Face the Music: Can an Album's Cover Art Predict its Popularity?

Jessica Joy, Rachna Rawalpally, Samyu Vakkalanka, Marie Vaughan
DSAN 6600
April 29, 2025

Abstract: Album covers serve as a critical visual interface between artists and listeners, shaping first impressions and setting the emotional tone for the music. In this study, we developed an artificial neural network to investigate whether artistic styles, colors, and thematic elements of album cover art can predict an album's popularity. We examined questions such as whether female artists display more sexualized imagery compared to male artists, whether sexualized covers attract more listeners, and how composition correlates with album success. Our findings reveal distinct visual patterns associated with popularity across artist genders and genres. While stylistic differences emerged, such as pop albums favoring vibrant imagery and rap/hip-hop albums tending toward darker, more complex designs, the existing popularity of the artist proved to be a major factor in predicting album success.

1. Introduction

Album covers serve as the visual gateway to an artist's music. They are often the first thing a potential listener encounters, whether browsing in stores or scrolling on digital platforms. Through cover art, artists can generate interest, establish expectations, and set the emotional tone for their music. As Emily Dolan notes in "Toward a Musicology of Interfaces," *"the album cover can be thought of as an instrument through which we interface with sound"* (Vad, 2021).

Previous research has explored the relationship between album cover art and musical attributes. For instance, Greenfield and Paintsil (2024) applied a DenseNet architecture to examine whether cover art could serve as an effective proxy for genre classification. Similarly, Mohan and Swamydoss (2024) developed music recommendation systems based on visual features extracted from album covers. However, there has been limited investigation into whether machine learning techniques can leverage album art to predict an artist's popularity. Joye and Fennis (2023) acknowledged this gap, conducting statistical analyses to study how visual characteristics of album covers influence music consumption, though without employing machine learning approaches.

This paper aims to address this research gap by developing an artificial neural network to predict album popularity based on cover art features. We specifically examine how elements such as imagery, text, composition, color, nudity, and visible skin are associated with album success. Our goal is to uncover patterns that could help emerging artists make more strategic design choices to enhance audience engagement and visibility in an increasingly competitive and saturated music landscape.

2. Data

2.1 Albums and Genres

To build our dataset, we first used the Spotify API to search for albums released in the past five years, collecting details such as album name, artist name, release date, and popularity, yielding about 1,500 albums. We then used Wikipedia to determine each artist's gender, narrowing the list to 860 artists with existing Wikipedia pages. Using these artists, we pulled additional Spotify data including genres, popularity, and follower count and retrieved up to five top albums per artist, compiling a final dataset of 3,289 albums. Since Spotify only provides artist-level genre information, we found album-specific genres by scraping Wikipedia infoboxes and, when necessary, using the Last.fm API as a fallback. We then standardized the varied genre labels from these sources into 16 broader categories to ensure consistency across the dataset.

2.2 Generated Columns

To extract features from album covers for training our ANN, we utilized two pretrained models and multiple Python libraries. For common object information, we utilized YOLOv8, a pretrained object detection model on the COCO (Common Objects in Context) dataset. We focused on detecting people, cars, and animals, and recorded their counts and confidence scores. To quantify how provocative or sexual an album cover appeared, we used NudeNet, a pretrained image classification model developed by the GitHub user "notAI-tech." NudeNet detects the presence of nudity and identifies specific types of exposed or covered body parts. After running our image dataset through NudeNet, we generated one-hot encoded columns for features such as "female breast exposed," "female breast covered," "belly exposed," "feet exposed," and "buttocks exposed," among others.

