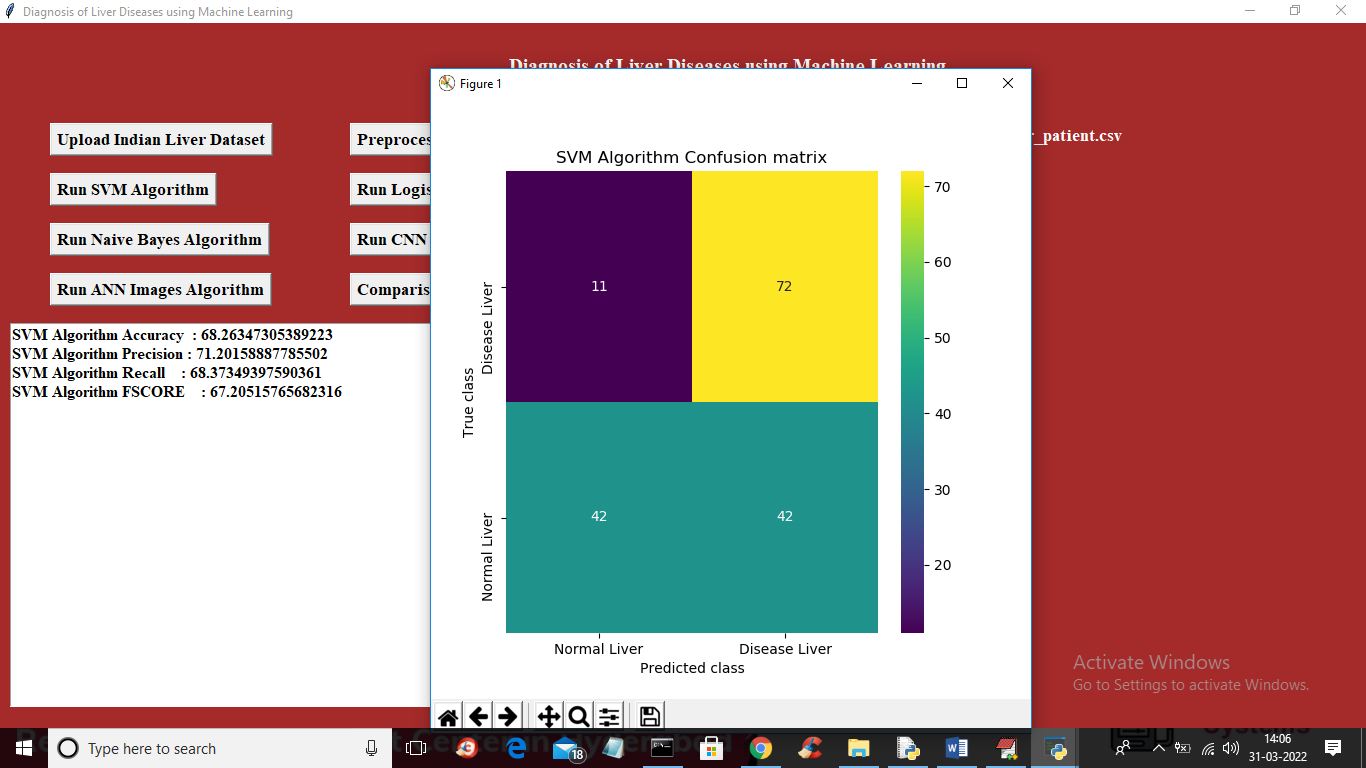
In above screen selecting and uploading dataset and then click on ‘Open’ button to get below output



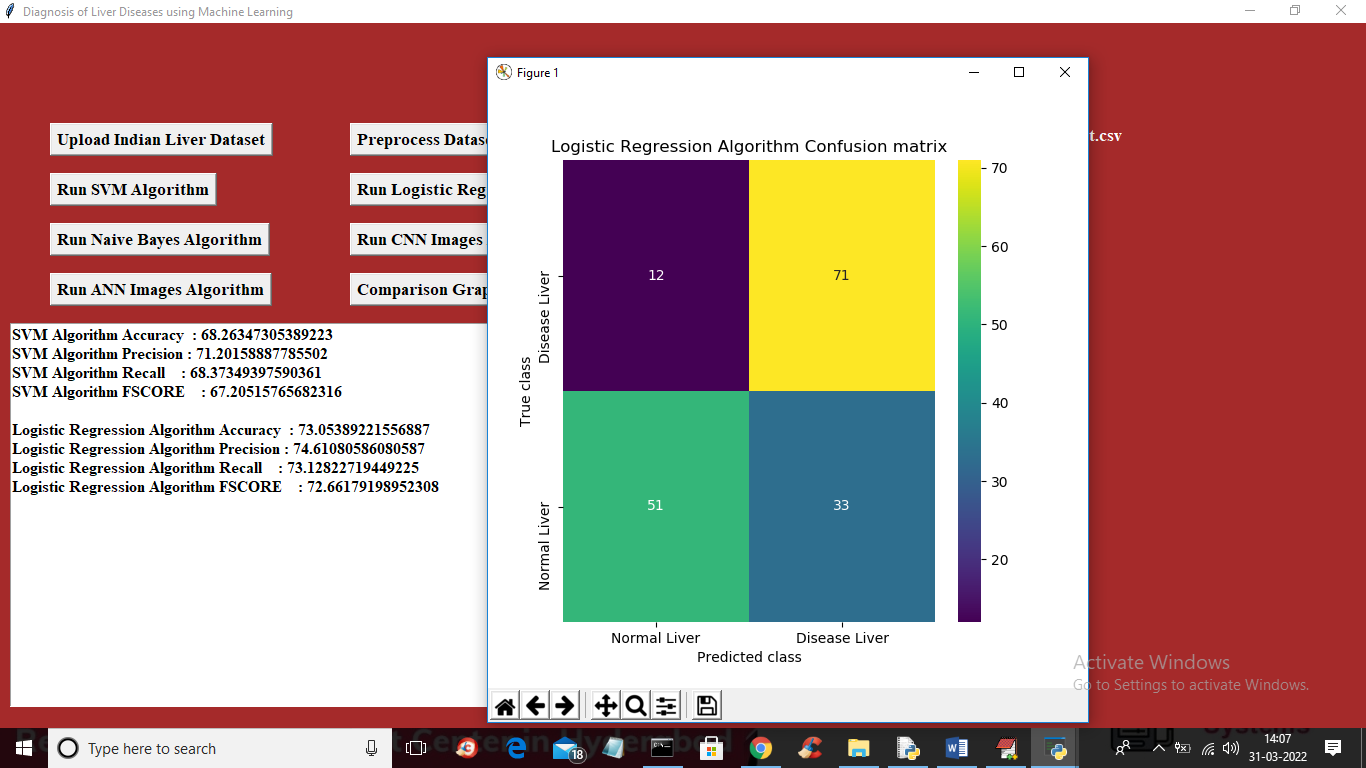
In above screen dataset loaded and in graph we can see number of records for normal as 1 and disease as 2 and in above screen we can see gender is displaying as English characters and ML will not take such character so we need to convert to integer