Chapter 5

Key Applications of Machine Learning in Music Data

Nigar Tuğbagül Altan Gülgün[1]

Abstract

Music has held an important place in human life throughout history and has continued to exist as a significant form of expression in cultural, social, and artistic contexts. With the advancement of technology, the ways in which music is produced, distributed, and consumed have undergone substantial transformation. In particular, with the widespread adoption of digital technologies, music listening platforms have evolved, and a vast number of musical works have become accessible in digital environments. In addition, the use of digital tools and software in music production processes has increasingly expanded.

The large volume of music data available in digital environments has made processes such as analyzing and classifying these data, discovering new music, and developing recommendation systems based on listener preferences increasingly important. In this context, machine learning methods play a significant role in the processing and interpretation of music data. Especially on digital music platforms where Western music is predominantly represented, numerous studies in the literature focus on topics such as the identification of musical genres, the analysis of song popularity levels, and the automatic classification of musical works.

In this section, the structure and representation methods of music data are examined from the perspective of machine learning (ML). Furthermore, recent research and example studies on the application of machine learning techniques in music analysis are reviewed. In this regard, the section aims to provide guidance for both researchers and practitioners by presenting a theoretical conceptual framework as well as practical examples.

1 Asst. Prof. Dr. Nigar Tuğbagül Altan Gülgün, Istanbul Okan University, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, tugbagul.altan@okan.edu.tr, 0000-0002-1597-7231.

https://doi.org/10.58830/ozgur.pub1236.c5006

In conclusion, this section serves as a concise resource that provides both theoretical knowledge and practical guidance for researchers and practitioners aiming to achieve specific research objectives using music data. By demonstrating how machine learning methods can be effectively applied in music research and applications, it also establishes a fundamental reference framework for future studies.

1. Introduction

This chapter explores the application of machine learning techniques to music data. It begins with an overview of historical development of machine learning (ML) and what music is. In the subsequent sections, existing music datasets, the characteristics of music data, and several studies are discussed. The chapter then examines existing music datasets, the inherent characteristics of musical data, and selected studies from the literature that have applied machine learning methods to various music-related tasks.

Research on processing and analyzing music data using machine learning continues to advance rapidly, driven by the growing availability of digital music resources and computational tools. In this context, the chapter is intended to serve as a comprehensive guide for researchers and practitioners seeking to apply machine learning techniques in the field of music analysis.

1.1. Born of ML and ML

In the 1940s, Warren McCulloch and Walter Pitts presented some of the earliest foundational work on neural networks. Their research introduced key system components, including linear threshold decision elements, and provided a logical formalization for representing various forms of behavior and memory. In 1947, they further described a network architecture framework associated with geometric transformations, contributing to the conceptual development of computational models of neural activity (Minsky and Papert,1969).

Frank Rosenblatt, widely regarded as a pioneer of early perceptron research, formulated three fundamental questions concerning information processing in biological systems. These questions addressed: (1) how information about the physical world is sensed or detected by a biological system; (2) in what form this information is stored or remembered; and (3) how information retained in storage or memory influences recognition and behavior. These questions were presented at the Cornell Aeronautical Laboratory in 1958 (Rosenblatt, 1958). Rosenblatt also illustrated the organization of a perceptron, which is represented in Figure 1.

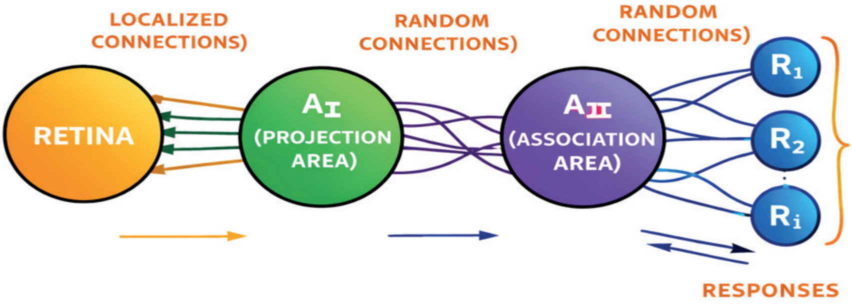

Figure 1. The organization of a typical photoperceptron

Figure 1 illustrates the organization of a typical photoperceptron—a perceptron designed to respond to optical patterns as stimuli—proposed by Frank Rosenblatt. In this model, Rosenblatt explained the interaction among units connected within a simple perceptron architecture.

In 1962, Rosenblatt further developed and demonstrated several theoretical foundations of the machines he referred to as perceptrons. However, the perceptron—considered the simplest form of a learning machine—was later critically analyzed by Marvin Minsky and Seymour Papert in 1969 (Minsky and Papert,1969). Their work emphasized the limitations of single-layer perceptrons and discussed aspects of parallel computation. They also presented a diagram illustrating their analysis, which is shown in Figure 2 to represent a fundamental concept of parallel computation.

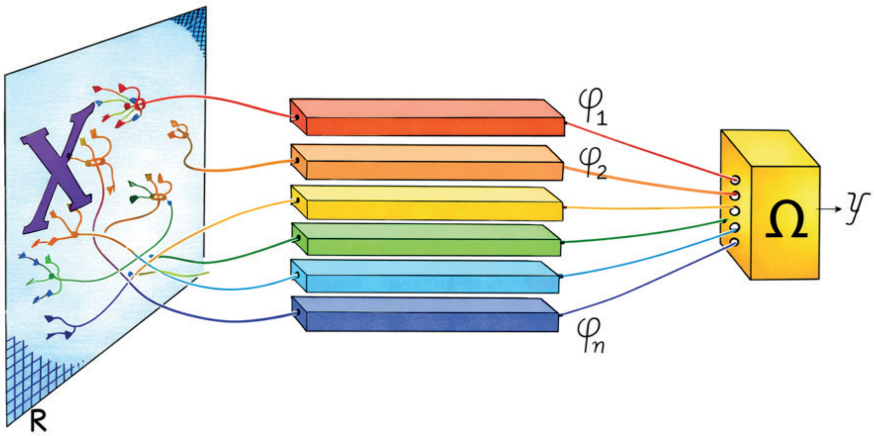

Figure 2. The simplest concept of parallel computation.

