Limitations of XGBoost in Predicting Material Parameters for Complex Constitutive Models

Pedro Prates[1,2*], *Dário* Mitreiro[1], and *António* Andrade-Campos[1,2]

[1]Centre for Mechanical Technology and Automation (TEMA), Department of Mechanical Engineering, University of Aveiro, 3810-193 Aveiro, Portugal
[2]Intelligent Systems Associate Laboratory (LASI), Portugal

Abstract. Machine learning models, particularly Extreme Gradient Boosting, have been explored for predicting material parameters in constitutive models that describe the plastic behaviour of metal sheets. While effective for simple constitutive models like Hill'48, their performance declines with more complex models such as the Cazacu-Plunckett-Barlat yield criterion. This study examines the influence of training dataset size and dimensionality reduction via principal component analysis on predictive accuracy. Results show that increasing the training dataset size leads to only marginal improvements, with testing coefficient of determination value plateauing at about 0.50, despite a consistently high training value of about 0.99999, indicating overfitting. Similarly, applying principal component analysis to the baseline model provided no significant enhancement. These findings suggest that simply expanding the dataset or reducing dimensionality is insufficient to address the complexities of CPB'06. Instead, alternative approaches such as advanced feature selection, hybrid ML-physics-based models, or regularization techniques may be required to improve generalization. Future work should explore methods integrating domain knowledge and physics-based modelling to enhance predictive accuracy for complex constitutive models.

Keywords: Parameter identification; Machine learning; Metal sheets; Numerical simulation.

1 Introduction

Accurate prediction of material behaviour under complex loading conditions is crucial for optimizing sheet metal forming processes. Numerical simulations, combined with advanced constitutive models, play a crucial role in capturing the anisotropic and hardening behaviour of metallic materials. Each model offers distinct advantages and challenges, shaped by its mathematical formulation and parameter dependencies. In recent years, data-driven approaches like machine learning (ML) have become popular alternatives for predicting constitutive parameters, driven by increased computational power and data availability [1].

ML-based regression techniques, particularly those utilizing Extreme Gradient Boosting (XGBoost) [2], have gained prominence as effective tools for calibrating material parameters from experimental or simulated data. Unlike ANN-based approaches commonly employed in previous studies [1], XGBoost offers enhanced interpretability, computational efficiency, and robustness to overfitting, especially beneficial when working with limited datasets. These methods facilitate faster and more efficient parameter identification, but their predictive accuracy is influenced by the complexity of the yield criterion and the availability of high-quality training data. Previous studies using XGBoost with the Hill'48 yield criterion demonstrated strong predictive capabilities, highlighting the potential of ML for material modelling [3, 4].

This work examines the performance of XGBoost-based ML models in predicting material parameters of Cazacu-Plunckett-Barlat (CPB'06) yield criterion [5]. A biaxial tensile test on a cruciform specimen is employed to generate heterogeneous stress and strain fields, representative of real-world forming conditions. Additionally, the study investigates how the size of the training dataset impacts prediction accuracy, revealing variations in model sensitivity to data availability and the complexity of the yield criterion.

The findings provide valuable insights into the limitations of ML-based parameter identification for CPB'06 yield criterion, contributing to the development of more efficient and reliable material behaviour predictions for industrial applications.

2 Methods and procedures

2.1 Numerical model

This study utilizes a biaxial tensile test on a cruciform sample as the selected mechanical testing approach. The sample's geometry, previously designed in earlier work [6], promote heterogeneous stress and strain fields, capturing a wide range of stress and strain paths commonly observed in sheet metal forming processes.

* Corresponding author: prates@ua.pt

The sample's geometry and in-plane dimensions are illustrated in Fig. 1. To optimize computational efficiency, and due to symmetries in the boundary conditions, sample geometry, and material behaviour, the numerical simulation considers only one-eighth of the sample. In fact, the highlighted grey region in Fig. 1 presents symmetry across the three orthogonal planes: $x = 0$ mm, $y = 0$ mm, and $z = 0$ mm. Displacement boundary conditions ensure equal displacements ($u_x = u_y = 2$ mm) at the ends of both sample arms. The numerical model employs a structured mesh comprising 564 C3D8R hexahedral elements, characterized by trilinear shape functions and reduced integration. Finite Element Analysis (FEA) simulations are conducted using ABAQUS CAE software [7], with each simulation divided into twenty equally spaced time-steps. Throughout these steps, the forces along the *0x* and *0y* directions, as well as the strain field components (ε_{xx}, ε_{yy}, and ε_{xy}), are recorded at every interval.

