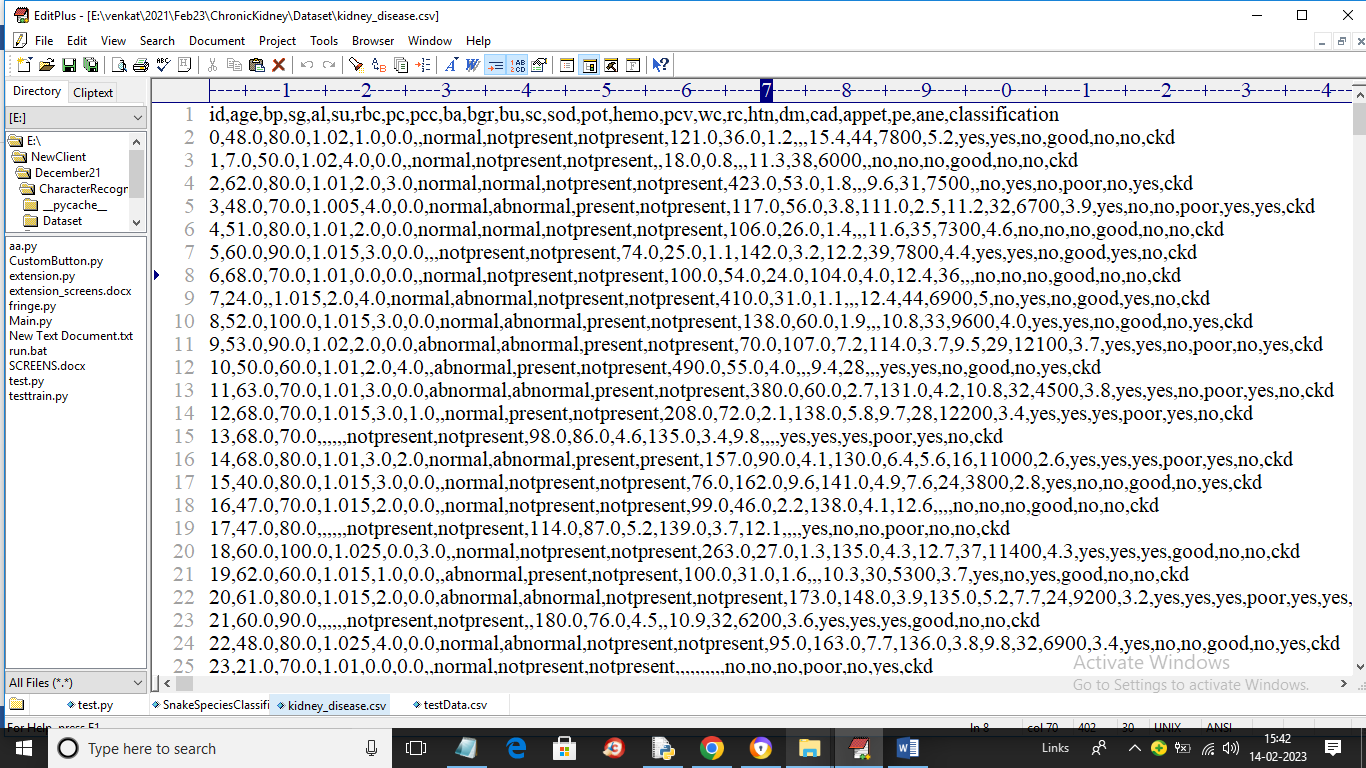
Chronic Kidney Disease Prediction using CNN, LSTM & Ensemble Model

In this project we are using various deep learning algorithms such as CNN, LSTM and Ensemble CNN with Random Forest to predict chronic kidney disease. To train all algorithms we have used Chronic Kidney Dataset from KAGGLE repository and below is the dataset URL

