MSc Project - Reflective Essay

Project Title:	Classification of Monophonic Guitar Notes Across Strings in Noisy Environments
Student Name:	Haruniya Rajan
Student Number:	220256861
Supervisor Name:	Miss Dalia Senvaityte
Programme of Study:	MSc Big Data Science

Introduction

Classifying monophonic guitar notes string-wise is a complex and nuanced challenge, particularly under noisy conditions. In a world where digital music technology is rapidly advancing, this project sought to contribute to the field by employing a Convolutional Neural Network (CNN), a deep learning architecture known for its efficacy in pattern recognition (LeCun et al. 1998). Utilizing the GuitarSet dataset (Xi et al. 2018), the project encompassed several stages, including data loading, preprocessing, feature extraction, model design, and evaluation. A key aspect was the introduction of noise through data augmentation, simulating real-world scenarios and enhancing the model's ability to generalize. The novelty of this approach, coupled with the personal challenges and growth experienced throughout the project, adds layers of complexity and intrigue to the endeavor. This reflective essay serves as an in-depth exploration of the project, elucidating the technical intricacies, methodological choices, and the scientific rigor that defined this endeavor. The insights and reflections contained herein offer a comprehensive view into a project that stands at the intersection of music theory, signal processing, and machine learning, providing a unique perspective in the ongoing dialogue in the field and reflecting a journey marked by innovation, learning, and perseverance.

The strength and weakness analysis

The project's strengths are prominently showcased in the classification report for guitar strings, achieving an overall accuracy of 79% for string classification. This success is a testament to the robustness of the applied methodologies, including the strategic use of CNN, renowned for handling image and audio data (Krizhevsky et al. 2012). The high precision for certain strings, especially the 6th string at 92%, underscores the model's adeptness at discerning intricate audio patterns. This was particularly satisfying for me, as it validated the hours spent fine-tuning the model and selecting the right features. The incorporation of noise during data augmentation was a masterstroke, simulating real-world noisy conditions and thereby enhancing the model's resilience and generalization capabilities. This is particularly crucial for real-world applications where pristine audio conditions are rare. The meticulously chosen feature extraction techniques, encompassing spectral contrast, chroma, and MFCCs, ensured a comprehensive and nuanced representation of the guitar sounds (Tzanetakis & Cook 2002). This rich representation was pivotal in empowering the model to distinguish between the nuanced differences of various strings, something that has been a persistent challenge in previous research. Furthermore, the hyperparameter tuning process played a crucial role in refining the model's performance, ensuring it was tailored to the specificities of the dataset. The high f1-scores for certain strings reflect the model's balanced performance in both precision and recall, a key metric in classification tasks. Additionally, the project's adaptability to different noise levels and its scalability to handle large datasets demonstrate its potential for broader applications.

Despite the commendable strengths, the project wasn't without its challenges, especially evident in the note classification, where the overall accuracy dropped to 29%. In the string classification, the highest recall was 86.66% for string 1, while the lowest was 72.50% for string 6. The highest precision was 89.73% for string 6, while the lowest was 70.61% for string 4.In note classification, the highest recall was 47.06% for F5, while the lowest was 0% for F♯5 and G♯5. The highest precision was 43.01% for G4, while the lowest was 0% for F♯5 and G♯5. These variations spotlight the model's inconsistencies, a common hurdle in the field of audio classification, hinting at the model's struggles with certain frequency ranges or overlapping harmonics inherent in guitar notes. This was a humbling reminder of the complexity of the task at hand and the limitations of even the most advanced models. The choice of CNN, while effective for string classification, might not have been the optimal choice for the intricate task of note differentiation (Choi et al. 2016). The noise augmentation, while beneficial for string classification, might have introduced challenges in note classification, making it challenging for the model to discern between genuine note variations and noise-induced fluctuations. These disparities might be indicative of overfitting or biases introduced by the GuitarSet dataset. The model's performance could vary with guitars of different makes, tonal qualities, or playing techniques, limiting its universal applicability. A more nuanced approach, such as combining CNN with other neural network architectures, might have been more adept at capturing the sequential nature of musical notes, a direction that offers exciting possibilities for future research.

Future Work

The possibilities for future work in this project are extensive and hold great promise for advancing the field of monophonic guitar note classification. One intriguing avenue is the exploration of different neural network architectures that might be more adept at capturing the subtleties of musical notes (Humphrey et al. 2012). This could include experimenting with recurrent neural networks (RNNs) or transformer models, which have shown potential in sequence modeling and might provide more nuanced note differentiation.

