**Multi-horizon Stock Market Prediction with Attention-based Transformers**

**1. Motivation and Introduction**

Financial market prediction remains one of the most challenging domains in applied machine learning, with significant implications for investment strategies, risk management, and economic policy. Traditional time series forecasting methods like ARIMA and statistical models have been the cornerstone of quantitative finance for decades. However, these approaches often struggle to capture the complex non-linear relationships and long-range dependencies inherent in financial data. Recent advancements in deep learning, particularly transformer architectures with self-attention mechanisms, have shown remarkable success in domains requiring temporal context understanding and pattern recognition.

The financial markets represent an ideal testing ground for these advanced architectures due to their inherent complexity, where price movements are influenced by a multitude of factors including economic indicators, company performance, market sentiment, and global events. While transformer models have revolutionized natural language processing and are beginning to make inroads in time series forecasting, their application to multi-horizon financial prediction remains relatively unexplored, especially in capturing regime shifts and incorporating external factors.

This research aims to bridge this gap by evaluating the effectiveness of transformer-based models with attention mechanisms for stock market prediction across multiple time horizons (daily, weekly, monthly) compared to traditional forecasting methods. By incorporating multiple prediction horizons, we can evaluate the models' ability to capture both short-term fluctuations and longer-term trends, which is crucial for different trading strategies. Additionally, we will explore the integration of sentiment analysis from financial news, providing a more comprehensive approach to market prediction that incorporates both quantitative price data and qualitative market sentiment.

The practical implications of this research extend beyond academic interest. More accurate and robust prediction models could enhance investment decision-making, improve risk management strategies, and potentially identify market inefficiencies. Furthermore, the interpretability aspects of attention mechanisms offer insights into which features and temporal patterns contribute most significantly to predictions, addressing the "black box" criticism often leveled at complex machine learning models in finance.

**2. Research Question**

Can transformer-based models with attention mechanisms outperform traditional time series models in predicting stock price movements across multiple time horizons (daily, weekly, monthly), and how does the incorporation of sentiment analysis affect prediction accuracy?

**3. Initial Review**

The application of machine learning to financial forecasting has evolved significantly in recent years. Hsu et al. (2020) conducted a comprehensive review of deep learning approaches for financial time series prediction, noting that while CNNs and RNNs have shown promise, they often struggle with capturing long-range dependencies crucial for financial forecasting. This limitation provides a natural opening for transformer architectures, which excel at modeling long-range dependencies through self-attention mechanisms.

Transformer models, first introduced by Vaswani et al. (2017) for machine translation, have been adapted for time series forecasting by Li et al. (2019) in their Informer model, which demonstrated superior performance over RNN-based models for long sequence time-series forecasting. However, their work did not specifically address the unique challenges of financial data, such as its high noise-to-signal ratio and non-stationarity.

Addressing financial applications more directly, Jiang et al. (2021) implemented a transformer-based model for stock trend prediction, reporting improvements over LSTM models. However, their work focused solely on daily predictions and did not explore multi-horizon forecasting or the integration of sentiment analysis.

The integration of textual data for financial prediction has been explored by Yang et al. (2020), who demonstrated that combining price data with financial news sentiment can improve prediction accuracy. Similarly, Sawhney et al. (2020) used BERT embeddings of financial news to enhance stock movement prediction, though they did not incorporate transformer architectures for the time series component.

A significant gap in the literature exists regarding the application of transformer models to multi-horizon financial forecasting with multimodal inputs. Additionally, the interpretability of these models in the financial context, particularly the attention weights and their relationship to market regimes, remains underexplored. This research aims to address these gaps by developing a comprehensive framework for multi-horizon stock prediction that leverages both quantitative price data and qualitative sentiment information.

