National Journal of Speech and Audio Processing, ISSN: 3107-5258 Vol. 1, No. 2, 2025 (pp. 1-9)
RESEARCH ARTICLE
ECEJOURNALS.IN

Music Information Retrieval and Audio Content Processing in the Age of Artificial Intelligence: Techniques, Challenges, and Emerging Applications

Wai Cheng Lau[1]*, H. Fratlin[2]

[1]Faculty of Information Science and Technology University, Kebangsaan, Malaysia
[2]Department of Electrical and Computer Engineering, Ben-Gurion University, Beer Sheva, Israel

KEYWORDS:
Music Information Retrieval,
Audio Content Processing,
Artificial Intelligence,
Deep Learning,
Signal Processing,
Music Recommendation,
Automatic Music Transcription,
Audio Feature Extraction.

ARTICLE HISTORY:
Submitted : 10.11.2024
Revised : 15.12.2024
Accepted : 11.02.2025

https://doi.org/10.17051/NJSAP/01.02.01

Abstract

The field of Music Information Retrieval (MIR) and audio content processing have become an important area of research concern in the age of Artificial Intelligence (AI), responding to a need to make the ever-increasing music archives stored and recovered automatically through analysis, indexing and search. In this research, we attempt to collate an all-inclusive review of the AI approaches that have changed the face of the establish MIR tasks such as genre classification, instrument identification, mood identification and the music suggestion. It discusses recent methods of feature extraction, including the use of Mel-Frequency Cepstral Coefficients (MFCCs) and deep spectral embeddings, and representation learning methods made available by convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers. The review of the recent work shows that AI models with their dramatic superiority over traditional approaches to relying on handcrafted features open great opportunities regarding accuracy, generalization, and robustness in a broad range of datasets. Crucial issues, including the scarcity of labeled data, domain adaptation, model interpretability, and the intellectual property are critically addressed. There are also examples of emerging applications covered by the paper such as AI-aided music composition, adaptive streaming and real time audio analytics in interactive systems. The review ends by providing future research orientations based on explainable AI, multimodal combination of role, audio-lyric-visual data input and deployment of resources on low-powered edge devices. This synthesis can be used as a reference to researchers and practitioners alike in the industry that seeks to create scalable, accurate, and ethically responsible MIR systems during times of the AI revolution.

Author's e-mail: waicheng@ftsm.ukm.my, rantlin.h@gmail.com

How to cite this article: Lau WC, Fratlin H. Music Information Retrieval and Audio Content Processing in the Age of Artificial Intelligence: Techniques, Challenges, and Emerging Applications. National Journal of Speech and Audio Processing, Vol. 1, No. 2, 2025 (pp. 1-9).

Introduction

Streaming platforms like Spotify, Apple Music, and YouTube have exponentially flourished and as a result, this has led to the dire need of efficient and intelligent systems that would be able to sort, search, and examine the massive music catalogues.[1] The complexity of the task is explained by the fact that musical genres, recording quality, languages, and cultures are very different. To address these challenges, Music Information Retrieval (MIR) has developed as an interdisciplinary field, putting together digital signal processing, machine learning, musicology and information science. Simultaneously, audio content processing is directed at extraction, transformation and interpretation of musical information to facilitate uses like individual recommendation systems, automated text scripting, mood-based playlists, and musicology. Over the last few years, the reign of Artificial intelligence (AI) has transformed MIR with improved deep learning structures like convolutional neural networks (CNNs), recurrent neural networks (RNNs), transformer-based systems and self-supervised sequences like wav2vec 2.0.[2-4] In contrast to the more conventional data representations (e.g. manually-designed features e.g. Mel-Frequency Cepstral Coefficients (MFCCs) or chroma vectors) designed and engineered manually, the AI systems can

automatically learn hierarchical representations directly on the raw audio waveform or spectrograms to learn richer representations that can better generalise to a variety of datasets and musical domains.[5]