First, the set of functions $\varphi_1(X)$, $\varphi_2(X),\ldots,$ $\varphi_n(X)$ is computed independently. The resulting values are then combine through a function Ω, which takes n arguments, to produce the final value of ψ.

Between 1950 and 1980, the development of artificial intelligence shifted direction due to limitations in computational power and data storage capacity. During the 1990s and early 2000s, the field increasingly adopted data-driven modeling approaches grounded in strong mathematical principles, particularly statistical learning theory, convex optimization, and feature engineering. Between 2012 and 2017, deep learning emerged as a highly influential paradigm, replacing manual feature engineering with end-to-end learning architectures capable of automatically extracting hierarchical representations from data. Since 2017, many research institutions and organizations have focused on developing models based on self-supervised learning, scaling laws, and multimodal architectures.

Today, machine learning (ML) methods are generally categorized into several primary learning paradigms: supervised learning, unsupervised learning, semi-supervised learning, self-supervised learning, and reinforcement learning. In supervised learning, algorithms are trained using labeled datasets. In contrast, unsupervised learning aims to discover patterns or structures in data without labeled outputs. Semi-supervised learning combines both labeled and unlabeled data to improve model performance. Reinforcement learning enables systems to learn optimal behaviors through interaction with an environment and feedback in the form of rewards or penalties. Self-supervised learning leverages inherent structures within unlabeled data to generate supervisory signals for model training. The major categories of ML are illustrated in Figure 3.

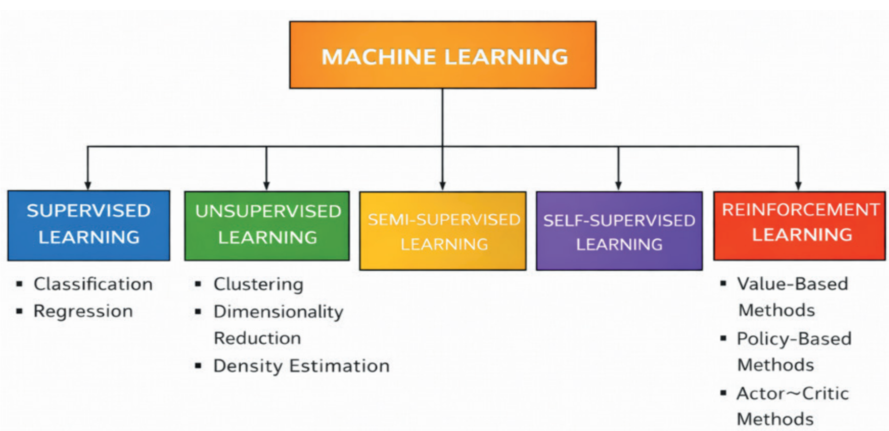

Figure 3. The major categories of ML

Figure 3 illustrates the general categories of machine learning (ML). Each category can be selected depending on the specific objectives and characteristics of the data. For example, supervised learning approaches are generally more suitable for labeled datasets such as tabular data, time-series data, images with annotations, text with labels, and audio with ground-truth information. In contrast, unsupervised learning methods are often preferred for analyzing high-dimensional tabular data, image and text embeddings, and sensor data, where explicit labels are not available.

In this chapter, we examine different types of music data that can be modeled and discuss various machine learning approaches used to process these data types across a range of applications and analytical objectives.

1.2. What is Music?

The ontological question "What is music?" has intrigued scholars across a wide range of disciplines for decades, including philosophy, musicology, sociology, cultural anthropology, history, archaeology, theology, mythography, biology, the humanities, cognitive science, neuroscience, and cultural studies, among others. According to widely recognized dictionaries—such as the Concise Oxford Dictionary, the Penguin Dictionary of Music, the Compact Oxford English Dictionary, and the Merriam-Webster Dictionary—the term music originates from the Greek word mousiké (μουσική), meaning "the art of the nine Muses." Music is commonly defined as the art or science of combining or arranging vocal or instrumental sounds to produce aesthetically pleasing compositions. Such compositions are typically characterized by beauty, form, harmony, melody, rhythm, unity, continuity, and emotional expression (Kokkidou, 2022).

Music is a highly multifunctional medium of communication and interaction, serving as a system through which individuals and communities engage with each other and with their environment. Across cultures, music fulfills a wide range of social, cultural, and emotional functions in both individual and collective contexts. These functions are accompanied by considerable variability in musical structures, practices, and phenomena, as well as by sets of rules, norms, and values that govern participation and shape music's role in human societies (Regelski 2020, Hargreaves et all. 2005). In essence, music serves as a vehicle for communication both among individuals and between individuals and society.

May Kokkidou conducted the Music Definition Project to explore the fundamental nature of music (Kokkidou, 2022). The study involved extensive collaboration with postgraduate music students, many of whom were practicing music teachers. The participant group comprised thirty-six women and twenty-one men, aged 18 to 49 years, with approximately 60% aged 30 or younger. This research prompted new and challenging questions, such as: "What is music for?" and "What does music do?" These questions highlight the importance of systematically investigating music to guide scientific research in the field.

Research on music spans a wide spectrum, from music perception and psychology to emotional responses to music and computational approaches using machine learning.

In this chapter, music data types for machine learning (ML), the key components of ML applied to music data, and numerous studies employing ML in music analysis are discussed. The aim is to provide foundational knowledge and practical guidance for researchers or practitioners who wish to learn or implement machine learning techniques for tasks such as classification, clustering, prediction, or achieving specific objectives using music data.

2. Music Data Types for ML

In 1965, at the age of 17, Ray Kurzweil developed one of the earliest computer programs capable of composing music in the style of classical composers. This program, written in Fortran, was demonstrated on the television show "What's My Secret?" and gained wider recognition following his appearance on the CBS program I've Got a Secret. Although Kurzweil's system did not employ neural networks, it is regarded as an early example of a symbolic AI system, relying on pattern-based rule systems for music composition.

In 1977, research on automated musical sound processing was conducted at the University of Michigan and published in the journal Automatic Music Transcription. The study utilized a low-pass filter at the Nyquist frequency and sampled musical signals at an effective rate of 4,000 samples per second to input music into a computer. The dataset consisted of flute recordings captured on reel-to-reel tape, which were subsequently converted into digital files using an analog-to-digital converter. Due to the memory limitations of computers at the time, each file was restricted to a duration of 60 seconds (Benetos et all., 2019).