<https://www.kaggle.com/datasets/mansoordaku/ckdisease>

In below screen we are showing dataset details



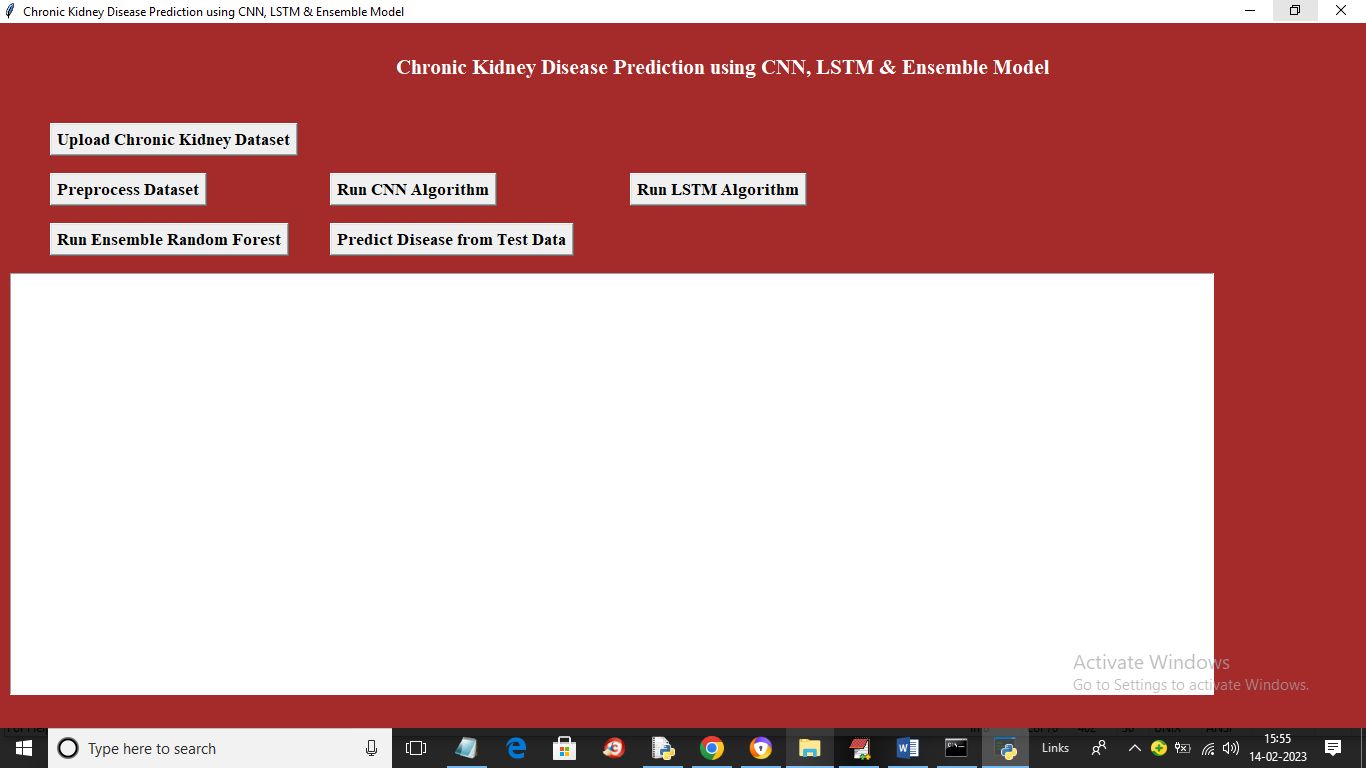
In above dataset screen first row contains dataset column names and remaining rows contains dataset values and in last column we have class label called ‘CKD or NO\_CKD’. So by using above dataset we will train all algorithms and then evaluate their performance in terms of accuracy, precision, recall, FCSORE and Confusion Matrix. In above dataset we can see dataset contains both non-numeric and numeric values but algorithms accept only numeric values so we need to convert all non-numeric values into numeric values by applying Label Encoder class.

To implement this project we have designed following modules

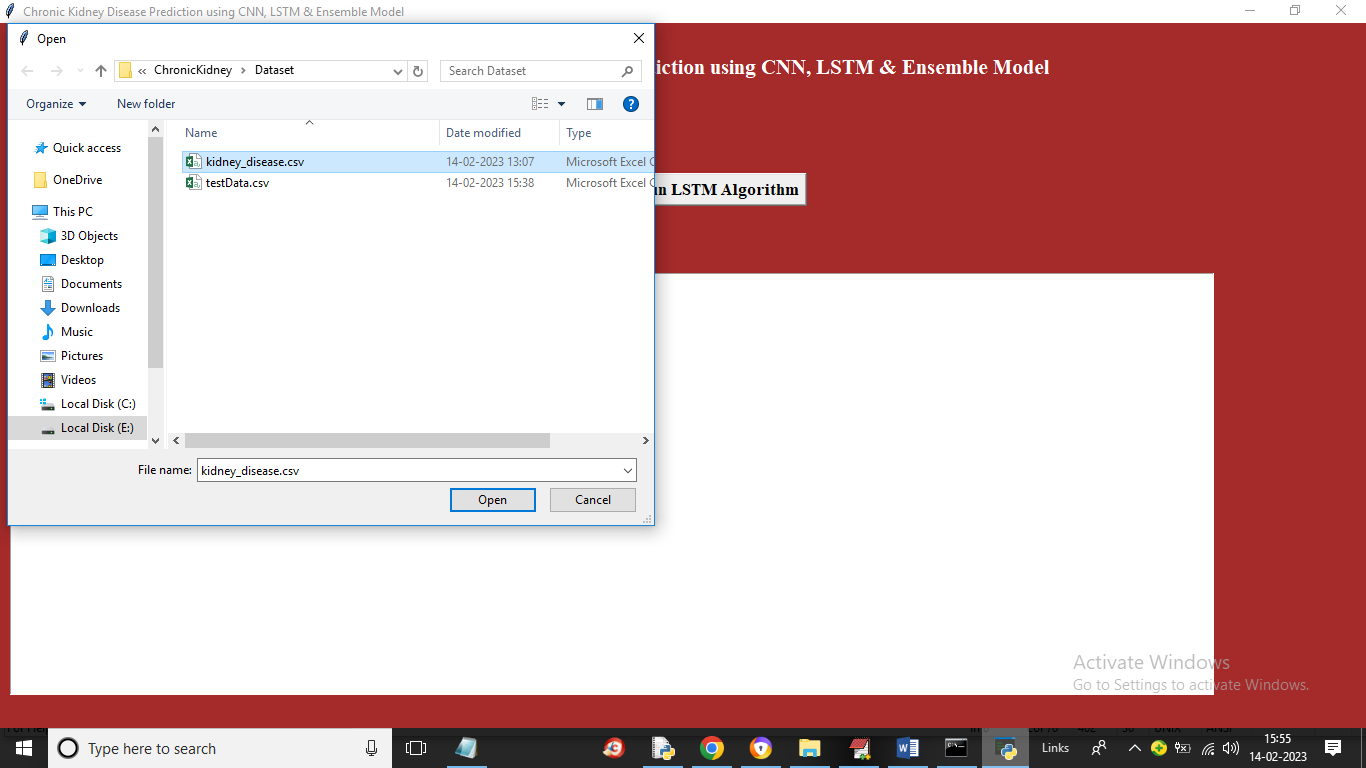
1. Upload Chronic Kidney Dataset: using this module we will upload dataset to application and then plot number of NON-CKD and CKD patients graph
2. Preprocess Dataset: using this module we will read dataset and then replace missing values with 0 and then normalize dataset values and then split dataset into train and test where application will be using 80% dataset for training and 20% for testing
3. Run CNN Algorithm: using this module we will input 80% training data to CNN algorithm to train a model and this model will be applied on 20% test data to calculate prediction accuracy
4. Run LSTM Algorithm: using this module we will input 80% training data to LSTM algorithm to train a model and this model will be applied on 20% test data to calculate prediction accuracy
5. Run Ensemble Random Forest: using this module we will extract trained optimized features from CNN and then retrain CNN optimized features with Random Forest algorithm to further improved accuracy and make its ensemble model
6. Comparison Graph: using this module we will plot accuracy graph between all algorithms
7. Predict Disease from Test Data: using this module we will upload test data and then ensemble model will predict weather test data is Normal or contains CKD disease

