Student Award Papers

Evaluation of machine learning performance on source separated passive acoustic recordings

Carter Paprocki, B.S.,M.S.[1], Morgan Crump, B.S.,M.S.,Ph.D.[2], Tyler Dare, B.S.,Ph.D.[1], Andrew Barnard, B.S.,M.S.,Ph.D.[3], Kristin Cody, B.S.,M.S.,Ph.D.[4], Peter Newman, B.S., M.S., Ph.D.[5], Derrick Taff, B.A.,M.A.,Ph.D.[6]

[1] Acoustics, Pennsylvania State University, [2] Recreation, Park, and Tourism Management, Pennsylvania State University, Pennsylvania State University, [3] Administration, Michigan Technological University, [4] Advanced Technology Innovation Pipeline, Naval Nuclear Laboratory, [5] Rubenstein School of Environment and Natural Resources, University of Vermont, [6] Recreation, Park, and Tourism Management, Pennsylvania State University

Keywords: 74.8 Frequency domain and spectral analysis, 56.1 Outdoors

https://doi.org/10.3397/1/377333

Noise Control Engineering Journal

Vol. 73, Issue 3, 2025

The United States National Park Service often deploys passive acoustic monitoring devices whose data are used to characterize the ecological health of the site and quantify the level of anthropogenic sources present. The analysis of the data is then used to inform conservation management policy that protects the natural soundscape of the National Parks. Traditional methods for analyzing the acoustic recordings involve human listeners annotating a subset of this data and labeling the acoustic sources present. Manual data processing is labor intensive and relies on human perception that varies between individuals, these limitations ultimately reduce the amount of data that can be analyzed. By augmenting the manual annotation of acoustic data with machine learning (ML) this research aims to increase the quantity of data that can be annotated while achieving higher annotation confidence for acoustic sources. Previous research has shown that separating the transient acoustic sources from the background acoustic environment reduces data complexity allowing for improved characterization of the sources. This paper will evaluate the performance of machine learning models using source separated data to understand the effects it has on data annotation accuracy, hyperparameter complexity, and ML model training time.

1. INTRODUCTION

Anthropogenic pollution refers to environmental contamination caused by human activities.[1] Since the Industrial Revolution, a variety of pollutants have been released into the environment at growing rates, including anthropogenic noise, which negatively impacts ecosystems, human health, and the broader environment.[1-4] Increasingly, anthropogenic noise is encroaching upon natural areas used for outdoor recreation, personal escapes, and connections with nature.[5-7]

The U.S. National Park Service (NPS) manages protected lands—including national parks, monuments, and historic sites—to conserve natural and cultural resources while providing recreational opportunities. Under the NPS Organic Act of 1916, the agency is required to protect acoustic resources to ensure future generations can experience natural soundscapes unimpaired. This commitment is reinforced by the 2006 NPS Management Policies, which mandate the preservation of park soundscapes to protect cultural and historic auditory experiences.[8] However, as anthropogenic noise alters the soundscapes of parks and protected areas (PPAs), both human experiences and environmental conditions within these spaces are changing.[6,9] Understanding these dynamic changes is crucial for effective park management.

1.1. Soundscapes in Parks and Protected Areas

Soundscapes in PPAs result from a complex interplay between natural sounds, human activity, and environmental factors. These areas feature diverse biotic and abiotic elements—such as bird calls, rustling leaves, and flowing water—shaped by time of day, season, weather, and topography.[10-15] Anthropogenic noise sources, including road traffic, aircraft, and visitor activity, can mask or alter these natural sounds.[16-18]

As human presence increases in and around PPAs, natural soundscapes are increasingly disturbed, contributing to biodiversity loss and habitat degradation.[6,19-23] Natural landscapes free from human-generated noise have become increasingly rare.[24,25] These shifts in soundscapes directly affect visitor experiences and satisfaction.[26,27]

1.2. Monitoring and Managing Acoustic Environments

Protecting acoustic environments requires ongoing monitoring, data analysis, and the integration of interdisciplinary research into park policies. A key component of NPS soundscape management is the inventorying and monitoring of acoustic conditions (NPS Management Policies, section C.5). This involves establishing baseline data to inform site-specific management plans, with ongoing efforts to expand the scope and efficiency of data collection.

NPS currently employs a passive acoustic monitoring system that records 25 consecutive days of audio (600 hours). Due to the extensive effort required for manual annotation, NPS selects a subset of 8 days, manually labeling 2.5% of the total recorded audio, as outlined in the Air Tour Management Plan.[28] This equates to analyzing 10-second intervals every two minutes, with trained human operators (EHOs) at the NPS listening lab annotating the data.[29] The manual annotation process is time-intensive, often requiring multiple weeks per site, limiting the capacity for large-scale soundscape conservation efforts.

