Computers in Industry 138 (2022) 103634

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Correlation-based feature extraction from computer-aided design, case study on curtain airbags design

Arjomandi Rad Mohammad[a,*], Kent Salomonsson[b], Mirza Cenanovic[a], Henrik Balague[c], Dag Raudberget[a], Roland Stolt[a]

[a] *Department of Product Development, School of Engineering, Jönköping University, Jönköping, Sweden*
[b] *Department of Mechanical Engineering, School of Engineering Science, University of Skövde, Skövde, Sweden*
[c] *Autoliv AB, Vårgårda, Sweden*

ARTICLE INFO

Article history:
Received 22 June 2021
Received in revised form 7 January 2022
Accepted 11 February 2022
Available online xxxx

Keywords:
Feature extraction
CAD/CAE
Parametric models
Medial Axis
Design Automation
Machine Learning
Regression Analysis
Curtain Airbag

ABSTRACT

Many high-level technical products are associated with changing requirements, drastic design changes, lack of design information, and uncertainties in input variables which makes their design process iterative and simulation-driven. Regression models have been proven to be useful tools during design, altering the resource-intensive finite element simulation models. However, building regression models from computer-aided design (CAD) parameters is associated with challenges such as dealing with too many parameters and their low or coupled impact on studied outputs which ultimately requires a large training dataset. As a solution, extraction of hidden features from CAD is presented on the application of volume simulation of curtain airbags concerning geometric changes in design loops. After creating a prototype that covers all aspects of a real curtain airbag, its CAD parameters have been analyzed to find out the correlation between design parameters and volume as output. Next, using the design of the experiment latin hypercube sampling method, 100 design samples are generated and the corresponding volume for each design sample was assessed. It was shown that selected CAD parameters are not highly correlated with the volume which consequently lowers the accuracy of prediction models. Various geometric entities, such as the medial axis, are used to extract several hidden features (referred to as *sleeping parameters*). The correlation of the new features and their performance and precision through two regression analyses are studied. The result shows that choosing sleeping parameters as input reduces dimensionality and the need to use advanced regression algorithms, allowing designers to have more accurate predictions (in this case approximately 95%) with a reasonable number of samples. Furthermore, it was concluded that using sleeping parameters in regression-based tools creates real-time prediction ability in the early development stage of the design process which could contribute to lower development lead time by eliminating design iterations.

1. Introduction

High-level technical products (such as airbags and jet engine components) that are characterized by having iterative and simulation-driven design processes, often suffer from long development lead time. These products are associated with having a large number of manual simulation loops in the early stages of the design process (Arjomandi Rad, 2020). Supporting tools for simulation-driven design include knowledge management systems to capture both tacit knowledge and knowledge objects and create an ability to mimic the reasoning of experts in simulation and design (Fatfouta and Le-Cardinal, 2021). Additional supporting tools include creating automated systems with CAD (Computer-aided design) and CAE (Computer-aided engineering) to save time and reduce human error. This common practice is widely investigated by the fields of design science and design automation (Chakrabarti and Blessing, 2016; La Rocca, 2012). However, running complex simulations repeatedly, even in an automated manner is time and energy-consuming and can be quite unfeasible. Increasing computational power, utilization of parallel or cloud-based techniques are other effective methods to decrease simulation time but they have not been used to solve the lead time problem since in product development one simulation iteration can be dependent on the previous iteration's simulation results.

* Corresponding author.
E-mail address: mohammad.rad@ju.se (A.R. Mohammad).

https://doi.org/10.1016/j.compind.2022.103634
0166-3615/

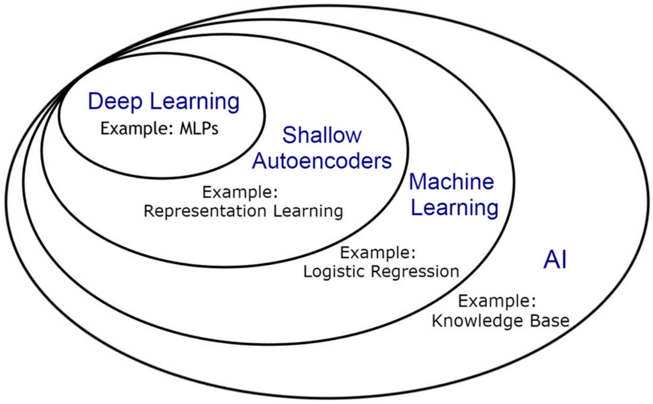

Fig. 1. A Venn diagram showing AI categorization with examples (Goodfellow et al., 2016).

Machine learning (ML) belongs to a larger family of algorithms that are part of the Artificial Intelligence (AI) branch, depicted in Fig. 1. ML is "A form of applied statistics with increased emphasis on the use of computers to statistically estimate complicated functions and a decreased emphasis on proving confidence intervals around these functions" (Goodfellow et al., 2016). Within the design processes as support, ML has been used to estimate the results of iterative assessment tasks. For instance, a method called *Kansei engineering* is a good example of the application of ML in the early and conceptual phases of product design and development. The method is used to generate the shape of the product by inputting specific design elements. The attempt tries to map the form of the product as the design variables to the feelings of consumers as an indication for the output (Fan et al., 2014; Wang et al., 2016).

Individual and independent pieces of information, being input into the system is called a *feature* in machine learning (Goodfellow et al., 2016). A large body of theoretical literature deals with considerations that ought to be made for selecting features when building predictors or classifiers. Selecting the best features has been done by measuring the relevance of the feature such as correlation coefficient, and ranking them (*filters*), assessing features' influence on the performance of the predictor (*wrapper*), or incorporating the feature selection as part of the training process (*embedded)* (Guyon et al., 2008). Feature construction (*feature extraction*) is a method that aims to build more compact features to increase prediction performance and *feature reduction* reduces the number of them intending to acquire better predictors by removing irrelevant and redundant features to defy the curse of dimensionality (Guyon and Elisseeff, 2003). All mentioned methodologies consist of many methods that are used as a pre-processing step in machine learning to improve prediction efficiency and accuracy. The proposed method in this paper of correlation-based feature extraction which is used for ranking features based on various correlation matrices (Guyon et al., 2008) is found effective when used in the process of selecting features and it best works with supervised learning methods (Hall, 1999).

Using data-driven approaches in the engineering design of consumer products has been reviewed recently (Chiarello et al., 2021). By identifying the tools, algorithms, and data sources that have been used in engineering design, the authors touch upon challenges and gaps that need to be tackled in the future. For instance, one of the listed challenges is "Identifying latent features (e.g. temporal features, behavioral features) hidden in CAD data". Indeed, real-time analysis of the design is a common practice but collecting analyses of the design and deriving performance or cost indicators, or in other words 'data mining' for new product development is still not addressed in the literature (Bertoni et al., 2017). The research gap which this paper tries to fill is to address common problems associated with using ML-based predictors in the design of consumer products, namely dimensionality and parameterization. The problem arises when creating configurations of the geometry using CAD model parameters as features (also known as inputs or variables). Often to produce samples, designers are required to fully define the CAD with many parameters and constraints and then to fully cover design space in turn, leads in having cumbersome training process.