SCREEN SHOTS

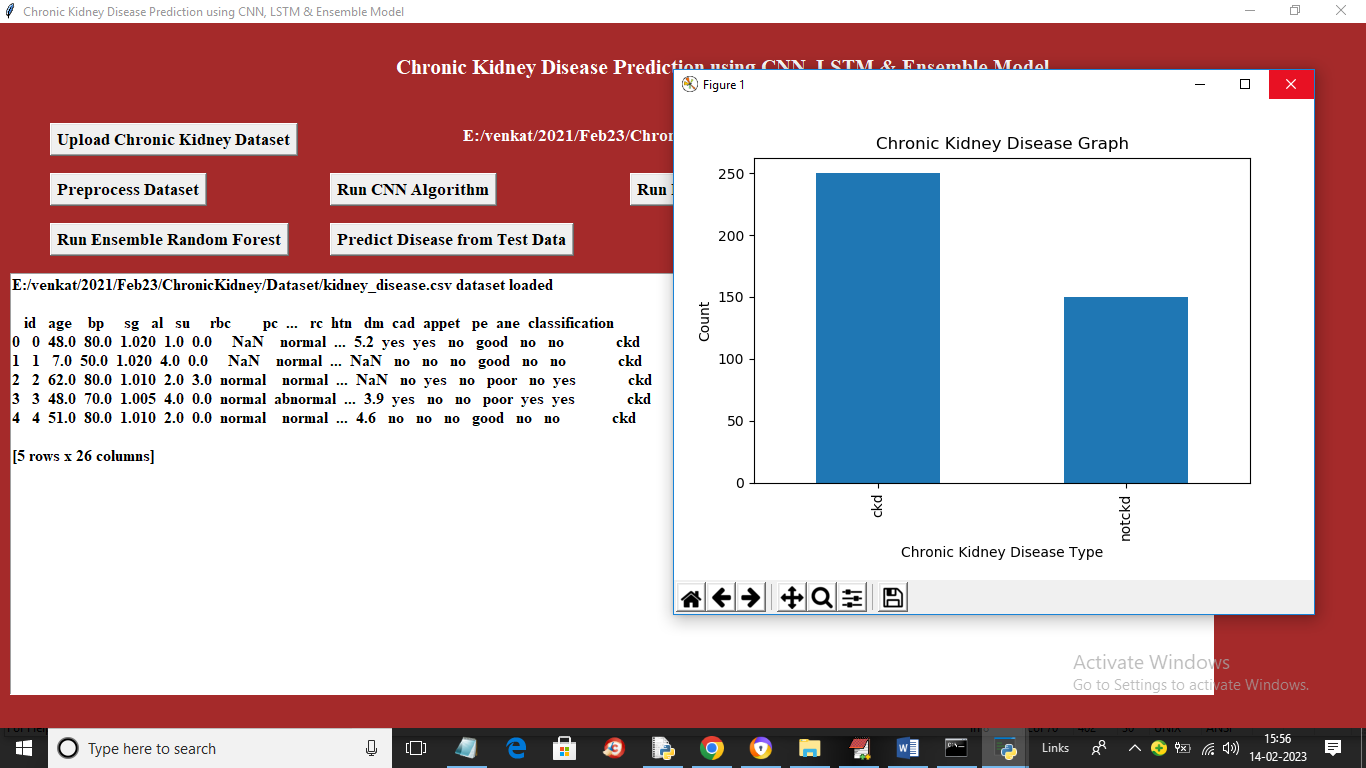
To run project double click on ‘run.bat’ file to get below screen