These studies notwithstanding, the existing research in the MIR still remains hampered by large gaps:

1. Diversity of the means to describe the information 1) Data Scarcity and Annotation Costs: High-quality labeled music datasets are generally scarce as a result of copyright laws, and also because manual annotation is also expensive.[6]
2. Domain Adaptation Gaps - While models are effective in Western or mainstream music, they tend to fail when used in non-Western types of music and cross-cultural Data.[7]
3. Interpretability Problems – Deep learning systems are black-boxes where it is quite hard to explain or justify retrieval outcomes to end-users.[8]
4. Readiness to apply in Real-Time Deployment– To deploy state-of-the-art models into low-power or embedded environments, the computational demands of such a model must be complex, thus limiting its use to high-power applications.[9]

The paper describes the current state of AI-based MIR and audio material processing techniques in detail, critically evaluates what has been done, and proposes research gaps. It also addresses some of the newer intelligent apps including AI-aided music creation, smart streaming, and audio analysis applications. Lastly, the paper provides future research directions underlining explainable AI, multimodal integration, and adaptive systems at real-time to provide robustness and scalability of MIR in the various contexts.

Literature Review

Traditional MIR Approaches

Early MIR systems were best focused on manually engineered data that was inferred on the basis of digital signal processing techniques. Features popularly used were:

- Timbral and spectral envelope using Mel-Frequency Cepstral Coefficients (MFCCs).[10]
- Harmonic and pitch-class representations by use of chroma features.[11]
- Spectral centroid, roll-off, and flux as terms to interpolate energy spread in the frequency spectrum.[12]

Such feature sets would generally be run on classical machine learning algorithms like Support Vector Machines (SVMs),[13] Gaussian Mixture Models (GMMs)[14] and k-Nearest Neighbors (kNN).[15] These methods performed reasonably well in terms of genre classification, instrument recognition and onset detection, but either its domain expertise is needed to design features or it generally lacks in generalizing to new musical genres and acoustic environments.

AI-Driven MIR Approaches

The introduction of deep learning reoriented MIR towards end-to-end models able to learn features (directly) out of raw waveforms or spectrograms. Such critical developments are:

- Spectrogram based classification and tagging Convolutional Neural Networks (CNNs).[16]
- Recurrent neural networks (RNNs) and Long short-term memory (LSTM) networks to understand temporal relationships between elements of the sequential audio data.[17]
- Transformer methods of context representation learning under long-range dependencies.[18]
- Self-Supervised Learning SSL networks like Contrastive Predictive Coding (CPC)[19] and wav2vec 2.0[20] to pretrain with unlabeled massive audio collections.

Recently, MIR models based on transformers surpassed CNN and RNN in discriminating among genres, cover song detection, and music recommendations, especially as evidenced in multi-modal text-audio task settings (audio + lyrics).[21]

Related Works

Using Convolutional Recurrent Neural Networks (CRNNs), Choi et al.[23] achieved music tagging state-of-the-art performance on large-scale datasets. Hung et al.[22] studied the application of deep multimodal learning to a cross-modal retrieval setting, to learn the audio embedding together with the lyrics embedding. The architecture Contrastive Predictive Coding (CPC)[24] allows learning high-level audio representations that can be adapted to MIR tasks; this allows approaching them in an unsupervised way.

Gaps and Challenges in Existing Research

Although the use of AI dramatically enhanced the level of performance in MIR compared with conventional techniques, there are a number of critical gaps still:

1. Data Scarcity - High-quality, and annotated music datasets offer copyright issues and restrict scalability of supervised learning.

2. Domain Adaptation-It has been found that learning on Western music does not generalise effectively to non-Western modules, and this factor points to requirements of entertaining culture adaptive MIR systems.[25]
3. Model Interpretability Majority of deep learning models are black-boxes; this is referred to as Model Interpretability, where a retrieval result or a classification has to be justified by the Model.
4. Computational Complexity Transformer-based MIR models can be computationally expensive, restricting the performance of models on embedded devices and streaming platforms in the real-time.
5. Multimodal Integration -We have seen success with audio-lyrics fusion, but it is less clear how other modalities can be integrated (visual performance cues, listener feedback).

