

PREDICTION OF PROPELLER NOISE IN LOW REYNOLDS NUMBER OPERATION USING MACHINE LEARNING MODELS

Marcos de Rosa Jacinto[1], Dominik Skrna[1] & Martin Berens[1]

[1]Technische Universität Wien. Karlsplatz 13, Vienna 1040, Austria.

Abstract

Propeller noise is one of the main challenges for the acceptance of Advanced Air Mobility (AAM) aircraft in society. The necessity of high-fidelity noise simulations for the accurate prediction of noise emissions slows down development processes and increases costs. In this paper, a Machine Learning (ML) algorithm is implemented to predict propeller noise based on high-fidelity simulation and experimental data. The influence of the algorithm parameters on performance is investigated and trained models are assessed on a test dataset. When applied for the prediction of SPL spectra, the best models achieve a mean absolute error as low as 2.5 dB across the entire frequency range. Models are also evaluated on their capability to predict noise from two sets of input parameters and determine which one is louder. Accuracy of 95% is achieved on this task.

Keywords: Aeroacoustics, Propeller, Noise, Machine Learning, Artificial Intelligence.

1. Introduction

The utilization of low-altitude airspace for transportation and economical growth has been a widely discussed topic in recent years. Tang [1] highlights how Low-Altitude Economy (LAE) can improve the economic systems as a whole through different mechanisms. Examples include leading to breakthroughs in communication and sensing technologies, boosting logistics of warehousing, allowing for precise agriculture operations and offering solutions to existing terrestrial limitations on delivery of goods, be them of geographical nature or due to congestion of current infrastructure. The author also emphasizes that, although electric vehicles for LAE are more efficient than diesel-fuelled ground vehicles, the higher energy demand and current dependency on fossil energy in certain regions may lead to higher emissions of greenhouse gases, thus a clean energy supply needs to be prioritized.

In addition to challenges regarding the energy source, the social acceptance of Advanced Air Mobility (AAM) aircraft also presents hurdles for their integration into society. According to an EASA study on the societal acceptance of Urban Air Mobility (UAM) in Europe, the main concerns relate to environmental impact, safety, security and noise emissions [2]. The latter introduces additional complexity to the development of AAM aircraft: to achieve high propulsive efficiency at low speeds, several vehicle concepts rely on multiple propellers or rotors for the generation of lift and/or thrust, which are known for contributing to high levels of noise [3], [4].

Precise determination of the noise signature of an aircraft requires high-fidelity fluid simulations and experimental validation, both of which are too expensive to conduct in a conceptual design phase. With the increased availability of computational resources in recent years, novel Machine Learning (ML) based methods for predicting propeller noise are being explored with the ambition of achieving lower assessment time during early design phases. Li and Lee [5] developed an ML model based on selected rotor characteristics, namely tip Mach number, pitch angle, solidity and radius to predict its Overall Sound Pressure Level (OASPL) at chosen locations. Low-fidelity data based on XFOIL [6] and Blade Element Momentum Theory were used for training. Thurman and Zawodny [7] trained

the same model on ANOPP and ANOPP2 [8] codes to predict tonal and broadband noise during hover operation. Both studies concluded that the ML algorithm can predict noise trends observed with low-fidelity assessment.

A major limitation of low-fidelity methods is their inability to identify complex phenomena such as turbulence induced noise, blade tip vortex noise and the formation of Laminar Separation Bubbles (LSB) for propellers operated at low Reynolds number. The presence of LSB is a known issue that can lead to increases in broadband noise [9].

In this paper, the use of ML algorithms for the estimation of propeller noise in low Reynolds number operation using high-fidelity computational and experimental data is investigated.

2. Methodology

The objective of this research is to implement a neural network model to predict specific noise metrics for a propeller according to its primary geometrical parameters and aerodynamic performance.

Figure 1 illustrates an arbitrary Sound Pressure Level (SPL) spectrum for a propeller at an observer location near the rotational plane. From this graph, two main regions can be defined, one over lower frequencies in which tonal noise is dominant and characteristic peaks at the Blade Passing Frequency (BPF) and its harmonics are observed, and another, over higher frequencies, in which broadband noise prevails. To model the low-frequency region, the SPL at the first five BPF harmonics are selected. Meanwhile, for the frequency range above 1000 Hz, the SPL at one-twelfth octave band intervals are derived up to 16 kHz. The reason for banding the latter region is to reduce the statistical noise which is observed when the data is displayed at its maximum frequency resolution. The simplified representation of the acoustic spectrum amounts to a total of 54 data points, and is illustrated in Fig. 1. A round marker is used to indicate tonal noise at the first five BPF harmonics and a step plot is adopted to display the one-twelfth octave bands.

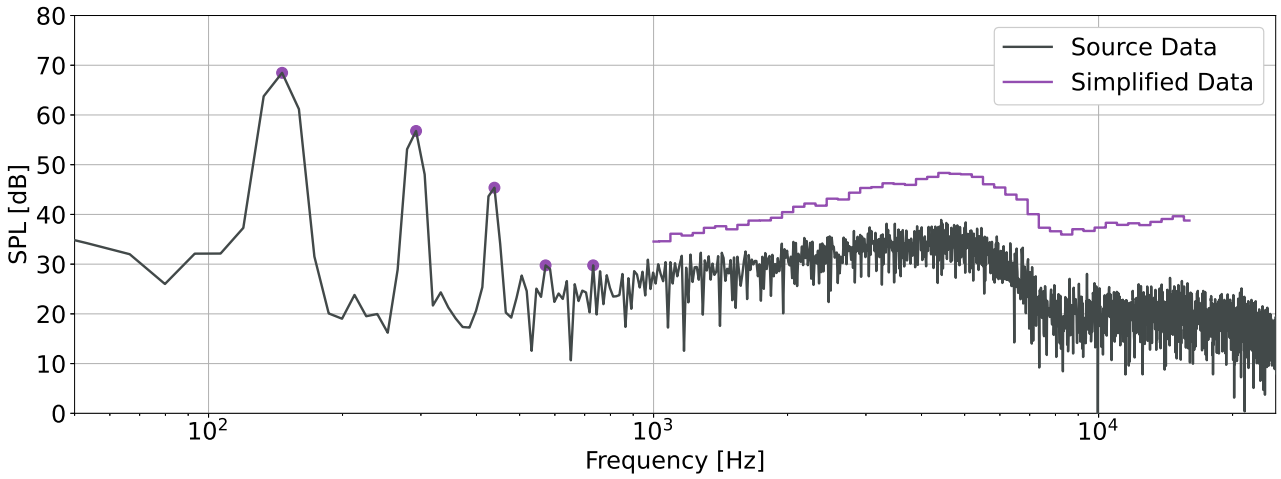


Figure 1 – Example of a propeller's SPL spectrum as in the source measurement data (black) and its simplified representation (purple) for application in a neural network.

The noise observed from a propeller depends on its geometrical and operational parameters, as well as receiver location. For the purpose of this study, the parameters below have been selected for investigation:

1. **Diameter [in]:** Determines the area on which the propeller acts on the air. Larger diameter values present a higher potential for thrust generation. At a fixed thrust condition, it leads to a lower disk loading, enabling operation at lower rotational speeds.

2. **Pitch [in]:** Theoretically represents how many inches the propeller would move forward through a solid medium after a full rotation. It's calculated using the incidence angle of the blade relative to the rotational plane at a radial location of 75% of the total radius. The blade incidence angle has a direct impact on the pressure coefficients on the blade and boundary layer development.

Lower values will lead to lower thrust output, while excessively high pitch may lead to flow separation. For fixed thrust operation, a higher pitch may allow operation at lower rotational speeds at the cost of higher torque demand.

3. **Number of Blades [-]:** Impacts the propeller's characteristic Blade Passing Frequency. For a fixed thrust output and blade chord distribution, more blades increase the propeller's solidity and decrease the disc loading, promoting reduced noise emissions.

4. **Solidity [-]:** The ratio between total blade area projected onto the rotational plane and disc area determined by the propeller's radius. It depends primarily on the number of blades, their chord distribution and the radius.

5. **BPF [Hz]:** Characteristic frequency and respective harmonics at which tonal noise is observed. It is given by the propeller's rotational speed in Hertz multiplied by the number of blades.

6. **Rotational Speed [RPM]:** Operational parameter which determines the circumferential velocity distribution along the blade according to radial position. In combination with flow properties and local chord lengths, it determines the Reynolds numbers on which the blade operates. As aforementioned, the rotational speed directly influences the BPF.

