Student Name: Vinay Kotadiya

Student ID: 23101562

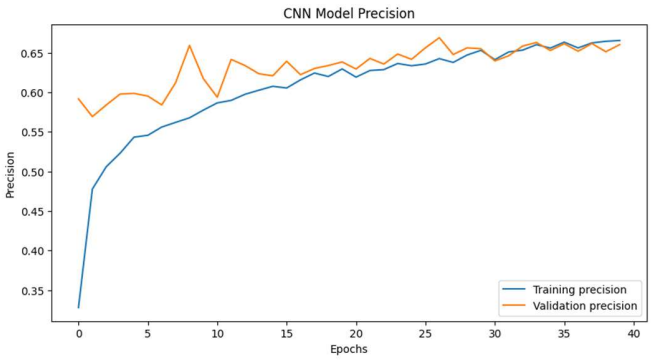
Introduction

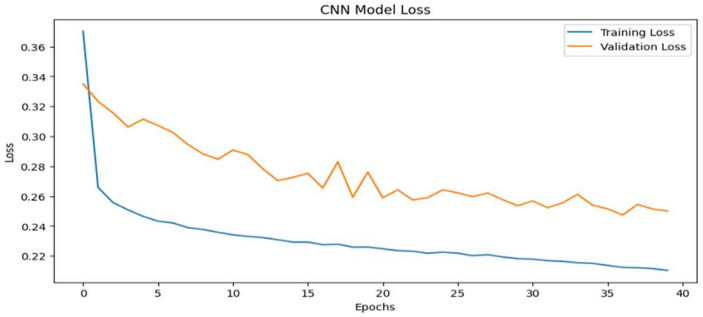
Deep learning has also, in recent years, significantly improved automatic classification of content in the various kinds of media. One such hard problem is multi-label classification where an object or instance can be in more than one class simultaneously. For this project, we address movie genre classification, where a single movie will be assigned multiple genres such as Action, Drama, and Thriller. To this end, we employed two different types of neural networks. A Convolutional Neural Network (CNN) was trained on movie posters, and a Long Short-Term Memory (LSTM) network was trained on movie summaries. The goal was to train and test both models, compare and analyze their performance, and make observations about their strengths, weaknesses, and areas of improvement. This report summarizes the complete experimental procedure, results, in-depth analysis, and final conclusions.

Model Training Process

Every one of the models was developed, trained, and validated individually. In the CNN, posters were normalized and resized to a standard size. In the LSTM, overviews were tokenized, padded, and converted into word sequences. Genre labels were encoded as multi-hot vectors for both of the models. Binary cross-entropy loss was utilized for multi-label classification, and the Adam optimizer was utilized. Model checkpointing and early stopping on validation loss were used to preserve the best model. Models were loaded prior to evaluation to ensure use of the optimum version. Both models were tuned heavily to prevent overfitting and maximize generalization to new movies.

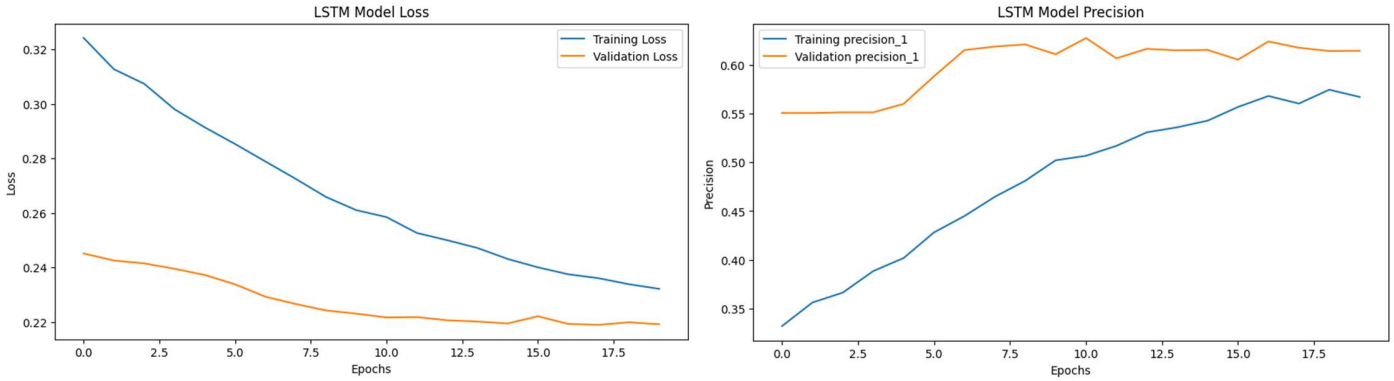
CNN Model Training and Validation





The CNN model was trained for over 40 epochs, and training and validation performance both improved linearly. Training loss started off around 0.37 and decreased gradually towards around 0.21, and validation loss also decreased and leveled off around the 20th epoch, which shows the model was picking up well without overfitting. Training accuracy increased significantly from 0.33 to around 0.67, and validation accuracy remained close to the training accuracy throughout training. This implies that the model generalized well to new samples. Generally, the training of the CNN model was successful with good accuracy and stability.