Being obliged to follow the standard parameterization convention through the designing process naturally limits the designer and suppresses creative solutions because then the designers will be forced to follow the same parameterization convention that is used when training sets are created. Therefore, in complex geometries, designers usually avoid using any constraints and parameterized CAD models, because doing so will either limit the ability to manipulate the design shape or result in having a sparse training set in the design space. Another practical problem is that CAD designers are not the same as the CAE simulation engineers and they might not sit in the same company or work environment. This becomes an issue if a higher level of competence in each area is needed which is often the case for a high-level technical product. This fact triggers a back-and-forth work between several engineers or departments in the company and thus negatively influences the development lead time of the design processes. The conceptual phase of a design process encompasses an evaluation stage (Pahl and Beitz, 2013) in which having an independent prediction tool (A live prediction model) could make CAD designers aware of the consequences of their decisions on CAE results. This evaluation stage can significantly increase the development speed by avoiding the costly simulation loops, thus such a tool can fill the aforementioned gap. Therefore, the presented concept in this paper is an effort to overcome the high dimensionality in engineering design that is one of the common problems in applying data science in engineering design (Chiarello et al., 2021).

This paper introduces a correlation-based feature extraction application in CAD for regression-based machine learning algorithms. After an introduction that outlines the existing gap, the next section explores the related works, and the case study used throughout this paper is introduced in the third section. Using finite element simulations in the next section, a parametric study is performed to study the effect of each CAD parameter on the volume. Separately, the latin hypercube sampling method is used to generate and study a group of 100 design samples and their volumes as an output. It was shown that CAD parameters alone, would not lead to effective prediction accuracy. In the fifth section, with utilizing the concepts of different geometric entities (such as area, circumference, or the medial axis), new parameters referred to as *sleeping parameters* are defined and studied as a performance indicator for the inflated curtain airbag. It was demonstrated that new features have better correlations with the volume. And they can be extracted from geometry without any need for model parameterization which maintains freedom in design. Two regression analyses performed in the sixth section, compare and validate the performance of extracted parameters in a regression model by showing the ability of these parameters in reducing the prediction error margins. The discussion in the last section explains the effectiveness of the sleeping parameters, such as the ones studied in this paper. This will allow designers to build simple but accurate regression models with a low number of features and sample points (small training set).

2. Related works

Statistical approximating methods in engineering design (e.g. response surface methodology, Taguchi methods, neural networks, inductive learning, and kriging) has been used for a long time to address computation-intensive design problems (Simpson et al., 2001; Sun and Wang, 2019; Wang and Shan, 2007). One application of these modeling techniques is to build regression models to reduce the number of simulation iterations. But their focus is on reducing computational cost rather than reducing problem dimension. Thus, modeling techniques in engineering design do not address high dimensionality problems, and the literature in this section is very scarce. (Wang and Shan, 2007). The first metamodeling techniques to tackle High-dimensional Expensive Black box (HEB) problems utilized radial basis functions with a high dimensional model representation which basically offers an explicit function expression and thus shows the contribution of each design parameter (Shan and Wang, 2010). Since then, other metamodeling techniques that combat the curse of dimensionality are published, for instance, by improving kriging surrogates of high-dimensional design models by partial least squares dimension reduction (Bouhlel et al., 2016), by using convolutional neural networks (CNN) to study over hundred-dimensional and strong-nonlinear product design problems (Li et al., 2017), etc. This area of research aims to find better metamodeling techniques with proposing sophisticated algorithms yet due to the scope of this paper, they are not reviewed extensively in this section.

The common ground for any metamodeling method is that they require pre-performed simulation data or experiment data as input for the approximation and this makes most of them a data-driven approach. A data-driven solution is not only about the size of data under study but also more about decisions making based on data analysis and interpretation. This can be inferred from the consensus definition for the data-driven approach "Using computational systems to extract knowledge from structured and unstructured data" (Chiarello et al., 2021). As discussed in the Introduction, using CAD model parameters as input with any data-driven approach to mapping CAD input to CAE output could pull forward problems such as dimensionality and parameterization.

To overcome the mentioned problems many studies have tried to simplify the geometry to reduce the number of parameters. For instance, Wang et al. digitally created simple n-fold symmetric shapes representing cohesive contact zone in stereo-lithography (SLA) as input for a neural network. For output vector calculation they used finite element simulation and proposed a network that enabled a fast prediction model for stress distribution in the separation of the 3D printed part during the pull-up process of the bottom-up SLA technique. (Wang et al., 2018). Additionally, Yoo et al. used a large number of CAD models and CAE results for training a deep learning algorithm. The presented framework is applied to the road wheel design process, in seven stages; starting from 2D generative design based on topology optimization principles and ending in analysis and visualization of the CAE results (Yoo et al., 2020). In this framework, the ML algorithm first generates a large number of 3D CAD models by minimizing the distance from a reference design, and later the designers manually verify and select suitable designs for further modifications. Using a simple design geometry that has only three design parameters and applicability of the framework on large datasets are among the limitation of this study. Though, a simple geometry, as well as a less computationally expensive CAE method (modal analysis), contributes to the success of this approach. As mentioned earlier, running thousands of finite element simulations to build large machine learning databases will require a huge computational power. Just, for instance, assuming 5 min simulation run time for 10,000 simulations requires 35 days of run time. Other simplification methods also exist in literature such as data mining design methodology (Du and Zhu, 2018) which was suggested for high-dimensional design problems and it essentially simplifies the design space by shrinking the changing interval of the design parameters, using the decision tree technique. Approaches with simplifying geometry or abstracting problem dimensions are limiting the design (Sun and Wang, 2019) in various forms and this means the designer needs to follow certain limiting rules to have a feasible design case.

As another approach to overcome the mentioned problems, some studies in the literature have increased the number of variants drastically. For example, Ramnath et al. investigated the potential of applying data science into engineering design (Ramnath et al., 2019, 2020) by an automated method for creating big datasets of 3D CAD models. Several approaches, such as design catalogs/tables, user-defined features, and other variant creation techniques are used for generating many variants (60,000) of an automotive hood in CAD software. Their framework was based on creating a variety of configurations and then filtering out the infeasible cases which fail during CAD work (Ramnath et al., 2019). A workflow for correlating geometric configurations according to several performance and safety requirements was also demonstrated to attain validity for created training data set (Ramnath et al., 2020). However, these two studies do not propose any automatic FEA model for mesh generation, and corresponding simulations were not performed. More examples from increasing the size of samples can be found in the literature. Secco et al. used input parameters such as the wing planform, airfoil geometry, and flight condition as inputs to construct a neural-network-based prediction model for aerodynamic coefficients of transport airplanes. The output was calculated for a huge library (100,000) with a full-potential multiblock structured code with an average time of 21.8 s per airplane (Secco and de Mattos, 2017). This could have not been possible to perform if the computational time was over one minute for each simulation run because of exhaustive run time.