2.1. Forms of Music Data

Music data can take several forms, including audio signals (e.g., waveforms, spectrograms), symbolic representations (e.g., MIDI files, sheet music), and metadata (e.g., genre, artist, tempo, mood, lyrics). Each type of music data requires distinct machine learning techniques tailored to its structure and characteristics.

2.1.1. Audio signals (waveforms, spectrograms)

The two most common methods for visualizing audio data are waveforms and spectrograms, both of which are widely used by machine learning models to analyze and interpret sound, including speech. Raw audio files are typically stored in formats such as WAV, MP3, or FLAC. In machine learning, raw audio is often represented as a time-series of amplitudes, with typical applications including genre classification and instrument recognition.

Audio representations in the time–frequency domain, such as Short-Time Fourier Transform (STFT), Mel-spectrograms, or Mel-Frequency Cepstral Coefficients (MFCCs), are displayed as 2D arrays (time × frequency). These representations are particularly useful for applications such as music transcription, emotion recognition, and other tasks that require analysis of both temporal and spectral characteristics of sound.

The piano-roll is a simplified representation that encodes only the pitches of musical notes and their temporal arrangement. It is typically presented as a two-dimensional graph, with pitch displayed along the vertical axis and time along the horizontal axis. Figure 4 illustrates three complementary representations of the opening of W. A. Mozart's *Piano Sonata No. 5, KV. 283, II. Andante*: a piano-roll (top), a spectrogram (middle), and a waveform (bottom).

The **piano-roll** provides a symbolic depiction of the music, showing the precise pitch and duration of each note, which facilitates analysis of melodic, harmonic, and rhythmic structures. In contrast, the **spectrogram** represents the frequency content of the sound over time, with intensity or color indicating amplitude, allowing detailed examination of harmonic content, timbre, and articulation. Finally, the **waveform** displays the raw audio signal, with amplitude plotted over time, making it particularly useful for observing overall dynamics and temporal variations in loudness (Pricop and Iftene, 2024).

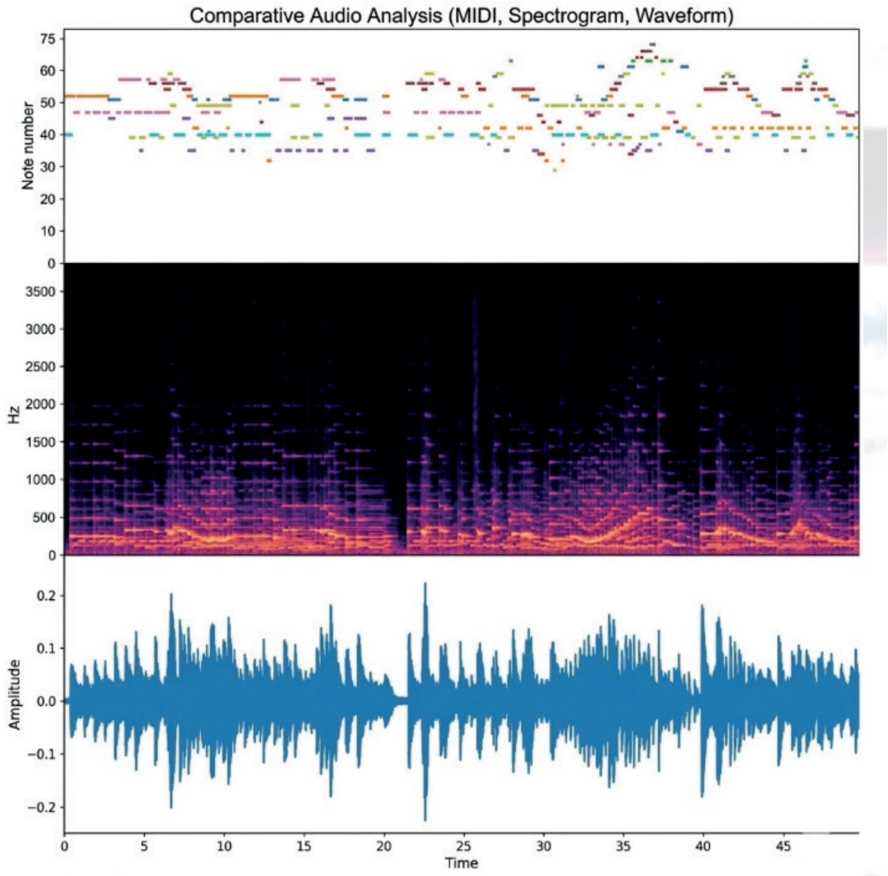

Figure 4. Three different representations of the beginning of W. A. Mozart Piano Sonata No. 5, KV. 283, II. Andante: piano-roll, spectrogram, waveform (from top to bottom, respectively) (Pricop and Iftene, 2024)

2.1.2. Symbolic data (MIDI, sheet music)

A collection of musical events, or notes, constitutes symbolic music data, which typically includes information about timing (onset, offset, duration), pitch, volume, and instrumentation. Various formats are used to store music files digitally, with one of the most widely recognized being the Standard MIDI (Musical Instrument Digital Interface) format. Introduced in August 1982, MIDI is a protocol designed to facilitate communication between keyboards and synthesizers, based on the Western musical interval system. In this system, the smallest interval, the semitone, corresponds to a single step in MIDI (Gedik, 2013; Plasser and Widmer,2023).

Standard MIDI Files are composed of discrete events, each consisting of two key components: a MIDI time, which specifies the timing of the event, and a MIDI message, which conveys the musical action (e.g., note on, note off, control change).

Standard MIDI Files contain events. Standard MIDI Files contain two components:

"a MIDI time, and a MIDI message".

The list of midi events: 00 90 40 40 00 90 43 40 81 00 80 43 00 00 90 45 40 81 00 80 45 00 00 80 40 00 00 90 3C 40 00 90 47 40 81 00 80 47 00 00 90 48 40 81 00 80 48 00 00 80 3C 40 for an example of a score in Figure 5 (Airy, S., Parr, J. M. 2001; de Oliveira, H., Oliveira, R., 2017; Heckroth, 1998).

