An economic workflow for training machine learning models to classify acoustic data

Peter Glorie (1), Dr. Saeed Shehnepoor (1)

(1) Wood, Perth, Australia

Abstract – The increasing popularity and availability of AI has accelerated the development of machine learning based solutions for analysis of acoustic monitoring data. Machine Learning has been demonstrated to be effective in multiple applications including environmental noise monitoring and determining contributions to worker noise exposure. However, there is a significant cost to create effective models, due to the large data labelling requirement, iterative approach, specialist skills and tools required to review data for labelling or output performance.

While ML methods have been found to add value to large-scale or long-term monitoring programs (e.g. for above ground mining operations), applying ML to shorter term data sets (e.g. short-term construction) may not be economic due to the time and specialist skills required to develop an effective model. A significantly less labour-intensive development process would unlock the benefits of ML analysis for a broader range of use cases in industry.

This paper presents a workflow solution that bridges the gap between data science and acoustics, empowering users to harness the power of AI and analyse large data sets, without requiring coding or data science skills. The solution broadens the applications for ML analysis of acoustic data in industry by enabling accurate ML models to be developed much more efficiently.

This paper will outline how an acoustician can efficiently create accurate ML models that classify large and complex acoustic data.

1 INTRODUCTION

Advancements in computational power, the availability of data, and the emergence of high performing algorithms and technology stacks are increasing the opportunities for Artificial Intelligence (AI) to be applied to industrial problems. This includes opportunities to automate analysis of acoustic monitoring data.

Machine Learning (ML), a type of AI that involves training algorithms to recognise patterns, has been demonstrated to be effective and efficient in multiple applications. These include identifying the contribution of sources to environmental noise levels for construction or industrial operations and determining contributions to worker industrial noise exposure.

ML models are trained on large datasets that contain classified samples (referred to as ‘labelled data’) to recognise the patterns that define each class. Sufficiently trained ML models can accurately predict the class of previously unseen data (referred to as ‘inference’). ML can substantially reduce the specialist labour required for large-scale or long-term acoustic monitoring programs, since the complex analysis of the data can be largely automated to extract contributions of sources of interest to overall levels. It also makes accurate real-time analysis of incoming data practical.

Examples of how ML can add value to acoustic monitoring programs include:

- Construction noise monitoring, where ML can be used to classify emergent noise events (e.g. events that exceed an Lmax criterion): Events that are not attributed to construction noise such as bird vocalisations or sirens can be automatically filtered out, while events associated with construction activities such as impact noise from pile driving or trucks revving can be sent through to a supervisor as an alert.

- Mine site noise monitoring, where ML can be used to exclude extraneous noise from time series data and data classified as mining noise can be aggregated for assessment against Leq or L10 criteria. Exceedances of the criteria result in noise alerts. This approach has been demonstrated to reduce false alerts by more than 95% when compared to standard alerting approaches.

- Workplace noise assessment, for classification and quantification of the types of noise sources contributing to exposure within noise dosimetry data. This information can be used to focus and quantify the benefits of noise control efforts.

ML methods have been found to add value to large-scale or long-term monitoring programs, however, the benefit in applying ML to shorter term data sets is limited. This is due to the investment in time, tools and specialist skills required to train and apply ML models (e.g. construction or well drilling). Or, where the variance in data is so great that multiple models are required (e.g. workplace dosimetry data). A significantly less labour-intensive workflow would unlock the benefits of ML analysis for a broader range of use cases in industry.

2 REQUIREMENTS OF HIGH PERFORMING ML MODELS

The performance of ML models often involves a delicate balance between accuracy, dataset size, and speed. Accuracy refers to how well the model predicts or classifies data, which typically improves with larger datasets that provide more information for training and more detailed models that take longer to train. Larger datasets also require more computational resources and time to process, potentially slowing down the model’s performance. There can be significant costs involved in creating large, labelled datasets and in training detailed ML models. Balancing these factors is crucial for developing effective and economic ML models.