Several Python libraries were used to extract additional visual features from the album covers, including color detection, skin percentage, textual features, and composition features. The colorsys library was used to convert RGB values to HSV color space, allowing colors to be mapped to named color categories. In addition, the ColorThief library processed each image to extract the top five dominant colors. Further color analysis included transforming each cover into RGB, HSV, and grayscale arrays to compute brightness and its variability. We were then able to flag images as vibrant, muted, and warm or cool color palettes. To estimate the percentage of the album cover occupied by skin, images were converted to HSV color space, and the number of pixels falling within a predefined skin color range was calculated. This count was then divided by the total number of pixels in the image to obtain the skin coverage percentage. This metric provided insight into how prominently close-up images of faces or bodies appeared on the album covers. For text-based features, we applied optical character recognition (OCR) to analyze the function of words on the album cover. The EasyOCR library allowed us to record distinct text elements, search for keywords like "explicit", and calculate a text coverage ratio to quantify how much of the cover is occupied by text. Lastly, we used OpenCV's Canny edge detection library to obtain album composition features. The cover was turned into a grayscale image, and the algorithm detects edge pixels around shapes, textures, and other details to calculate an edge density score. A higher score was classified as a complex design, while a lower score reflected a simpler, more minimal look. Our final dataset contained 100+ features about the album cover that we explored in our analysis.

3. Models

3.1 Convolutional Neural Network

We approached the task of predicting album popularity from cover images as a binary classification problem. Instead of predicting the exact popularity score, we created a binary target variable by thresholding the Spotify popularity metric at 65, labeling albums as "Top" (1) or "Not Top" (0). This allowed us to focus on identifying visual patterns linked to higher success. Album covers were resized to 64×64 pixels for efficiency while preserving detail, and pixel values were normalized to [0,1] to stabilize training. Using TensorFlow's tf.data API, we organized the images and labels into a dataset, split 80% for training and 20% for validation, and batched the data in groups of 32. To improve generalization, we applied data augmentation to the training set, including random flips, rotations, and zooms, encouraging the model to focus on meaningful visual features rather than surface-level patterns.

Our convolutional neural network (CNN) consisted of three convolutional blocks followed by fully connected layers. Each block included 32 filters with 3×3 kernels, L2 regularization to prevent overfitting, batch normalization, ReLU activation for non-linearity, 2×2 max pooling to reduce dimensions, and 30% dropout for additional regularization. The dense layers flattened the feature maps, applied a fully connected layer with 64 neurons and ReLU activation, and concluded with a single sigmoid-activated neuron for binary classification. The model was trained using binary cross-entropy loss, the Adam optimizer, and accuracy as the evaluation metric. We implemented early stopping with a patience of three epochs and a maximum of 50 epochs, although training typically stopped earlier. By combining L2 regularization, batch normalization, dropout, and augmentation, we aimed to maximize model robustness given the dataset's limited size and the complexity of the prediction task.

3.2 Artificial Neural Network

For this model, we also framed album popularity as a binary classification problem, labeling anything above 65% popularity as "popular" and the rest as "not popular". To prepare our inputs, we began by standardizing all features using StandardScalar to ensure they are on the same scale. Then, SMOTE (Synthetic Minority Oversampling Technique) was used to upsample the minority class due to our class imbalance. Since we had over 100 features as our input, feature selection was done through a preliminary simple Random Forest to obtain feature importance and retain all features above the median. This left us with about 30 features and dramatically reduced dimensionality while preserving the most informative features.

Our artificial neural network architecture began with ingesting these selected features and a Gaussian Noise layer that injects small perturbations of noise during training to help with regularization. Next, we pass the data through two fully connected hidden layers: the first layer with 112 neurons and the second with 64 neurons, both using ReLU activations to introduce nonlinearity. Each dense layer was regularized with an L1 penalty of lambda=0.001 to encourage sparsity. The dense layers were then followed by batch normalization to stabilize the activations and speed up convergence. After normalization, we applied a dropout layer with a rate of 50 percent. This layer randomly zeros out half the neurons during each update to prevent the model from memorizing a path, reducing overfitting. The final output layer is a single neuron with a sigmoid activation that outputs the probability of an album cover being popular or not. The model was compiled with an Adam optimizer at a learning rate of 0.001. Instead of binary cross-entropy as our loss function, we used focal loss, which performs better for imbalanced classes. Focal loss extends standard cross-entropy and forces the model to focus on challenging, misclassified cases, dynamically down-weighting easy examples. Hyperparameter tuning resulted in an alpha of 0.6, controlling the weight of the minority class, and a gamma of 2, controlling the downweighting of well-classified examples. Early stopping and a custom SMOTE callback were utilized to ensure that training is stopped before overfitting occurs.

