Online Journal of Music Sciences, 2024, 9(2), 405-414
https://doi.org/10.31811/ojomus.1580523

Research Article ♪ **Araştırma Makalesi**

Real-time chord identification application: Enabling lifelong music education through seamless integration of audio processing and machine learning

Gerçek zamanlı akor tanımlama uygulaması: Ses işleme ve makine öğreniminin kusursuz entegrasyonuyla ömür boyu müzik eğitimini mümkün kılıyor

Nihan Özbaltan[1]

[1] İzmir Bakırçay University, İzmir, Türkiye.

ABTRACT

Lifelong music education is critical need for all with a particular focus on adult learners and seniors. One of the difficulties in music education is identifying chords accurately. This is a preliminary study to develop a chord identification application using Artificial Intelligence (AI) technologies. I seek to answer the key research question of how audio processing algorithms and deep learning models can be used to provide real-time, accurate and user-friendly chord recognition that meets the diverse needs of adult learners and senior citizens. Our overall goal is to create an application that not only assists with chord identification, but also fosters a lifelong love of music and learning. My methodology is based on the principles of adult and senior education initiatives and includes the following key steps: using ready-made datasets for audio processing and feature extraction, transforming waveforms into mel spectrograms, and preparing and extending the datasets where necessary. I then train and optimise deep learning models, such as various convolutional neural network (CNN) architectures, to achieve high accuracy in chord recognition. By using advanced technologies and adhering to the principles of lifelong learning, our research aims to enhance the musical journey of individuals throughout their lives, contributing to both personal enrichment and cognitive well-being.

Keywords: lifelong music education, audio processing, machine learning, adult learners, artificial intelligence

ÖZ

Ömür boyu müzik eğitimi, özellikle yetişkin öğrenciler ve yaşlılar olmak üzere herkes için kritik bir ihtiyaçtır. Müzik eğitimindeki zorluklardan biri akorları doğru bir şekilde belirlemektir. Bu, Yapay Zeka (AI) teknolojilerini kullanarak bir akor tanımlama uygulaması geliştirmek için yapılan bir ön çalışmadır. Ses işleme algoritmalarının ve derin öğrenme modellerinin yetişkin öğrencilerin ve yaşlı vatandaşların çeşitli ihtiyaçlarını karşılayan gerçek zamanlı, doğru ve kullanıcı dostu akor tanıma sağlamak için nasıl kullanılabileceği konusundaki temel araştırma sorusunu yanıtlamayı amaçlıyorum. Genel hedefimiz yalnızca akor tanımlamaya yardımcı olmakla kalmayıp aynı zamanda müzik ve öğrenmeye yönelik ömür boyu bir sevgiyi de teşvik eden bir uygulama yaratmaktır. Metodolojim yetişkin ve yaşlı eğitim girişimlerinin ilkelerine dayanmaktadır ve aşağıdaki temel adımları içerir: ses işleme ve özellik çıkarma için hazır veri kümelerini kullanma, dalga formlarını mel spektrogramlarına dönüştürme ve gerektiğinde veri kümelerini hazırlama ve genişletme. Daha sonra akor tanımada yüksek doğruluk elde etmek için çeşitli evrişimli sinir ağı (CNN) mimarileri gibi derin öğrenme modellerini eğitiyor ve optimize ediyorum. İleri teknolojileri kullanarak ve ömür boyu öğrenme ilkelerine bağlı kalarak, araştırmamız bireylerin yaşamları boyunca müzik yolculuğunu geliştirmeyi, hem kişisel zenginleşmeye hem de bilişsel refaha katkıda bulunmayı hedefliyor.

Anahtar kelimeler: yaşam boyu müzik eğitimi, ses işleme, makine öğrenimi, yetişkin öğrenciler, yapay zeka

Nihan Özbaltan — nihan.ozbaltan@bakircay.edu.tr
Geliş tarihi/Received: 07.11.2024 — Kabul tarihi/Accepted: 10.12.2024 — Yayın tarihi/Published: 30.12.2024

1. INTRODUCTION

Learning to play a musical instrument has many benefits, enriching the mind and soul by improving cognitive skills such as memory, attention and problem solving, as well as academic performance (Lippolis et al., 2022; Preda-Uliţă, 2016). Beyond intellectual growth, musical education reduces stress, builds self-esteem and emotional expression, and promotes social interaction, teamwork, communication skills and empathy, making it a valuable pursuit in today's fast-paced, distraction-filled world (Menglibekovich, 2024; Roden et al., 2021). In the field of music education, the promotion of lifelong learning among adults and seniors is increasingly recognised as central to personal enrichment and cognitive well-being (Bowles, 1991; Choo & Choi, 2023; François et al., 2015; Nataliia, 2019; Tsugawa, 2022).

