An Empirical Study Using Machine Learning to Analyze the Relationship Between Musical Audio Features and Psychological Stress

Harini Anand[a], Shalini Kammalam Srinivasan[b], Hasika Venkata Boggarapu, Arti Arya[c] and Richa Sharma[d]

Department of Computer Science and Engineering, PES University, Bangalore, India

{*harini.anand, shalinikammalam, hasikabv*}*@gmail.com, artiarya@pes.edu, richasharma@pes.edu*

Keywords: Music, Psychological Stress, XAI, Classification, SHAP, Integrated Gradients, Feedforward Neural Network, Audio Techniques, Danceability, Energy, Loudness, Mode, Speechiness, Acousticness, Instrumentalness, Liveness, Tempo.

Abstract: Music plays a vital role in regulating emotions and mental well-being, influencing brain function and stress levels. This study leverages Explainable AI (XAI) techniques, specifically SHapley Additive exPlanations (SHAP) and Integrated Gradients, to analyze the impact of scientifically backed audio features—such as Danceability, Energy, Acousticness, etc on stress classification. Using a Feedforward Neural Network, we achieved a 0.96 accuracy in categorizing music preferences into "Stressed," "Not-stressed," and "Borderline" states. The classifier operates effectively across languages and genres, enhancing its versatility for detecting Psychological Stress by providing interpretable insights.

1 INTRODUCTION

Music significantly impacts brain function and structure, influencing areas related to emotion, motivation, and anticipation (Vuust et al., 2022). It can modulate heart rate and breathing, thereby affecting our stress levels and overall mental state. Beyond its role as a source of entertainment, music is an integral part of our lives, playing a crucial role in mental health and well-being by affecting emotions, moods, and other such psychological states. Research has shown that music preferences and listening strategies are linked with the psychological welfare of listeners, as well as stress and internalized symptomatology. However, studies examining the time-varying nature of music consumption in terms of acoustic content and its association with users' well-being, remain scarce. Music's power to shift and regulate mood makes it a useful tool for managing emotions. For instance, during periods of stress, individuals often rely on music to impact their moods and alter affective states (Adiasto, 2022). Focusing on predictive accuracy over model interpretability has a possibility of resulting in a gap of transparency in the decision-making process, which is important in crucial use cases such as healthcare. This underscores the extending need for explainable AI (XAI) approaches in psychological prediction and diagnosis. The motivation of the research is to explore the time-varying nature of music consumption and its acoustic content in relation to users' well-being, also understanding the potential of music as a tool for emotion regulation and its implications for mental health. Additionally, the study addresses the challenge of bridging the gap between predictive accuracy and interpretability in AI models used for psychiatric diagnosis and prediction. Initial efforts relied on a genre-based approach but lacked robustness for non-English songs. By focusing on these universal audio features, the classifier is designed to operate effectively across all languages and genres of music, ensuring a broad, culturally inclusive application.

This research contributes to the field by conducting a empirical study to analyze the association between music consumption patterns and psychological well-being, developing an explainable AI model that balances predictive accuracy with interpretability for detecting stress levels, and identifying specific acoustic features of music that are linked to changes in mood and mental state.

[a] https://orcid.org/0009-0002-1477-3197
[b] https://orcid.org/0009-0000-0231-356X
[c] https://orcid.org/0000-0002-4470-0311
[d] https://orcid.org/0000-0002-4539-7051

Anand, H., Srinivasan, S. K., Boggarapu, H. V., Arya, A. and Sharma, R.
An Empirical Study Using Machine Learning to Analyze the Relationship Between Musical Audio Features and Psychological Stress.
DOI: 10.5220/0013249900003890

In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) - Volume 3*, pages 1031-1039
ISBN: 978-989-758-737-5; ISSN: 2184-433X

2 RELATED WORK

2.1 Literature Survey

Recent research underscores the significant role of music in emotional regulation and mental well-being. According to (Stewart et al., 2019), individuals with tendencies toward depression often use music for mood regulation, consciously selecting compositions that help manage their emotions. Music has been shown to influence internalized symptomatology and depression, indicating a strong connection between listening strategies and psychological well-being. This connection is especially pronounced among young people, for whom music serves as a critical emotional outlet (McFerran, 2014). These findings highlight the potential of music as a non-pharmacological intervention for mental health issues, particularly in the context of emerging adulthood, which is characterized by significant life transitions and stressors (Anderson et al., 2003).

Explainable AI (XAI) techniques are pivotal in providing insights into how machine learning models make predictions, especially in complex domains such as healthcare and psychology. These techniques help decrease the gap between human understanding and model predictions, enhancing trust and interpretability. However, much of this research prioritizes predictive accuracy over model interpretability, which can be problematic in healthcare applications where transparency is crucial (Lin, 2011).

