International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

MUSIC EMOTION RECOGNITION USING DEEP NEURAL NETWORKS AND LIBROSA

Kulsum[1], Dr. C. Berin Jones[2]
[1]PG Scholar, Department of CSE, Shadan Women's College of Engineering and Technology, Hyderabad, abdullatabbu@gmail.com
[2]Professor, Department of CSE, Shadan Women's College of Engineering and Technology jonesberin@gmail.com

ABSTRACT

In order to arrange, search, and suggest music on contemporary platforms, the categorization of musical emotions is crucial. The deep emotional content included in music may not be fully captured by traditional models, which frequently rely on raw audio or textual data. In order to effectively classify musical emotions, we suggest a Convolutional Neural Network (CNN)-based model in conjunction with Librosa for feature extraction. Mel-frequency cepstral coefficients (MFCCs), chroma characteristics, spectral contrast, and tonette representations are among the significant audio features extracted from music signals using Librosa in the suggested method. By preserving timbral, harmonic, and rhythmic elements important to emotion identification, these aspects offer a condensed and useful representation of the audio. From these collected features, hierarchical patterns are subsequently learned using the CNN model. While pooling layers highlight prominent emotional patterns and minimize dimensionality, convolutional layers automatically identify local correlations in the acoustic information. By removing the need for manually created feature combinations, this deep learning approach enables the model to successfully generalize across a variety of music samples. The suggested approach is able to capture intricate emotional links in music by fusing CNNs' pattern-learning capabilities with Librosa feature extraction. This method provides a reliable and scalable solution for automated music emotion classification, supporting real-world platform applications including playlist creation, music analytics, and music recommendation.

1. INTRODUCTION

Human mood, behavior, and cognitive processes are all influenced by music, which is a universal language that expresses a broad spectrum of emotions. Applications like playlist creation, music analytics, and tailored recommendation systems have made it more crucial to identify certain emotions in music. Conventional methods for classifying musical emotions frequently rely on written metadata or unprocessed audio signals, which may not be able to fully capture the complex emotional nuances present in the music. Furthermore, handcrafted traits might not generalize well across a variety of musical genres and can be time-consuming to extract. Automated methods for successfully addressing these issues have arisen with the development of machine learning and deep learning techniques. In order to categorize musical emotions, we present a hybrid framework in this paper that blends Convolutional Neural Networks (CNN) with Librosa-based feature extraction. Mel-frequency cepstral coefficients (MFCCs), chroma characteristics, spectral contrast, and tonnetz features are among the significant audio representations that may be extracted using Librosa, a robust audio processing package. These qualities provide a condensed yet informative input for emotion identification by capturing the timbral, harmonic, and rhythmic aspects of music. The CNN model is then used to automatically identify correlations that suggest emotional content by learning hierarchical patterns from these variables. CNN's pooling layers emphasize prominent patterns related to emotions and reduce dimensionality, while convolutional layers effectively record local dependencies. By doing away with the requirement for manual feature engineering, this deep learning method enables the model to generalize over a wide range of musical genres and styles. The suggested approach successfully models intricate correlations between audio properties and perceived emotions by fusing CNN's pattern learning capabilities with Librosa's feature extraction. The framework is appropriate for real-world music platforms since it provides automated and scalable classification. Additionally, this method improves the reliability and accuracy of music emotion identification systems, offering listeners, content producers, and streaming services useful information. The system can adjust to vast and varied music datasets by utilizing deep learning, which enhances user pleasure and recommendation quality. In the end, this initiative advances analytics and intelligent music management, allowing for emotionally conscious applications on digital music platforms.

2. EXISTING SYSTEM OF PROJECT

- The current methods for classifying musical emotions mostly use multimodal data, lyrics, or audio to identify and categorize emotions. Conventional machine learning methods or deep learning models like Convolutional Neural

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Networks (CNNs), Recurrent Neural Networks (RNNs), or hybrid models are frequently used by these systems.

- In order to ascertain the emotional tone of the music, these models concentrate on extracting acoustic characteristics like pace, pitch, and timbre as well as linguistic characteristics like mood and keywords.
- Although these methods have attained a respectable level of accuracy, they usually consider music as discrete data and disregard the larger context in which it is produced and listened to.
- For instance, performers may lean toward certain genres, listeners have subjective emotional reactions and preferences, and composers may adhere to various musical styles that elicit particular feelings.
- The depth and precision of emotion classification are reduced when these human-centric factors are disregarded. Because of this, these systems frequently have trouble generalizing across a variety of music genres and user settings, even with improvements in feature extraction and modeling.