In the field of adult music learning, it is important to consider the unique constraints that adults face: a different way of absorbing new knowledge, limited time for hobbies, and a lower willingness to engage in activities without visible benefits (Dascălu et al., 2014; François et al., 2015).

One solution to these challenges is online learning, which offers a convenient and cost-effective way for adults to learn to play a musical instrument. Computer-mediated solutions provide the flexibility and accessibility needed to accommodate the busy schedules and specific needs of adult learners (Gvozdevskaia, 2021). Online platforms typically offer a broad range of courses and programs that might not be available locally, enhancing educational and career prospects. Many distance learning programs allow students to progress at their own pace, accommodating different learning speeds and styles, which helps in better absorption and retention of information (Boon, 2024; Upitis et al., 2012). Engaging with these platforms also helps students develop essential technological skills, preparing them for the modern workforce.

With the advancement of information technology, online music education has become a mainstream educational method (Ma et al., 2024). Computer technology is increasingly integrated into various teaching and educational processes. The importance of various educational tools and artificial intelligence in the educational landscape is growing and becoming crucial for overall educational success (Konecki, 2023). The need for this approach was underlined by the outbreak of COVID-19, which forced music teachers to move to online teaching. The sudden shift from traditional classroom settings to online education highlighted the critical need for effective digital learning management. Teachers, administrators, students and parents had to adapt quickly to new technologies and teaching methods. This experience underscored the importance of accountability in ensuring that educational goals are met despite the challenges posed by remote learning environments. The pandemic highlighted the need for robust distance learning strategies, comprehensive training and ongoing support to maintain the quality and continuity of education in unprecedented times (Canavar & Titrek, 2024).

A significant challenge for learners of all ages is the accurate identification of musical chords during performance, a skill that is crucial for understanding and interpreting musical compositions (Mukherjee, Dhar, Ghosh, et al., 2020; Rao & Feng, 2023). This challenge is compounded by the diverse learning needs and cognitive profiles of adult and older learners (François et al., 2015). To address these challenges, artificial intelligence (AI) technologies propose the development of a chord identification application (Hrybyk & Kim, 2010; Mukherjee, Dhar, Paul, et al., 2020; Yan, 2022). The discovery of intricate melodic motifs, chord progressions, and rhythmic patterns through musicological research provided deeper insights into musical works (Liu & Dai, 2023). By seamlessly integrating audio processing and machine learning techniques, I provide real-time and accurate chord recognition specifically tailored to the learning characteristics of lifelong learners. Such an innovation not only supports continuous skill development, but also fosters a deep-rooted appreciation for music throughout an individual's life.

Figure 1

Audio classification techniques using deep learning (Zaman et al., 2023)

