Filling the Gaps: ML Tabular Regression for Missing Material Properties in Data-Scarce Engineering Contexts

Davide Di Stefano[1], Pascal Salzbrenner[1], Junyan He[2], Adarsh Chaurasia[2]
[1]Ansys UK Ltd, Cambridge, CB1 7EG, UK Country
[2]Ansys Inc, Canonsburg, PA 15317, USA

Davide.distefano@synopsys.com

Abstract (Arial 10)

AI and machine learning (ML) are playing an increasingly important role in materials science. However, much of the attention has focused on idealized applications such as crystal structure prediction or ML interatomic potentials. Although often very advanced, these techniques often fail to address the day-to-day challenges faced by engineers working with real materials and real constraints.
In industrial settings, engineers frequently need material properties to perform thermal stress or performance critical simulations, only to find that the required value is missing, inconsistent, or measured under incompatible conditions. Measuring the appropriate data experimentally is costly and slow. Consequently, available datasets are small, heterogeneous, and tied to processing routes or microstructural states that differ from the engineer's application. As a result, engineers often rely on crude estimates or ad hoc substitutions with “similar” materials, introducing uncertainty and potential design risk.
We explore how modern ML techniques, combined with domain aware data augmentation and careful uncertainty quantification, can make far better use of the data we already have. By leveraging ensemble-based regression, tabular foundation models, microstructure aware featurization, and curve-based transformations, we show that it is possible to estimate missing properties with quantified confidence and broaden the usable evidence base for legacy or incomplete materials systems.
Using a thermal stress analysis of an exhaust manifold as a representative case, Fig 1, we show how the common practice of compensating for missing material properties by selecting a “similar” material can lead to significant inaccuracies. By contrast, ML models trained on available data can provide more reliable, uncertainty aware estimates of the required properties. While not flawless, we demonstrate that ML based estimates surpass ad hoc substitutions, giving engineers a stronger basis for simulation and design choices. Crucially, this not only reduces unnecessary over engineering and costly redesign cycles but also enables engineers to make confident use of a wider range of existing materials.

Figures

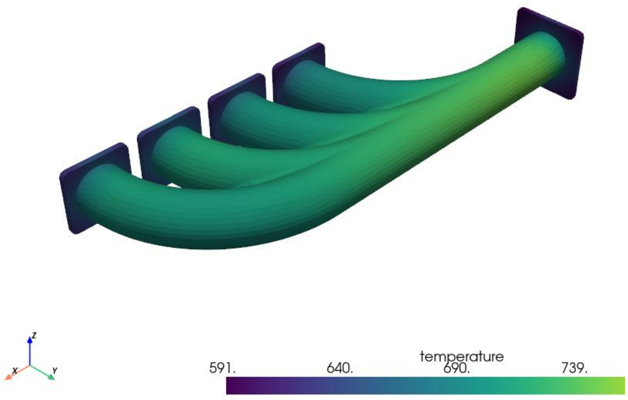

Figure 1. Steel exhaust manifold, temperature distribution obtained from a CFD simulation