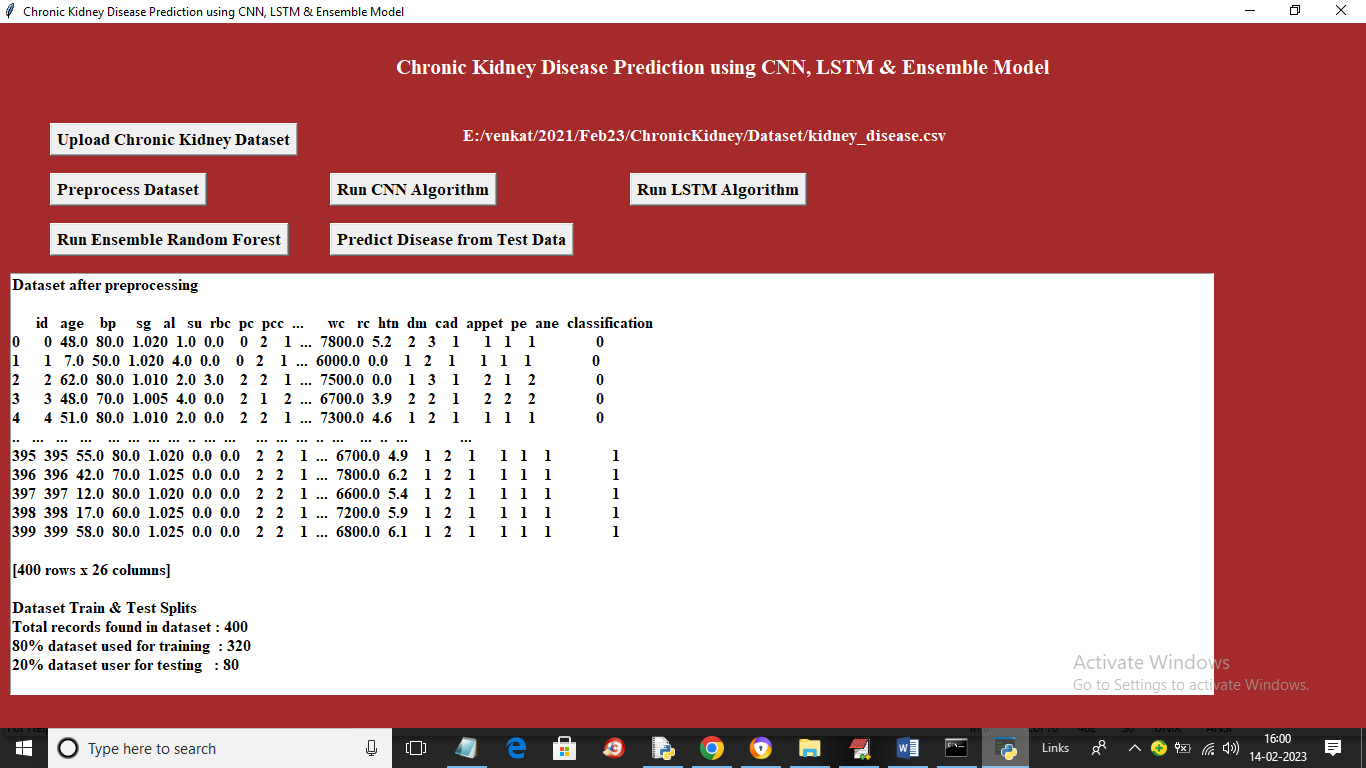
In above screen click on ‘Upload Chronic Kidney Dataset’ button to upload dataset and get below output screen