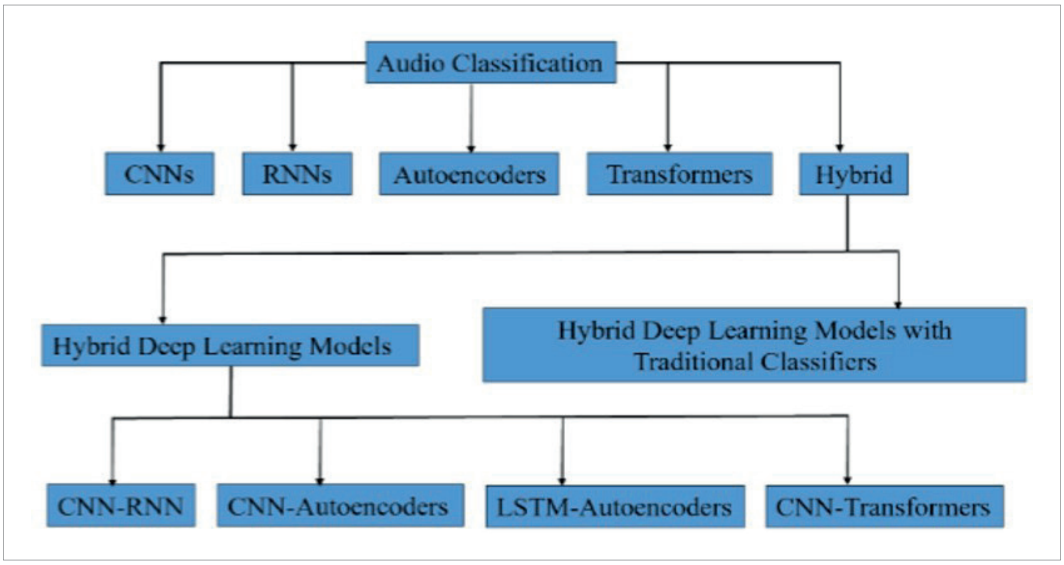


Figure 1 shows the recent advancements in audio classification using deep learning models, including CNNs, RNNs, autoencoders, transformers, and hybrid models (Zaman et al., 2023). The process typically involves converting 1D audio signals into 2D spectrograms using techniques like STFT, Mel-spectrograms, and Log-Mel spectrograms, which are then used as inputs for deep learning models. Their review covered feature extraction, deep learning architectures, datasets, performance, and application areas. CNNs are effective for tasks requiring spatial feature extraction, such as music genre and environmental sound classification. RNNs excel in capturing temporal dependencies for tasks like speech and audio-sequence classification. Autoencoders aid in unsupervised feature learning and dimensionality reduction. Transformer-based methods are valued for their concise and meaningful audio representations. Hybrid models combine multiple architectures to leverage their strengths, providing comprehensive feature extraction and capturing both spatial and temporal dependencies in audio data.

Identifying musical instruments from audio excerpts using a deep convolutional neural network is a challenge. First, employ continuous wavelet transform of audio signals via Morse wavelet to create two-dimensional feature maps, which are then processed by a robust CNN (Dutta et al., 2022). Deep Convolutional Neural Networks (DCNNs) is used in evaluating the effectiveness of for Environmental Sound Classification (ESC) tasks (Mushtaq & Su, 2020). The potential of deep learning in Music Information Retrieval (MIR) and suggest promising avenues for future advancements in the field (Kristian et al., 2024).

Audio classification is more complex than image or text classification due to the need for extensive pre-processing. Magla et al. addresses this challenge by using an envelope function with a specific threshold to eliminate non-informative audio segments. Pre-processed audio from different musical instruments is then subjected to feature extraction and classified using deep convolutional neural networks (CNNs) and Recurrent Neural Networks (RNNs). The proposed approach demonstrates effectiveness in audio classification for smart devices, achieving results comparable to state-of-the-art techniques (Mangla et al., 2022).

Giri investigates the classification of acoustic instruments using Convolutional Neural Networks (CNNs) with a dataset from Kaggle that includes recordings of piano, violin, drums, and guitar (Giri & Radhitya, 2024). The dataset consists of 700 samples each for guitar, percussion, and violin, and 528 samples for piano in the training set, with 80 samples per instrument in the test set. Features extracted using the librosa package include mel spectrograms, MFCCs, and other spectral and non-spectral characteristics. The study evaluates various CNN configurations using three feature sets: spectral-only, non-spectral-only, and a combined set. The combined feature set achieves the highest performance with a validation accuracy of 71.8% and a training accuracy of 76.9%, outperforming non-spectral features at 68.4% and spectral-only features at 69.3%. These results highlight the benefits of using an extensive feature set for accurate instrument classification.

This study bridges the gap between the necessity of lifelong music education and the practical challenges of chord identification, particularly for adult learners and seniors. The use of AI-driven audio processing and deep learning techniques ensures the development of a real-time, precise, and intuitive chord recognition tool. This innovation not only supports continuous skill development but also nurtures a deep appreciation for music, aligning with our commitment to lifelong learning and personal growth through music.

This introduction outlines our research's rationale, emphasizing the educational imperative of enhancing musical literacy among adult learners and seniors. It highlights the potential of AI-driven solutions to bridge gaps in musical education by offering intuitive tools that facilitate learning and engagement across diverse demographics. The subsequent sections delve into a dataset and our methodology, focusing on data utilization, algorithm development all aimed at achieving robust educational outcomes in lifelong music education contexts. Ultimately, this study seeks to contribute to the broader discourse on integrating technology with educational goals to enrich individuals' musical journeys over their lifetimes.

2. DATASET

This study uses publicly available dataset from Kaggle which is Acoustic Guitar Notes (n.d.). The Acoustic Guitar Notes Dataset consists of nearly 1,500 recordings, capturing every possible note on a standard 6-string guitar up to the 16th fret, with additional D2 and D#2 notes to accommodate the popular Drop D tuning. The notes span from D2, the lowest note, to G#5, the highest, covering a frequency range of 73.42 Hz to 830.61 Hz. Each note class includes at least 24 recordings.

Recordings are evenly split between two guitars: a Walden G551E steel-string guitar and a Yamaha CM-40 classical guitar with nylon strings. Each 2-second recording is captured at a 44.1 kHz sampling frequency and converted to mono format.

All notes are played directly, without additional techniques like hammer-ons or slides. To introduce variation, slight differences in playing style are included, labeled with a three-character identifier at the end of each recording's title (Table 1). While these variations are not consistent enough for training purposes, the identifiers are included for completeness and potential exclusion of undesirable sounds.

