**Introduction:-**

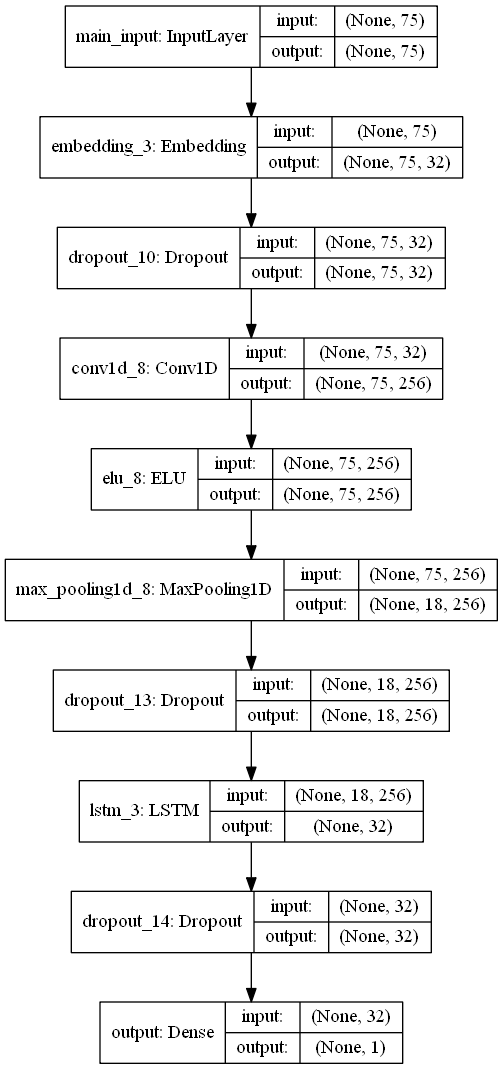
Because of the colossal growth of internet, most of the users have changed their preference from traditional shopping, banking etc. to online mode. This paved the way for a lot of cybercrimes including phishing. The attackers try to extract sensitive/personal details such as user ID, passwords and debit card/credit card information by disguising themselves as reliable websites. Identifying whether the Uniform Resource Locator (URL) of a website is legitimate or phishing is a difficult task because it exploits the user’s vulnerabilities. Although many products are available for detecting phishing websites, they are just making use of heuristic approach and black lists and hence they can’t prevent phishing in a more effective way. Then, a lot of machine learning techniques have been proposed. But, feature extraction was a time consuming process. In this work, an ensemble model that uses LSTM and CNN has been implemented.

**Novelty:-**

Most of the existing work has been carried out using Convolutional Neural Networks initially, and then there had been many works using Convolutional Neural Networks with either word level or character level features or both of them. Then, they used Recurrent Neural networks with either LSTM or BiLSTM. In this, an ensemble model combining Convolutional Neural Networks and LSTM has been used with effective hyper parameter tuning to achieve the best performance. And, one more strength of this work is that it has been carried out with a massive data set containing nearly 2,00,000 URLs, which is balanced also.

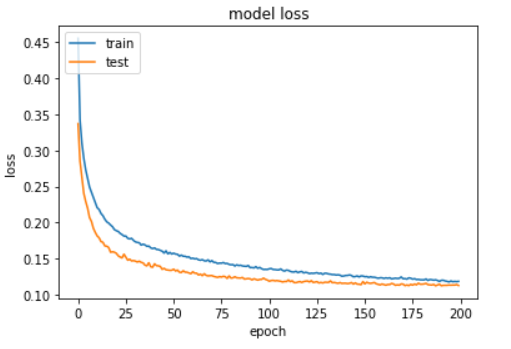
**Model:-**

A deep learning solution to phishing website detection has been implemented. The solution uses an ensemble model, meaning it combines LSTM and CNN to detect phishing websites. The dataset is balanced perfectly and has approximately 2 lakh URLs. The data distribution plot is also shown in the project. The URLs are preprocessed. Then, they are fed into the input layer of the model. The dimension of an input instance is 75x32. The solution has been achieved after testing the model by tuning the hyper parameters and choosing the best performing model from it. The model has been trained for 200epochs. The model diagram is shown below.



**Learning curve:-**

The learning curve is shown below. It can be observed that as the number of epochs raises, the validation error decreases gradually. After that the validation error increases. That’s the reason of fixing 200 epochs.



**Performance Evaluation:-**

The achieved accuracy is 96% and the precision for the legitimate and phishing class are 95% and 97% respectively which is far better than the existing solutions.