Finally, to boost our results, we took an ensemble approach and combined our ANN with an independently tuned XGBoost classifier. We simply averaged the predictions of both models to yield a final prediction for each album. This approach utilized XGBoost's ability to capture complex feature interaction and ANN's capacity for learning deep, nonlinear representations. This ensemble model consistently outperformed either model alone, achieving a higher AUC, precision, and recall on the test set, demonstrating more stable performance and better generalization.

4. Results

4.1 Convolutional Neural Network

Our CNN model achieved limited success in predicting album popularity from cover art, with a classification accuracy of approximately 56%—only slightly better than random guessing—and an AUC of 0.51. As shown in the confusion matrix, the model correctly classified 273 unpopular and 85 popular albums but misclassified 300 albums overall.

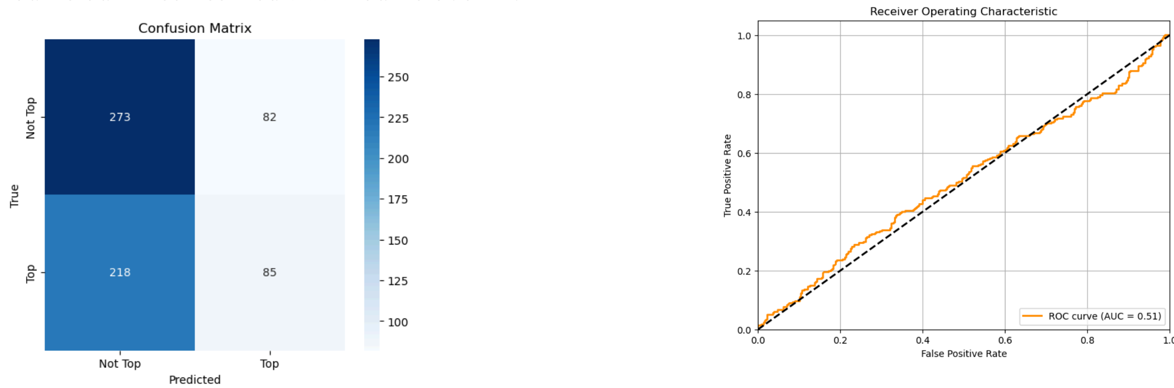


Figure 1: Confusion matrix and ROC curve for CNN

This model performed poorly because of the small dataset size (3,289 albums), which is far below the number typically needed for CNNs to learn meaningful visual patterns. Also, album covers have highly diverse and subjective artistic styles, making it hard for the model to consistently link visual elements to commercial success. These results suggest that cover art alone may not provide sufficient information to predict album popularity, as visual aesthetics are only one piece of a much broader and more complex consumer decision-making process.

4.2 Artificial Neural Network

In our initial iterations, we saw a large improvement over the baseline CNN performance. To boost performance even further, we explored a better way to incorporate artist popularity into our model. Due to Spotify's black box nature of their metrics, we were unsure how artist and album popularity was calculated and feared introducing data leakage into our model. However, after examining the correlation between artist and album popularity, we determined that artist popularity is a valid feature. The relationship isn't perfect since many well-known artists release albums that underperform, so including it adds valuable, independent information.

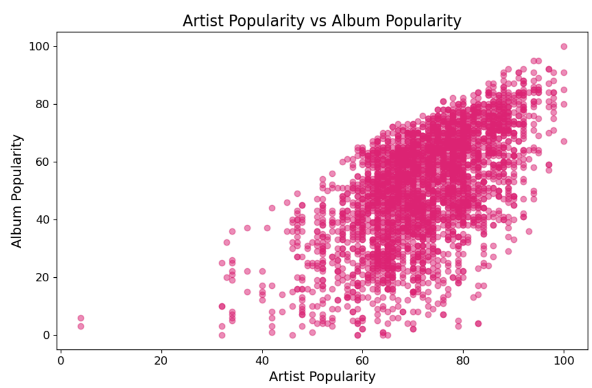


Figure 2: Scatter plot of Artist Popularity and Album Popularity

After including artist popularity in our model, our performance increased. When evaluated on the test set, the ensemble achieved the highest AUC of 0.81, compared with 0.81 for the neural network and 0.80 for XGBoost. The ensemble had an accuracy of 0.76 at the optimal threshold of 0.50.