Table 1

Identifiers on the Acoustic Guitar Notes dataset

The first character indicates the string type:	The second character indicates the plucking method:	The third character indicates how the note was sounded:
's' for steel string	'p' for pick (plectrum)	'n' for a normal ring-out
	'f' for finger or thumb	'l' for a louder than normal note
	'n' for nail	'm' for a note muted early with the palm

3. METHOD

A signal is a representation of how a particular quantity varies over time. In audio, this quantity is air pressure, which fluctuates as sound waves move through the air. To capture these variations digitally, I sample the air pressure at fixed intervals, typically at a rate of 44.1 kHz (or 44,100 samples per second). This process creates a digital waveform of the audio signal, allowing it to be analysed, modified, and interpreted with computer software.

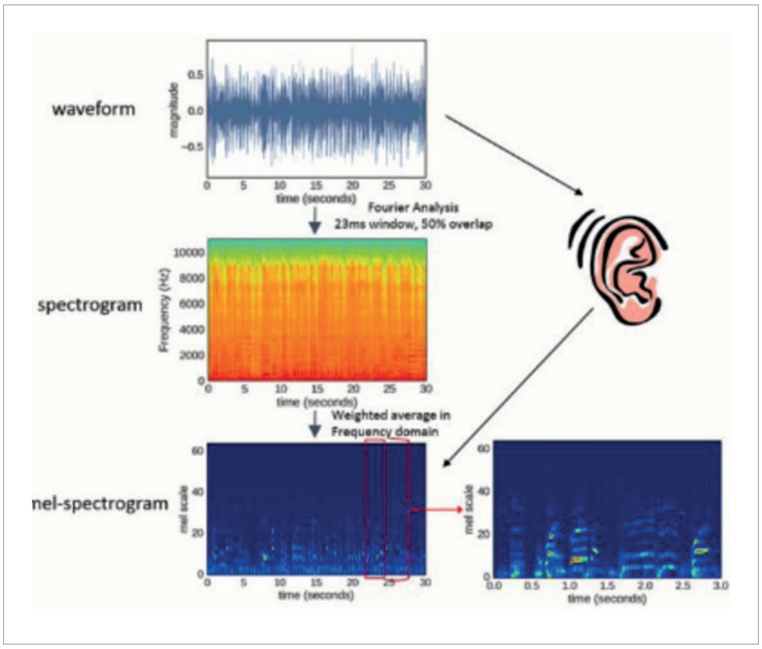
An audio signal is made up of multiple single-frequency sound waves combined. When I sample the signal over time, I capture only the resulting amplitudes at each interval. The Fourier transform, a mathematical operation, allows us to break down the signal into its individual frequencies and their respective amplitudes, effectively converting the signal from the time domain to the frequency domain. This transformation produces a spectrum, which displays the intensity of each frequency component in the signal.

A spectrogram is essentially a sequence of Fast Fourier Transforms (FFTs) displayed over time, showing how the amplitude or loudness of different frequencies varies. When creating a spectrogram, certain adjustments are applied: the y-axis (frequency) is converted to a logarithmic scale, while colour represents amplitude on a decibel scale, also logarithmic. These transformations align with the way humans perceive sound, as we are most sensitive to a specific range of frequencies and amplitudes.

In 1937, Stevens, Volkmann, and Newman developed the mel scale, which is a perceptual pitch scale designed so that equal distances in pitch correspond to equal perceptual differences for listeners (Stevens et al, 1937). To convert standard frequencies to the mel scale, a specific mathematical transformation is applied. This conversion is essential for analysing and interpreting sound in a way that reflects human hearing. A mel spectrogram is a form of spectrogram in which the frequencies are mapped to the mel scale, making it particularly useful for sound analysis related to human perception (Dong, 2018). Figure 2 shows the way from waveform to mel-spectogram.

Figure 2

From waveform to mel-spectogram (Dong, 2018)



3.1. Data Preprocessing

Loading Audio Data: The librosa library (McFee, 2015) is utilized to load the audio files. It is essential to specify the sample rate (`sr`) as `None` to maintain the original sampling rate.

Converting to Mel Spectrogram: Instead of employing raw audio data, we transform the audio files into Mel spectrograms. A Mel spectrogram is a visual representation of audio data that is more suitable for processing by a neural network. These spectrograms are generated using the librosa library.

