Genetic Syndromes Classification Report

1. Introduction

This report presents the results of genetic syndrome classification using KNN with Euclidean and Cosine distance metrics.

The objective is to analyze the performance of these metrics using image embeddings, generating visualizations, and evaluating the model with various metrics.

2. Classification Results

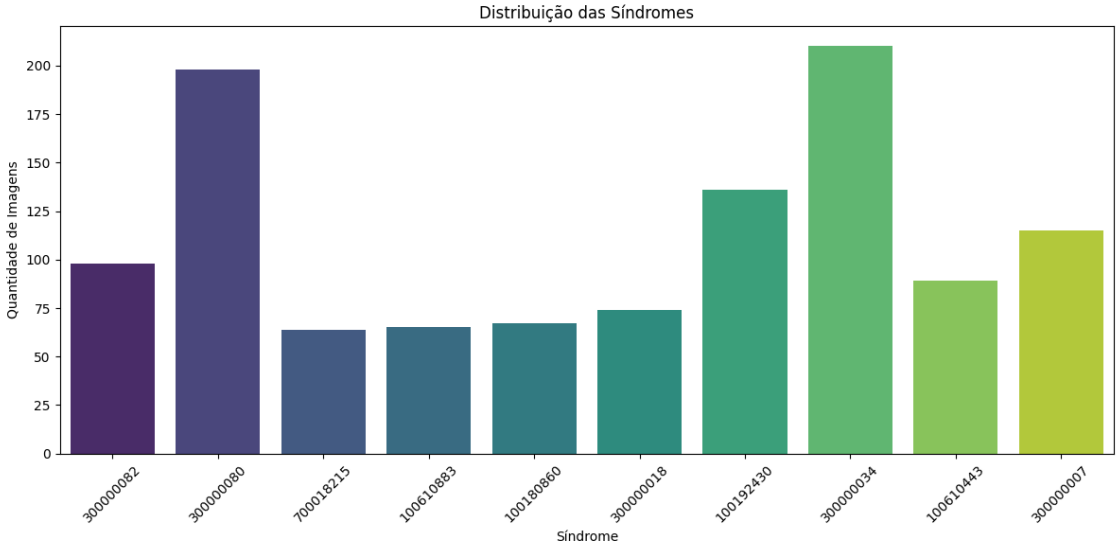
Performance Metrics Comparison

Metric	Euclidean	Cosine
Best K	15	7
F1-Score	0.75	0.80
AUC ROC	0.95	0.96
Top-k Accuracy	0.92	0.93

3. Visualizations and Analyses

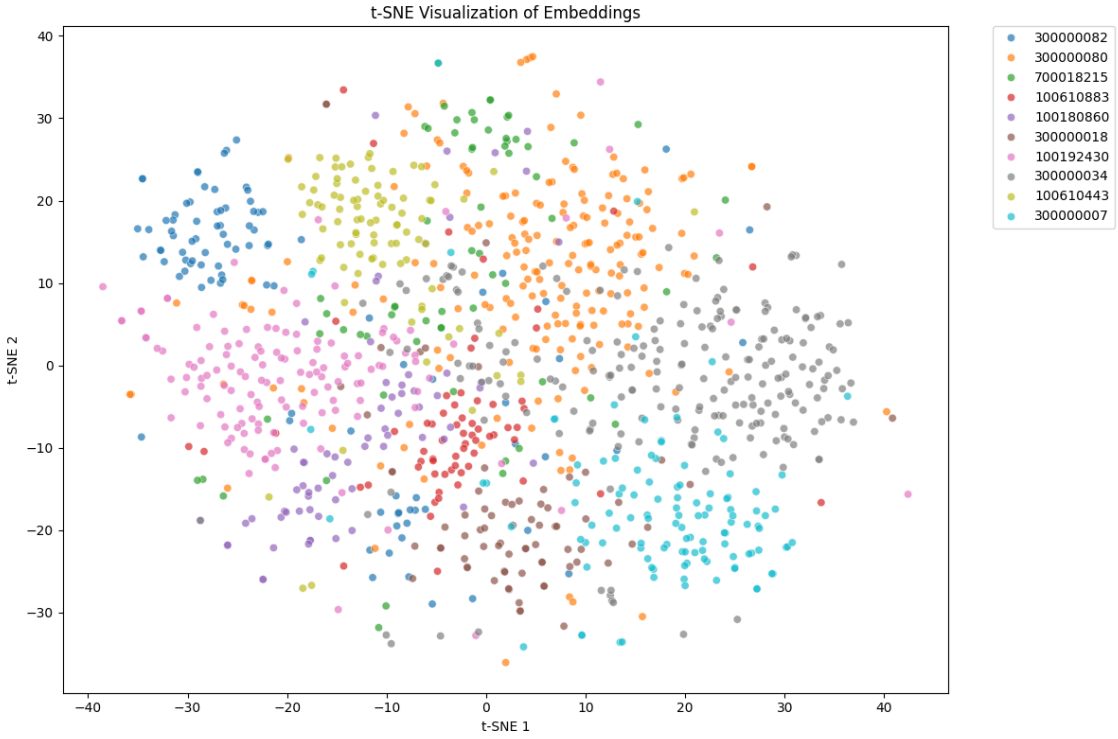
Syndrome Distribution:

This graph shows the number of images per syndrome. Note the class imbalance, which can impact model performance.