EXISTING SYSTEM DISADVANTAGES:

- Limited Generalization with Noisy or Sparse Data
- High Computational Cost for Big Graphs
- Potential Missing of Subtle Emotional Nuances Not Captured in Graph Relations

3. RELATED WORK

A. CLASSIFICATION OF MUSIC EMOTIONS

Three primary categories can be used to group early research on music emotion classification: (1) audio signal-based methods; (2) text-based methods; and (3) multimodal methods. Feature extraction from the audio data is a common step in methods that rely on audio signals [20]. For categorization tasks, machine learning algorithms first used manually designed audio features including the zero-crossing rate [21], spectral centroid [22], and Mel spectrograms [23]. In order to extract features for genre classification, certain techniques have recently used spectrogram representations of audio data as input to Convolutional Neural Networks (CNNs) [24]. It is important to understand that copyright constraints may restrict the availability of audio information, suggesting that approaches that only use audio aspects may have drawbacks and may not be very generalizable. Many researchers have shifted their attention to textual data, like as lyrics and audience review texts, in order to extract linguistic elements for genre categorization due to the possible unavailability of audio information [25]. Using comment texts to enhance information is one example.

B. Neural Graph Networks

GNNs use the flow of these features to predict labels by distributing and integrating node information across several neural network layers. Spectral graph convolution is used by the Graph Convolutional Network (GCN) described in [26] to estimate the aggregate of input from nearby nodes.On the other hand, the Graph Attention Network (GAT) [27] collects data from the surrounding environment using attention weights, which are computed from the similarity of features across nodes. DeepWalk [28] uses local data collected by random walks to determine the network's representation. In order to create node representations, the Heterogeneous Graph Neural Network (HGNN) uses relationships to collect data about various node kinds. After processing feature embeddings in the learning route using the attention algorithm, the Hierarchical Attention Network (HAN) [17] fuses the collected semantic data once more using the attention mechanism to produce the feature embedding of semantic fusion.
In Table 1, we have classified the aforementioned approaches and enumerated their benefits and drawbacks.

C. Initial

1) Heterogeneous Graph
The definition of a heterogeneous graph is G = (V, E, A, R, φ, ϕ), where V represents the sets of nodes and E represents the sets of edges. Where |A+ R| >2. φ: V → A is a node type mapping function, and ϕ: E → Ris is an edge type mapping function. A is the sets of node types, and R is the sets of edges.
A heterogeneous graph made up of different node types (music, listener, singer, and composer) and relationships (audience relationship between music and listener, singing relationship between music and singer, composition relationship between music and composer) is shown in FIGURE 2(b).
META-PATH (2)
A path in a heterogeneous graph that represents a composite connection relationship R = R1 R2 ◦ · · · ◦ Rl between A1 and Al+1, where ◦ indicates the composition operator on relationships, is called a meta-path: A1 R1−→ A2 R2−→... Rl−→ Al+1.
For instance, three meta-paths are depicted in FIGURE 2(c):
Composer-music-composer, singer-music-singer, and listener-music-listener. Here, singer-music-singer denotes two singers performing a song together, listener-music-listener denotes two listeners listening to a song

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

together, and composer-music composer denotes two composers co-writing a song.

4. PROPOSED SYSTEM

Using auditory signals, the suggested approach is intended to automatically categorize musical emotions. The first step is Librosa-based feature extraction, which transforms unprocessed audio into useful representations like tonnetz, chroma characteristics, spectral contrast, and MFCCs. These elements provide a solid basis for identifying emotions by capturing important aspects of the music, such as pitch, timbre, and rhythm. The system can efficiently handle a variety of music samples thanks to preprocessing, which guarantees that the features are standardized and normalized. After that, the system uses a CNN-based classification module to use the retrieved features to learn hierarchical patterns. While pooling layers minimize dimensionality and emphasize the most prominent emotional patterns, convolutional layers automatically identify local correlations in the audio information. Real-world applications like music recommendation, playlist creation, and customized music analytics are supported by the system's modular design, which enables it to grow to massive music datasets and adapt to various emotion categories.

