International Journal of Scientific Research in Computer Science and Engineering
Research Paper
Vol.10, Issue.5, pp.37-44, October (2022)
E-ISSN: 2320-7639

A Convolutional Neural Networks (CNN) Approach to Music Genre Classification

Abdulsalam Auwal Jamilu[1*], Lawal Abubakar[2], Ubaidullah Abdallah [3]

[1]Dept. Computer Science, Ahmadu Bello University, Zaria, Nigeria
[2]Dept. Physics, Bayero University Kano, Nigeria
[3]Dept. *ICT*, F M C K, Nigeria

[*]*Corresponding Author: abdulsalamjauwal@fmckeffi.gov.ng, Tel.: +234-07036501948*

Available online at: www.isroset.org

Received: 25/Aug/2022, Accepted: 28/Sept/2022, Online: 31/Oct/2022

Abstract— Music is becoming easier to access through the internet, and musical applications like Spotify and Apple Music have common services that help their customers automatically classify music into different genres, the classification of music genres is a fundamental step in developing a powerful music recommendation engine.With the escalating number of music available digitally on the internet, there is a growing demand for the systematic organization of audio files and thus a rise in the interest in automatic music genre classification. Moreover, detecting and grouping music in a similar genre is a keen part of the music recommendation system and playlist that are personalized to soothe listeners' mood and their unique music taste. However, Convolutional Neural Networks have appeared to be accurate in classifying music into different genres. Over the last decade, Convolutional Neural Networks have achieved breakthroughs in domains ranging from pattern recognition, image processing and voice recognition. For the Convolution Neural Network model to be able to classify music into different genres there would be a need for pre-processing of data by converting the raw audio into Mel-spectrograms. These features that have been extracted would then be used for training and classification. Additionally, Mel-spectrograms are visual, and CNN works better with images.
This research focuses on a review, in the identification of music genres. Music Information Retrieval (MIR) can make it easier to identify essential information like trends, popular genres, and performers.

***Keywords*—** Convolution Neural Network, Deep learning, Classification, Music genre classification.

I. INTRODUCTION

The classification of music genres is a critical undertaking with several real-world applications; As the amount of music released daily continues to increase, particularly on Streaming services like Spotify and YouTube [1]. In 2016, Spotify added over 500,000 songs (567,693), and in 2017, it added over a million more (934,265) to its music catalogue [2]. As a result, any device with an Internet connection can access the over 35 million music that are available to Spotify subscribers [3]. Hence the need for accurate meta-data for database management and search/storage purposes rises concurrently [1]. The ability to instantly identify songs in any given playlist or album by genre is a critical feature for any music streaming service, and the statistical analysis potential provided by accurate and full labeling of music and audio is seemingly unlimited [4]. Figure 1 below describe a music recommendation system utilizing artificial intelligence, based on listening history, listener emotion, and duration of the day.

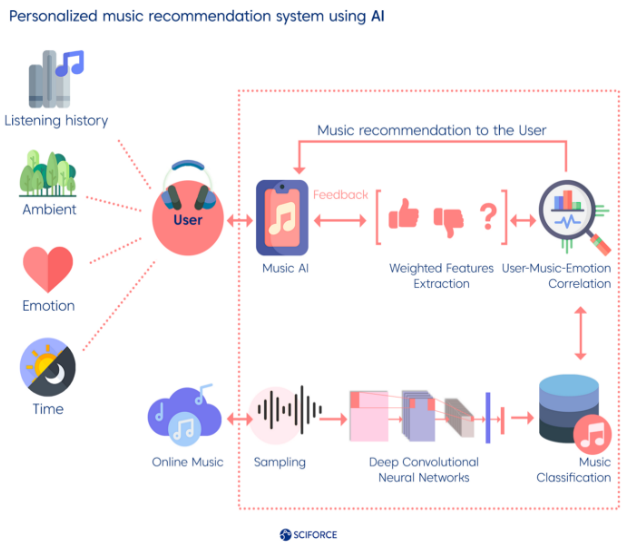

Figure 1. Personalized music recommendation system using Artificial Intelligence [5]

This research focuses on a review, in the identification of music genres. Music Information Retrieval (MIR) can make it easier to identify essential information like trends, popular genres, and performers. Yu et al went on to extract each song from both datasets into 18 smaller pieces in 3 seconds with a 50% overlaps [3]. Among the most popular musical genres, its members frequently have things in common that are related to the instrumentation, rhythmic pattern, and pitch content of the music [4]; The need to organize and organize the ever-growing amount of music files that are available digitally on the internet is driving the importance of automatically extracting music characteristics [4]. One of the services that music content distribution companies will offer to attract in customers is automatic music analysis [4]. Creators and consumers interact and respond to one another, creating momentum, with a music collection that expands by tens of thousands of new creative works each day.