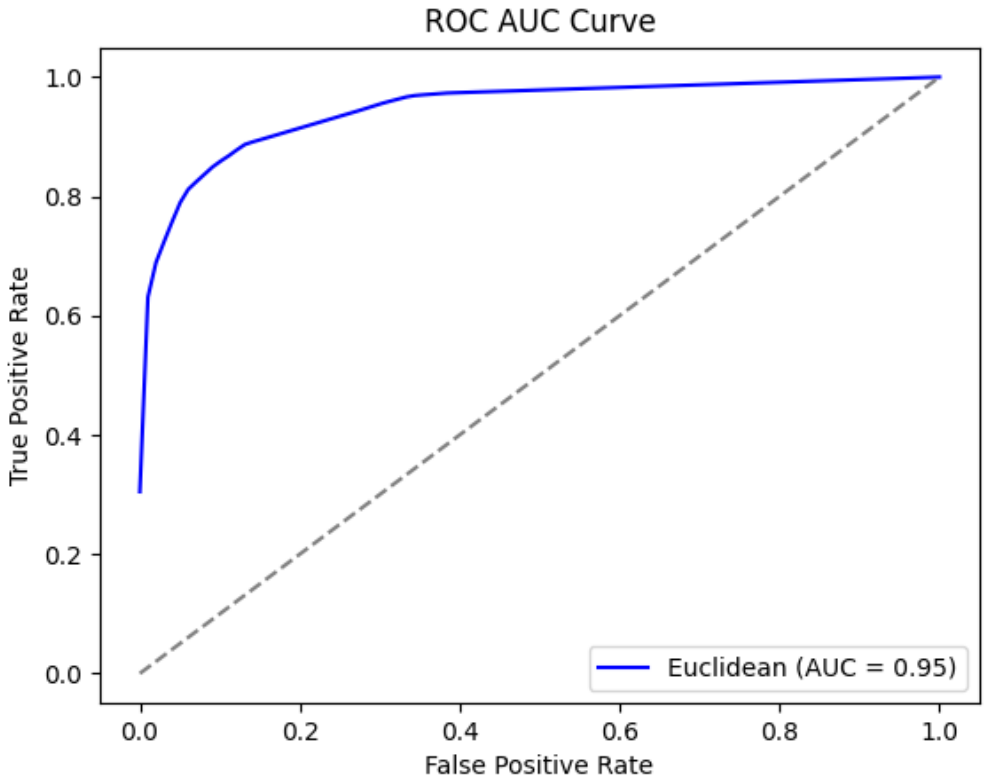
t-SNE Visualization:

t-SNE projection reduces the dimensionality of embeddings to 2D, revealing possible clusters that aid classification.



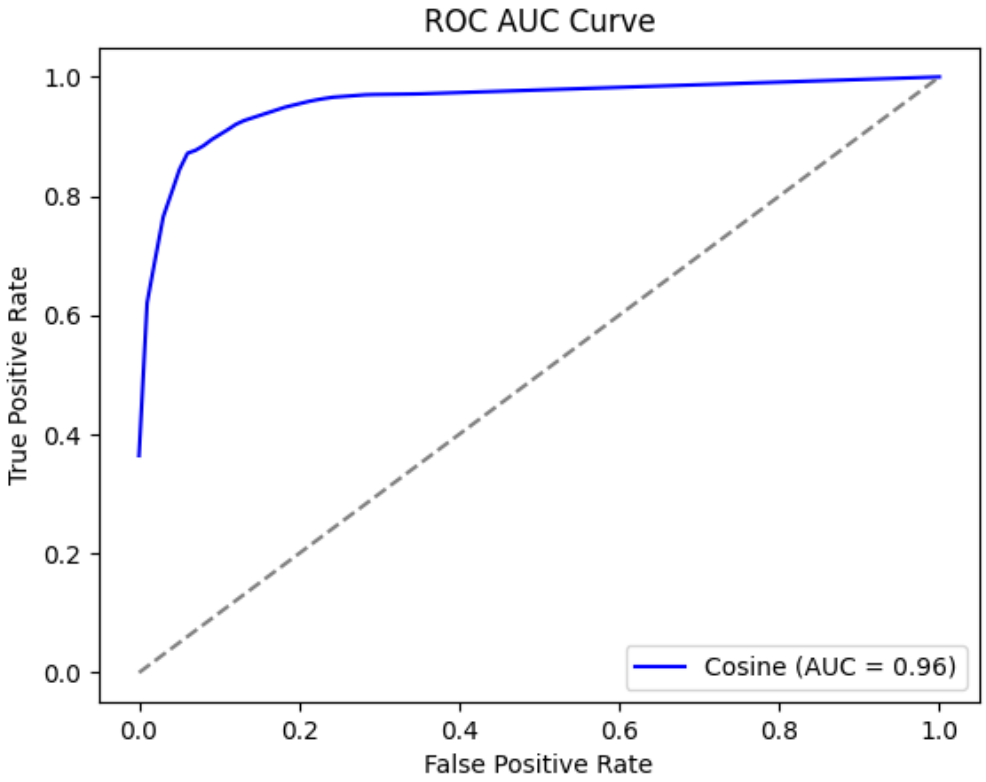
ROC AUC Curve - Euclidean:

The ROC AUC curve shows the relationship between True Positive Rate and False Positive Rate using Euclidean distance.



ROC AUC Curve - Cosine:

The ROC AUC curve shows the relationship between True Positive Rate and False Positive Rate using Cosine distance.



4. Conclusion

The Cosine distance metric showed better performance compared to Euclidean due to its sensitivity to angular directions in high-dimensional spaces.

Future analyses may include other classification algorithms and data augmentation techniques to address class imbalance.