Advantages of the Proposed System

The capacity to create realistic, emotion-rich feature representations; the ability to capture complex, non-linear relationships in data; the ability to improve classification accuracy by synthesizing additional training samples; the ability to perform robustly even with limited or imbalanced datasets; and the ability to learn the joint influence of singers, composers, and listeners without relying on graph structures

5. MODULES

1. Data Collection:

This module collects a variety of musical recordings from publicly accessible music emotion datasets, streaming services, and open sources. The appropriate emotional category—such as cheerful, sad, angry, or calm—is identified on each song. To guarantee that the model can generalize across many musical genres, the dataset contains a variety of genres. An adequate depiction of emotions is ensured via proper collection, allowing the model to learn efficiently.

2. Interpreting the Data:

The main goal of this module is to examine the gathered music data in order to comprehend its features. Distributions of emotion classes, track lengths, and audio quality are examined using statistical analysis and visualizations. Additionally, this stage aids in identifying dataset imbalances that may affect model performance, such as underrepresented emotions. Preprocessing and feature extraction techniques are guided by analysis. **3. Data Preparation**:

Preprocessing gets unprocessed audio data ready for model input and feature extraction. This entails normalizing audio signals, padding or cutting recordings, and transforming them into a constant sample rate. It is also possible to apply quiet removal and noise reduction. Each track is represented in a condensed, informational way by using Librosa to extract features like MFCCs, chroma, spectral contrast, and tonnetz.

4. Adjusting the Model:

The CNN's hyperparameters, including the number of convolutional layers, filter sizes, pooling techniques, batch size, and learning rate, are adjusted during fine-tuning. Overfitting is avoided by using strategies like regularization and dropout. This module makes sure the model gets the best accuracy and does a good job of generalizing to new music samples.

5. Model Effectiveness:

This module uses metrics including accuracy, precision, recall, F1-score, and confusion matrices to assess the trained CNN's performance. Efficiency is evaluated in terms of processing requirements and prediction speed. The system's suitability for large-scale or real-time deployment on music platforms is ensured by model efficiency.

6. Predictive Outcomes:

The method forecasts the emotional labels of brand-new or unheard music in the last module. Against verify performance, predicted emotions are compared against real labels. Graphs or dashboards are used to illustrate the results, allowing users to decipher musical emotional patterns. This module shows how the model can be used in playlist creation and music recommendation systems.

6. TECHNIQUES USED IN PROJECT

6.1 EXISTING TECHNIQUE:

A Heterogeneous Graph Neural Network (HGN), which models many node types and relationships in a graph structure, is used in the current system. HGN depicts singers, composers, and listeners as nodes in the context of music emotion classification, and their interactions—such as genre preferences, emotional expression, or teamwork—as edges. Using methods like attention mechanisms and meta-path learning to handle heterogeneous input, the algorithm gathers information from its neighbors to learn meaningful feature representations for every node. HGN seeks to enhance categorization performance by capturing the intricate relational structure between music creators and

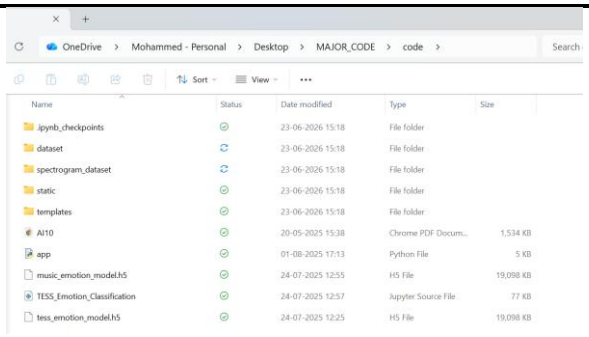
International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

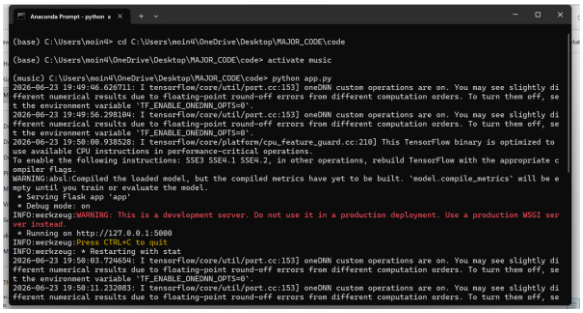
consumers. Even while HGN is good at learning structured relationships, it may have trouble generalizing when dealing with sparse or unbalanced data. Additionally, it may not accurately depict the underlying emotion-generative process in music because it strongly relies on graph building and metadata availability, particularly when emotional cues are subtle or poorly reflected in the graph structure.