Along with the music, lyrics have also been used to categorize music into genres and subgenres [6]. One technique to currently catalogue music content on the internet is through genre classes, which are often manually developed by human professionals [4]. The process of defining a music genre is quite subjective; The culture of the listener’s has a significant impact, due to the fact that rock, pop, and disco are far more popular and actually originated in Europe, North America, and South America, respectively, these people have a much greater understanding of these genres of music; As a result, these individuals can identify certain genres considerably more quickly [6]. Automatically classifying musical genres has the ability to automate this procedure and serve as a vital piece of a comprehensive system for retrieving music information from audio signals; Additionally, it offers a framework for designing and analyzing features for characterising musical material, Most suggested audio analysis techniques for music use these features as the basis for similarity retrieval, classification, segmentation, and audio thumb nailing [4].

On the other hand, research on machine audition, or computer systems that understand auditory input, has thus far primarily concentrated on spoken language recognition; However, one of the prerequisites for an intelligent system is having the capacity to recognize a variety of events in a certain environment; A crucial component of an intelligent system that functions in the actual world is the ability to comprehend varied acoustic information in addition to verbal and visual information [7].

In this study, a novelty solution to the automatic music genre classification problem is presented. The proposed approach uses Convolutional Neural Networks (CNNs). Due to recent outstanding achievements in fields including image processing, speech recognition, and artificial intelligence/human-machine interface, convolutional neural networks (CNNs) have drawn a lot of attention (e.g., Go playing); Originally developed for image processing, CNNs have recently produced convincing outcomes for both audio processing and music information retrieval (MIR) [8].

Convolutional Neural Networks (CNNs) are like conventional Artificial Neural Networks (ANNs) in that they are made up of neurons that learn to optimize themselves; Each neuron will continue to receive an input and carry out an operation (such as a scalar product followed by a non-linear function), which is the foundation for ANNs; The network as a whole will continue to express a single perceptual score function from the input raw image vectors to the final output of the class score (the weight). The loss functions linked with the classes will be in the final layer, and all of the standard techniques created for conventional ANNs are still applicable; The fact that CNNs are primarily employed in the field of pattern detection within images is the only distinguishing feature between CNNs and conventional ANNs [9].

Additionally, deep learning approaches like convolutional neural networks typically do not have to be supplied with hand-crafted features, as they are able to "learn" the features from the data; A Deep learning model can identify and extract useful features for a specific classification task; In fact, deep learning models that perform feature extraction and classification outperform models that classify manually extracted features by a large margin [10].

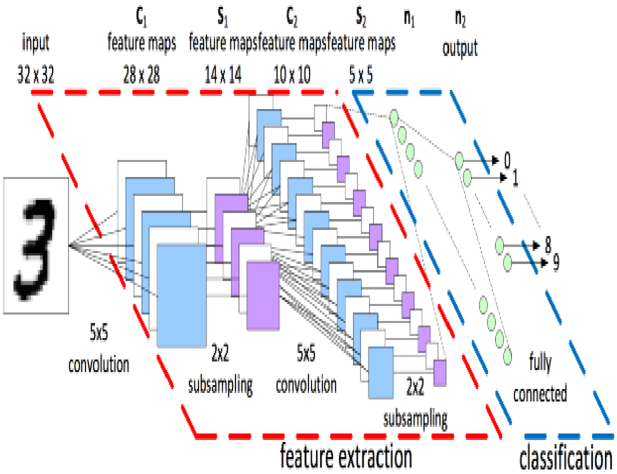

Figure 2. CNNs architecture for music genre classification [11]

Any realistic music classification tool must include novelty detection since some songs may not fit into any of the classes that the system supports; In these situations, it may make more sense to designate the kind of the song as unknown rather than assigning it an incorrect label; Novelty in music class detection is the recognition of new or unknown data or signals that a machine learning system was unaware of during training [12].

In a spectrogram, a signal is displayed in 2D with time on the x-axis and frequency on the y-axis; Volkmann and Newman created the MEL scale as a pitch measurement unit in 1937 because humans struggle to perceive frequencies on a linear scale [13]. In contrast to the commonly used Hz scale, this scale makes no distinction between lower frequencies like 400 and 800 Hz and higher frequencies like 7600 and 8000 Hz; It has been discovered that humans are more adept at spotting differences in lower frequencies than in higher ones; To do this classification task, most researchers use log-based MEL Spectrograms, which are spectrograms with MEL scale on the y-axis; By doing this, it would be possible to model a neural network for genre recognition with the benefit of preventing any information loss[14].