Figure 5. An example of a score

MIDI files are often difficult for humans to read and edit. Additionally, bit-level errors can corrupt an entire file, and transferring files between different applications can be challenging. One of the primary applications of machine learning for music data is automatic music transcription, which is essential for processing and analyzing musical information.

In (de Oliveira and Oliveira, 2017), the authors present an automatic piano transcription system that converts polyphonic audio recordings into musical scores using a machine learning model. The system transcribes audio recordings (WAV files) from the MIDI Aligned Piano Sounds (MAPS) database into standard music notation. For this purpose, the corresponding MusicXML files are used to represent the musical scores.

The Automatic Music Transcription Pipeline described in the study is illustrated in Figure 6. This work is a seminal study demonstrating how audio signals can be converted into symbolic music data. The transcription process begins with multipitch detection in MIDI format, followed by MIDI quantization, which converts the detected pitches into a structured musical score representation (Benetos et al., 2019).

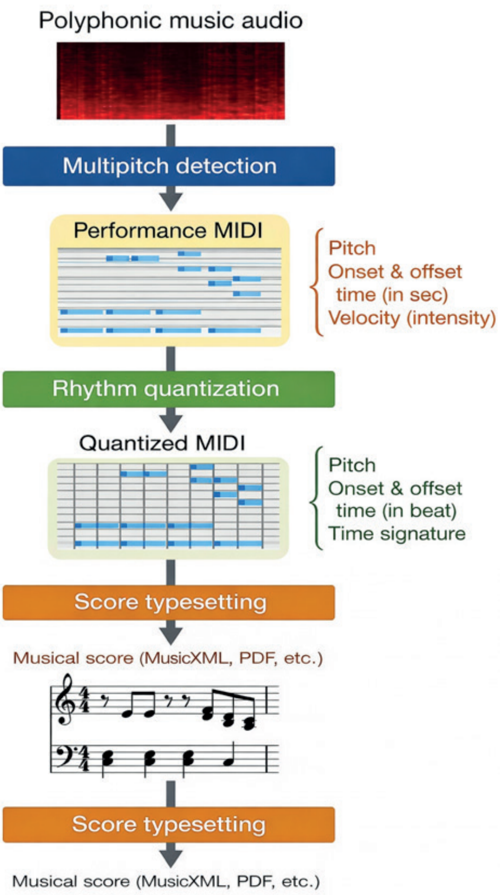

Figure 6. The Automatic Music Transcription Pipeline

MusicXML is an open, flexible, and platform-independent XML-based file format with a hierarchical, tag-based structure that is also human-readable. It is specifically designed for representing symbolic Western music notation. A MusicXML file can store a wide range of musical information, including notes (pitch, duration), rests, chords, measures and bar lines, clefs, key and time signatures, dynamics, articulations, slurs, ties, as well as structural elements such as voices, instrument definitions, tempo markings, repeats, and codas.

Importantly, MusicXML encodes musical score information rather than audio, meaning it specifies what to play rather than the actual sound waveform. Other examples of XML-based music markup languages include ChordML, MCML (Music Contents Markup Language), EMNML (Extensible Music Notation Markup Language), MML (Music Markup Language), MNML (Music Notation Markup Language), MEI (The Music Encoding Initiative), MX-IEEE 1599, WEDELMUSIC [13] and Lakh MIDI Dataset (Raffel, 2016).

MusicXML is widely supported by various applications, including sheet music software, music editors, educational programs, and music databases. An example of the MusicXML format is illustrated in Figure 7. The MusicXML format is easy to understand and can be readily converted or translated into other music notation formats.

```
<?xml version="1.0" encoding="UTF-8"?>
<score-partwise version="3.1">
  <part-list>
    <score-part id="P1">
      <part-name>Music</part-name>
    </score-part>
  </part-list>

  <part id="P1">
    <measure number="1">
      <attributes>
        <divisions>1</divisions>
        <key><fifths>0</fifths></key>
        <time><beats>4</beats><beat-type>4</beat-type></time>
        <clef><sign>G</sign><line>2</line></clef>
      </attributes>

      <note>
        <pitch><step>C</step><octave>4</octave></pitch>
        <duration>4</duration>
        <type>whole</type>
      </note>

    </measure>
  </part>
</score-partwise>
```

Figure 7. Example of MusicXML format

2.1.3. Metadata (Genre, artist, tempo, mood, lyrics)

Lyrics is textual format and most ML application with it are listed as Lyrics is Lyrics is textual format, genre classification, hit song prediction, lyric generation, identifying common topics in songs. While music metadata type with genre, artist and/ or tempo formats is represented numerical/ sequential, user interaction data type with play counts, ratings is numerical/ sequential. On the other hand, most application in Ml for both is recommendation systems. MIDI (Musical Instrument Digital Interface) is a technical standard that enables electronic musical instruments, computers, and other digital devices to communicate and synchronize with each other. Importantly, MIDI does not transmit audio; instead, it conveys digital instructions specifying how music should be played.

In Music Information Retrieval (MIR) and computational musicology, "music data" exists at multiple levels, ranging from raw acoustic signals to symbolic representations and user-context metadata. The optimal choice of music data depends on the research objectives.

For example:

Playlist recommendation often uses the AotM-2011 dataset (McFee and Lanckriet, 2012).

Singing voice separation employs several datasets:

MIR-1K, designed specifically for singing voice separation, contains 1,000 song clips and has been widely used in multiple studies (Hsu et al., 2011; Hsu et al., 2015).

ccMixter (Liutkus and Rafii, 2017), a community-driven collection of shared and open music.

MUSDB18, which includes 150 full-length tracks (~10 hours) across multiple genres, along with isolated stems for drums, bass, vocals, and other instruments. MUSDB18 is organized into a training set ("train") of 100 songs and a test set ("test") of 50 songs.

In (Prabhakar and Lee, 2023), three datasets—GTZAN (Tzanetakis and Cook, 2002), ISMIR, and MagnaTagATune (collected via the TagATune game and music from the Magnatune label—were evaluated using a combination of Bidirectional Long Short-Term Memory (BLSTM) networks and a Graph Convolutional Network, achieving an accuracy of 93.51%.