Future work could also involve developing more sophisticated noise handling techniques, possibly employing denoising autoencoders or specialized filtering methods (Vincent et al. 2008) could be explored in depth, possibly leading to a new standard in noise-robust audio classification. This direction could have broad applications beyond guitar note classification, impacting various fields such as speech recognition and environmental sound analysis. The feature extraction techniques used in the project could be further enriched by incorporating additional audio features or leveraging state-of-the-art deep learning-based feature extraction methods. This could enhance the model's ability to capture complex audio characteristics, paving the way for more accurate and versatile audio analysis tools.

The success in string classification under noisy conditions opens exciting doors for real-time applications. One promising avenue is the development of a real-time guitar tuning application that functions during live performances. Unlike existing tuning apps that are used prior to a performance, this innovative application could provide continuous tuning adjustments as the musician plays, ensuring optimal sound quality throughout the performance. This technology could revolutionize how musicians interact with their instruments, allowing for more dynamic and responsive playing experiences. Additionally, an educational tool that provides instant feedback on playing techniques could be developed, tailored to different skill levels and musical genres, making it accessible and valuable to a wide range of users. Collaboration with professional musicians, music educators, and sound engineers could lead to more targeted and impactful applications of the technology. Engaging in interdisciplinary research and development could foster innovation and ensure that the technological advancements align with the real-world needs and aspirations of the music community.

Work that could have been conducted with more time

With more time, the project could have delved into several promising areas that would have enriched the research and its potential applications. Collecting and curating a more diverse dataset, encompassing guitars of different makes, tonal qualities, and playing techniques, would have allowed for a more comprehensive understanding of the complexities involved in guitar sound analysis. This could have included collaborations with musicians and instrument manufacturers to ensure a wide-ranging and authentic dataset.

The optimization for low-latency processing could have been a significant focus, exploring cutting-edge algorithms and hardware solutions to enable real-time applications such as live performance assistance or interactive music creation tools (Wang et al. 2014). This direction could have bridged the gap between research and real-world usability, making the technology more accessible to artists and educators.

Ethical considerations, such as data privacy and user consent, could have been more thoroughly investigated, ensuring that the project aligns with best practices in research ethics. Sustainability considerations, including energy-efficient computing and responsible data management, could have been integrated into the project's design, reflecting a commitment to environmental responsibility.

The hyperparameter tuning process could have been further refined, and additional experiments with various model architectures and feature extraction techniques could have been conducted. Exploring ensemble methods or multi-modal learning could have provided insights into how different models complement each other, leading to more robust and versatile solutions. In the context of guitar sound analysis, multi-modal learning could have included different modalities such as audio signals, visual representations of finger placements, and tactile feedback from the strings. These complementary data sources could have been integrated to create a more comprehensive model that captures the intricate dynamics of guitar playing.

The project could have also explored partnerships with educational institutions or music technology companies to pilot the developed models in real-world scenarios. Engaging with industry stakeholders and academic researchers could have facilitated valuable feedback and fostered innovation, ensuring that the technology meets the needs and expectations of end-users. Furthermore, the development of a user interface would have been a necessary step before conducting more in-depth user studies. Understanding how musicians and educators interact with the technology, and identifying areas for improvement from a user experience perspective, would have required this interface. This human-centered approach could have ensured that the technology is not only technically sound but also user-friendly and aligned with the creative process of music-making.

Critical Analysis of the relationship between theory and practical work

The relationship between theory and practical work in this project is a complex interplay that has shaped the outcomes and insights gained. The theoretical foundations of Convolutional Neural Networks (CNNs) and their application in audio processing guided the initial approach (LeCun et al. 1998). The theory posited that CNNs, with their ability to capture spatial hierarchies and patterns, would be well-suited for classifying guitar strings and notes. However, the practical implementation revealed a nuanced picture.

In the practical implementation of the project, a critical evaluation of the process and robust troubleshooting skills were essential. Challenges arose in various stages, from data preprocessing to model tuning and noise augmentation. The inconsistencies within the dataset required statistical techniques to rectify, while the model's initial overfitting necessitated the application of various regularization techniques and architectural adjustments. The innovative noise augmentation process required iterative testing and refinement to balance real-world simulation with pattern recognition. The significant drop in note classification accuracy led to a detailed analysis of the model's errors, providing insights into its limitations and directions for future improvement. This continuous evaluation, adaptation, and problem-solving were instrumental in achieving the project's goals and provided valuable lessons in resilience, adaptability, and innovation.

In the realm of string classification, the theory held strong, with the CNN model achieving a commendable accuracy. The theoretical understanding of how noise can mimic real-world conditions guided the successful implementation of noise augmentation, enhancing the model's robustness. The choice of features, grounded in the theoretical understanding of audio signal processing, contributed to this success (Tzanetakis & Cook 2002).