1.3. Machine Learning for Acoustic Analysis

Machine learning has proven effective in detecting and annotating acoustic events in passive acoustic recordings. Since 2012, the use of machine learning in eco-acoustics has expanded significantly, improving the efficiency of long-term acoustic analysis.[30] Bioacoustics studies have successfully used machine learning to identify specific sound sources, such as wildlife vocalizations or anthropogenic noise.[31,32] However, broad-scale identification of multiple acoustic sources remains complex and resource-intensive.

To improve annotation efficiency without relying solely on large deep learning models, this research explores robust adaptive de-noising techniques to extract meaningful acoustic features. Effective preprocessing is essential, as machine learning models perform best with high-quality input data. Proper preprocessing minimizes noise and allows models to generalize to new, unprocessed data, reducing overfitting.[33]

This study evaluates the implementation of spectral subtraction as a technique for separating transient acoustic events from non-stationary background noise. Evaluation of spectral subtraction will be done by comparing CNN models trained on a denoised spectrogram dataset as well as CNN models trained using the original spectrograms. The goal of this research is not to design a robust generalized CNN but to determine whether there is an improvement in CNN performance when using spectral subtraction as a preprocessing step. By categorizing acoustic data into distinct classification domains, we aim to simplify input data, accelerate model training, and enhance multiclass classification accuracy.[34] Low signal-to-noise ratios (SNR) are known to impair the performance of machine learning techniques, including convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and deep learning neural networks.[35-37] Addressing these challenges will improve automated soundscape analysis, supporting NPS efforts to monitor and preserve natural acoustic environments more effectively.

2. METHOD

The goal of this research was to evaluate the impact of spectral subtraction as a preprocessing step in acoustic source classification using a convolutional neural network (CNN). Specifically, we aimed to determine whether spectral subtraction improves classification accuracy and to quantify its effect on CNN training time. Previous research has shown that spectral subtraction can effectively denoise audio up to -6 dB SNR while allowing the spectrograms separated into component transient signal and background noise spectrograms.[38] The dataset consisted of five transient acoustic sources, representing the most common transient signals in the NPS dataset: bird song, people talking, jet aircraft, propeller aircraft, and helicopters. To increase the complexity of the classification task and expand the training dataset, we injected each signal with seven diverse types of real-world and synthetic noise at different signal to noise ratios (SNR): traffic, wind, rain, insect chorus, generator, white noise, and pink noise. Then a CNN was trained on this dataset using the three spectrogram representations original, transient, and background.

2.1. Classification Dataset Creation

Due to the inherent complexity of real-world sounds which at any moment contain multiple acoustic sources with overlapping frequency content, it was determined that generating a synthetic dataset using a subset of audio from the NPS archive would provide a scalable single class dataset. For each of the data classes five sections of 2-minutes of audio were selected where the target class was the dominate acoustic feature through the entire 2-minute recording. Some of the recordings did require additional filtering such as the aircraft audio due to them occasionally containing an erroneous bird call. However, the filter order was limited to ensure minimal manipulation of the signal phase and existing noise due to the recording device. With the audio cleaned, various forms of background noise at 6 dB and 0 dB SNR levels were added to the recordings using the data augmentation technique noise injection. In order to properly scale each noise recording to the corresponding signal the noise weighting value β was calculated using Eqn. 1.

$$SNR = 10 * \log_{10}\left(\frac{mean\left(A^2_{signal}\right)}{\beta * mean\left(A^2_{noise}\right)}\right) \quad (1)$$

Noise injection is a proven tool used for data augmentation and has been shown to increase prediction accuracy and generalization of CNN machine learning models when a training dataset has a limited number of samples.[39] For these reasons, the five noise sources were chosen that represent common environmental background noise found in the NPS dataset. Two sources of synthetic noise were also selected white and pink noise due to their noise profiles having spectral energy in the two frequency ranges of interest broadband and low frequency, respectively. The total number of noise records for each class was equal to the transient sources and consisted of five 2-minute recordings. The number of original samples for each transient acoustic

Table 1. Number of samples for each transient class with different background noise profiles.

Class	Original Sample Size	Background Noise						
		Wind	Rain	Traffic	Generator	Insect	White	Pink
Bird	120	1200	1200	1200	1200	1200	1200	1200
Speech	120	1200	1200	1200	1200	1200	1200	1200
Jet	120	1200	1200	1200	1200	1200	1200	1200
Propeller	120	1200	1200	1200	1200	1200	1200	1200
Helicopter	120	1200	1200	1200	1200	1200	1200	1200

source class and the total number of samples generated with each from each noise source is shown in Table 1.

2.2. Spectral Subtraction Denoising

Spectral subtraction has been shown to amplify the signal-to-noise-ratio and provide effective detection and denoising for acoustic sources at exceptionally low SNR. Compared to traditional methods of band pass filtering and wavelet filtering, spectral subtraction is more effective due to it not relying on a set frequency band and can effectively denoise sources that have overlapping frequency content with the background noise. Additionally spectral subtraction can be localized using a moving average of the noise estimation and allow for denoising of transient signals in nonstationary background noise. Spectral subtraction also does not require microphone array data, which is needed for signal cross correlation techniques, meaning it can be a powerful denoising tool for bioacoustics and passive acoustic monitoring studies that often only deploy one or two microphones. Due to these many strengths spectral subtraction was chosen for this research.