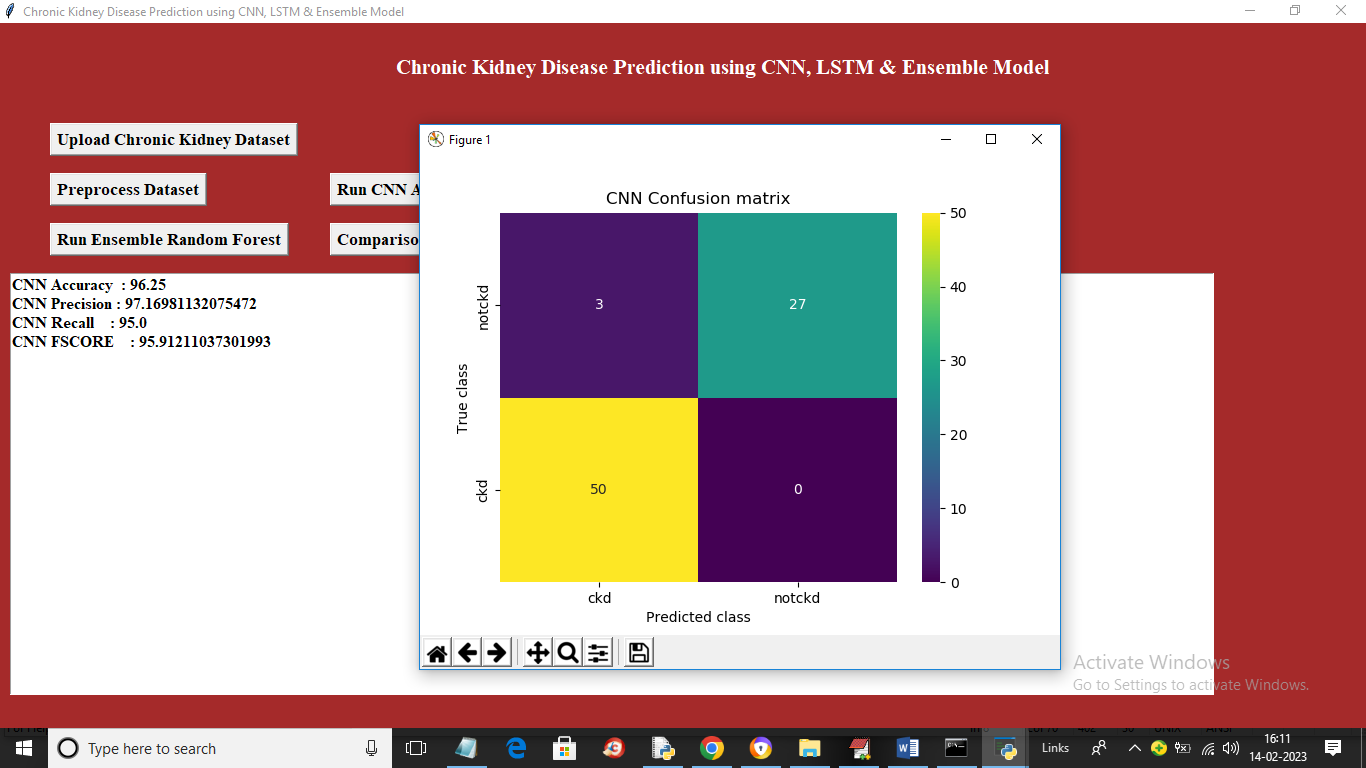
In above screen selecting and uploading dataset file and then click on ‘Open’ button to load dataset and get below output



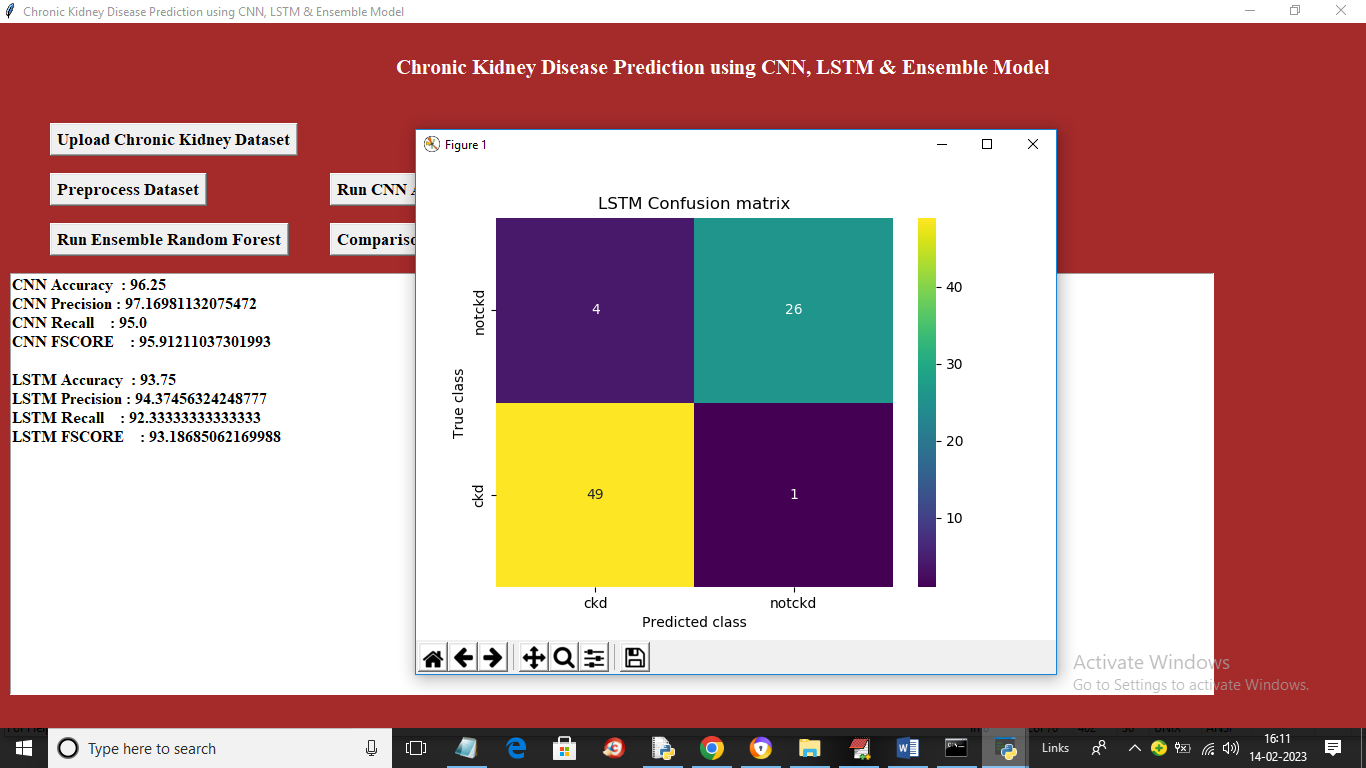
In above screen dataset loaded and in graph x-axis contains ‘CKD or NON-CKD’ labels and y-axis represents count and using this graph we are showing number of CKD and NON-CKD patients available in dataset. In above screen we can see dataset contains both numeric and non-numeric data and by applying label encoding class we will convert non-numeric data to numeric data by clicking on ‘Preprocess Dataset’ button and get below output



