Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-20174-2

Is a poster a strong signal of film quality? evaluating the predictive power of visual elements using deep learning

Thaís Luiza Donega e Souza[1] · Caetano Mazzoni Ranieri[2] · Anand Panangadan[5] · Jó Ueyama[3] · Marislei Nishijima[4]

Received: 3 June 2024 / Revised: 15 August 2024 / Accepted: 28 August 2024

Abstract
A film is considered an experience good, as its quality is only revealed after consumption. This situation creates information asymmetry before consumption, prompting producers, who are aware of their film's quality, to search for methods to signal this. Economic literature specifies that a signal to disclose a product's quality must be strong, meaning only producers of good-quality films can effectively utilize such a signal. However, a poster represents the most economical signal, and all producers, regardless of film quality, have access to this option. To study whether a poster can signal film quality, we first apply a low-dimensional representation of poster images and cluster them to identify quality-related patterns. We then perform a supervised classification of films into economically successful and unsuccessful categories using a deep neural network. This is based on the hypothesis that higher quality films tend to sell more tickets and that all producers invest in the highest quality poster services. The results demonstrate that a film's quality can indeed be predicted from its poster, reinforcing its effectiveness as a strong signal. Despite the proliferation of advanced visual media technologies, a simple yet innovative poster remains an effective and appealing tool for signaling film information. Notably, posters can classify a film's economic success comparably to trailers but with significantly lower processing costs.

Keywords Movie theater · Graphical design · Film advertisement · Autoencoder · Clustering analysis · Deep learning

Highlights

Is a poster a strong signal of film quality? A deep learning application on images
Thaís Luiza Donega e Souza, Caetano Mazzoni Ranieri, Anand Panangadan, Jó Ueyama, Marislei Nishijima

- Signaling Power of Posters: Analyzes whether film posters can signal movie quality and predict box office success.
- Machine Learning Techniques: Applies clustering and deep learning to study visual patterns in 16,000 film posters.

Extended author information available on the last page of the article

- Marketing Effectiveness: Discusses film posters as cost-effective marketing tools in the entertainment industry
- Wider Implications: Indicates potential applications for similar analyses in other areas of visual marketing.

1 Introduction

The economic theory classifies a film as a classic example of an "experience good" because consumers can only truly assess its quality after viewing it [1]. This previous information asymmetry poses a financial risk for film producers [2], who rely heavily on advertising to signal the quality of their films to consumers [3, 4]. This signaling is crucial, especially in the first week of release when ticket sales peak. Studios often allocate over half of their production budget to marketing campaigns [5] to create awareness and generate interest in the film, influencing consumer perceptions of its quality and increasing box office success. Usually, studios use multiple merchandise strategies, such as trailers, big stars, influencers or posters, to sign film quality.

According to the signaling theory [6], a signal must be strong enough to reveal a product's quality effectively. Spence [7] states that a signal of quality is strong when only producers of good-quality products can signal. This is because if both good-quality and bad-quality producers can signal, consumers cannot distinguish between good and bad products despite the signal.

This study investigates whether posters can be considered a strong signal of film quality, as defined under the economic concepts of asymmetric information and uncertainty. It utilizes machine learning and deep learning techniques on poster images to predict the financial success of films.

Since film posters are inexpensive to produce and accessible to all filmmakers, they serve as a universal signaling tool, regardless of a film's quality. This widespread use complicates the ability to differentiate quality, as discussed by [7] in his signaling theory. This raises the question of whether film posters serve as a weak signal of quality. Rather than simply accepting this premise, we explore whether posters can effectively predict financial success in films. The underlying hypothesis is that if a poster can successfully convey the quality of a film and the film is indeed of high quality, then it should correlate with higher box office revenues; conversely, poor quality should correlate with lower revenues. Additionally, we assume that all filmmakers purchase the highest quality poster services, as posters are relatively inexpensive compared to other film media. Considering the complexity of human visual perception, we believe that a deep learning analysis of images could mimic this perceptual process to some extent. We define financial success as films whose revenues, adjusted for inflation, exceed the average revenue of our dataset, which includes approximately 16,000 film posters from movies released in the United States between 1980 and 2020.

Numerous studies in marketing have utilized visual elements and deep learning to extract various types of information. For example, [8] and [9] applied a Convolutional Neural Networks (CNN) to analyze online images shared on social networks, revealing insights into consumer characteristics. Wang et al. [10] used background photos from social media in conjunction with a hierarchical Bayesian model to uncover consumer preferences in clothing. Pérez-Núñez et al. [11] proposed an innovative approach to enhance recommender systems by leveraging images and machine learning to better understand user preferences. Additionally,

[12] evaluated the effectiveness of three computer vision models-YOLOV2, Google Cloud Vision, and Clarifai-in analyzing brand-related visual content from social media, identifying which models are most effective for different aspects of this analysis.

Our interest in film posters as quality signals is rooted in several key factors, including their low cost of production. First, posters have long been a primary advertising medium in the movie industry. Second, they compete visually in the market and directly inform consumers about a film's content [13]. Third, unlike film trailers, posters can be displayed in various locations such as billboards, newspapers, malls, and theater lobbies. Fourth, they serve as a quick and direct source of information for undecided individuals. Fifth, modern catalogs of video streaming and On-Demand platforms often rely on film posters for their visual applications. Additionally, posters are designed to grab the attention of potential viewers and convey essential information about the film's theme, genre, actors, and storyline [14]. According to [15], iconic film posters are exemplary in illustrating these characteristics (see Fig. 1) .

As methodological strategy, our first step was to explore existing visual patterns in posters related to film revenues (domestic and worldwide grosses), genre, and year of release in the United States, using clustering algorithms and proposed adapted metrics. These metrics were motivated by the linear combination of intra-cluster compactness and inter-cluster separation, generating two custom measures for extrinsic evaluation able to address the significant imbalance in class sizes; *Normalized within-cluster homogeneity* and *Normalized between-clusters completeness*. All poster images were previously compressed with an autoencoder [16–18], and had their dimensionality reduced using Principal Component Analysis (PCA) [19]. The second step involved classifying film posters based on binary financial success or failure, measured based on box office revenues. For this classification task, we utilize Convolutional Neural Network (CNN) architectures from the EfficientNet family [20].

To our knowledge, this study is the first to apply deep learning techniques to explore the signaling quality of film posters. By introducing a novel application of deep learning algorithms, our work contributes to the economic signaling literature and opens avenues for extending these methodologies to other forms of simple visual advertisements, such as book covers, game covers, and music album covers. Furthermore, the intermediate representations generated by our models could be leveraged to enhance movie recommendation systems and provide consumers with valuable insights into film quality, drawing on findings from [21].

Fig. 1 Iconic movie posters. Source: IMDB

The remainder of this paper is organized as follows. Section 2 provides a comprehensive literature review, including related works on automated poster analysis and analyses based on other types of media for marketing. Section 3 details the data and methods used in this study, including data gathering, the generation of an abstract representation using an autoencoder, the clustering techniques employed for analyzing the relationships between these representations, and the classification procedures employed. Section 4 presents and discusses the results, especially the relationships between the clusters and different features from the movies, along with quantitative analysis and results from the classification models. The paper concludes with the main findings and implications in Section 5.

2 Related work

The extensive information available on the Internet, combined with advancements in deep learning for computer vision, has piqued researchers' interest in exploring diverse advertising media within the creative industry [20, 22–24].

Previous studies in automated film poster analysis have primarily concentrated on genre-specific characteristics. Ivasic-Kos et al. [25] attempted to automatically classify film posters into genres using low-level features such as color histograms, edges, and faces. Employing methods such as Distance ranking, KNN, and Naïve Bayes, they achieved unsatisfactory results when attempting to classify 1,500 movie posters into six classes. Ivasic-Kos and Pobar [26] expanded on this approach by including global low-level features like GIST features, dominant color features, color moments, and Classemes, along with a high-level image descriptor trained on other datasets. This enhanced approach improved their previous results, achieving an F1-score of 0.32.

Sirattanajakarin and Thusaranon [27] classified 1,700 movie posters into genres. They proposed a novel framework to extract semantic features manually (such as theme, emotion, composition, palette color, and the number of males/females) and then processed them with a multi-label classifier. Their best performance was 24% accuracy for 18 genres. Chu and Guo [28] used a CNN for object detection followed by a shallow Neural Network to classify 8,000 movie posters into 23 genres. They applied data augmentation for minority classes to balance the dataset. They achieved an overall accuracy of 18.73%. Nambiar et al. [29] included film reviews and poster images to classify 43,000 films into 19 genres. They used the VGG16 and ResNet architectures for images and GloVe and Word2Vec for reviews. The authors demonstrated that neural processing methods were better than conventional classifiers, and using both image and text features improved prediction performance (an F1-Score of 0.67). Other studies [30–32] also worked with movie posters using deep neural networks for classification according to genre.

Recently, recommendation system researchers have begun to include films' visual identity in their systems. Chen et al. [33] showed that the inclusion of posters has improved the performance of recommendation systems and that there is still room to widen their use.

There is a limited exploration of automated methods linking commercial success to visual content. Wang et al. [34] proposed a multimodal approach through trailers, synopses, cast, and hand-crafted post-release features to predict the box office revenue of 466 films in the Chinese movie market. They processed video trailers using Google-LeNet, LSTM-NN, and Siamese Network with user ratings as the target classification classes. This approach achieved 58% accuracy.

Ko et al. [35] predicted movie success by analyzing movie trailer content (audio and video) and corresponding YouTubers' "react" videos. They extracted low-level features, such as intensity, color, shot, motion, OpenSmile, face detection, from trailer frames and react videos for 112 films in the American market. The metadata was responsible for 70% of the accuracy in their results while adding video content improved it by up to 8%.