The specific algorithm used in this research was a modification of the work done by Verteleskaya and Simak and the full derivation of the method can be found in a previous publication.[38,40] The method developed by Verteleskaya and Simak used a voice activity detector in order to detect which vertical frames of a spectrogram contained the acoustic signal of interest. The voice activity detector was replaced by a Shannon entropy threshold detector in this research. Shannon Entropy was first described by the father of information theory C.E. Shannon in a 1948 publication.[41] Shannon entropy is the measure of information contained in a signal, wherein noise contains the most amount of information due to it being a stochastic variable has equal probability of occurring as all other values. Once a signal is contained within the noise the entropy and information of the signal drops due to there being a higher probability the measured value is a part of the signal. The calculation of Shannon entropy from a probability distribution is shown in Eqn. 2.

$$H[X] = -\sum_{i=1}^{I} P(x_i) * log_2(P(x_i)) \tag{2}$$

When this is applied to a spectrogram, each vertical frame of the power spectra is normalized to ensure the summable area under the curve is equal to one shown in Eqn. 3. Where the normalized power spectrum is represented by $\widehat{Y_N}[m]$ and N is the index of the spectrogram frame and m is the index of the frequency.

$$\widehat{Y_N}[m] = \frac{Y_N[m]}{\int_{k=1}^{K} Y_N[m]\ dm} \tag{3}$$

Normalization also benefits the algorithms' ability to detect sources due to it not relaying on total energy contained in each frame but the relative information in each frame meaning detection will not be bias to transient sources with higher energy which is a common limitation in energy threshold detection methods. The normalized power spectrum $\widehat{Y_N}[m]$ is then substituted into equation 2 as the probability distribution and one is able to calculate the Shannon entropy of each spectrogram frame. The instantaneous Shannon entropy is then used to calculate the standard deviation and a moving average. This research found that a 20 second moving average with the standard deviation of the 2-minute recording was effective parameters for denoising. The threshold for detecting spectrogram frames that contained signal was calculated by subtracting ½ the standard deviation of the 2-minute recording from the 20 second moving average. If the instantaneous value was below the threshold value it was determined that this frame contained signal and should not be included in the local noise estimation. This calculation is shown in Eqn. 4.

$$if\ \widehat{H}[N] < \frac{1}{T}\sum_{i=1}^{T} \widehat{H}_{x-i+1} - \alpha * \sigma \ \ L[N] = 0 \tag{4}$$

The noise logical vector $L[N]$ is set to zero when the instantaneous Shannon entropy is less than the threshold value denoting the spectrogram frame contains transient signal and is set to one when the instantaneous value is greater than the threshold value denoting the spectrogram frame is noise and should be included in the noise estimation. This process is illustrated in Fig. 1, where a recording of bird song is passed into the denoising algorithm, it is shown that dips in the in the Shannon entropy align with the bird calls thus an estimation of the localized noise is calculated. The calculation of the localized noise is similar to the entropy threshold value, as it takes the frames of the spectrogram where the noise logic is 1 and calculates a moving average of the corresponding power spectra. This calculation of the local noise estimate is shown in Eqn. 5 where the local noise power spectra is $\widehat{D_N}[m]$, and L is the number of spectrogram frames determined to be noise in the 20-second moving window.

