Application of Machine and Deep Learning Techniques for Acoustic Load Predictions on the VEGA-E Launcher

[1]De Paola*; [1]Camussi; [1]Di Marco; [1]Stoica; [1]Capobianchi; [2]Marongiu; [2]Beretta; [2]Paglia;

[1]Department of Engineering, Roma Tre University, Via Vito Volterra, 62, 00146 Rome, Italy
[2]AVIO SpA, Via Latina, 00034, Colleferro, Italy

Keywords: Machine Learning, Deep Learning, Launch Vehicles

Abstract. Artificial intelligence (AI) can be used to optimize the prediction of pressure fluctuations over the external surfaces of aerospace launchers, minimizing the need for extensive wind tunnel tests. In the present research, various machine learning (ML) and a deep learning (DL) approach were tested and validated on the VEGA-E launcher. These methods were applied to data expressed both as pressure signals and in the frequency domain, represented as spectra. The techniques evaluated included decision trees, Gaussian Process Regression (GPR), Support Vector Machines (SVMs), artificial neural networks (ANNs), linear regression, and ensemble methods such as bagged and boosted trees. The models were trained using experimental data from an extensive wind tunnel test campaign, conducted to support the design of the VEGA-E launcher and to provide wall pressure fluctuations under various configurations. The primary objective of this study was to identify the most suitable algorithm capable of processing such complex datasets efficiently and providing reliable predictions for both time-domain and frequency-domain data. Different statistical indices, including the root mean square error (RMSE) and the correlation coefficient (R-squared), were used to evaluate the performance of the tested models. Results highlight the potential of ML and DL-based models to optimize experimental campaigns and enhance the design process for aerospace launch vehicles.

1. Introduction

During the ascent phase, compressible flows over aerospace vehicles generate pressure fluctuations that induce vibrations, posing risks to payload integrity and accelerating structural fatigue. Predicting these loads is crucial for optimizing launcher design and minimizing costs. Traditionally, wind tunnel testing and semi-empirical models have been employed, but their high cost and time-consuming nature drive the search for alternative predictive methods. Extensive studies have analyzed wall-pressure fluctuations, turbulence-driven phenomena, and aerodynamic loads through statistical models and experimental campaigns. Notable contributions include Bull's review [1], Panda et al.'s investigation on hammerhead vehicles [2], and Lee & Sung's multi-array pressure measurements [3]. Further research explored separated flows, pressure-velocity correlations, and wall pressure propagation [4,5,6]. Experimental wind tunnel studies have also provided valuable insights into buffeting phenomena and aerodynamic forces on launchers [7,8,9,10]. Building on extensive research on the VEGA family launchers, this study aims to investigate and identify an appropriate machine learning-based predictive model for acoustic loads evaluation. ML techniques, particularly supervised learning, have demonstrated success in various engineering applications, including jet noise prediction [12], CFD-acoustics correlation [13], and far-field noise estimation [14]. ANNs and deep learning have been employed in fluid mechanics for turbulent flow predictions and wall-shear stress modeling [15,16,17]. Despite advancements, ML-based pressure load prediction for launchers remains challenging due to complex physical phenomena and limited experimental data. This study evaluates multiple ML algorithms—decision trees, regression models, SVMs, GPR, and ANNs—trained on an extensive experimental database from VEGA-E wind tunnel tests. Performance is assessed using RMSE, MSE, and R-squared, with validation via the 'Hold Out' method (20% validation set). Sensitivity analysis on dataset size further refines model robustness. Section II details the VEGA-E use case, Section III presents key results, and Section IV provides conclusions.