**4. Data Sources & Statistics**

For this research, we will utilize several complementary datasets:

1. **Yahoo Finance API**:
   * Daily stock price data (Open, High, Low, Close, Volume) for 100 companies from the S&P 500 index spanning from January 2015 to December 2024
   * Approximately 250,000 data points (100 stocks × 2500 trading days)
   * Data will be split into training (70%, 2015-2021), validation (15%, 2022-2023), and testing (15%, 2024) sets
2. **Alpha Vantage API**:
   * Macroeconomic indicators including GDP, unemployment rates, inflation metrics, and interest rates
   * Approximately 1,000 data points (10 indicators × 100 monthly observations)
   * Will be used as additional features to provide economic context
3. **Financial News Articles**:
   * Data from Kaggle's "Financial News Sentiment Analysis" dataset containing 4,000+ news articles with sentiment labels
   * NASDAQ's Market News API for real-time financial news (approximately 50,000 articles)
   * Features extracted will include sentiment scores, entity mentions, and topic classifications
4. **Market Regime Indicators**:
   * VIX index data (volatility)
   * TED spread (credit risk)
   * Yield curve data
   * Will be used to identify different market regimes for analysis

The dataset will be preprocessed to handle missing values, normalize features, and align temporal data from different sources. Technical indicators (e.g., Moving Averages, RSI, MACD) will be calculated from price data and included as additional features. For sentiment analysis, we will use pretrained financial language models to extract sentiment scores from news articles related to the stocks in our dataset.

**5. Methods**

We will implement and compare the following methods:

**Traditional Machine Learning Models**:

1. **ARIMA (AutoRegressive Integrated Moving Average)**:
   * A classical time series forecasting method that combines autoregression, differencing, and moving average components
   * Will serve as a baseline for comparison
   * Hyperparameters will include p, d, q values optimized using grid search
   * Implementation will use statsmodels library in Python
2. **XGBoost with Time-based Features**:
   * An ensemble method using gradient boosting that has shown strong performance in many forecasting tasks
   * Will incorporate time-based features such as technical indicators, lagged values, and seasonality components
   * Hyperparameters to be optimized include learning rate, max depth, number of estimators
   * Implementation will use the XGBoost library in Python

**Deep Learning Models**:

1. **Temporal Fusion Transformer (TFT)**:
   * A state-of-the-art transformer architecture specifically designed for multi-horizon forecasting
   * Includes specialized components for temporal data: variable selection networks, gated residual networks, and multi-head attention mechanisms
   * Capable of handling mixed variables (categorical and continuous) and providing interpretable attention weights
   * Implementation will use PyTorch
2. **Time Series Transformer with Sentiment Integration**:
   * A modified transformer architecture that incorporates both price data and sentiment features
   * Will use a dual-attention mechanism to attend to both price patterns and relevant sentiment information
   * Will be implemented with uncertainty quantification through probabilistic outputs
   * Implementation will use PyTorch and Hugging Face's Transformers library

For each model, we will conduct extensive hyperparameter tuning using Bayesian optimization to ensure fair comparison. Additionally, we will implement ensemble methods combining predictions from different models to explore potential performance improvements.

**6. Evaluation Methods**

The evaluation of our models will be comprehensive, focusing on both prediction accuracy and practical applicability:

**Regression Metrics**:

* Mean Absolute Error (MAE): To measure the average magnitude of errors without considering direction
* Root Mean Squared Error (RMSE): To emphasize larger errors, which is particularly important in financial applications
* Mean Absolute Percentage Error (MAPE): To provide a scale-independent measure of prediction accuracy

**Direction Accuracy Metrics**:

* Directional Accuracy: Percentage of correctly predicted price movement directions
* F1 Score: Harmonic mean of precision and recall for direction prediction
* Matthews Correlation Coefficient (MCC): A balanced measure that works well even with imbalanced classes

**Financial Performance Metrics**:

* Sharpe Ratio: Risk-adjusted return measure for trading strategies based on model predictions
* Maximum Drawdown: Largest peak-to-trough decline in portfolio value
* Profit & Loss (P&L): Simulated trading performance using model predictions

**Model Interpretability Assessment**:

* Attention Weight Analysis: Evaluating which temporal patterns and features receive the highest attention
* SHAP (SHapley Additive exPlanations) Values: To understand feature importance and contribution to predictions
* Regime-based Performance Analysis: Evaluating model performance across different market regimes

Each model will be evaluated on predictions across multiple time horizons (1-day, 5-day, 20-day), allowing us to assess how performance varies with forecast horizon. Statistical significance tests will be conducted to determine if performance differences between models are statistically significant.

**Bibliography**

Hsu, M. W., Lessmann, S., Sung, M. C., Ma, T., & Johnson, J. E. (2020). Bridging the divide in financial market forecasting: Machine learners vs. financial economists. Expert Systems with Applications, 61, 215-234.

Jiang, W., Zhang, L., & Wu, C. (2021). Transformer-based deep learning models for financial time series forecasting. Neurocomputing, 438, 89-101.

Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y. X., & Yan, X. (2019). Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. Advances in Neural Information Processing Systems, 32.

Sawhney, R., Agarwal, S., Wadhwa, A., & Shah, R. R. (2020). Deep attentive learning for stock movement prediction from social media text and company correlations. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

Yang, L., Xu, Z., & Xiang, J. (2020). Incorporating financial news for stock trend prediction with attention-based recurrent neural network. Expert Systems with Applications, 163, 113792.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12), 11106-11115.

**Multimodal Learning for Early Detection of Chronic Disease Progression**

**Motivation and Introduction**

Chronic diseases represent a significant global healthcare challenge, with conditions like diabetes, cardiovascular disease, and chronic kidney disease affecting millions worldwide. Early detection and accurate progression prediction are crucial for effective intervention and improved patient outcomes. Despite advances in medical diagnostics, current approaches often rely on isolated data sources and fail to capture the complex, multifaceted nature of disease progression.

Healthcare data exists in various modalities - structured clinical measurements, unstructured clinical notes, medical imaging, and temporal patient histories. Each modality provides unique but complementary information about a patient's condition. Traditionally, these data types have been analyzed separately, potentially missing valuable cross-modal patterns that could improve diagnostic accuracy and progression prediction.

Multimodal learning techniques offer a promising approach by integrating heterogeneous data sources to create more comprehensive patient representations. Recent advances in deep learning have demonstrated success in combining text, tabular, and image data in various domains. However, significant challenges remain in healthcare applications, including temporal alignment of different modalities, handling missing data, and developing interpretable models that can support clinical decision-making.

This research aims to develop and evaluate novel multimodal learning approaches for chronic disease progression prediction, with a particular focus on addressing the challenges of data integration, temporal modeling, and interpretability. By leveraging the complementary information across clinical notes, laboratory results, and imaging data, we aim to create more accurate and robust predictive models that can detect subtle signs of disease progression before conventional clinical markers show significant changes.

**Research Question**

How can multimodal learning frameworks effectively integrate clinical notes, laboratory results, and imaging data to improve early detection and progression prediction of chronic diseases?

**Initial Review**

Recent literature has explored various aspects of multimodal learning in healthcare, though significant gaps remain in developing comprehensive frameworks for chronic disease progression.

Rajkomar et al. (2018) demonstrated the potential of deep learning for predicting a range of clinical outcomes using electronic health records. Their approach achieved impressive performance but primarily focused on structured data rather than truly multimodal integration. This highlights the need for more comprehensive data fusion techniques.

Chen et al. (2021) proposed a multimodal approach combining clinical notes and structured data for mortality prediction. They utilized BERT for text encoding and demonstrated improved performance over unimodal approaches. However, their work did not address the temporal aspects of disease progression or incorporate imaging data.

Huang et al. (2020) explored multimodal fusion of imaging and clinical data for Alzheimer's disease progression prediction. Their model showed promising results but lacked mechanisms for handling missing data and temporal alignment, which are common challenges in real-world healthcare settings.

Alsentzer et al. (2019) introduced ClinicalBERT, a domain-specific language model for clinical text. While not directly addressing multimodal learning, their work provides valuable foundations for processing clinical notes within a multimodal framework.

Norgeot et al. (2019) demonstrated the value of temporal modeling in predicting disease progression, focusing on rheumatoid arthritis. Their approach highlighted that capturing temporal patterns significantly improves predictive performance, but did not explore multimodal integration.

Shickel et al. (2018) provided a comprehensive review of deep learning approaches for electronic health records, identifying key challenges and opportunities. Their analysis emphasizes the need for improved methods for handling missing data and temporal dependencies across modalities.

Xiao et al. (2022) proposed a transformer-based multimodal approach for integrating clinical time series and notes but did not address the integration of imaging data or explore the interpretability aspects crucial for clinical adoption.