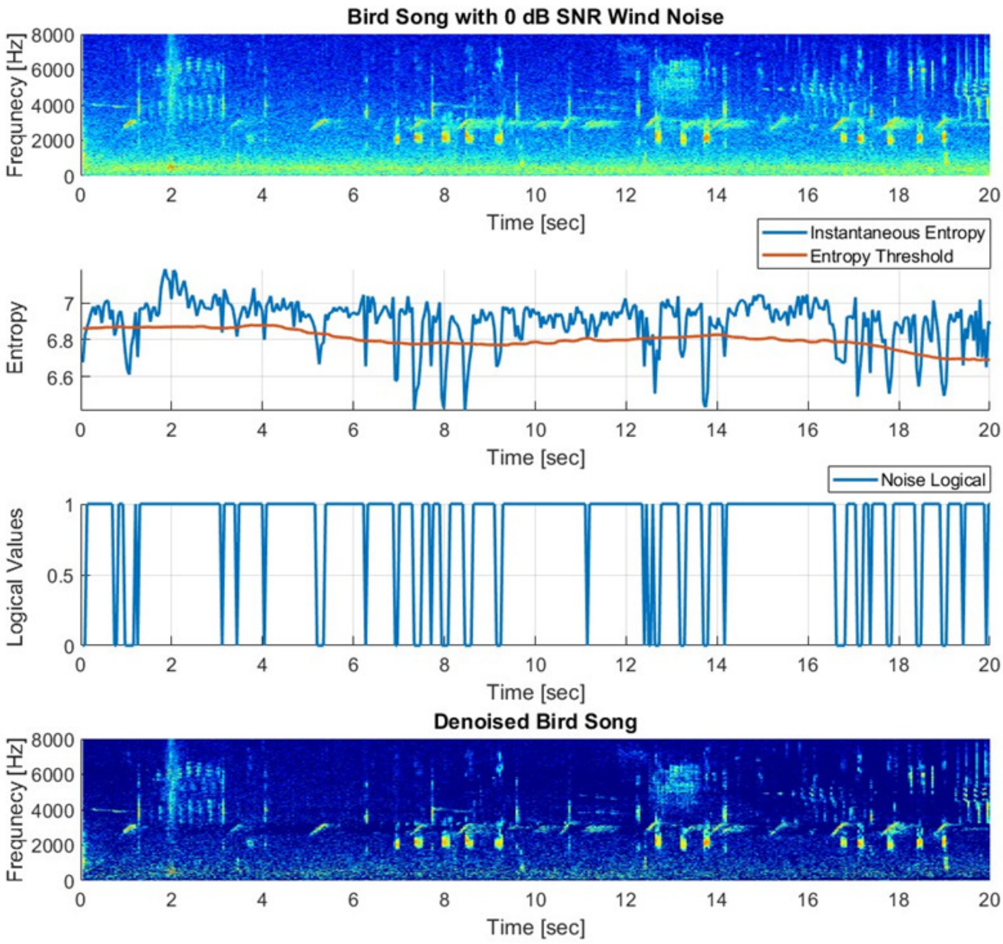

Fig. 1. Instantaneous Shannon entropy dips below the threshold value when tonal content is present due to the frame containing less information compared to the average, allowing for effective denoising to occur.

$$\widehat{D_N}[m] = \frac{1}{\sum L} \sum_{i=1}^{L} Y_{N-i+1}[m] * L[N] \tag{5}$$

In order to calculate the final denoised spectrogram shown in Fig. 1, the local noise estimate $\widehat{D_N}[m]$ is compared to the magnitude of the input spectrogram. When the magnitude of the input spectrogram minus the local noise estimate is greater than the noise floor value, then that bin has the noise estimate subtracted from the magnitude of the input spectrogram. If this value is less than the noise floor value, then that bin was set to the noise floor. The noise floor is calculated by multiplying a user input δ by the local noise estimate, it was found that a value of $\delta = 1 * 10^{-5}$ works well. This is shown in Eqn. 6.

$$\widehat{S}_N[m] \begin{cases} if\ Y_N[m] - \widehat{D}_N[m] > \delta * \widehat{D}_N[m] \\ \qquad = Y_N[m] - \widehat{D}_N[m] \\ else \\ \qquad = \delta * \widehat{D}_N[m] \end{cases} \tag{6}$$

Using Spectral Subtraction, one is able to take the original input spectrogram data and then spit it into its component transient signal spectrogram and then subtract the signal spectrogram from the input spectrogram to generate a background spectrogram. These three spectrogram representations were then parsed into 5 second intervals and then passed into the CNNs to perform classification.

2.3. Machine Learning Implementation

In this research, the goal was to determine the effect of using spectral subtraction as a preprocessing step on machine learning performance, specifically in the classification of transient signals or injected noise. To achieve this, 20 CNN networks were trained with different percentages of the training data, ranging from 50% to 90% shown in Table 2. The validation percentage was fixed at 10% to ensure fair evaluation of all models, and the validation data was used to determine when to stop training using MATLAB's 'ValidationPatience' parameter.[42] This method helps prevent overfitting by halting training when the validation loss stops improving after a set number of intervals.

Given the large dataset size, a minibatch size of 32 was used to ensure the data could be processed efficiently without exceeding the system's memory. The CNN architecture, illustrated in Fig. 2, begins with an image input layer of size 400x400x1, corresponding to the dimensions of a 5-second spectrogram. Since the audio files were already normalized between -1 and 1 during the creation of the noise-injected dataset, no further normalization was applied at the input

Table 2. Convolution Neural Network training procedure.

Input Data	Classification	Percentage of Training Data				
Original	Signal	50	60	70	80	90
Original	Background	50	60	70	80	90
Signal	Signal	50	60	70	80	90
Background	Background	50	60	70	80	90

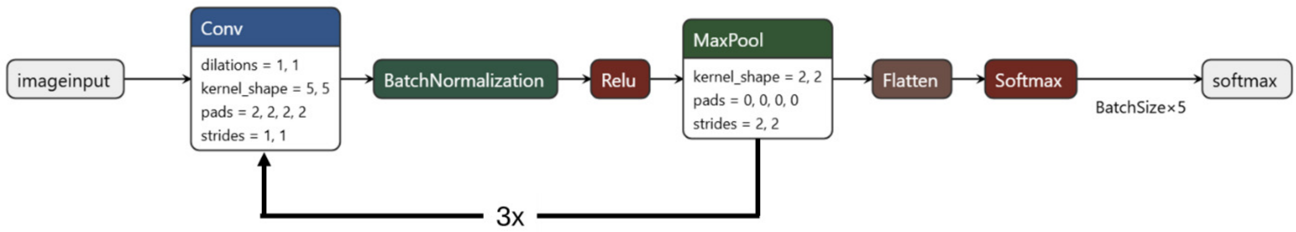

Fig. 2. CNN network used for four classification tasks.

layer, preserving the dB scale representation of the spectrograms.