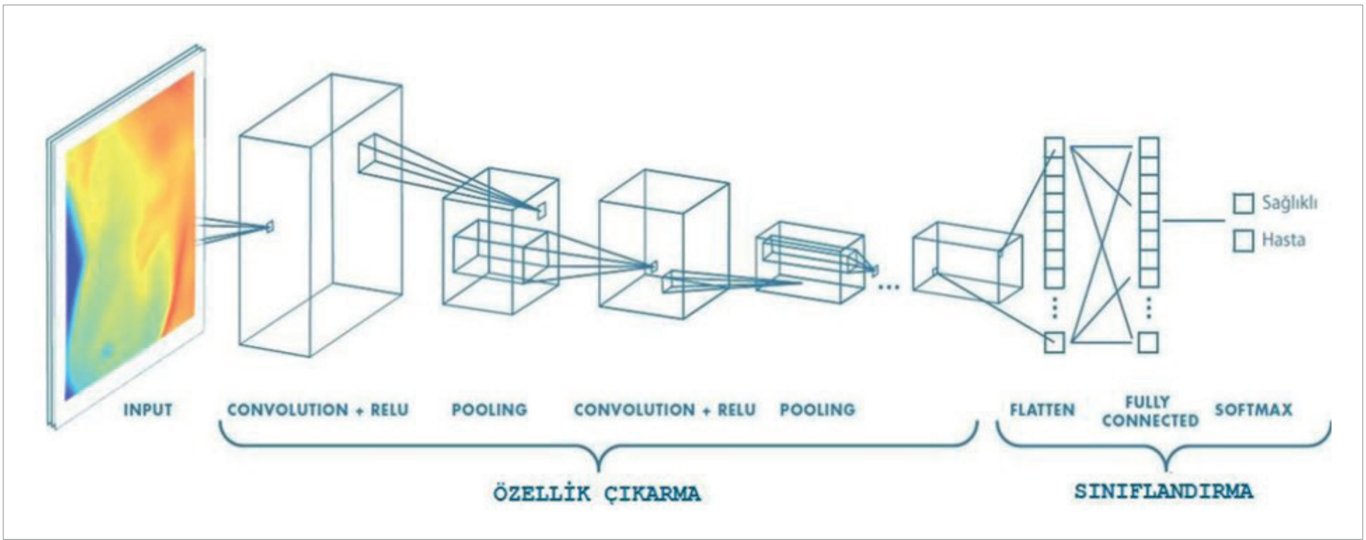
Resizing Spectrograms: To ensure uniformity among all spectrograms, they are resized to a target shape (e.g., 128×128 pixels). This step is critical to guarantee compatibility with the neural network.

3.2. Model Architecture

Deep learning models have proven to be highly effective tools across various domains of signal processing, such as speech signal processing, image processing, and time-series analysis (Mavaddati, 2024). CNN's (see Figure 3) are extensively used in various applications, such as speech recognition, audio classification, music recommendation, and audio source separation (isolating individual audio sources within a single recording), among others. They have transformed the field of audio classification, unlocking a wide range of possibilities. Although CNNs work with audio or speech signals, they generally do not operate directly on raw one-dimensional (1D) signals. Instead, audio data is preprocessed into 2D representations, which can then be used as inputs to the CNN model (Zaman et al., 2023).

Figure 3

Structure of CNN



Two neural network architectures are designed for classification tasks, both incorporating a sequence of convolutional layers, ReLU activation functions, max pooling layers, and a fully connected layer that concludes with a softmax activation for output classification. The first architecture, referred to as Model-1, includes batch normalization layers following each convolutional layer to stabilize and accelerate the training process by normalizing the inputs to each layer. The second architecture, Model-2, mirrors the structure of Model-1 but excludes batch normalization layers. This distinction allows for a comparative analysis of the impact of batch normalization on model performance, convergence speed, and generalization capabilities.

Model 1

Network Architecture

1. Input layer: Input shape corresponds to the shape defined in input_shape.
2. Convolution layer: 32 different 3×3 filters with a stride of 1, activation 'relu'.
3. Batch Normalization layer: Normalizes the outputs of the previous layer.
4. Convolution layer: 64 different 3×3 filters with a stride of 1, activation 'relu'.
5. Batch Normalization layer: Normalizes the outputs of the previous layer.
6. Max pooling layer: 2×2 pooling.
7. Dropout layer: 25% dropout rate.
8. Convolution layer: 128 different 3×3 filters with a stride of 1, activation 'relu'.
9. Batch Normalization layer: Normalizes the outputs of the previous layer.
10. Max pooling layer: 2×2 pooling.
11. Dropout layer: 25% dropout rate.
12. Flatten layer: Flattens the input to 1D.
13. Fully connected layer: 256 neurons, activation 'relu'.
14. Dropout layer: 50% dropout rate.
15. Output layer: Number of neurons equal to the length of classes,

Batch Normalization (BN) is an auxiliary layer used to mitigate the issue of exploding and vanishing gradients, which commonly occur when optimizing deep neural networks (Ioffe & Szegedy, 2015). BN normalizes the output of the previous layer, ensuring well-behaved gradients. When training with Batch Normalization, each training example is processed along with other examples in the mini-batch, resulting in the network no longer producing deterministic values for a given training example. Our experiments revealed that this variability enhances the network's generalization capabilities. While Dropout (Srivastava et al., 2014) is commonly used to mitigate overfitting, I discovered that in a batch-normalized network, Dropout can either be omitted or its strength significantly reduced.