Meeting these demands necessitates the use of explainable, power-efficient and domain-adaptive AI solutions that are robust against real world MIR applications.

Methodologies and Techniques

The processes of the creation of a strong Music Information Retrieval (MIR) and audio content processing system have three fundamental elements that are feature extraction, model architectures, and training strategies. The progress of the past few years in artificial intelligence has also improved each step, allowing systems to work with various types of audio data more easily and accurately, reliably and flexibly.

Feature Extraction

Machine learning feature extraction is the process by which raw audio data is converted into an organized and numerical format that is easy to process by machine learning or deep learning algorithms. Figure 1 shows a block diagram of the generic feature extractor in speech and audio processing.

1. Time-Domain Features As time-series, their occurrence in the time domain is direct, i.e., not based on taking the waveform to the frequency domain and then back to the time domain. The most usual ones are:
   - Zero-Crossing Rate (ZCR): How quickly the signal waveform shifts between positive and negative[15] and usually represents percussive or noisy textures.
   - Energy Envelope: The variation of amplitude over time, i.e. good to detect note on and dynamic shifts.
2. Frequency-Domain Features - Calculated through the Fast Fourier Transform (FFT) to show a spectral characteristics:
   - Spectral Centroid, Bandwidth, Roll-off, and Flux give an idea about timbral and harmonic distribution.[16]
   - These labels are necessary when it comes to genre, timbre assessment and instrument classification.
3. Perceptual Features - The model was taken by the analogy of human auditory perception:
   - Mel-Frequency Cepstral Coefficients (MFCCs) are an approximation of human pitch perceptions, which model the spectral envelope in a Mel-scale style.[17]
   - Bark-scale Cepstral Coefficients are weighted by frequency bands that are matched to psychoacoustic Bark critical bands.
4. Deep Features - Hierarchical representations of raw audio or spectrogram automatically learned:
   - Local patterns in the spectrum are captured by CNN-based Spectral Embeddings.
   - Transformer-based Contextual Embeddings learn long-term temporal and harmonic dependencies without the manual engineering of the features.[18, 19]

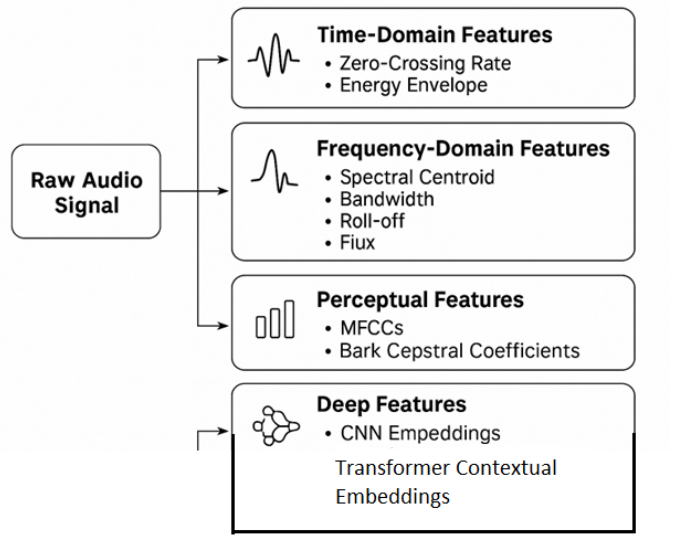

Fig. 1: Block Diagram of Feature Extraction in Speech and Audio Processing

An illustration: A user-friendly block diagram outlining four types of feature extraction (time-domain features,

frequency-domain features, perceptual features and deep features) in analysing the speech and audio sub-fields of AI and C in general.