There are numerous approaches that reduce the cost of model creation including but not limited to improving model scalability with generalised models (e.g., one model for many monitoring locations), unsupervised data labelling (e.g. crowd-sourcing). However, these approaches often result in underperforming models with uncertainty in the outputs, resulting in either incorrect results or repeated human effort to review outputs. For example, in an industrial monitoring scenario, coarse filters aimed at detecting a class of noise may result in a larger number of false positives and lead to unnecessarily conservative curtailment of operations.

High performing ML models, that analyse large acoustic data sets more reliably than a human specialist, with acceptable accuracy, are the key to unlock efficiency benefits. Such models require supervised labelling and need to be custom-built to suit the specific application, objective, and location. To be economically viable, the workflow for creating a custom model needs to be largely automated yet flexible enough to adapt to specific applications without requiring coding or data science skills.

The essential components for production of high-performing ML models are large accurately labelled data, a diverse and balanced training dataset, and use of thoroughly tested methods for feature extraction and model architectures. Additionally, demonstrating a model’s performance requires extensive validation on unseen data that comprehensively represents the expected dataset. Sub-optimal performance is resolved by iterating the training process.

ML workflows involve numerous steps to build and deploy models. In general, for acoustics ML, the steps are the same as a typical ML process for other classification tasks, such as for images. However, unlike image classification tasks, the development of acoustic ML models presents unique challenges because a) listening to the data takes time, b) there is often no context in the data, c) many types of sound can be similar and d) sound is often a combination of different sources. Labelling therefor requires acoustic expertise to distinguish between similar sound sources and make judgements on the appropriate class, whilst maintaining consideration of the model objectives. This necessitates both specialised skills in analysis of acoustic data and specialised tools for presenting the data to the analyst. There is a limited availability of commercial tools specific to acoustics ML. Consequently, creating effective and accurate machine learning models using conventional methods is too labour intensive.

3 AUTOMATING THE WORKFLOW

The machine learning workflow includes key steps to build and deploy models. These include data collection, data cleaning/manipulation, data labelling, feature engineering, model training/testing, model validation and model deployment/inference. Wood has developed noiseAI/studio (‘Studio’), an application for acoustics ML that integrates and automates many of the steps that would typically be completed by a data scientist. Studio provides a graphical interface and tools that enables an acoustic expert to complete the full ML workflow, with a high level of control, faster and at a lower cost. The sections below describe the ML workflow within Studio and how these steps are enhanced.

3.1 Data labelling

For a person to accurately infer the type of sound present in acoustic data, it is often necessary for this data to be presented in multiple ways. Studio presents data to the user in four formats: audio spectrograms, audio playback, noise level plots and 1/3 octave spectra (refer Figure 1). These visualisations enable the user to make informed decisions for labelling sections of the data based on the time it is present in the audio.

