**Robust Intelligent Malware Detection Using Deep Learning**

Now-a-days to detect cyber-attack we are using static and dynamic analysis of request data. Static analysis is based on signature which we will match existing attack signature with new request packet data to identify packet is normal or contains attack signature. Dynamic analysis will use dynamic execution of program to detect malware/attack but dynamic analysis is time consuming. To overcome from this problem and to increase detection accuracy with old and new malware attacks author is using machine learning algorithms and evaluating prediction performance of various machine learning algorithms such as Support Vector Machine, Random Forest, Decision Tree, Naïve Bayes, Logistic Regression, KNearest Neighbours and Deep Learning Algorithms such as Convolution Neural Networks (CNN) and LSTM (Long Short Term Memory). In all algorithms CNN and LSTM giving better performance.

To implement this paper and to evaluate machine learning algorithms performance author is using binary malware dataset called ‘MALIMG’. This dataset contains 25 families of malware and application will convert this binary dataset into gray images to generate train and test models for machine learning algorithms. This algorithms converting binary data to images and then generating model so they are called as MalConv CNN and MalConv LSTM and other algorithm refers as EMBER. Application convert dataset into binary images and then used 80% dataset for training model and 20% dataset for testing. Whenever we upload new test malware binary data then application will apply new test data on train model to predict malware class. In dataset total 25 families of malware we can see and below are their names.

'Dialer Adialer.C','Backdoor Agent.FYI','Worm Allaple.A','Worm Allaple.L','Trojan Alueron.gen','Worm:AutoIT Autorun.K',

'Trojan C2Lop.P','Trojan C2Lop.gen','Dialer Dialplatform.B','Trojan Downloader Dontovo.A','Rogue Fakerean','Dialer Instantaccess',

'PWS Lolyda.AA 1','PWS Lolyda.AA 2','PWS Lolyda.AA 3','PWS Lolyda.AT','Trojan Malex.gen','Trojan Downloader Obfuscator.AD',

'Backdoor Rbot!gen','Trojan Skintrim.N','Trojan Downloader Swizzor.gen!E','Trojan Downloader Swizzor.gen!I','Worm VB.AT',

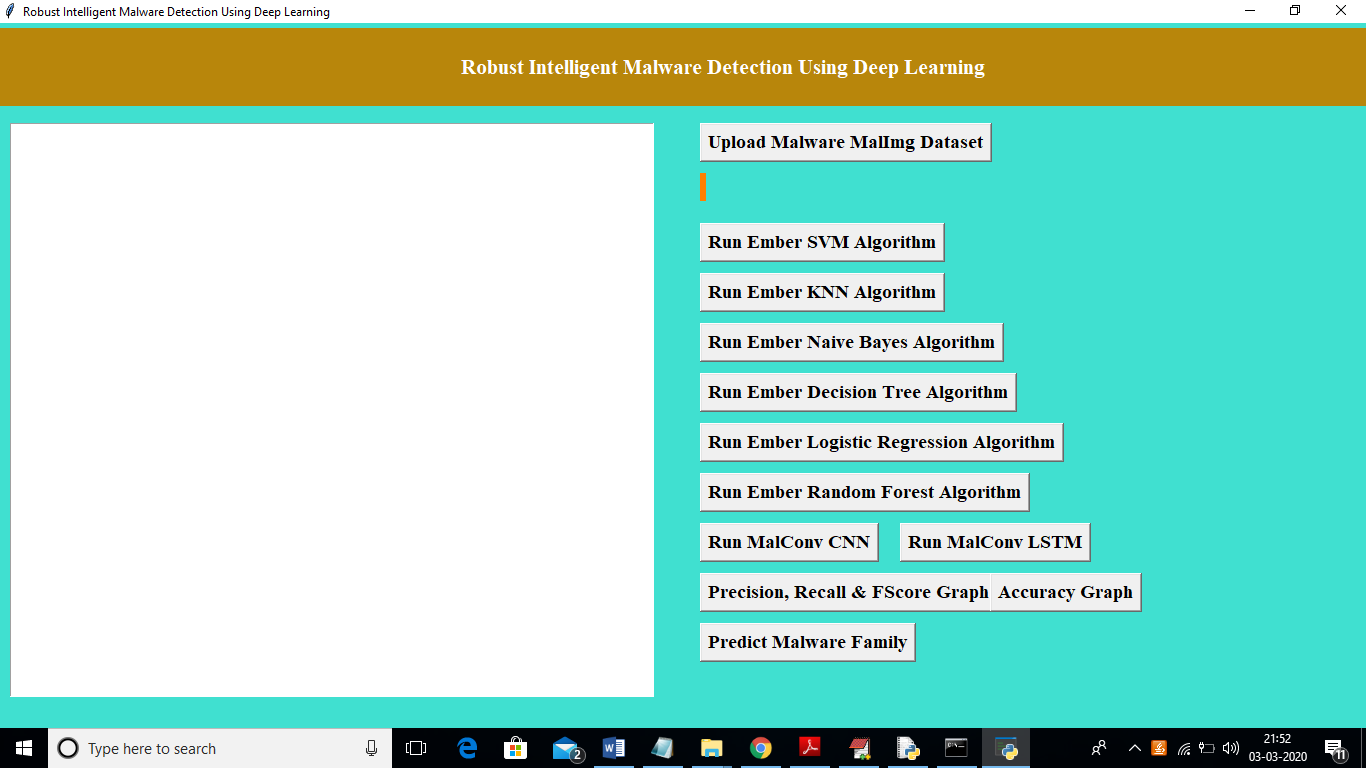
'Trojan Downloader Wintrim.BX','Worm Yuner.A'