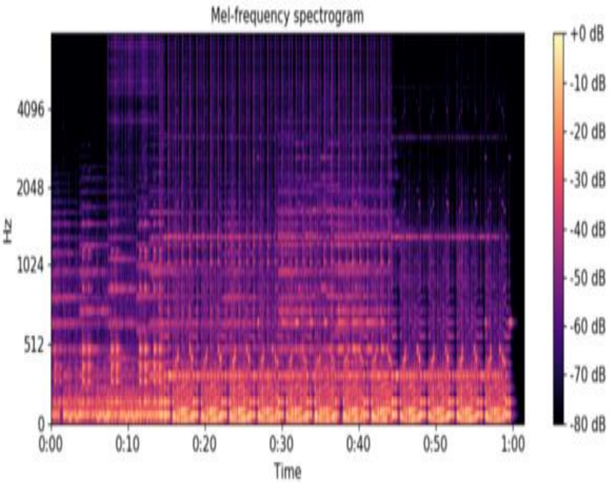

Figure 3. Music Spectrogram Data in one minute [11]

Mel-spectrogram is a useful tool for identifying and visualizing hidden audio elements. Using properties extracted from pictures, a CNNs model may carry out tasks like classification and recognition.

After the introduction, this paper is divided into 2 sections. Section II provides background on recognition and understanding musical sounds; and the section discusses the literature related to our study and Section III with conclusions and future directions.

II. RELATED WORK

A Brief History of Artificial Intelligence in Music Creation

The idea of artificial intelligence (AI) was originally introduced in 1956 at the Dartmouth Society. It is a part of the interdisciplinary field of natural science, social science, and technical science; It is a new science and technology that mimics, enhances, and extends robots or software by researching human intelligence before applying it to many fields [15].

Human intelligence is fundamentally creative, making creativity a difficult problem for artificial intelligence (AI) to solve; It is too important to ignore, even for technologically oriented AI, as creative programmes may prove very helpful in the real-world; Furthermore, cognitive science-related AI models can assist psychologists in comprehending how it is feasible for human minds to be creative [16]. The nature of creative processes is radically changing because of new technology; Computers are now a necessary component of many creative practices. The study of creating software that displays actions that could be considered creative in humans is known as "computational creativity" [17]. Although the phrase "artificial intelligence" is relatively new, the idea is quite old. Some people might be surprised to learn that musical automata and automatic music devices have existed for hundreds of years [18]. Appolonius of Perga created one of the earliest known automatic musical instrument designs for a wind instrument (247-205 B.C); Also, a mechanical spinet piano and drum set were invented and constructed in the 1500s by none other than Leonardo da Vinci.

The nineteenth century saw the development of large (27 in.) music boxes based on punched-metal-disk technology, which eventually gave rise to the orchestrion, an automated orchestra [18]. In 1930, the mechanical era had come to an end and the electrical era had already begun; Electromechanical methods (audio discs, audio tape, etc.) were used to store and reproduce recorded music. The development of sequence-controlled calculators and finally the creation of the first entirely electronic computer, the ENIAC, in 1945 were both influenced by electromechanical switching mechanisms [18].

Modern research on AI and music began in the 1950s, In the 1950s, an early attempts at computer-generated music with a focus on algorithmic music generation surfaced; The Manchester Mark II computer, invented by pioneers like Alan Turing, made it possible for computational systems to identify, produce, and analyze music, opening a variety of new study opportunities in the field of music intelligence [19]. Also In the 1960s, Electronic keyboard instruments and synthesizers were first created by applying intelligent technology to keyboard instruments, such as the electronic organ and keyboard-style electronic synthesizer; These devices can mimic the sound of other musical instruments and save pre-programmed melodies and patterns. It includes an intelligent feature and is simple to use [15].

Many projects aimed at automatic music detection started to emerge in the early and mid-1970s [18]. A. Ashton reported on a minicomputer-based system that let users play keyboard music and have the computer "remember" their performance before showing it in standard musical notation on a graphical display [20]. Also, a system for converting live music into piano-roll notation was built by Berlin mechanic Hohlfeld in 1774; To create common musical notation, this piano-roll notation was then simply converted [18]. Ya yue [32] shows in table 1, a comparison between different note detection methods.

Table1. Comparison of different note detection methods.

Method	Precision	Recall
BP	87.50	85.20
RBF	91.20	86.70
LSTM	94.10	91.90
CNN-BiLSTM	96.30	93.30

Current research into Artificial Intelligence and Music are taking several different directions, some of these concepts are [21]:

- the intelligent composer’s assistant
- responsive instruments,
- analysis and generative modeling of music,
- Recognition and understanding of musical sound.