**Data Sources & Statistics**

**MIMIC-IV Dataset:**

* Format: Structured (CSV) and unstructured (text) EHR data
* Size: ~40,000 patients, ~50,000 hospital admissions
* Modalities: Clinical notes, laboratory measurements, vital signs, medications, diagnoses
* Relevant statistics: Will focus on subsets of patients with chronic conditions (diabetes, heart failure, chronic kidney disease), with approximately 5,000-8,000 patients per condition category
* Train/validation/test split: 70%/15%/15%

**UK Biobank (if accessible):**

* Format: Structured clinical data, imaging data, genetic data
* Size: ~500,000 participants
* Modalities: Clinical measurements, MRI scans, genetic information
* Relevant statistics: Will focus on participants with longitudinal data and evidence of chronic disease development
* Train/validation/test split: 70%/15%/15%

**PhysioNet-MIMIC-CXR Dataset:**

* Format: DICOM images and associated metadata
* Size: ~377,000 chest X-rays from ~65,000 patients
* Modalities: Chest X-ray images, associated clinical notes
* Relevant statistics: Will identify patients with both imaging data and corresponding clinical notes in MIMIC-IV
* Train/validation/test split: 70%/15%/15%

These datasets provide complementary information that will enable thorough evaluation of multimodal learning approaches. The MIMIC-IV dataset offers rich longitudinal clinical data, while the imaging components provide crucial visual information about disease states. The temporal alignment between these datasets will be a key preprocessing step.

**Methods**

**Traditional Machine Learning Methods:**

1. **Random Forests:** Random Forests will be implemented as a baseline model due to their robustness to overfitting and ability to handle mixed data types. Features will be extracted from each modality separately (TF-IDF vectors from clinical notes, statistical features from lab results, and pretrained CNN features from images) and concatenated for input to the model. Hyperparameter tuning will focus on optimizing tree depth, number of estimators, and feature selection methods.
2. **Gradient Boosting Machines (GBM):** GBM will be used as our second traditional ML approach, specifically XGBoost or LightGBM implementations. These models excel at handling complex relationships in heterogeneous data and have shown strong performance in healthcare applications. Similar to Random Forests, we will use feature concatenation for multimodal integration, with careful attention to feature importance analysis for interpretability.

**Deep Learning Methods:**

1. **Multimodal Transformer Architecture:** We will develop a transformer-based architecture that can effectively integrate information across modalities while maintaining temporal awareness. The model will consist of:
   * ClinicalBERT encoder for processing clinical notes
   * Temporal convolutional network for processing time series lab data
   * ResNet-based feature extractor for imaging data
   * Cross-modal attention mechanisms for information fusion
   * A temporal transformer encoder to model disease progression over time

This architecture will address the key challenges of temporal alignment and missing data through specialized attention mechanisms and carefully designed training procedures.

**Additional Methodological Considerations:**

* Implementation of missing data imputation techniques specific to each modality
* Development of interpretability methods to provide clinical insights
* Approaches for handling class imbalance in disease progression labels
* Temporal alignment strategies for synchronizing data across modalities

**Evaluation Methods**

**Performance Metrics:**

* Area Under the ROC Curve (AUC-ROC): Primary metric for evaluating overall discrimination ability, especially important for handling class imbalance in disease progression
* Area Under the Precision-Recall Curve (AUC-PR): Provides complementary information to AUC-ROC, particularly valuable for imbalanced datasets
* Sensitivity and Specificity: Important for clinical relevance, with particular emphasis on early detection sensitivity
* Temporal metrics: Evaluation of prediction lead time (how far in advance the model can predict progression events)

**Ablation Studies:** We will conduct comprehensive ablation studies to assess the contribution of each modality and component of our multimodal framework. This will involve training models with various combinations of modalities and architectural components, allowing us to quantify the added value of multimodal integration and identify the most informative data sources for different disease types.

**Interpretability Analysis:** Model interpretability will be evaluated through both quantitative measures (e.g., feature importance scores, attention weights) and qualitative assessment by clinical experts. We will develop visualization techniques to illustrate how different modalities contribute to specific predictions, enhancing trust and facilitating clinical validation.

**Baseline Comparisons:** Performance will be compared against both unimodal approaches and existing clinical risk scores for the targeted chronic diseases. This will provide context for assessing the practical value of our multimodal approach compared to current clinical practice.