Neuroscientific and clinical studies provide compelling evidence for the therapeutic potential of music. For example, (Juslin and Sloboda, 2010) discusses how music therapy can be an effective treatment for different mental health conditions, including depression, autism, schizophrenia, and dementia. The therapeutic advantages of music are due to its ability to affect tough neuro-biological processes in the brain, thereby modulating emotions and alleviating anxiety. Furthermore, the use of Music Information Retrieval (MIR) algorithms in analyzing audio features like tempo and rhythm offers a promising approach to understanding how different types of music can impact stress levels and overall mental health. This integration of technology and neuroscience paves the way for innovative interventions that harness the power of music to improve psychological well-being (Saarikallio, 2007).

The study by (Gujar, 2023) investigates the correlation between music and mood using machine learning techniques and reveals that certain musical elements like key, tempo, and mode are linked to specific moods, providing valuable insights for enhancing mental health through music. The study by (Ahuja, 2019) analyses the mental stress among college students using machine learning algorithms to evaluate the impact of exam pressure and internet usage on their well-being.

The study by (Erbay Dalli, 2023) elaborates on how multiple-session music interventions can be employed as a nursing strategy to manage anxiety levels in ICU patients. Hearing music on a regular basis, prevents CUMS-induced oxidative stress in the hippocampus, prefrontal cortex, and serum of mice. This paper by (Gu, 2023) indicates that in mouse experiments, hearing music reduces stress-induced anxiety and depression-like behaviors. Music has the ability to regain preventing oxidative stress, neurotrophic factor deficits, hypothalamus-pituitary-adrenal axis homeostasis, and inflammation.

After an elaborate literature survey, as highlighted in the previous section, the following research gaps underscore the pressing need for further exploration into the intricate correlation between audio features in music and psychological stress:

- Existing studies focus primarily on genre based stress detection and lack robustness for regional and foreign songs.
- Insufficient exploration of how specific audio features, such as tempo, liveness, and danceability, influence stress psychopathology.
- A lack of research applying XAI (Explainable AI) techniques in healthcare diagnosis and prediction, which are needed to ensure transparency and interpretability in AI models.

2.2 Our Contributions

Since the research gaps point to a substantial opportunity to explore the potential for music, through these individual audio features, we aim to implement certain key techniques, mentioned below that shall serve as a preventive and therapeutic tool for reducing psychological stress.

- Building a high performing classifier using Machine Learning and train it specifically on Music Audio Features to detect Psychological Stress.
- Incorporate Explainable AI Techniques to demonstrate interpretability.
- Validate the classifier on unseen data using different types of linguistic music datasets.

3 PROPOSED APPROACH

The proposed approach is as depicted in Figure 1.

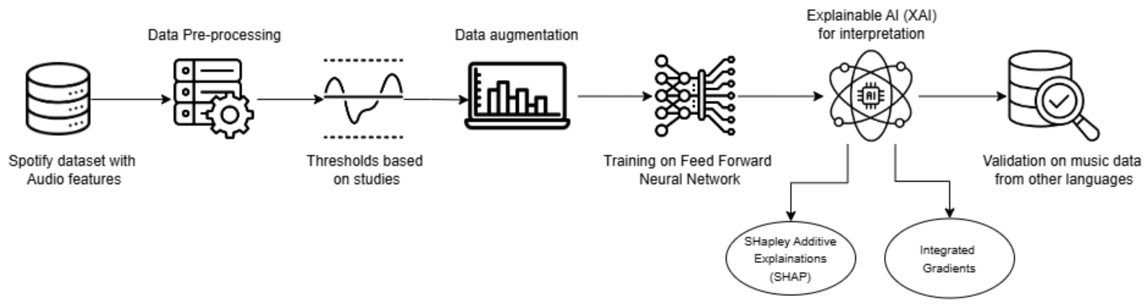

Figure 1: Proposed Approach to Establish a Relation between Music and Psychological Stress.

3.1 Dataset Description

The dataset is a comprehensive collection of 15,150 classic hits from 3,083 artists, spanning a century of music history from 1923 to 2023 sourced from Kaggle. This diverse dataset is divided into 19 distinct genres, showcasing the evolution of popular music across different eras and styles. Each track in the dataset is of songs from Spotify with audio features such as Danceability, Energy, Loudness, Mode, Speechiness, Acousticness, Instrumentalness, Liveness, and Tempo, offering detailed insights into the acoustic properties, rhythm, tempo, and other musical characteristics of each track in the dataset.

3.2 Threshold Validation

To utilize the audio features in the dataset for our intended objective of exploring the influence of music on psychological stress, several clinical and research based studies were used, where researchers exploratorily inferenced the connection between music features and users' preferred recovery-related feelings while hearing and after hearing to self-selected music.