Model 2

Network Architecture

Input layer: Input shape corresponds to the shape defined in input_shape.
Convolution layer: 32 different 3×3 filters with a stride of 1, activation 'relu'.
Convolution layer: 64 different 3×3 filters with a stride of 1, activation 'relu'.
Max pooling layer: 2×2 pooling.
Dropout layer: 25% dropout rate.
Convolution layer: 128 different 3×3 filters with a stride of 1, activation 'relu'.
Max pooling layer: 2×2 pooling.
Dropout layer: 25% dropout rate.
Flatten layer: Flattens the input to 1D.
Fully connected layer: 256 neurons, activation 'relu'.
Dropout layer: 50% dropout rate.
Output layer: Number of neurons equal to the length of classes, activation 'softmax'.

For 2D layers/filters, the first dimension corresponds to the mel-scale and the second dimension corresponds to the time. All hidden layers use RELU activation functions, the output layer use softmax function, and the loss is calculated using cross-entropy function. Dropout and L2 regularization were used to prevent extreme weights. The model is implemented using Keras (2.0.1) (Chollet, 2015) with TensorFlow as backend and trained on a single GTX-1070 using stochastic gradient descent.

The chosen loss function for optimization is empirical cross-entropy, a standard choice for classification tasks. This function measures how well the estimated posterior distribution (output from the SoftMax layer) aligns with the ground truth labels. The Adam optimizer, widely used for training neural networks, is employed in this study.

4. RESULTS, CONCLUSİONS AND RECOMMENDATİONS

Two models are evaluated by changing epoch and batch size and check their accuracy, training time and loss. An epoch refers to one complete pass through the entire training dataset. Here, models were trained for 10 or 15 epochs. Batch size refers to the number of training examples utilized in one iteration. Here, the batch sizes used are 16 and 32. Accuracy measures of how well the model is performing. Loss is a measure of the model's error. Lower values indicate better performance.

According to Table 2, Model 1 achieves higher accuracy (84.19%) with 15 epochs compared to 10 epochs (70.79%), but the training time is longer. Model 2 achieves extremely high accuracy (99.31%) with 15 epochs. However, with 10 epochs, the accuracy is very low (2.75%). The loss is very low with 15 epochs (0.0311), indicating excellent performance. With 10 epochs, the loss is slightly lower (0.0275), but this does not correspond to higher accuracy, indicating potential overfitting or other issues with the training process.

All in all, Model 1 performs better with more epochs and a larger batch size, balancing accuracy and training time. Model 2 performs exceptionally well with 15 epochs and a smaller batch size but performs poorly with 10 epochs and a larger batch size. The choice between models and hyperparameters (epochs and batch size) significantly affects the training outcomes, including accuracy, training time, and loss.

Tablo 2

Models Summary

Model	Epoch	Batch Size	Accuracy	Training Time	Loss
1	15	16	0.8419243693351746	2028.3262300491333	0.42452532052993774
1	10	32	0.7079038023948669	1182.4379615783691	3.2317662239074707
2	15	16	0.993127167224884	2022.086627960205	0.03112957440316677
2	10	32	0.027480177581310272	1298.1670281887054	0.027480177581310272

According to Table 1, Model 1 showed a more consistent performance, with a higher accuracy achieved when trained for more epochs and with a smaller batch size. The increase in training time for more epochs is justified by the significant improvement in accuracy. Model 2 demonstrated that it could achieve near-perfect accuracy with 15 epochs and a batch size of 16, indicating its potential for high precision. However, its performance drastically declined with fewer epochs and a larger batch size, suggesting sensitivity to training configurations. Both models confirmed that a higher number of epochs generally leads to better performance, although this also results in longer training times. The low loss values for Model 2 suggest excellent performance, but the discrepancy between the loss and accuracy with 10 epochs indicates possible overfitting or other training issues.

Model 1 is recommended for scenarios where consistent performance is desired across different training configurations. Model 2 is suitable for applications where extremely high accuracy is required, provided that careful tuning of training parameters (such as epochs and batch size) is performed.

Using more epochs (15 or higher) is recommended for both models to ensure better learning and higher accuracy. A smaller batch size (e.g., 16) appears to be beneficial, particularly for Model 2, in achieving higher accuracy.

Investigate the drastic drop in accuracy for Model 2 with fewer epochs and a larger batch size to understand the underlying causes. Explore other hyperparameter tuning methods and regularization techniques to further optimize both models.