All above names are the malware families

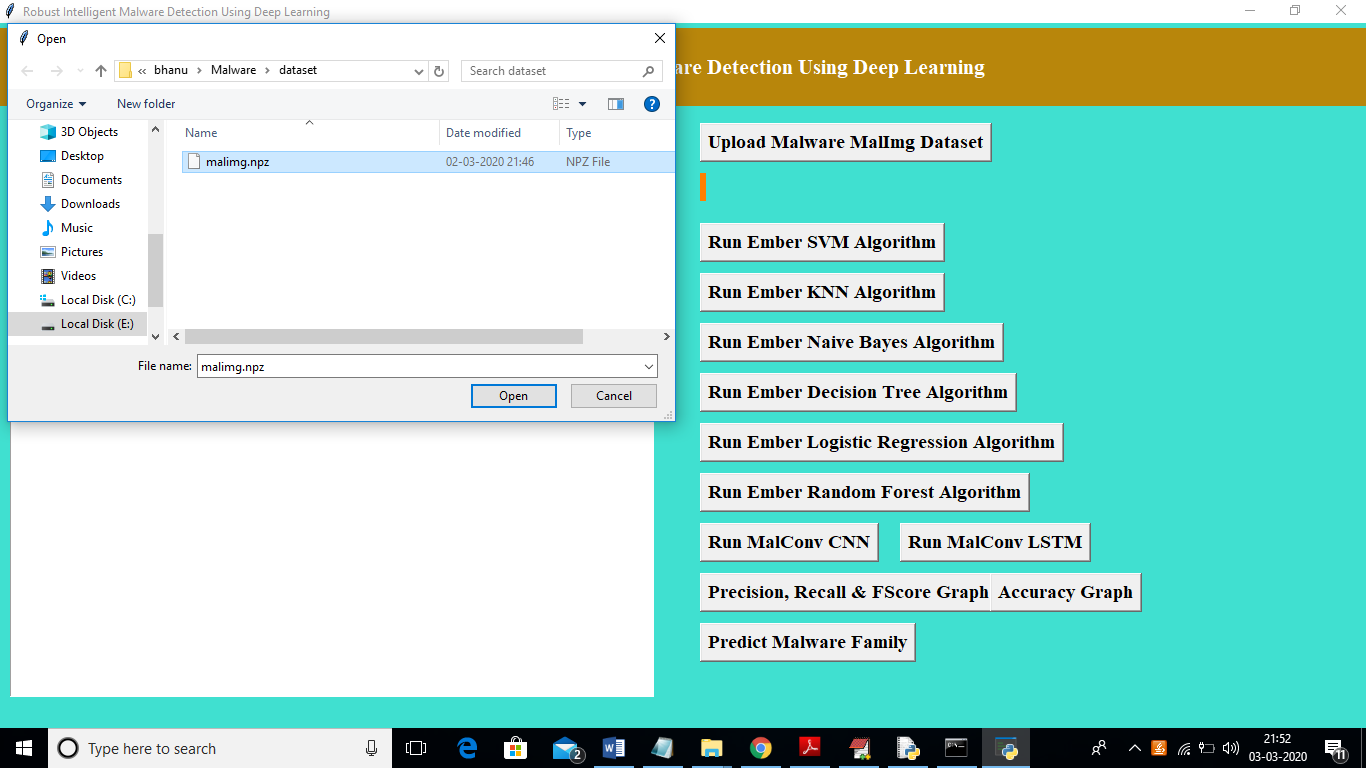
All new malware test files I saved inside images folder and you can upload this files to predict it class. All algorithms details you can read from paper.

Screen shots

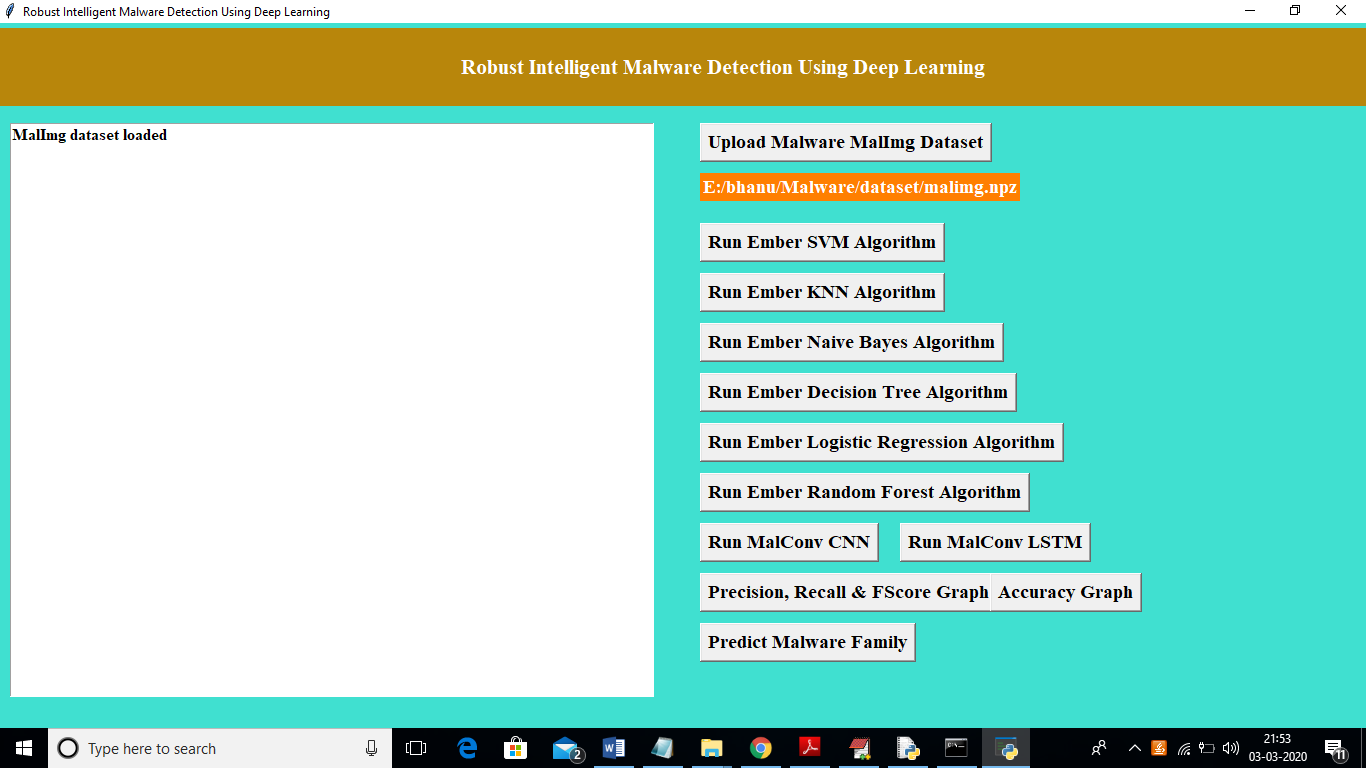
To run this project double click on ‘run.bat’ file to get below screen