The connection between film posters and box office revenues, as found in the literature, is typically regarded as an additional feature in classification systems rather than the sole poster determinant of film box office revenues. For instance, [36] and [15] utilized a VGG16-CNN to categorize movie posters into user rating classes. They then combined the output class probabilities with other metadata (such as genre, budget, cast, user rating, and critic reviews) to train the neural network. Their systems were evaluated by classifying 3,807 movies into six box office classes. The performances were similar, with accuracies of 55.03% and 52.20%, respectively. The inclusion of poster information led to an improvement of 2-5%.

Using an alternative measure of film success, [37] attempted to predict the award success from a film poster. They used 3,844 posters and three levels of feature extraction (handcrafted, mid-level, and deep-level) to predict the winners of four major awards. For the mid and deep levels, they evaluated different deep learning networks, including PlaceNet (a deep autoencoder network for visual loop closure detection), EmotionNet, AlexNet, VGG, and late fusion combining networks. The highest accuracy was obtained with EmotionNet, even better than using only low-level features.

In related markets outside of film, some studies focus solely on images to reveal insights. For instance, [38] investigated genre classification in book covers using pre-trained neural networks on a dataset of 57,000 covers and over 30 genre classes. Their work revealed learned features by genre, such as colors, the presence of objects, persons, and text. They found that some book covers exhibited unexpected patterns, such as food books having similar covers to children's books. Jolly et al. [21] expanded this analysis by including the Single Shot Multibox Detector and the Efficient and Accurate Scene Text Detector, both pre-trained on specialized datasets, which resulted in slightly better overall results, although some genres were more accurately identified than others. Zhu et al. [39] took a different approach by aligning book covers with their corresponding scenes in movie adaptations. They used a Long Short Term Memory (LSTM) neural network for texts and contexts, and Convolutional Neural Networks (CNNs) with the MSCoCo dataset for images, covers, and scenes.

Studies such as those by [25–27], and [28] have made significant strides in classifying film posters by genre using unstructured data and convolutional neural networks, demonstrating the potential for automated genre classification based on visual elements. On the other hand, research by [34] and [35] combined multimodal data like trailers, synopses, and cast information with visual content to predict movie success, highlighting the value of diverse data sources in allowing for the prediction of films' financial success.

However, these previous works also have limitations. Studies such as [26] and [29] focused on genre-specific classification, which may not be generalized well to other tasks like predicting economic success. Conversely, approaches that rely on multimodal data, such as [34] and [40], often require substantial data processing resources. Moreover, none of these previous works assessed the relationships between film posters and financial success.

In this study, we strive to offer a more comprehensive approach to understanding film posters' predictive power regarding the films' economic success by employing deep learning architectures specifically designed for image analysis and clustering techniques. Our work seeks to generalize beyond genre classification by providing an exploratory study on the visual features of the posters, as well as determining whether movie posters can offer insights into a film's financial success.

3 Data and methods

We adhere to the economic literature's framework, inspired by [7], to examine whether a film poster serves as a strong signal of quality. Consequently, our study investigates if a film poster can effectively predict a film's financial success. It is worth noting that a high-quality film poster may be associated with high box office revenues, while a low-quality film poster may indicate lower earnings. To achieve this, we utilize autoencoders to generate a lower-dimensional representation of film posters and employ unsupervised learning methods, such as clustering, to explore the dataset. This exploration helps us understand the relationship between the most prominent graphical features on the posters and film characteristics, such as genre, year, and financial success. Subsequently, we identify the most significant features in posters to predict a film's commercial success using supervised learning with deep neural networks.

We utilized machine learning models in three main steps. First, a simple CNN-based autoencoder was used to generate embeddings, which are abstract representations of each poster image. These embeddings were then used as descriptors. In the second step, three clustering algorithms were assessed to identify patterns while aggregating visual features based on the embeddings. Lastly, a general-purpose CNN was employed to classify the films based on their financial success, using the raw poster images as inputs.

In regard to the autoencoder used for generating the embeddings, it was a simple yet effective unsupervised method for building representations without making assumptions about the content of the images. Other options could involve self-supervised learning algorithms such as SimCLR, MoCo, and BYOL [41]. However, these techniques would add complexity to the analysis and might make the results dependent on the specific algorithm used. Since the study focused on the results of the clustering algorithms, introducing a more complex method for representation learning could unnecessarily complicate the analysis and make the subsequent steps less manageable, as different clustering algorithms would need to be applied to different sets of embeddings generated by various algorithms.

The selection of clustering techniques aimed to consider the various heuristics that can be used to develop a clustering algorithm, resulting in different sets of clusters generated through different data clustering approaches. The study analyzed three types of clustering algorithms: centroid-based (i.e., k-means), hierarchical (i.e., agglomerative), and graph-based (i.e., spectral) clustering algorithms. Although density-based clustering was initially considered in the experimental design, it was excluded due to inconclusive results in exploratory experiments.

To classify the poster images based on their financial success, we utilized a well-established family of CNNs for classification, the EfficientNet [20]. Our goal was to determine if any indication of financial success could be identified through a deep learning architecture, and a single state-of-the-art architecture proved sufficient in this respect, as evidenced by the results presented. While we explored other architectures, including larger models from the same family, during preliminary analyses, we found no significant improvements in accuracy. We also acknowledge that a lower-scale model such as EfficientNetB0 would make the results more reproducible.

3.1 Data

Our dataset comprises 16,013 film posters, along with their release year and genre, spanning the United States from 1980 to 2019 (July) collected from the Internet Movie Database - IMDB[1]. We use only one poster for each film, the most prominently featured poster listed on IMDb for each film released in the US. In general, posters vary according to the country in which the film is released. Additionally, the dataset includes information on each film's box office revenues, which serve as our measure of financial success, obtained from Box Office Mojo[2], deflated by the Consumer Price Index (CPI).

To maintain consistency and eliminate variations in poster design across different countries and cultures, we focused solely on films from the United States, given its substantial market share. Thus, our dataset comprises all posters available at IMDB from releases in the United States from 1980 to July 2019.

We transformed the financial variables (domestic and worldwide gross) into binary categorical variables by thresholding them according to the mean value of the sample, see Table 1, which shows the distribution of both financial variables and the mean threshold.

The pre-processing of film posters is handled during the training and inference pipelines of the neural network, in which the images are resized to 256×256 (despite its original shape) and normalized into the [0,1] interval using min-max scaling for each color channel. Additionally, each poster is associated with corresponding metadata from Box Office Mojo, including film revenues, genre, and year of release. This is achieved using unique identification codes that are consistent across both IMDb and Box Office Mojo. To account for economic variations over time, the film revenues are adjusted for annual inflation corresponding to the year of the film's release.

Figure 2 presents the distributions of log domestic and worldwide film grosses. While Fig. 3 illustrates the distribution of film success according to genres (a) and year of release (b).

3.2 Data exploration

To investigate the correlation between visual features in film posters and three target variables – financial success (measured by domestic and worldwide gross), release year, and genre – we conducted an analysis using unsupervised learning on the image dataset. Our methodological strategy involved developing a machine-learning data processing pipeline to learn an abstract representation solely from the images. We utilized an autoencoder to learn embeddings from the images without labeled data. We then applied classical clustering techniques to identify similarities between these embeddings and analyzed the coherence of the resulting clusters with respect to each target variable.

3.2.1 Abstract representation

Using a custom Convolutional Neural Network (CNN) to generate the embeddings, the model was designed as an autoencoder for a simple and effective approach to dimensional reduction [16]. The components of an autoencoder are the **encoder**, which comprises a sequence of layers that subsequently reduce the data dimension down to a fixed-length feature vector, and the **decoder**, which implements the reverse operations on the feature vector to

1 www.boxofficemojo.com

2 www.imdb.com

Table 1 Financial variables distributions based on the sample mean

Class	Observations	Mean	Unsuccessful (0)	Sucessful (1)
Domestic Gross	16013 (100%)	19.96 M	12758 (79.7%)	3255 (20.3%)
Worldwide Gross	16013 (100%)	45.99 M	13211 (82.5%)	2802 (17.5%)

restore the original input with as little loss as possible. Once the autoencoder is trained, the encoder produces the embeddings to be used as an abstract, or the lower-dimensional image representation.

Figure 4 shows the designed architecture of the autoencoder. The RGB images were resized to 256 $\times$ 256 before being fed to the neural network. The encoder comprised five layers, with 3 $\times$ 3 convolutional kernels interspersed with 2 $\times$ 2 max-pooling operations. We implemented a decoder with the exact reverse operations to train the neural network. After the model was trained, the encoder output (i.e., the 8 $\times$ 8 $\times$ 256 internal representation) was used in our analyses.

This representation was flattened, forming an array with 16,384 positions. We then applied the Principal Component Analysis (PCA) transform and kept the smallest number of principal components to preserve 90% of the variance, ending up with an embedding composed of 136 dimensions, suitable for traditional clustering techniques.

3.2.2 Clustering

The embeddings of the poster images are used as inputs to three clustering algorithms: K-Means [42], agglomerative clustering [43], and spectral clustering [44].

