Title:
**Beyond Traditional Feature Selection in Computer-Aided Design: Feature-Mining in Engineering Design**

Authors:
Mohammad Arjomandi Rad, radmo@chalmers.se, Chalmers University of Technology
Massimo Panarotto, massimo.panarotto@chalmers.se, Chalmers University of Technology
Ola Isaksson, ola.isaksson@chalmers.se, Chalmers University of Technology

Keywords:
Thin-walled tubes, Crashworthiness, Feature extraction, Data-driven design, Simulation-driven design, CAD/CAE, Engineering design

DOI: 10.14733/cadconfP.2024.345-350

Introduction:
Building predictive models for assessing a design's performance is a standard practice in engineering design. Traditional Computer-Aided Design (CAD) parameterization, while instrumental in many engineering applications, presents inherent challenges when tailored for predictive modeling, particularly in the realm of surrogate or metamodels. One of the primary issues is the high-dimensional nature of CAD-defined parameters. This issue introduces unnecessary complexity to prediction models, especially with the increase in the problem size, making predictive models cumbersome and potentially reducing their accuracy. As a result, it will be incredibly expensive to explore design space with more complex systems [1]. With more data availability and increased complexity of modern products, the information overload can overwhelm any design process [2], [3].

Additionally, conventional CAD parameterization tends to be deterministic and might not easily accommodate the variability and uncertainty inherent in real-world applications, which is crucial for robust prediction models [4]. For example, in case of design change (such as an assembly relationship or a sketch dimension removal or addition) could make a surrogate model useless because the dimension that was being utilized as a feature in the machine learning might not be used in the geometry as a result of the drastic change in the shape.

Furthermore, CAD parameters are frequently derived based on design intent [5] meaning designers just want to ensure that their idea translates accurately into a digital (and ultimately physical) object. Yet model predictive power is often brought in late during design iterations. So most likely, designers unintentionally make decisions that lead to the inclusion of irrelevant or redundant features that do not reveal as problems until the later phases of product development [6]. In design space exploration, not only does this increase the computational demands but can also obscure meaningful relationships (between input and output), hindering the model's generalizability and performance. Thus, there's a pressing need for refined and purpose-driven parameterization techniques that align more closely with the requirements of predictive models.

In this paper, by feature, we mean variables in the predictive model dataset that are connected to the geometry and CAD work. To avoid mixing, we use the sleeping parameters convention that was introduced recently by the authors [7]. CAD features are historically used for building meta/surrogate models [8], [9]. These features are dependent on the geometry and how the designer shapes it, which brings in all the mentioned problems, such as rigid parameterized models, high dimensionality, unintentional complexity by following design intent, and so on. Sleeping parameters are defined in contrast to conventional CAD parameterization. Sleeping parameters are defined as mined features that are extracted from the geometry of the design but are not coupled with them. They can be constructed, extracted, selected, and then processed even if the geometry loses or gains drastic changes [7]. The aim is to make the end predictive model more flexible by engineering high-quality features.

Studied Case:

Within the automotive industry, the analysis of structural components is crucial, given their role in vehicle safety and performance. Many automakers use a repetitive design process to evaluate the performance of these designed beams. As the final design needs to be integrated with other systems, these design iterations can take up to many years [10]. Therefore, being able to predict the performance of these beams is of utmost importance. To show such ability in this paper, we use cross-section geometries for the beams that already exist in the literature. [11]. Figure 1 shows 46 geometries that are a simplified representation of a Toyota RAV4's frame[10]. Using these images of the geometries, a similar scaled curve is extracted for each shown cross-section.