Basic Convolutional Neural Networks (CNNs) components

According to Khan, CNN is one of the most used ML techniques today, especially for applications involving computer vision; CNN can learn representations from the grid-like data, and most recently, it demonstrated a significant performance boost in numerous ML applications [22]. CNNs' ability to reduce the number of ANN parameters is their most advantageous feature. As deep learning technology advances, more and more deep learning techniques are being used in a variety of contexts, including speech recognition, image classification, image recognition, and machine translation, among others. Convolutional Neural Networks (CNN) are a type of deep learning technique that are frequently employed in the computer vision industry.

The term 'convolution' is a mathematical process that combines two signals to create a third signal; A convolution is an integral that describes how much one function, g, overlaps another, f, as they are shifted over one another. As a result, its "blends" two functions together [23]. The CNN architecture is a collection of several convolution, activation, pooling, and other layers that are organized into a single skeleton.

A Neural structure called a Convolution Neural Network model has several layers, each of which is made up mostly of individual neurons. CNN has evolved into being highly productive in learning theoretical features thanks to more shrewd organizational structures. Three significant characteristics of CNN are weight sharing, spatial inspection, and proximity association. CNN provides all the information needed to build the model in an organized, and well calculated, manner. Figure 4 illustrates CNN's organizational structure.

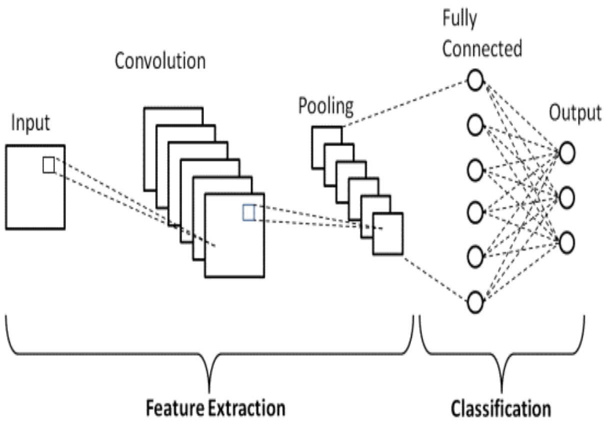

Figure 4. Layer Architecture of CNN [24]

Each neuron in a completely connected layer is entirely linked with every other neuron inside this layer below, as the name indicates. In the final stages of a CNN, when it is desired to leverage the characteristics discovered by the prior layers to produce predictions, fully connected layers are frequently employed.

To create novel architectures and achieve improved performance, the arrangement of CNN components is crucial. The function of these elements in a CNN architecture is briefly discussed in this section.

1) Convolutional layer

This layer is the initial layer that is utilised to extract the various features from the input images. In this layer, a filter of a specific size M x M and the input image conduct the mathematical process of convolution. By sliding the filter over the input image, the dot product between the filter and the input image's components is calculated according to the filter's size (M x M). The result is known as the Feature map, which provides details about the image, including its corners and edges. To learn more features from the input image, subsequent layers will later receive this feature map as input. After applying the convolution operation to the input, CNN's convolution layer passes the output to the following layer. Due to their ability to preserve the spatial link between the pixels, convolutional layers in CNN are quite beneficial.

2) Pooling layer

A Convolutional Layer is typically followed by a Pooling Layer. The main goal of this layer is to reduce the convolved feature map's size to lower the computational costs. This is accomplished by reducing the connections between layers and individually operating on each feature map. There are many Pooling operations, depending on the approach utilized. It essentially summarizes the features produced by a convolution layer. The largest component in Max Pooling is obtained from the feature map. The average of the components in a predefined sized Image portion is determined via average pooling. Sum Pooling determines the overall total of the components in the predefined section. In most cases, the Pooling Layer acts as a link between the Convolutional Layer and the Fully Connected Layer. For the networks to recognize the features on their own, this CNN model generalizes the characteristics that the convolution layer extracted. The computations in a network are likewise decreased with the aid of this.

3) Activation function

The activation function is one of the important parameters of the CNN model. They are utilised to learn and approximate any form of continuous and complex link between network variables. Simply said, it determines which model information should move forward and which should not at the network's end. By doing so, the network gains nonlinearity. There are various frequently used activation functions, including the ReLU, SoftMax, tanH, and Sigmoid functions. There is a particular use for each of these functions. Sigmoid and SoftMax functions are preferred for a CNN model of binary classification, while for multi-class classification, SoftMax is typically utilised. Clearly said, activation functions in a CNN model decide whether to activate a neuron. It determines via mathematical processes whether the input to the work is crucial or not for prediction.