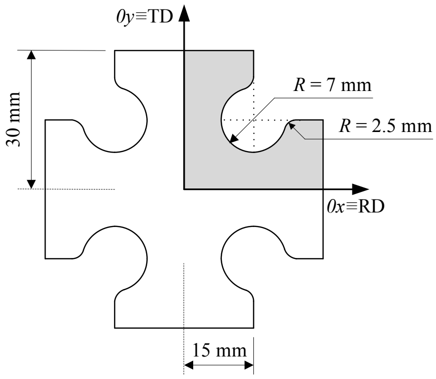

Fig. 1. Geometry and in-plane dimensions of the biaxial tensile test on a cruciform sample [6]. The specimen thickness is 1 mm.

The material model assumes isotropic elastic behaviour, governed by Hooke's law (with Young's modulus E = 210 GPa and Poisson's ratio v = 0.3), and isotropic hardening described by Swift law [8] under an associated flow rule. The CPB'06 yield criterion can be written as follows:

$$\left(|\Sigma_1| - k\Sigma_1\right)^a + \left(|\Sigma_2| - k\Sigma_2\right)^a + \left(|\Sigma_3| - k\Sigma_3\right)^a = Y^a , \quad (1)$$

where Σ_1, Σ_2, and Σ_3 are the principal values of the transformed deviatoric stress tensor $\mathbf{\Sigma = C\sigma'}$ where **C** is a constant fourth-order tensor containing 9 independent anisotropy parameters, given by:

$$\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & & & \\ C_{12} & C_{22} & C_{23} & & & \\ C_{13} & C_{23} & C_{33} & & & \\ & & & C_{44} & & \\ & & & & C_{55} & \\ & & & & & C_{66} \end{bmatrix} . \quad (2)$$

The Swift law describes the yield stress evolution during plastic deformation as follows:

$$Y = K\left[\left(\frac{\sigma_0}{K}\right)^{1/n} + \bar{\varepsilon}^p\right]^n , \quad (3)$$

where $\bar{\varepsilon}^p$ is the equivalent plastic strain and σ_0, K, and n are material parameters.

2.2 Data synthesis and model training

Using the Latin Hypercube Sampling (LHS) method, a total of 10550 simulations were generated. Each simulation had an average computation time of 10 seconds, utilizing a 13th Generation Intel® Core™ i9-13900 24-core processor (2.00–5.60 GHz) with 32 GB of RAM. The input space and step sizes for the material parameters are detailed in Table 1. Numerical simulations of the cruciform tensile test were performed while maintaining the same geometry, boundary conditions and elastic properties. 4750 out of 10550 simulations were selected as they exhibited no decrease in load during the simulation, meaning they did not surpass the necking point before the 20th time step. Then, the dataset was randomly split into two sets: 4000 samples were used for training and 750 for testing. Each of the training and test sets consist of two matrices: a feature matrix and a target matrix. The feature matrix has shape $n_{\text{samples}} \times n_{\text{features}}$, where n_{samples} is the total number of samples (i.e. numerical simulations) of each set (n_{samples}=4750) and n_{features} is the total number of features (20 timesteps×(2 forces + 3 strain components × 564 elements) = 33880 features). The target matrix has shape $n_{\text{samples}} \times n_{\text{outputs}}$, where n_{outputs} is the total number of model outputs (i.e. material parameters – n_{outputs}=9). The training set was used to train the Extreme Gradient Boosting (XGBoost) regression algorithm, with additional details available in [1]. Most hyperparameters remained at default settings, except for the *learning rate* (0.02), *max_depth* (15), and *n_estimators* (1000) [2]. The test set was used to evaluate the model's performance in predicting material parameters.

Table 1. Material parameters, ranges, and step sizes for dataset generation.

Material Parameters	Range	Step
C_{22}	-0.5 – 2.5	0.0001
C_{33}	-0.5 – 2.5	0.0001
C_{12}	-1.5 – 1.5	0.0001
C_{13}	-1.5 – 1.5	0.0001
C_{23}	-1.5 – 1.5	0.0001
C_{66}	-1.5 – 1.5	0.0001
K [MPa]	280 – 700	0.01
σ_0 [MPa]	120 – 300	0.01
n	0.100 – 0.300	0.001

2.3 Sensitivity analysis

The predictive performance of trained ML models is influenced by multiple factors, including training dataset size, dimensionality reduction, feature selection, and hyperparameter tuning. This work focuses on the influence of training dataset size and dimensionality reduction.