2. Test case description and procedures

The database originates from an aeroacoustic test campaign conducted on a scaled model of the VEGA-E launcher at the Trisonic Wind Tunnel of INCAS. This blow-down type tunnel provides airflow for a

limited duration, depending on the specified test conditions. It operates across three-speed regimes subsonic, transonic, and supersonic, covering a Mach range from 0.4 to 3.5.
Pressure fluctuations were recorded using 22 Miniature Pressure Transducer KULITE XCQ-062 sensors, which were flush-mounted at various spatial positions along the model's surface to capture a comprehensive description of the phenomenon. The locations of all pressure transducers are shown as red dots in Figure 1. The signals analyzed were acquired from a total of 22 sensors positioned along the x-direction. The transducers' output signals were recorded using an in-house conditioning system, which also provided signal amplification and anti-aliasing filtering. The pressure data were sampled at a rate of 10^5 samples per second, with a fixed acquisition time of 2.5 seconds for all cases.
The acquired data were processed in terms of Sound Pressure Level (SPL), computed for each sensor over a frequency range from 0 Hz to 50 kHz. When expressed in one-third octave bands, this results in 28 band center values. The operating conditions tested and used for ML training are:

- Mach Number (Ma): 0.8, 0.85, 0.9, 0.95, 0.98;
- Angle of Attack (α): 0°, 5° 10°;
- Roll Angle (φ): 0°, 45°, 90°;

Figure 1*:* Sketch of the pressure transducer position along the launcher model.

For prediction purposes, the SPL was not considered as a continuous function; instead, individual values for each frequency band were used. The parameters involved in the analysis are as follows:

- **Features:** Mach number (Ma), angle of incidence (α), sensor position along the launcher (X), and center band frequencies (f).
- **Predictor:** Sound Pressure Level (SPL).

Several supervised ML algorithms were trained and tested, categorized into the following five families:

- Regression algorithms [18].
- Decision trees [19,20,21].
- Artificial Neural Networks (ANNs) [22].
- Support Vector Machines (SVMs) [23].
- Gaussian Process Regression (GPR) [18].

For both the validation and testing phases, the statistical indices considered were Root Mean Square Error (RMSE) and R-squared. For all algorithms, a portion equivalent to 10% of the dataset was excluded from the training phase and used for a subsequent testing phase. This subset was randomly selected and kept consistent across all tested approaches. During the training phase, all algorithms were validated using the 'Hold Out' (HO) method, allocating 20% of the dataset for validation. This method is a widely used technique for evaluating the performance of ML models by randomly dividing the dataset into two separate subsets: the training set and the validation set. Each model is trained on the training set, where it learns the underlying patterns and relationships within the data. The trained model is then evaluated both on the validation set and on the test set, providing an estimate of its performance on unseen data and helping to assess its generalization ability.

3. Results

The analysis was conducted using MATLAB and its built-in regression algorithms. For all tested models, the default hyperparameters were initially used, as they provide a solid starting point for evaluation. For the sake of brevity, the specific hyperparameters are not detailed here; however, further information can be provided by the authors upon request. The results, in terms of statistical indices evaluated on both the validation and test sets, are summarized in Table 1. The models are ranked based on increasing RMSE values on the test set. Notably, the approaches with the lowest RMSE values are decision tree-based methods, followed by GPR and wide, three-layer neural networks, which exhibit significantly higher errors.

Additionally, tree-based algorithms also demonstrate excellent performance across other metrics, such as the R-squared indicator, which is close to the ideal value (R = 1). In particular, the best results were obtained using decision tree ensemble models, which combine multiple decision trees—each trained on different subsets of the training data—to enhance prediction accuracy. The ensemble method aggregates individual predictions, leading to improved overall performance.

Table 1: ML algorithms' performances evaluated on the validation and test set.