4) Batch normalization

Batch normalization is a method for normalizing activations in deep neural network's intermediary layers. BN is a preferred deep learning technique because of its propensity to increase accuracy and speed up training. Nevertheless, despite its immense success, there is still little agreement over the precise cause and method of these advancements. During convolutions, we have shared filters that follow the input's feature maps (in images, the feature map is generally the height and width). Every feature map uses the same filters. The output can then be normalized and distributed over the feature maps in a suitable manner. In other words, this implies that each full feature map is generated along with the normalization values. Each feature would have a separate mean and standard deviation in a typical Batch Norm. In this case, the mean and standard deviation for each feature map will be the same for all the features it contains.

5) Dropout

Usually, over-fitting in the training dataset might result from all features being connected to the FC layer. When a certain model performs so well on training data that it has a negative effect on the model's performance when applied to new data, this is known as over-fitting. To solve this issue, a dropout layer is used, in which a small number of neurons are removed from the neural network during training, reducing the size of the model. 30% of the nodes in the neural network are randomly removed when the dropout threshold of 0.3 is passed. Dropout enhances the performance of a machine learning model by preventing over-fitting by simplifying the network. As part of training, dropout help remove neurons from neural networks.

6) Fully connected layer

Finally, The Fully Connected (FC) layer, which connects the neurons between two layers, is made up of the weights and biases as well as the neurons. These layers make up the final few layers of a CNN Architecture and are often positioned before the output layer. The input image from the earlier layers is flattened and supplied to the FC layer in this. The flattened vector then proceeds through a few additional FC layers, where the standard processes for mathematical functions happen. At this point, the classification process gets started. Since two fully connected layers would perform better than a single connected layer, two layers are connected. These levels in CNN lessen the need for human supervision.

Popular CNNs Architecture

Over time, CNNs have substantially improved, largely because of increased processing power, fresh concepts, and experiments, as well as the global deep learning community's general enthusiasm. Figure 5 describe some popular architectures described below:

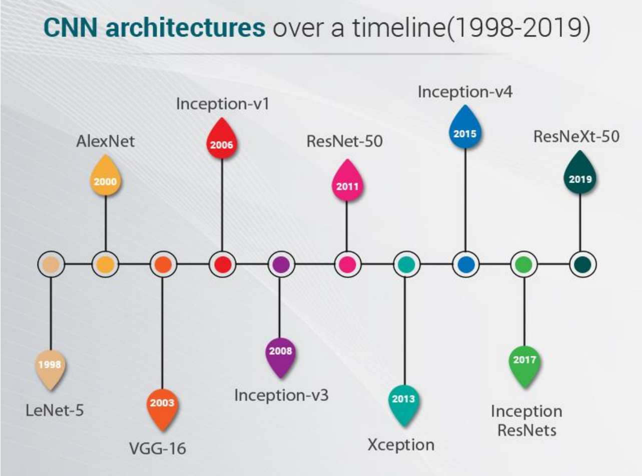

Figure 5. CNN Architectures Timeline (1998-2019) [25]

1) LeNet

Yann LeCun created the first Convolutional Networks applications that were effective in the 1990s. The LeNet architecture, used to read zip codes and other data, is the most well-known of them [22].

2) AlexNet

Alex Krizhevsky, Ilya Sutskever, and Geoff Hinton's AlexNet was the first project that helped Convolutional Networks in Computer Vision gain popularity. When the AlexNet was entered in the 2012 ImageNet ILSVRC challenge, it fared much better than the second runner-up (top 5 error of 16 percent compared to runner-up with 26 percent error). The Network's architecture was very similar to LeNet's, but it was deeper, larger, and contained Convolutional Layers piled on top of one another (previously it was common to only have a single CONV layer always immediately followed by a POOL layer) [22].

3) GoogleNet

A Convolutional Network from Google's Szegedy et al. took first place in the ILSVRC 2014 competition. Its primary contribution was the creation of an Inception Module, which drastically decreased the number of parameters in the network (4M, compared to AlexNet with 60M).

Furthermore, the top layers of the ConvNet in this research use Average Pooling rather than Fully Connected layers, removing a significant number of parameters that do not appear to be important. The Google Network also has a number of successors, the most recent being Inception-v4 [22].

Traditional network architectures only used stacked convolutional layers; more recent architectures explore creative and original techniques to build convolutional layers to increase learning effectiveness. These architectures offer general architectural guidelines that machine learning practitioners can alter to solve a range of computer vision issues. These architectures can be used as rich feature extractors for image classification, object recognition, image segmentation, and several other sophisticated applications.