A substantial body of research in the literature focuses on machine learning applications using digital music platforms, which predominantly feature

Western music. Among these, the GTZAN Genre Dataset is one of the most widely used. It consists of 1,000 audio recordings evenly distributed across 10 music genres, with each recording having a duration of 30 seconds. Due to its structured format and balanced genre distribution, the GTZAN dataset has been extensively employed for tasks such as music genre classification, audio feature extraction, and algorithm benchmarking. As a result, a large number of studies applying various classification algorithms to the GTZAN dataset are available in the literature.

In recent years, the widely used music streaming platform Spotify has made portions of its music-related data publicly accessible, facilitating academic research. Similar to many large-scale music datasets, the Spotify dataset predominantly features Western music and does not include Turkish-language compositions. The dataset contains 42,305 songs across 15 genres, including Trap, Techno, Techhouse, Trance, Psytrance, Dark Trap, Drum and Bass (DnB), Hardstyle, Underground Rap, Trap Metal, Emo, Rap, RnB, Pop, and Hip-hop.

This dataset has been employed in numerous studies for classification and prediction tasks. For example, some research has focused on predicting the popularity scores of songs based on musical features and metadata (Rosenblatt, 1958). Additionally, several studies have explored genre classification and feature-based clustering using this dataset (Kokkidou, 2022; Regelski, BBB; Hargreaves, 2005; Benetos et al., 2019; Gedik, 2013; Plasser and Widmer, G., 2023; Shibata et al., 2021). Despite the abundance of research conducted using Western music datasets, studies focusing on the makam classification of Turkish classical music remain limited in the literature. One of the major resources containing Turkish music is the Dunya CompMusic Dataset, developed within the framework of the CompMusic Project. This dataset contains audio recordings, musical scores, metadata, and annotations specifically prepared for computational analysis. It includes approximately 6,500 audio recordings corresponding to around 412 hours of music. In addition, the dataset contains 2,200 musical scores accompanied by detailed metadata, including the title of the piece, makam information, and instrument data (McFee and Lanckriet, 2012).

The Dunya CompMusic platform is not merely an archive for listening; it is a structured academic resource developed to support research in Music Information Retrieval (MIR) and computational musicology. Within the CompMusic/Dunya ecosystem, several specialized sub-datasets have been created to address different research objectives.

One of the most notable of these is the SymbTr Turkish Makam Music Symbolic Dataset, a symbolic dataset specifically designed for the analysis of Turkish makam music. The dataset comprises 2,200 musical pieces, representing approximately 150 different makams, nearly 100 rhythmic patterns (usul), and around 50 musical forms. In total, it contains approximately 865,000 musical notes, corresponding to roughly 80 hours of nominal playback duration (Hsu, 2011).

The dataset provides musical data in multiple formats, including Text, MusicXML, PDF, MIDI, and MU2, enabling diverse analytical approaches and computational experiments. Detailed information regarding the dataset's characteristics and structure is available on the official dataset website.

2.2. ML in Music

A search in Scopus using the keywords "Machine Learning" and "Music", limited to Computer Science and restricted to articles, abstracts, and keywords from 2015 to 2026, yielded 3,025 documents. A similar search with the keywords "Deep," "Learning," and "Music" returned 3,451 documents, while a search using "Machine," "Learning," "Music" resulted in 3,396 documents. These numbers indicate that interest in the application of machine learning to music has increased steadily over the years.

The question "Can machines think?" (Koul, 2019) is widely considered one of the fundamental questions that led to the emergence of Artificial Intelligence (AI) as a field of study. The capability of a machine to imitate intelligent human behavior is referred to as **Artificial Intelligence (AI)**. AI encompasses computer systems or sets of algorithms designed to perform tasks that typically require human intelligence, such as learning from data, recognizing speech or images, making decisions, solving problems, and understanding natural language (Samuel, 1959; Mycka and Mańdziuk, 2025).).

Machine Learning (ML) is a subfield of artificial intelligence that enables computers to learn patterns from data and improve their performance without being explicitly programmed for every task. Machine learning algorithms analyze data, identify underlying patterns, and use these patterns to make predictions or decisions.

Deep Learning (DL) is a specialized subfield of machine learning and representation learning that relies on artificial neural networks with multiple processing layers. These networks learn hierarchical representations of data, allowing models to capture complex patterns and multiple levels of abstraction (Samuel, 1959).

Figure 8 represents the relationship between AI, ML, and DL.

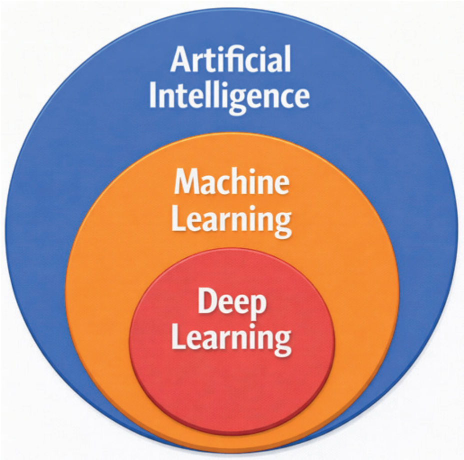

Figure 8. The Relationship between AI, ML and DL

From Figure 8, Machine learning is a subset of artificial intelligence (AI), while deep learning—which includes techniques such as Natural Language Processing (NLP), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs)—is a specialized subset of machine learning. Both machine learning and deep learning can be broadly categorized into two main types: supervised learning, which relies on labeled data, and unsupervised learning, which identifies patterns from unlabeled data.

2.3. Basic Steps in Machine Learning

The basic steps involved in developing a machine learning solution are illustrated below:

Problem Definition: Clearly define the task to be addressed and determine the most suitable type of machine learning approach, such as supervised, unsupervised, or reinforcement learning.

Data Collection: Gather relevant data from reliable sources. In music-related applications, this may include audio files, symbolic representations (e.g., MIDI), and associated metadata.

Data Preprocessing – Prepare the dataset by handling missing values, normalizing or standardizing features, and partitioning the data into training, validation, and test sets.

Feature Extraction and Engineering – Convert raw data into informative features suitable for machine learning algorithms. For instance, Mel-Frequency Cepstral Coefficients (MFCCs) are commonly employed for audio analysis.