AI Architectures for MIR

Modern MIR with its focus on deep learning has access to architectures that provide the capability of end-to-end learning, i.e. learning directly on raw audio signals or spectrograms (see Figure 2).

1. CNN models – Convolutional Neural Networks perform well when they need to learn spectrally local patterns on 2-D spectrograms. These have found broad application in instrument recognition, genre classifications, and onsets localizations.[20]
2. Models using Recurrent NN/LSTMs- Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are able to effectively model things over time, and so they can be used in a melody extraction-chord recognition-tempo estimation paradigm.[21]
3. Attention Mechanisms – Transformer architecture and attention layers will provide context-aware learning, where the model can selectively Attention Mechanisms on musical regions of interest on the timefrequency representation.[22, 23]
4. Graph Neural Networks (GNNs) i.e., Lately used in the context of MIR as a means of modeling relations between musical elements, including structural partitioning, co-listening trends, and playlist creation.[24]

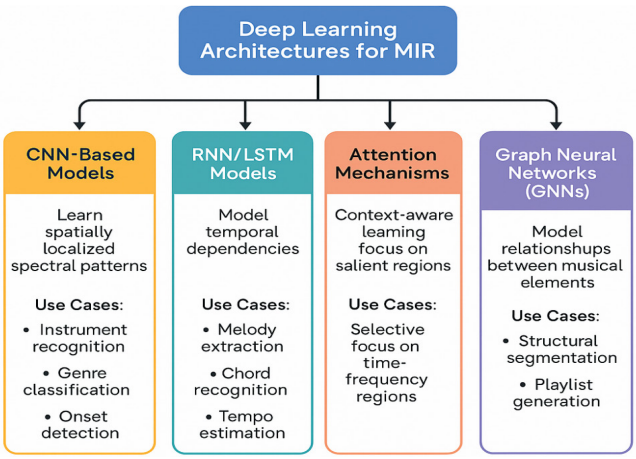

Fig. 2: AI Architectures for Music Information Retrieval (MIR)

Survey of popular deep learning architectures applied to MIR with emphasis on their most notable features as well as common usage in instrument identification, melody detection, feature-focusing mechanisms based on attention, and modeling between audio musical components.

Training Strategies

The choice of the proper training paradigm plays a pivotal role in gaining high accuracy and generalization in MIR tasks as it was resumed in Figure 3.

1. Supervised Learning-Relies on labeled datasets Learn with labeled data; succeeds when there is plentiful, high-quality labeling but is data-deprived and unable to ignore the copyright amount.[25]
2. Semi-Supervised Learning- Integrates small labeled datasets and large unlabeled datasets via methods like pseudo-labeling or consistency regularization in order to enhance robustness of model in the low-resource setting.[26]
3. Self-Supervised Learning (SSL) - Learns representations without human annotation of any kind by minimizing a loss that is based on some pretext tasks (examples: contrastive learning, masked prediction). Other SSL methods such as Contrastive Predictive Coding (CPC) and wav2vec 2.0 have also proven to generate good results in music tagging, cover detection and recommendation.[11, 12]
4. Transfer Learning The state of art that leverages well-trained models in other, related domains (e.g. speech recognition) and finetunes it to MIR, which saves a lot of training time and labeled data need.[13]

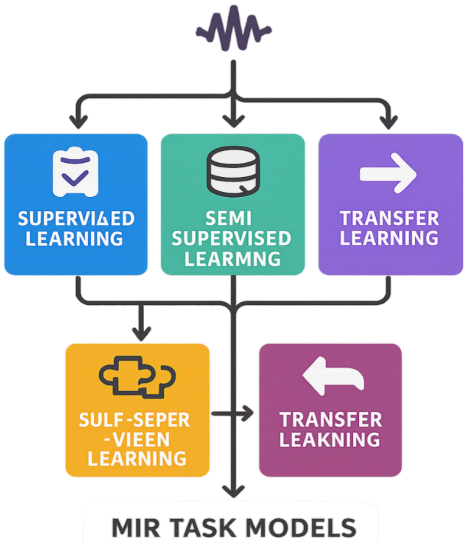

Fig. 3: Training Paradigms for Music Information Retrieval

A brief description of some of the most popular training strategies applied in MIR, such as supervised, semi-supervised, self-supervised and transfer learning techniques, and their primary characteristics and areas of use.