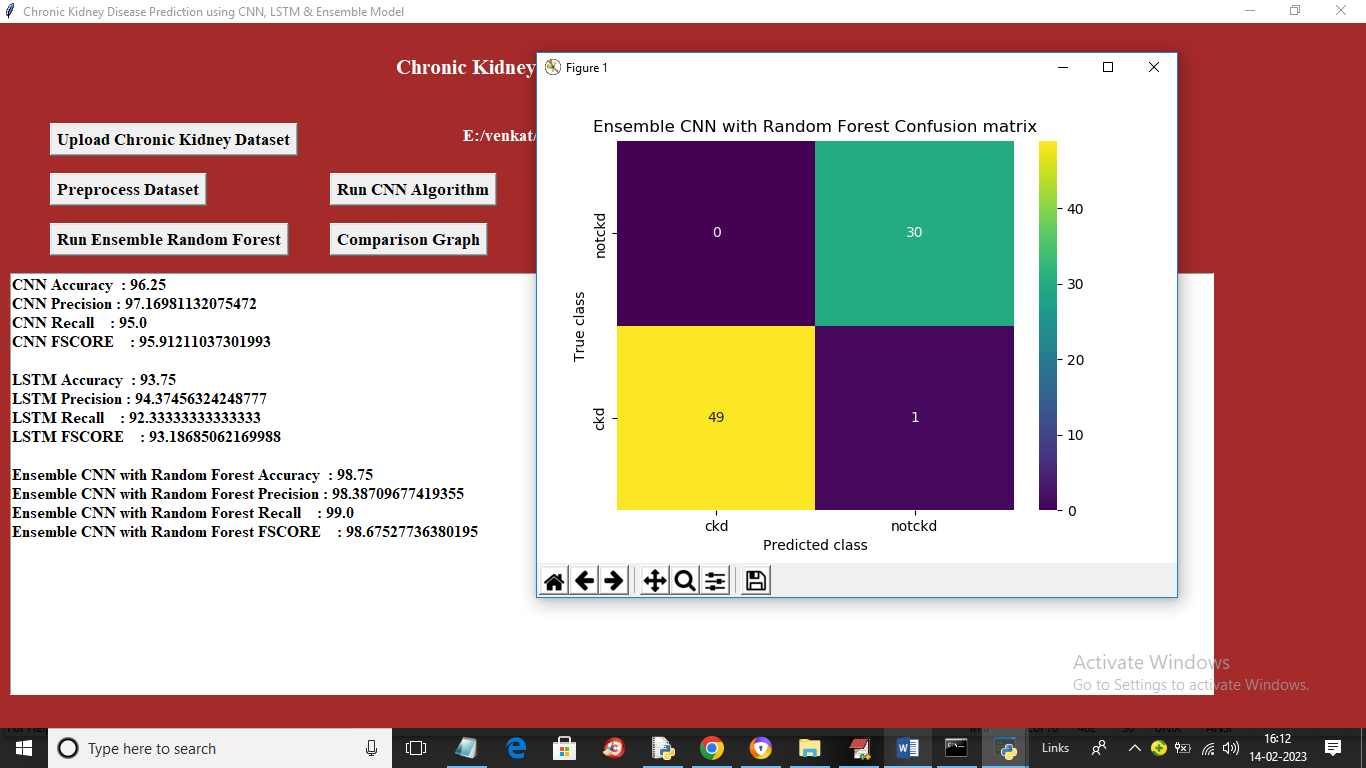
In above screen we can see all dataset values converted to numeric format and in last lines we can see dataset contains 400 records and then using 320 records for training and 80 records for testing. Now dataset is ready and now click on ‘Run CNN Algorithm’ button to train CNN and get below output