The input is then passed through a convolution layer with 32 filters, followed by a batch normalization layer. The batch normalization normalizes the activations before they are passed through the activation function. A max pooling layer is applied to reduce the spatial dimensions of the feature maps, identifying the most critical features while improving computational efficiency. This sequence of convolution, normalization, and pooling is repeated three times. Finally, the output is fed into a fully connected layer, followed by a softmax function for classification. This CNN architecture aims to effectively capture the important features of the spectrograms while preventing overfitting and improving training efficiency. The optimization function implemented is the Adam optimizer, which uses a stochastic gradient-based approach to update the weights and biases of the convolutional layers, as well as the learnable parameters (gamma and beta) in the batch normalization layers. The Adam optimizer adapts learning rates for each parameter individually using first and second moment estimates, improving convergence stability and efficiency in deep CNN models.[43]

Once training was complete, the convolutional neural networks (CNNs) were evaluated using the entire synthetic dataset to assess their overall performance for both background and transient source classification tasks. Their relative performance was measured based on several key metrics, including the final training accuracy they achieved, which reflects how well the models fit the training data. Additionally, the number of iterations required for the validation loss to stop decreasing was analyzed, as this indicates the point at which the models stopped improving and potentially began overfitting. The number of iterations was also used to gauge how fast the model learned. The validation accuracy was also considered, providing insight into how well each CNN generalized to new data during model training. Finally, their overall accuracy on the entire dataset was examined to determine their ultimate classification performance across all available samples. By comparing these metrics, a comprehensive understanding of each model's strengths and weaknesses was obtained.

3. RESULTS

This research demonstrates that spectral subtraction preprocessing enhances CNN classification performance by improving generalization, increasing training accuracy, and reducing the number of iterations required for validation loss convergence. A comparison of CNN classification accuracy and training efficiency, as a function of the percentage of training data used, is shown in Fig. 3.

While the highest validation accuracy was achieved by the CNN model trained on the original dataset (92.4%) compared to the spectral subtraction model (90.6%), the spectral subtraction models exhibited a more consistent improvement across all training dataset sizes. On average, CNNs trained with spectral subtraction preprocessing achieved a 3% higher validation accuracy, suggesting improved generalization to unseen data. Furthermore, the spectral subtraction models mitigated the accuracy drop observed in the original dataset as the training dataset size increased.

In terms of training accuracy, CNNs trained with spectral subtraction data saw a more significant improvement compared to those trained on the original spectrograms. On average, spectral subtraction models exhibited a 6.25% higher training accuracy, further supporting the idea that this preprocessing step enhances feature extraction for classification. Additionally, spectral subtraction models demonstrated increased training efficiency, converging 7% faster than models trained on the original dataset. This increase in training efficiency is particularly demonstrated when comparing the models trained on 70% of the dataset, where spectral subtraction models achieved comparable valida-