In one of their studies (Adiasto et al., 2023a), 470 participants took a survey where the users indicated the type of music she or he would pick to de-stress from a theoretically stressed situation. Using Data analysis techniques such as split-sample procedures, a k-medoid cluster analysis was held to identify audio feature commonalities between songs that were self-selected by the users. In addition to this, several regression analyses were also done to cross check and deduce the connection of musical audio features and preferred recovery psychological states and emotions.

Analyses in (Adiasto et al., 2023b) revealed the role played by positive emotions in the stress reduction process, it is safe to conclude that music's audio features have recovery potential under the condition that it draws out a favourable emotional response. Based on this base theory, the expansive domain of music emotion recognition (MER) hints which songs are the most impactful for stress reduction and recovery.

Studies on music emotion recognition (Duman et al., 2022) have explored various combinations of musical audio features, including tempo (the speed of a song), pitch (the frequency of a particular note or sound), and timbre (the overall quality or color of a song). These features are analyzed in relation to the valence (i.e., the positivity of an emotion, with higher valence values indicating more positive emotions) and arousal components of self-reported musical emotions. Compared to the baseline, participants' preferred music choices to reduce stress, indicated significantly higher levels of energy, danceability, and acousticness.

In addition, tempo has been found to positively correlate with both emotional valence and arousal, demonstrating that songs with faster tempos tend to be associated with emotions that are more positive and have higher levels of arousal.

Table 1 lists the scientifically validated thresholds using various audio features of Spotify songs, which we make use of in our proposed approach.

3.3 Dataset Pre-Processing & Augmentation

Valence as a feature was dropped due to low variability across the dataset (i.e., most of its values were similar), as it would contribute little information to the model.

To handle class imbalance, rule-based synthetic data generation (He et al., 2008) is used, where we utilize a hybrid technique approach combining boundary-based sampling (Liu et al., 2007) with Gaussian noise injection (Zhang et al., 2018). This is the most suitable approach when the dataset has predefined class boundaries. The dataset distribution after data augmentation is as seen in Table 2.

Table 1: Threshold Distribution of Stress Level Class Labels Based on Music Audio Features.

Feature	Not-Stressed (0)	Stressed (1)	Borderline (2)
Danceability	0.5 – 0.998	0 – 0.38	0.39 – 0.49
Energy	0.6 – 0.99	0.3 – 0.59	0 – 0.299
Mode	1	0	0
Speechiness	0.01 – 0.07	0.31 – 1	0.071 – 0.3
Acousticness	0.35 – 1	0 – 0.3	0.31 – 0.34
Instrumentalness	0 – 0.18	0.21 – 1	0.19 – 0.2
Liveness	0 – 0.1	0.2 – 1	0.11 – 0.2
Tempo (BPM)	60 – 130	181 – 220	130 – 180

Table 2: Class Distribution in Full, Training, and Testing Datasets.

Dataset	Stressed	Borderline	Not Stressed
Full Dataset	5602	5935	5848
Training Set	4482	4748	4678
Testing Set	1120	1187	1170

3.4 Model Architecture & Training

3.4.1 Why Feed-Forward Neural Network (FFNN)

A Feed-Forward Neural Network (FFNN) was selected for this classification problem due to its efficiency in modeling structured, non-sequential data and capturing complex patterns. FFNNs are particularly effective in handling data like audio features, where each feature represents a specific aspect of the input without any inherent temporal or sequential dependencies.

The FFNN architecture also allows for flexibility with activation functions and regularization techniques, making it ideal for addressing potential overfitting, and ensuring the model generalizes well to unseen data. By using FFNN, we benefit from its ability to map high-dimensional input data to meaningful output probabilities for each class. Other State of the Art Models did not outperform the chosen architecture of choice for our use case.

3.4.2 Model Architecture

The model was designed using three dense layers to learn the complex relationships among the input features. Each layer incorporates Dropout and Batch Normalization techniques to eliminate overfitting and improve generalization. Dropout with a rate of 0.7 was applied after each dense layer to randomly set a section of the input units to 0 during training, effectively preventing the model from relying too heavily on any one specific feature. Batch normalization was employed to normalize the activations in each layer, stabilizing the learning process and improving convergence speed. Softplus activation functions were used in the intermediate layers due to their smooth gradient properties, which facilitate stable training. Finally, the output layer employs the softmax activation function to produce probabilities for each of the three classes, making it suitable for a multi-class classification problem.

3.4.3 Training Configuration

Our model was compiled using the Adam optimizer with a learning rate of 0.001, a commonly used optimizer known for its efficient handling of sparse gradients and adaptive learning rates. The categorical cross-entropy loss function was used, as it is the standard for multi-class classification problems.