More data-driven approaches exist in literature with less emphasis on geometrical CAD model parameters as features. Rahman et al. used the designer's sequential design behavioral data stored in the design action log file (.JSON) of a CAD program to train a machine-learning algorithm and predict the next stage in the process as immediate design action (Rahman et al., 2019). This approach is a novel way of using CAD software as a data source and clearly emphasizes the gap in engineering design literature to explore alternative ways of using CAD to extract features for data-driven approaches (Chiarello et al., 2021). Yet, based on performed literature study, extraction of hidden features based on CAD geometry has not been proposed. But many applications of feature extraction and feature reduction exist in other domains and each of them is an independent research topic backed with a substantial number of publications. Such topics are vibration analysis and signal processing with the aim of condition monitoring in mechanical systems like bearings or gears (Caesarendra and Tjahjowidodo, 2017) or mechanical defect prediction models using either supervised or unsupervised learning methods (Kondo et al., 2019) or image processing and pattern recognition where the number of features requires a lot of preprocessing on the input images (Kumar and Bhatia, 2014) or electronic circuits design automation where feature extraction is being practiced on generating new circuit topology (structure) with reusing learned patterns (Huang et al., 2021). Sensitivity analysis methods such as Principal Component Analysis (Yuce et al., 2014) or analysis of variance (ANOVA) (Khalkhali et al., 2017) together with Taguchi are widely used dimension reduction strategies to select the most important parameters and reduce the dimensionality of the model. Yet, considering how large the number of parameters and constraints in a real CAD model and how small their effect on simulation output can be, running higher-order Taguchi arrays add up to existing computational complications.

3. Studied case

Since its invention in the early 1990s, the side curtain airbags have become an important part of vehicle restraint systems and they are widely used to prevent serious injuries by increasing head

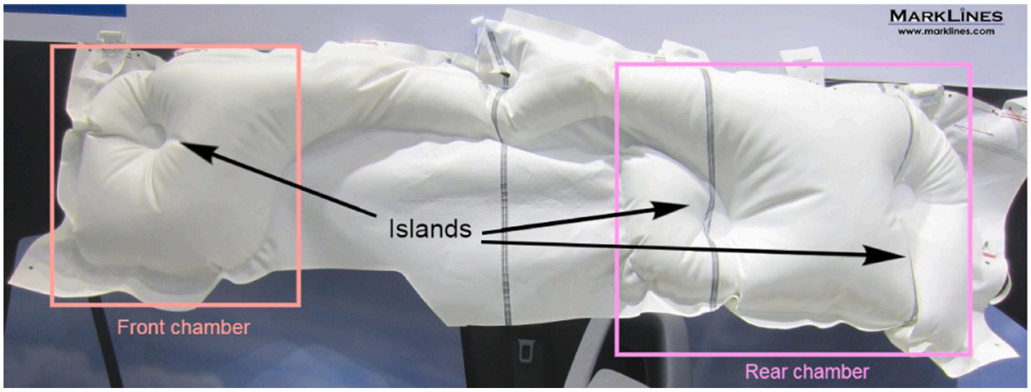

Fig. 2. Typical curtain airbag design for sedan class.
(with copyright permission from MarkLines Co.).

protection for both front and rear seat occupants. For the US market, one measurement is the safety criteria (Federal motor vehicle safety standards, 2011) that include measuring the amount that a human head can go out of the widows (in millimeters) and is called, Ejection Mitigation (EjM). One of the principal requirements for inflatable curtains is to shield the occupants from intruding objects (Evans and Leigh, 2013). The coverage area must be communicated to airbag manufacturers (Suppliers) from car manufacturers (OEMs) as requirements to be met during the design process. This is considered for various occupants and positioning of car pillars, roof rail, door glass, occupant seats, and components in front of occupants such as dashboard and steering wheel. A typical curtain airbag design for a sedan class is shown in Fig. 2. The *front and rear chambers* (marked in the figure) are responsible for cushioning the head of the occupant by filling the space between the head and windows. Each chamber has one or two so-called *islands* (inner sewing lines) to prevent the cushion from becoming a balloon (or like a pillow). The non-inflated fabric around the chambers is responsible for holding the integrity of the whole bag and it protects the occupants from broken glass and other intruding objects at the time of the crash.

The number of islands and the size of chambers depends on the size of the designed bag which in turn is affected by the size of the car. So, for a large SUV with three rows of seats, another chamber could be added to the bag shape and a higher capacity of inflators might be necessary to fill in the bag. Likewise, for a small coupe vehicle, designers could design one big chamber instead of two and consequently choose a lower capacity. It also influences the cost of the bag as suppliers (airbag manufacturers/sellers) tend to sell the bigger capacities while OEMs (car manufacturers/buyers) try to settle on a smaller one. This is because it is easier to meet the safety requirements with bigger capacities, but it will be also more costly. The design process continues back and forth until a design case meets all requirements (Dix et al., 2012).

The curtain airbag design process is characterized as being iterative and simulation-driven (Arjomandi Rad, 2020). Meaning that from early conceptual phases designers are front-loading simulations to meet requirements such as volume and EjM, etc. Finite element simulations are used with separate simulation models (varying in complexity) for each requirement. Fig. 3 shows how coverage requirement (req.) is met first in a CAD environment and then volume and EjM are calculated within a separate finite element simulation model. Looping between volume and coverage or looping between EjM and coverage could happen as frequently as 50–60 loops and the design process will continue until all three are satisfied. This looping between requirement gates in the design process is a common workflow for components in the automobile industry (Fatfouta and Le-Cardinal, 2021). Considering the time spent by engineers for pre-processing, processing, and post-processing, the importance of an automated prediction model in early phases is highlighted.

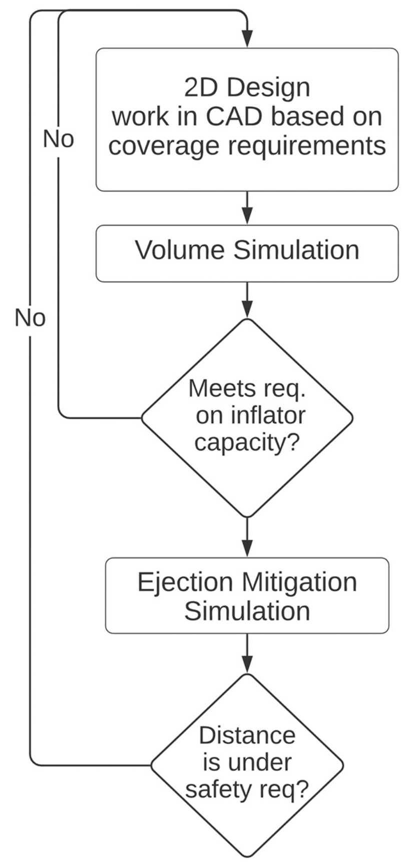

Fig. 3. The inflatable curtain design process in the conceptual phase.