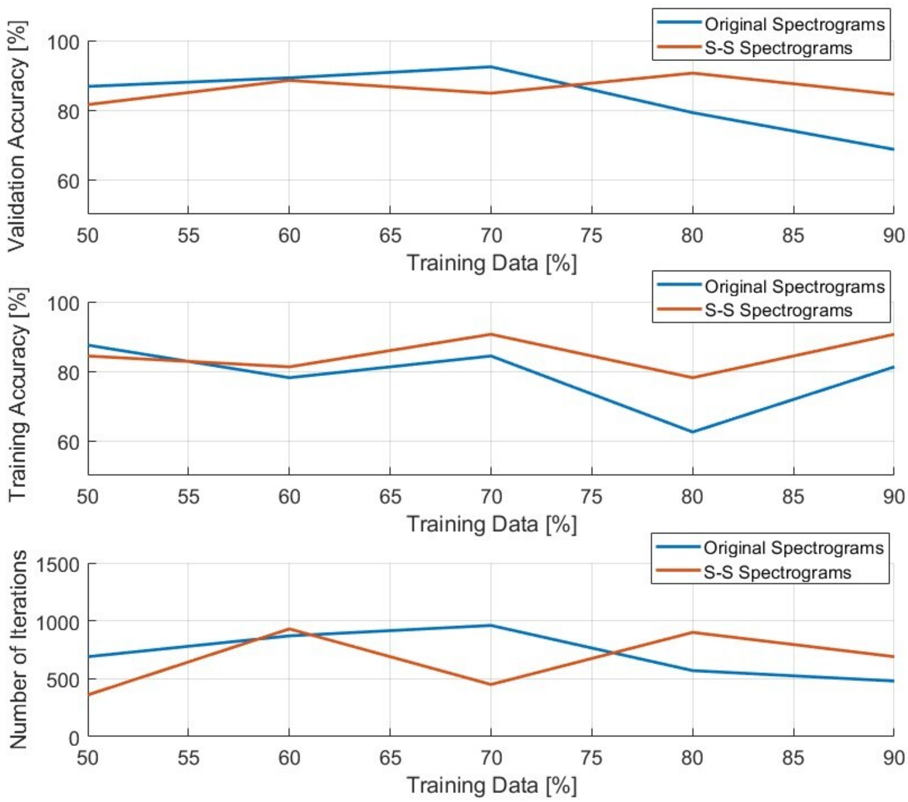

Fig. 3. Comparison of CNN models trained on original and spectral subtraction (S-S) data.

tion and training accuracy but required 50% fewer iterations to converge.

To evaluate how the models performed in classifying individual transient signal classes, this research compared the confusion matrices of the best-performing spectral subtraction and original dataset models. Model selection was based on validation accuracy, with the spectral subtraction model trained on 80% of the training data and the original model trained on 70% of the training data chosen for comparison. These models were then presented with the entire dataset training, validation, and test dataset to determine their overall classification accuracy. It was found that the spectral subtraction model achieved an overall classification accuracy of 90.7% whereas the original model achieved a slightly higher 92.4% classification accuracy. As shown in Fig. 4, this drop in overall accuracy was mostly due to the propeller classification where it saw a drop in accuracy of 8.6% compared to the original model and often chose Helicopter and Jet classes. Further investigation into the noise sources causing the propeller misclassification showed that white noise and pink noise caused the majority of misclassifications. This trend held true when evaluating all of the noise sources associated with misclassifications for the spectral subtraction model.

4. DISCUSSION AND CONCLUSIONS

This work aimed to understand whether implementation of spectral subtraction as a preprocessing step improved CNN classification performance. It was observed that on average the CNN model trained using spectral subtraction data did perform better having higher validation and training accuracy while needing fewer training iterations to converge. However, the CNN trained using the original data did provide the highest overall performance when comparing individual models, it is believed that this is due to the way spectral subtraction handles longer duration transient events such as aircraft overflight. Where the algorithm could be treating important learnable features of the aircraft overflights as a background source and passing these features to background noise spectrogram representation. It was also observed that the noise classes associated with the majority of miss-classifications using the spectral subtraction models were due to the synthetic noise added to the models. It is believed that this synthetic noise caused extensive masking at low frequency where most of the source classes have predominant frequency content.

Building on these findings, future research will aim to refine the spectral subtraction technique and investigate its application to real-world passive audio recordings. Testing this technique on diverse, real-world environments—ranging from quiet backcountry sites to noise-heavy urban settings—will allow for a comprehensive evaluation of the al-

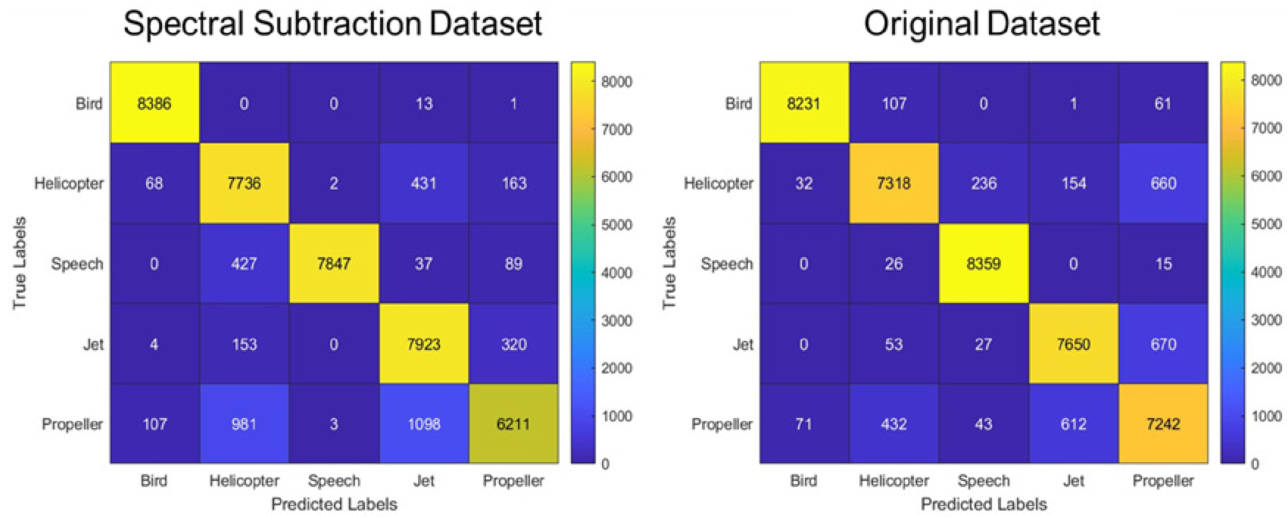

Fig. 4. Comparison of source classification accuracy using the best CNN models.

gorithm's robustness in different acoustic conditions. This will also assist in understanding how well it preserves transient events during the denoising process.