Model Selection – Choose an appropriate machine learning algorithm based on the problem and dataset characteristics. Typical options include Support Vector Machines (SVM), Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN).

Model Training – Train the selected model using the training dataset and optimize relevant hyperparameters to enhance predictive performance.

Model Evaluation – Evaluate the trained model on unseen data using appropriate performance metrics, such as accuracy, F1-score, or Root Mean Square Error (RMSE), and refine the model as necessary.

Deployment – Integrate the validated model into real-world applications, such as music recommendation systems, emotion recognition tools, or genre classification platforms.

Monitoring and Maintenance – Continuously monitor model performance in the operational environment and retrain or update the model as new data becomes available to ensure sustained effectiveness.

Figure 9 illustrates the fundamental steps, along with brief descriptions, that must be considered when developing a machine learning solution.

Figure 9. Basic steps in ML

2.4. Music Data Types for ML

Music Information Retrieval (MIR) is a research and technological field dedicated to analyzing, understanding, and extracting information from music data using a variety of techniques, including signal processing, machine learning, and data science. Within MIR, common tasks include genre classification, beat and tempo tracking, beat detection, key and chord recognition, instrument recognition, and music similarity, clustering, etc. Figure 10 provides a summary of the various types of music data utilized in machine learning, along with the machine learning techniques most commonly applied to each type.

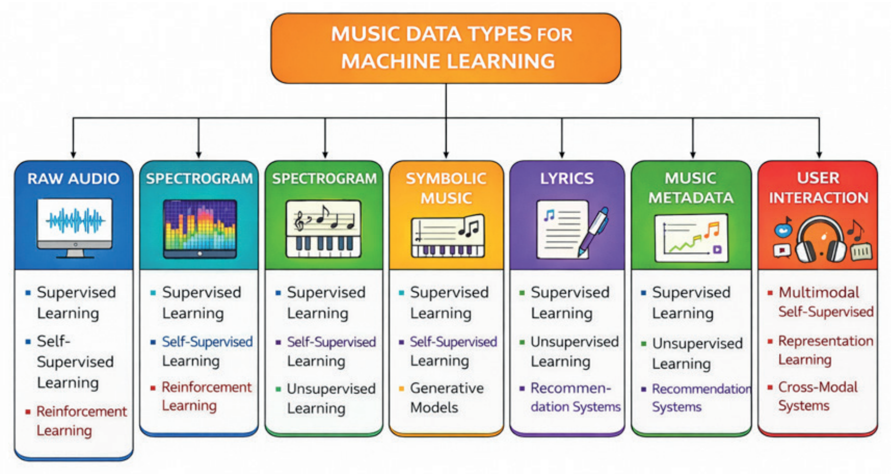

Figure 10. Music Data Types for ML

2.5. Key Applications of ML in Music

In the music domain, machine learning (ML) methods are frequently applied to music data to achieve specific objectives. ML supports a wide range of applications, which can be structured into a taxonomy of music data types and corresponding tasks. Key applications include:

- Music classification for listener recommendations or predicting user preferences for new music.
- lyric analysis, automatic music generation, including composition of original music or emotion-based music generation aimed at evoking specific emotional responses.
- Genre categorization and style modeling, as well as mood detection.
- Automatic music transcription, converting audio recordings into symbolic music notation.
- Score following, which tracks a live performance in relation to a written score.

These objectives can involve partial or fully integrated ML applications, depending on the complexity of the task.

There are several motivations for researchers to study music classification. For instance, songs can be categorized based on the emotional responses they evoke using emotion-based algorithms, or organized to help listeners discover similar music that matches their preferences. In other words, classification enables the identification of potential new favorite songs in a time-efficient manner.

The literature includes numerous studies on music classification, each pursuing different objectives and employing a variety of dataset types.

Machin

77e learning has had a profound impact on music research, providing innovative approaches to music generation, style modeling, and genre classification. Advances in machine learning models have made it possible to create music that emulates human compositional techniques, opening new opportunities for creativity and artistic innovation (Tang et al., 2018).

Table 1 summarizes selected studies from the literature, including the machine learning approaches employed, their research objectives, and their applications to various types of music data.

Table 1. Several representative studies from the literature, highlighting the types of ML, used, their objectives.

The Used Keywords of Related Researches	**Objectives**	**References**
Convolutional Neural Networks (CNN) Support Vector Machines (SVMs)	Categorize music into genres, moods, or other attributes based on audio features	Purwins et al., 2019; Puig, 2019; Schedl, 2019; Briot et al., 2017; Costa et al., 2017; Senac et al., 2022; Lu et al., 2022; Martín-Gutiérrez et al., 2022
Deep Learning (Music Deep Learning)	Automatic generator	He and Zhan, 2025
Deep Learning (Music Deep Learning)	Singing voice detection	Monir et al., 2022
Deep Learning (Music Deep Learning)	Instrument recognition	Choi et al., 2017
Deep Learning (Music Deep Learning)	Music emotion recognition, classification	Han et al., 2022; Zhao et al., 2018
Deep Learning (Music Deep Lear+ning)	Transcription	Sturm, 2016

Deep Learning (DL) - Convolutional Neural Network (CNN)	The Live emotion based music recommendation	Pooja et al., 2026
Deep Learning (DL)	Artists classification; Music classification	Mycka and Mańdziuk,2026
Deep Belief Networks (DBNs) Recurrent Neural Networks (RNNs) Variational Autoencoders (VAEs) Long Short-Term Memory (LSTM) Networks and Transformers	Music transcription	Rafii et al., 2018; Monir et al., 2022; Choi et al., 2017; Han et al., 2022; Sturm et al., 2016; Casini and Roccetti,20182018
Deep Neural Networks (DNNs) Machine Learning (ML) Long Short-Term Memory (LSTM)	Advanced music generation	Pricop and Iftene, 2024
Long Short-Term Memory (LSTM) Recurrent neural network (RNN) Markov Chains	Automatic music generation	Cretu et al., 2022
Deep Learning (Music Deep Learning) Long Short-Term Memory (LSTM)	Music generation	Chou and Peng, 2019
Support Vector Machine (SVM) Random Forest (RF) Artificial Neural Network (ANN)	Music classification	Dai and Huang, 2022
Artificial Neural Network (ANN) Convolutional Neural Network (CNN)	Classify the music based on certain instruments	Shreevathsa et al., 2020
Machine Learning (ML)	Automatic evaluation system for piano music performance	Xiao, 2026
Machine Learning (ML)	Cultural music classification	Abebe et al., 2026
Machine Learning (ML)	The Music Emotion Recognition (MER)	Louro et al., 2026
Unsupervised Disentanglement Strategies	Controllable Music Generation	Ibáñez-Martínez et al., 2026

3. Conclusion

In recent years, machine learning techniques have increasingly shaped the analysis and processing of musical data. The rapid expansion of digital music platforms and the availability of large-scale datasets have enabled researchers to employ advanced computational methods across a wide range of music-related tasks. In particular, machine learning has been extensively applied in areas such as Music Information Retrieval (MIR), classification of artists, genres and moods, recommendation systems, automatic music generation and music emotion recognition, etc.