Let X be the set of film poster embeddings in the dataset, which includes N=16,013 images in our sample. Let V be a set of categorical variables v that can follow a pattern found in

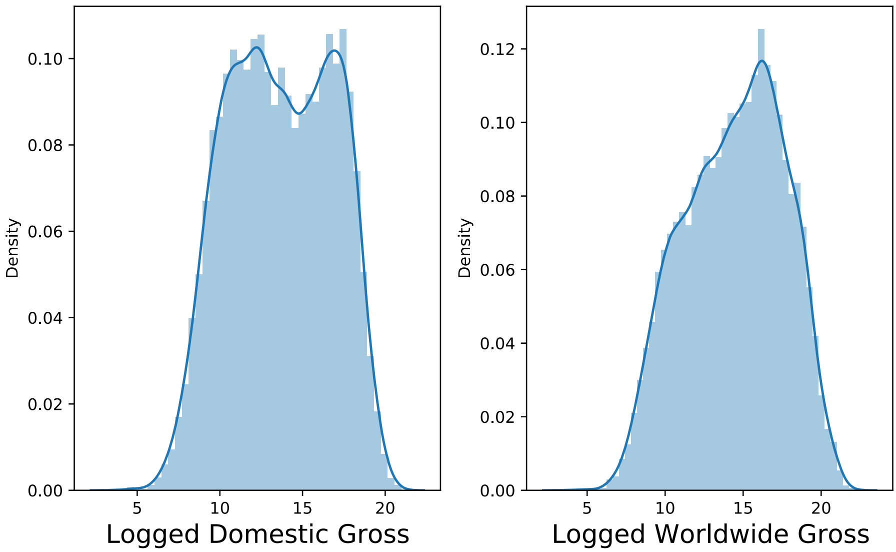
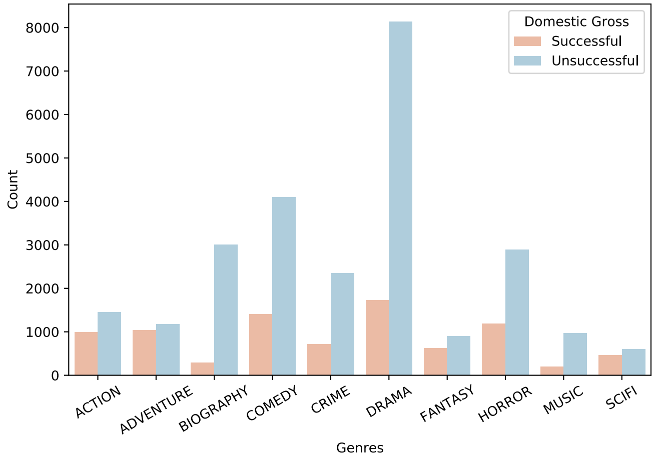
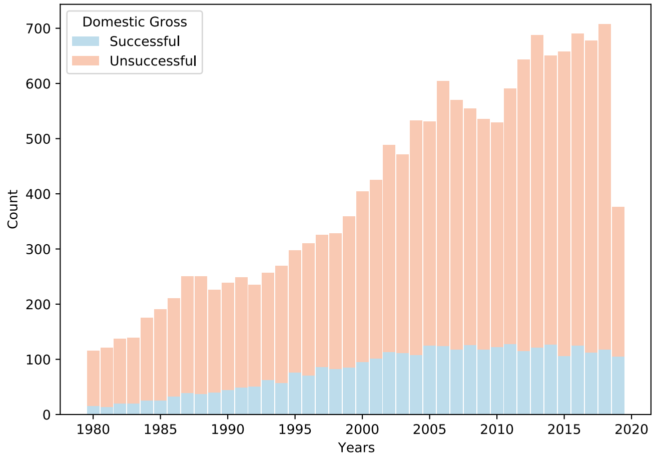


Fig. 2 Film Log Domestic (Left) and Worldwide Gross (Right) distributions



(a)



(b)

Fig. 3 Distributions of (a) genre, and (b) release year

the resulting clusters. Each variable $v \in V$ has categorical values $t = 1, 2, 3 \ldots T$ and each category of a variable has N_t size. A clustering algorithm learns an underlying function $F_v : X \rightarrow t$ for each categorical variable in V, associating the images in X to the values in t. Thus, the patterns that can emerge in resulting clusters come from the categorization of categorical variables v into their categories t in different clusters c. We are particularly interested in $v =$ "domestic gross", $v =$ "worldwide gross", $v =$ "genre", and $v =$ "release year". Table 2 lists the interest variables, categories, and number of observations.

Using the observed categories t of the interest variables, we can externally evaluate the clustering results based on how coherent the groups are for such categories. Therefore, for each interest variable in V, the resulting clusters were analyzed according to the proportion of instances belonging to each category.

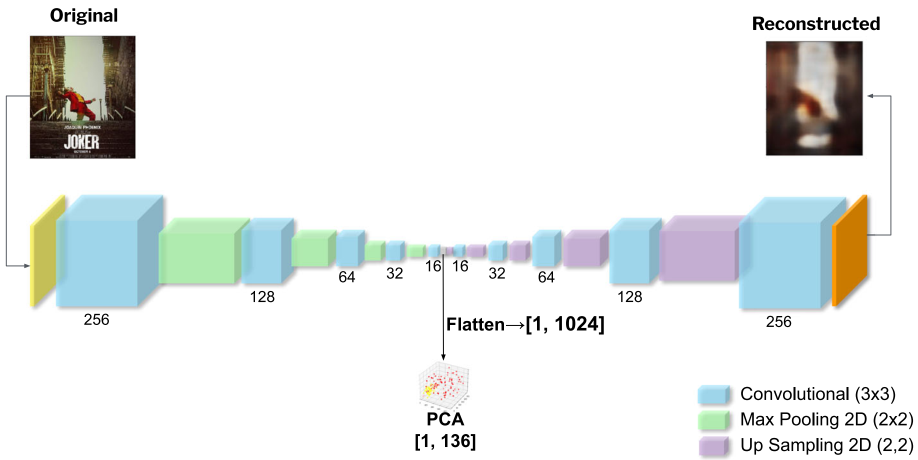


Fig. 4 CNN-based autoencoder designed to extract features from the image. The encoded representation, an $8 \times 8 \times 256$ structure, is flattened and reduced again using PCA, resulting in the embedding used in our analyses

The dataset shows a significant imbalance across the categories of the variables of interest. For domestic gross and domestic gross, the data was categorized in "Unsuccessful" and "Successful" following an approximately 80%/20% distribution (see Table 1). As film genres are multi-label (i.e., a film can be assigned with more than one genre), we replicate a film poster observation by its number of genres adding up to more than the number of film posters on the dataset and their categorical imbalance can be seen in Fig. 3a. There is also a severe imbalance in the number of films in each genre. The release year category is also imbalanced, as film production has been increasing through the years (see Fig. 3b).

The large imbalances in category sizes within the variables render both homogeneity and completeness unsuitable for analyzing this dataset of film posters [45]. These are the most frequent measures for extrinsic evaluation of clustering and other common Cluster Validation Indexes [46]. To deal with this problem, we propose a custom external evaluation measures to normalize for class and cluster imbalances, as defined in Table 3. These measures are normalized and weighted bay the sizes of categories, clusters and sample size, addressing

Table 2 Variables of interest and its corresponding category along its observations

Variable	Categories	Observations
v = Domestic Gross	$\{t_1$: Successful, t_2 : Unsuccessful$\}$	16,013 (100%)
v = Worldwide Gross	$\{t_1$: Successful, t_2 : Unsuccessful$\}$	16,013 (100%)
v = Genre[1]	$\{t_1$: drama, t_2 : comedy, t_3 : horror, t_4 : biography, t_5 : crime, t_6 : action, t_7 : adventure, t_8 : fantasy, t_9 : musical, c_{10} : sci-fi$\}$	34,293 (100%)
v = Year	$\{t_1$:(1980:1984), t_2 :(1985:1989),. . ., t_8 :(2015:2019)$\}$	16,013 (100%)

[1]Genres are multi-labeled, we thus replicated film posters to account for all genres, reaching 34,293 film observations according to genre

Table 3 Normalized measures for clustering evaluation

Variable	Description
N	Number of film posters
N_t	Number of film posters in a category t of a variable v
s_c	Number of film posters in a cluster c
$s_{t,c}$	Number of film posters in a category t and cluster c
Within-cluster Measures - Homogeneity	
$W_{t,c} = \frac{s_{t,c}}{s_c}$	Percentage Within-cluster
$H_{t,c} = \frac{W_{t,c}}{N_t/N}$	Normalized within-cluster homogeneity
Between-clusters Measures - Completeness	
$B_{t,c} = \frac{s_{t,c}}{N_t}$	Percentage Between-clusters
$C_{t,c} = \frac{B_{t,c}}{N_t/N} \cdot \frac{s_c}{N}$	Normalized between-clusters completeness

two requirements: (i) the clusters specialized in recognizing certain classes, i.e., it contains more elements from a given class than the overall prevalence of this class; and (ii) the classes adhered to specific clusters, i.e., most elements from a given class fell into one or a few clusters. We call the first measure $H_{t,c}$ of "normalized within-cluster homogeneity" and the second , $C_{t,c}$ of "normalized between-clusters completeness".

3.3 Classification

To classify the posters according to the film financial success and identify patterns in the images that have discriminatory power, we also used deep neural networks from the EfficientNet family in a typical training/evaluation workflow [20], as shown in Fig. 5. EfficientNet family models have achieved greater accuracy with far fewer parameters related to other CNNs. As our dataset is too small for learning all the weights from a random initial state, we used pre-trained weights. Through tests, the Student Noise weights, a semi-supervised learning approach that improves the performance of EfficientNet models [47], was chosen.

Performing a 10-fold cross-validation to evaluate the resulting models of binary classification task (success or unsuccessful), we appended a single neuron at the end of the neural network, with a Sigmoid activation function. Next, we used the Adam optimizer and set the learning rate to $2e^{-}3$. Each fold was trained for 50 epochs, with a patience of 20 for early-stopping. Only the best model from each fold was considered for evaluation.

4 Results and discussion

4.1 Unsupervised learning

This section presents the experiments and results obtained from applying the clustering methods, as well as an exploratory analysis of these results using our within-cluster and between-cluster measures.