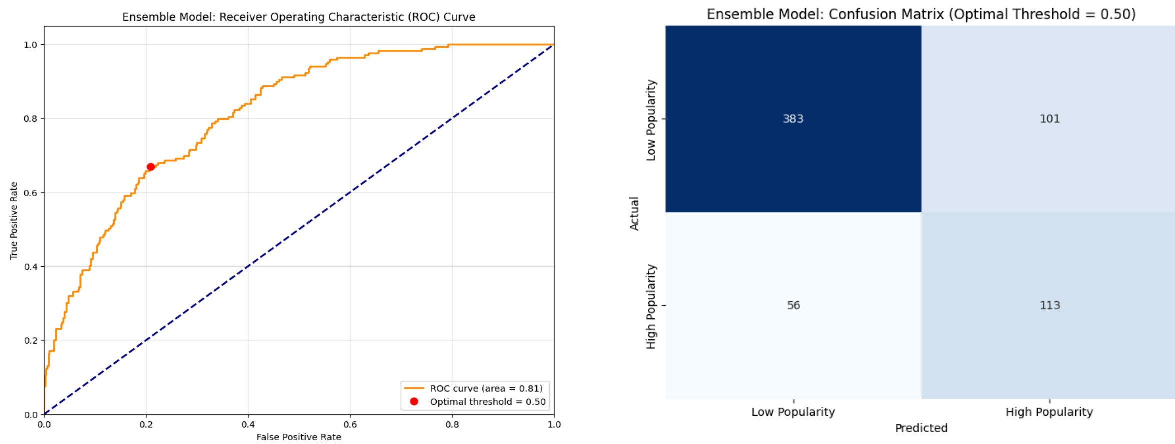


Figure 3: ROC curve and confusion matrix for ensemble model

Finally, using our tuned XGBoost model, we extracted feature importances. Artist popularity ranks highest, which is unsurprising. This is followed by composition-related characteristics such as the album's minimalism and visual attributes, including the presence of a face on the cover, the amount of text, color vibrance, and percentage of skin, all of which help answer our original questions.

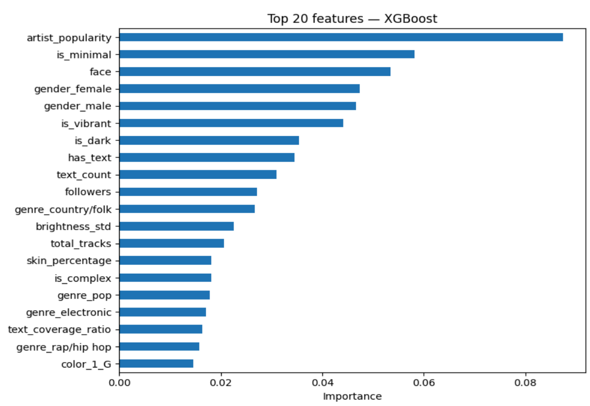
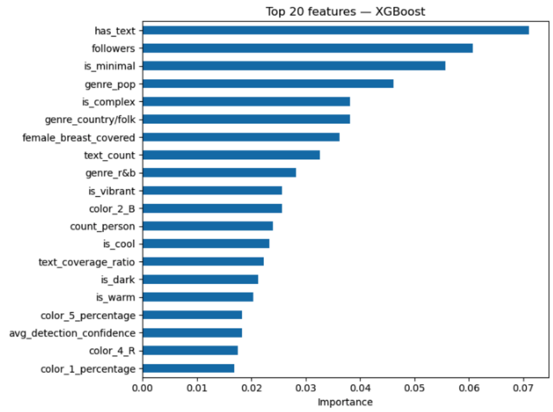