Model Type	RMSE (Validation)	RMSE (Test)	R-Squared (Validation)	R-Squared (Test)
Bagged Trees	1,122	0,967	0,937	0,956
Fine Tree	1,077	0,998	0,942	0,953
Medium Tree	1,374	1,270	0,906	0,924
Rational Quadratic GPR	1,397	1,274	0,902	0,924
Coarse Tree	1,970	1,939	0,806	0,823
Exponential GPR	2,156	2,028	0,768	0,806
Wide Neural Network	4,451	2,311	0,010	0,749
Matern 5/2 GPR	2,529	2,549	0,680	0,694
Squared Exponential GPR	2,588	2,612	0,665	0,679
Trilayered Neural Network	2,599	2,748	0,663	0,645
Fine Gaussian SVM	2,813	2,857	0,605	0,616
Medium Gaussian SVM	3,423	3,439	0,415	0,443
Least Squares Regression Kernel	3,237	3,543	0,476	0,409
Bilayered Neural Network	3,922	3,716	0,231	0,350
SVM Kernel	3,670	3,795	0,327	0,322
Cubic SVM	3,812	3,912	0,274	0,280
Quadratic SVM	4,164	4,258	0,134	0,147
Coarse Gaussian SVM	4,170	4,271	0,131	0,141
Stepwise Linear	4,470	4,573	0,002	0,015
Interactions Linear	4,470	4,578	0,002	0,014
Medium Neural Network	4,470	4,583	0,002	0,011
Linear	4,470	4,584	0,002	0,011
Narrow Neural Network	4,470	4,584	0,002	0,011
Efficient Linear SVM	4,493	4,600	-0,008	0,004
Efficient Linear Least Squares	4,469	4,602	0,002	0,003

Robust Linear	4,496	4,603	-0,010	0,002
Linear SVM	4,496	4,604	-0,010	0,002
Boosted Trees	6,787	6,840	-1,302	-1,203

An example of the predictive performance of selected trained algorithms is presented in Figure 2. Each plot compares the measured and predicted wall pressure SPL as a function of frequencies normalized by the first center-band frequency in the one-third octave band representation. The reported cases correspond to the signal recorded by the sensor located on the ogive at Ma=0.85, α=5°, φ=0°. The graphs support the findings based on statistical indices, confirming that the bagged tree algorithm provides the most accurate predictions. However, it is worth noting that the GPR with a rational quadratic kernel also achieves highly precise predictions at certain points, even surpassing the tree-based algorithm in some cases. Nevertheless, where the GPR model deviates from the measured signal, the errors tend to be larger, leading to an overall lower performance compared to the bagged tree approach.
The predictive capability of the selected algorithms is further assessed by analyzing the global quantity Cp_{rms}. This parameter is widely used as an indicator of the overall acoustic power of the pressure signals and it is closely related to the integral of the previously presented pressure spectra depending on both the sensor position along the launcher and the flow conditions.
Figure 3 compares measured and predicted values of Cp_{rms} at Ma=0.85, α=5°, φ=0° for the models illustrated in Figure 2. As expected, the bagged tree and the GPR with a rational quadratic kernel outperform the other models, providing highly accurate predictions.
Similar results are observed across the entire dataset, including all tested flow conditions and roll angles, though these additional cases are not reported here for brevity.