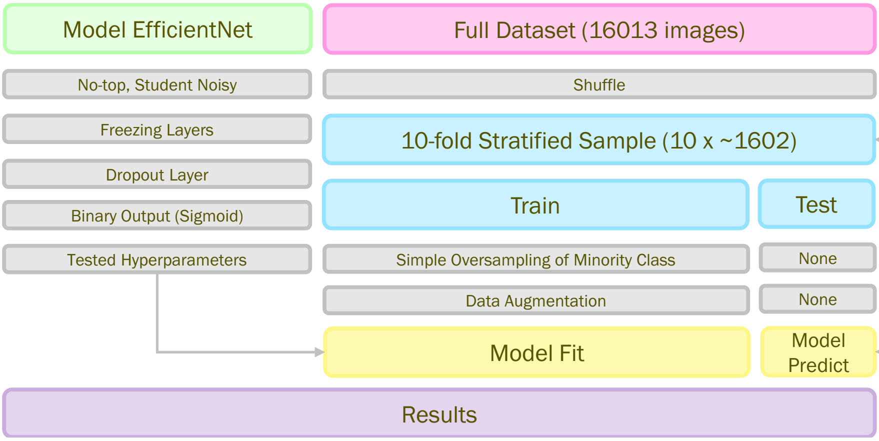


Fig. 5 Workflow for the supervised learning approach. Data from the images are shuffled and randomly split into 10 partitions so that 10 models would be trained and evaluated with one partition for validation/testing, and all the others for training. Oversampling was used before training to compensate for the class imbalance. The EfficientNet architecture was followed by a binary classification layer, in order to generate a classification output

4.1.1 Clustering results

Three traditional clustering methods - K-Means, agglomerative, and spectral clustering - were applied to the embeddings derived from the autoencoder network followed by the PCA transform (Fig. 4). Given that a movie poster integrates various pieces of information, we utilized an exploratory approach to determine the optimal number of clusters (k). We do not anticipate that an unsupervised method can isolate a single feature from the posters. Instead, we systematically analyze the clustering outputs to explore what the methods can reveal about high-level classifications and how these characteristics might interconnect.

We initiated our analysis using the elbow method for K-Means [48], which suggested $k = 6$. We then used it as the first guess for the dendrogram analysis, after which we chose $k \in 4, 5$ for agglomerative clustering. For spectral clustering, we used Eigengap Heuristics that resulted in $k \in \{1, 2, 3, 4, 6\}$ [44]. Using the values of k for K-Means and agglomerative clustering, we run experiments for $k \in \{4, 6\}$. Table 4 illustrates the methods, the number of clusters (k), their graphic representation, and their clusters' representation.

As part of our clustering analysis, we explored the clusters visually. Tables 5 to 7 present snippets from a representative random sample of 10 film posters for each pair of clustering methods and k value. This exploratory analysis indicates that a distinct pattern has emerged from the clustering efforts: the colors present in the posters of each cluster appear to be closely associated with their respective movie genres [30, 32].

Note that we will refer to the K-means method with $k = 6$ as KMeans-6, agglomerative with $k = 4$ and $k = 5$ as Agglomerative-4 and Agglomerative-5, and to the spectral with $k = 4$ and $k = 6$ as Spectral-4 and Spectral-6.

Table 5 suggests a pattern of colors correlated with k-means' clusters, especially in clusters c_1 and c_5, which correspond to horror and comedy genres. For the remaining clusters, the patterns of colors are less evident (e.g., cluster c_2 suggests hot colors, such as red, orange, and yellow). However, there was not a clear relationship between the sampled images in the other clusters and particular genres.

Table 4 Summary of the configurations and the distributions of the results for each clustering technique and the number of clusters

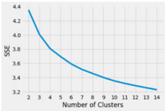
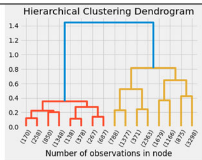
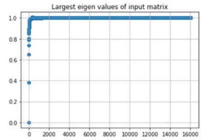
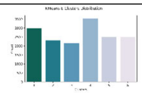
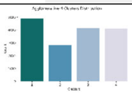
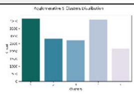
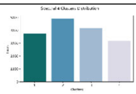
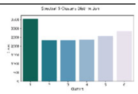
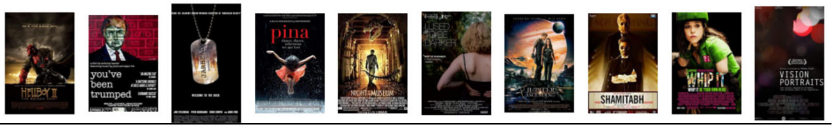
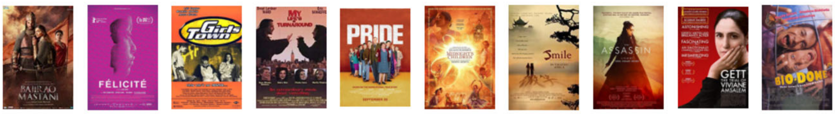
	KMeans	Agglomerative		Spectral	
Number of clusters k	6	4	5	4	6
Method to determine k	Elbow	Dendrogram Analysis		Eigengap Heuristics	
Graphical representation					
Clusters distribution					

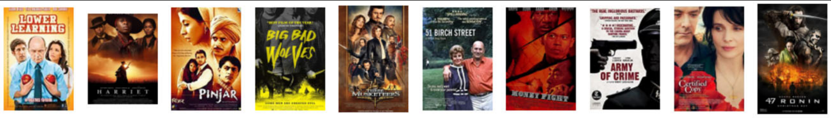
Table 5 Randomly selected film posters for each of 6 **K-Means** clusters

	Random sample of 10 posters
c_1	
c_2	
c_3	
c_4	
c_5	
c_6	

Results for Agglomerative (Tables 6-a and b) and Spectral (Tables 7-a and b) clustering are similar with respect to colors.

All experiments consistently resulted in clusters characterized by specific color tones-dark, light, reddish, and bluish. Given that certain colors are strongly associated with specific

Table 6 Randomly selected film posters for each of 4 **(a)** and 5 **(b)** of **agglomerative** clusters

(a) Random sample of 10 posters with $k = 4$	
c_1	
c_2	
c_3	
c_4	

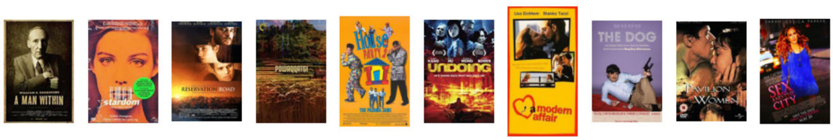
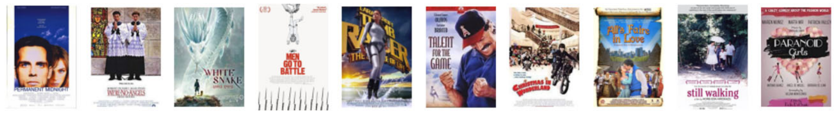
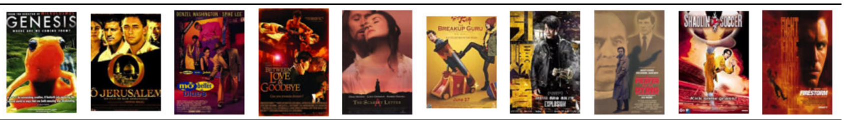
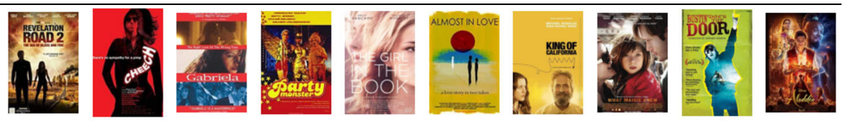
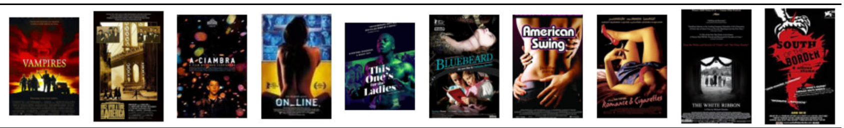
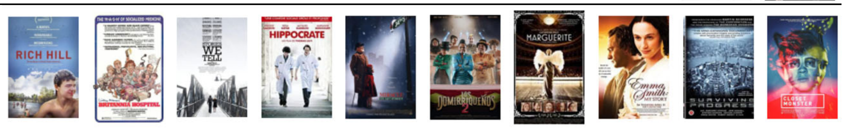
(b) Random sample of 10 posters with $k = 5$	
c_1	
c_2	
c_3	
c_4	
c_5	

Table 7 Randomly selected film posters for each of 4 **(a)** and 6 **(b)** **Spectral** clusters

(a) Random sample of 10 posters with $k = 4$	
c_1	
c_2	
c_3	
c_4	

(b) Random sample of 10 posters with $k = 6$	
c_1	
c_2	
c_3	
c_4	
c_5	
c_6	

genres, posters from films within these genres proved easier to cluster than others. Table 8 provides a summary of this exploratory analysis and the clustering outcomes.

A further avenue for exploration involves the study by [49], which links colors to emotions. However, due to the absence of a direct metric for quantifying emotions in our current framework, this intriguing aspect will be deferred to future research initiatives.

4.1.2 Quantitative analysis

Qualitative analysis involves to correlate the resulting clusters with the four interest variables, summarized in Table 2, and calculate the within-cluster and between-clusters measures shown in Table 3. The objective is to evaluate whether the resulting clusters of poster embeddings, based solely on features learned through an unsupervised approach, carry information about categories of our interest variables. For brevity, the following section shows only the results of KMeans-6 method.

- *Example 1: Kmeans-6 and domestic gross*

Figure 6 depicts the within-cluster and between-cluster measures resulting from KMeans-6 with respect to the financial success, measured by film's domestic gross' features. Results on worldwide gross are very similar and not reported. Figure 6a shows that cluster c_1 favors successful films, while cluster c_2 mainly rejects them. Results shown in Fig. 6c support this correlation. In addition, clusters c_1 and c_4 include mainly successful films, as shown in Figs. 6b, c, and d.