By leveraging large-scale musical datasets and sophisticated computational models, machine learning methods facilitate the systematic analysis and interpretation of musical information. These approaches enable the extraction of meaningful features from audio signals, symbolic music representations, and associated metadata. Such capabilities support diverse analytical tasks, including music genre recognition, mood detection, popularity prediction, and automatic transcription.

Accordingly, machine learning has become a critical tool for both academic research and practical applications in music analysis and technology. Its adoption has contributed to the development of more efficient music analysis systems and intelligent digital music services, advancing the capabilities of computational approaches in understanding and interacting with music.

Research on the application of machine learning techniques to music data continues to advance rapidly. It is hoped that this chapter will serve as a useful resource for researchers seeking to pursue studies in this area.

References

Abebe, M., Endashew, L., Heikkonen, J., Kanth, R., & Mohapatra, S. K. (2026). Decoding cultural music classification with machine learning and segment length analysis. *SN Computer Science*. https://doi.org/10.1007/s42979-026-04747-6 Benetos, E., Dixon, S., Duan, Z., & Ewert, S. (2019). Automatic music transcription: An overview. *IEEE Signal Processing Magazine, 36*(1), 20–30. https://doi.org/10.1109/MSP.2018.2869928

Airy, S., & Parr, J. M. (2001). MIDI, music and me: Students' perspectives on composing with MIDI. *Music Education Research, 3*(1), 41–49. https://doi.org/10.1080/14613800020029941

Bishop, G. (n.d.). *Music datasets for machine learning*. *Medium*. Retrieved March 8, 2026, from https://gail-bishop.medium.com/music-datasets-for-machine-learning-a6cd8d707340

Briot, J.-P., Hadjeres, G., & Pachet, F.-D. (2017). Deep learning techniques for music generation—A survey. *arXiv*. https://arxiv.org/abs/1709.01620

Casini, L., Marfia, G., & Roccetti, M. (2018). Some reflections on the potential and limitations of deep learning for automated music generation. In *IEEE 29th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)* (pp. 27–31).

Choi, K., Fazekas, G., Cho, K., & Sandler, M. (2017). A tutorial on deep learning for music information retrieval. *arXiv*. https://arxiv.org/abs/1709.04396

Costa, Y. M. G., Oliveira, L. S., & Silla, C. N. (2017). An evaluation of convolutional neural networks for music classification using spectrograms. *Applied Soft Computing, 52*, 28–38.

Cretu, B.-A., Vranceanu, A., Cojocariu, A., Simionescu, C., & Iftene, A. (2022, August 8–10). *Music generation using neural nets*. In *Proceedings of the International Conference on Innovations in Intelligent Systems and Applications (IEEE INISTA 2022)* (pp. 1–6). Biarritz, France: IEEE.

Dai, Z., & Huang, X. (2022). Electronic dance music classification based on machine learning methods. In *2022 International Conference on Electronics and Devices, Computational Science (ICEDCS)* (pp. 351–354). IEEE. https://doi.org/10.1109/ICEDCS57360.2022.00084

de Oliveira, H., & Oliveira, R. (2017). Understanding MIDI: A painless tutorial on MIDI format. *arXiv*. https://doi.org/10.48550/arXiv.1705.05322

Essinger, S. D., & Rosen, G. L. (2021). *An introduction to machine learning for students in secondary education*. Drexel University, Department of Electrical & Computer Engineering. https://exampleurl.edu

Gedik, A. (2013). Türk müziği icralarının analizi için bir MIDI projesi. *Yedi, 10*, 81–92. https://izlik.org/JA45NM54LS

Gray, B., Geis, R., O'Connell, T., Babyn, P., Koff, D., & Shabana, W., et al. (2018). Canadian Association of Radiologists white paper on artificial intelligence in radiology. *Canadian Association of Radiologists Journal, 69*, 120–135. https://doi.org/10.1016/j.carj.2018.02.002

Han, D., Kong, Y., Han, J., & Wang, G. (2022). A survey of music emotion recognition. *Frontiers of Computer Science, 16*(6), 1–11.

Hargreaves, D. J., MacDonald, R., & Miell, D. (2005). How do people communicate using music? In D. Miell, R. MacDonald, & D. J. Hargreaves (Eds.), *Musical communication* (pp. 1–26). Oxford University Press.

He, Y., & Zhan, Y. (2025). Application of machine learning in automatic generation of intangible cultural heritage music scores. In *Proceedings of the International Conference on Artificial Intelligence and Computer Engineering* (pp. 2502–2505). IEEE. https://doi.org/10.1109/ICAACE65325.2025.11019496

Heckroth, J. (1998). *A tutorial on MIDI and wavetable music synthesis* (Application Note AN27). Crystal Semiconductor Products Division, Cirrus Logic, Inc.

Honi, P. (2023). *Exploring the synergy between generative AI, data and analytics in the modern age*. TechRxiv. https://doi.org/10.36227/techrxiv.24045792

Hsu, C.-L., Jang, J. S. R., & Wang, D. (2011). A trend estimation algorithm for singing pitch detection in musical recordings. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Hsu, C.-L., Jang, J. S. R., & Wang, D. (2010). On the improvement of singing voice separation for monaural recordings using the MIR-1K dataset. *IEEE Transactions on Audio, Speech, and Language Processing, 18*(2), 310–319.