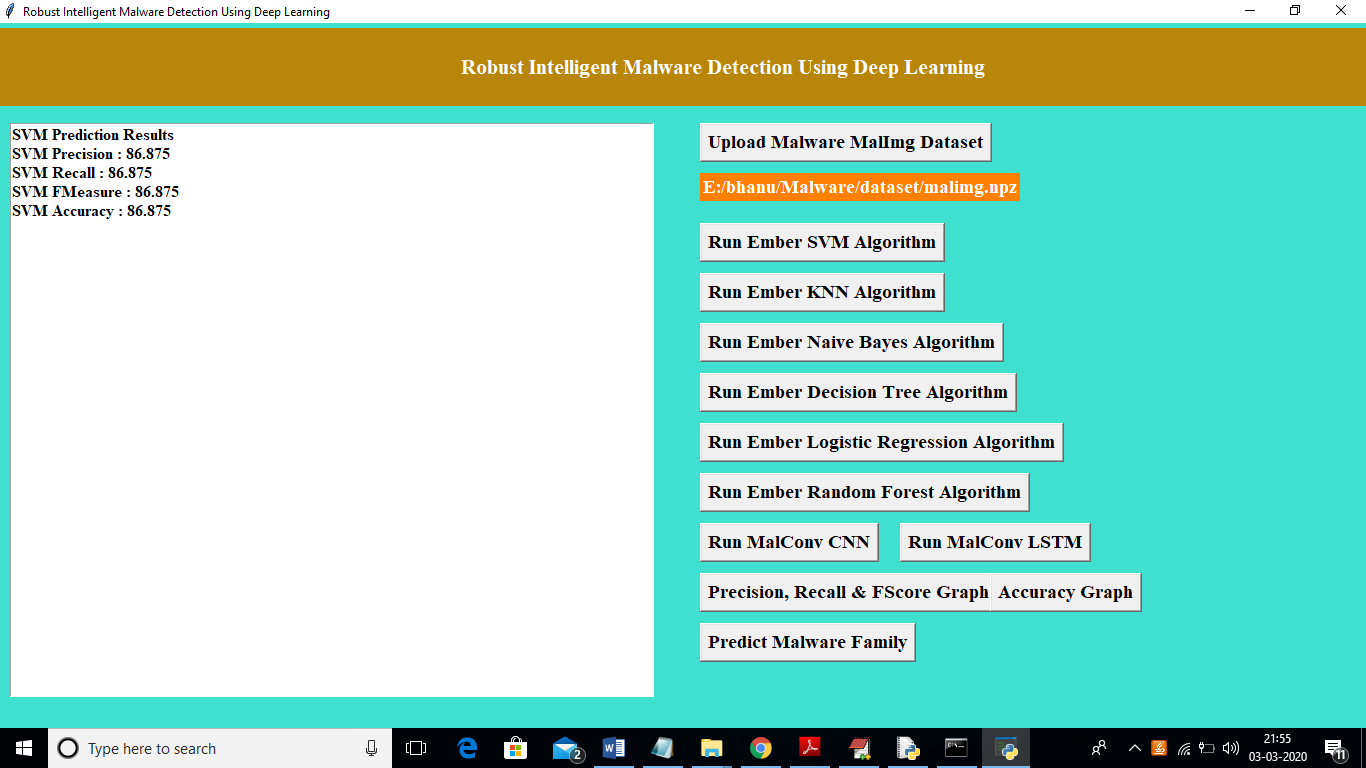
In above screen click on ‘Upload Malware MalImg dataset’ button to upload dataset



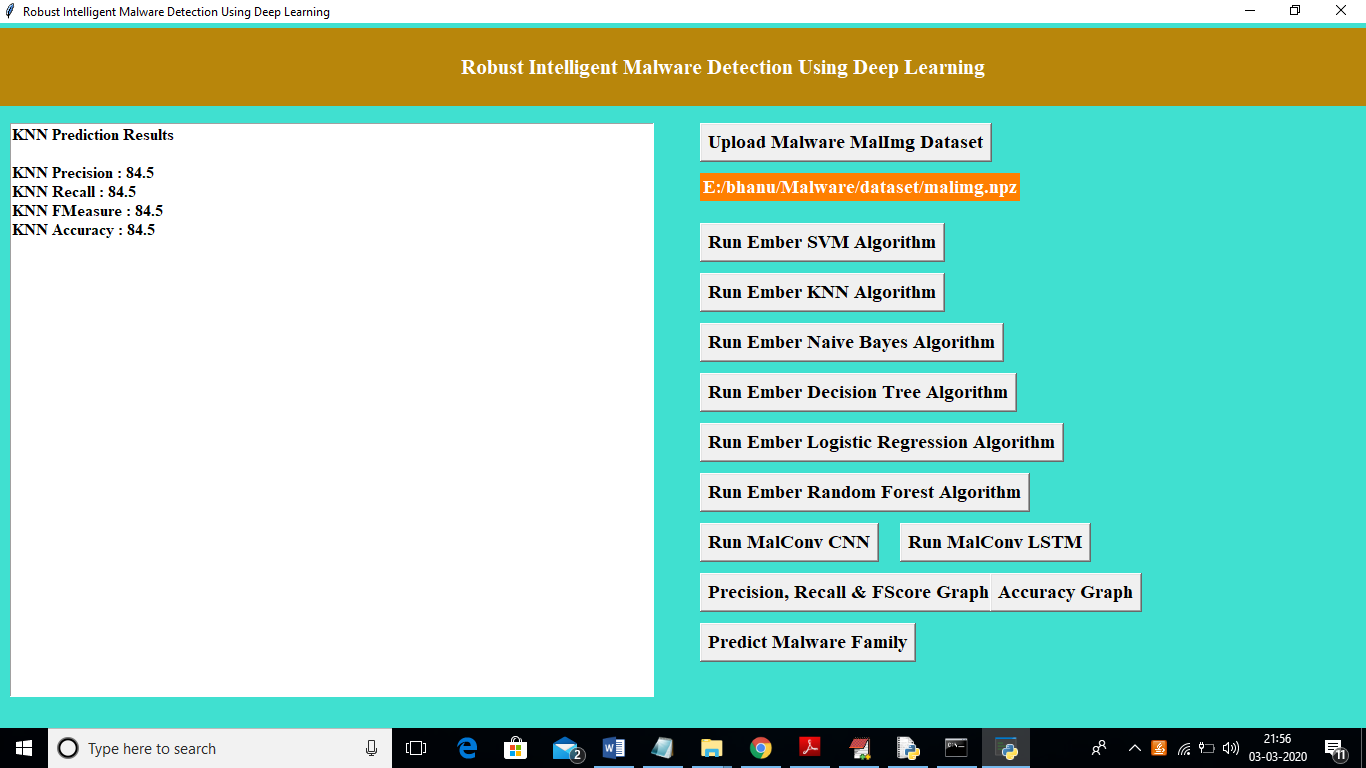
In above screen I am uploading ‘malimg.npz’ binary malware dataset and after uploading dataset will get below screen