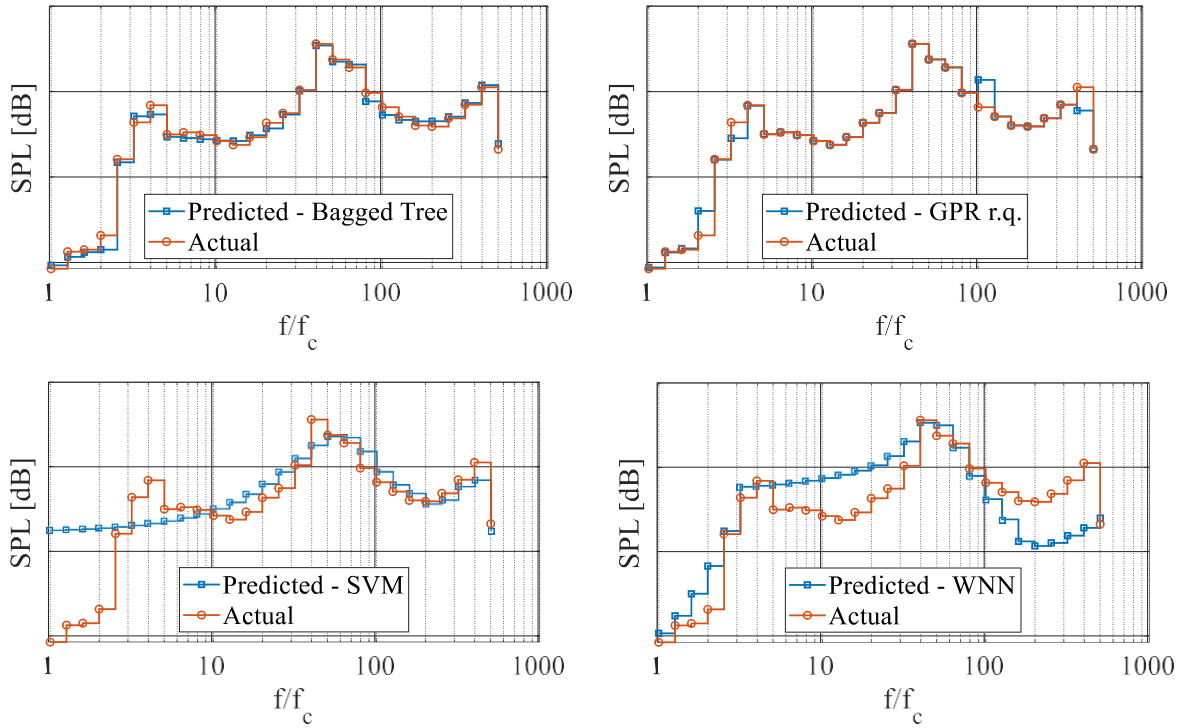

Figure 2: Comparison of models' prediction. Ma=0.85, α=5°, φ=0°, sensor on the ogive.

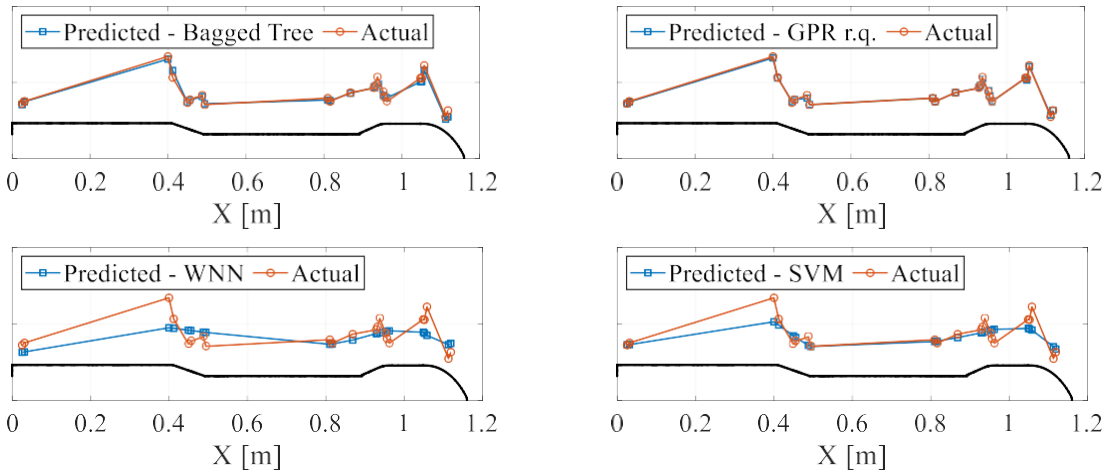

Figure 3: Measured Cp_{rms} compared with different model predictions. Ma=0.85, α=5°, φ=0°.

Finally, an initial attempt was made to predict the pressure signal at a sensor excluded from the training set using a deep learning approach. Specifically, an LSTM network was built to learn the evolution of pressure signals based on the test conditions (Mach and α). This model was trained to recognize patterns in the data and enabled the prediction of a new signal at a position not included in the original dataset. An example of the obtained result is shown in Figure 4. It is observed that the shape is accurately reconstructed, while the signal is underestimated, with a maximum error of 10 dB.