There has been a lot of research performed to study curtain airbags. Song et al. introduced three types of simplified models in curtain airbags mainly intending to save modeling and simulation time (Song et al., 2011). It was argued that due to a variety of requirements from different safety organizations, there is a need for different configurations with regards to positioning impactors and dummies in the simulation models. In a different study, the airbag shape design has also been studied (Chavare et al., 2013) to increase the performance of curtain airbags by applying Knowledge-Based Engineering (KBE) methodology on the packaging and positioning dummy and seat with respect to a set of interior design requirements. The result was a curtain airbag module configuration that ensures protection zones for occupant body regions. Furthermore, they used 3D clearance/interference volume between dummy, seat, and vehicle side combined to arrive at the airbag cushion geometry, volume, and inflated chamber size. Yun et al. presented the curtain airbag design procedure (Yun et al., 2014) with two phases to first select important parameters affecting head impact criteria (HIC) using sensitivity analysis, and second to optimize the function based on Taguchi orthogonal arrays to minimize HIC. Another study (Park, 2017) aims to establish a design procedure and an optimization process for airbags using CAE techniques mainly to minimize development time. Parameterized airbag shape and morphing techniques were used to generate surrogate sled models. The direct optimization method (not meta-model-based) is used for airbag shape optimization regarding multiple load cases.

In this paper, to study how parameters are affecting the volume of the airbag and to find out the most influencing parameters on the simulation output, a subsystem of an actual airbag is considered as a generic prototype. The shape of the designed prototype is inspired by the front chamber of an actual curtain airbag depicted in Fig. 2 and it holds most of the features necessary for full analysis of an actual curtain airbag shape. Fig. 4 shows the designed prototype with 14 selected parameters which is being studied in this paper. Moreover, Table 1 presents the names of the parameters which are tagged with numbers to ease the lookup.

Design parameterization guidelines including the *Independence axiom* and *Information axiom* were used when selecting the mentioned 14 parameters (Suh, 1998). The Independence axiom maintains the independence of design intent, and the Information axiom attempts to minimize the information content of the design intent (Chang, 2016). To meet the independence axiom all 14 parameters are bound into an interval that allows them to change between a minimum and a maximum range. Table 1 presents all selected parameters and their associated bounds. To meet the information axiom all parameters are selected in a way that allows all possible design cases with a minimum number of parameters. For example, among all the variations of parameters that could have described an island's shape and position, 5 parameters are selected based on their ability to create the most frequently used shapes and the rest of the parameters are assumed constant. For instance, the radii and the control points for the curves on top of the island are assumed constant because they are rarely changed in the actual curtain airbag design.

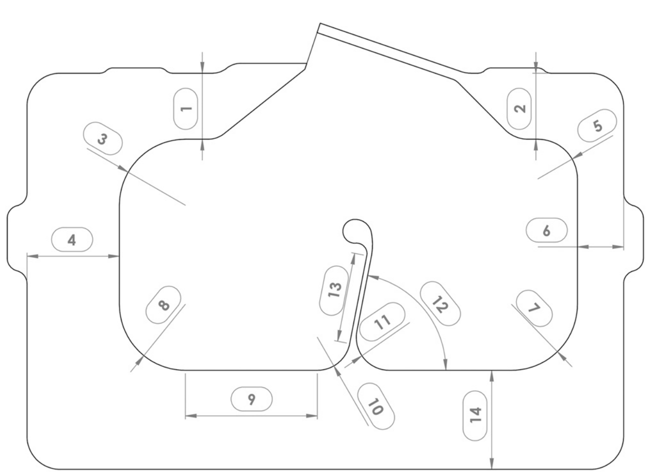

Fig. 4. Studied prototype representing all curtain design features.

Table 1
Selected 14 CAD parameters and their varying bounds.

	Parameter Name	Min (mm)	Max (mm)
1	Offset1	50	100
2	Offset2	50	100
3	Radius1	40	135
4	Offset12	40	135
5	Radius2	40	60
6	Offset22	35	70
7	Radius3	40	100
8	Radius4	60	100
9	IslandOffset	100	250
10	IslandR1	10	50
11	IslandR2	10	50
12	IslandAngle	40°	130°
13	IslandLength	150	220
14	IslandBottom	40	180

4. Problem definition, sampling techniques

This section deals with identifying the problem area. To study parameters and investigate each parameter's effect, two studies have been carried out. *One factor at a time* study, where only one parameter at a time is changed, and *latin hypercube* study where all the parameters are changed with the help of a latin hypercube sampling method. In both studies utilizing macro tools in CATIA® and Visual Basic for Applications (VBA) programming, a parameterized CAD model is modified based on each design sample and the generated geometry is exported as a '.igs' file. The choice for this file format was based on the experience of the designers in the industry. As for the simulation technique, there are several finite element (FE) codes in the literature to simulate airbag deployment but studying which is superior over the other is out of this paper's scope. Yet for the current paper, the *uniform pressure method* is used mostly because it is faster to create and run and it gives sufficient precision for early design phases. This method assumes uniform pressure and temperature everywhere inside the airbag. This is a close approximation of the airbag after it is fully inflated and stabilized so the airbag geometry is considered without any fold (Zhang et al., 2004). In this way, All the generated geometries in both studies are used in this finite element analysis (FEA). Meshing is done with ANSA® and generated key files are submitted to the LS-DYNA solver. Post-processing is carried out with META® and the volume-time curve is extracted for each design sample. Fig. 5 shows one of the curves as an example for the volume-time ratio.

In all the simulations the pressure is raised to 40 kPa with a smooth step function and then kept constant until the simulations end in 100 ms (ending criteria). From Fig. 5 it is clear that the maximum volume (recorded as 33.7 Liters) is reached in the last 20 ms of the simulation. At this time where the maximum volume is reached, the pressure is maintained constant at 40 kPa. The run time for each design sample ranges between 10 and 15 min, depending on

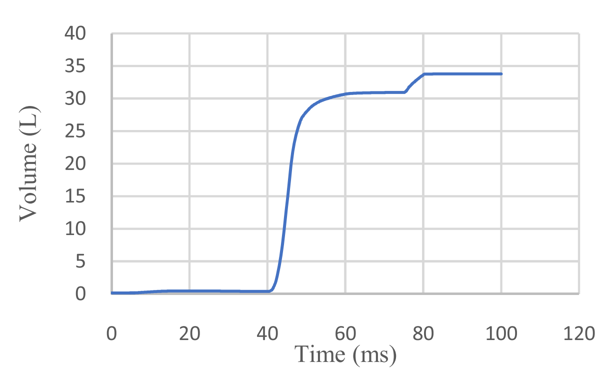

Fig. 5. Volume - time ratio for one of the performed simulations.