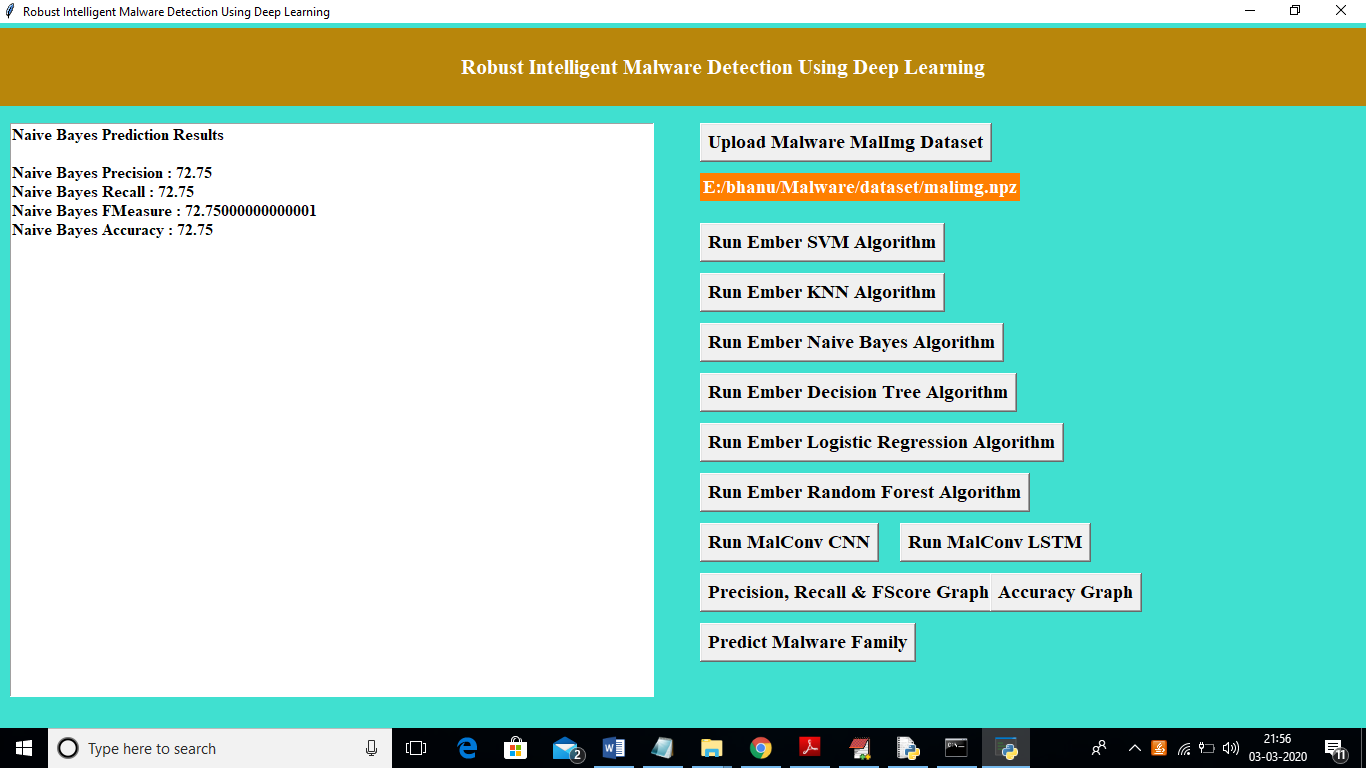
Now click on ‘Run Ember SVM algorithm’ button to read malware dataset and generate train and test model and then apply SVM algorithm to calculate its prediction accuracy, FSCORE, Precision and Recall. If algorithm performance is good then its accuracy, precision or recall values will be closer to 100.



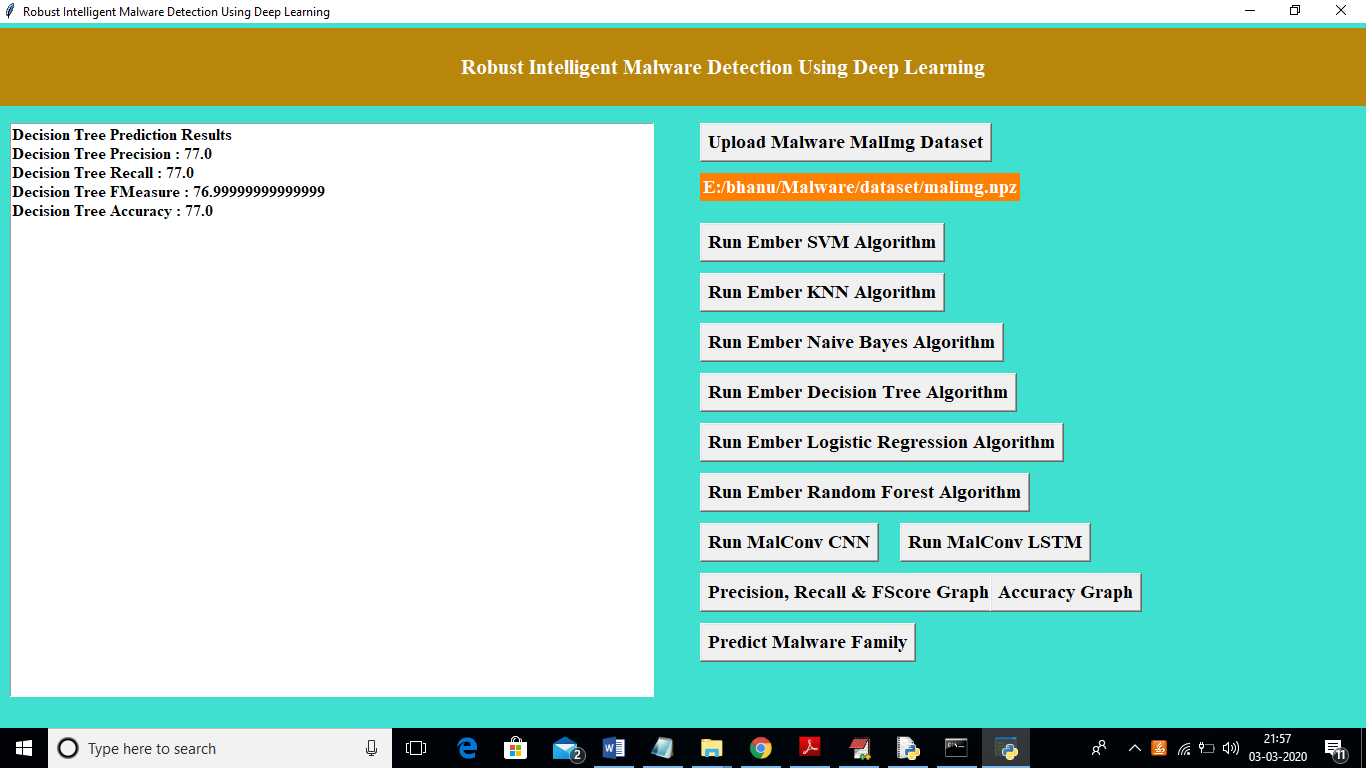
In above screen we got SVM precision, recall and fSCORE. Now click on ‘Run Ember KNN Algorithm’ button to get its performance