code by applying preprocessing technique so close above graph and then click on ‘Preprocess Dataset’ button to get below output



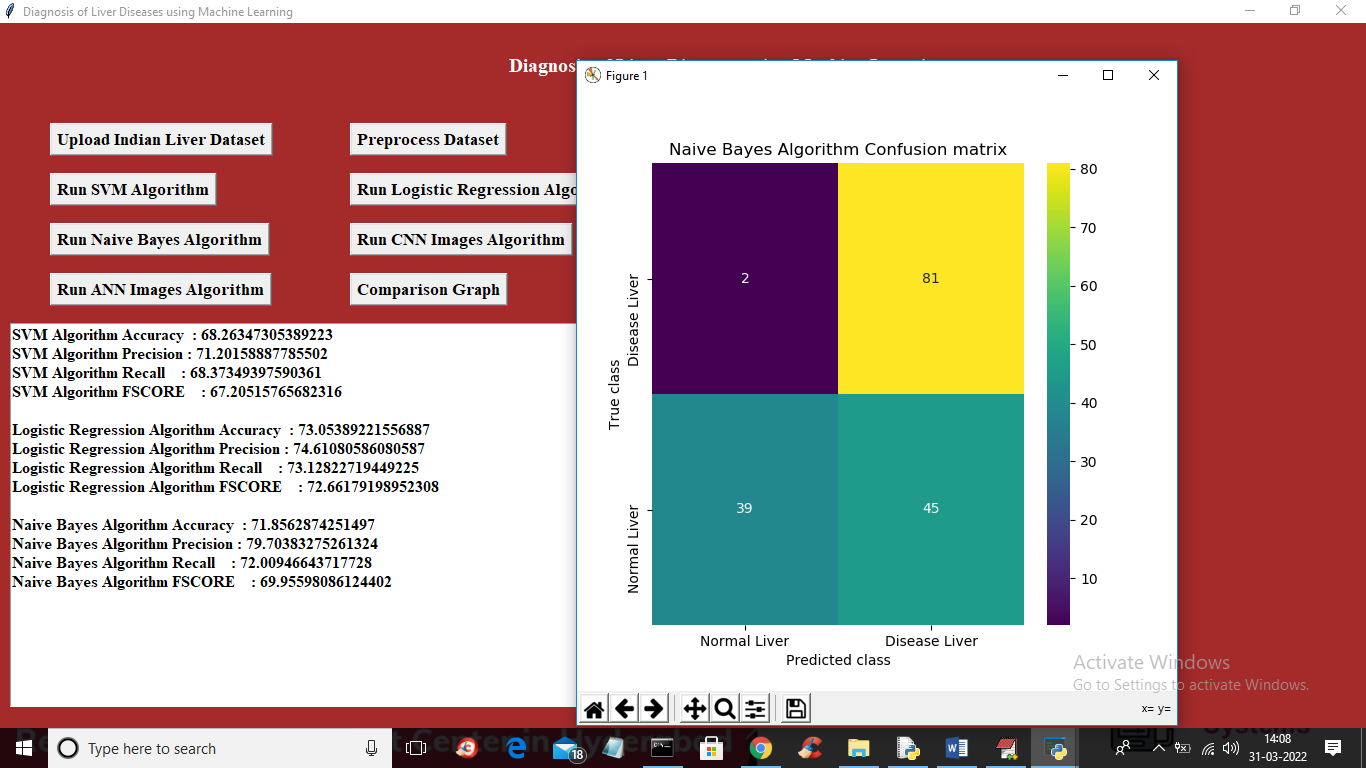
In above screen dataset preprocessing completed and we can see before applying under and over sample disease records were 167 and after applying its increase to 416 and we can see dataset split size and now dataset is ready so click on ‘Run SVM Algorithm’ button to get below output