Ensure adequate computational resources to handle longer training times when using more epochs, especially for applications requiring high precision. For real-time or resource-constrained environments, balance the trade-offs between accuracy and training time by selecting appropriate model configurations based on specific needs. By following these recommendations, the models can be effectively utilized for chord identification, contributing to the broader goal of enhancing lifelong music education through advanced AI technologies.

Ethical approval

Open source data was used, does not require an ethics committee.

Author contribution

Study conception and design: NÖ; data collection: NÖ; analysis and interpretation of results: NÖ; draft manuscript preparation: NÖ. All authors reviewed the results and approved the final version of the article.

Source of funding

The authors declare the study received no funding.

Conflict of interest

The authors declare that there is no conflict of interest.

Etik kurul onayı

Açık kaynaklı veri seti kullanılmıştır ve etik kurul onayı gerektirmemektedir.

Yazarlık katkısı

Çalışmanın tasarımı ve konsepti: NÖ; verilerin toplanması: NÖ; sonuçların analizi ve yorumlanması: NÖ; çalışmanın yazımı: NÖ. Tüm yazarlar sonuçları gözden geçirmiş ve makalenin son halini onaylamıştır.

Finansman kaynağı

Yazarlar, çalışmanın herhangi bir finansman almadığını beyan etmektedir.

Çıkar çatışması

Yazarlar, herhangi bir çıkar çatışması olmadığını beyan etmektedir.

REFERENCES

Acoustic Guitar Notes. (n.d.). https://www.kaggle.com/datasets/mohammedalkooheji/guitar-notes-dataset/data

Boon, İ. E. T. (2024). Self-regulated learning skills in instrument education: A qualitative study. *International Journal of Education and Literacy Studies, 12*(1), 106-114. https://doi.org/10.7575/aiac.ijels.v.12n.1p.106

Bowles, C. L. (1991). Self- expressed adult music education interests and music experiences. *Journal of Research in Music Education, 39*(3), 191-205. https://doi.org/10.2307/3344719

Canavar, S., & Titrek, O. (2024). The impact of COVID-19 pandemic on school administrators' psychology, family life, and work life (İznik sample). In *Proceedings of the International Conference on Education Studies* (pp. 76-96). https://doi.org/10.2991/978-94-6463-380-1_8

Chollet, F. (2015) Keras. GitHub. https://github.com/fchollet/keras

Choo, S. H., & Choi, J. H. (2023). The status and satisfaction level of instrumental music education on adult music education. *Korean Journal of Research in Music Education, 52*(3), 109-130. https://doi.org/10.30775/KMES.52.3.109

Dascălu, M.-I., Coman, M., Postelnicu, R., & Nichifor, C. (2014). Learning to play a musical instrument in adulthood: Challenges and computer-mediated solutions. *Procedia - Social and Behavioral Sciences, 142*, 23-28. https://doi.org/10.1016/j.sbspro.2014.07.639

Dong, M. (2018). Convolutional neural network achieves human-level accuracy in music genre classification. arXiv. http://arxiv.org/abs/1802.09697

Dutta, A., Sil, D., Chandra, A., & Palit, S. (2022). CNN based musical instrument identification using time-frequency localized features. *Internet Technology Letters, 5*(1), e191. https://doi.org/10.1002/itl2.191

François, C., Grau-Sánchez, J., Duarte, E., & Rodriguez-Fornells, A. (2015). Musical training as an alternative and effective method for neuro-education and neuro-rehabilitation. *Frontiers in Psychology, 6,* 475. https://doi.org/10.3389/fpsyg.2015.00475

Giri, G. A. V. M., & Radhitya, M. L. (2024). Musical instrument classification using audio features and convolutional neural network. *Journal of Applied Informatics and Computing*, *8*(1), 226-234.

Gvozdevskaia, G. A. (2021). Methods of maintaining the mental activity of students when learning to play musical instruments online (on the example of learning to play the piano). *Musical Art and Education, 9*(3), 66-80. https://doi.org/10.31862/2309-1428-2021-9-3-66-80

Hrybyk, A., & Kim, Y. (2010). Combined audio and video analysis for guitar chord identification. In *Proceedings of the 11th International Society for Music Information Retrieval Conference, ISMIR 2010*.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv. http://arxiv.org/abs/1502.03167

Konecki, M. (2023). Adaptive drum kit learning system: Impact on students' learning outcomes. *International Journal of Information and Education Technology, 13*(10), 767-773. https://doi.org/10.18178/ijiet.2023.13.10.1959

Kristian, Y., Zaman, L., Tenoyo, M., & Jodhinata, A. (2024). Advancing guitar chord recognition: A visual method based on deep convolutional neural networks and deep transfer learning. *ECTI Transactions on Computer and Information Technology (ECTI-CIT), 18*(2), 235-249.