Additionally, further optimization of the spectral subtraction algorithm will be explored by examining how user-defined parameters, such as the averaging window length and noise estimation weights, influence the retention of key features in transient signals. Having a deeper understanding of this process will help inform whether previously defined transient events such as aircraft overflights are better handled as background noise.

In conclusion, this research shows that spectral subtraction is a powerful preprocessing step that has measurable improvements for CNN classification tasks. Spectral subtraction allowed the CNN networks to learn meaningful classification features faster while providing higher accuracy classifications given larger more robust datasets.

5. ACKNOWLEDGEMENTS

I would like to thank Rickover Fellowship Program in Nuclear Engineering for the continued mentorship and funding of my research. Also, I would like to acknowledge the National Park Service who are the sponsors of this research and have provided the dataset I use in my research. I would also like to thank my advisors Dr. Barnard and Dr. Dare who have been invaluable in guiding and mentoring my research.

Submitted: July 31, 2025 EDT. Accepted: August 01, 2025 EDT. Published: August 31, 2025 EDT.

REFERENCES

1. Arihilam NH, Arihilam EC. Impact and control of anthropogenic pollution on the ecosystem–A review. *Journal of Bioscience and Biotechnology Discovery*. 2019;4(3):54-59. doi:10.31248/JBBD2019.098

2. Barber JR, Crooks KR, Fristrup KM. The costs of chronic noise exposure for terrestrial organisms. *Trends in ecology & evolution*. 2010;25(3):180-189. doi:10.1016/j.tree.2009.08.002

3. Jariwala R, Upadhyaya I, George NV. Robust equalizer design for adaptive room impulse response compensation. *Applied Acoustics*. 2017;125:1-6.

4. Bragdon CR. *Noise Pollution: The Unquiet Crisis*. University of Pennsylvania Press; 2016. doi:10.2307/j.ctv4v31dk

5. Goines L, Hagler L. Noise pollution: a modem plague. *South Med J*. 2007;100(3):287-294.

6. Bronzaft AL. Noise pollution: A hazard to physical and mental well-being. In: *Handbook of Environmental Psychology*. ; 2002:499-510.

7. Buxton RT, McKenna MF, Mennitt D, et al. Noise pollution is pervasive in US protected areas. *Science*. 2017;356(6337):531-533. doi:10.1126/science.aah4783

8. Olson S, Reid P, eds. *Protecting National Park Soundscapes*. National Academies Press; 2013.

9. National Park Service. *National Park Service Management Policies*.; 2006.

10. Barber JR, Burdett CL, Reed SE, et al. Anthropogenic noise exposure in protected natural areas: estimating the scale of ecological consequences. *Landscape ecology*. 2011;26:1281-1295.

11. Peña L, Monge-Ganuzas M, Onaindia M, De Manuel BF, Mendia M. A holistic approach including biological and geological criteria for integrative management in protected areas. *Environmental management*. 2017;59:325-337. doi:10.1007/s00267-016-0781-4

12. Song X, Lv X, Yu D, Wu Q. Spatial-temporal change analysis of plant soundscapes and their design methods. *Urban Forestry & Urban Greening*. 2018;29:96-105.

13. Cumming GS, Allen CR. Protected areas as social-ecological systems: perspectives from resilience and social-ecological systems theory. *Ecological applications*. 2017;27(6):1709-1717.

14. Krause B. Anatomy of the soundscape: evolving perspectives. *Journal of the Audio Engineering Society*. 2008;56(1/2):73-80.

15. Pijanowski BC, Villanueva-Rivera LJ, Dumyahn SL, et al. Soundscape ecology: the science of sound in the landscape. *BioScience*. 2011;61(3):203-216.

16. Francomano D, Gottesman BL, Pijanowski BC. Biogeographical and analytical implications of temporal variability in geographically diverse soundscapes. *Ecological indicators*. 2020;112:105845. doi:10.1016/j.ecolind.2019.105845

17. Doser JW, Hannam KM, Finley AO. Characterizing functional relationships between anthropogenic and biological sounds: a western New York state soundscape case study. *Landscape ecology*. 2020;35:689-707.

18. Francis CD, Newman P, Taff BD, et al. Acoustic environments matter: Synergistic benefits to humans and ecological communities. *Journal of environmental management*. 2017;203:245-254.