To further combat overfitting and ensure model generalization, we applied early stopping, which monitors the validation loss and halts training if it does not improve for five consecutive epochs. This prevents unnecessary training and helps preserve the best weights. Additionally, the ReduceLROnPlateau callback was implemented to reduce the learning rate by a factor of 0.5 if the validation loss stagnates for two epochs, promoting better convergence. These training techniques, in combination with the model architecture, lead to improved performance on the test set while minimizing overfitting.

4 RESULTS & DISCUSSIONS

4.1 Model Performance

The feedforward neural network achieved a training accuracy of **96%** with a corresponding training loss of **0.2850** and demonstrated strong performance during training and validation phases. Figure 2 depicts Training and Validation loss curves demonstrating model performance, and Figure 3 suggests Training and Validation accuracy curves demonstrating model performance, thus proving it not only fits the training data well but also generalizes excellently to unseen data.

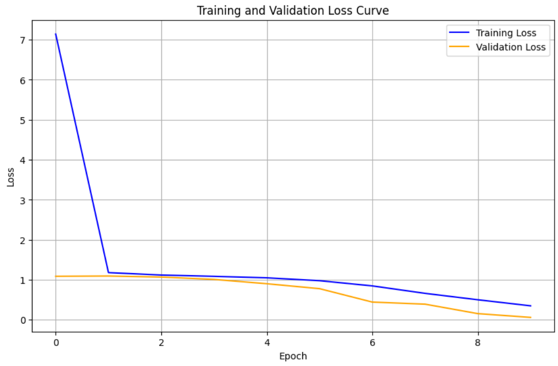

Figure 2: Training and validation loss curves demonstrating model performance.

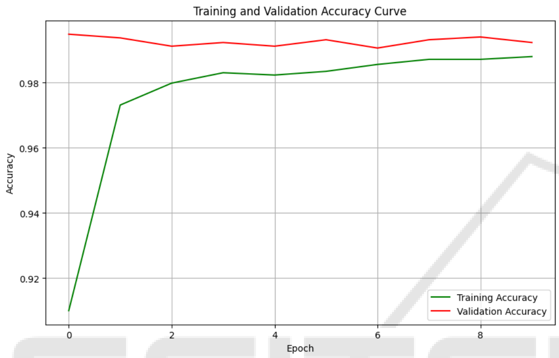

Figure 3: Training and validation accuracy curves demonstrating model performance.

4.2 Explainable AI (XAI)

Explainable AI (XAI) is very essential for interpreting the predictions given by a feedforward neural networks (FFNNs) and addressing their "black box" nature. This helps the stakeholders to better understand, trust and validate model predictions especially in domains like healthcare.

SHapley Additive exPlanations (SHAP) and Integrated Gradients are the two XAI techniques used in our study to analyze the results of FFNN. These methods help to interpret the influence of each feature on the model's output thereby providing insights into the internal workings of the model.

SHapley Additive exPlanations (SHAP). SHAP (Lundberg, 2017) assigns an importance value called shapley value to each feature for individual predictions. This importance value represents both the magnitude and direction of a feature's impact thus helping us to interpret feature influence at a more granular and instance-specific level. Taking all possible feature combinations into consideration, SHAP calculates feature importance for each input feature.

The SHapley value for feature i, $\phi_i(f)$, is defined as in Equation (1):

$$\phi_i(f) = \sum_{F \subseteq N \setminus \{i\}} \frac{|F|! \cdot (|N| - |F| - 1)!}{|N|!} [f(F \cup \{i\}) - f(F)] \tag{1}$$

Here:
- N is the set of all features.
- F is a coalition of features excluding feature i.
- $|F|$ is the number of features in coalition F.
- $f(F)$ is the model's prediction for the instance with only features in set F.
- $f(F \cup \{i\})$ is the model's prediction for the instance with features in set $F \cup \{i\}$.
- $\phi_i(f)$ is the SHapley value for feature i, capturing its marginal contribution to the prediction in all possible coalitions.

By computing the contribution of feature i across all coalitions, SHAP provides a complete view of how each feature influences the model's output. The interpretability it offers helps stakeholders pinpoint which features are driving predictions in individual cases bringing in transparency in FFNN predictions.

Integrated Gradients. Unlike SHAP which focuses on instance-level feature contributions, Integrated Gradients (Sundararajan, 2017) provide a global perspective on feature relevance by measuring the cumulative effect of gradients along a path from a baseline w' to the input w. The baseline is typically chosen as a neutral input such as a black image for vision models or a zero vector for text models. By integrating gradients along this path it capture the sensitivity of the model's output with respect to each feature helping explain the model's behavior over the entire input space.