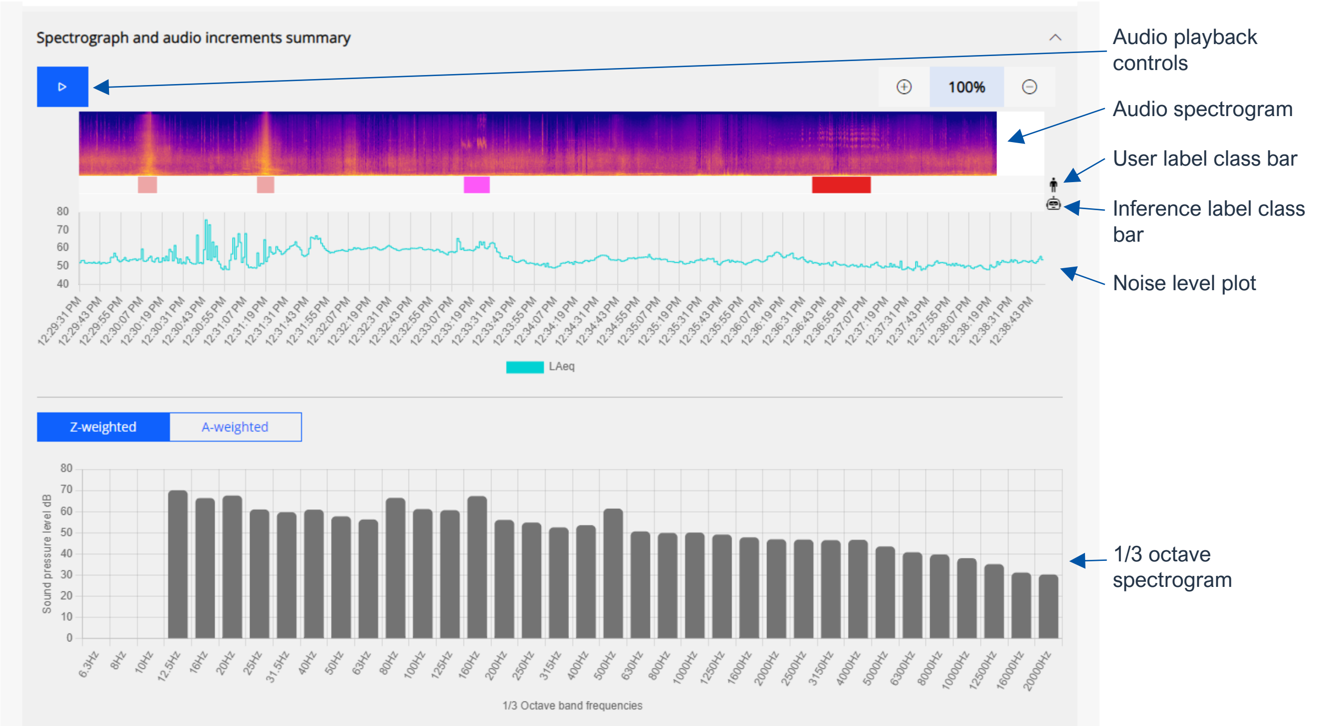


Figure 1 – Studio data visualisations

Within the application, classes for labelling can be customised and defined. Classes can be single (i.e. for binary classification) or multi class. A balanced label dataset is important for effective models, so statistics are provided on the amount of data labelled so that users can identify if a class is significantly over or underrepresented in the data set and balance the representation of classes.

Effective models also require a suitably large amount of labelled data. In general, the quantity of labelled data for each class depends on many factors including but not limited to:

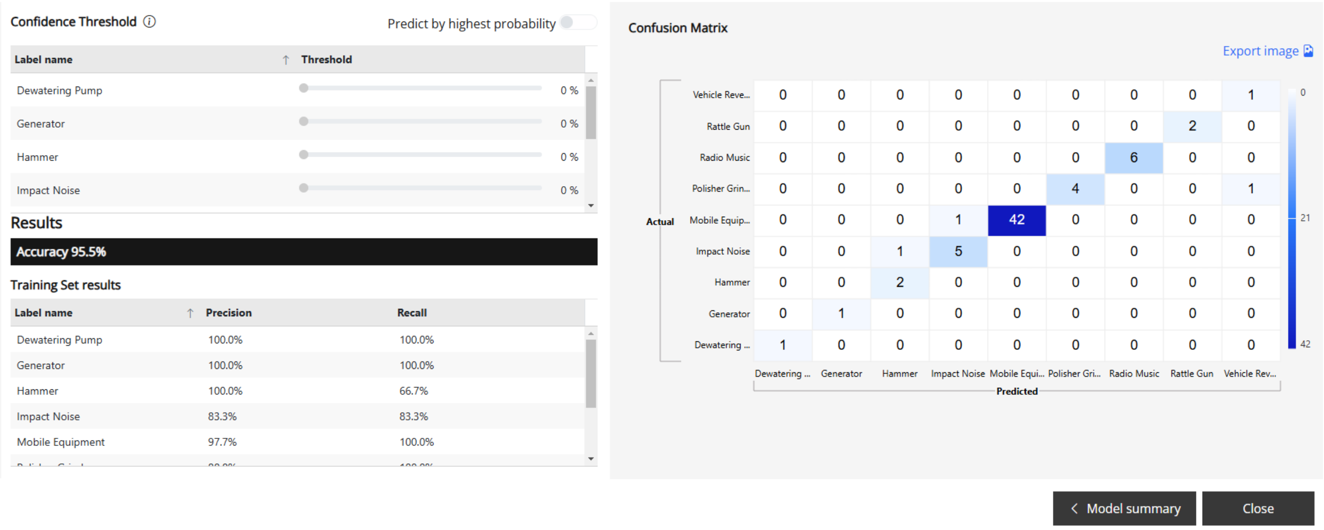
- Complexity of the classification task: Complexity increases where there are subtle differences that need to be considered for classification. Higher complexity will require more labelled data to achieve high accuracies.
- Number of classes: A larger number of classes means a larger amount of data needs to be labelled. This may be proportionally linear as each class will require a minimum amount of labelled data, or where complexity increases with the additional classes, more labelled data may be required.
- Presence of variations in the data: Data must cover the range of variation in a class, so more data is required for classes with higher variation.