7. **Rotational Speed Relative Standard Deviation [-]:** Statistical dispersion of the rotational speed around its mean value during data acquisition. This parameter is included with the objective of allowing the model to determine if higher RPM oscillations lead to higher noise emissions due to unsteady phenomena. It also serves as identifier for the model regarding the source of the data, because samples obtained from computational simulation present an RPM standard deviation equal to zero.

8. **Thrust [N]:** Amount of force generated by the propeller in the axial direction. It is directly linked to the pressure distribution present on the blades. Higher thrust values result in higher loading noise, assuming other geometrical parameters are constant [10].

9. **Thrust Relative Standard Deviation [-]:** Statistical dispersion of thrust around its mean value during data acquisition. Present in both the computational and experimental data.

10. **Torque [Nm]:** Amount of torque generated by the propeller in the rotational axis. It is directly linked to the pressure distribution on the blades. Propellers may present similar thrust but different torque values due to a larger extent of flow separation and consequently a stronger turbulent wake, which may increase broadband noise emissions [10].

11. **Torque Relative Standard Deviation [-]:** Statistical dispersion of torque around its mean value during data acquisition. Present in both the computational and experimental data.

12. **Directivity Angle [deg]:** The position of the receiver is relevant as noise from dipole sources propagates more strongly in the rotational plane and is weaker along the rotational axis, where monopole/quadrupole noise is dominant. This parameter is used to allow the model to learn information regarding noise directivity.

The twelve parameters above serve as inputs to a neural network model, whose output corresponds to the 54 data points that represent the noise signature of a given propeller at a specific operational condition and observer location. Parameters such as propeller advance ratio and observer distance are currently not included because of limitations of the experimental test facility. Therefore, all data points are given for a propeller being operated in static condition, representative of hovering flight, with observers located at an arc of 1.8 meters from the propeller centre.
The data used for training the ML model consists of a blend of high-fidelity data obtained via computational simulations and experimental tests. Tab. 1 lists the propellers for which data was collected and Fig. 2 shows their frontal view in scale, allowing for a comparison between the different geometries.

Table 1 – Propellers for which data have been acquired and respective source.

Propeller	Diameter [in]	Pitch [in]	Blades	Solidity	Simulation	Laboratory
APC	16	10	2	0.0706	-	✓
APC	18	8	2	0.0758	-	✓
APC	18	12	2	0.0716	-	✓
Super ElectroLine	19.5	12.25	3	0.0971	-	✓
APC	20	13	2	0.0718	-	✓
Falcon	20	13	3	0.1106	-	✓
Mejzlik	22	7.4	2	0.0802	✓	✓
LFS Baseline	24	7.9	2	0.0694	✓	-
LFS 7-Series	24	7.9	2	0.0695	✓	-
NeoBlade	24	14.2	3	0.0786	✓	✓
Super ElectroLine	24	10	2	0.0747	-	✓
Falcon	24	10	2	0.0843	-	✓
Falcon	24	10	3	0.1224	-	✓
Mejzlik	24	8	2	0.0761	-	✓
T-Motor	24.2	7.9	2	0.0906	✓	✓

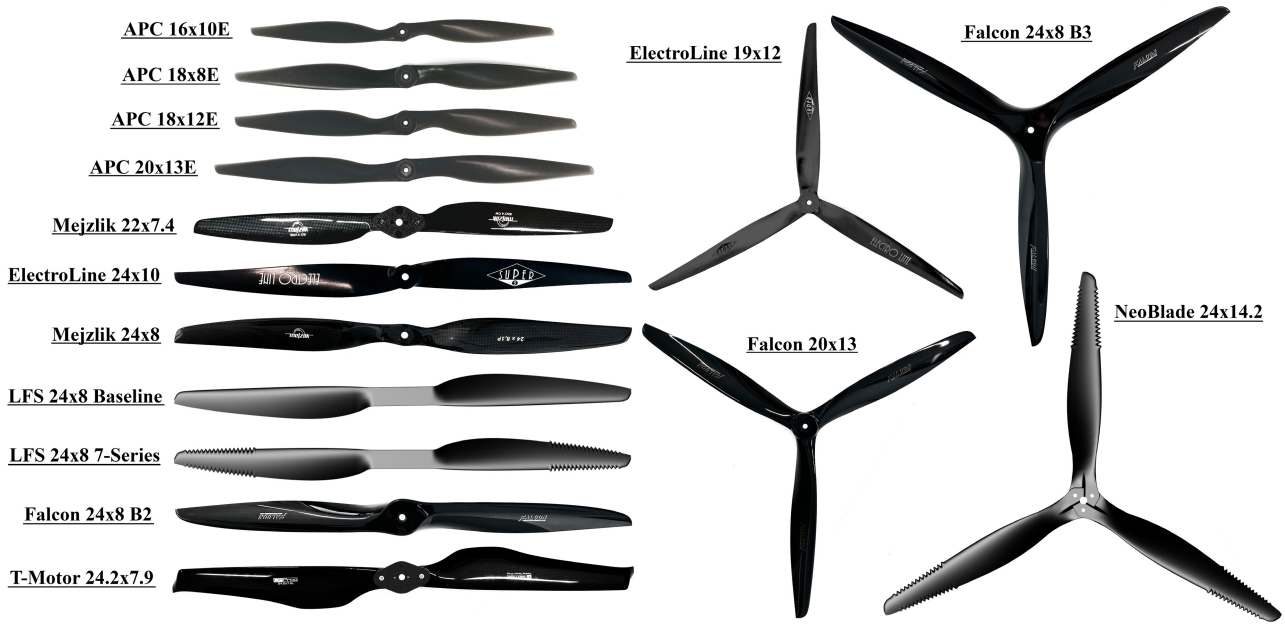
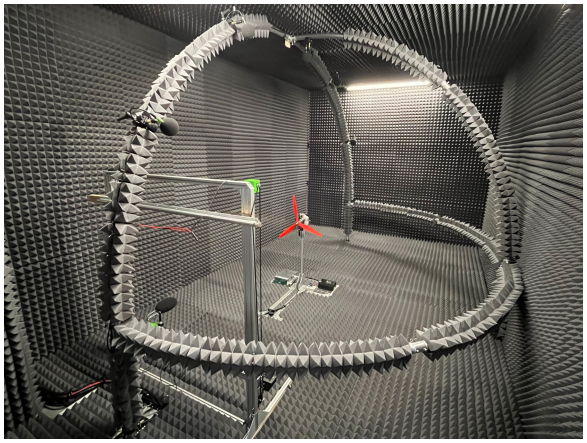


Figure 2 – Frontal view of propellers used for data collection.

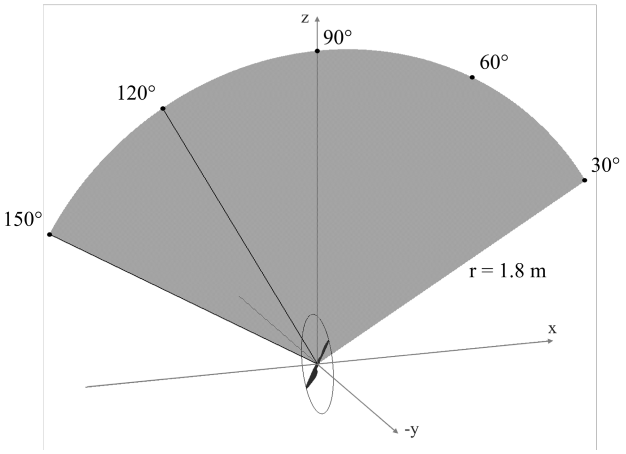
2.1 Experimental Setup

Laboratory experiments to measure aerodynamic performance and noise characteristics of propellers are conducted using a propeller test rig in a sound measurement room at the Technische Universität Wien, in which the walls are covered with foam pyramids of 50 mm base width and 80 mm height to reduce acoustic reflections (Figure 3a). Five 1/2 inch G.R.A.S. 46AE microphones are located on an arc of 1.8 m radius, centred on the propeller, in intervals of 30° from 30° to 150°. They have a frequency range of 3.15 Hz to 20 kHz and a sensitivity of 50 mV/Pa. Acoustic measurements are carried out using the Brüel & Kjær Pulse LabShop software with a resolution of 8 Hz, for the range of 8 Hz to 25.6 kHz. The acoustic pressure is averaged over 10 periods of 0.125 seconds using a Hanning window with 50% overlap, resulting in a total acquisition time of 0.6875 seconds. The data recorded is exported as a time series of acoustic pressure p. A Fast Fourier Transfer is then applied to convert the signal to the frequency domain and the acoustic pressure was converted to SPL using

Eq. 1, with reference pressure $p_{ref} = 20\ \mu Pa$.