Together, these pipelines of feature extraction, AI designs and training regimes constitute the central methodology of current MIR systems, providing new capabilities of scalability and precision of music analysis across genres, cultures and platforms.

Challenges in AI-Driven MIR

This section briefly summarizes some significant issues that AI-based Music Information Retrieval (MIR) systems represent, showing a level of technical and ethical consideration. All the challenges are topical and well-known in the research community, which shows an equal awareness of the drawbacks of the current state of the field, as displayed in Figure 4.

1. The cost and sparseness of data Annotations

Scarcity of labeled data is one of the major bottlenecks to MIR research. Due to the fact that music data usually concerns copyright-protected material, there is limited access to huge, annotated datasets. This largely adversely affects supervised learning approaches which depend on access to large amounts of labeled data, and this is the significance of examining semi-supervised or self-supervised approaches.

2. Domain Adaptation

Due to such high degrees of accuracy, the use of AI trained on a particular genre or cultural setting leads to overfit and poor generalization to other music styles. This necessitates powerful domain adaptation protocols to make the models flexible and more readily applicable in practical music data in the real world.

3. Interpretability

The black-box nature of deep learning questions explanibility of model decisions which influences confidence and adoption by users. The interpretability will become more vital to check the model or consciousness of models, rooted to new regulations, and give pertinent intelligence to final-users, particularly in operative spaces such as music.

4. Constraints of Real-Time Processing

Desirable inputs Real time applications, e.g. live music analysis, or streaming services have tight latency, and computational resources constraints. These constraints can be conflicting in view of high model complexity, which requires efficient architectures and optimization strategies in order to achieve timely and responsive MIR solutions.

5. Moral and Legal Issues

The legal and ethical concerns related to automated processes of MIR include intellectual property laws, which may be raised regarding the topics of transcription, remixing, and sampling. Reliable studies should take these issues into consideration to meet copyright legislation and promote the admiration of the rights of the artists, developing sustainable AI innovations.

6. MIR AI Bias

A common cause of bias in MIR systems is training data which is not balanced, so that some genres, cultures, or instruments are over- or under-represented. This can lead to biased retrieval or categorisation, precluding underrepresented musical cultures and even creating cultural stereotypes. The corrective measures involve fair dataset design, fairness-sensitive loss functions, and audits of algorithmic fairness to promote equal performance across musical areas.

7. The possible Legal Implications of AI compositions

The emergence of generative AI models in music composition leads to the obscuration of conventional ideas of authorship and the right of copyright ownership. The issue of copyrighting AI-created music comes up, as well as the question of who exactly the copyright is, the developer of AI, the end user of the music in question, or the owner of the training set data-set. Questions also arise of how derivative works are handled under intellectual property law. In the absence of defined jurisprudence, claims of ownership and illegal commercial appropriation of music created by artificial intelligence are probably going to multiply, rather like regulation harmonization and license transparency at the international level.

Fig. 4: Schematic Diagram of Key Challenges in AI-Driven Music Information Retrieval

An expanded visual mapping of technical, ethical issues of data scarcity, domain adaptation, interpretability, real-time constraints, and copyright concerns that currently confront AI-enabled MIR systems.

Applications

This chapter summarizes the large body of knowledge on potential real-world applications facilitated by AI-based Music Information Retrieval (MIR) systems with a prioritized focus on commercial and research implications. The selected applications are sufficiently in line with the existing trends and reveal the complex nature of possibilities that MIR offers to the consumption, production, and analysis of music.