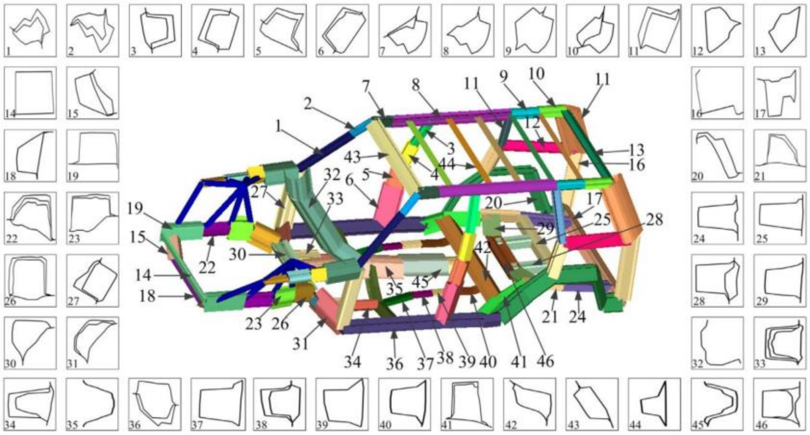

Fig. 1: An example of frames in BIW with many different beam geometries

To be able to show the predictability, SEA and PCF of each one of these tubes are simulated under lateral load as shown in Figure 2. To this end, a dynamic explicit simulation is used with a semi-automatic mass scaling in a process. Several output variables are requested from the simulation, all these variables are requested every 5e-5 second of the simulation. The reaction force is the first variable requested; this is measured at every node in the tube's boundary condition. The other two variables are the velocity of the wall displacement, since the wall touches the tube at the beginning of the simulation the measured amount of wall displacement is the same amount the tube displaces [12].

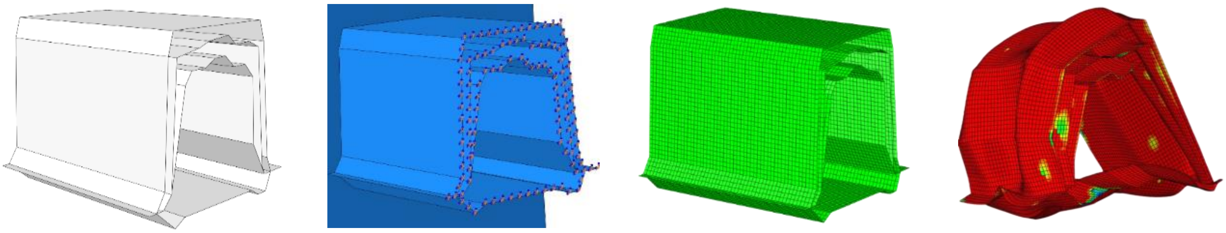

Fig. 2: The finite element process of one example cross-section.

All cross-sections are simulated to read out their crashworthiness characteristics, such as PCF and SEA. The verification of the FEM model was done with the results of published literature [13]. Many of the geometries that are simulated have spot welds between the different plates. To simplify the FEM process, they are considered as rigid nodes as suggested in the literature [14]. Since the performance of the welds is not of interest in this study, this choice does not affect the final conclusions.

Extracting sleeping Parameters:

This paper suggests using an alternative geometric representation of a shape as a means for data mining before building a predictive machine learning model. The medial axis or the "skeleton" of a shape is the set of all points inside the shape that have equal distance to two or more points on the shape's boundary. The medial axis of a simple polygon is closely related to the Voronoi diagram constructed from its edges. [7]. The shape of the medial axis is unique for each polygon and it can be used to restore a shape based on its medial axis as well. In general, one can consider the medial axis of a shape as an alternative geometric representation of the shape but in a lower dimension. This is because the medial axis of a 3d shape is a 2d surface, and the medial axis of a 2d shape reduces down to a curve. The convex or concaveness of the shape is not of importance for our use case aim because it just determines the direction and position of the medial axis. However, the shaper must be closed, and indeed, three of the cross-sections that are not closed polygons (numbers 16, 32, and 35) are removed from our analyses.