Studio also enables adjustment of the frequency and time ranges for classification – key parameter definitions to ensure the relevant features that are included for classification are optimised.

3.2 Model training

Various ML models can be adapted to perform on acoustic data. Studio implements ML models that have been thoroughly tested for this purpose and can perform classification well on a diverse range of acoustic problems.

Training a model with Studio typically takes less than one minute. Following model training the user is presented with a confusion matrix for review (example provided in Figure 2), an effective initial check of the model's performance on each class that indicates areas where there is confusion between classes. Performance metrics are available also, including precision (the model's ability to predict a class correctly), recall (the model's ability to identify relevant class samples), and overall accuracy (the model's ability to identify samples correctly)



Confidence Threshold

Label name	Threshold
Dewatering Pump	0 %
Generator	0 %
Hammer	0 %
Impact Noise	0 %

Results

Accuracy 95.5%

Training Set results

Label name	Precision	Recall
Dewatering Pump	100.0%	100.0%
Generator	100.0%	100.0%
Hammer	100.0%	66.7%
Impact Noise	83.3%	83.3%
Mobile Equipment	97.7%	100.0%

Confusion Matrix

	Dewatering ...	Generator	Hammer	Impact Noise	Mobile Equi...	Polisher Gri...	Radio Music	Rattle Gun	Vehicle Rev...
Vehicle Reve...	0	0	0	0	0	0	0	0	1
Rattle Gun	0	0	0	0	0	0	0	2	0
Radio Music	0	0	0	0	0	0	6	0	0
Polisher Grin...	0	0	0	0	0	4	0	0	1
Mobile Equip...	0	0	0	1	42	0	0	0	0
Impact Noise	0	0	1	5	0	0	0	0	0
Hammer	0	0	2	0	0	0	0	0	0
Generator	0	1	0	0	0	0	0	0	0
Dewatering ...	1	0	0	0	0	0	0	0	0

Figure 2 – Model performance review in Studio, with example confusion matrix for occupational noise source classification. Values indicate number of samples predicted vs actual. High values on the diagonal are ideal as this demonstrates the model is predicting classes correctly.

At this point in the workflow, the user has the option to define the confidence threshold for each class and see the result in the confusion matrix. Where class performance needs further improvement the user has some options, including (but not limited to): reviewing label accuracy, labelling more data, consolidating classes or further performance testing of the model on further data (described in section 3.3).

Studio enables the user to train multiple ML models. The performance of different models can be compared, and different models can be used for different data sets (e.g. where different monitoring locations require separate models to achieve high performance).

3.3 Inference

Inference is where the model is used to classify unseen data. Inference may be used initially to test the model's performance on a small portion of data, as a further validation step. Once the model performance is deemed adequate, inference can be undertaken on large data sets for rapid analysis.

Studio enables the user to analyse imported data using trained models. The results from inference are presented alongside the noise level graphs and audio spectrograms, so that the user can easily review performance. Where there are multiple trained models that need to be compared, the user can promptly switch between different model inference result sets.