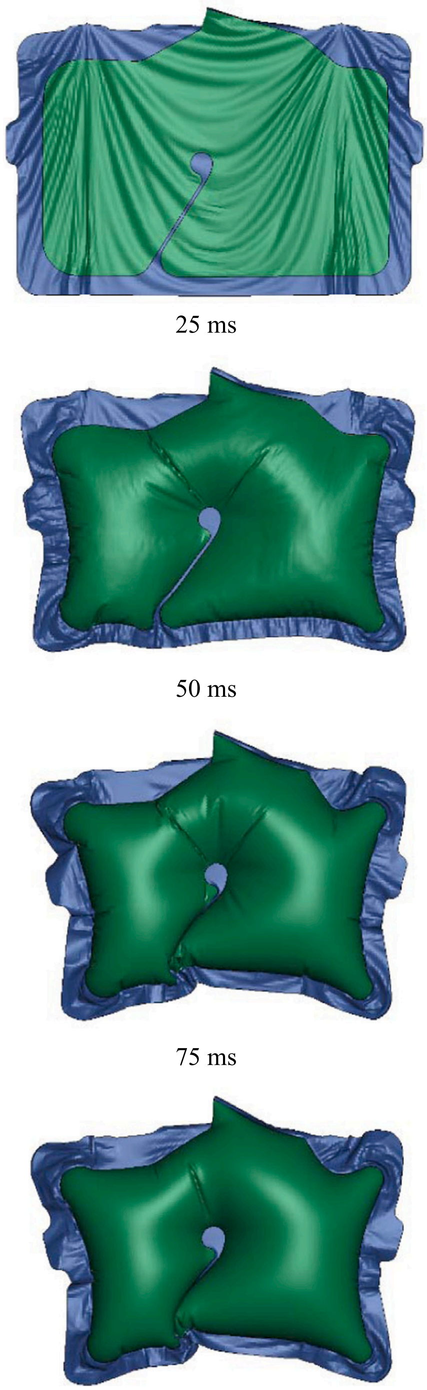

Fig. 6. Studied prototype simulation in a 25 ms step time.

the size of the bag. Fig. 6 shows one of the simulations performed on a 25 ms interval.

In the *One factor at a time* study, 14 aforementioned parameters (Table 1) on the geometry have been altered in five steps, so the results for volume change are shown in Fig. 7 for the total of 70 simulations. During the study, only one parameter is changed, and others are kept constant. In this figure, the vertical axis shows the volume in terms of liter and the horizontal axis represents the 5 changes in each parameter, normalized with respect to the average of the five steps (each parameter's average). The length of steps and in horizontal axis shows how much that parameter has been changed in comparison to others. For example, the smaller length ('*Radius2*') shows that the parameter had a smaller boundary to change in CAD due to the defined constraints. The slope of the curves shows the sensitivity of the volume to each parameter's change. For instance, '*Offset12*' and '*Offset22*' have more or less the same sensitivity. As it can be inferred from the figure, length-like parameters (e.x. *Offset1, Offset2, etc.*) are linearly correlated with the volume and the radius-like parameters (e.x. *Radius1, Radius2, etc.*) are correlated quadratically. As expected, the more a parameter is affecting the area of the geometry, the more it is changing the volume. Another interesting effect happens with the parameter '*IslandOffset*' and '*IslandAngle*' where the volume with their increase first decreases and then increases. This behavior can be explained considering the thickness of the bag (how much it becomes inflated).

It should be mentioned that throughout this paper, for evaluating the correlation between two sets, Pearson coefficient (R^2) which is a popular method in machine learning for ranking features and filtering them has been used (Guyon et al., 2008). This correlation coefficient for a feature with values x and output with values y is defined as Eq. (1).

$$R^2 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \quad \sum_i (y_i - \bar{y})^2}} \tag{1}$$

Where $\bar{x}$ is the mean of feature data set, and $\bar{y}$ is the mean of the output data set. If R^2 is 0, it means that there is no correlation, and input parameters cannot predict the value of the output. Similarly, if its value is 1, it means that input parameters will always be successful in predicting the output. The R^2 value is always $0 < R^2 < 1$ and for this criterion 0.9 or above is considered as excellent precision, 0.8 or above good, and in some cases, 0.6 or above is considered satisfactory (Rad and Khalkhali, 2018).

In the second study, a latin hypercube is used to generate design samples. A python module called 'diversipy' (Wessing, 2018) is used as an implemented version of latin hypercube to create 100 normalized design samples between the aforementioned bounds for the parameters. All cases are mapped into desired intervals and using CATIA knowledge ware bench work, a parameterized model is modified with a VBA script, and all the CATIA models are generated the same as described before. Fig. 8 illustrates the three most correlated parameters (out of 14) with the volume for 100 design samples generated with the latin hypercube. As can be seen, these CAD parameters are correlating with the volume with the R^2 value of 0.037, 0.018, 0.044 which means almost no correlation. Other parameters are also similar to the depicted ones, so they were not included as they don't yield more information. This lack of correlation among all CAD model parameters and the simulation output can be explained by the effect of each parameter on the volume change. As it can be inferred from Fig. 7 some parameters have so small effect (for example, '*Radius 3*' or '*Radius 4*'), and some have very high (for example, '*IslandLength*' or '*IslandAngle*'). If one parameter with high impact increases the volume and one other parameter with low impact reduces it. The effect of the one with low impact will be faded or maybe even out with each other when they are changed together. This problem with these parameters will affect the accuracy of the machine learning regression model negatively because the regression function will not be able to properly map inputs to the output.

In other words, the algorithm might have difficulties finding a meaningful relation between input and output. Consequently, using low correlated features will require either having large training sets or having a higher number of features. As mentioned in the introduction, filtering methods are a group of feature selecting methods that are based on the increased correlation between features and the output (or label tags in classification problems) (Cai et al., 2018). The feature extraction on CAD proposed in the next