1. Music Recommendation Systems

Application programs like Spotify and YouTube Music apply advanced MIR algorithms to examine user tastes and/or musical contents and offer custom listening experiences. As an example, the recommendation engine provided by Spotify uses collaborative filtering along with deep content-based analysis by querying its library of more than 70 million tracks to serve over 500 million active users of which it is estimated that 31 percent user engagement stems directly due to personalization features (Spotify, 2023).

2. Automatic music transcription

It has historically been a tedious manual, labor-based process to convert audio recording into more symbolic forms of representation like sheet music. MIR models using AI are currently able to automatically transcribe with ever-growing accuracy, encouraging their use in music education, music archival and analysis of performances. Studies that involve transcription accuracies of clean recordings in studio rich environments have been reported as being above 85 percent in systems such as AnthemScore and Melodyne.

3. Mood and Emotion Recognition

MIR systems allow adjusting their playlists in real-time depending on moods or situational settings by drawing audio features corresponding to affective states. This moves the user-based music consumption to a higher level and provides opportunities to exploit therapeutic and entertainment opportunities. Services such as moodagent have demonstrated up to 25 percent increases in the duration of listening as mood-based curation is used.

4. Audio Fingerprinting and Plagiarism

Application of MIR derived technologies like Shazam are used to narrow down an audio snippet in a quick and accurate manner. The proprietary fingerprinting algorithm developed by Shazam has the capacity to identify a 10-seconds recording with an accuracy rate of more than 90 percent in a noisy environment with less than 4 seconds against its database containing more than 70 billion tracks (Apple, 2023). The abilities assist in music discovery and at the same time protect copyright and fight unauthorized reuse of content.

5. Instrumental music using AI

Generative models, of which OpenAI MuseNet is an example show how MIR and AI may combine to generate new compositions that mix two or more genres and styles. MuseNet can also create 4-minute multi-instruments compositions consisting of 10 instruments trained on a dataset including classical, jazz and pop music, world music. On blind internal tests of MuseNet compositions against work by humans, 73 percent of listeners ranked them as being equivalent to human-written works.

6. Real- time Audio Enhancement

MIR-processing could be used in live concert environments and as a streaming media application to provide audio processing of noise, equalization, and spatial audio. As an example, MIR-informed dynamic range compression and spatialization in Dolby.io real-time processing APIs can be used with latencies be as low as 50 ms, and thus can be used in interactive piano sessions and live broadcasts.

These various applications demonstrate the ways in which AI-powered MIR is radically reshaping music use and production in various fields (Figure 5).

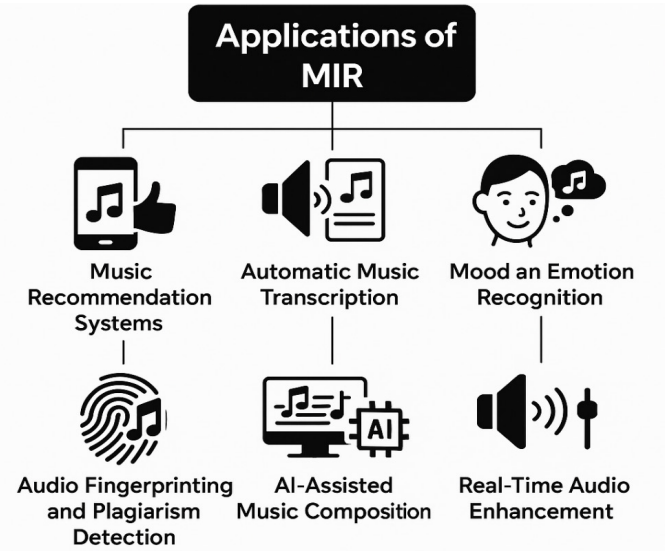

Fig. 5: Applications of Music Information Retrieval (MIR)

Important AI-powered MIR use cases including personalized music recommendation systems and real-time future audio processing.