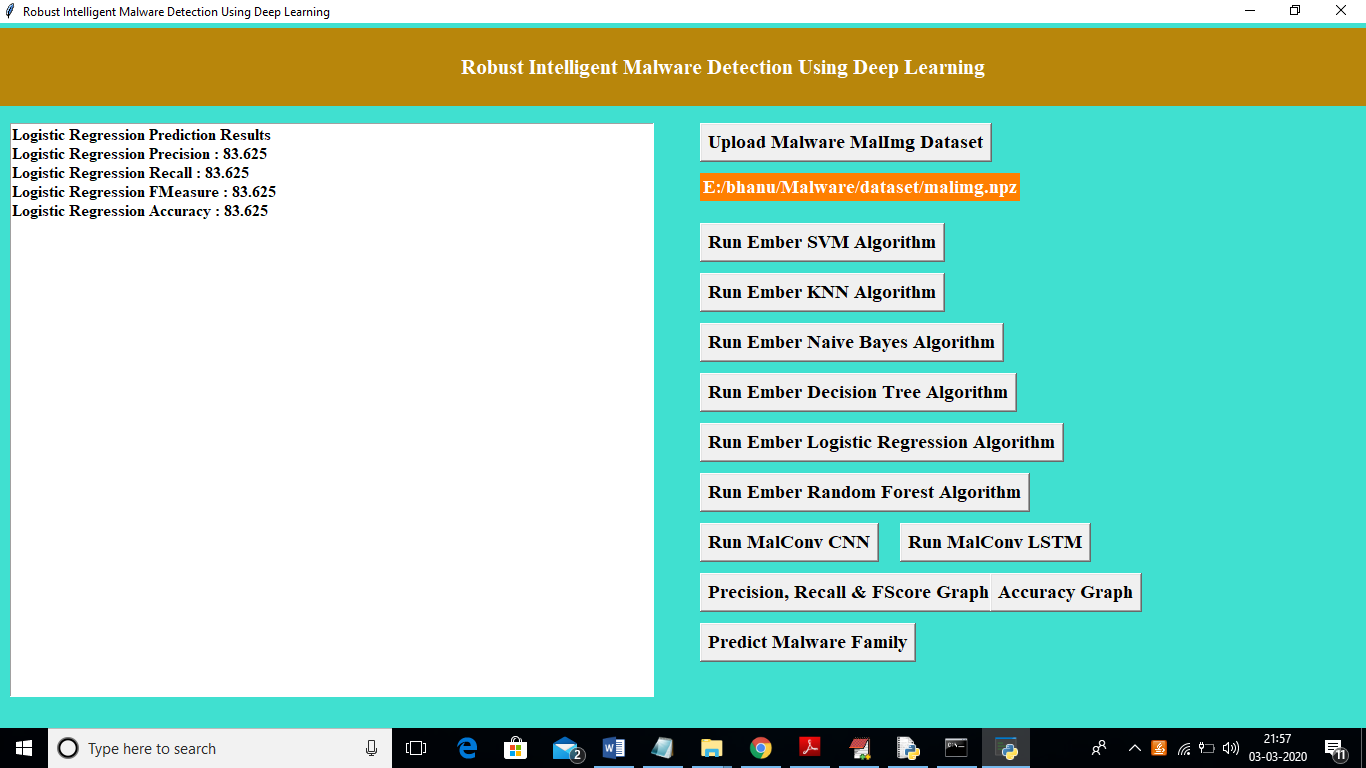
In above screen we got KNN details and now click on ‘Naïve Bayes’ button to get its performance details



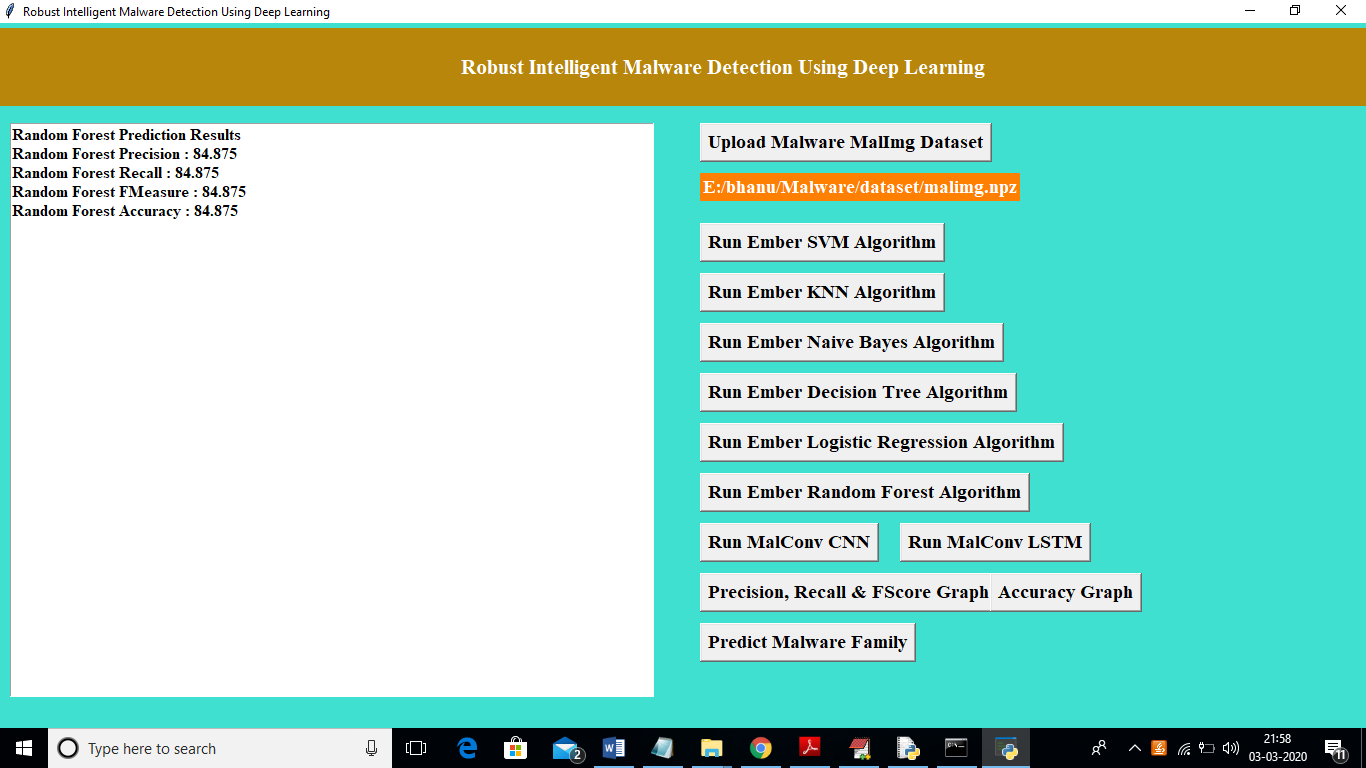
In above screen we got naïve bayes details and now click on ‘Decision Tree’ button to get its performance details