Where the user is running inference on a larger quantity of data (e.g. days or weeks of noise monitoring data), the task can be queued and will run in the background, so that the user can complete other tasks. The time taken to complete inference is approximately 5 to 10 minutes for a 6-8 hour sample of noise data analysed.

Outside of Studio, the model can be implemented for inference within an automated solution such as real-time noise monitoring.

4 PERFORMANCE

The accuracy achievable with machine learning model classification can be improved greatly with optimised methods, and the cost of producing these models can be reduced with streamlined processes. However, the accuracy and cost of a machine learning model also largely depends on the given set of data and the task required. For example, a basic and low-cost model may achieve perfect precision and recall in identifying the sound of an alarm, were the sound is unique and dominant, and nearly impossible to confuse with anything else.

However, if the goal is to distinguish the difference in sounds between different types of construction equipment (i.e. dozers, excavators, loaders), this would require a more complex model and a larger training data set that captures the diversity of the noise sources. Accuracy would be substantially lower compared to the alarm example as there will be many situations where these machines sound very similar. There may be situations where the noise sources are concurrent, adding more complexity and cost, and further limiting accuracy.

Generally, in classifying sounds, if a human can easily tell the difference between classes, then higher accuracy metrics are possible.

Accuracy metrics should also be tailored to the task at hand. Reverting to the construction example, an accurate classification of the type of equipment may not be important. Rather, it may be more important that the model detects construction noise when it is above the allowable limit, with a high level of recall.

The models used in Studio have achieved high levels of accuracy for noise monitoring. Table 1 provides two example scenarios, along with the model performance achieved for each. In both examples, the model objectives are achieved and the subsequent use of the models on large data sets provide results of high value.

Table 1. Examples of ML model performance achieved using noiseAI/Studio

Use case	Objectives	Model Performance	Comments
Occupational noise dosimetry in a plant maintenance workshop	Determine the contributions of noise source exposure so that noise controls can be prioritised	Performance on 1s samples across 9 source classes: • Accuracy: 97.3% Accuracy is higher when noise sources are louder and dominant.	Cumulative exposure results across a work shift are determined within 2 dB when compared to manually analysed data.
Environmental noise monitoring nearby to a mine site in a rural setting	Generate alerts when mining noise exceeds the 1-hour L_{eq} alert limit. A high level of recall is required (avoid false negatives, reduce false positives)	Performance on 1s samples for detection of mining noise: • Recall: 96.5% • Precision: 91.5% Recall is prioritised when tuning the model. Accuracy is higher when mining noise is louder and dominant.	Aggregate 1-hour L_{eq} results achieve the objectives in detection of mining noise exceedances.

5 CONCLUSIONS

The opportunities for industry to gain greater insights from the use of ML on noise monitoring data are myriad. However, existing methods for creating effective and high performing models are labour intensive, due to the large data labelling requirement, iterative approach, specialist skills and tools required to review data for labelling or review of inference performance. This significantly limits the economic applications for ML analysis of acoustic data.

noiseAI/studio provides a workflow for building, validating and applying effective machine learning models for classification of acoustic data. The entire process can be done easily and with a fast turn-around time by one person, without the need for coding or data science skills. Streamlined ML methods are implemented, that are optimised for acoustics and are able to produce accurate models for noise monitoring.

noiseAI/studio improves the economics of using AI on noise data, broadening the potential applications and use cases.

ACKNOWLEDGEMENTS

Development of noiseAI/studio was a huge and complex undertaking, shared by many within Wood including (but not limited to), Ben Turner (Perth, Studio product owner), Kane Alexander (Perth, data engineer), Ross Kilhams (UK, software development lead), Alice Macdonald (UK, senior technology QA engineer) and Cameron Smith (UK, digital solutions engineer).

Thankyou to those who helped review the paper including Matthew Pettersson (Perth, principal consultant), Samuel Turner (Melbourne, noiseAI product manager) and Jim McLoughlin (Perth, principal consultant).

Finally, development of Studio would not have been possible without significant investment. Thank you to Wood for their willingness to invest in our vision for AI everywhere.