Figure 4: Feature importance XGBoost model

4.2.1 Gender and Genre

Using the ANN and XGBoost models, we further explored the influence of gender and genre on album art. By analyzing the top features from the XGBoost model, we found that minimalism and the presence of people were strong indicators across both genders. Key differences emerged: female albums tend to feature covered body parts, while male albums often showcase faces and people. Focusing on the two most popular genres, pop and rap/hip-hop, we observed similarities and distinct contrasts. Both genres commonly incorporate text, but pop albums stand out with faces, a minimal aesthetic, and vibrant colors. In contrast, rap/hip-hop albums tend to feature muted tones, more complex designs, and a focus on bold imagery.

Female Artists: Male Artists:



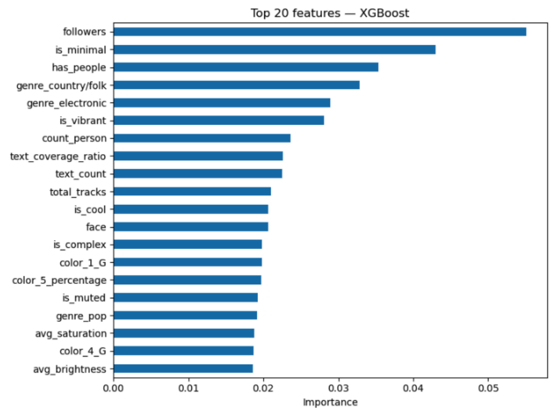
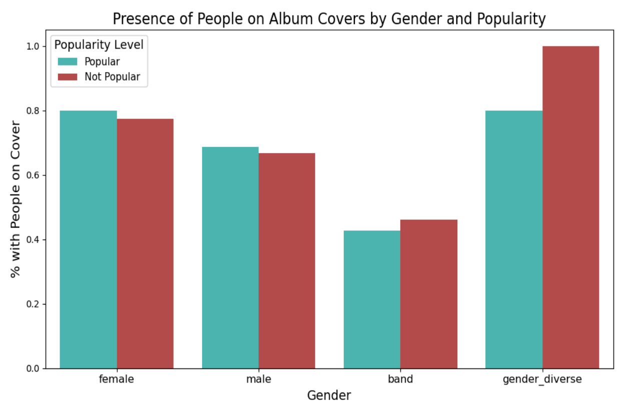


Figure 5: Feature Importance XGBoost model for Gender (female and male)



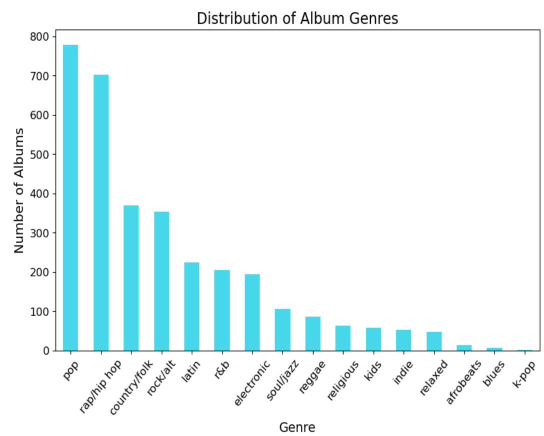
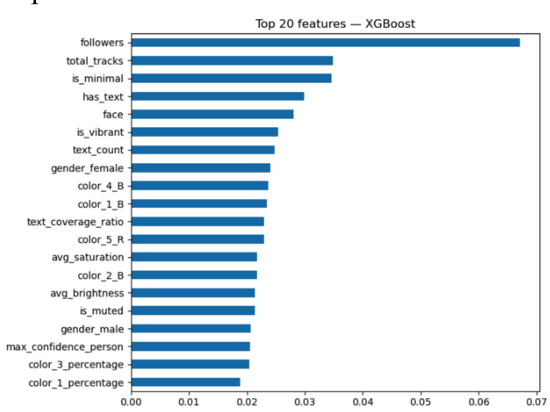