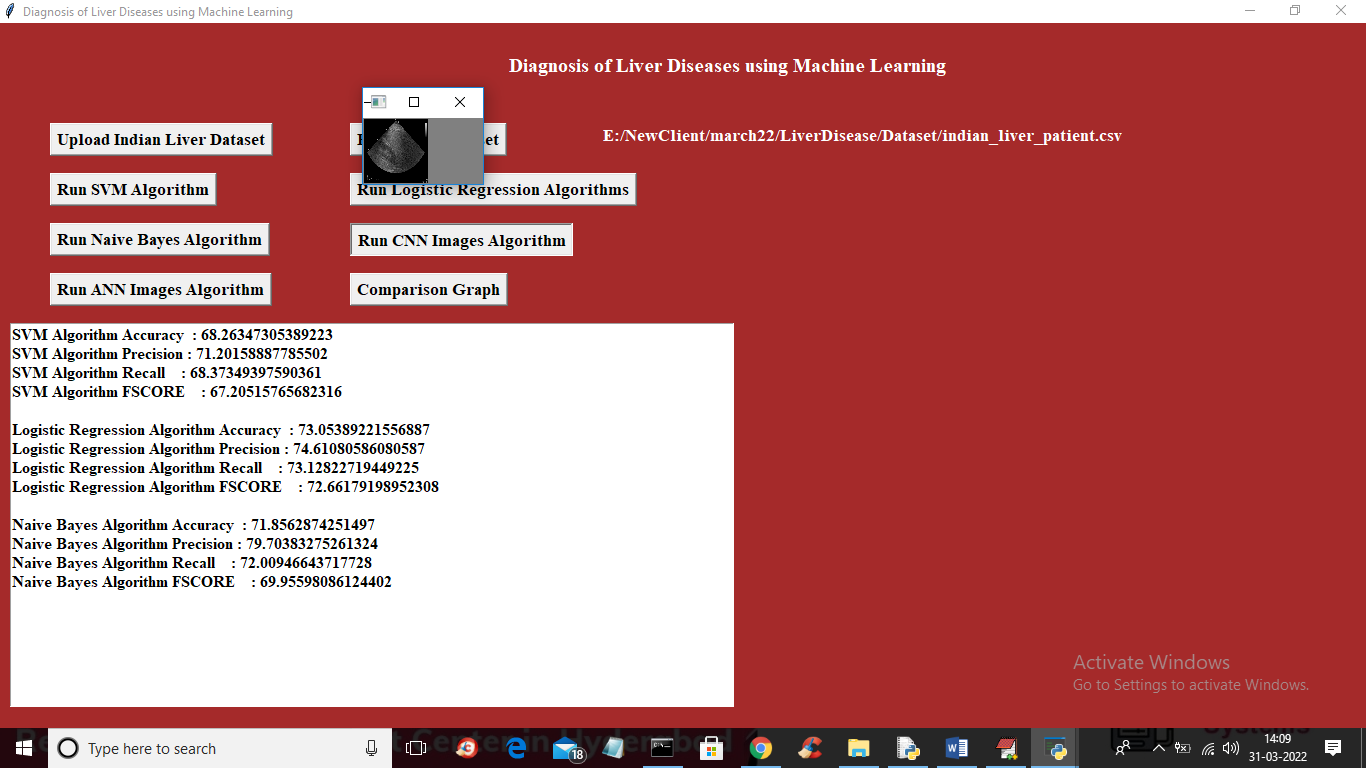
In above screen SVM is trained and we got its accuracy as 68% and in confusion matrix graph we can see SVM predicted 42 records as Normal CORRECTLY and 11 predicted as Incorrect. Now close above graph and then click on ‘Run Logistic Regression Algorithm’ button to get below output



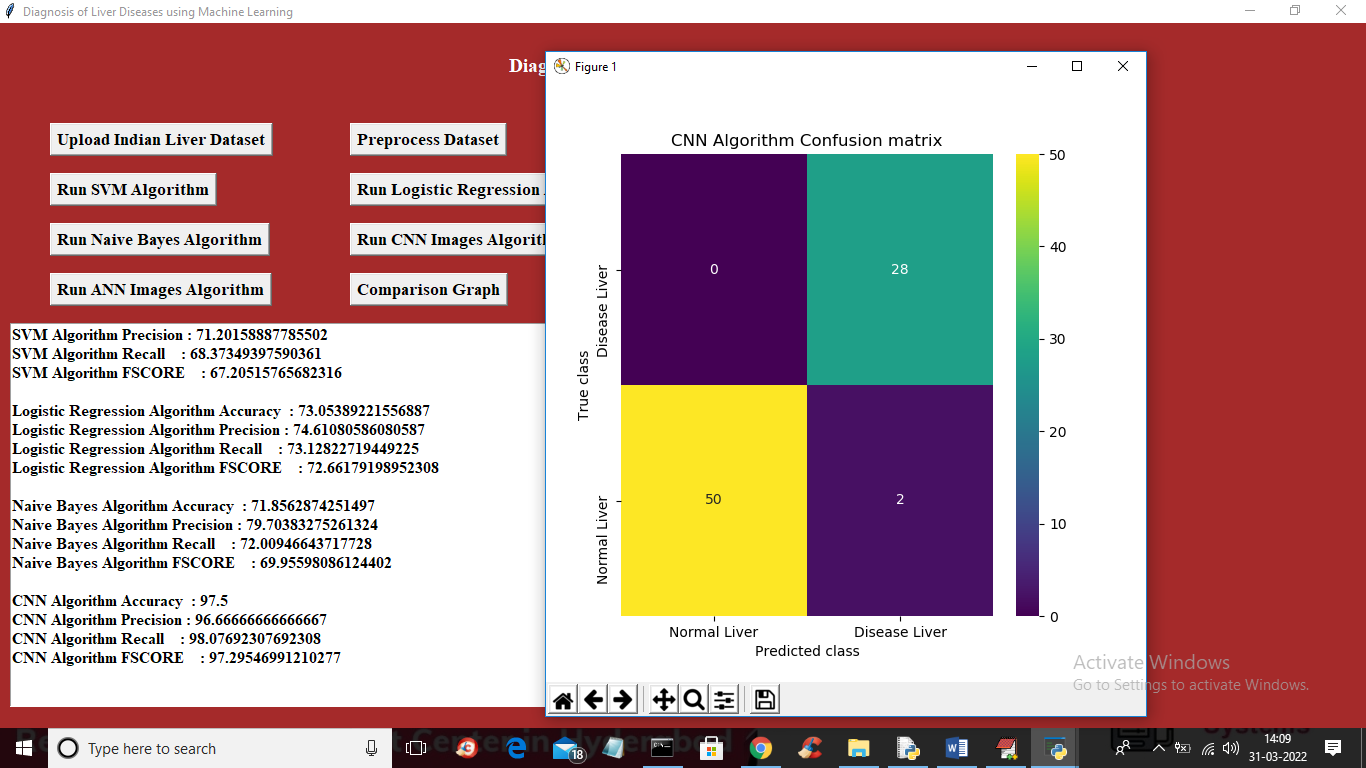