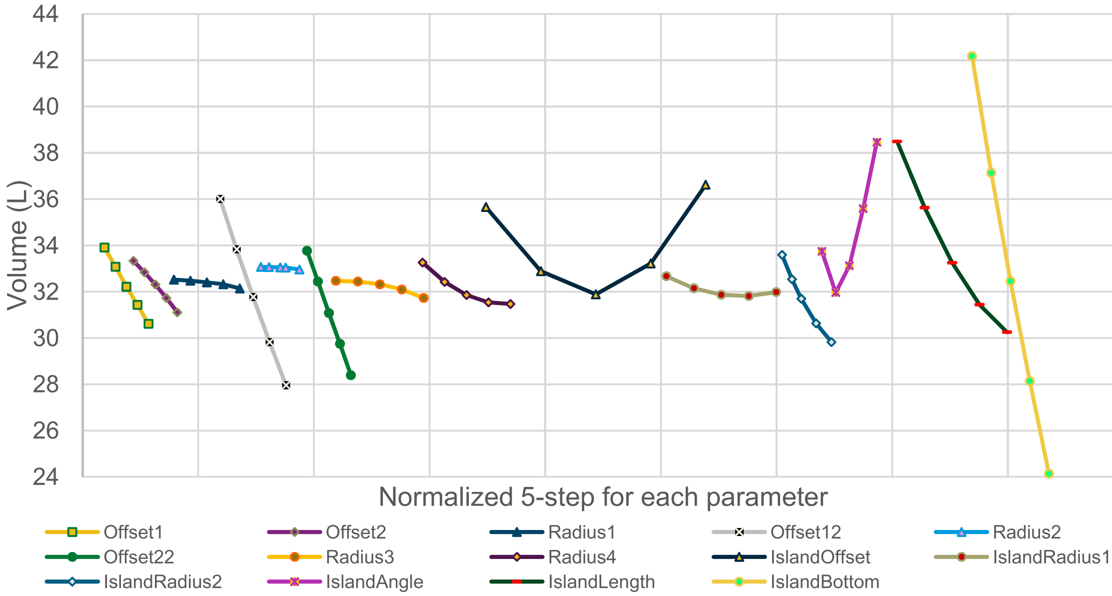

Fig. 7. One factor at a time study on fourteen selected parameters.

section is using the filtering method as a base. Some new parameters are extracted and ranked according to firstly being independently (from the parameterization) measurable and secondly having a high correlation with the volume as output. This will ensure their impact when they are used to create a machine learning regression model and will increase the accuracy of such a model which is proposed and verified in the last section.

5. Sleeping parameters

The parametric study in the previous section shows how much area can be effective in the final volume output, which is no surprise. Yet, it refers to more possible geometrical entities associated with the airbag shape which can show a correlation with the volume. Therefore, a more in-depth study is carried out on finding more parameters with the same characteristics, and the results are provided in this section. Overall, underlying parameters of this kind that can be obtained independently and fast from the geometry are an example of a feature extraction application for machine learning in the CAD environment. In this paper, the term *sleeping parameters* is used to address them and more examples of such parameters are presented further in the text. They are called 'sleeping' because they are not primely linked to the CAD model parametrization, and are derived from the model without any special treatment before or during the design process. The extraction can simply be done after the design is finished and when the design is ready for the simulation stage. To gather a handful of Sleeping parameters, a workflow as illustrated in Fig. 9 has been used during the next sections. In this figure, if a geometric entity or an extracted parameter satisfies two conditions, it is added to a non-dimensional array (Ndarray). Later, we use all extracted parameters to train a regression model and measure its accuracy. This is looped until we reach the expected accuracy.

5.1. Area and circumference

Using the 100 samples acquired from the latin hypercube in the previous section, a CATIA VBA script is used to read out the area of each model. The correlation between area and volume is studied and depicted in Fig. 10 (top). The figure also demonstrates the correlation coefficients between two sets. The amount of correlation in this figure makes the area an interesting parameter to estimate the volume in the early design phases. The figure also shows when the area is increased so does the volume. It has been proven that the increase in volume is always greater than the increase in the surface area (Emert and Nelson, 1997). This is true for cubes, spheres, or any other polyhedron object whose size is increased without changing its shape (only undergo geometric change and not topological change). Additionally, two upper and lower boundary lines are obtained and shown in the figure with their corresponding functions. Using the 70 samples from the parametric study in the previous section and the same script Fig. 10 (bottom) is acquired, which shows the correlation between the volume and the area for mentioned samples. In this figure, the purple points show the area of the design samples in

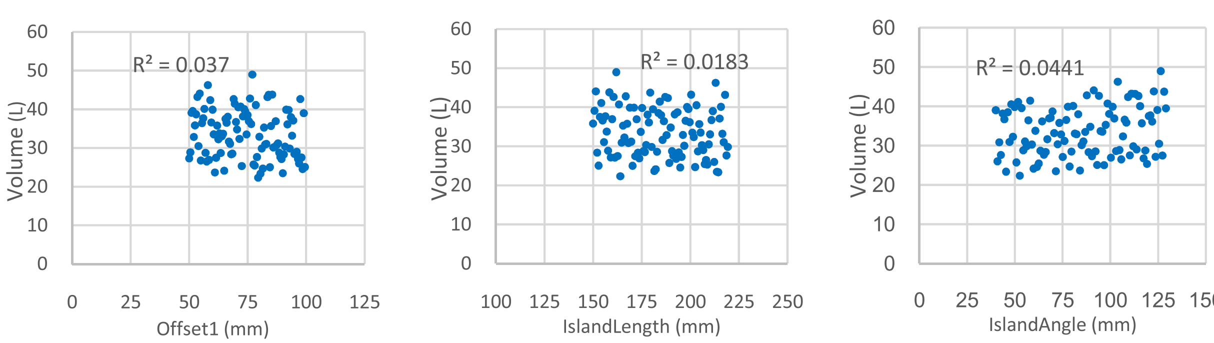

Fig. 8. Three example parameters out of 14 in the second study with 100 design samples.

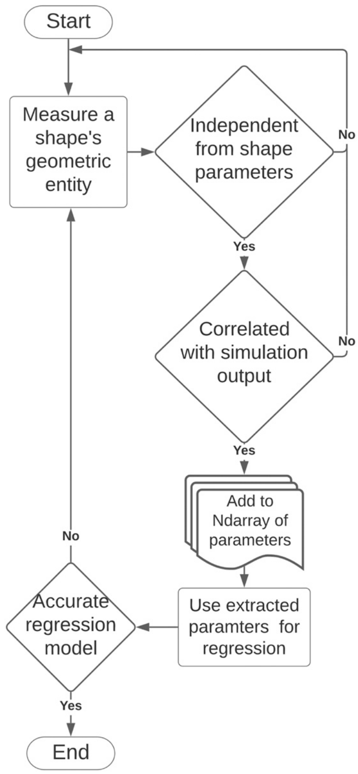

Fig. 9. The process used in this section to extract Sleeping parameters.

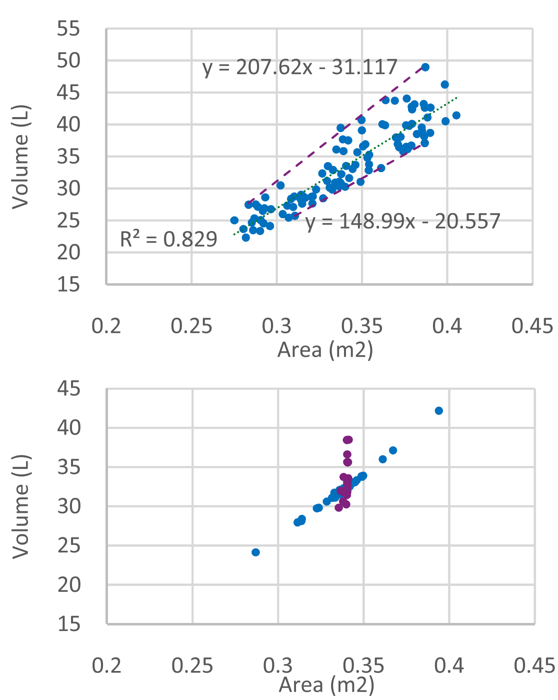

Fig. 10. Correlation between area and volume (top: latin hypercube study, bottom: One factor at a time study).

which only island-associated parameters have been changed. These parameters are namely *IslandOffset*, *IslandR1*, *IslandR2*, *IslandAngle*, and *IslandLength*. Blue points show samples associated with the change of other parameters. Two differentiated sets of parameters clearly display a different behavior with the volume.