Mel-Spectrograms

Each frame of the spectrum (energy/amplitude spectrum) in the Mel-Spectrum contains a short-time Fourier transform (STFT), converting the linear frequency scale to the logarithmic Mel-scale. The STFT is then applied to the filter bank to produce the Eigen vector, whose Eigen values can be roughly interpreted as that of the distribution of signal energy on the Mel-scale frequency; To train the convolutional neural networks with recognition, first convert the audio data into 15-seconds-long data, and then we turn all the data into Mel-spectrograms [26]. To detect audio, which typically has complicated properties, usable information must be extracted. One of the effective techniques for audio processing is the Mel spectrogram, and each audio sample is sampled at 8 kHz. Figure 6 illustrates the process for determining the genre of music from audio data.

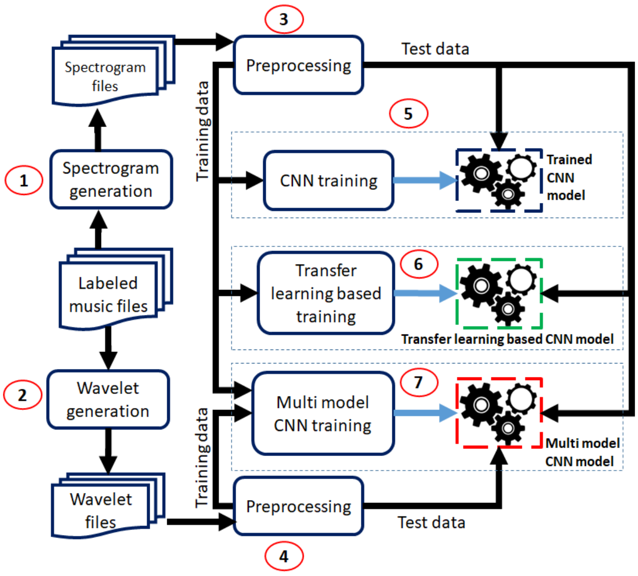

Figure 6. Image Classification Process Flow [27]

For Convolution Neural Network model to be able to classify the music into the different genres, it needs to undergo data pre-processing by converting the raw audio into Mel-spectrograms. The features that have been extracted would then be used for training and classification. More so, Mel-spectrograms are visual, and CNN works better with images.

III. PROSPECT WITH THE COMBINATION OF ARTIFICIAL INTELLIGENCE AND MUSIC

The trend of fusing technology and the arts is inevitable, and "AI and music" has slowly permeated many facets of singing, playing, and composition. In the modern era of music AI technology, the foundations of generative modeling and algorithmic composition have dynamically expanded into higher-level research and even into the music industry. The contribution of AI music intelligence to the creative process has substantially increased with the introduction of more experimental algorithms and deeper neural networks [28].

While AI technologies give researchers and artists alike a wider range of options, we emphasize in this contribution the significance of recognizing the historical context in which these technologies evolved and developed as well as the increasingly yucomplicated ramifications of their interactions with various levels of human civilizations [29]. Like how computers and software tools in the 1960s affected media art and cultural production in general, AI brings up new possibilities for creation; Regarding generative art, including music, the improved capacity to not only identify patterns from big datasets and build models that may simulate human decisions but also to produce "realistic" outputs is relevant [30]. Solanki *et al.* [33] in their study table 2, shows different musical instrument used in experiment with the result of training and testing audio results obtained.

Table 2. Musical instruments training and testing labels

Instruments	Training (n)	Testing (n)
Cello	388	111
Clarinet	505	62
Flute	451	163
Acoustic guitar	637	535
Electric guitar	760	942
Organ	682	361
Piano	721	995
Saxophone	626	326
Trumpet	577	167
Violin	580	211
Voice	778	1044

Commercial uses of automatically generated audio files, as well as contemporary music composition techniques like data signification and machine learning-based composition, put into question what constitutes creativity as well as human-centric viewpoints on music and other cultural disciplines [30].The use of deep learning in AI music is related to the fact that it is a growingly industry-led initiative that includes a subject of enormous industrial interest, even though it frequently overlaps with academic research. Given that powerful GPU cloud infrastructure and vast musical data sets were important in the rise of deep learning-based AI music, large tech corporations are well-positioned to lead related research. Although frequently referred to as something distinct, along the lines of assistive, augmentative, or transformative, these studies entail automating portions of the musician's profession that were previously performed manually [30].

In terms of the creative process itself, the subjects of AI music research include explorations of new forms of expression, helpful technologies, and convenience in the sense of automating human work. Some of these topics are more centered on the former, while others are more focused on the latter, including the commercialization of automated processes and a growth in the accessibility of a variety of creative skills [30].