The Integrated Gradient for feature i, $\mathrm{IG}_i(x)$, is defined as in Equation (2):

$$\mathrm{IG}_i(w) = (w_i - w'_i) \times \int_{\alpha=0}^{1} \frac{\partial G(w' + \alpha \times (w - w'))}{\partial w_i} \, d\alpha \tag{2}$$

Here:
- w is the input for which we calculate attributions.
- w' (or w_0) is the baseline input, serving as a reference point for comparison.
- w_i is the i-th feature value of the input x.
- w'_i is the i-th feature value of the baseline input w'.
- α is a scaling factor ranging from 0 to 1,
- G is the function representing the model (e.g., a neural network), mapping inputs to outputs.

Integrated Gradients help explain how much each feature contributes to a model's output by capturing the average gradient along the path from baseline to input. This provides insights into the overall influence of each feature across the model, rather than individual predictions.

Together, SHAP and Integrated Gradients provide complementary insights into Feed Forward Neural Network predictions. While SHAP explains feature impact on a case-by-case basis, Integrated Gradients capture a broader view of feature relevance across different instances.Using both methods can lead to a deeper understanding of the model behavior, particularly in sensitive applications like healthcare, where interpretability is crucial for ethical considerations, transparency, and trust. This XAI interpretability framework provides stakeholders with granular details and global trends, fostering a more comprehensive understanding of model decisions and enhancing confidence in AI-assisted healthcare predictions.

4.2.1 Not Stressed Class

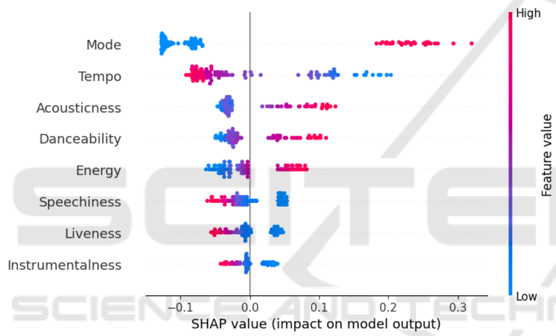

Figure 4: XAI SHAP Plot for Not Stressed Class.

For ’Not Stressed’ class as seen in Figure 4, Tempo emerged as one of the most influential features exhibiting a clear inverse relationship with the ”not stressed” classification. Higher tempo values, represented by pink dots in the SHAP plot, consistently showed negative SHAP values, indicating that faster-paced music significantly decreases the likelihood of a ”not stressed” classification. On the other hand, lower tempo values (blue dots) displayed positive SHAP values, suggesting that slower-paced music contributes positively to the ”not stressed” prediction.

Danceability showed a notable pattern where higher values (pink dots) on the positive SHAP value side indicated that more danceable tracks are more likely to be classified as ”not stressed”. This suggests that songs with stronger rhythmic elements and regular patterns may contribute to a less stressful listening experience. Acousticness showed a more complex distribution with moderate values (purple dots) having a positive impact on the ”not stressed” classification. This indicates that songs with balanced acoustic elements are more likely to be classified as not-stressful compared to those at either extreme of the acousticness spectrum.

Energy values clustered around the center of the SHAP value range, with a slight positive skew for higher energy levels, suggesting a positive association with the ”not stressed” classification. Features such as Speechiness and Instrumentalness exhibited minimal impact on the model’s predictions, as evidenced by their tight clustering around zero SHAP values. The analysis also revealed that Liveness had a slight negative impact when at higher values, indicating that highly live recordings are less frequently associated with the ”not stressed” classification. This could potentially be attributed to the more unpredictable and dynamic nature of live performances.

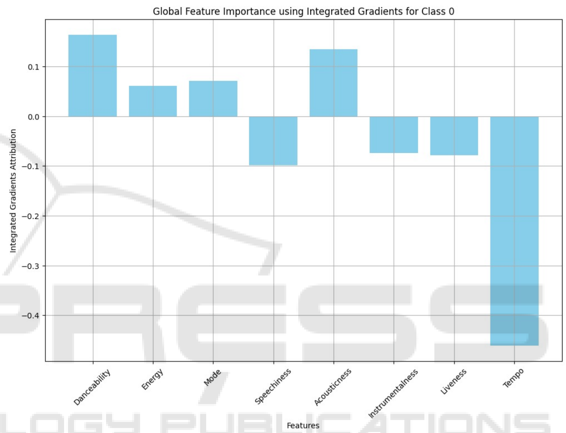

Figure 5: Integrated Gradients for Not Stressed class.