- *Example 2: Kmeans-6 and genre*

For genres, Fig. 7a shows that cluster c_5 best groups comedy films and cluster c_1, horror, which confirms the color preference mentioned in Table 4 and Fig. 7c. These clusters also reject "opposite" colors/genres, i.e., cluster c_1 also contains mainly comedy and adventure, while cluster c_5, does *not* include horror films. This result may be explained by the visual design that has been typically associated with these two genres. Cluster c_4 best groups the posters according to action, fantasy, and sci-Fi genres, and also includes most of the Sci-Fi films.

We also split each genre according to the domestic gross' categories (success and unsuccessful) and calculated the within $H_{t,c}$ and between metrics $C_{t,c}$ for these subsamples, as shown in Fig. 8. In Fig. 8a, we report the $H_{t,c}$ measures per cluster, while Fig. 8b reports the same measure of the domestic gross variable for each genre. Adventure and sci-Fi films are more accurately classified into successful or unsuccessful films. Cluster c_3 is best at distinguishing successful from unsuccessful films as cluster c_1 diverges the most from the tendency of the overall bar (Fig. 8b).

- *Example 3: Kmeans-6 and Year of film release*

We grouped the release year variable in 5-year buckets to show the variation with year more clearly. As shown in Fig. 9a, cluster c_3 has a significant portion of poster designs of old films, while cluster c_6 typically includes newer films. On the other hand, clusters c_1 and c_2 include the 1990-2004 designs (Fig. 9a and c). This result could indicate the changes in movie poster design over the years. In Fig. 9b and d, we observe that among all 80's movies (oldest ones), cluster c_4 includes the majority. Table 9 summarizes the main results of the clustering experiments.

Table 8 Summary of the visual inspection of the clustering experiments

Method/Cluster	c_1	c_2	c_3	c_4	c_5	c_6
KMeans 6 Clusters	Dark tones, black; Horror	Mixed, hot tones (yellow, orange, red)	Light, mixed	Black, blue and dark tones; Action	Light, Bright, white; Comedy	Mixed, Light
Agglomerative 4 Clusters	Dark, black, red, blue (strong); Action, Horror	Mixed, Bluish and grey; Fantasy	Hot tones (yellow, orange, red)	light, bright, white		
Agglomerative 5 Clusters	Hot tones, yellow mainly	Mixed, light blue and grey; Fantasy	Mixed, dark Blue and grey	light, bright, white; Comedy	Dark, black, red, blue (strong); Action, Horror	
Spectral 4 Clusters	Mixed	Dark, black	Mixed, light	Dark, Hot tones		
Spectral 6 Clusters	Dark, black	Dark, mixed	Hot tones	Dark, Hot tones	Mixed	Light, Bright, White, Cold

The first row represents cluster labels, i.e, from c_1 to c_6, while the first column stands for methods. For example, cluster c_1 of KMeans-6 has a higher prevalence of posters with dark tones and is typically related to horror films

Fig. 6 Within-cluster, $W_{t,c}$, $H_{t,c}$, and between-clusters, $B_{t,c}$, $C_{t,c}$, measures for Kmeans-6 of ***domestic gross***. Heatmaps in charts (a) and (c) display deviations from the sample mean: lighter colors indicate proximity to the mean, while intense blue or red signifies a cluster's effectiveness in identifying a category by concentrating or excluding instances. Bar charts (b) and (d) highlight the differences between cluster and overall bars, with greater differences suggesting more effective clustering

Overall, certain patterns emerged from the clustering analyses. Successful films tend to be clustered in the same cluster as action films along with crime, sci-fi, and horror genres. Clusters that reject unsuccessful films tend to reject comedy films. No correlation is observed between the film release year and the box office revenues; this is an expected result since there is a continuum of film releases yearly. Regarding the number of clusters in the same method, we observe that, in general, the results are similar. Using the largest value of k resulted in repeated clusters that are similar to one cluster found when using the smallest k, and the signal for some categories weakened as k was increased. Regarding genre, the methods tend to perform similarly to the grouping of action and horror and/or action and sci-fi. However, while KMeans-6 groups comedy in one cluster, Agglomerative-4 groups comedy with adventure and fantasy in a unique cluster. We detect an inverse relationship between action/horror films and comedy in all methods.

None of the methods successfully grouped the posters of unsuccessful movies. However, all methods formed clusters that captured similarities among successful film posters, often explicitly excluding those from unsuccessful films. Thus, while no distinct pattern emerges exclusively for unsuccessful movie posters, a clear distinction exists between the posters of successful and unsuccessful films.

Lastly, Table 4 suggests that each method performs better in recognizing one of the four interest variables; Spectral-4 was slightly better in recognizing different design patterns in the year of film release, while Agglomerative-4 and -5 were poorer. This is expected, since

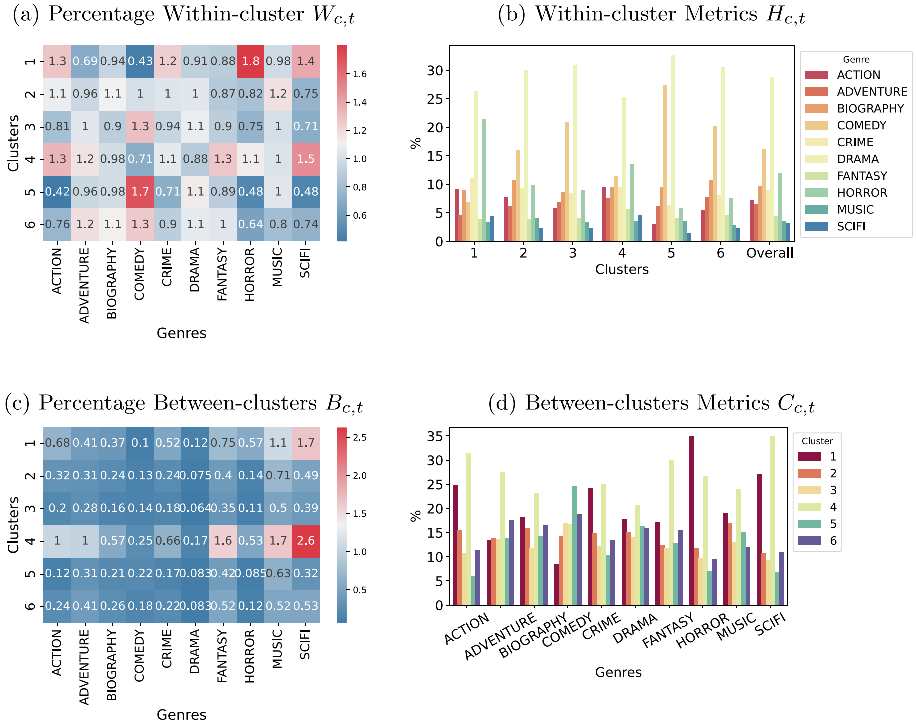


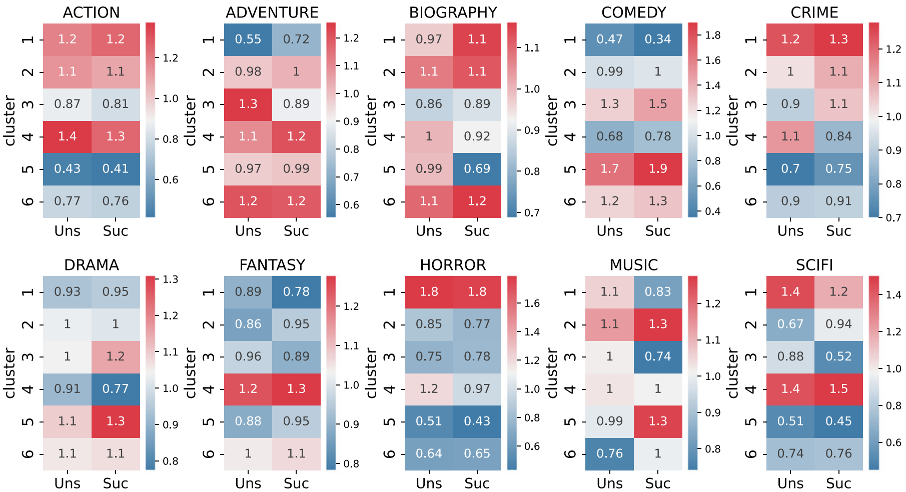
Fig. 7 Within-cluster and between-clusters results for Kmeans-6 with respect to the ***genre*** variable, according to the measures defined in Table 3. In the heatmaps in charts (a) and (c), observe the divergence from the sample mean: the lighter the color, the closer the proportion of instances in a cluster is to the sample mean; the more intense, blue or red, the better the cluster is in identifying a category by concentrating or rejecting instances from it. In bar charts (b) and (d), observe the difference between the Cluster bar and the Overall bar: the greater the difference, the less random the clustering

there is no single "best" clustering method, but each is best suited to a particular interest variable [17, 50].

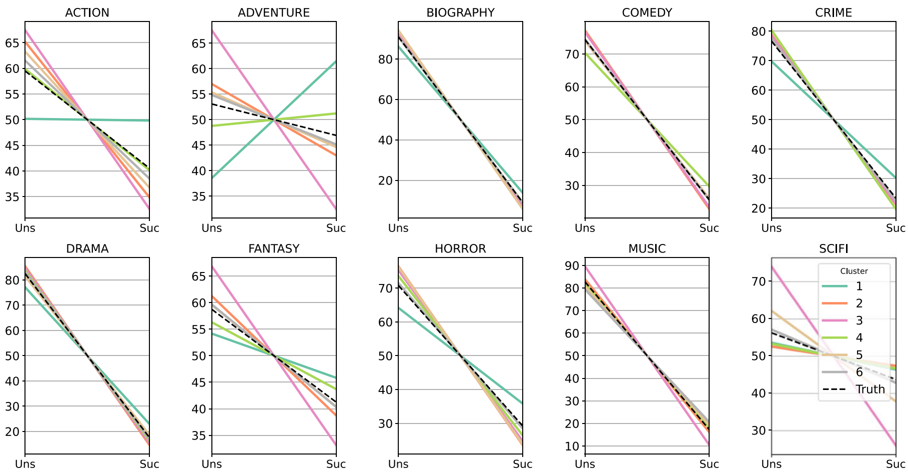
4.2 Supervised learning

The supervised learning analysis aimed to uncover film poster patterns related to financial success.