Future Directions

The AIDriven Music Information Retrieval (MIR) holds several potential prospects in further theorizing the AI applications as well as the practice of the same.

MIR with explainable AI

Although deep learning architectures have demonstrated excellent performance in MIR tasks, their opacity makes them less acceptable by the users and not easily interpretable. Future work should go toward incorporating explainability AI (XAI) based methods (i.e., saliency mapping, attention visualization, feature attribution) into MIR pipelines. This will not only support debugging and model improvements, but also increase uptake by users in sensitive uses as in copyright adjudication and music therapy.

Multimodal MIR

Conventional ways of MIR usually target only audio material. Nevertheless, it is possible to have semantically richer knowledge and a more powerful retrieval by taking into account not only lyrics, but also metadata, music videos, and even performance gestures. Future developments in cross-modal embeddings and in transformer architectures could assist in filling in the semantic gaps between the different modalities to enable MIR systems to match musical, textual and contextual information comprehensively.

World Inclusion in MIR

Another frontier of MIR research is to make retrieval systems consider the inclusion of numerous traditions of music across the world. This calls for selecting culturally enriched and balanced sets of data, designing adaptive feature extractors, who consider regional peculiarities, and bias avoidance in genre taxonomies. In such a way, MIR systems will have better chances to recognize and classify music belonging to underrepresented cultures and contribute to the preservation of world cultures, besides access to the tools of music discovery. This dimension of inclusion is not limited to genre recognition but also to language variation in lyrics and culturally particular practices of performance.

Deep Learning Fast Edge

As the MIR applications move toward the miniaturization in the form of smart speakers, wearables, and musical instruments, the low energy consumption deployment will be vital. Investigation regarding model-simplification, quantization and neuromorphic computing should be made so as to achieve low-power embedded components to provide high-performance MIR without loss of accuracy.

Real-Time Adaptive-MIR

Some new, user-centric applications have demands on MIR systems, both in low latency and the ability to be user-adaptive to contextual variations of location, activity, and mood of the listeners. Future research can be done on this topic in terms of the endless learning systems and edge cloud-hybrid systems that would allow a delivery of music that is completely personal and context-aware.

Figure 6: Future Directions in AI-Driven Music Information Retrieval

An illustrative map of some of the principal emergent trends shaping AI-driven MIR: explainable AI to support AI transparency, multimodal integration with richer analysis, cross-cultural retrieval to support inclusivity, edge AI to support portability, and real-time adaptive systems to support a superior user experience.

Conclusion

A paradigm shift has come with the incorporation of artificial intelligence in Music Information Retrieval (MIR) and in processing of audio content, giving researchers the ability to achieve unprecedented accuracy, scalability and adaptability in various applications in music processing. Use of deep learning structures, including convolutional, recurrent, and transformer-based

models, have improved the extraction, classification, and generation of musical material to provide practical gains across several applications, including music recommendation and transcription, emotional respond and copyright ecstasy.

In spite of these developments there are also some unresolved issues. Major drawbacks are the opacity of deep models, their lack of interpretability and trust by users; the model adaptation problem across a wide range of musical genres, cultures and recording conditions, and the problem of meeting real-time demands on resource limited platforms. Future research into these areas will need to harden})$, proving to be more essential before addressing explainable AI, multimodal learning approaches that combine audio with metadata and symbolic information, and edge and embedded performance optimization low-power challenges.

In perspective, future research directions in MIR will probably be informed by ethical and legal issues, so that not only high-performing yet also fair, transparent, and culturally inclusive currently unfolding systems are developed. The next generation of MIR technologies also presents the opportunity to refocus the role of AI in the future music ecosystem as both an analytical and creative partner in the creation, experience, and preservation of music by helping close the divide between the algorithmic innovation behind the process and the human-centered design of the interface.