19. Quinn CA, Burns P, Gill G, et al. Soundscape classification with convolutional neural networks reveals temporal and geographic patterns in ecoacoustic data. *Ecological Indicators*. 2022;138:108831. doi:10.1016/j.ecolind.2022.108831

20. Vitousek PM, Mooney HA, Lubchenco J, Melillo JM. Human domination of Earth's ecosystems. *Science*. 1997;277(5325):494-499. doi:10.1126/science.277.5325.494

21. Chapin FS III, Zavaleta ES, Eviner VT, et al. Consequences of changing biodiversity. *Nature*. 2000;405(6783):234-242. doi:10.1038/35012241

22. Rands MR, Adams WM, Bennun L, et al. Biodiversity conservation: challenges beyond 2010. *science*. 2010;329(5997):1298-1303.

23. Wrightson K. An introduction to acoustic ecology. Soundscape. *The journal of acoustic ecology*. 2000;1(1):10-13.

24. Dumyahn SL, Pijanowski BC. Soundscape conservation. *Landscape ecology*. 2011;26:1327-1344. doi:10.1007/s10980-011-9635-x

25. Mennitt D, Fristrup K, Nelson L. A spatially explicit estimate of environmental noise exposure in the contiguous United States. *The Journal of the Acoustical Society of America*. 2015;137(4_Supplement):2339-2340. doi:10.1121/1.4920539

26. Slabbekoorn H. Soundscape ecology of the Anthropocene. *Acoustics Today*. 2018;14(1):42-49.

27. Anderson GS, Rapoza AS, Fleming GG, Miller NP. Aircraft noise dose-response relations for national parks. *Noise Control Engineering Journal*. 2011;59(5):519-540. doi:10.3397/1.3622636

28. Miller NP. The effects of aircraft overflights on visitors to US National Parks. *Noise Control Engineering Journal*. 1999;47(3):112-117. doi:10.3397/1.599294

29. Schmidt WB. *The Analysis and Protection of the Natural Soundsscape in National Parks*. National Park Service, Natural Resource Stewardship and Science Directorate; 2000.

30. Urazghildiiev IR, Clark CW. Detection performances of experienced human operators compared to a likelihood ratio based detector. *The Journal of the Acoustical Society of America*. 2007;122(1):200-204. doi:10.1121/1.2735114

31. Nieto-Mora DA, Rodríguez-Buritica S, Rodríguez-Marín P, Martínez-Vargaz JD, Isaza-Narváez C. Systematic review of machine learning methods applied to ecoacoustics and soundscape monitoring. *Heliyon*. 2023;9(10). doi:10.1016/j.heliyon.2023.e20275. PMID:37790981

32. Sethi SS, Jones NS, Fulcher BD, et al. Combining machine learning and a universal acoustic feature-set yields efficient automated monitoring of ecosystems. *bioRxiv*. Published online 2020:865980. doi:10.1101/865980

33. Ibrahim AK, Zhuang H, Chérubin LM, Schärer-Umpierre MT, Erdol N. Automatic classification of grouper species by their sounds using deep neural networks. *J Acoust Soc Am*. 2018;144(3):EL196-EL202. doi:10.1121/1.5054911

34. Brunton SL, Kutz JN. *Data-Driven Science and Engineering*. 2nd ed. Cabridge University Press; 2022. doi:10.1017/9781009089517

35. Mossad OS, ElNainay M, Torki M. Deep Convolutional Neural Network with Multi-Task Learning Scheme for Modulations Recognition. In: *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*. ; 2019:1644-1649. doi:10.1109/IWCMC.2019.8766665

36. Fahlman E, Ostlund F. *Data Complexity and Its Effect on Classification Accuracy in Multi Class Classification Problems*. Degree Project, Dept. Comp. Sci. and Eng., KTH ROYALINSTITUTE OF TECHNOLOGY; 2022.

37. Paraskevas I, Potirakis SM, Rangoussi M. Natural soundscapes and identification of environmental sounds: A pattern recognition approach. In: *2009 16th International Conference on Digital Signal Processing*. ; 2009:1-6.

38. Paprocki CA, Barnard A, Dare T, Cody K, Newman P. Acoustic source separation of transient events from background noise. In: *31st International Conference on Noise and Vibration Engineering (ISMA2024)*. ; 2024.

39. Hu R, Hu K, Wang L, et al. Using Deep Learning to Classify Environmental Sounds in the Habitat of Western Black-Crested Gibbons. *Diversity*. 2024;16(8):509. doi:10.3390/d16080509

40. Verteletskaya E, Simak B. Noise Reduction Based on Modified Spectral Subtraction Method. *IAENG International Journal of Computer Science*. 2011;38.

41. Shannon CE. A mathematical theory of communication. *The Bell System Technical Journal*. 1948;27:379-423. doi:10.1002/j.1538-7305.1948.tb00917.x

42. MathWorks. *trainingOptions (Deep Learning Toolbox)*. The MathWorks, Inc.; 2024. https://www.mathworks.com/help/deeplearning/ref/trainingoptions.html

43. Kingma DP, Ba J. Adam: A method for stochastic optimization. Published online 2014.