The Integrated Gradients analysis for not stressed class as in Figure 5 strongly validates our SHAP value findings. Notably, Danceability shows the highest positive contribution, aligning with our SHAP interpretation where higher danceability values positively influenced ”not stressed” predictions. Similarly, Tempo displays the strongest negative contribution (-0.45), which perfectly corroborates our SHAP analysis where higher tempo values pushed predictions away from the ”not stressed” class. The moderate positive contributions of Energy and Mode, along with the negative impacts of Instrumentalness and Liveness, also mirror the patterns observed in the SHAP visualization, reinforcing the reliability of our feature importance interpretations.

These findings provide valuable insights into the musical characteristics that contribute to a song being classified as ”not stressed,” with tempo and danceability emerging as particularly significant predictors in the model’s decision-making process.

4.2.2 Stressed Class

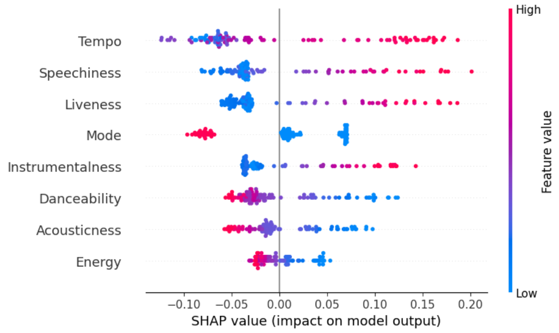

Figure 6: XAI SHAP Plot for Stressed Class.

For 'Stressed' class as observed in Figure 6, Tempo demonstrated the most substantial influence, showing a strong positive correlation with stress classification. The visualization indicates that higher tempo values (pink dots) are predominantly positioned on the positive SHAP value side, suggesting that faster-paced music significantly increases the likelihood of a song being classified as "stressed". This aligns with the intuitive understanding that rapid tempos may induce heightened arousal states. Speechiness emerged as the second most influential feature, with lower values (blue dots) clustered on the negative side and higher values (pink and purple dots) extending toward positive SHAP values. This suggests that songs with greater vocal presence and spoken word content are more likely to be classified as stressed. Liveness showed a notable distribution pattern where higher values (pink dots) extended into positive SHAP values, indicating that songs with stronger live performance characteristics tend to be classified as more stressful. This could be attributed to the increased ambient noise and audience participation typical in live recordings.

Mode exhibited an interesting pattern with higher values (pink dots) concentrated on the negative side, suggesting that major mode songs are less likely to be classified as stressed, while lower values (blue dots) on the positive side indicate minor mode songs contribute to stress classification. Instrumentalness displayed a scattered pattern with moderate to high values suggesting that highly instrumental tracks have some association with stress classification.

Danceability showed an inverse relationship, with higher values (pink dots) concentrated on the negative SHAP value side, indicating that more danceable tracks are less likely to be classified as stressed. Acousticness and Energy demonstrated more modest impacts, with relatively tight clustering around zero, though both showed slight tendencies toward negative SHAP values for higher feature values.

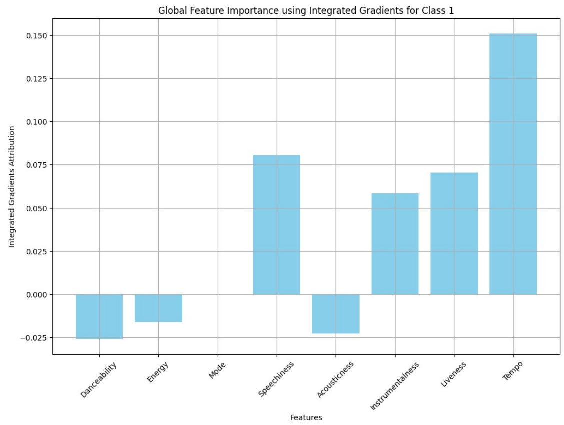

Figure 7: Integrated Gradients for Stressed Class.

The Integrated Gradients as seen in Figure 7 validates our SHAP analysis with Tempo emerging as the strongest predictor (0.15) for stressed music. Speechiness and Liveness show notable positive contributions, aligning with their positive SHAP values. The negative impacts of Danceability and Energy further confirm our SHAP interpretations of their inverse relationship with stress classification. The results suggest that faster, speech-heavy, and live performance elements are more strongly associated with stressed music classification, while danceable and major mode characteristics tend to oppose this classification.

4.2.3 Borderline Class

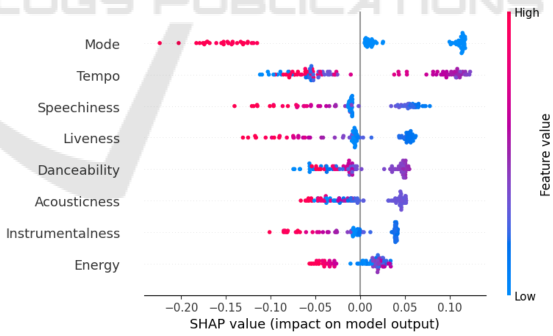

Figure 8: XAI SHAP Plot for Borderline Class.