**Bibliography**

1. Rajkomar, A., Oren, E., Chen, K., et al. (2018). Scalable and accurate deep learning with electronic health records. npj Digital Medicine, 1(1), 1-10.
2. Chen, M., Pellegrini, E., Schaffter, T., et al. (2021). Deep learning for preterm birth prediction: Systematic review, current trends, and future directions. Artificial Intelligence in Medicine, 115, 102060.
3. Huang, Y., Xu, J., Zhou, Y., et al. (2020). Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network. Frontiers in Neuroscience, 14, 259.
4. Alsentzer, E., Murphy, J.R., Boag, W., et al. (2019). Publicly available clinical BERT embeddings. arXiv preprint arXiv:1904.03323.
5. Norgeot, B., Glicksberg, B.S., Trupin, L., et al. (2019). Assessment of a deep learning model based on electronic health record data to forecast clinical outcomes in patients with rheumatoid arthritis. JAMA Network Open, 2(3), e190606.
6. Shickel, B., Tighe, P.J., Bihorac, A., et al. (2018). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE Journal of Biomedical and Health Informatics, 22(5), 1589-1604.
7. Xiao, C., Choi, E., & Sun, J. (2022). Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. Journal of the American Medical Informatics Association, 29(2), 371-384.

**Cross-lingual Music Genre Classification from Lyrics and Audio Features**

**Motivation and Introduction**

Music genre classification is a fundamental problem in music information retrieval (MIR) that has traditionally focused on either audio features or lyrical content, rarely integrating both modalities effectively. Most existing solutions primarily target English-language music, creating significant gaps in classification accuracy for non-English music. This limitation becomes increasingly problematic in our globalized music consumption environment, where streaming platforms expose listeners to diverse cultural content.

This research addresses the critical need for cross-lingual music genre classification systems that can effectively categorize music regardless of language or cultural origin. While recent advances in multilingual transformers have shown promise in natural language processing tasks, their application to music genre classification—particularly when combined with audio features—remains underexplored. Current approaches often struggle with cultural nuances in genre definitions, as what constitutes "rock" or "hip-hop" may vary significantly across cultures and languages.

This research proposes a novel multimodal approach that leverages both acoustic properties and lyrical content across languages. By combining state-of-the-art multilingual language models with acoustic feature extraction, this study aims to develop a more robust classification system that performs consistently across linguistic and cultural boundaries. Such a system would not only advance MIR technology but also support more inclusive music recommendation systems, potentially discovering cross-cultural musical connections previously obscured by language barriers.

**Research Question**

How can cross-lingual NLP models combined with audio feature extraction improve music genre classification accuracy across different languages and cultural contexts compared to unimodal and monolingual approaches?

**Initial Review**

Recent research in music genre classification has evolved along two primary paths: audio-based and lyrics-based approaches, with limited exploration of their integration. Dhanaraj and Logan (2005) pioneered early work combining audio and lyrical features using simple SVM classifiers, but their study was limited to English content. More recently, Oramas et al. (2018) demonstrated that deep learning models can effectively combine audio and text for music genre classification, showing a 4% improvement over unimodal approaches, though still primarily focused on English lyrics.

In the cross-lingual domain, Lin et al. (2021) applied multilingual BERT to song lyrics for emotion classification across five languages, revealing that pre-trained multilingual models can transfer sentiment knowledge across languages, though their work did not address genre classification specifically. Parallel to this, Castellón et al. (2021) identified cultural biases in genre definitions across languages, showing how the same acoustic features might signal different genres in different cultural contexts.

For audio processing, Pons et al. (2018) introduced musically-motivated CNNs for spectrogram analysis that outperformed traditional MFCCs for genre classification. Building on this, Won et al. (2020) developed attention mechanisms for music tagging that dynamically focus on discriminative temporal segments, showing particular efficacy for genres with distinctive rhythmic patterns.

Despite these advances, there remains a significant gap in research addressing truly multimodal, cross-lingual genre classification that accounts for cultural variations in genre definitions. This project aims to bridge this gap by leveraging recent advances in both domains while explicitly addressing cultural contextualization of genres.

**Data Sources & Statistics**

1. **Million Song Dataset (MSD)**: A collection of audio features and metadata for one million contemporary popular music tracks.
   * Format: HDF5 files containing pre-computed audio features
   * Size: 280 GB of feature data
   * Statistics: Contains tracks spanning 44,745 unique artists with genre labels derived from AllMusic tags
2. **GTZAN Dataset**: A standard benchmark dataset for music genre classification.
   * Format: 30-second audio clips in .wav format
   * Size: 1000 audio tracks (100 per genre)
   * Statistics: 10 genres with equal distribution (classical, blues, country, disco, hip-hop, jazz, metal, pop, reggae, rock)
3. **Genius Lyrics API**: For extracting multilingual lyrics corresponding to tracks in the MSD and GTZAN datasets.
   * Format: JSON responses containing song lyrics
   * Coverage: Estimated to provide lyrics for approximately 60-70% of tracks in the MSD
   * Languages: Primary focus on English, Spanish, French, German, Portuguese, and Japanese lyrics
4. **LyricsTranslate Dataset**: A supplementary dataset of human-translated lyrics.
   * Format: Plain text
   * Size: Approximately 50,000 song translations across multiple language pairs
   * Statistics: Will be used for validation and testing of cross-lingual models