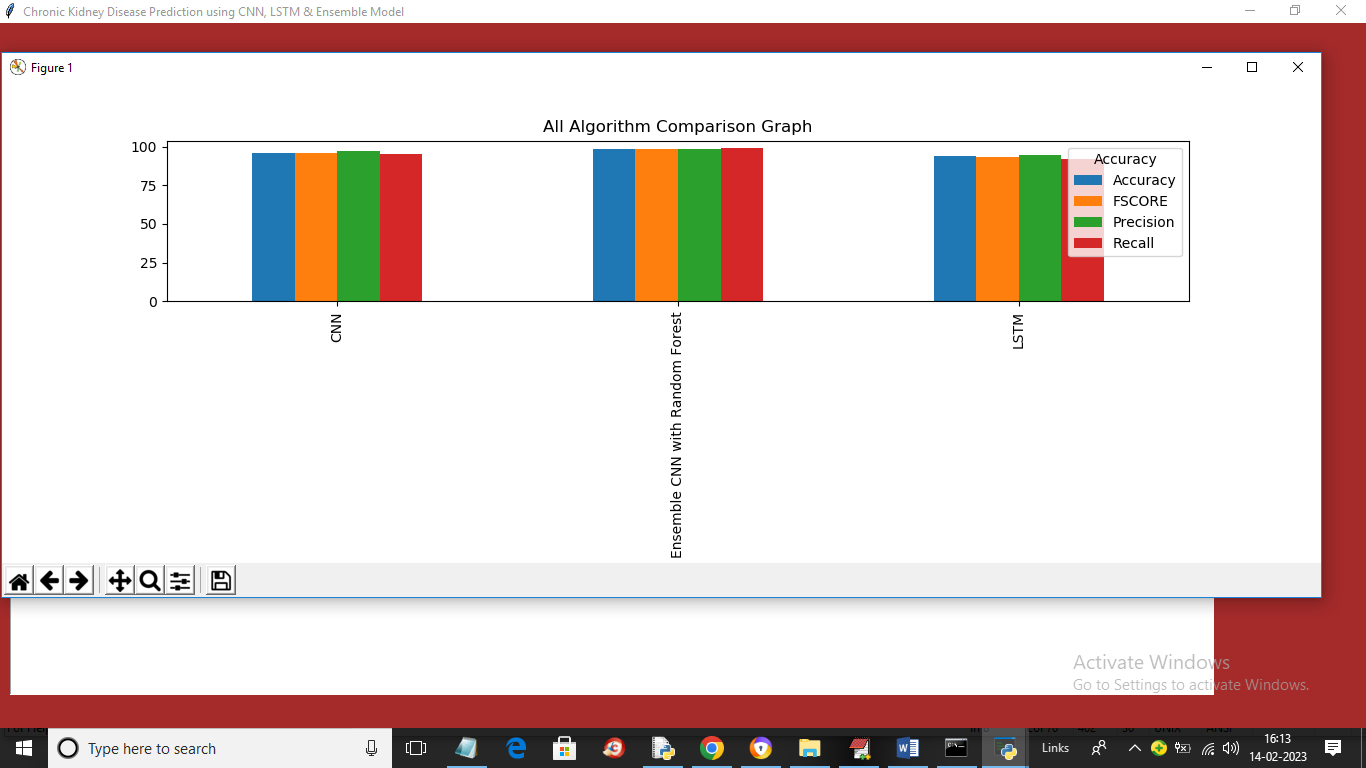
In above screen with CNN we got 96% accuracy and we can see other metrics also like precision, recall and FSCORE. CNN is one of the optimized algorithm so we may get accuracy between 94 to 100% for different runs. In above confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels and different colour boxes represents Correct prediction count and blue colour boxes contains incorrect prediction count which are only 3. Now close above graph and then click on ‘Run LSTM Algorithm’ button to train LSTM and get below output



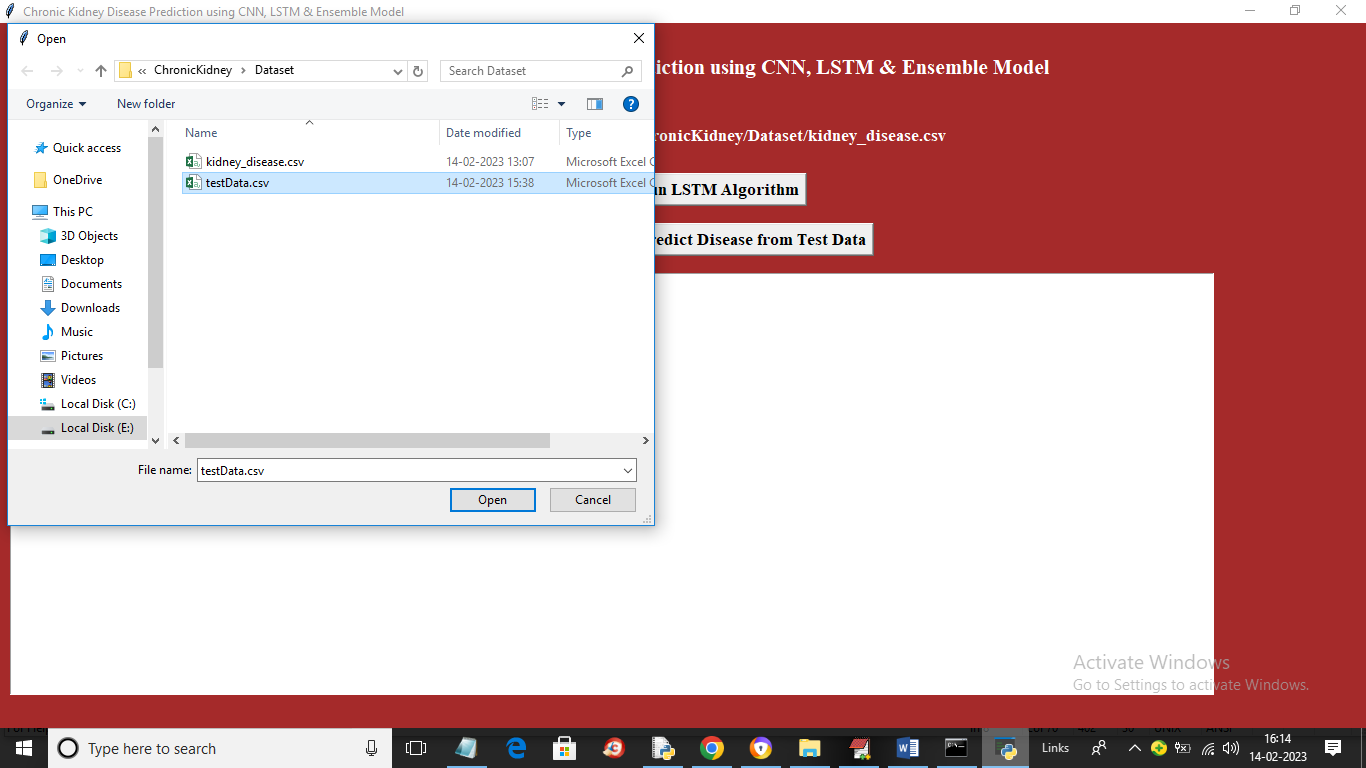
In above screen with LSTM we got 93% accuracy and now click on ‘Run Ensemble Random Forest’ algorithm button to train ensemble CNN with random forest and get below output