All images were resized to meet the requirements of the neural network input layer (256 × 256 for EfficientNetB0 with 3 RGB channels) and re-scaled (1./255). They were shuffled and split into training and test sets for stratified 10-fold cross-validation. For the training set, we performed oversampling of the minority class, i.e., we duplicated positive samples randomly and employed data augmentation on them. The experiment workflow is illustrated in Fig. 5. All experiments were performed in Google Collaboratory, which provides a virtual machine computing environment with GPUs and high RAM memory.



(a) Percentage Within-cluster, $W_{t,c}$, of domestic gross per genre. Note the divergence from the sample mean: the lighter the color, the closer the proportion of instances in a cluster is to the sample mean; the more intense, blue or red, the better the cluster is in identifying a category by accepting or rejecting instances from it.



(b) Percentage Between-cluster metric $H_{t,c}$ of domestic gross per genre. The proportions of unsuccessful and successful movies in each cluster (which always sum to 100%) are connected with a line. Note the slopes of these lines. The dashed line refers to the overall proportion of unsuccessful and successful movies in the dataset. The more the line of a cluster deviates from this dashed line, the better the cluster is in accepting instances from one of the categories (i.e., unsuccessful or successful).

Fig. 8 Kmeans-6 results for domestic gross grouped by genre

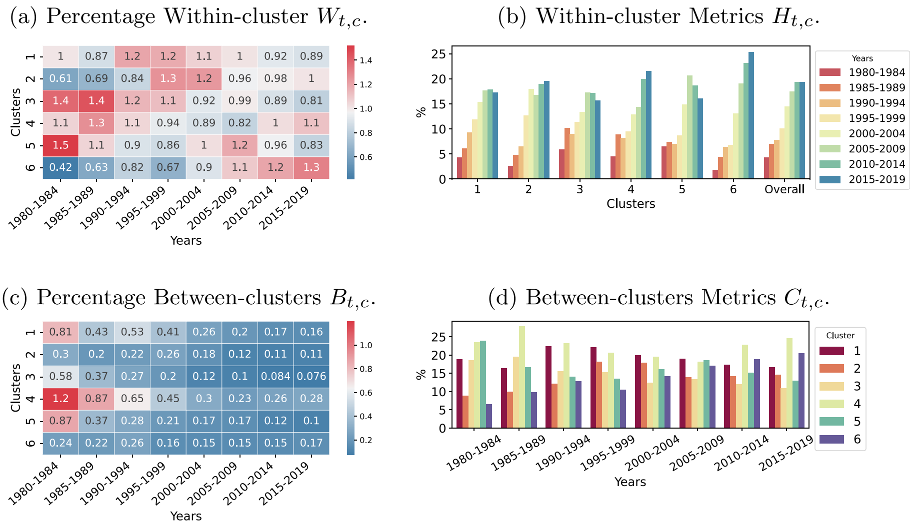


Fig. 9 Within-cluster and between-clusters results for Kmeans-6 with respect to the ***release year*** variable (five-year group), according to the measures defined in Table 3. In the heatmaps in charts (a) and (c), observe the divergence from the sample mean: the lighter the color, the closer the proportion of instances in a cluster is to the sample mean; the more intense, blue or red, the better the cluster is in identifying a category by concentrating or rejecting instances from it. In bar charts (b) and (d), observe the difference between the Cluster bar and the Overall bar: the greater the difference, the less random the clustering

Table 9 Summary of results of the unsupervised analysis

		KMeans-6						Agg-4				Agg-5					Spectral-4				Spectral-6					
Variable	**Value**	c_1	c_2	c_3	c_4	c_5	c_6	c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4	c_5	c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4	c_5	c_6
Domestic Gross	**Success**	++	o	o	+	o	o	++	o	- -	o	-	o	+	o	++	+	++	o	- -	- -	+	-	++	o	o
	Unsuccess	-	-	o	o	o	o	-	o	o	o	o	o	o	o	-	o	o	o	o	o	o	o	o	o	o
Genre	**Action**	+	o	-	+	- -	-	++	o	+	- -	o	o	++	- -	o	o	+	- -	o	o	o	o	+	-	- -
	Adventure	-	o	o	+	o	+	o	++	o	o	o	+	o	o	- -	+	o	o	o	o	o	o	-	+	o
	Biography	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Comedy	- -	o	+	-	++	+	- -	++	o	++	o	+	- -	++	- -	o	- -	++	o	-	o	o	- -	o	++
	Crime	+	o	o	o	o	o	+	o	o	o	o	o	o	o	+	o	o	o	o	o	o	o	+	o	o
	Drama	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
	Fantasy	o	o	o	+	o	o	o	++	o	o	o	+	o	o	o	o	o	o	o	o	o	o	o	+	-
	Horror	++	o	-	o	- -	-	++	- -	o	- -	o	- -	o	- -	++	-	++	- -	o	o	-	-	++	-	- -
	Music	o	+	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	-	o	o	o	o
	Sci-Fi	+	-	-	++	- -	-	++	o	o	- -	o	o	++	- -	+	o	++	- -	-	o	o	-	+	+	- -
Release year	**1980-1984**	o	o	++	o	++	- -	o	- -	o	++	o	- -	o	++	o	- -	+	++	-	++	- -	o	o	o	++
	1985-1989	o	o	++	+	o	-	o	o	o	o	o	o	o	o	o	o	+	+	-	++	o	o	o	o	+
	1990-1994	+	o	+	o	o	-	o	o	o	o	o	o	o	o	+	o	+	o	o	++	o	o	o	-	o
	1995-1999	+	+	o	o	o	- -	o	o	o	o	o	o	o	o	+	-	+	o	+	o	o	+	+	o	o
	2000-2004	o	+	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	+	o	o	+	+	o	o
	2005-2009	o	o	o	o	+	o	o	o	o	o	o	o	o	o	o	o	o	o	o	-	o	o	+	o	o
	2010-2014	o	o	o	o	o	+	o	o	o	o	o	o	o	o	o	+	o	o	o	o	+	o	o	+	o
	2015-2019	o	o	o	o	o	+	o	+	o	o	o	+	o	o	o	+	o	-	o	+	+	o	o	o	o

Where: (- -) strongly rejects/repeals; (-) weakly rejects/repeals; (o) neutral; (+) weakly attracts; (++) strongly attracts. The symbols refer to how much the cluster could group and/or reject one feature class based on our within-cluster measure. For instance, KMeans-6 cluster c_1 includes mainly posters from successful movies and rejects posters from unsuccessful ones

4.2.1 Baseline: majority class classifier

Since we are analyzing a novel dataset without an established baseline in the published literature, we addressed the significant imbalance within the dataset by computing baseline results using a "dummy" majority class classifier, which predicts all films as unsuccessful. Tables 10 and 11 summarize our baseline results.

4.2.2 EfficientNet B0 classifier

We fine-tuned a pretrained EfficientNet B0 neural network on our dataset. To evaluate its performance, we experimented with an optimized threshold by analyzing the ROC curve and tested it with a balanced test sample, by randomly under-sampling the majority class. Figure 10 shows an example (fold 3, lowest loss) of train and test snippets followed by its train history. Table 12 presents the overall results for this experiment.

4.3 Discussion

Our unsupervised clustering analysis, which highlighted genres and colors as prominent features, proved particularly adept at identifying genres due to the existing correlation between genre and color. Aligning with previous research that has demonstrated a strong link between film genres and color schemes in posters, this relationship suggests that color could serve as a distinctive attribute for further exploration [25, 26, 31]. Moreover, while no prior study has leveraged clustering in this context, our findings indicate that it is a viable method for pinpointing broad film genres like horror, action, and comedy. Although our clustering did not reveal a definitive pattern for classifying posters based on financial success, the process yielded a latent representation that was instrumental for the subsequent supervised learning phase, underlining the value of this analytical step.

From the supervised learning approach, the EfficientNetB0 results indicate correct classification posters according financial success with an accuracy of 73% (Table 12). In Experiment C, Precision and Recall values are more evident, indicating strong learning even with the large imbalance (Tables 12 and 13). This result suggests that posters are strong signals of film quality since they can be used to classify financial success and unsuccessful films [7].

This pioneering study is the first to use deep learning to classify film posters based solely on their graphical characteristics according to their financial success. It compares favorably with other studies that utilize multiple features beyond posters (Fig. 11). Additionally, the research analyzes a substantial dataset of 16,000 film posters from releases in the United States, covering a broad period from 1980 to 2019.

Table 10 Results for the "dummy" baseline classifier, which consisted of always predicting the majority class (i.e., it always predicted a given input poster as belonging to an unsuccessful movie)

Set	Metric	Mean (std)
Train	Accuracy	50 (0.00)
Test	Accuracy	79.7 (0.02)
	Recall	0 (0.00)
	Precision	0 (0.00)
	AUC	50 (0.00)
	F1 measure	0 (0.00)

The accuracy is equal to the prevalence of the majority class in the dataset

Table 11 Confusion matrix for the dummy classifier

	Predicted negative	Predicted positive
Actual negative	12,758	0
Actual positive	3,255	0

"Positive" refers to the "successful" category, while "negative" refers to the "unsuccessful"

To contextualize our findings, it's noteworthy to compare them with [36] and [15], who achieved an accuracy of 63.15% by classifying box office revenues using multiple inputs along with 3,807 film posters. This is 10% lower than our accuracy, which relies exclusively on film posters.