In the 'Borderline' class as in Figure 8, Mode emerged as one of the most distinctive features, showing a clear bimodal distribution. Lower values (blue dots) appeared significantly on the positive SHAP value side, while higher values (pink dots) were concentrated on the negative side. This suggests that minor mode songs (represented by lower values) have a stronger association with borderline stress classification, while major mode songs tend to push predictions away from this category. Tempo displayed an interesting distribution with both high (pink dots) and moderate (purple dots) values showing positive SHAP values, indicating that mid to higher-tempo songs contribute to borderline stress classification. The pattern suggests that tempo has a more nuanced impact compared to its role in clear-cut stressed or not-stressed classifications.

Speechiness demonstrated a pattern where lower values (blue dots) showed positive SHAP values, suggesting that songs with less vocal content are more likely to be classified as borderline stressed. This contrasts with its impact on definitive stress classifications, indicating a unique characteristic of borderline cases.

Liveness and Danceability showed relatively modest impacts, with slight clustering around zero but extending into both positive and negative SHAP values. This suggests these features play a more subtle role in borderline classification compared to their influence on definitive stress categories. Acousticness exhibited a pattern where moderate values (purple dots) showed slight positive SHAP values, indicating that balanced acoustic characteristics might contribute to borderline classification.

Instrumentalness and Energy displayed relatively concentrated distributions near zero, with Energy showing a slight tendency toward positive SHAP values for moderate (purple) feature levels, suggesting these features have minimal but consistent impacts on borderline classification.

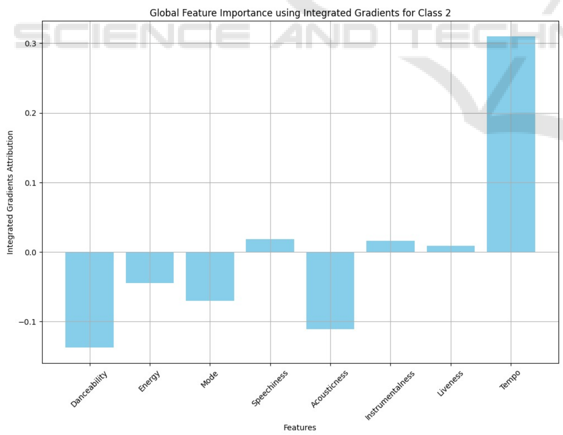

Figure 9: Integrated Gradients for Borderline Class.

The Integrated Gradients for borderline stress class in Figure 9 shows Tempo with the highest positive contribution (0.3), matching our SHAP findings where mid-to-high tempo values indicated borderline stress. Danceability shows the strongest negative impact (-0.13), while Mode and Acousticness display moderate negative contributions. The results suggest that borderline stress music often combines elements typically associated with both stressed and not-stressed categories, creating a distinct musical profile for this intermediate classification.

4.3 Model Validation on External Datasets

4.3.1 Validation Technique and Its Importance

Validating an audio classifier using external datasets from different regional languages is a crucial step in assessing the robustness and generalization capabilities of the model. By testing the model on datasets that it hasn't seen before—especially those that vary in linguistic, acoustic, and cultural characteristics—the evaluation process goes beyond standard train-test splits. This technique helps ensure that the classifier isn't overfitting to the specific features of the training data but instead is capable of accurately categorizing diverse audio inputs. In this context, validating with Tamil and Hindi music datasets allows for a comprehensive evaluation, considering different tonal qualities, rhythms, and cultural nuances in music, ultimately providing confidence in the model's versatility and performance.

4.3.2 Performance on External Music Dataset

When evaluated on the Hindi music dataset, the audio classifier achieved a high test accuracy of **98.63%** with a minimal test loss of **0.0478**. This indicates that the model effectively distinguishes between stressed, not-stressed, and borderline music categories, even when exposed to new, linguistically rich audio content. The high accuracy reflects the model's ability to generalize its learned thresholds for stress classification beyond its training set, capturing the nuanced features of Hindi music that may influence emotional and stress responses. The model demonstrated an even higher level of accuracy on the Tamil music dataset, achieving a test accuracy of **99.43%** and a remarkably low test loss of **0.0148**. This impressive result suggests that the classifier not only adapts well to the phonetic and rhythmic distinctiveness of Tamil music but also maintains consistency in identifying the stress levels associated with different tracks.

5 CONCLUSION AND FUTURE SCOPE

In conclusion, our study demonstrates that among the wide range of music audio features such as Dancability, Tempo, Energy, and Acousticness etc. are particularly effective in reducing psychological stress.