In above screen with Logistic Regression we got accuracy as 73% and in confusion matrix 51 records correctly predicted and 12 incorrectly predicted and now close above graph and then click on ‘Run Naïve Bayes Algorithm’ button to get below output



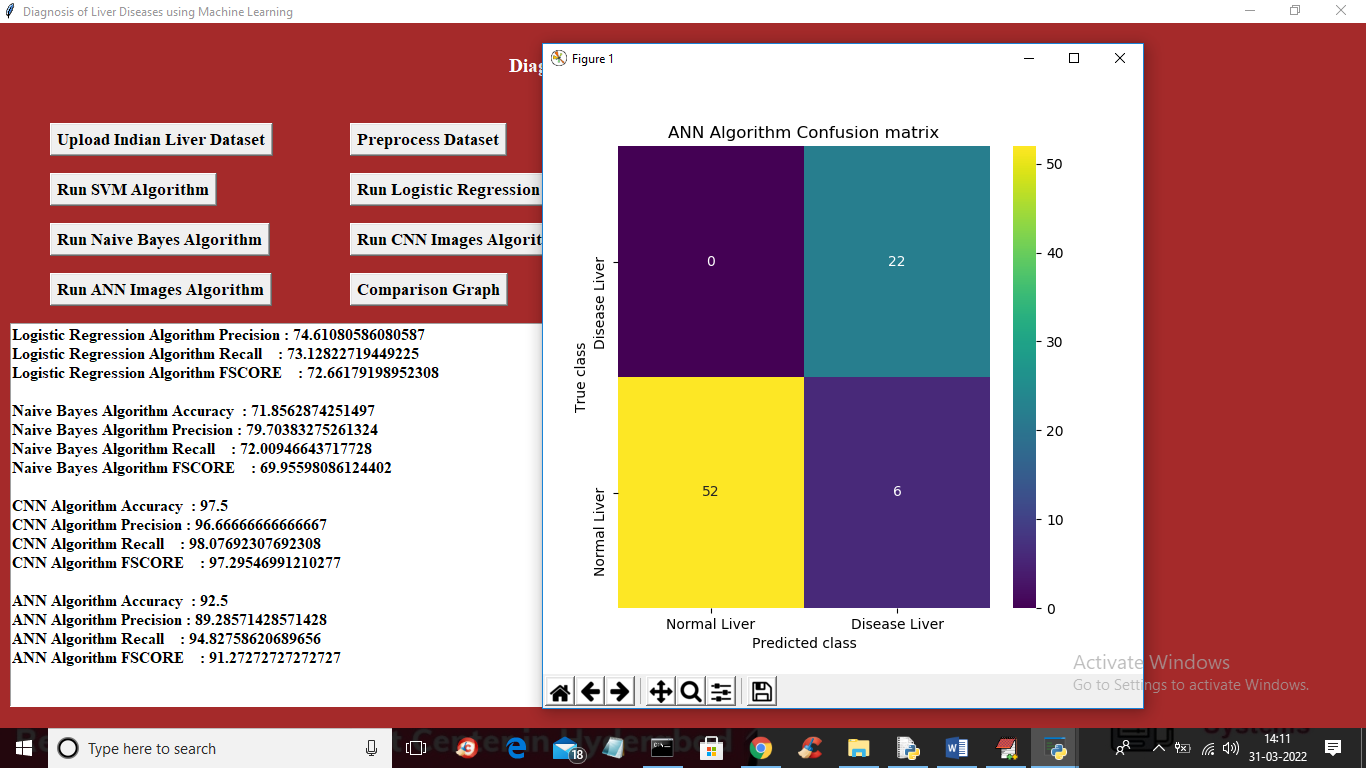
In above screen with Naïve Bayes we got 71% accuracy and now click on ‘Run CNN Algorithm’ button to get below output



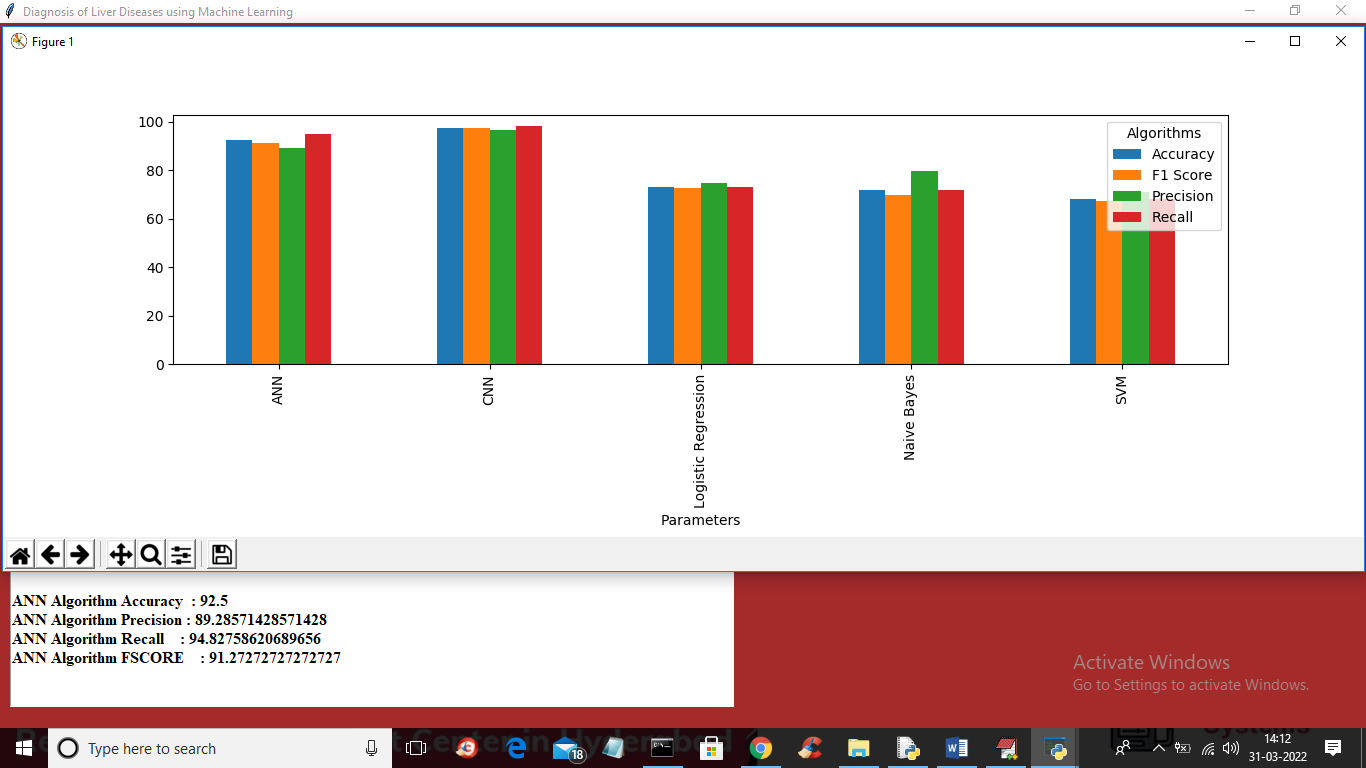
In above screen CNN training on images and we are displaying one processed image so close above showing image to get below output



In above screen with CNN we got 97% accuracy and in confusion matrix we can see only 2 records are incorrectly predicted as NORMAL but its Disease image and now close above graph and then click on ‘Run ANN Algorithm’ button to get below output



In above screen with ANN we got 92% accuracy and in confusion graph we can see ANN predicted 6 records incorrectly so for Indian Liver dataset Logistic regression gave better accuracy and for image CNN gave better accuracy and now click on ‘Comparison Graph’ button to get below output



In above graph x-axis represents algorithm names and y-axis represents accuracy and other values and in each different colour bar represents different metric such as accuracy, precision, recall and FSCORE.