Compared to studies analyzing film trailers, which involve higher data processing costs, posters remain valuable. For instance, [40] analyzed 474 trailers and traditional film metadata from releases in the United States between 2010 and 2014 to predict revenues, finding that trailers only increased the explained variance by 6%. Meanwhile, [51] examined the emotions displayed by characters in 131 film trailers, using these observations to predict financial returns with accuracies of 77% and 79%, depending on the model used. Thus, despite working with a large number of images in a trailer, the additional images do not lead to better accuracy compared to posters.

It is important to emphasize that our study aimed to determine whether a poster can signal a film's quality. This premise is based on the idea that higher-quality films tend to achieve better sales and that producers universally opt for the highest-quality posters due to their cost-effectiveness. Our results indicate that a simple poster can effectively communicate a film's quality. As a cost-effective medium, posters do more than just serve as promotional material; their ability to convey quality information may stem from the fact that high-quality posters are intrinsically linked to the film's storyline and other key characteristics.

A potential limitation of our study is that it does not exhaustively analyze all visual elements present in film posters, as our research is confined to images sourced from the internet. Future research could enhance the accuracy of our model by examining high-resolution film poster images in greater detail, utilizing inpainting techniques, as suggested by [52], including textual content and facial features.

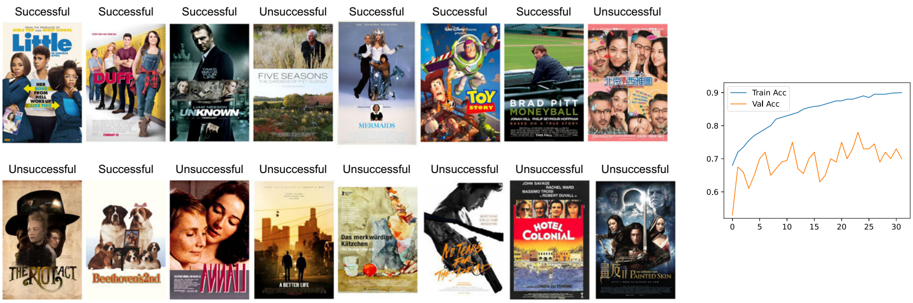


Fig. 10 Samples of posters from the training (top) and test (bottom) sets, followed by train history. In the graph, the average training and validation accuracy is plotted against the training epoch for the EfficientNet B0 architecture

Table 12 Results for EfficientNet B0 model for classifying whether a poster belongs to a successful or unsuccessful movie

Metric	Train Set Mean (std)	A - Evaluate Test Set (Threshold = 0.5) Mean (std)	B - Evaluate Test Set (Optimal Threshold = 0.4023) Mean (std)	C - Evaluate Balanced Test Set[1] (Threshold = 0.5) Mean (std)
Loss	0.2731 (0.0146)	0.5468 (0.025)	–	0.648 (0.075)
Accuracy	88.5 (0.758)	72.997 (2.201)	67.11	67.787 (2.0795)
Recall	0.893 (0.641)	0.613 (5.371)	0.73	0.6129 (5.3716)
Precision	0.879 (0.86)	0.396 (2.419)	0.35	0.70496 (2.0511)
Auc	95.46 (0.486)	76.171 (1.759)	69.4	74.856 (1.9674)
F1-Score	88.59 (0.734)	48.12 (3.336)	47.52	65.57 (2.969)

[1]652 observations. The cross-entropy loss (optimized while training; the lower, the better), the accuracy and the other classification metrics are shown for each condition. We considered the metrics obtained from the training set, the test set with a standard threshold of 0.5, the test set with the optimal threshold of 0.4023, and an artificially balanced test set

Another limitation stems from the possibility that some film producers might opt not to invest in high-quality poster services, despite their relatively low cost. This could introduce potential discrepancies in our findings. Therefore, future research should investigate the causality between the quality of film posters and consumer decisions regarding film consumption.

Finally, our study pioneers the use of machine learning to empirically test economic theories of signaling, expanding the scope of inquiry to evaluate the effectiveness of signals in environments where all producers are capable of signaling. Moreover, it concentrates on a remarkably simple signal-a single image-to extract complex information about film quality, demonstrating its effectiveness.

5 Conclusion

This study marks a pioneering effort to evaluate film quality signals within a dataset of over 16,000 film posters, utilizing both unsupervised and supervised deep learning methods. It advances the economic literature on signaling by introducing a novel method to ascertain whether a simple image-a poster-can effectively indicate film quality. To this end, deep learning algorithms are employed to identify patterns in posters that correlate with the financial

Table 13 Confusion matrices for test sets A, B, and C (different threshold values)

	A − 0.5		**B − 0.4022**		**C − 0.5 (balanced)**	
	Predicted negative	**Predicted positive**	**Predicted negative**	**Predicted positive**	**Predicted negative**	**Predicted positive**
Actual negative	9,694	3,064	8,363	4,395	2,446	809
Actual positive	1,260	1,995	871	2,384	1260	1,995

“Positive" refers to the “successful" category, while “negative" refers to the “unsuccessful"

(a) Popular successful movies.

(b) Popular surprisingly successful movies.

Fig. 11 A sample of 20 successful film posters as classified using the EfficientNetB0 model. We show a set of successful movies with high budget (a - expected to be successful) and a set of successful movies with low budget (b - surprisingly successful)

success of their corresponding films. The core hypothesis of this research is that the quality of films, as reflected in their box office returns, can be deduced from these visual cues. This expands the scope of investigations into signaling, in line with Spence's signaling theory, recognizing that although quality signals are universally accessible, they can still effectively differentiate product quality.

Additionally, when compared with other studies, particularly those using trailers, the ability of a single poster to correlate with economic success is notably strong. This is especially significant because we rely solely on the film poster information, without incorporating other film metadata, to classify economic success.

The paper contributes significantly to computational research by establishing a robust methodology for using images to signal quality, applicable not only to movie posters but also to other experiential products such as book and music album covers. It introduces new analytical metrics for evaluating clustering outcomes from complex visual data, paving the way for interdisciplinary research approaches.

For film producers, this study highlights a practical takeaway: in an era dominated by visual media, a well-designed poster remains a powerful and appealing tool for signaling film quality, emphasizing the need for meticulous design and strategic use.

Acknowledgements This work was supported by the Coordination of Improvement of Higher Education Personnel (CAPES), and the São Paulo Research Foundation (FAPESP), grants 2021/10921-2 and 2013/07375-0.

Author Contributions **Thais Luiza Donega e Souza**: Conceptualization, Methodology, Software, Data curation, Formal analysis, Investigation, Writing - original draft, Visualization. **Caetano Mazzoni Ranieri**: Conceptualization, Investigation, Methodology, Validation, Writing - review & editing. **Anand Panangadan**: Methodology, Validation, Writing - review & editing. **Jó Ueyama**: Methodology, Validation, Writing - review & editing, Supervision, Funding acquisition. **Marislei Nishijima**: Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Funding This work was supported by the Coordination of Improvement of Higher Education Personnel (CAPES), for Thaís Luiza Donega e Souza, and the São Paulo Research Foundation (FAPESP), grants 2021/10921-2 and 2013/07375-0, for Caetano Mazzoni Ranieri.

Availability of data and materials The data that support the findings of this study are available from IMDB and BoxOfficeMojo sites. Restrictions may apply to the availability of these data. The collection and organization of the data studied are available from the authors upon reasonable request.
Code for data cleaning and analysis is available at https://colab.research.google.com/drive/15iWjGKxG5BbRYWVvQgQ6WSs738zUpV9s?usp=sharing, https://colab.research.google.com/drive/1GrNDAL8N4unurF4uDLBF9Eu08kApgtPU?usp=sharing, https://drive.google.com/file/d/1ecki2qqbAvJDuSWwWxsB7g4dZzdHW3wh/view?usp=sharing, and https://colab.research.google.com/drive/131ECgPpGdjKgqC7k1KabAu3H2_QDFFSU?usp=sharing.

Declarations

Competing interests The authors have no financial or proprietary interests in any material discussed in this article.

References

1. Varian HR (2000) Buying, sharing and renting information goods. J Ind Econ 48(4):473–488. https://doi.org/10.1111/1467-6451.00133
2. Walls WD (2005) Modeling movie success when "nobody knows anything": Conditional stable-distribution analysis of film returns. J Cult Econ 29(3):177–190. https://doi.org/10.1007/s10824-005-1156-5
3. Akdeniz MB, Talay MB (2013) Cultural variations in the use of marketing signals: A multilevel analysis of the motion picture industry. J Acad Market Sci 41(5):601–624. https://doi.org/10.1007/s11747-013-0338-5
4. Akerlof GA (1978) The market for "lemons": Quality uncertainty and the market mechanism. In: Uncertainty in Economics, Elsevier, ???, pp 235–251. https://doi.org/10.1016/B978-0-12-214850-7.50022-X
5. Zhang W, Skiena S (2009) Improving movie gross prediction through news analysis. In: 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, vol 1. pp 301–304. https://doi.org/10.1109/WI-IAT.2009.53. IEEE
6. Spence M (1978) Job market signaling, uncertainty in economics. Q J Econ 87(3):355–374
7. Spence M (2002) Signaling in retrospect and the informational structure of markets. Am Econ Rev 92(3):434–459. https://doi.org/10.1257/00028280260136200
8. Tous R, Gomez M, Poveda J, Cruz L, Wust O, Makni M, Ayguadé E (2018) Automated curation of brand-related social media images with deep learning. Multimed Tools Appl 77(20):27123–27142. https://doi.org/10.1007/s11042-018-5910-z
9. You Q, Bhatia S, Luo J (2016) A picture tells a thousand words-about you! user interest profiling from user generated visual content. Signal Process 124:45–53. https://doi.org/10.1016/j.sigpro.2015.10.032
10. Wang M, Li X, Liu Y, Chau P, Chen Y (2024) A contrast-composition-distraction framework to understand product photo background's impact on consumer interest in e-commerce. Decis Support Syst 178:114124. https://doi.org/10.1016/j.dss.2023.114124
11. Pérez-Núñez P, Díez J, Luaces O, Remeseiro B, Bahamonde A (2023) Users' photos of items can reveal their tastes in a recommender system. Inf Sci 642:119227. https://doi.org/10.1016/j.ins.2023.119227
12. Nanne AJ, Antheunis ML, Van Der Lee CG, Postma EO, Wubben S, Van Noort G (2020) The use of computer vision to analyze brand-related user generated image content. J Interact Mark 50(1):156–167. https://doi.org/10.1016/j.intmar.2019.09.003
13. Schroeder J (2002) Visual Consumption. Routledge Milton Park, Abingdon-on-Thames, Oxfordshire, England, UK
14. King E (2003) Movie Poster. Mitchell Beazley London
15. Zhou Y, Zhang L, Yi Z (2019) Predicting movie box-office revenues using deep neural networks. Neural Comput & Appli 31(6):1855–1865. https://doi.org/10.1007/s00521-017-3162-x
16. Xie J, Girshick R, Farhadi A (2016) Unsupervised deep embedding for clustering analysis. In: International Conference on Machine Learning, pp 478–487. PMLR
17. Guérin J, Thiery S, Nyiri E, Gibaru O, Boots B (2021) Combining pretrained cnn feature extractors to enhance clustering of complex natural images. Neurocomputing 423:551–571. https://doi.org/10.1016/j.neucom.2020.10.068