2.3.1 Training set size

The training set was divided into multiple subsets to assess the performance of the machine learning models with different amounts of training data. Subsets of 500, 1000, 1500, 2000, 2500, 3000, and 3500 samples were used, along with the full training set of 4000 samples. Model training was performed for each subset using the same hyperparameters, and each trained model was evaluated on the 750 samples of the testing set.

2.3.2 Dimensionality reduction

In this study, Principal Component Analysis (PCA) was applied to reduce the dimensionality of the dataset, aiming to enhance the predictive performance of the ML models. PCA is a powerful technique that transforms the original dataset into a new coordinate system, where each axis represents a direction of maximum variance in the data. The objective of this analysis was to determine how many principal components would yield the best model performance and why performance improves or deteriorates as more components are added. To achieve this, various PCA component counts were explored, ranging from a single component to 4000, matching the total number of samples in the full training dataset. Model training was performed for each PCA-transformed dataset using the same hyperparameters, and each trained model was evaluated on the 750 samples of the testing set.

2.4 Performance evaluation

In this work, prediction performance of the model was evaluated using the testing set by comparing the estimated and true values of the material parameters using the coefficient of determination (R^2), given by:

$$R^2 = \frac{1}{9}\sum_{j=1}^{9}\left(1-\frac{\sum_{i=1}^{750}\left(y_{\text{test},\,i}^{(j)}-\hat{y}_{\text{test},\,i}^{(j)}\right)^2}{\sum_{i=1}^{750}\left(y_{\text{test},\,i}^{(j)}-\overline{y}_{\text{test}}^{(j)}\right)^2}\right), \tag{4}$$

where $y_{\text{test},\,i}^{(j)}$ is the true value of parameter j for sample i of the test set, $\hat{y}_{\text{test},\,i}^{(j)}$ is the predicted value of parameter j for sample i of the test set, and $\overline{y}_{\text{test}}^{(j)}$ is the mean of the actual values for parameter j.

3 Results and discussion

Fig. 2 presents the sensitivity analysis of training dataset size on the testing R^2 metric. The analysis indicates that testing R^2 improvements were gradual and eventually plateaued at approximately 0.502 for the model trained with the full 4000-sample training set, referred to as the baseline model. Regardless of the training set size, the training R^2 remains consistently high (approximately 0.99999, not shown in Fig. 2), while the testing R^2 drops significantly to 0.502 for the baseline model, indicating overfitting. This performance gap underscores the complexity of the CPB'06 yield criterion, highlighting the limited but gradual improvement in predictive accuracy with increasing training set size. This suggests that while larger datasets contribute to better generalization, their impact is modest, indicating that alternative strategies, such as dimensionality reduction, improved feature selection or hybrid ML-physics-based models, may be required to achieve significant performance gains. In particular, the higher parameter count and the nonlinear dependencies between material parameters introduce significant challenges for model training. In contrast, previous studies using the Hill'48 yield criterion [3-4] reported consistently high R^2 values, reflecting the ML model's ability to generalize more effectively due to Hill'48's simpler structure and lower parameter count. The relatively low R^2 values obtained for CPB'06 may also result from parameter insensitivity, indicating that identifying and excluding weakly influential parameters could enhance generalization.

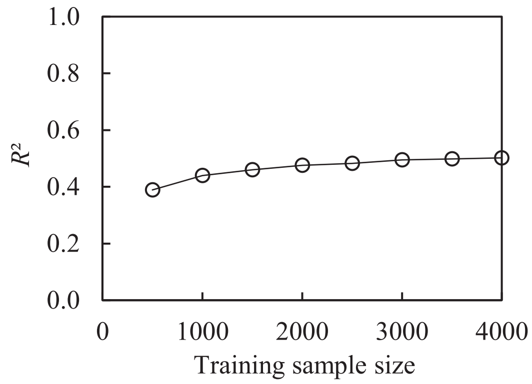

Fig. 2. Sensitivity analysis of training sample size on the R^2 metric for the test set performance.