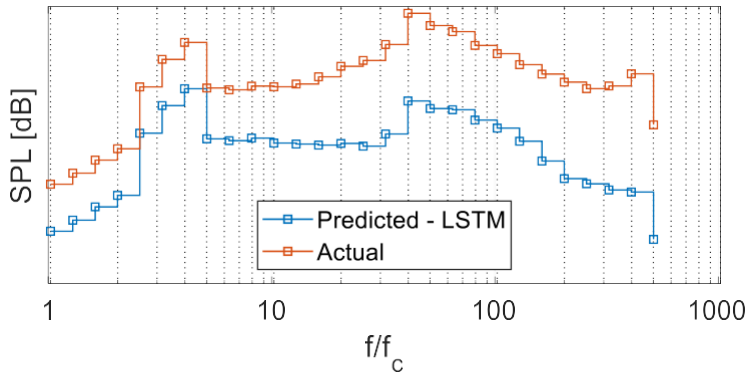

Figure 4: Comparison between the spectrum of the LSTM-predicted signal and the measured signal.

4. Conclusions

This study evaluates the capability of various ML algorithms to predict pressure fluctuations on the surface of a scaled VEGA-E launcher model. The training and testing data were obtained from pressure sensors mounted on the model during wind tunnel experiments conducted under different flow conditions. The dataset allowed for the assessment of multiple ML techniques, including regression models, decision tree ensembles, and artificial neural networks.

Among the tested approaches, ensemble tree-based methods, particularly the bagged tree algorithm, demonstrated the best predictive performance for key statistical quantities such as the sound pressure level (SPL) and the root-mean-square pressure coefficient. The bagged tree model consistently achieved the lowest root-mean-square error (RMSE), outperforming Gaussian Process Regression (GPR), Support Vector Machines (SVM), and ANNs. Its superior performance is attributed to its ability to mitigate variance through bootstrap aggregation, making it robust against noise and capable of generalizing well to untested flow conditions. It is worth noting that the GPR with a rational quadratic kernel also achieves precise predictions at certain points, even surpassing the tree-based algorithm in some cases. Nevertheless, where the GPR model deviates from the measured signal, the errors tend to be larger, leading to an overall lower performance compared to the bagged tree approach.

In contrast, artificial neural networks, while known for their ability to model complex relationships, exhibited higher prediction errors in this application. This may be due to their sensitivity to hyperparameter tuning and the relatively limited size of the experimental dataset, which could have constrained their generalization capability. Similarly, SVM struggled with the complexity and variability of wall pressure fluctuations, leading to reduced predictive accuracy.
Overall, the results confirm the effectiveness of ML techniques in predicting wall pressure fluctuations, with the bagged tree algorithm emerging as the most promising approach due to its accuracy, robustness, and generalization capability. Despite these promising findings, further validation is necessary across different geometries and flow conditions. The inherent challenges of turbulent flows and varying launcher configurations must be addressed to develop a more versatile predictive framework. Future research will extend the methodology to additional datasets, including transonic and supersonic regimes, and explore hybrid approaches that combine ML with traditional modeling techniques to further improve accuracy and applicability.