(a) Propeller test rig available at the Aircraft Systems Research Group lab.



(b) Observer locations for experimental and computational setup.

Figure 3 – Experimental Setup.

$$L\ [dB] = 20 \cdot \log\left(\frac{p}{p_{ref}}\right) \tag{1}$$

The SPL values for the one-twelfth octave bands are calculated through the sum of the sound energy within the boundaries of each fraction octave band, as shown by Eq. 2, where L_i stands for the SPL at bands in frequency resolution and $L_{12^{th}}$ the SPL value of the new one-twelfth octave band.

$$L_{12^{th}}\ [dB] = 10 \cdot log\left(\sum_i 10^{\frac{L_i}{10}}\right) \tag{2}$$

The Tyto Robotics Flight Stand 15 is used to operate the propeller, allowing measurements of up to 150 N of thrust and 8 N.m of torque. The operational data obtained is averaged over each acoustic acquisition period and their respective standard deviations are calculated.

2.2 Computational Setup

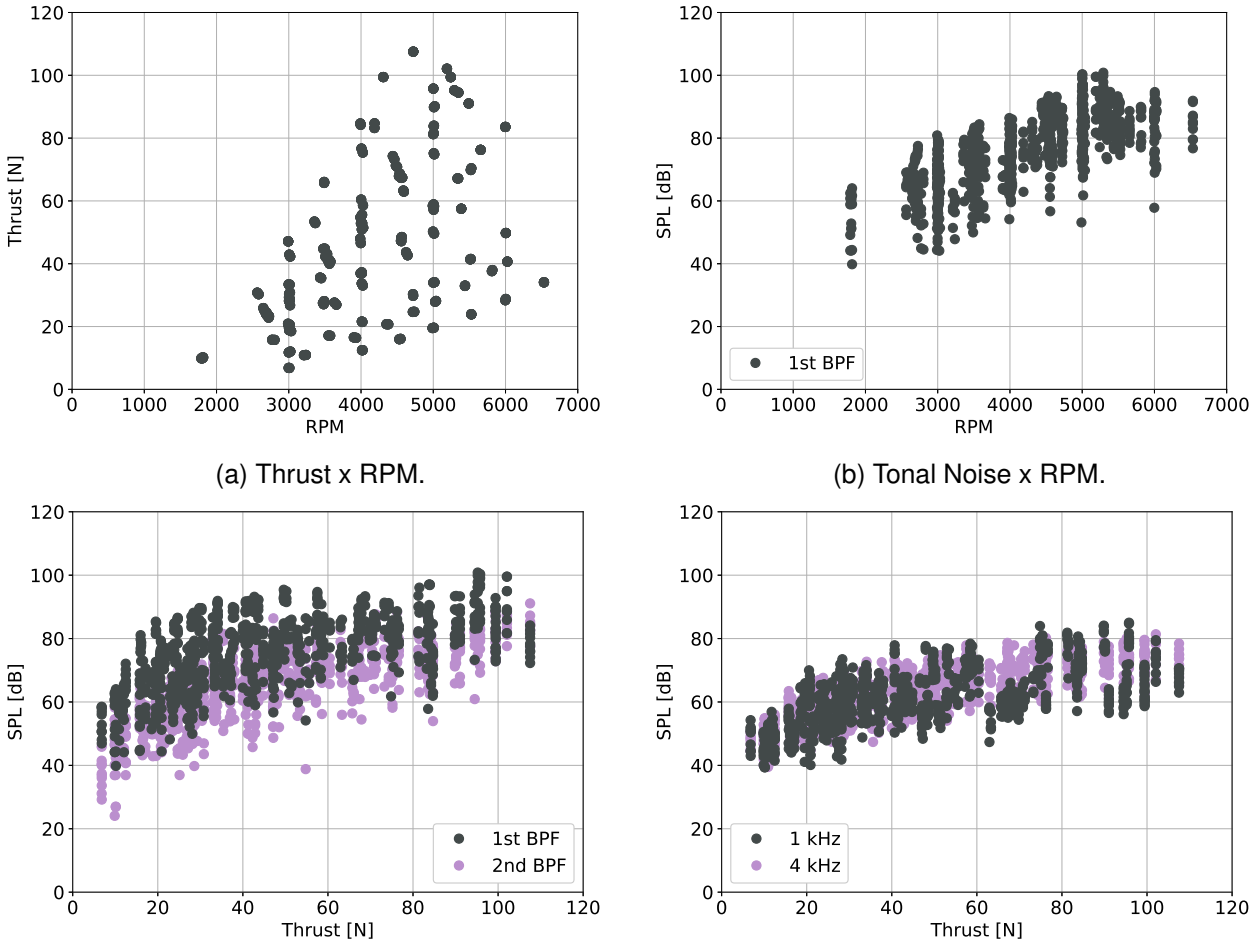
Computational simulations are carried out using the software SIMULIA PowerFLOW [11], based on the Lattice-Boltzmann Method. It performs Very Large Eddy Simulations (VLES), resolving larger turbulence flow scales while modelling unresolved ones using the renormalization group (RNG) form of the k-epsilon equations.

The setup already used by the authors in [12], [13] is adopted in the current research. It exploits the porous formulation of the Ffowcs Williams-Hawkings (FW-H) analogy to predict noise at far-field probe locations based on high-fidelity pressure fluctuations in the near field. The temporal settings allow for capturing SPL spectra in the frequency range of 13.3 Hz to 25 kHz with a resolution of 13.3 Hz. Data acquisition begins after the convergence of thrust and torque values and is carried out for a period equivalent to ten propeller rotations. Observer locations of 30° to 150° in intervals of 30° on the XZ plane at a distance of 1.8 meters from the centre of the propeller, as illustrated in Figure 3b, have been selected to provide consistency with the experimental setup.

2.3 Data Exploration and Pre-Processing

A total of 1000 samples are available, 110 of them are acquired via computational simulations and 890 through experimental tests. Figure 4 displays the relationship between selected input and output model parameters. Figure 4a shows the operational envelope contained in the data with respect to thrust output and RPM. The highest thrust values are achieved for rotational speeds around 4200 to 5200 RPM for propellers with high diameter and solidity. The points towards higher RPM values originate from smaller propellers as they have lower torque, which enables operation at higher speeds.

Figure 4b illustrates the tonal noise emission at the 1^{st} BPF versus the operating RPM. The tonal noise shows a tendency to increase until 5000 RPM and then slightly decreases, indicating a correlation between thrust and tonal noise. This is confirmed in Figure 4c, which directly compares the tonal noise at the 1^{st} and 2^{nd} BPFs with thrust. The lower boundary of the scatter increases nearly linearly with thrust, while the upper boundary shows an asymptotic behaviour towards higher thrust. It is observed that the noise emission is higher at the first BPF than in the second BPF. Finally, Figure 4d shows the relationship between thrust and high frequency noise at 1 and 4 kHz. Once again the SPL measured increases with thrust, however no significant difference can be noticed between the two frequencies analysed, with the exception of a slight decrease in noise around 60 N to 75 N at 1 kHz relative to 4 kHz.



(a) Thrust x RPM.

(b) Tonal Noise x RPM.

(c) Tonal Noise x Thrust.

(d) Broadband Noise x Thrust.

Figure 4 – Data Visualisation.

During the process of training a neural network, data is accessed for two main tasks: training and testing. During training, the model uses the data to adjust its parameters (weights and biases). During testing, the model evaluates its performance on unseen data by computing metrics such as error or accuracy, without updating its parameters. It is therefore of importance to guarantee that learning data is not used for testing and vice-versa to prevent the model from falsely indicating high predicting performance due to its bias towards the training data. To meet this constraint, the dataset is randomly split into two segregated groups, 80% of the samples are allocated to training and the remaining 20% to testing.

After the dataset is split, the input parameters are standardized, i.e. their values are scaled down to a similar range to avoid that parameters with a higher order of magnitude, e.g. RPM $\approx 10^3$, have a larger impact into the predictions than parameters with lower order of magnitude, e.g. Solidity $\approx 10^{-2}$. This is achieved by removing the population's arithmetic mean μ from each sample value x_i and dividing the result by the population standard deviation σ, as seen in Eq. 3, where z_i represents

the standardized value.

$$z_i = \frac{x_i - \mu}{\sigma} \tag{3}$$

2.4 Neural Network Architecture

Neural networks consist of multiple layers of interconnected nodes. Each node processes incoming values by applying weights and biases, and then passes the result through an activation function to produce an output that is forwarded to the next layer. This process is exemplified on Fig. 5, on which the operations of a node is displayed on the left side, and a neural network with *j* hidden layers and *k* nodes per hidden layer is illustrated on the right side.