To construct a medial axis of shape Rhino Grasshopper is used. The saved STEP files (from the previous section) were imported into Rhino Grasshopper using the yellow part of the script shown in Figure 3. The figure shows after importing the geometry the data goes through components for creating a boundary surface and then the surface splits it into several segments in another component depending on how many sections exist in the geometry. The blue section in the figure reconstructs the medial axis of each section and then combines them with a series of components for later analysis. This is shown with two output components in the figure namely “Radius of circles” and “Medial axis segments

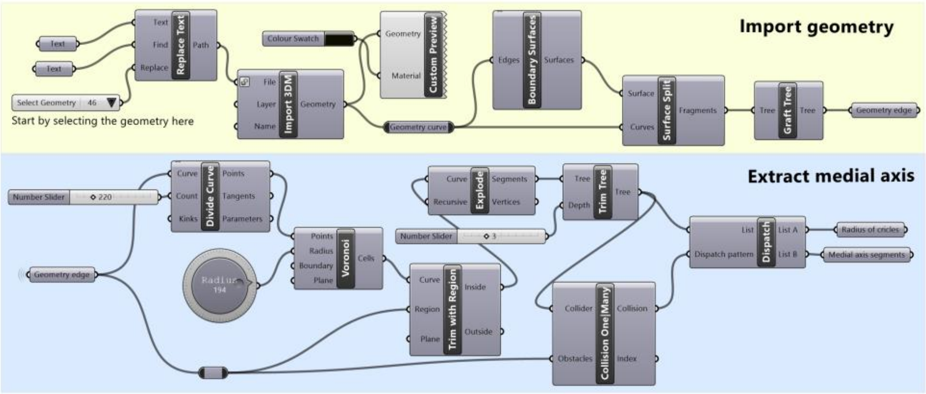


Fig. 3: The visual representation of the Rhino grasshopper code.

The steps for constructing the medial axis of a shape are as follows. First, the boundary is divided into equally distanced points (the number can be adjusted with a slider component), and then a circle is grown on each cell to create Voronoi cells. Here, every point in the circumference acts as a seed for these Voronoi cells. When these cells reach each other (as a result of increasing the size of cells), they create the desired medial axis.

Figure 4 illustrates the results of the applied process and the acquired medial (in blue line) axis and radius of the circles (in gray lines) for one, two, and three segmented cross sections from left to right. The black curves correspond to the geometry’s outer circumference, and as is shown, cross-sections have different numbers of segments. It can be argued that while the medial axis is coupled to the shapes' intricate geometry, the constructing radiuses are related to the regions of the shape and thus can hold regional-based information for end predictive models.

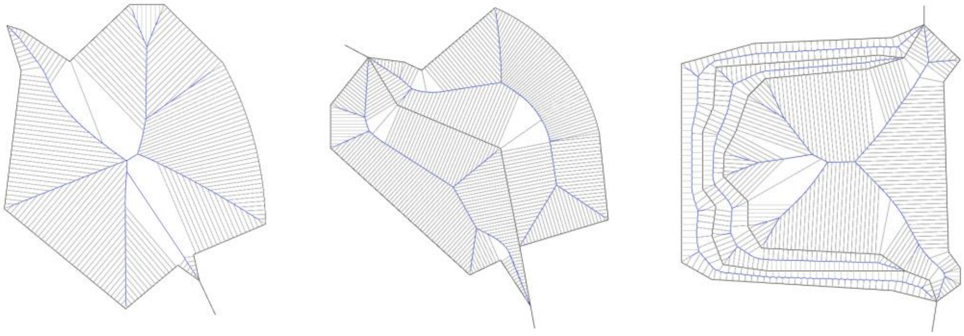

Fig. 4: Several geometries (3 out of 46) after the Voronoi operation.

Different types of features

The region-based information can be extracted by averaging or summing the length of all gray lines in the shape. Thus, averaging and summing gives us two features that are called “Average. circle radius” and “Width information,” respectively. The average circle radius is indirectly related to the shape's overall size and scale. While it doesn't give the exact area, it provides a sense of whether shapes are predominantly large or small within the dataset. If size matters for the predictive model, this feature could be a useful input.

The medial axis (blue curve) shown in Figure 4 can be used to extract several features, such as the “length of the medial axis” and “number of branch points.” The length of the medial axis can offer insight into the overall size and extent of the shape. The number of branch points, on the other hand, reflects the complexity of the shape's internal structure.