Generic approaches for music elements at higher or lower levels, such as melody and polyphony, rhythm, accompaniment, score, lyrics, etc., are available through freely downloadable software tools like Google's Magenta. Products from businesses like LANDR and iZotope provide AI-based automatic mastering and mixing [30]. A growing number of start-up businesses, including Jukedeck and Amper, provide hybrid methods that can be utilized by the public to create whole audio tracks that, for example, can be used as movie background music [30].

IV. LIMITATION WITH THE COMBINATION OF ARTIFICIAL INTELLIGENCE

As opposed to perceiving AI as a tool, some musicians believe AI will have a significant negative impact on the music industry. Because such technology is still so new, the music it currently produces may not be what we want to hear. It's not very good, to put it simply. This implies that it can still take hours to create a "decent" song, at which point many would argue that the user would as well have committed the time in making anything themselves. A computer could never replicate the work (and human touch) of a true musician, according to those who have made their own music or even listeners of authentic, artistic music.

Since the beginning of time, music has played a significant role in human history. The oldest piece of music that has ever been discovered dates back about 3,400 years. Songs have been used for a very long time to convey messages and folk tales, covering anything from societal ethics to global history. Music is frequently viewed as being too priceless to be transferred to technology since it is seen by many as being such an inherently human expression. It almost seems sacrilegious to consider a machine producing "random" music that hasn't been painstakingly made by an artist.

Most musical genres in our world have qualities that go well beyond simple structural play; they engage the intellect and use a wide range of expressive techniques, rich allusions to the outside world and the body, different meanings, higher- order thought, embodiment, or even humour. These kinds of situations are what make music, in all its forms, uniquely human and pertinent to people and their societies; If creating such music is the goal of artificial creativity, it follows that the problem of general human-like music creation (as opposed to a specialized problem setting like specific style replication) should be viewed as AI-complete. In other words, the complete modeling of music's potential is not an issue for partial AI, but rather calls for human-level cognition and general intelligence [31].

V. CONCLUSION AND FUTURE SCOPE

This study, discussed how the Convolutional Neural Networks, (CNNs) and Mel-spectrogram, works to classify music genres in general. The convolution layer, which takes up most of the network's time, is the most crucial layer in CNN. The quantity of levels in a network has an impact on its performance as well. However, on the other side, the network testing and training process takes longer as the number of tiers rises. Today, CNN is regarded as a powerful machine learning technique for a variety of applications, including speech recognition, face detection, and image and video recognition, sites like Spotify and YouTube swiftly make playlists and recommend the next song based on our listening habits. The kind, mood, and/or genre of music we listen to have an impact on these patterns. Most of the factors that affect the choice are related to the song's "sound," while other factors also have a role. For instance, a pop song may seem faster than, say, a romantic song; depending on certain metrics that characterize these features, music streaming services group similar sounding or feeling tracks into the similar genre. So, categorizing music according to genres can assist in playlist curation, filtering unfavorable content, and suggesting the next songs to listeners.

As technology advances and the music business learns how to use artificial intelligence (AI) as a supplement to human creativity, our world will continue to sound better and better every year. There are still obstacles to overcome, just like with any other technical development.

In the nearest future, CNN will be an essential source of research for multiple tries on different music datasets in order to raise objective perspectives on the performance of genre detection.

REFERENCES

[1] D. A. Huang, A. A. Serafini, and E. J. Pugh, "Music Genre Classification," CS229 Stanford, **2018**.

[2] L. Aguiar and J. Waldfogel, "Platforms, promotion, and product discovery: Evidence from Spotify playlists," National Bureau of Economic Research, **2018**.

[3] Lam Hoang."Literature Review about Music Genre Classification. In Woodstock "18: ACM Symposium on Neural Gaze Detection, June 03–05, **2018**, Woodstock, NY . ACM, New York, NY, USA, 3 pages.

[4] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on speech and audio processing*, **vol. 10, no. 5, pp. 293-302, 2002.**

[5] Sciforce. "3 reasons to use ai in music industry, April 26, ." **2020**.

[6] M. Serwach and B. Stasiak, "GA-based parameterization and feature selection for automatic music genre recognition," *17th International Conference Computational Problems , **2016***

[7] K. Kashino, K. Nakadai, T. Kinoshita, and H. Tanaka, "Organization of hierarchical perceptual sounds," in *Proc. 14th Int. conf. On Artificial Intelligence*, **1995, vol. 1: Citeseer, pp. 158-164.**

[8] M. Dörfler, R. Bammer, and T. Grill, "Inside the spectrogram: Convolutional Neural Networks in audio processing," in 2017 international conference on sampling theory and applications (SampTA), **IEEE, pp. 152-155, 2017.**

[9] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," arXiv preprint arXiv:1511.08458, **2015.**

[10] L. Nanni, S. Ghidoni, and S. Brahnam, "Handcrafted vs. non-handcrafted features for computer vision classification," Pattern Recognition, **vol. 71, pp. 158-172, 2017.**

[11] Octaviano. "Music Genre Classification using Convolutional Neural Network." April 26, **2022**.