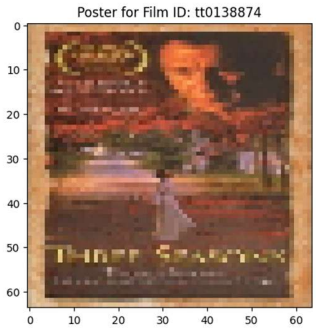
LSTM Model Training and Validation



The LSTM model trained for 20 epochs, with its performance constantly improving during training. The training loss started at approximately 0.32 and showed a constant decreasing trend epoch by epoch, finishing with a lower level at the end of the training. The validation loss started at about 0.24 and decreased steadily as well, with a constantly lower level than the training loss during the training. This steady decline in validation and training losses is an indicator that the model was learning properly with no sign of overfitting. Validation performance was consistent and steady, which is a sign of good generalization to new data. Overall, the LSTM model was behaving well and stable during training.

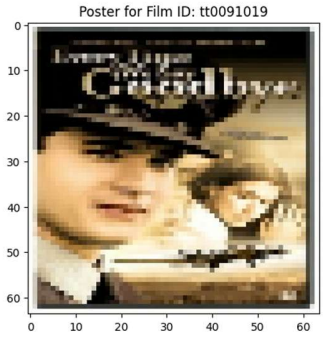
Critical Evaluation

Both models had clear strengths. Standard genres like Action, Comedy, Drama, and Fantasy were captured very well. LSTM was assisted by the fact that summaries incorporate genre-defining terms like battle, romance, or adventure that made it quite easy for LSTM to predict accurately. CNN performance was more dependent on the visual cues and was occasionally confused by dark or unclear poster graphics. But CNN worked well for genres where the poster images were distinctive, i.e., Science Fiction and Animation. Confusion matrices and metric curves indicate both models worked very well but the LSTM model worked a bit better than the CNN model consistently. In conclusion, a fusion of both poster images and text summaries together in a hybrid model would be able to improve performance in future work. This project clearly demonstrates the power of deep learning for multi-label classification tasks and the importance of choosing the right modality for the problem being tackled.



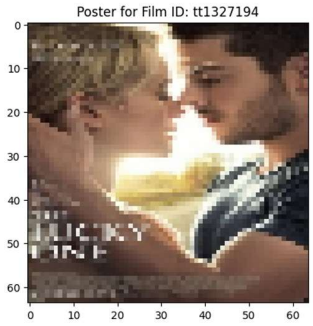
Overview 1:

In the first summary, CNN and LSTM models both correctly placed "Drama" at their top predictions. CNN also predicted "Romance" and "Comedy," which suggests lightness, and LSTM towards "Comedy" and "Crime." Although the true genre was simply "Drama," both models captured the emotional gravity of the story. Their minor difference is a sign of possible multiple interpretations of the mood of the story.



Overview 2:

In the second summary, both CNN and LSTM models predicted "Drama" and "Romance" in their top three genres. Their predictions overlap nearly exactly with the true genres of "Drama," "Romance," and "War." Both models did, however, miss the "War" element, which is very central to the setting of the tale. Nonetheless, the romantic and emotional focus was correctly captured by both models.



Overview 3

For the third prediction, CNN had predicted "Drama," "Romance," and "Comedy," as expected of the emotional and upbeat trajectory set out. LSTM's prediction was cut off but had started on a similar trend. The actual genre wasn't completely listed, but "Drama" and "Romance" seem fitting as per the storyline. In most cases, CNN was correct with the emotional mood, and LSTM showed partial agreement before cut off.

Conclusion

In this project, we have trained and tested two different deep learning models for multi-label movie genre classification based on posters and summaries. CNN and LSTM models produced notable results and showcased good ability to generalize on unseen data. The LSTM model performed slightly better than CNN, especially for classes where textual information played a vital role. Through close inspection of prediction samples and metric curves, we have identified strengths and weaknesses of each approach. Future work could explore combining the two modalities to achieve even stronger results.