Perimeter offers a basic size estimator. While it doesn't directly capture intricate details, it provides a general size reference for comparison between shapes. This can be valuable for a predictive model that involves any kind of normalization or adjustment based on size. In topology, "handles" refer to the number of holes in the shape. A higher number of handles correlates with a more complex boundary, potentially indicating cavities or indentations.

Another feature, the *compactness ratio* of a polygon shape is a well-known metric for 2d and 3d shapes and is an intrinsic property of objects [15]. This is calculated as the ratio of the area of a shape to the area of its bounding circle. For the same area, shapes that deviate significantly from being circular will have lower compactness ratios. This feature can offer insights into how well a shape fills its space. Hybrid features can help differentiate shapes with similar areas but drastically different contours.

Validating the extracted features:

After any feature mining process, it is crucial to employ feature selection techniques to determine which one of the extracted features has the strongest predictive power. Different correlation techniques can be used to study the quality of the extracted features. We aim to use different types of techniques to account for both linear relations in the data as well as nonlinear relations in this paper.

To be able to describe the complexity of the relation, a simple linear relationship that is derived from fitting a linear line to the data, is used. This parameter can be extracted from the score of the linear regression model in Keras. Table 1 shows the linear regression score as the first quality check metric for all extracted features. This score mainly tells us how well a simple linear model explains the variability in one variable based on changes in another variable.

| | Linear regression score | | Pearson correlation | | Spearman correlation | |
|---|---|---|---|---|---|---|
| | *SEA* | *PCF* | *SEA* | *PCF* | *SEA* | *PCF* |
| *Length of Medial Axis* | 0.83 | 0.89 | -0.91 | 0.94 | -0.93 | 0.92 |
| *Width Information* | 0.26 | 0.23 | -0.50 | 0.48 | -0.57 | 0.58 |
| *Num. of Handels* | 0.68 | 0.77 | -0.82 | 0.88 | -0.81 | 0.81 |
| *Branching Points* | 0.70 | 0.79 | -0.83 | 0.89 | -0.86 | 0.85 |
| *Shape Perimeter* | 0.87 | 0.98 | -0.93 | 0.99 | -0.99 | 0.98 |
| *Avg. Circle Radius* | 0.34 | 0.30 | 0.58 | -0.54 | 0.58 | -0.58 |
| *Shape Compactness* | 0.57 | 0.52 | 0.75 | -0.72 | 0.74 | -0.74 |

Tab. 1: Correlation between mined features and two FEM outputs.

Parametric Correlations are used when data is assumed to follow a normal distribution. For example, the Pearson Correlation Coefficient is the most common measure that captures the strength and direction of a linear relationship between two continuous variables. There are also nonparametric Correlations, such as Spearman, which is rank-based and is used when data is not normally distributed or contains outliers. Spearman Correlation (Spearman's rho) measures the strength and direction of a monotonic relationship between two variables (continuous or ordinal). A monotonic relationship means the variables consistently increase or decrease together, but not necessarily at a constant rate. Both Pearson and Spearman correlations range between [-1,1]. -1 means perfect negative correlation, +1 means perfect positive correlation, and zero means no correlation.

Conclusion:

The paper seeks to generalize the concept of sleeping parameters as an alternative way of using CAD to extract features for data-driven approaches in engineering design. Feature extraction has been used in data science to improve the quality of the inputs as a preprocess for machine learning. By combining the medial axis representation with Voronoi-derived circle radii, we obtain a rich set of features that capture both the skeletal and regional properties of complex shapes. These features offer valuable insights for data mining and predictive tasks across diverse engineering domains. The methodology is showcased on a crashworthiness case and is an example of how the medial axis can create new features that correlate with design performance. We conclude this concept can also improve the engineering application of data science. This will enable much more efficient mapping between input and output and will make the design loops independent of the parameterization. By leveraging this approach, engineers and designers can enhance the efficiency of design processes, facilitate iterative loops, and improve the accuracy of regression models.