The datasets will be pre-processed into standardized formats with consistent genre taxonomy mapping across languages. For training and validation, the data will be split into 70% training, 15% validation, and 15% test sets, stratified by genre and language to ensure balanced representation.

**Methods**

**Traditional Machine Learning Methods**

1. **Support Vector Machines (SVM) with TF-IDF**:
   * Will be applied to lyrical content with TF-IDF vectorization across multiple languages
   * Language-specific stopword removal and stemming will be implemented
   * Kernel selection (linear, polynomial, RBF) will be optimized through cross-validation
   * This approach establishes a strong baseline for text-only classification
2. **Random Forests on Acoustic Features**:
   * Will utilize handcrafted acoustic features including MFCCs, spectral contrast, chromagrams, tempograms, and onset strength
   * Feature importance analysis will identify universal vs. culturally-specific acoustic markers of genres
   * Hyperparameter optimization will include tree depth, number of estimators, and feature subset selection
   * This approach provides a strong baseline for audio-only classification

**Deep Learning Methods**

1. **Multimodal Transformer Architecture**:
   * Will integrate XLM-RoBERTa for multilingual lyrics processing with CNN-based audio feature extraction
   * The architecture will include:
     + A pre-trained XLM-RoBERTa encoder for lyrics in multiple languages
     + A parallel CNN pathway for spectrogram processing inspired by Pons et al. (2018)
     + Cross-attention mechanisms to create unified representations
     + A classification head with genre taxonomy mapping for different cultural contexts
   * This approach aims to leverage the strengths of state-of-the-art NLP and audio processing techniques while facilitating cross-modal and cross-lingual learning

**Evaluation Methods**

**Classification Performance Metrics**

Performance will be evaluated using precision, recall, F1-score, and accuracy, with macro-averaging to account for potential genre imbalance. Confusion matrices will be analyzed to identify cross-cultural genre conflations. Additional evaluation will include language-specific performance comparison to identify potential biases or performance gaps across languages.

**Cross-lingual Transfer Assessment**

The model's ability to transfer knowledge across languages will be assessed through zero-shot and few-shot evaluation on languages not seen during training. Performance degradation rates will be measured when moving from monolingual to cross-lingual settings, with ablation studies to quantify the contribution of audio features in bridging language gaps.

**Multimodal Integration Analysis**

The effectiveness of multimodal integration will be evaluated through comparison with unimodal baselines and through attention visualization techniques that reveal which modality (audio or lyrics) contributes more to classification decisions for different genres and languages. This will provide insights into genre characteristics that transcend or are bound by linguistic and cultural contexts.

**Bibliography**

1. Castellón, C. E., Figueras, V. G., & Serra, X. (2021). "Cultural biases in music genre perception: A cross-cultural study." Transactions of the International Society for Music Information Retrieval, 4(1), 21-32.
2. Dhanaraj, R., & Logan, B. (2005). "Automatic prediction of hit songs." In Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), 488-491.
3. Lin, Y., Yang, H., & Lee, C. (2021). "Cross-lingual transfer learning for music emotion recognition using multilingual BERT." In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 586-590.
4. Oramas, S., Nieto, O., Barbieri, F., & Serra, X. (2018). "Multi-label music genre classification from audio, text, and images using deep features." In Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), 23-30.
5. Pons, J., Nieto, O., Prockup, M., Schmidt, E. M., Ehmann, A. F., & Serra, X. (2018). "End-to-end learning for music audio tagging at scale." In Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), 637-644.
6. Won, M., Choi, K., & Serra, X. (2020). "Semi-supervised music tagging transformer." In Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), 535-541.
7. Hernandez-Olivan, C., & Beltran, J. R. (2022). "Music genre classification: A multi-modal deep learning approach." IEEE Access, 10, 39263-39276.