The activation function introduces non-linearity into a neural network, enabling it to model more complex relationships. It decides whether a node should pass its output forward based on the weighted sum of its inputs and biases. In this research, the Rectified Linear Unit (ReLU) function is used, which outputs zero for negative inputs and retains positive values unchanged. ReLU offers advantages such as avoiding vanishing gradients, requiring less computational effort since it does not involve exponential calculations, and enabling sparsity in the network. [14].

As a consequence of the input and output data definition in the previous sections, the number of nodes on the input and output layers is predetermined, the former consists of 12 nodes, while the latter contains 54 nodes. The number of hidden layers and their respective number of nodes are defined through a parametric analysis. Different layouts are trained and tested, and the hyperparameters of the model with the best performance are selected.

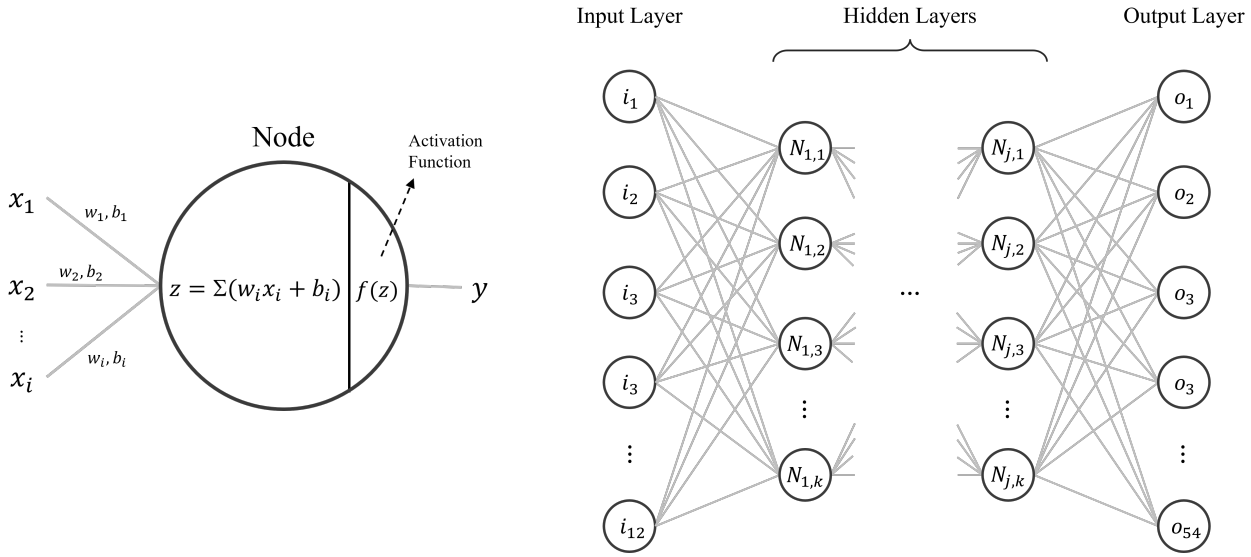


Figure 5 – Example of a hidden node functionality (left) and of a neural network layout with *12* input nodes, *j* hidden layers with *k* nodes each and *54* output nodes (right).

The predictions made by the model during training are compared to actual target values via loss functions to determine the errors between them. Based on the loss values, an optimizer updates the model's weights and biases for the next iteration. This process is repeated for a chosen number of iterations, i.e. epochs. For a model that performs regression, the most common loss functions adopted are Mean Absolute Error (MAE), Mean Squared Error (MSE) and Huber. The first one computes the absolute difference between true value y_i and predicted value $\hat{y}_i$, as displayed in Eq. 4. While it is more robust to outliers, it can result in slower convergence due to its constant gradient. The MSE calculates the square of the difference between y_i and $\hat{y}_i$ (Eq. 5), which can lead to faster convergence, but is more sensitive to outliers because of the quadratic term. The Huber loss serves as a compromise between MAE and MSE, it computes the error depending on the offset of the predicted value to the real value and a user defined δ, as shown in Eq. 6. The three loss functions are tested in this research to determine which one leads to better model performance.

$$l_{MAE} = |y_i - \hat{y}_i| \tag{4}$$

$$l_{MSE} = (y_i - \hat{y}_i)^2 \tag{5}$$

$$l_{Huber} = \begin{cases} 0.5 \cdot (y_i - \hat{y}_i)^2, & \text{if } |y_i - \hat{y}_i| < \delta \\ \delta \cdot (|y_i - \hat{y}_i| - 0.5 \cdot \delta), & \text{otherwise} \end{cases} \tag{6}$$

The optimizer chosen is *Adam*, an algorithm for first-order gradient-based optimization of stochastic objective functions [15]. The magnitude of the corrections the optimizer applies to the weights and biases can be controlled through the hyperparameter *Learning Rate (LR)*, which is included in the investigation of model settings. While a higher LR may lead to faster convergence, it might also cause instability during training and overshooting of optimal minima, which in consequence might lead to failure in convergence. Lower values of LR promote a more stable training with slower convergence, however excessively low LR may prevent convergence altogether. Therefore, the selection of an appropriate LR is fundamental to guarantee that convergence is achieved during training.
Another hyperparameter included in the training settings study is the *batch size*. It determines how many data samples are used to determine weights and biases corrections. If the batch size is set to a value n_{batch}, the errors over n_{batch} samples are taken into account before applying corrections to the model. Adopting low batch sizes promote better generalization, while large batch sizes may lead to higher stability in the training process due to more stable gradients.
In this research, a 5-fold cross validation approach is used to train the models. As shown in Fig. 6, the training dataset is randomly divided into 5 groups, from which 4 groups are used for training and the remaining group is used for validation. In the next step, a different group is selected for validation, while the remaining 4 groups are used for training. This is repeated until every group has been assigned as the validation group. At the end of this process, five models of identical settings are trained and their performance is compared. On a well distributed dataset, models with good hyperparameters are expected to have a low average error and deviation across the folds. The best model from the cross validation training is assessed using the test dataset and the results are used for the final evaluation of the hyperparameters.

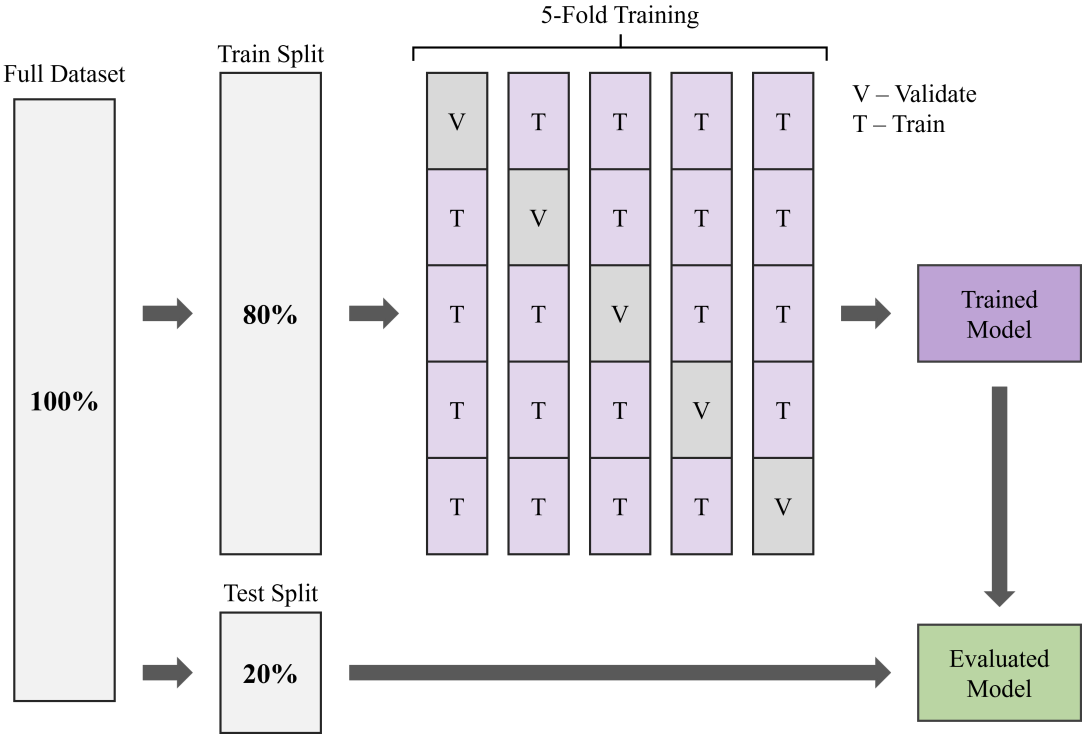


Figure 6 – Illustration of 80/20 dataset split and 5-fold training approach.