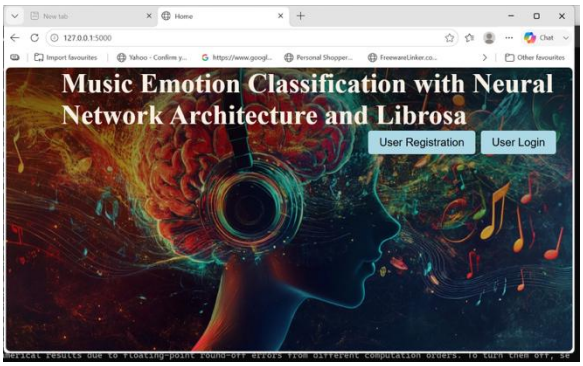
6.2 PROPOSED TECHNIQUE USED OR ALGORITHM USED:

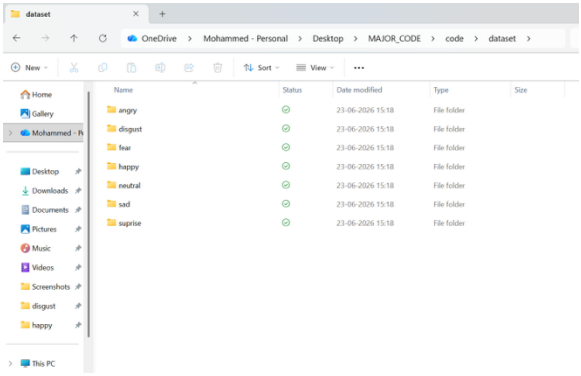
The suggested algorithm initially extracts pertinent audio elements from music songs using Librosa. Chroma characteristics indicate harmonic content, spectral contrast highlights variations between peaks and troughs in the spectrum, and MFCCs capture the spectral envelope and timbral qualities of sound. The CNN can concentrate on emotion-related traits rather than unprocessed audio waves thanks to these properties, which create a condensed yet expressive input. After that, a CNN architecture is used to categorize emotions and learn hierarchical feature representations. While pooling layers simplify and preserve important emotional cues, convolutional layers find patterns in the input information.. The final emotion category is predicted by fully connected layers that integrate the learnt patterns. In order to ensure that the system can generalize across a variety of musical genres and emotional expressions, it is trained using supervised learning with backpropagation. This robust and scalable method for automatic music emotion classification combines deep learning with feature extraction.