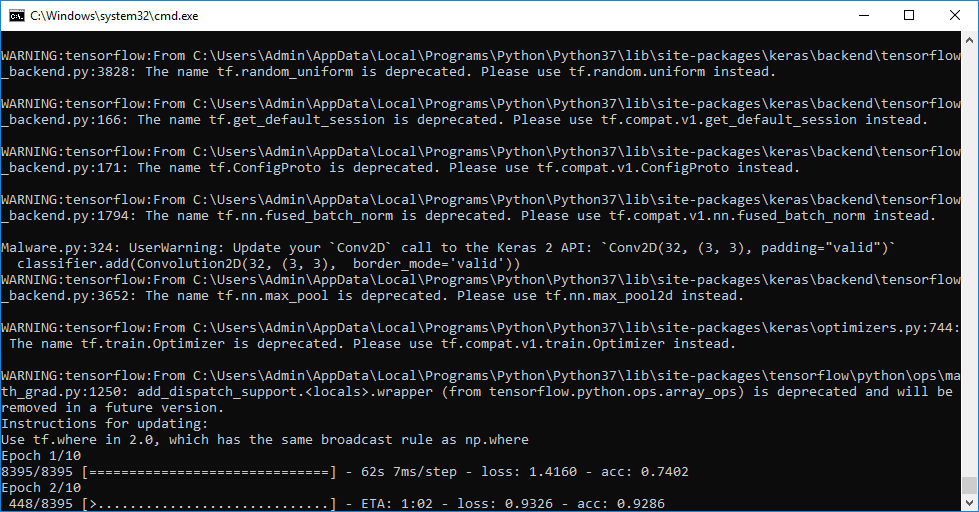
In above screen we got decision tree details and now click on ‘Logistic Regression’ button to get its details



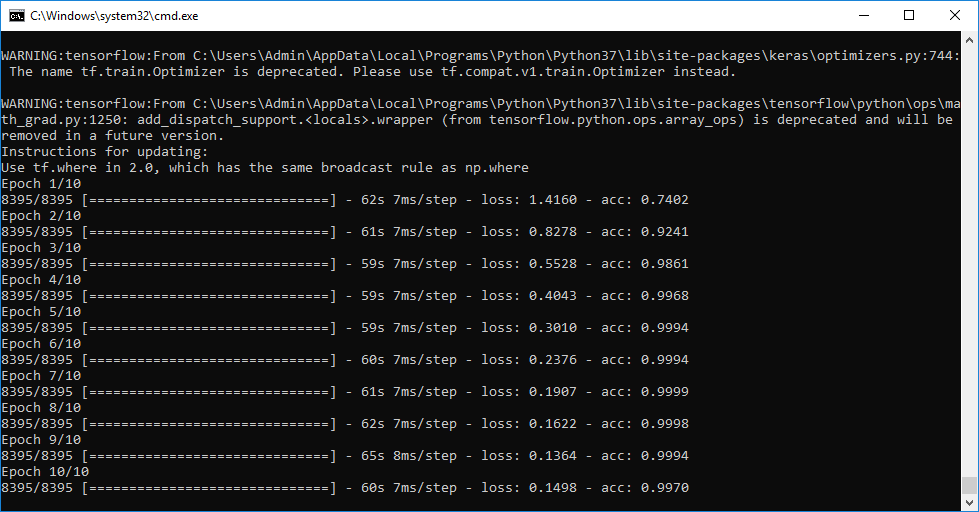
In above screen we got logistic regression details and now click on ‘Run Random Forest’ button to get its performance



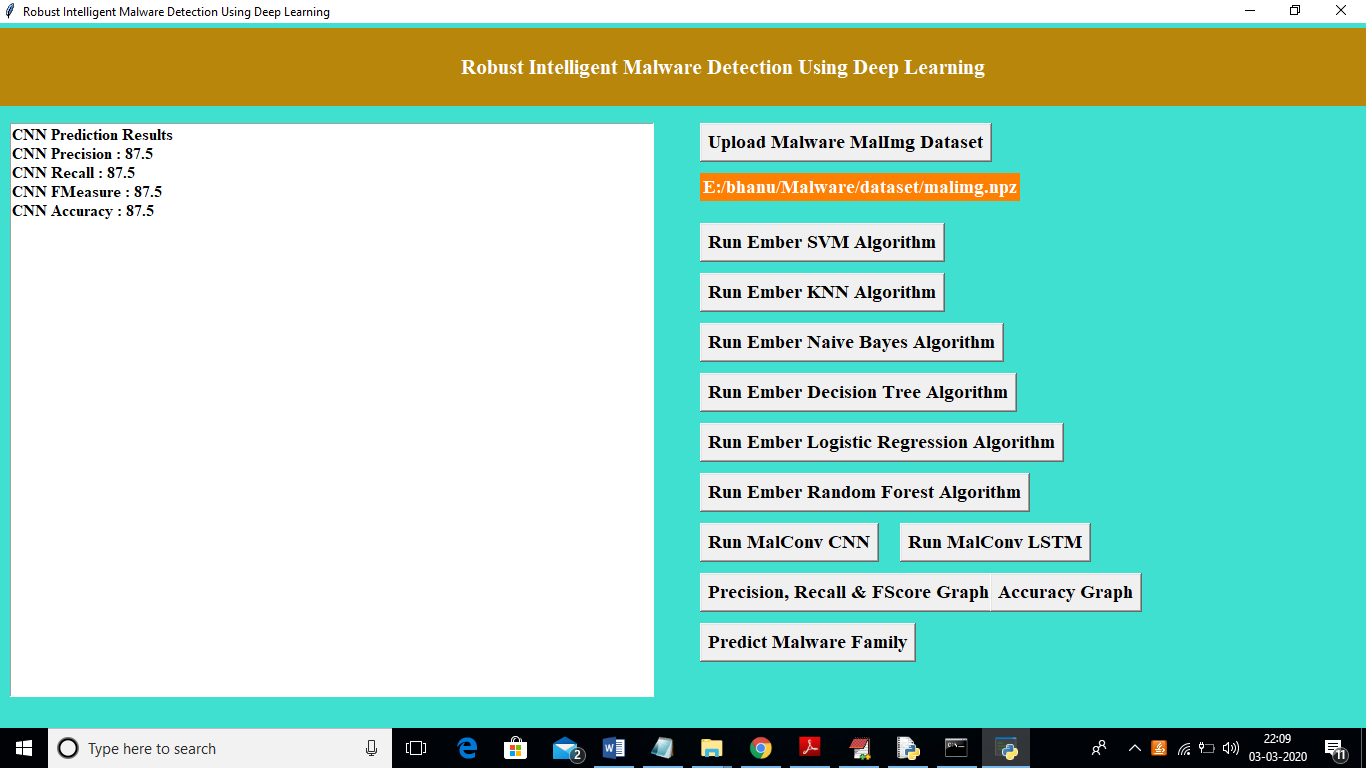
In above screen we got random forest details and now click on ‘Run MalConv CNN’ button to get its performance details. CNN may take 10 minutes to complete execution and u can check its ongoing processing in black console