As another instance, the circumferences of the 100 bags (same as area) are studied, and the results for correlation between

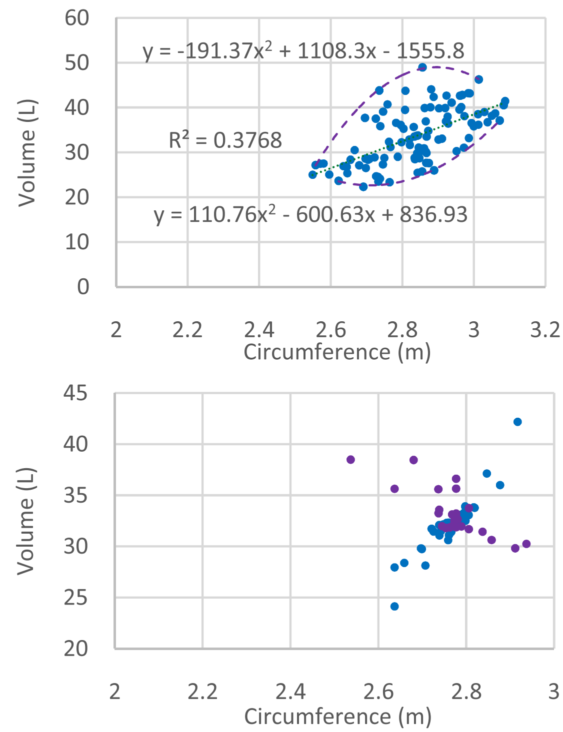

Fig. 11. Correlation between circumference and volume (top: latin hypercube study, bottom: One factor at a time study).

circumference and volume are shown in Fig. 11 (top). Upper and lower bounds are quadratic and with small deviation at the beginning and end of the plot (small and large circumferences) and large deviation in middle range circumferences.

Moreover, to find out the reason behind observed behavior, the circumference of the design samples from the parametric study, are separated into two sets in the same way with the same color code as illustrated before and the results are presented in Fig. 11 (bottom). Interestingly, the same island parameters, namely *IslandOffset*, *IslandR1*, *IslandR2*, *IslandAngle*, and *IslandLength* are the reasons for this behavior. In this figure, purple points are associated with the simulations where the island parameters are changed, and the blue points are depicting the rest of the simulations (where other parameters are changed). To find out more parameters other geometric entities have been taken into account in the next section.

5.2. Using medial axis to extract more parameters

To find out parameters that can represent thickness, the medial axis length of the bag shape in 2D is studied. The medial axis (also known as the topological skeleton) is a fundamental geometrical entity, first proposed by Blum (Blum, 1967) to describe a shape. Utilizing this concept allows representing a virtual shape by geometric location of the center of circles inscribed inside instead of its outer boundary. The medial axis is represented in a 2D planar as a line and in 3D, as a surface. What follows is the mathematical definition of the medial axis for a 2D object (a simple polygon). Let *G* denotes the boundary of a 2D object, then the medial axis *M(G)* is defined by a set of points like *y* (see Equation 2) which is tangent to the boundary G at two unique points like *x* and *x′*. And as such, these points must be equidistant to the medial axis point y. This distance can be measured by using a distance function $d_x(y)$ which shows the distance between any *x* and *y* (Ramamurthy and Farouki, 1999) as shown and Fig. 12.

$$\text{Distance function}: d_x(y) \overset{\text{def}}{=} \min_x \|x - y\| \tag{2.1}$$

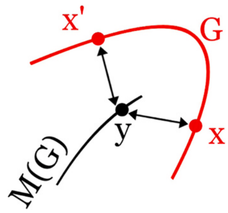

Fig. 12. Medial Axis definition.

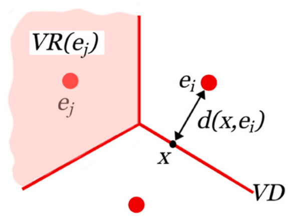

Fig. 13. A simple Voronoi diagram defined by three sites.

$$\textit{Medial Axis}: M(G) \stackrel{\text{def}}{=} \{y;\ \exists\ x \neq x' \in X,\ d_x(y) = \|x - y\| = \|x' - y\|\} \tag{2.2}$$

Voronoi diagram (VD) which is another fundamental geometrical entity associated with a closed bounded planar domain, can be used to obtain the medial axis. As defined in the mathematical definition shown in Equation 3. Let e_i denotes a nonempty site in a collection of sites, E in the space $\mathbb{R}^2$. The Voronoi regions $VR(e_i)$, is the set of all points whose distance to e_i is not greater than the distance to any other site $e_j \in E$, where $j \neq i$. Consider a distance measure $d(x, A)$ shown in Fig. 13 which denotes the distance between point x and the subset A, this is typically the Euclidean distance (Fabbri et al., 2002). The Voronoi region is then given by the definition

$$\textit{Voronoi regions}: VR(e_i) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^2,\ d(x, e_i) \le d(x, e_j),\ \forall\ e_i \neq e_j\} \tag{3.1}$$

The Voronoi diagram is then given by the union of the boundaries of the Voronoi regions

$$\textit{Voronoi diagram}: VD(E) \stackrel{\text{def}}{=} \bigcup_i \partial VR(e_i) \tag{3.2}$$

Voronoi diagram and its close associates such as the medial axis are typically grouped under the term 'Skeletons'. They have been extensively studied and used in a wide range of applications such as shape matching, surface reconstruction, dimensional reduction (Suresh, 2003), morphing (Ilies, 2006), mesh generation (Yan et al., 2013), etc.

In this paper, the Voronoi component in commercial software Rhino/Grasshopper has been used to create the medial axis of the 100 design samples generated by the latin hypercube sampling method. Grasshopper is a visual programming tool that has been used in several applications, such as in design science to solve design automation transparency problems by displaying input/output relations on a canvas (Heikkinen, 2021). Using the built-in component for python scripting in grasshopper, each design sample is imported as a '.igs' file and the boundary of the shape which is a closed-loop curve is split into 300 points (Fig. 14-A). Using these points as the center of circles for Voronoi regions and by increasing the radius, it is possible to create Voronoi regions with all the center points in the edge of the geometry (Fig. 14-B–E). Within higher radii, two meeting regions create a curve that has two close points on the geometry and correctly approximates the medial axis (Fabbri et al., 2002). Through a selection procedure in grasshopper, the medial axis is isolated and measured for its length (Fig. 14-F). The length of the medial axis is studied to find out the correlation with the volume of the bag when being inflated.