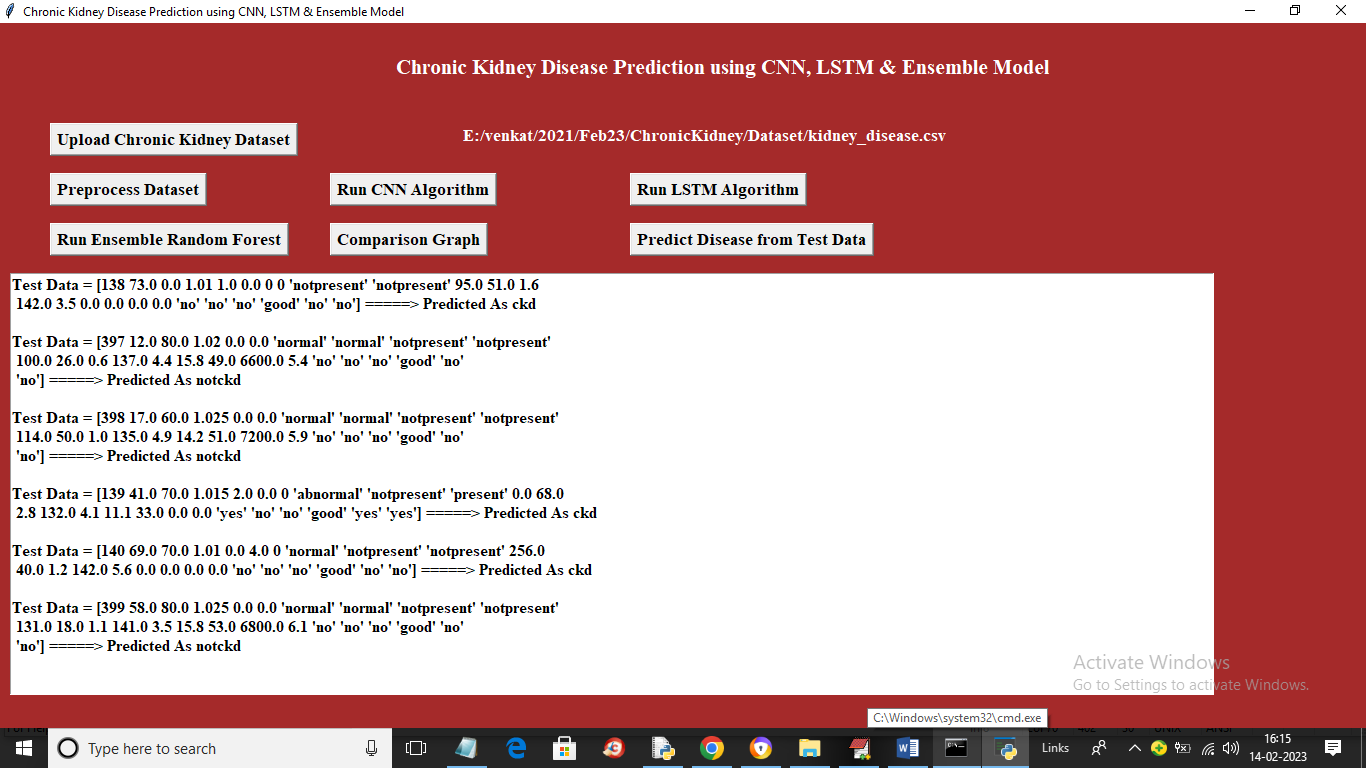
In above screen with Ensemble CNN Random Forest we got accuracy as 98% and we can see improvement in recall and FSCORE compare to CNN and LSTM. Now click on ‘Comparison Graph’ button to get below graph



In above graph x-axis represents algorithm names and y-axis represents accuracy and other metrics in different colour bar and in all algorithms Ensemble CNN with Random Forest got high accuracy and now close above graph and then click on ‘Predict Disease from Test Data’ button to upload test data and get below output



In above screen selecting and uploading testData.csv file and then click on ‘Open’ button to load dataset and get below output



In above screen in square bracket we can see test data and then after =🡺 arrow symbol we can see predicted values as ‘CKD or NOTCDK’