Figure 6: Bar graphs showing gender and genres

Pop Albums: Rap/Hip-Hop:



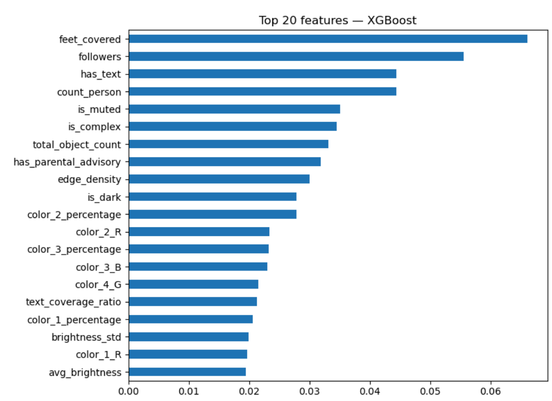


Figure 7: Feature Importance XGBoost for Genre (pop and rap/hip-hop)

5. Limitations

Our study faced key limitations that impacted data collection and interpretation. Spotify's increasing API restrictions, including reduced access to detailed track, album, and genre data, made it harder to analyze music popularity patterns and industry trends, while also limiting visibility for artists and analysts. Rate limits further shrank our dataset, weakening our model's ability to detect meaningful patterns in album cover features. Additionally, relying on Spotify's streaming-based popularity metric rather than physical sales was a limitation because cover art plays a stronger role in physical purchases. These challenges highlight the need for more transparent data access to support research and innovation across the music ecosystem.

6. Conclusion

Our main goal was to determine whether women show more skin on their album covers than men, and how that might relate to album popularity. While we found that women, on average, display more skin than men, the difference was relatively small and did not significantly drive album popularity. Instead, the artist's popularity was the most significant predictor of album popularity. However, we did observe essential gender differences. Specifically, albums by popular female artists tended to perform slightly worse in popularity compared to those by popular male artists. We also noted stylistic differences across genres: for example, pop albums generally featured more vibrant, colorful designs, whereas rap/hip-hop albums leaned toward darker, more complex imagery.

These findings have important implications for both artists and marketers. Artists can use this information to better understand what visual trends are successful within their genre, potentially helping guide their album art choices. Marketers, particularly those promoting lesser-known or emerging artists, could focus on increasing artist visibility and engagement since artist popularity is the strongest driver of success. Overall, our results highlight how gender biases persist within album cover art and music promotion, and emphasize the need to identify and address these biases when marketing artists.

We can see several interesting ways to expand on this work in further work. One avenue would be incorporating social media data, such as artists' Instagram activity, post engagement, and marketing strategies, to better understand how online presence influences album popularity. Another potential direction would be analyzing the evolution of album cover art over time, for instance, investigating whether the rise of streaming has changed how artists design their covers in the digital age. Exploring these areas would help deepen our understanding of the intersection between gender, marketing, and success in the music industry.

7. References

Dolan, Emily. "The Album Cover." *Journal of Popular Music Studies*, vol. 33, no. 3, 2021, pp. 11–18, https://doi.org/10.1525/jpms.2021.33.3.11.

Lin, Jasmine, and Peter Zeng. "Album Covers Deserve Some Attention." *CS231N: Convolutional Neural Networks for Visual Recognition*, Stanford University, Spring 2024, https://cs231n.stanford.edu/2024/papers/album-covers-deserve-some-attention.pdf.

Joye, Yannick, and Bob M. Fennis. "How the Eye Buys Music: The Influence of Visual Aesthetic Design on Music Purchase Intentions." *Psychology & Marketing*, vol. 41, no. 1, 2024, pp. 3–16, https://doi.org/10.1002/mar.21959.

Sharma, Deepak. "Image-Based Music Recommendations." *International Research Journal*, vol. 4, no. 2, 2024, pp. 32–41, https://www.irjweb.com/Image%20Based%20Music%20Recommendations.pdf.