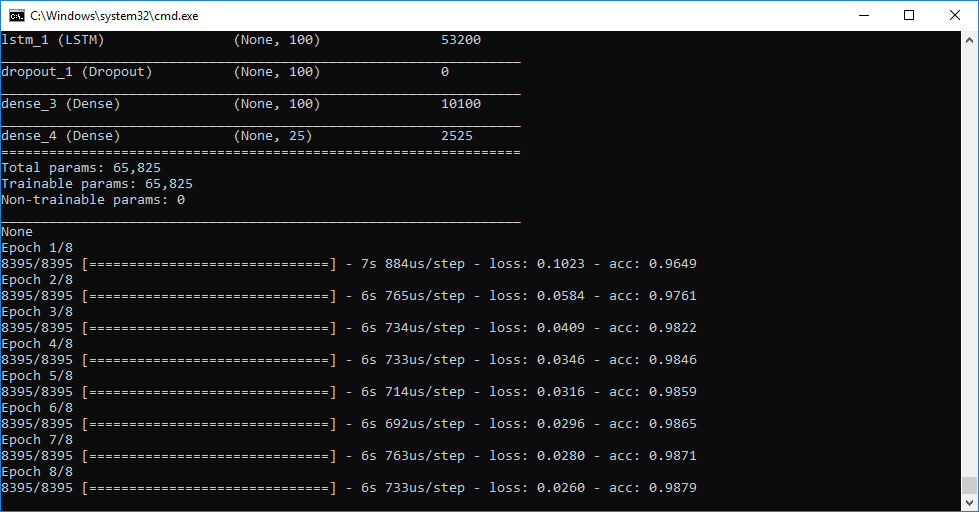
In above console it will take 10 epochs iteration and for each iteration it calculate accuracy for 8395 malware data. So u need to wait till all 10 epochs completed then u will get its performance details

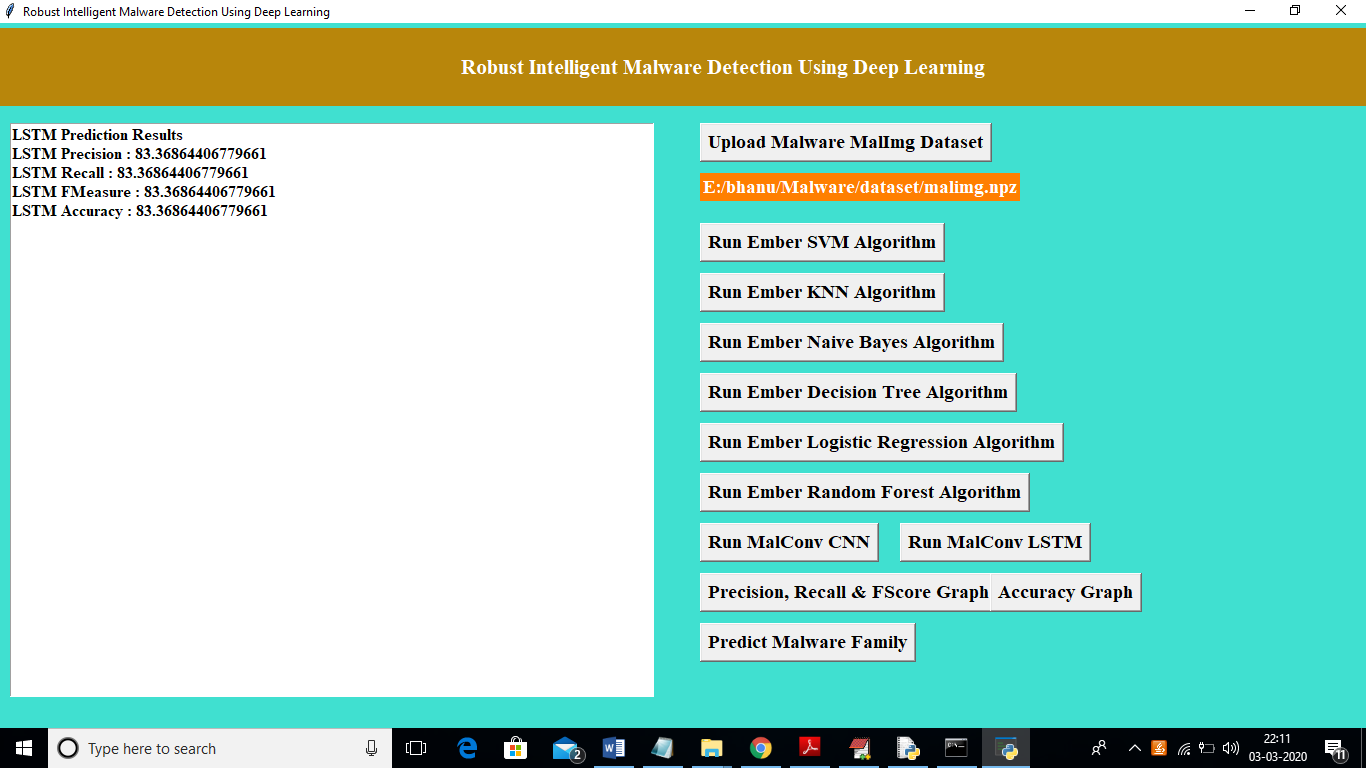


In above screen we can see CNN complete all 10 epochs and after that we will get accuracy details in main screen



In above screen we got CNN performance values and now click on ‘Run MalConv LSTM’ button to run LSTM algorithm. Similar to CNN LSTM also take 10 minutes and u can see ongoing process in below screen





In above screen we can see LSTM details now click on ‘Precision, Recall & FScore’ button to get comparson graph for all metrics and all algorithms