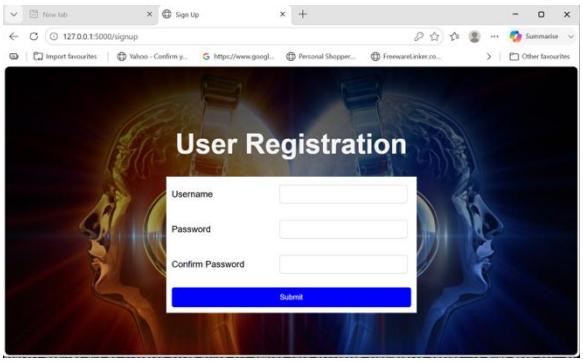
7. RESULT



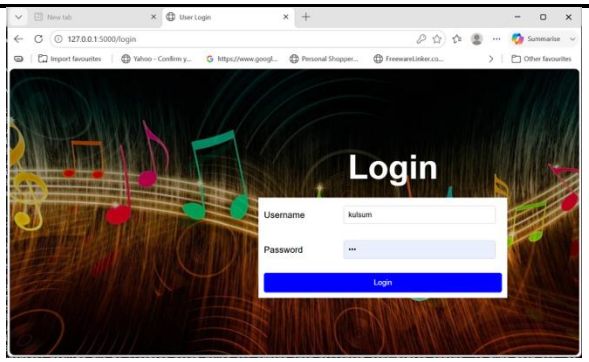


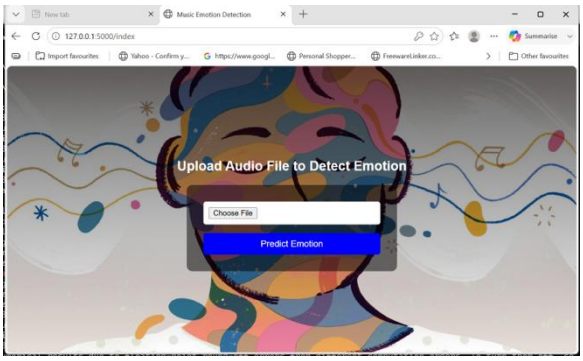


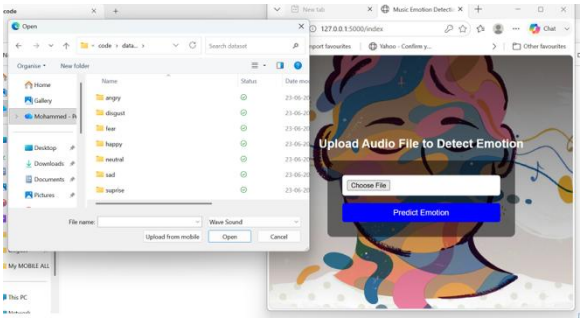


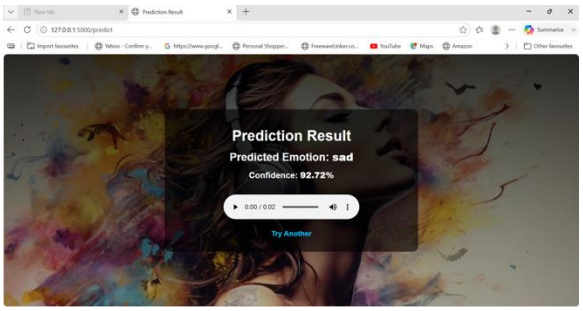


International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper









8. CONCLUSION

Personalized suggestions, intelligent playlist creation, and increased user interaction are made possible by music emotion classification, which is an essential part of contemporary music platforms. This project demonstrated a deep learning system for efficient emotion recognition that blends Convolutional Neural Networks (CNN) with Librosa-based feature extraction. The timbral, harmonic, and rhythmic components of music were captured by using Librosa to extract audio properties such MFCCs, chroma, spectral contrast, and tonnetz. Next, CNN was used to automatically identify emotional cues in the collected data and learn hierarchical patterns. The hybrid method guarantees versatility across a variety of genres and styles and does away with the necessity for bespoke features. Through training, refinement, and assessment, the model showed that it could accurately classify emotions and generalize effectively. In addition to addressing difficulties in music information retrieval, this study advances the expanding subject of affective computing. The created framework can be included into actual systems for analytics and music recommendation, enhancing user happiness. The system has the potential to become a scalable, real-time solution for emotion-aware music applications with additional research and improvements, making it a significant contribution to the development of intelligent music technology.

9. FUTURE ENHANCEMENT

By using multimodal data—such as song lyrics, metadata, and user feedback—in addition to audio aspects, the suggested music emotion categorization system can be improved. The capacity to identify temporal and contextual trends can be enhanced by integrating sophisticated deep learning models, such as CNN-LSTM hybrids or Transformers. Generalization can be improved with a larger, more varied dataset that spans several languages and cultures. While the music is playing, emotions can be categorized using real-time streaming analysis. Pre-trained audio models can be used in transfer learning to improve performance with little data. Subsequent iterations can concentrate on identifying mixed or subtle emotions rather than just discrete categories. By highlighting the features that contribute most to predictions, explainable AI approaches can enhance interpretability. End users can receive emotion-aware recommendations quickly through mobile and cloud-based deployment. Its usefulness will increase by integration with well-known music platforms. Ultimately, based on emotional conditions, the system can develop into a fully customized music helper.