[12] N. Scaringella, G. Zoia, and D. Mlynek, "Automatic genre classification of music content: a survey," *IEEE Signal Processing Magazine,* **vol. 23, no. 2, pp. 133-141, 2006.**
[13] N. Scaringella, G. Zoia, and D. Mlynek, "Automatic genre classification of music content: a survey," IEEE Signal Processing Magazine, **vol. 23, no. 2, pp. 133-141, 2006.**
[14] J. Mehta, D. Gandhi, G. Thakur, and P. Kanani, "Music Genre Classification using Transfer Learning on log-based MEL Spectrogram," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC),* **IEEE, pp. 1101-1107, 2021.**
[15] F. Yang, "Artificial intelligence in music education," in *2020 International Conference on Robots & Intelligent System (ICRIS),* **IEEE, pp. 483-484, 2020.**
[16] M. A. Boden, "Creativity and artificial intelligence," *Artificial intelligence,* **vol. 103, no. 1- 2, pp. 347-356, 1998.**
[17] S. Colton, R. L. De Mantaras, and O. Stock, "Computational creativity: Coming of age," *AI Magazine,* **vol. 30, no. 3, pp. 11-11, 2009**.
[18] C. Roads, "Artificial intelligence and music," *Computer Music Journal,* **vol. 4, no. 2, pp. 13-25, 1980.**
[19] C. Roads, "Artificial intelligence and music,"Computer Music Journal**, vol. 4, no. 2, pp. 13-25, 1980.**
[20] A. C. Ashton, "Electronics, music and computers," University of Utah Salt Lake City, **1970**.
[21] C. Roads, "Research in music and artificial intelligence," *ACM Computing Surveys (CSUR),* **vol. 17, no. 2, pp. 163-190, 1985.**
[22] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial intelligence review,* **vol. 53, no. 8, pp. 5455-5516, 2020.**
[23] E. W. Weisstein, "Convolution," https://mathworld. wolfram. com/, **2003.**
[24] MK Gurucharan. "Basic cnn architecture: Explaining 5 layers of convolutional neural network." **2022**
[25] AISmartz. "CNNs architectures over a timeline **(1998-2019)."**
[26] Q. Zhou et al., "Cough recognition based on mel- spectrogram and convolutional neural network," Frontiers in Robotics and AI, **p. 112, 2021.**
[27] Sawan Rai."Music genres classification using deep learning techniques." **2021**
[28] Dillon Ranwala. "The evolution of music and ai technology."https://wattai.github.io/blog/music_ai_evolution/ **July 2020**
[29] B. Caramiaux and M. Donnarumma, "Artificial intelligence in music and performance: a subjective art-research inquiry," in Handbook of Artificial Intelligence for Music: **Springer, pp. 75-95, 2021.**
[30] A. Koh, "Music for AI Reports: Dual Prospects in Music Production," **2018**.
[31] M. Rohrmeier, "On Creativity, Music's AI Completeness, and Four Challenges for Artificial Musical Creativity," Transactions of the International Society for Music Information Retrieval, **vol. 5, no. 1, 2022.**
[32] Ya Yue, "Note Detection in Music Teaching Based on Intelligent Bidirectional Recurrent Neural Network", Security and Communication Networks, **vol. 2022, Article ID 8135583, 9 pages, 2022.**
[33] Solanki, Arun & Pandey, Sachin. Music instrument recognition using deep convolutional neural networks. International Journal of Information Technology. 14, **2019**

AUTHORS PROFILE

Abdulsalam A. Jamilu had Bachelor of Science from Ahmadu Bello University, Nigeria in 2014 and Master of Science from same University in year 2021. He is currently working as a Programme Analyst at Federal Medical Center Keffi, Nigeria since 2018. He is a member of Datascience Nigeria since 2020, He has published research papers in reputed international journals including International Journal Of Innovative Research & Development and it's also available online. His main research work focuses on Machine Learning, Data science and Artificial Intelligence. He has 3 years of teaching experience and 4 years of Research Experience.

Lawal abubakar had Bachelor of Science from Ahmadu Bello University, Nigeria in 2017 and currently a Postgraduate student at Bayero University Kano, Nigeria. He has published research papers in reputed international journals including International Journal Of Innovative Research & Development. His main research work focuses on condensed matter Physics, quantum computing and Artificial Intelligence.