Conversely, the practical work in note classification exposed some gaps between theory and implementation. While the CNN was theoretically a sound choice, the practical challenges of distinguishing subtle tonal differences between notes were not fully anticipated. The theoretical promise of noise augmentation as a means to simulate real-world conditions became a double-edged sword in practice, complicating the note classification task.

Furthermore, the project's reliance on the GuitarSet dataset (Xi et al. 2018), while theoretically sound, introduced practical limitations. The assumption that the dataset would generalize across different guitars and playing styles was not tested in this project, as the GuitarSet dataset contains only one guitar. This limitation highlights the importance of considering the diversity and representativeness of the dataset in future work, and it serves as a reminder of the potential biases and limitations that can arise from relying on a single source of data.

Awareness of Legal, Social, Ethical Issues, and Sustainability

The project's development and implementation were guided by a keen awareness of legal, social, ethical, and sustainability considerations. Legally, the use of publicly available datasets and adherence to copyright laws ensured compliance with intellectual property rights (Lessig 2004). The project's focus on guitar note classification has social implications, potentially aiding in music education and democratizing access to learning resources. However, the potential misuse of the technology in applications like automated music creation raises ethical questions about creativity and originality.

From an ethical standpoint, transparency in methodology and acknowledgment of limitations were prioritized to maintain integrity in research (Resnik 2011). The potential biases in the dataset and the model's performance were openly addressed, reflecting a commitment to ethical research practices.

Ethical considerations, such as data privacy and user consent, could have been more thoroughly investigated, ensuring that the project aligns with best practices in research ethics. Sustainability considerations, including energy-efficient computing and responsible data management, could have been integrated into the project's design, reflecting a commitment to environmental responsibility.

Sustainability was considered in the project's design, with an emphasis on efficient algorithms and computational resources. The choice of CNN, while computationally intensive, was balanced with optimization techniques to minimize energy consumption. Specifically, techniques such as pruning, quantization, and efficient model architecture selection were employed to reduce the computational demands of the CNN. This not only made the model more energy-efficient but also aligned with broader goals of environmental sustainability (Patterson et al. 2016). The project's potential applications in enhancing music education and appreciation align with broader social sustainability goals, promoting cultural enrichment and inclusivity.

Conclusion

The journey of classifying guitar notes string-wise using a CNN algorithm under noisy conditions has been both challenging and enlightening. The project's success in achieving a high accuracy in string classification, coupled with the insights gained from the weaknesses in note classification, has laid a solid foundation for future exploration and improvement. The reflective analysis of strengths, weaknesses, future possibilities, the critical relationship between theory and practical work, and the awareness of legal, social, ethical issues, and sustainability has provided a comprehensive understanding of the project's multifaceted dimensions.

The project not only stands as a testament to the power of machine learning in audio analysis but also serves as a reminder of the complexities and responsibilities that come with technological innovation. The lessons learned from this project extend beyond the technical realm, encompassing broader themes of creativity, ethics, and societal impact. The experience has been a valuable stepping stone, fostering growth as a researcher and technologist, and inspiring continued pursuit of innovation and excellence in the field of audio signal processing.

References

Choi, S. et al. (2016), 'Automatic tagging using deep convolutional neural networks', arXiv preprint arXiv:1606.00298 .

Humphrey, E. et al. (2012), Learning feature representations for music classification, in '18th International Conference on Digital Signal Processing (DSP)'.

Krizhevsky, A. et al. (2012), 'Imagenet large scale visual recognition challenge', arXiv preprint arXiv:1409.0575 .

LeCun, Y. et al. (1998), 'Gradient-based learning applied to document recognition', Proceedings of the IEEE **86**(11), 2278–2324.

Lessig, L. (2004), Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity, Penguin.

Patterson, J. et al. (2016), Deep Learning: A Practitioner's Approach, O'Reilly Media, Inc.

Resnik, D. B. (2011), 'What is ethics in research & why is it important?'.

Tzanetakis, G. & Cook, P. (2002), 'Musical genre classification of audio signals', IEEE Transactions on Speech and Audio Processing **10**(5), 293–302.

Vincent, E. et al. (2008), 'Extracting document structure from plain text via template-based learning', Pattern Recognition **41**(12), 3845–3859.

Wang, Y. et al. (2014), 'End-to-end text recognition with convolutional neural networks', Pattern Recognition **47**(12), 4042–4055.

Xi, Q. et al. (2018), Guitarset: A dataset for guitar transcription, in 'ISMIR'.