The training algorithm has been implemented in Python using the package *PyTorch* [16] to set up the neural network architecture. The models are trained in the Austrian Scientific Computing (ASC) cluster using a Xeon Platinum 8174 CPU with 24 cores.
In total, 252 neural network layouts have been trained as part of the hyperparameters investigation in a two-phase approach. In the first one, 216 architectures are trained to identify trends in performance. In the second phase, the hyperparameters are adjusted accordingly and 36 new architectures are introduced. Table 2 lists the hyperparameters investigated in each phase.

Table 2 – Hyperparameters investigated.

Hyperparameter	Phase 1	Phase 2
Number of epochs	500, 800	1000
Batch size	80	16, 32, 64, 128
Number of hidden layers	2, 3, 4, 6	4, 6, 8
Size of hidden layers	64, 128, 256	64, 128, 256
Optimizer	Adam	Adam
Learning rate	0.01, 0.001, 0.0001	0.01
Loss criterion	MAE, MSE, Huber	Huber

2.5 Model Evaluation

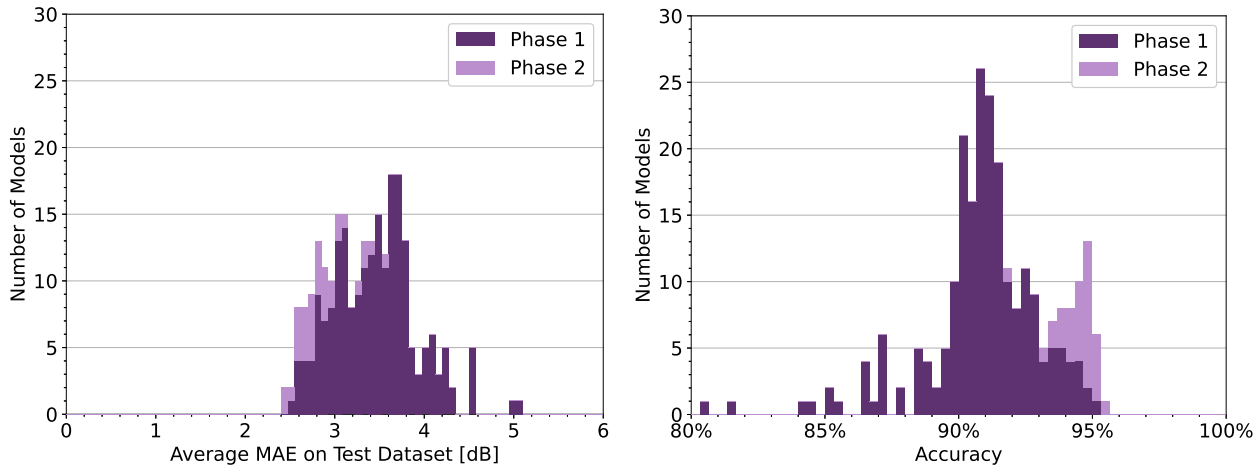
The final model selection follows two evaluation criteria, best statistical performance on the regression task and highest accuracy on a classification task. In the former, models are ranked according to their average MAE and standard deviation across all samples in the test dataset. For the latter, the test samples available are paired with one another and the models are used to predict which of them has the higher noise emissions. On a test dataset with n samples, a total of $0.5 \cdot (n^2 - n)$ unique pairs can be determined for the evaluation of accuracy. The decision for the loudest sample is made based on their acoustic energy across their SPL spectra.
To measure the certainty of the classification predictions, equation 7 is introduced, in which the confidence value Δ is given by the difference between the mean value of the SPL spectra of samples A and B, $(\hat{\mu}_A - \hat{\mu}_B)$, multiplied by the sign of the the difference between the real spectra, $(\mu_A - \mu_B)$. When the predicted and real data agree on which sample is louder, the equation returns a positive value. If the prediction differs from the real data, Δ is a negative value. Finally, a larger absolute value of Δ indicates that the model has a higher confidence on its classification.

$$\Delta = (\hat{\mu}_A - \hat{\mu}_B) \cdot \frac{(\mu_A - \mu_B)}{|\mu_A - \mu_B|} \tag{7}$$

3. Results

An overview of all models' performances is visualised on Fig. 7 through stacked histograms of the mean loss on the validation dataset for the regression task (a) and of the accuracy for the classification task (b). The regression histogram shows that most models achieve an MAE between 3 and 4 dB, while the best of them present errors of around 2.5 dB. It is also seen that the models trained in the second phase achieve a better performance on average, however the lowest error is improved only marginally. Regarding the accuracy, two peaks are observed, the highest one at around 91% corresponds to the mode of models trained in the first phase. The lower peak at 95% represents the mode on the second phase. Similarly to the regression, the models in the second phase achieve better performance on average, with the highest accuracy showing a slight increase.
In Fig. 8, the performance of the best models trained on the regression and classification tasks is displayed. Figure 8a displays violin plots for the highest ranking models according to the lowest MAE on the validation dataset. Model 238 achieves the lowest test error of 2.445 ± 0.779 *dB*, a narrower distribution relative to the other models visualized, and smaller outliers. Figure 8b shows that Model 243 achieves the highest performance on the classification task, with an accuracy of 95.37% over 19800 samples. Model 227 displays excellent performance on both criteria, with the third lowest test

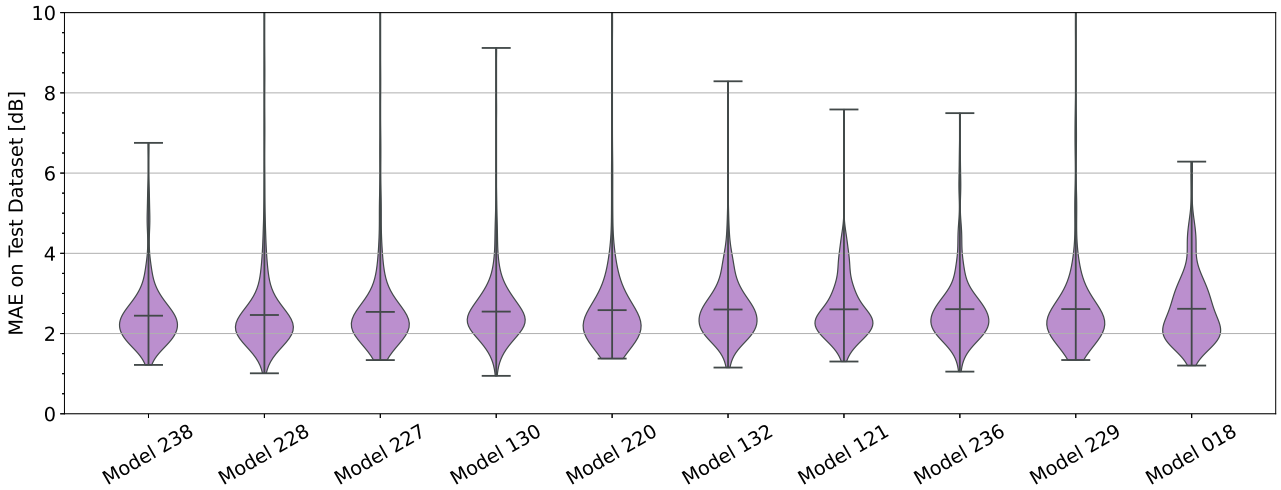
error at 2.540 ± 1.089 *dB* and the fourth best accuracy at 95.22%. From this image, it is noticeable that the top 10 models on the classification task have nearly identical performance, with differences in the order of tenths of a percent.