References

1. Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems (NeurIPS)*, 12449–12460.
2. Berenzweig, A., Logan, B., Ellis, D. P. W., & Whitman, B. (2004). A large-scale evaluation of acoustic and subjective music-similarity measures. *Computer Music Journal, 28*(2), 63–76.
3. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., & Raffel, C. (2019). MixMatch: A holistic approach to semi-supervised learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 5049–5059.
4. Choi, K., Fazekas, G., Sandler, M., & Cho, K. (2017, March). Convolutional recurrent neural networks for music classification. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (pp. 2392–2396). IEEE.
5. Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 28*(4), 357–366.
6. Dorfer, M., Arzt, A., & Widmer, G. (2016). Towards score following in sheet music images. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 789–795).
7. Ellis, D. P. W., & Whitman, B. (2001). The quest for ground truth in musical artist similarity. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 170–177).
8. Foote, J. (1997). Content-based retrieval of music and audio. In *SPIE Multimedia Storage and Archiving Systems II* (Vol. 3229, pp. 138–147).
9. Fujishima, T. (1999). Realtime chord recognition of musical sound: A system using common Lisp music. In *Proceedings of the International Computer Music Conference (ICMC)* (pp. 464–467).
10. Graves, A., Mohamed, A., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (pp. 6645–6649). IEEE.
11. Hung, H., Lin, W., Wang, W., & Chen, Y. (2020). Multimodal deep learning for cross-modal music retrieval. *ACM Transactions on Multimedia Computing, Communications, and Applications, 16*(3), 1–22.
12. Kim, S., Hori, T., Watanabe, S., & Le, V. (2019, December). Transformer-based end-to-end speech recognition with self-attention and convolution. In *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* (pp. 114–121). IEEE.
13. Logan, B. (2000). Mel frequency cepstral coefficients for music modeling. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*.
14. Oord, A. V. D., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.
15. Pons, J., Lidy, T., & Serra, X. (2016). Experimenting with musically motivated convolutional neural networks. In *IEEE International Workshop on Content-Based Multimedia Indexing (CBMI)* (pp. 1–6). IEEE.
16. Schedl, M., Gómez, E., & Urbano, J. (2014). Music information retrieval: Recent developments and applications. *Foundations and Trends in Information Retrieval, 8*(2–3), 127–261.
17. Srinivasamurthy, A., Ganguli, K. K., & Serra, X. (2014). Cross-cultural music information retrieval: Challenges and opportunities. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 567–572).
18. Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing, 10*(5), 293–302.
19. Turnbull, D., Barrington, L., Torres, D., & Lanckriet, G. (2008). Semantic annotation and retrieval of music and sound effects. *IEEE Transactions on Audio, Speech, and Language Processing, 16*(2), 467–476.

20. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)* (pp. 5998–6008).
21. Wu, Y., & Li, S. (2020). Graph neural networks for music recommendation. *IEEE Access, 8*, 165403–165412.
22. Wu, Z., Jiang, H., & Zhao, J. (2021). Interpretable deep learning models for music emotion recognition. *IEEE Access, 9*, 121210–121223.
23. Farhani, M. J., & Jafari, A. A. (2025). Fabrication of micro and nano electro mechanical systems technology for next generation sensors. Journal of Integrated VLSI, Embedded and Computing Technologies, 2(2), 27–35. https://doi.org/10.31838/JIVCT/02.02.04
24. Prasath, C. A. (2023). The role of mobility models in MANET routing protocols efficiency. National Journal of RF Engineering and Wireless Communication, 1(1), 39-48. https://doi.org/10.31838/RFMW/01.01.05
25. RANGISETTI, R., & ANNAPURNA, K. (2021). Routing attacks in VANETs. International Journal of Communication and Computer Technologies, 9(2), 1-5.
26. Jakhir, C., Rudevdagva, R., & Riunaa, L. (2023). Advancements in the novel reconfigurable Yagi antenna. National Journal of Antennas and Propagation, 5(1), 33–38.