Hsu, C.-L., Wang, D., Jang, J. S. R., & Hu, K. (2011). A tandem algorithm for singing pitch extraction and voice separation from music accompaniment. *IEEE Transactions on Audio, Speech, and Language Processing*.

Ibáñez-Martínez, L., Nkama, C., Poltronieri, A., Serra, X., & Rocamora, M. (n.d.). *Evaluating disentangled representations for controllable music generation*. Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain.

Kokkidou, M. (2022). *Music definition and music education: Many perspectives, many voices, many questions.*

Koul, A. (2019). What is AI? In *Practical deep learning for cloud, mobile, and edge: Real-world AI and computer vision projects using Python, Keras, and TensorFlow* (pp. 6–7). O'Reilly Media.

Liutkus, A., Fitzgerald, D., & Rafii, Z. (2015). Scalable audio separation with light kernel additive modelling. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 76–80).

Louro, P. L., et al. (2026). MERGE: A bimodal audio-lyrics dataset for static music emotion recognition. *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/TAFFC.2026.3668037

Merriam, A. P. (1964). *The anthropology of music*. Northwestern University Press.

Merriam-Webster. (n.d.). *Artificial intelligence*. In *Merriam-Webster.com dictionary*. Retrieved March 7, 2026, from https://www.merriam-webster.com/dictionary/artificial%20intelligence

Martín-Gutiérrez, D., Penaloza, G. H., Belmonte-Hernandez, A., & Garcia, F. A. (2020). A multimodal end-to-end deep learning architecture for music popularity prediction. *IEEE Access, 8*, 39361–39374. https://doi.org/10.1109/ACCESS.2020.2979145

McFee, B., & Lanckriet, G. R. (2012). Hypergraph models of playlist dialects. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 343–348).

Minsky, M., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. MIT Press.

Monir, R., Kostrzewa, D., & Mrozek, D. (2022). Singing voice detection: A survey. *Entropy, 24*(1), 114.

Mycka, J., & Mańdziuk, J. (2025). Artificial intelligence in music: Recent trends and challenges. *Neural Computing and Applications, 37*, 801–839. https://doi.org/10.1007/s00521-024-10555-x

Mycka, J., & Mańdziuk, J. (2026). Classification of piano performers with deep learning models. *Journal of Computational Science*. https://doi.org/10.1016/j.jocs.2026.102804

Plasser, M., Peter, S., & Widmer, G. (2023). Discrete diffusion probabilistic models for symbolic music generation. *arXiv*. https://doi.org/10.48550/arXiv.2305.09489

Pooja, M., Kavyashree, H. L., Pratham M., & Meghana, G. (2026). Live emotion based music recommender system. In *Proceedings of the 2026 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)* (pp. 1–6). IEEE. https://doi.org/10.1109/IITCEE67948.2026.11394620

Pricop, T.-C., & Iftene, A. (2024). Music generation with machine learning and deep neural networks. *Procedia Computer Science, 246*, 1855–1864. https://doi.org/10.1016/j.procs.2024.09.692

Puig, J. P. (2019). *Deep neural networks for music and audio tagging* (Doctoral dissertation, Universitat Pompeu Fabra, Barcelona, Spain).

Raffel, C. (2016). *Learning-based methods for comparing sequences, with applications to audio-to-MIDI alignment and matching* (Doctoral dissertation). Columbia University.

Rafii, Z., Liutkus, A., Stöter, F.-R., Mimilakis, S. I., FitzGerald, D., & Pardo, B. (2018). An overview of lead and accompaniment separation in music. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26*(8), 1307–1335.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65*(6), 386–408. https://doi.org/10.1037/h0042519

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development, 3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Schedl, M. (2019). Deep learning in music recommendation systems. *Frontiers in Applied Mathematics and Statistics, 5*, 44.

Shreevathsa, P. K., Harshith, M., R. M. A., & Ashwini. (2020). Music instrument recognition using machine learning algorithms. In *International Conference on Computation, Automation and Knowledge Management (ICCAKM)* (pp. 161–166). IEEE.

Sturm, B. L., Santos, J. F., Ben-Tal, O., & Korshunova, I. (2016). *Music transcription modelling and composition using deep learning*. arXiv. https://arxiv.org/abs/1604.08723

Tang, A., Tam, R., Cadrin-Chênevert, A., Guest, W., Chong, J., Barfett, J., Chepelev, L., Cairns, R., Mitchell, J., Cicero, M., Gaudreau Poudrette, M., Jaremko, J., Reinhold, C., Gallix, B., Gray, B., Geis, R., O'Connell, T., Babyn, P., Koff, D., & Shabana, W. (2018). Canadian Association of Radiologists white paper on artificial intelligence in radiology. *Canadian Association of Radiologists Journal, 69*, 120–135. https://doi.org/10.1016/j.carj.2018.02.002

The MagnaTagATune Dataset. (n.d.). *MagnaTagATune dataset*. Retrieved March 6, 2026, from https://mirg.city.ac.uk/codeapps/the-magnatagatune-dataset

Tudor-Constantin Pricop, & Iftene, A. (2024). Music generation with machine learning and deep neural networks. *Procedia Computer Science, 246*, 1855–1864. https://doi.org/10.1016/j.procs.2024.09.692

Tzanetakis, G., & Cook, P. (2002). *GTZAN genre dataset*. Retrieved March 8, 2026, from http://marsyas.info/downloads/datasets.html

URL-1: https://gail-bishop.medium.com/music-datasets-for-machine-learning-a6cd8d707340 (Date Accessed: 22 March 2026).

URL-2: https://zenodo.org/records/3532216, (Date Accessed: 02 Mart 2026).

URL-3: https://ismir2004.ismir.net, (Date Accessed: 06 March 2026).

URL-4:http://marsyas.info/downloads/datasets.html, (Date Accessed: 06 March 2026).

URL-5: https://www.merriam (Date Accessed: 06 March 2026).

Xiao, X. (2026). Design of an automatic evaluation system for piano music performance based on machine learning algorithms. In *Proceedings of the 2025 International Conference on Artificial Intelligence, Virtual Reality and Interaction Design* (pp. 583–587). Association for Computing Machinery. https://doi.org/10.1145/3777730.3777825.