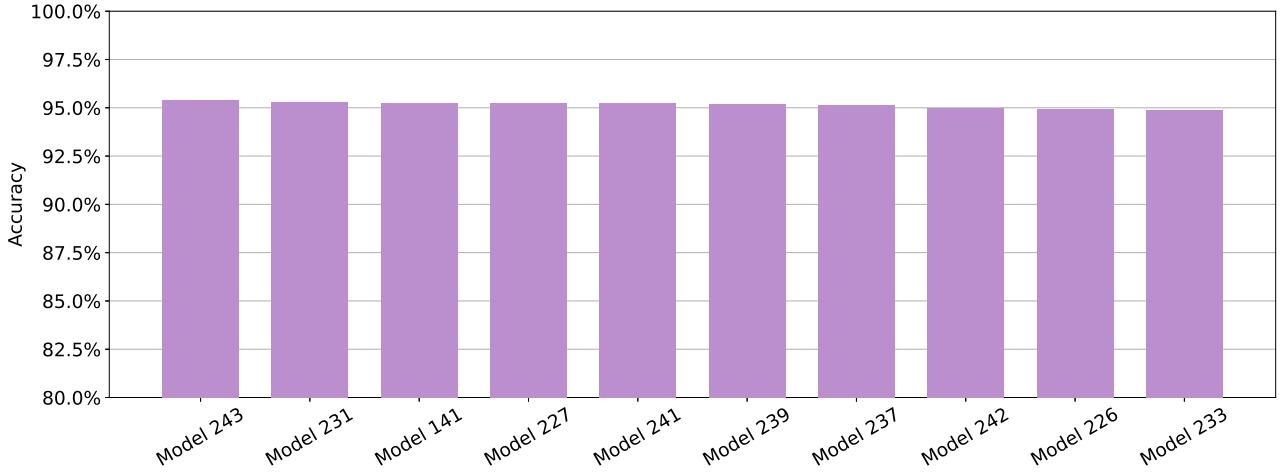
(a) Regression task.

(b) Classification task.

Figure 7 – Histograms of models' performances.



(a) Best models on regression task.



(b) Best models on classification task.

Figure 8 – Best models according to performance on respective task.

The configurations for the top three models in each task are available on Tab. 3. It can be inferred that the most promising neural network hyperparameters consist of higher numbers of epochs and hidden layers, a learning rate of 0.01, and the Huber loss function. The best models showed batch size values in the medium range of the investigated domain, and good results have been obtained for all sizes of hidden layers in this study, indicating a lower dependency on the latter.

Table 3 – Configuration of the best performing models.

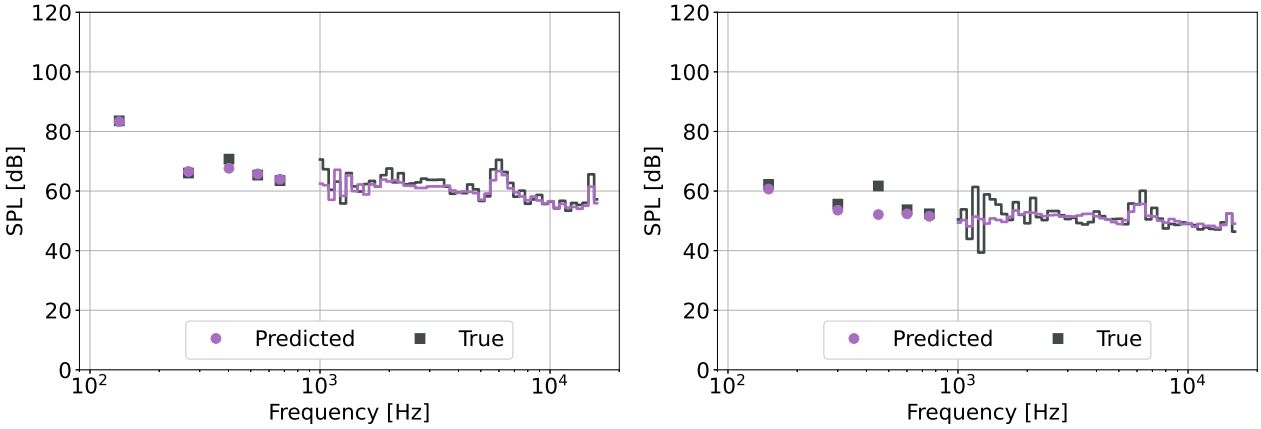
Parameters	Mod. 238	Mod. 228	Mod. 227	Mod. 243	Mod. 231	Mod. 141
Number of epochs	1000	1000	1000	1000	1000	800
Batch size	64	32	32	64	32	80
Hidden layers	6	4	4	8	6	6
Size of hidden layers	64	256	128	256	256	128
Optimizer	Adam	Adam	Adam	Adam	Adam	Adam
Learning rate	0.01	0.01	0.01	0.01	0.01	0.01
Loss criterion	Huber	Huber	Huber	Huber	Huber	Huber
Mean Validation Loss	2.445	2.463	2.540	2.727	2.888	2.952
Standard Deviation	0.779	1.034	1.089	1.196	1.343	1.094
Accuracy	94.85%	94.72%	95.22%	95.37%	95.27%	95.25%

3.1 Regression Task

In this section, the sound spectra predictions of model 238, the best performing model on the regression task, are assessed. For this purpose, Fig. 9 shows a comparison between true and predicted data for test samples in which the respective error lies within the first quartile (a), second quartile (b), third quartile (c) and fourth quartile (d) of all samples. Those represent different propellers, operational conditions and observers, illustrating the generalisation capability of the model. In Fig. 9a, it is observed that the predicted values nearly match the experimental data, capturing well the noise emissions both in the tonal and broadband noise regions. Figures 9b and 9c show slightly more dispersion either in the tonal noise or in the broadband region, but overall still serve as good predictions of the actual noise emissions. At last, Fig. 9d displays significant deviations after the second BPF and shows that the model underpredicts the broadband noise for this sample.

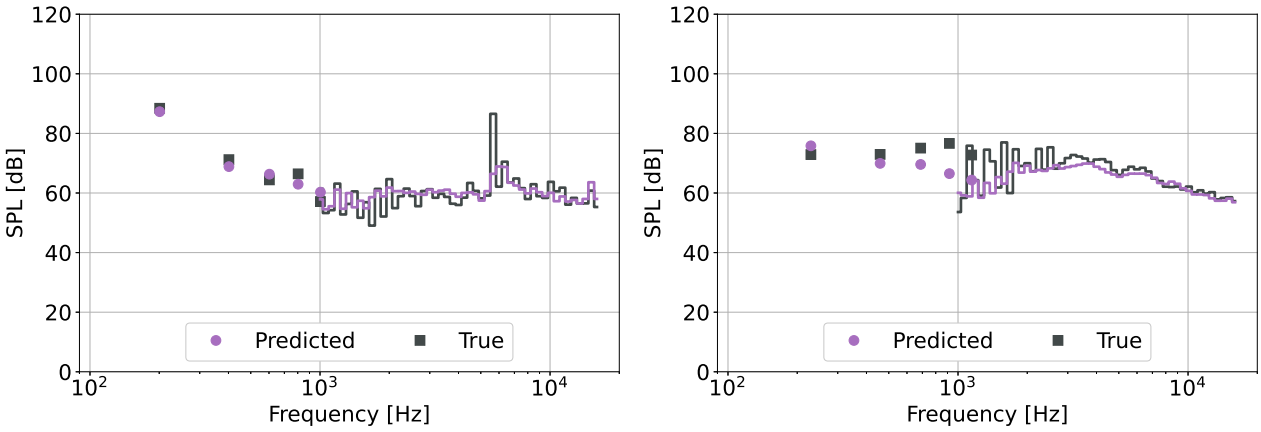
It must be noted that the spectral data obtained experimentally contain biases which limit the performance of the model. For example, in Fig. 9c there's a dominant tone at approximately 5600 Hz, which is due to motor noise at this rotational velocity and not due to propeller noise sources. The frequency resolution at which the data is recorded also presents itself as a limitation for the exact determination of tonal noise in cases where the BPF harmonics are not an integer multiple of the resolution. For example, if a propeller has a BPF of 164 Hz and the data acquisition frequency resolution is 8 Hz, the acoustic energy at the 1^{st} BPF will be split into the bands at 160 Hz and 168 Hz, leading to an under prediction of the tonal noise. Despite these limitations, the model shows satisfactory performance on the test dataset. A particular strength of the model is that it always predicts the noise at the first BPF to be higher than at subsequent harmonics. This agrees with theoretical knowledge of propeller tonal noise emissions.

To investigate the performance on model 238 relative to selected input parameters, Fig. 10a displays a scatter of the model's MAE versus the thrust generation for each sample on the test dataset. The figure also differentiates between samples obtained computationally or experimentally. It is apparent that for the experimental samples, the MAE has a lower spread below 80 N, while above this threshold, a slightly higher dispersion is observed. One reason for this is the larger amount of samples in the 0 to 80 N range, which increases the accuracy of the model in this region. From Fig. 10a, it also becomes clear that the highest errors are associated to samples obtained computationally. This result can be linked to the lower number of samples gathered via computational simulations, which hinders the model's performance on those. In Fig. 10b, the MAEs of model 238 relative to observer angles are displayed. The model has a consistent performance across the five locations investigated and, therefore, has been able to identify directivity patterns.



(a) Falcon 24x10 B2 - RPM 4027 - Obs. 90°.