REFERENCES

[1] L. Zhou, ‘‘Cultivation of artistic expression in college music and vocal music teaching,’’ Art Perform. Lett., vol. 4, no. 12, pp. 43–49, 2023.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[2] S. Ding, ''Research on the artistic expression of vocal music,'' in Proc. 2nd Int. Conf. Culture, Educ. Econ. Develop. Modern Soc. (ICCESE). Atlantis Press, 2018, pp. 663–665.
[3] A. Sabbadini, ''Opera on the couch: Music, emotional life, and unconscious aspects of music,'' Int. J. Psychoanalysis, vol. 104, no. 1, pp. 183–185, Jan. 2023.
[4] Q. Xianyang, ''Research of lens model in music emotional communication,'' BioTechnol, Indian J., vol. 10, p. 19, 2015.
[5] C. Nussbaum, A. Schirmer, and S. R. Schweinberger, ''Electrophysiological correlates of vocal emotional processing in musicians and non-musicians,'' Brain Sci., vol. 13, no. 11, p. 1563, Nov. 2023.
[6] J. J. Campos-Bueno et al., ''Emotional dimensions of music and painting and their interaction,'' Spanish J. Psychol., vol. 18, p. E54, 2015.
[7] T. Fischinger, M. Kaufmann, and W. Schlotz, ''If it's mozart, it must be good? The influence of textual information and age on musical appreciation,'' Psychol. Music, vol. 48, no. 4, pp. 579–597, Jul. 2020.
[8] X. Cai and H. Zhang, ''Music genre classification based on auditory image, spectral and acoustic features,'' Multimedia Syst., vol. 28, no. 3, pp. 779–791, Jun. 2022.
[9] B. Wilkes, I. Vatolkin, and H. Müller, ''Statistical and visual analysis of audio, text, and image features for multi-modal music genre recognition,'' Entropy, vol. 23, no. 11, p. 1502, Nov. 2021.
[10] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu, and X. Liu, ''A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection,'' IEEE Trans. Instrum. Meas., vol. 71, pp. 1–14, 2022.
[11] Y. Dong, Q. Liu, B. Du, and L. Zhang, ''Weighted feature fusion of convolutional neural network and graph attention network for hyperspectral image classification,'' IEEE Trans. Image Process., vol. 31, pp. 1559–1572, 2022.
[12] D. Pathak and U. S. N. Raju, ''Content-based image retrieval for super-resolutioned images using feature fusion: Deep learning and hand crafted,'' Concurrency Comput., Pract. Exper., vol. 34, no. 22, 2022, Art. no. e6851.
[13] C. Yuan, Q. Ma, J. Chen, W. Zhou, X. Zhang, X. Tang, J. Han, and S. Hu, ''Exploiting heterogeneous artist and listener preference graph for music genre classification,'' in Proc. 28th ACM Int. Conf. Multimedia, Oct. 2020, pp. 3532–3540.
[14] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, ''A comprehensive survey on graph neural networks,'' IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 1, pp. 4–24, Jan. 2021.
[15] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, ''Graph neural networks: A review of methods and applications,'' 2018, arXiv:1812.08434.
[16] W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, and D. Yin, ''Graph neural networks for social recommendation,'' in Proc. World Wide Web Conf., 2019, pp. 417–426.
[17] X. Wang et al., ''Heterogeneous graph attention network,'' in Proc. World Wide Web Conf., 2019, pp. 2022–2032.
[18] R. Bing, G. Yuan, M. Zhu, F. Meng, H. Ma, and S. Qiao, ''Heterogeneous graph neural networks analysis: A survey of techniques, evaluations and applications,'' Artif. Intell. Rev., vol. 56, no. 8, pp. 8003–8042, Aug. 2023.
[19] C. Shi, ''Heterogeneous graph neural networks,'' in Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2019, pp. 793–803.
[20] V. Mounika and Y. Charitha, ''Mood-Enhancing music recommendation system based on audio signals and emotions,'' in Proc. Int. Conf. Inventive Comput. Technol. (ICICT), Apr. 2023, pp. 1766–1772.
[21] X. Song et al., ''Automatic recognition of uterine contractions with electrohysterogram signals based on the zero-crossing rate,'' Sci. Rep., vol. 11, no. 1, p. 1956, 2021.
[22] B. Baris, M. E. Cek, and D. G. Kuntalp, ''Modulation classification of MFSK modulated signals using spectral centroid,'' Wireless Pers. Commun., vol. 119, no. 1, pp. 763–775, Jul. 2021.
[23] D. H. Rudd et al., ''Leveraged mel spectrograms using harmonic and percussive components in speech emotion recognition,'' in Proc. Pacific Asia Conf. Knowl. Discovery Data Mining. Cham, Switzerland: Springer, 2023, pp. 392–404.
[24] Y. Zhang, G. Kolkman, and H. Watanabe, ''Phase repair for time domain convolutional neural networks in music super-resolution,'' 2023, arXiv:2306.11282.
[25] N. J. O'Leary, ''The tempest presented by the lord Denney's players, and: The tempest presented by the Cincinnati Shakespeare company,'' Shakespeare Bull., vol. 35, no. 3, pp. 487–495, 2017.
[26] T. N. Kipf and M. Welling, ''Semi-supervised classification with graph convolutional networks,'' 2016, arXiv:1609.02907.
[27] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, ''Graph attention networks,'' 2017, arXiv:1710.10903.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[28] B. Perozzi, R. Al-Rfou, and S. Skiena, ‘‘Deepwalk: Online learning of social representations,’’ in Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2014, pp. 135–144.