Fig. 3 illustrates the impact of applying Principal Component Analysis (PCA) to the baseline model on the R^2 metric. The analysis shows an initial improvement in predictive accuracy as the number of PCA components increased, reaching a peak of 0.501 at 100 components. Beyond this point, performance gradually declined, with R^2 decreasing to 0.337 when all 4000 components were used. These results indicate that PCA effectively captures the most significant variance in the dataset, improving generalization when a moderate number of components are retained. However, applying PCA to the baseline model did not yield a substantial improvement, suggesting that the original feature space already contained the most relevant information for prediction. The decline in performance at higher component counts suggests that excessive dimensionality reintroduces noise and redundant features, potentially exacerbating parameter insensitivity and collinearity effects. These findings highlight the limited impact of PCA in overcoming the apparent predictive performance threshold of the XGBoost model.

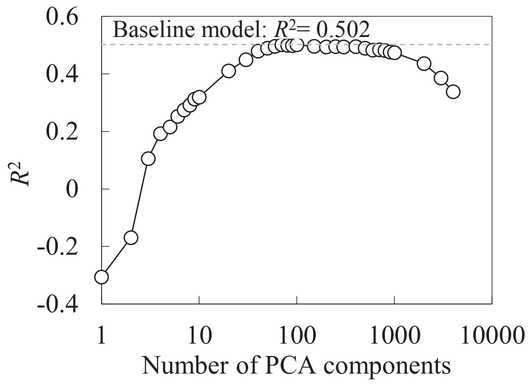

Fig. 3. Effect of the number of PCA components on the R^2 metric after applying PCA to the baseline model.

4 Conclusions

The findings of this study highlight the limitations of standard machine learning approaches in predicting material parameters for complex constitutive models such as CPB'06. While increasing the training dataset size led to marginal improvements of the predictive performance, the performance plateaued at $R^2 \approx 0.50$ suggests that simply expanding the dataset may not be sufficient to overcome the challenges posed by the high parameter count and nonlinear dependencies of CPB'06. Similarly, dimensionality reduction through PCA provided no significant enhancement, reinforcing the idea that feature redundancy and collinearity are not the primary bottlenecks in model performance.

These results indicate that alternative strategies – such as advanced feature engineering, tailored regularization techniques, or hybrid ML-physics-based models – may be necessary to achieve meaningful improvements in predictive accuracy. Future work should explore methods that better incorporate the underlying physics of material behaviour, potentially bridging the gap between data-driven and physics-based approaches to enhance the generalization capability of machine learning models for complex constitutive modelling.

Acknowledgements

The authors gratefully acknowledge the financial support of the Portuguese Foundation for Science and Technology (FCT) and by UE/FEDER through the programs CENTRO 2020 and COMPETE 2020, UIDB/00481/2020 and UIDP/00481/2020-FCT, DOI 10.54499/UIDB/00481/2020 (https://doi.org/10.54499/UIDB/00481/2020) and DOI 10.54499/UIDP/00481/2020 (https://doi.org/10.54499/UIDP/00481/2020), CENTRO-01-0145-FEDER-022083, and LA/P/0104/2020.
It was also supported by projects 2022.05783.PTDC–FCT (https://doi.org/10.54499/2022.05783.PTDC), 2022.02370.PTDC (https://doi.org/10.54499/2022.02370.PTDC) and 2023.14606.PEX (https://doi.org/10.54499/2023.14606.PEX), funded by Portuguese Foundation for Science and Technology.

References

1. A. Marques, T. Parreira, A. Pereira, B. Ribeiro, P. Prates, Int. J. Solid. Struct. 303, 113024 (2024)
2. T. Chen, C. Guestrin, *XGBoost: A scalable tree boosting system*, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17 August 2016, San Francisco, USA (2016)
3. P. Prates, J. Pinto, J. Marques, J. Henriques, A. Pereira, A. Andrade-Campos, Mater. Res. Proc. **41**, 1807 (2024)
4. P. Prates, D. Mitreiro, A. Marques, A. Pereira, A. Andrade-Campos, Mater. Res. Proc. (to be published)
5. O. Cazacu, B. Plunkett, F. Barlat, Int. J. Plast. **22**, 1171 (2006)
6. J. Martins, A. Andrade-Campos, and S. Thuillier, Int. J. Solids Struct. **172**, 21 (2019)
7. Dassault Systèmes. Abaqus 2017 documentation (2017)
8. H. Swift, J. Mech. Phys. Solids, **1**, 1 (1952)