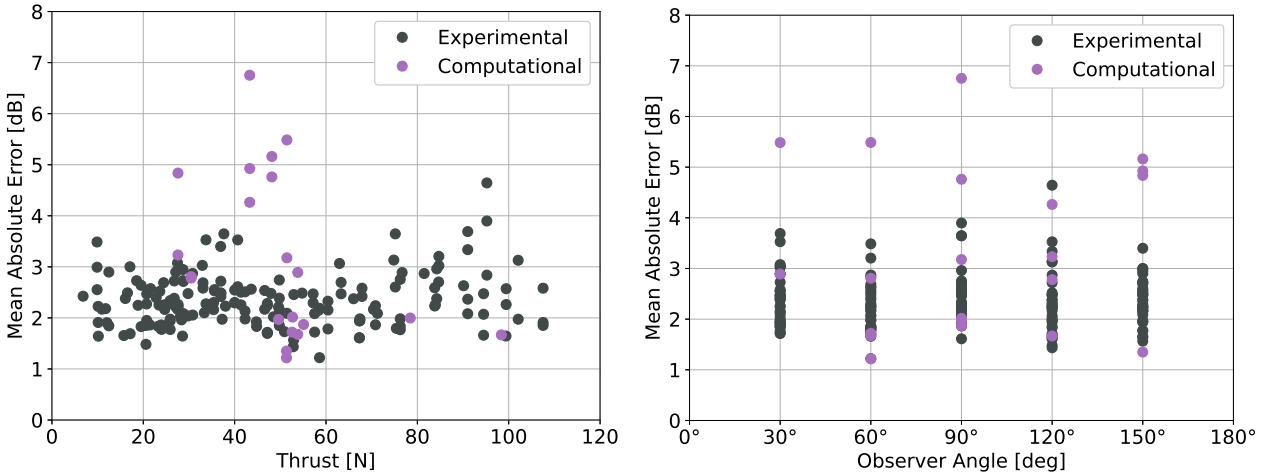
(b) ElectroLine 19x12 - RPM 3002 - Obs. 60°



(c) APC 18x8E - RPM 6025 - Obs. 90°.

(d) Falcon 20x13 - RPM 4586 - Obs. 30°.

Figure 9 – SPL spectra comparison for Model 238.



(a) Performance relative to propeller thrust.

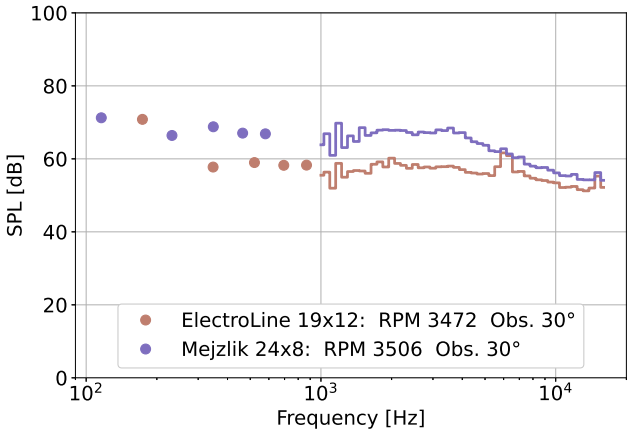
(b) Performance relative to directivity.

Figure 10 – Performance of model 238 relative to inputs.

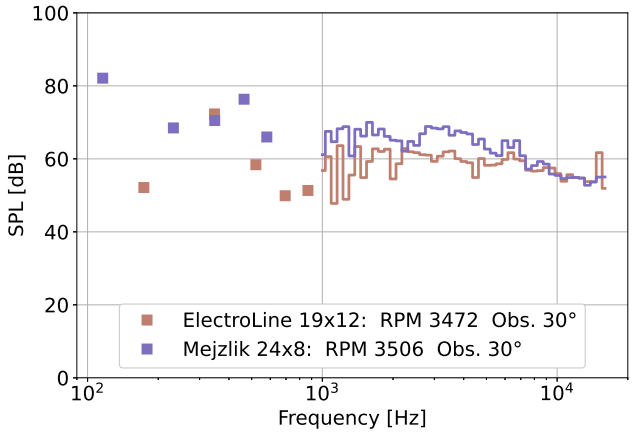
3.2 Classification Task

To assess the SPL predictions of the model which performed best on the classification task, Fig. 11 compares the output of Model 243 against the true noise spectra for three pairs of samples. The first (ID 9256) represents two distinct propeller geometries operating at similar rotational speeds. The second pair (ID 10663) consists of propellers of different geometries operating at a similar thrust level. Finally, the last pair (ID 15884) is composed by two propellers of the same diameter operating

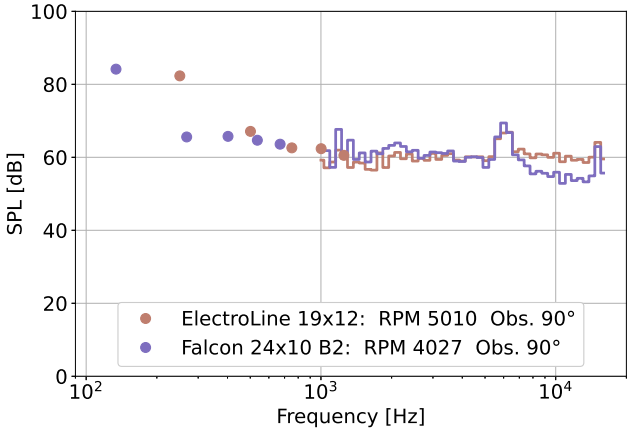
at similar rotational speeds.
Figures 11a and 11b show the predicted and actual output, respectively, for pair 9256. In this example, it is clear that the sample for the Mejzlik 24x8 propeller at 3506 RPM for an observer at 30° is louder. Although the model does not capture well the exact tonal and broadband noise present in the measured data, it is able to correctly indicate which sample emits more sound. Particularly of interest is the SPL at the 1^{st} BPF for the ElectroLine 19x12 sample. The laboratory data in Fig. 11b indicates it has peak at around 52 dB, which is uncharacteristically lower than the 2^{nd} and 3^{rd} BPFs, thus this is assumed to be an outlier. According to model 243, however, the SPL at the 1^{st} BPF for this propeller sample is predicted at 72 dB, a value that is more realistic.



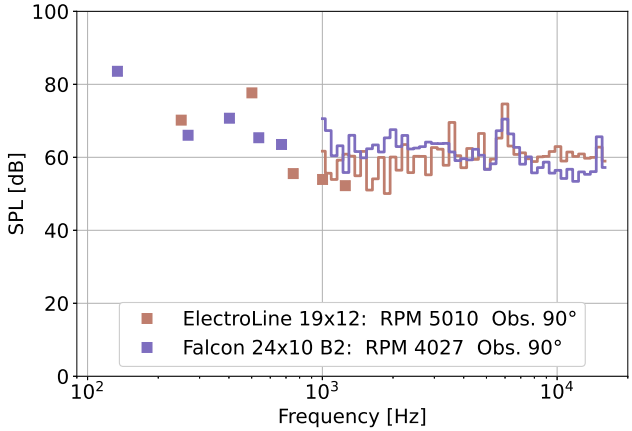
(a) Predicted values - Pair 9256.



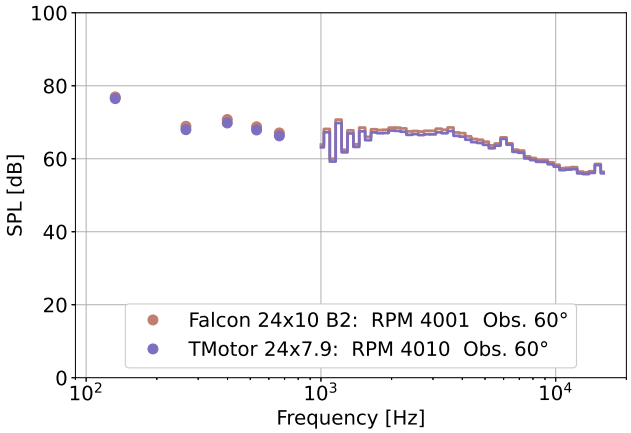
(b) True values - Pair 9256.