5. References

[1] Bull, M.K. Wall-pressure fluctuations beneath turbulent boundary layers: Some reflections on forty years of research. Journal of Sound and Vibration 1996, 190, 299–315, https://doi.org/10.1006/jsvi.1996.0066.
[2] Panda, J.; Garbeff, T.J.; Burnside, N.J.; Ross, J.C. Unsteady pressure fluctuations measured on a hammerhead space vehicle and comparison with Coe and Nute's 1962 data. International Journal of Aeroacoustics 2018, 17, 70–87. https://doi.org/10.1177/1475472X17743626.
[3] Lee, I.; Sung, H.J. Multiple-arrayed pressure measurement for investigation of the unsteady flow structure of a reattaching shear layer. Journal of Fluid Mechanics 2002, 463, 377–402. https://doi.org/10.1017/S002211200200890X.
[4] Goody, M.C.; Simpson, R.L.; Chesnakas, C.J. Separated flow surface pressure fluctuations and pressure-velocity correlations on prolate spheroid. AIAA journal 2000, 38, 266–274. https://doi.org/10.2514/2.953.
[5] Camussi, R.; Guj, G.; Marco, A.D.; Ragni, A. Propagation of wall pressure perturbations in a large aspect ratio shallow cavity. Experiments in Fluids 2006, 40, 612–620. https://doi.org/10.1007/S00348-005-0101-X.
[6] Leclercq, D.J.; Bohineust, X. Investigation and modelling of the wall pressure field beneath a turbulent boundary layer at low and medium frequencies. Journal of Sound and Vibration 2002, 257, 477–501. https://doi.org/10.1006/jsvi.2002.5049.
[7] Herron, A.J.; Crosby, W.A.; Reed, D.K. Overview of the space launch system ascent aeroacoustic environment test program. 54th AIAA Aerospace Sciences Meeting 2016. https://doi.org/10.2514/6.2016-0543.
[8] Rainey, A.G. Progress on the launch-vehicle buffeting problem. Journal of Spacecraft and Rockets 1965, 2, 289–299. https://doi.org/10.2514/3.28174.
[9] Camussi, R.; Guj, G.; Imperatore, B.; Pizzicaroli, A.; Perigo, D. Wall pressure fluctuations induced by transonic boundary layers on a launcher model. Aerospace Science and Technology 2007, 11, 349–359. https://doi.org/https://doi.org/10.1016/j.ast.2007.01.004.
[10] Caruana, D.; Mignosi, A.; Corrège, M.; Pourhiet, A.L.; Rodde, A.M. Buffet and buffeting control in transonic flow. Aerospace Science and Technology 2005, 9, 605–616. https://doi.org/10.1016/j.ast.2004.12.005.
[11] Graham, W.R. A comparison of models for the wavenumber-frequency spectrum of turbulent boundary layer pressures. Journal of Sound and Vibration 1997, 206, 541–565. https://doi.org/10.1006/jsvi.1997.1114.
[12] Dowdall, J. Applying Machine Learning to Jet Noise Prediction. In Proceedings of the Acoustics and Urban Air Mobility Technical Working Group Meetings, 2018. https://doi.org/GRC-E-DAA-TN62141.
[13] Shah, M. Machine Learning Models for Jet Noise Analysis. Master's thesis, The Ohio State University, 2019. Advisor: Gaitonde, Datta.
[14] Brown, C.A.; Dowdall, J.; Whiteaker, B.; McIntyre, L., A Machine Learning Approach to Jet-Surface Interaction Noise Modeling. In AIAA Scitech 2020 Forum; 2020. https://doi.org/10.2514/6.2020-1728.
[15] Kim, J.; Lee, C. Prediction of turbulent heat transfer using convolutional neural networks. Journal of Fluid Mechanics 2020, 882, A18. https://doi.org/10.1017/jfm.2019.814.
[16] Lee, S.; You, D. Data-driven prediction of unsteady flow over a circular cylinder using deep learning. Journal of Fluid Mechanics 2019, 879, 217–254. https://doi.org/10.1017/jfm.2019.700.
[17] Le Clainche, S.; Rosti, M.; Brandt, L. A data-driven model based on modal decomposition: application to the turbulent channel flow over an anisotropic porous wall. Journal of Fluid Mechanics 2022, 939, A5. https://doi.org/10.1017/jfm.2022.159.
[18] Mitchell, T.M.; Mitchell, T.M. Machine learning; Vol. 1, McGraw-hill New York, 1997.
[19] Breiman, L. Classification and regression trees; Routledge, 2017.
[20] James, G.; Witten, D.; Hastie, T.; Tibshirani, R. An Introduction to Statistical Learning: with Applications in R; Springer Texts in Statistics, Springer New York, 2013.
[21] Dietterich, T.G. Ensemble Methods in Machine Learning. In Proceedings of the Multiple Classifier Systems, Berlin, Heidelberg, 2000; pp. 1–15.
[22] Bishop, C. Neural Networks For Pattern Recognition; Vol. 227, Oxford University Press, 2005.
[23] Cristianini, N.; Shawe-Taylor, J. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods; Cambridge University Press, 2000.