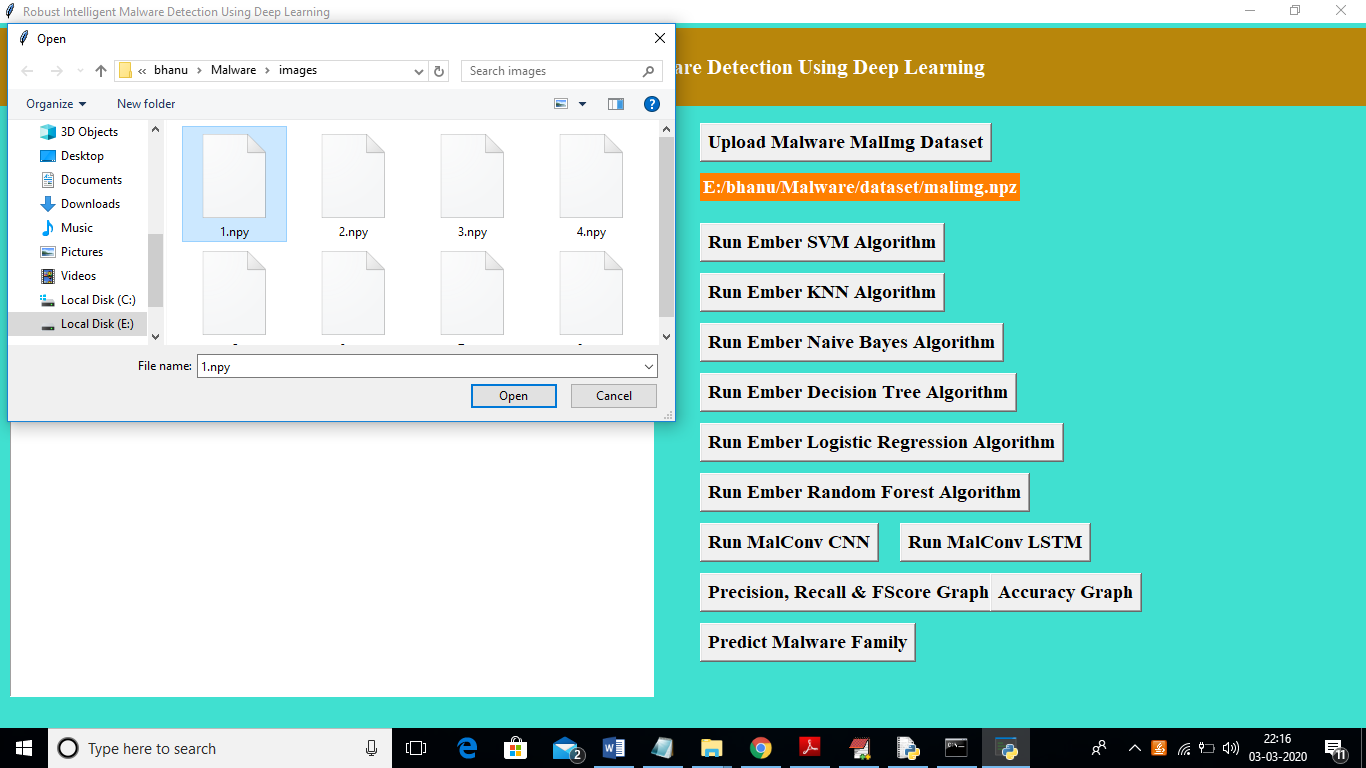


In above screen we can see precision graph for all algorithms and CNN get better performance. In above graph x-axis represents algorithm name and y-axis represents precision value and now close above graph to get recall graph

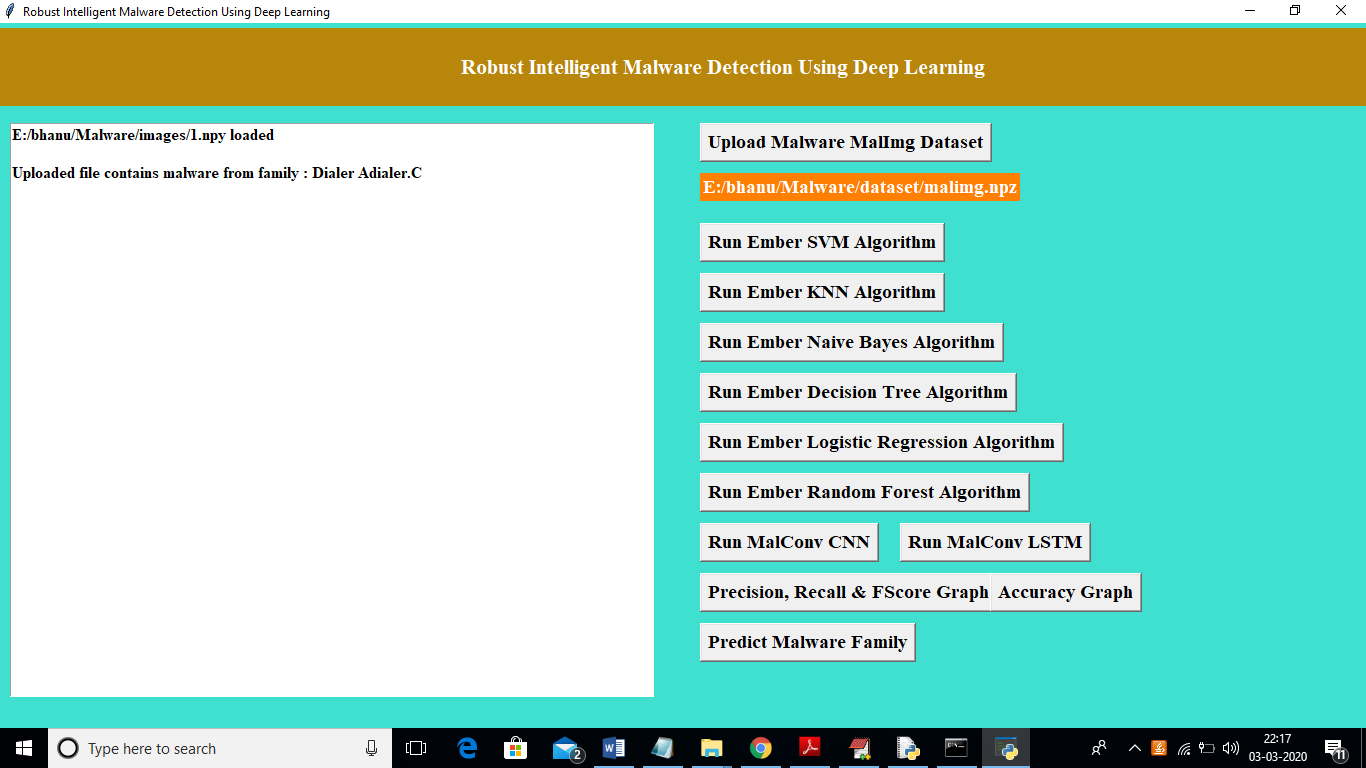
Now click on accuracy button to get accuracy graph



Now click on ‘Predict Malware Family’ button and upload binary file to get or predict class of malware



In above graph I am uploading one binary file called 1.npy and below is the malware prediction of that file



In above screen we can see uploaded test file contains ‘Dialer Adialer.C’ malware attack. Similarly u can upload other files and predict class