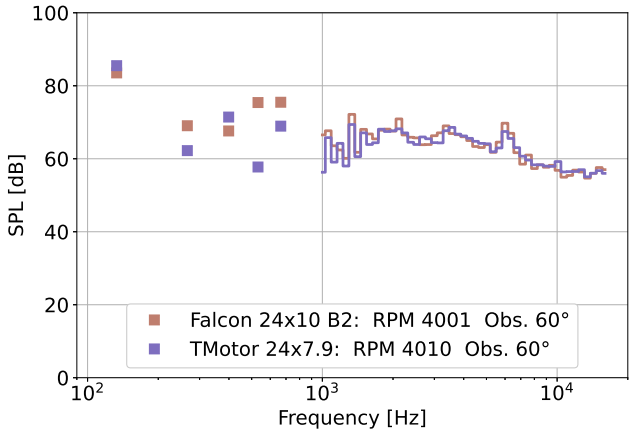
(c) Predicted values - Pair 10663.



(d) True values - Pair 10663.



(e) Predicted values - Pair 15884.



(f) True values - Pair 15884.

Figure 11 – SPL spectra comparison for Model 243.

In Figs. 11c and 11d the samples in pair 10663 are investigated. Model 243 shows an overall good regression for the noise output of the two-bladed propeller Falcon 24x10 at 4027 RPM, however, it tends to overpredict the noise emissions for the ElectroLine 19x12 at 5010 RPM. It can be seen that the predicted SPL at the BPF harmonics for both samples are similar, this could have been caused by the fact that the propellers are being operated at the same thrust output. In spite of this, the broadband trends of each sample have been maintained, for example, the Falcon 24x10 has lower noise until 4 kHz, but higher sound emissions after 6 kHz.
Finally, for Figs. 11e and 11f, the model predicts nearly identical SPL spectra for both samples in pair 15884. This is expected considering the propellers have similar geometrical and operational parameters. The experimental data confirm that those samples have comparable noise emissions, especially in the broadband region, while the main differences lie in the BPF harmonics.
Table 4 shows the confidence values for the three pairs in Fig. 11, as well as the model's predicted and true classifications on which sample is louder. It is seen that for pair 9256, the model computes a high positive Δ, indicating correct prediction with good confidence. For pair 10663, the model has a Δ equal to -0.14, therefore, it incorrectly predicted which sample is louder with a low absolute confidence value. This indicates that the model predicted similar overall noise levels for both samples, although the shape of the SPL spectra is significantly different between them, as seen in Fig. 11d. With regard to pair 15884, a Δ of 0.74 is found, therefore the prediction made by the model is correct, however, the low confidence value signals that the samples have a similar noise output.

Table 4 – Confidence values for predicted classifications of louder sample by Model 243.

Pair	Sample A	Sample B	Pred.	Truth	Δ
9256	ElectroLine 19x12 - RPM 3472	Mejzlik 24x8 - RPM 3506	B	B	6.82
10663	ElectroLine 19x12 - RPM 5010	Falcon 24x10 B2 - RPM 4027	A	B	-0.14
15884	Falcon 24x10 2B - RPM 40001	TMotor 24x7.9 - RPM 4010	A	A	0.74

4. Conclusion

The application of neural networks to predict noise emissions from propellers in low Reynolds number operation is investigated in this paper. It expands on the available literature by including high-fidelity data from computational simulations and laboratory measurements in the machine learning training process to address limitations in noise prediction inherent to low-fidelity acoustic methods. Several neural network configurations are assessed in a hyperparameter study, and the best models are selected for further analysis on their potential to correctly estimate a propeller's SPL spectra at a given operational condition and directivity angle, and on their potential to compare two propeller samples and correctly classify which of them has higher noise emissions.
The best models trained are able to predict the SPL spectra of propellers with a mean absolute error as low as $2.5\ dB$ across the frequency spectrum investigated, showing a good match for tonal noise and broadband noise tendencies. The models showed good generalisation, and outliers observed in the real data were not present in the predicted data. The results indicate a loss of performance in samples derived from computational simulations, which is linked to the fact that the dataset is unbalanced, as only 11% of the samples represent computational data. The models also showed satisfactory performance in the task of comparing propeller samples to determine which one is louder. The majority of models achieved an accuracy of over 90% on this task, while the best ones correctly classified the samples in over 95% of the cases.
These results indicate that ML methods have the potential to accurately predict propeller noise. Neural network models could be employed, for example, in the early design phase of AAM aircraft to estimate propeller noise according to its operational conditions and serve as a tool to aid in propeller selection. The data in this paper is limited to propellers that achieve up to 110 N of thrust, consequently, the models are expected to have poorer performance outside this boundary. Therefore, to achieve a mature stage of usability, training models on a larger and more diverse dataset is highly recommended to expand the design range on which they can be applied.

The research presented in this paper could be further complemented by a feature importance analysis of the input parameters. This would aid in the identification of which inputs are more or less dominant for the noise prediction. The available dataset could also be extended by including more comprehensive parameters, such as chord and twist distributions along the blade, and airfoil coordinates.

5. Contact Author Email Address

Author: Marcos de Rosa Jacinto, PhD candidate at TU Wien
E-mail: marcos.jacinto@tuwien.ac.at

6. Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.

Acknowledgements The Aircraft Systems Research Group at TU Wien is partially funded by the Austrian Federal Ministry for Innovation, Mobility and Infrastructure (BMIMI) by means of the Endowed Professorship for Innovative Aviation Technologies. This research contains data developed through the FFG Take-Off project *NeoBlade*. The computational simulations to obtain propeller data and the training of the machine learning models within this research have been conducted with support of the Austrian Scientific Computing (ASC) infrastructure. The authors express their gratitude towards Joy Haegemans, PhD candidate at the Katholieke Universiteit Leuven, for her support on the implementation of the neural network algorithm and for sharing her knowledge on the topics of Machine Learning and Artificial Intelligence.

References

[1] Tang, J., "Can low altitude economy development bring economic and environmental dividends?" *Journal of Air Transport Management*, vol. 131, Mar. 2026.

[2] EASA, "Study on the societal acceptance of Urban Air Mobility in Europe," European Union Aviation Safety Agency, Tech. Rep., 2021.

[3] Alabi, B. N. T. et al., "Breaking the sound barrier: Investigating noise challenges in the development of Advanced Air Mobility," *Transportation Research Record: Journal of the Transportation Research Board*, Dec. 2025.

[4] Bauranov, A. and Rakas, J., "Designing airspace for Urban Air Mobility: A review of concepts and approaches," *Progress in Aerospace Sciences*, vol. 125, Aug. 2021.

[5] Li, S. and Lee, S., "A machine learning-based fast prediction of rotorcraft broadband noise," in *AIAA AVIATION 2020 FORUM*, American Institute of Aeronautics and Astronautics, Jun. 2020.

[6] Drela, M., "Xfoil: An analysis and design system for low reynolds number airfoils," in *Low Reynolds Number Aerodynamics*, T. J. Mueller, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 1989.

[7] NASA Langley Research Center, Thurman, C., and Zawodny, N., "Aeroacoustic characterization of optimum hovering rotors using Artificial Neural Networks," in *Proceedings of the Vertical Flight Society 77th Annual Forum*, Virtual Conference: The Vertical Flight Society, May 2021, pp. 1–10.

[8] *Aircraft NOise Prediction Program*. NASA Langley Research Center.

[9] Grande, E., Ragni, D., Avallone, F., and Casalino, D., "Laminar separation bubble noise on a propeller operating at low Reynolds numbers," *AIAA Journal*, vol. 60, no. 9, pp. 5324–5335, Sep. 2022.

[10] Made, J. E. and Kurtz, D. W., "A review of aerodynamic noise from propellers, rotors, and lift fans," 1970.

[11] *PowerFLOW 2025 User's Guide*. Dassault Systèmes Simulia Corp, 2025, p. 135.

[12] Skrna, D., De Rosa Jacinto, M., and Berens, M., "Aeroacoustics investigation of propeller leading-edge tubercles applied to Advanced Air Mobility," in *ICAS proceedings 34th congress of the International Council of the Aeronautical Sciences*, Florence, Sep. 2024.

[13] Skrna, D., De Rosa Jacinto, M., and Berens, M., “Investigation of the combined application of leading-edge tubercles and trailing-edge serration on AAM propellers in terms of aeroacoustics using Lattice-Boltzmann Method,” in *AIAA SCITECH 2025 Forum*, Orlando, FL, Jan. 2025.

[14] Glorot, X., Bordes, A., and Bengio, Y., “Deep sparse rectifier neural networks,” in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, G. Gordon, D. Dunson, and M. Dudík, Eds., ser. Proceedings of Machine Learning Research, vol. 15, Fort Lauderdale, FL, USA: PMLR, Nov. 2011, pp. 315–323.

[15] Kingma, D. P. and Ba, J., *Adam: A method for stochastic optimization*, Jan. 2017.

[16] *PyTorch*, version 2.12.0.