Lippolis, M., Müllensiefen, D., Frieler, K., Matarrelli, B., Vuust, P., Cassibba, R., & Brattico, E. (2022). Learning to play a musical instrument in middle school is associated with superior audiovisual working memory and fluid intelligence: A cross-sectional behavioral study. *Frontiers in Psychology, 13,* 982704. https://doi.org/10.3389/fpsyg.2022.982704

Liu, X., & Dai, Y. (2023). Virtual computer systems in AI-powered music analysis: A comparative study for genre classification and musicological investigations. *Journal of Information Systems Engineering and Management, 8*(4), 23395. https://doi.org/10.55267/iadt.07.14016

Ma, R., Deeprasert, J., & Jiang, S. (2024). Toward lifelong learning: Modelling willingness of chinese older adults learning music via social media. *Theory and Practice, 2024*(5), 3067-3081. https://doi.org/10.53555/kuey.v30i5.3388

Mangla, P., Arora, S., & Bhatia, M. P. S. (2022). Intelligent audio analysis techniques for identification of music in smart devices. *Internet Technology Letters, 5*(2), e268. https://doi.org/10.1002/itl2.268

Mavaddati, S. (2024). A voice activity detection algorithm using deep learning in the time-frequency domain. *Neural Computing and Applications,* 1-15. https://doi.org/10.1007/s00521-024-10795-x

McFee, B., Raffel, C., Liang, D., Ellis, D. P. W., McVicar, M., Battenberg, E., & Nieto O. (2015). librosa: Audio and music signal analysis in Python. In *Proceedings of the 14th Python in Science Conference* (pp. 18-24). https://doi.org/10.25080/Majora-7b98e3ed-003

Menglibekovich, B. M. (2024). Benefits of learning an instrument: Exploring the advantages. *Journal of Education, Ethics and Value, 3*(5), 140-143.

Mukherjee, H., Dhar, A., Ghosh, M., Obaidullah, S. M., Santosh, K. C., Phadikar, S., & Roy, K. (2020). Music chord inversion shape identification with LSTM-RNN. *Procedia Computer Science, 167*, 265-274. https://doi.org/10.1016/j.procs.2020.03.327

Mukherjee, H., Dhar, A., Paul, B., Obaidullah, S. M., Santosh, K. C., Phadikar, S., & Roy, K. (2020). Deep learning-based music chord family identification. In S. Santosh (Ed.), *Advances in intelligent systems and computing* (Vol. 1034, pp. 169-178). Springer. https://doi.org/10.1007/978-981-15-1084-7_18

Mushtaq, Z., & Su, S. F. (2020). Environmental sound classification using a regularized deep convolutional neural network with data augmentation. *Applied Acoustics, 167*, 107389. https://doi.org/10.1016/j.apacoust.2020.107389

Nataliia, S. (2019). Lifelong learning: Music adult education. *Continuing Professional Education Theory and Practice, 1*, 17-22. https://doi.org/10.28925/1609-8595.2019.1.1722

Preda-Uliţă, A. (2016). Improving children's executive functions by learning to play a musical instrument. *Bulletin of the Transilvania University of Brașov, Series VIII: Performing Arts, 9*(2), 37-47.

Rao, Z., & Feng, C. (2023). Automatic identification of chords in noisy music using temporal correlation support vector machine. *IAENG International Journal of Computer Science, 50*(2), 179-185.

Roden, I., Friedrich, E. K., Etzler, S., Frankenberg, E., Kreutz, G., & Bongard, S. (2021). Development and preliminary validation of the Emotions While Learning an Instrument Scale (ELIS). *PLoS ONE, 16*(8), e0255019. https://doi.org/10.1371/journal.pone.0255019

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929-1958.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America*, 8, 185-190. https://doi.org/10.1121/1.1915893

Tsugawa, S. (2022). Intergenerational music teaching and learning among preservice music teachers and senior adult musicians. *Research Studies in Music Education, 44*(1), 52-69. https://doi.org/10.1177/1321103X20977541

Upitis, R., Abrami, P. C., & Brook, J. (2012). Learning to play a musical instrument with a digital portfolio tool. *Journal of Instructional Pedagogies, 9*, 1-12.

Yan, H. (2022). Design of online music education system based on artificial intelligence and multiuser detection algorithm. *Computational Intelligence and Neuroscience, 2022*(1), 1-12. https://doi.org/10.1155/2022/9083436

Zaman, K., Sah, M., Direkoglu, C., & Unoki, M. (2023). A survey of audio classification using deep learning. *IEEE Access, 11*, 106620-106649. https://doi.org/10.1109/ACCESS.2023.3318015