18. Diallo B, Hu J, Li T, Khan GA, Liang X, Zhao Y (2021) Deep embedding clustering based on contractive autoencoder. Neurocomputing 433:96–107. https://doi.org/10.1016/j.neucom.2020.12.094
19. Ringnér M (2008) What is principal component analysis? Nat Biotechnol 26(3):303–304. https://doi.org/10.1038/nbt0308-303
20. Tan M, Le Q (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, pp 6105–6114. PMLR
21. Jolly S, Iwana BK, Kuroki R, Uchida S (2018) How do convolutional neural networks learn design? In: 2018 24th International Conference on Pattern Recognition (ICPR), pp 1085–1090. https://doi.org/10.1109/ICPR.2018.8545624. IEEE
22. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25**
23. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations (ICLR), pp 1–14
24. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 1–9
25. Ivasic-Kos M, Pobar M, Mikec L (2014) Movie posters classification into genres based on low-level features. In: 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp 1198–1203. https://doi.org/10.1109/MIPRO.2014.6859750. IEEE
26. Ivasic-Kos M, Pobar M (2017) Multi-label classification of movie posters into genres with rakel ensemble method. In: International Conference on Innovative Techniques and Application of Artificial Intelligence, pp 370–383. https://doi.org/10.1007/978-3-319-71078-5_31. Springer
27. Sirattanajakarin S, Thusaranon P (2019) Movie genre in multi-label classification using semantic extraction from only movie poster. In: Proceedings of the 2019 7th International Conference on Computer and Communications Management, pp 23–27. https://doi.org/10.1145/3348445.334847
28. Chu WT, Guo HJ (2017) Movie genre classification based on poster images with deep neural networks. In: Proceedings of the Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes, pp 39–45
29. Nambiar G, Roy P, Singh D (2020) Multi modal genre classification of movies. In: 2020 IEEE International Conference for Innovation in Technology (INOCON), pp 1–6. https://doi.org/10.1109/INOCON50539.2020.9298385. IEEE
30. Mangolin RB, Pereira RM, Britto AS, Silla CN, Feltrim VD, Bertolini D, Costa YM (2020) A multimodal approach for multi-label movie genre classification. Multimed Tools Appl 1–26. https://doi.org/10.1007/s11042-020-10086-2
31. Simões GS, Wehrmann J, Barros RC, Ruiz DD (2016) Movie genre classification with convolutional neural networks. In: 2016 International Joint Conference on Neural Networks (IJCNN), pp 259–266. https://doi.org/10.1109/IJCNN.2016.7727207. IEEE
32. Wehrmann J, Barros RC (2017) Movie genre classification: A multi-label approach based on convolutions through time. Appl Soft Comput 61:973–982. https://doi.org/10.1016/j.asoc.2017.08.029
33. Chen X, Zhao P, Xu J, Li Z, Zhao L, Liu Y, Sheng VS, Cui Z (2018) Exploiting visual contents in posters and still frames for movie recommendation. IEEE Access 6:68874–68881. https://doi.org/10.1109/ACCESS.2018.2879971
34. Wang Z, Zhang J, Ji S, Meng C, Li T, Zheng Y (2020) Predicting and ranking box office revenue of movies based on big data. Inf Fusion 60:25–40. https://doi.org/10.1016/j.inffus.2020.02.002
35. Ko MY, Li JL, Lee CC (2019) Learning minimal intra-genre multimodal embedding from trailer content and reactor expressions for box office prediction. In: 2019 IEEE International Conference on Multimedia and Expo (ICME), pp 1804–1809. https://doi.org/10.1109/ICME.2019.00310. IEEE
36. Zhou Y, Yen GG (2018) Evolving deep neural networks for movie box-office revenues prediction. In: 2018 IEEE Congress on Evolutionary Computation (CEC), pp 1–8. https://doi.org/10.1109/CEC.2018.8477691. IEEE
37. Matsuzaki Y, Okayasu K, Imanari T, Kobayashi N, Kanehara Y, Takasawa R, Nakamura A, Kataoka H (2017) Could you guess an interesting movie from the posters?: An evaluation of vision-based features on movie poster database. In: 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), pp 538–541. https://doi.org/10.23919/MVA.2017.7986919. IEEE
38. Iwana BK, Rizvi STR, Ahmed S, Dengel A, Uchida S (2016) Judging a book by its cover. arXiv preprint arXiv:1610.09204. https://doi.org/10.48550/arXiv.1610.09204
39. Zhu Y, Kiros R, Zemel R, Salakhutdinov R, Urtasun R, Torralba A, Fidler S (2015) Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In: Proceedings of the IEEE International Conference on Computer Vision, pp 19–27

40. Tadimari A, Kumar N, Guha T, Narayanan SS (2016) Opening big in box office? trailer content can help. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 2777–2781. https://doi.org/10.1109/ICASSP.2016.7472183. IEEE
41. Grill JB, Strub F, Altché F, Tallec C, Richemond P, Buchatskaya E, Doersch C, Avila Pires B, Guo Z, Gheshlaghi Azar M et al (2020) Bootstrap your own latent-a new approach to self-supervised learning. Adv Neural Inf Process Syst 33:21271–21284
42. Sculley D (2010) Web-scale k-means clustering. In: Proceedings of the 19th International Conference on World Wide Web pp. 1177–1178. https://doi.org/10.1145/1772690.1772862
43. Steinbach M, Karypis G, Kumar V (2000) A comparison of document clustering techniques
44. Von Luxburg U (2007) A tutorial on spectral clustering. Stat Comput 17(4):395–416. https://doi.org/10.1007/s11222-007-9033-z
45. Rosenberg A, Hirschberg J (2007) V-measure: A conditional entropy-based external cluster evaluation measure. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pp 410–420
46. Rendón E, Abundez I, Arizmendi A, Quiroz EM (2011) Internal versus external cluster validation indexes. Int J Comput Commun 5(1):27–34
47. Xie Q, Luong MT, Hovy E, Le QV (2020) Self-training with noisy student improves imagenet classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 10687–10698
48. Cui M et al (2020) Introduction to the k-means clustering algorithm based on the elbow method. Acc Audit Finan 1(1):5–8
49. Rahmatabadi AF, Bastanfard A, Amini A, Saboohi H (2022) Building movie recommender systems utilizing poster's visual features: A survey study. In: 2022 10th RSI International Conference on Robotics and Mechatronics (ICRoM), pp 110–116. https://doi.org/10.1109/ICRoM57054.2022.10025210. IEEE
50. Hoef H, Warrens MJ (2019) Understanding information theoretic measures for comparing clusterings. Behaviormetrika 46(2):353–370. https://doi.org/10.1007/s41237-018-0075-7
51. Singh J, Goyal G (2019) Anticipating movie success through crowdsourced social media videos. Comput Hum Behav 101:484–494. https://doi.org/10.1016/j.chb.2018.08.050
52. Amirkhani D, Bastanfard A (2019) Inpainted image quality evaluation based on saliency map features. In: 2019 5th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), pp 1–6. https://doi.org/10.1109/ICSPIS48872.2019.9066140. IEEE

Authors and Affiliations

Thaís Luiza Donega e Souza[1] · Caetano Mazzoni Ranieri[2] · Anand Panangadan[5] · Jó Ueyama[3] · Marislei Nishijima[4]

✉ Caetano Mazzoni Ranieri
cm.ranieri@unesp.br

Thaís Luiza Donega e Souza
thais.donega@usp.com

Anand Panangadan
apanangadan@fullerton.edu

Jó Ueyama
joueyama@icmc.usp.br

Marislei Nishijima
marislei@usp.br

[1] School of Arts, Sciences and Humanities (EACH), University of São Paulo (USP), Rua Arlindo Bettio, 1000, 03828000 SP, São Paulo, Brazil

[2] Institute of Geosciences and Exact Sciences (IGCE), São Paulo State University (UNESP), Avenida 24 A, 1515, 13506692 Rio Claro, SP, Brazil

[3] Institute of Mathematical and Computer Sciences (ICMC), University of São Paulo (USP), Avenida Trabalhador Sao-Carlense, 400, 13566590 São Carlos, SP, Brazil

[4] Institute of International Relations (IRI), University of São Paulo (USP), Avenida Prof. Lucio Martins Rodrigues, Travessa do Labirinto, 286, 05508020 São Paulo, SP, Brazil

[5] Computer Science Department, California State University, 800 N State College Blvd, 92831 Fullerton, CA, USA