Leveraging these scientifically validated thresholds, we developed a robust audio classifier using a Feed-Forward Neural Network (FFNN) which efficiently categorizes music across all languages and genres into "Stressed," "Not-Stressed," and "Borderline" class labels by analyzing key features, which are linked to positive emotional valence and arousal. This universal audio feature-based approach surpasses genre-specific limitations, offering accurate and culturally inclusive stress classification across diverse languages and music styles. For future work, we aim to explore and incorporate advanced techniques that enhance the interpretability of our model.

ACKNOWLEDGEMENTS

We extend our deepest gratitude to Dr Keya Das, Department of Psychiatry at PES University Institute of Medical Sciences and Research for her invaluable contributions to this study. Her expertise, time, and insightful feedback were crucial in validating our approach and ensuring the success of this research.

REFERENCES

Adiasto, K., B. D. G. J. v. H.-M. L. M.-R. K. . G. S. A. E. (2022). Music listening and stress recovery in healthy individuals: A systematic review with meta-analysis of experimental studies. *PLoS One*, 17(6):e0270031.

Adiasto, K., van Hooff, M. L. M., Beckers, D. G. J., and Geurts, S. A. E. (2023a). The sound of stress recovery: an exploratory study of self-selected music listening after stress. *BMC Psychology*, 11(1):40.

Adiasto, K., van Hooff, M. L. M., Beckers, D. G. J., and Geurts, S. A. E. (2023b). The sound of stress recovery: an exploratory study of self-selected music listening after stress. *BMC Psychology*, 11(1):40.

Ahuja, R. & Banga, A. (2019). Mental stress detection in university students using machine learning algorithms. *Procedia Computer Science*, 152:349–353.

Anderson, C. A., Carnagey, N. L., and Eubanks, J. (2003). Exposure to violent media: The effects of songs with violent lyrics on aggressive thoughts and feelings. *Journal of Personality and Social Psychology*, 84(5):960–971.

Duman, D., Neto, P., Mavrolampados, A., Toiviainen, P., and Luck, G. (2022). Music we move to: Spotify audio features and reasons for listening. *PLoS One*, 17(9):e0275228.

Erbay Dalli, Ö., B. C. . Y. Y. (2023). The effectiveness of music interventions on stress response in intensive care patients: A systematic review and meta-analysis. *Journal of Clinical Nursing*, 32(11-12):2827–2845.

Gu, Y.-Y., Z. L.-L. T. X.-Y. Y. F.-X. L. L.-M. G. Y.-S. L. H. L. T.-Z. B. G.-Q. . F. Z.-Q. (2023). A spatial transcriptome reference map of macaque brain states. *Translational Psychiatry*, 13:220.

Gujar, S. S. & Reha, A. Y. (2023). Exploring relationship between music and mood through machine learning technique. In *Proceedings of the 5th International Conference on Information Management & Machine Intelligence*, pages 1–6.

He, H., Bai, Y., and Garcia, E. A. (2008). Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *Proceedings of the 2008 International Joint Conference on Artificial Intelligence*, pages 1322–1327.

Juslin, P. N. and Sloboda, J. A. (2010). *Handbook of Music and Emotion: Theory, Research, Applications*. Oxford University Press, Oxford, UK.

Lin, S. T., Y. P. L. C. Y.-S. Y. Y.-Y. Y. C.-H. M. F.-. C. C. C. (2011). Mental health implications of music: Insight from neuroscientific and clinical studies. *Harvard Review of Psychiatry*, 19(1):34–46.

Liu, X., Wu, J., and Zhou, Z. (2007). Generative oversampling for mining imbalanced datasets. *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 9–18.

Lundberg, S. M. & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, volume 30.

McFerran, K. S., G. S. H. C. . M. K. (2014). Examining the relationship between self-reported mood management and music preferences of australian teenagers. *Nordic Journal of Music Therapy*, 24(3):187–203.

Saarikallio, S. & Erkkilä, J. (2007). The role of music in adolescents' mood regulation. *Psychology of Music*, 35(1):88–109.

Stewart, J., Garrido, S., Hense, C., and McFerran, K. (2019). Music use for mood regulation: Self-awareness and conscious listening choices in young people with tendencies to depression. *Frontiers in Psychology*, 10:1199.

Sundararajan, M., T. A. . Y. Q. (2017). Axiomatic attribution for deep networks. In *International Conference on Machine Learning*, pages 3319–3328. PMLR.

Vuust, P., Heggli, O. A., Friston, K. J., et al. (2022). Music in the brain. *Nature Reviews Neuroscience*, 23:287–305.

Zhang, Y., Chen, J., Liao, L., and Yao, K. (2018). Random noise injection-based adversarial training for robust speech recognition. In *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3156–3160.