*Mohammad Arjomandi Rad*, https://orcid.org/0000-0002-7894-7734
*Massimo Panarotto*, https://orcid.org/0000-0001-5216-0944
*Roland Stolt*, https://orcid.org/0000-0001-6278-2499
*Ola Isaksson*, https://orcid.org/0000-0003-0373-3720

REFERENCES

[1] J. Camba, M. Contero, P. Company, and N. Hartman, The Cost of Change in Parametric Modeling: A Roadmap, in CAD'20, May 2020, 31–35. https://doi.org/10.14733/cadconfP.2020.31-35.

[2] H. Bang and D. Selva, iFEED: Interactive Feature Extraction for Engineering Design, in Volume 7: 28th International Conference on Design Theory and Methodology, Charlotte, North Carolina, USA: American Society of Mechanical Engineers, Aug. 2016, V007T06A037. https://doi.org/10.1115/DETC2016-60077.

[3] O. Isaksson and C. Eckert, Product development 2040, 2020, https://doi.org/10.35199/report.pd2040.

[4] P. Koch, J. Allen, F. Mistree, and D. Mavris, The Problem of Size in Robust Design, Feb. 1997. https://doi.org/10.1115/DETC97/DAC-3983.

[5] J. Kim, M. J. Pratt, R. G. Iyer, and R. D. Sriram, Standardized data exchange of CAD models with design intent, Comput.-Aided Des., 40(7), 2008, 760–777. https://doi.org/10.1016/j.cad.2007.06.014.
[6] T. Robinson et al., Computer-aided design model parameterisation to derive knowledge useful for manufacturing design decisions, Proc. Inst. Mech. Eng. Part B J. Eng. Manuf., 232(4), 2018, 621–628. https://doi.org/10.1177/0954405417708218.
[7] M. Arjomandi Rad, K. Salomonsson, M. Cenanovic, H. Balague, D. Raudberget, and R. Stolt, Correlation-based feature extraction from computer-aided design, case study on curtain airbags design, Comput. Ind., 138, 2022, 103634. https://doi.org/10.1016/j.compind.2022.103634.
[8] G. G. Wang and S. Shan, Review of Metamodeling Techniques in Support of Engineering Design Optimization, J. Mech. Des., 129(4), 2006, 370–380. https://doi.org/10.1115/1.2429697.
[9] T. W. Simpson, J. D. Poplinski, P. N. Koch, and J. K. Allen, Metamodels for Computer-based Engineering Design: Survey and recommendations, Eng. Comput., 17(2), 2001, 129–150. https://doi.org/10.1007/PL00007198.
[10] M. Arjomandi Rad, Data-driven and real-time prediction models for iterative and simulation-driven design processes, Jönköping University, School of Engineering, 2022. Accessed: Jan. 10, 2023. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-56577
[11] W. Zuo, Y. Lu, X. Zhao, and J. Bai, Cross-sectional shape design of automobile structure considering rigidity and driver's field of view, Adv. Eng. Softw., 115, 2018, 161–167. https://doi.org/10.1016/j.advengsoft.2017.09.006.
[12] A. Hedlund and D. Blom, Correlation-based analysis on thin walled tubes. 2022. Accessed: Mar. 15, 2024. [Online]. Available: https://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-58398
[13] D. Al Galib and A. Limam, Experimental and numerical investigation of static and dynamic axial crushing of circular aluminum tubes, Thin-Walled Struct., 42(8), 2004 1103–1137. https://doi.org/10.1016/j.tws.2004.03.001.
[14] Xiang, Y.; Wang, Q.; Fan, Z.; Fang, H.: Optimal crashworthiness design of a spot-welded thin-walled hat section, Finite Elem. Anal. Des., 42(10), 2006, 846–855. https://doi.org/10.1016/j.finel.2006.01.001.
[15] Santiago-Montero, R.; Bribiesca, E.; Santiago, R.: State of the art of compactness and circularity measures, Int. Math. Forum, 4, 2009, 1305–1335.