In this section, we will delve into the process of constructing a CNN model from scratch to determine its effectiveness in identifying the genre of movie posters. We will begin by discussing the data-processing steps we undertook to prepare our dataset.

Firstly, we noticed that the range of movie release dates spanned from 1874 to 2033, with some posters created well in advance of the actual release date. To ensure a more relevant dataset, we chose to exclusively utilize posters released after 1970, resulting in a total of 424,102 images. Additionally, we acknowledged the close similarities between horror and thriller movies, making it difficult to distinguish between the two genres. Both genres frequently rely on elements such as suspense, tension, and unexpected plot developments, and incorporate elements of fear, danger, and psychological manipulation. To streamline our analysis, we decided to focus solely on the horror genre. Furthermore, due to the limited number of available images, we excluded the western genre from our analysis. As a result, our dataset narrowed down to 9 genres: 'Action', 'Animation', 'Comedy', 'Documentary', 'Drama', 'Horror', 'Music', and 'Romance'.

Moving on to our CNN model, it consists of 12 convolutional layers, with an input image size of (345,230,3). The initial 9 layers have 32 filters, while the last 3 layers have 64 filters, all with a kernel size of (3, 3), ReLU activation, and padding to maintain the input and output size. Instance normalization is applied after each convolutional layer to enhance the learning process, and max-pooling layers are utilized for down sampling the feature maps. Following the flattening process, batch normalization is applied, and the data is passed into a dense layer with 32 units and ReLU activation. Dropout regularization with a rate of 0.25 is then employed to prevent overfitting. In the subsequent layer, batch normalization is applied again, and the data is passed through another dense layer with 64 units and ReLU activation. Dropout regularization with a rate of 0.5 is applied at this stage. Finally, a dense layer with the number of genres to classify and softmax activation is added, producing predicted probabilities for each genre. The model is compiled using the Adam optimizer, categorical cross-entropy loss, and accuracy as the evaluation metric.

For training and testing, we utilized a dataset consisting of 9000 images, with 7200 images used for training and 1800 images for testing. Each genre has an equal number of images to maintain balance. To ensure consistency, we set a random seed of 4749, as it yielded the best results across multiple trials. Initially, our aim was to classify 5 genres, but we encountered disappointing performance, with low accuracy, F1-score, and ROC curve indicating worse performance than random classification.

We hypothesized that the lack of consideration for time differences might contribute to this outcome. To test our hypothesis, we grouped movie poster release dates by decades, generating 6 indicator functions. These indicators were combined with the output of the flattening layer and fed into a new flattening layer, which in turn was connected to the dense layers for genre prediction. To enhance prediction performance, we narrowed our objective down to classifying only 3 genres.

To address this issue, we hypothesized that the lack of consideration for time differences might contribute to the outcome. We examined the evolution of movie poster design and found significant changes in color themes and aesthetics over time. Figure 2 showcases the evolution of movie poster design from the 1910s to the 2010s. Each poster represents a different decade, starting from the left with the 1910s, followed by the 1950s, 1990s, and finally the 2010s. The progression of poster design reveals a remarkable transformation in color themes. In the 1910s poster, the color palette is predominantly black-and-white, and the poster relies on contrasting tones and shading to convey its message. Moving to the 1950s, the poster embraces the emergence of Technicolor and presents a more realistic and vibrant color palette. By the 1990s, movie posters adopt bold and saturated color themes. The colors are intensified to capture attention and create a visually striking impact. Finally, in the 2010s poster, there is a shift towards more subtle and refined color variations. Gradients and nuanced color palettes are employed to add depth and dimension to the artwork.

To test our hypothesis, we grouped movie poster release dates by decades and generated 6 indicator functions. These indicators were combined with the output of the flattening layer and fed into a new flattening layer, which was then connected to the dense layers for genre prediction. Narrowing our objective down to classifying only 3 genres, we selected ["Action", "Drama", "Romance"], ["Comedy", "Drama", "Romance"], and ["Drama", "Horror", "Music"] as the genre combinations to focus on. After running the model multiple times, we observed an average improvement of 0.04 in accuracy when incorporating the additional indicator functions, confirming the significance of time differences in the classification process.

Among the three selected genre combinations, ["Comedy", "Drama", "Romance"] showed the best testing performance, with an area under the ROC curve of 0.620083, an F1 score of 0.455876, precision of 0.463122, and recall of 0.473333. We also experimented with training the model using movie posters released after 2000 only, grouping release dates by 5 years to generate 6 indicator functions. The results showed notable improvement for ["Action", "Drama", "Romance"], little improvement for ["Comedy", "Drama", "Romance"], and a setback for ["Drama", "Horror", "Music"].

Throughout our analysis, we found that constructing a CNN model from scratch had limited effectiveness in identifying the genre of movie posters. The model performed reasonably well for the classification of 3 genres, but the results were just average. As a result, we have decided to explore more complex models, such as pre-trained models and Vision Transformers, to identify multiple genres.

**Appendix, graphs and tables**

Fig1 A graph of a number of blue bars

Description automatically generated

Fig 2 A person in a long dress

Description automatically generated A person holding an object

Description automatically generated A person holding an object

Description automatically generated A video game cover with cartoon characters

Description automatically generated

Fig 3 A table with numbers and a number of names

Description automatically generated with medium confidence

Fig 4 A line graph with different colored lines

Description automatically generated A graph of a line graph

Description automatically generated with medium confidence A graph of a curve

Description automatically generated with medium confidence

Table 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Area under ROC | F1 score | Precision | Accuracy |
| Action, Drama, Romance (1970+) | 0.602477 | 0.426114 | 0.429030 | 0.444667 |
| Action, Drama, Romance (2000+) | 0.661677 | 0.465727 | 0.453608 | 0.448667 |
| Comedy, Drama, Romance (1970+) | 0.620083 | 0.455876 | 0.463122 | 0.473333 |
| Comedy, Drama, Romance (2000+) | 0.620150 | 0.455927 | 0.456363 | 0.464667 |
| Drama, Horror, Music (1970+) | 0.620150 | 0.455927 | 0.456363 | 0.464667 |
| Drama, Horror, Music (2000+) | 0.530219 | 0.417394 | 0.424437 | 0.414000 |