Fig. 15 shows the plot for the length of the medial axis in correlation with volume for 100 design samples created by the latin hypercube sampling method. As it can be seen, when the length of the medial axis increases so does the volume.

Although the derived medial axis length for each design sample due to its correlation with volume is a good feature to be used in regression models. However, it does not represent the largeness of the areas that are inflated 'balloon-ability'. Another drawback is that it is not sensitive to the small changes on the island. If the island moves a little bit toward the right side (change in parameter 9 of Fig. 4), it has a negligible effect on the length of the medial axis but since the left area gets bigger, the pressure could create a bigger balloon in that region and the volume could increase. To solve these problems the maximum radii among the circles that are used to generate the medial axis is derived and the result is plotted with respect to volume in Fig. 16. The figure shows 'maximum radius' is a better representative of the balloon-ability since it can have a better reflection of the island's movement on the volume, and thus it can be useful in the regression analysis.

As it is demonstrated in Fig. 14 medial axis is the collection of points that are created when two circles meet each other in the middle of the geometry. To study the characteristics of the circles that construct the medial axis, all circles inscribed in the geometry are drawn as illustrated in Fig. 17-A. As it is shown in this picture, the medial axis is exploded into equally distanced points and circles are drawn on the *x*-*z* plane to be tangent with the nearest edge. Then, all circles are flipped around the *x*-axis making their plane change from the *x*-*z* plane to *x*-*y*. This is like drawing them in a perpendicular plane to the bag's geometry which is shown in a perspective view in Fig. 17-B. The new circles prove that they are representing the volume better (since they are correlating better) than previously extracted parameters, this representation is clear in Fig. 17-C which is the top viewport of the same circles in Fig. 17-B. The top view (looking into the *x*-*y* plane from above) shows that whenever one side of the bag has an opportunity to get bigger, the size of the radius of the circles inscribed also grows, and this aligns well with depicting the degree to which the bag becomes balloon-like. Interestingly, the radius of the circles inscribed (see Fig. 17) in the geometry has a potential application to be extracted as an independent feature. However, for ease of the process, the circumferences of all the circles are calculated as a coefficient of the radius.

Additionally, the maximum radius from Fig. 15 only represents one side of the island (the bigger one), and thus, the effect of inscribed circles in the smaller chamber is not represented with this parameter. To address this problem, and to fully benefit from the radius of these circles, the sum of the circumference of all the circles inscribed in the geometry is plotted against the volume in Fig. 18.

To illustrate how much building a regression analysis on extracted features (sleeping parameters) can increase the efficiency and precision of the regression models, the next section presents a comparison and analyses on the performance of these parameters in regression models.

5.3. Regressions and analyses

The presented framework in the previous section for extracted sleeping parameters is all about finding independent (from parameterization), and measurable features that show a high correlation with the output volume. To illustrate this, a comparison between 3 sleeping parameters with the 3 CAD parameters is presented in Table 2 which shows a big improvement in the correlation of the extracted features with the volume as simulation volume.

In addition, a Multivariate Linear Regression (MLR) analysis together with the gradient descent method is used to create a prediction model for the simulation output (volume) using extracted

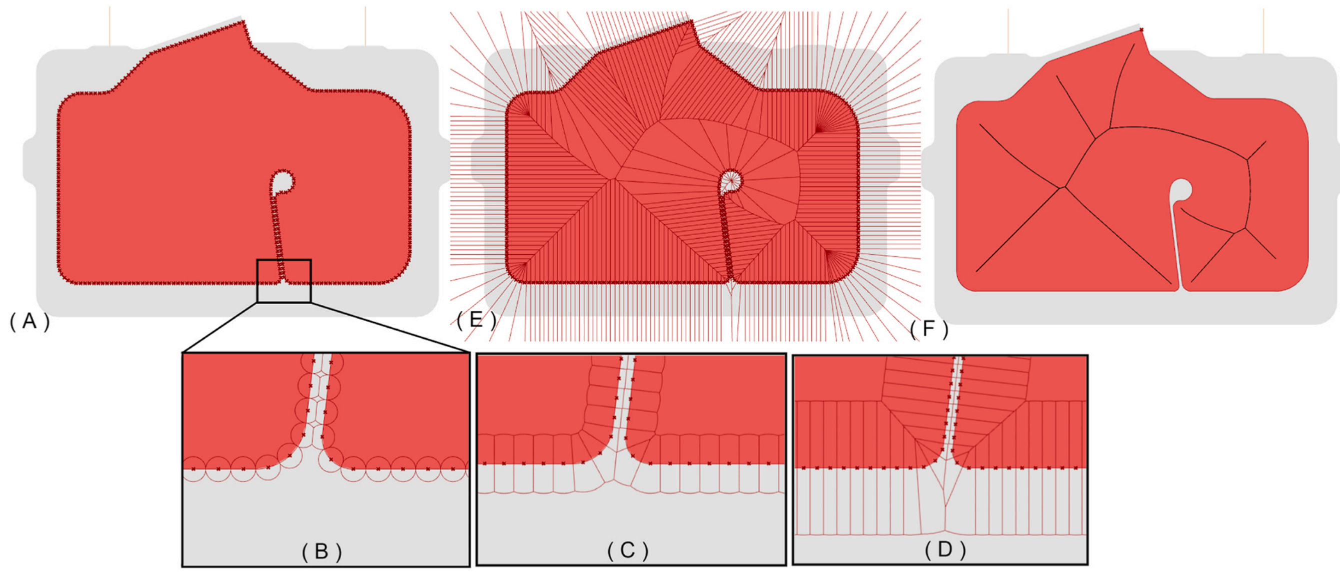

Fig. 14. Using Voronoi diagram to calculate medial axis in Rhino/Grasshopper.

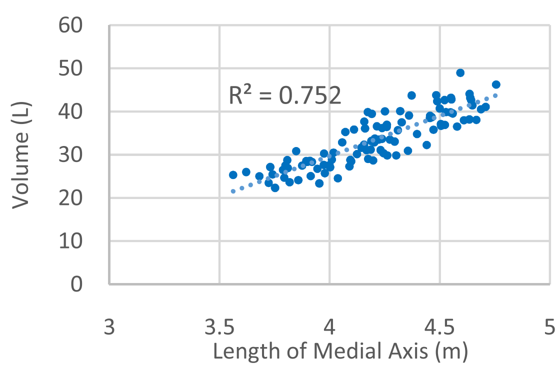

Fig. 15. Correlation between length of medial axis and volume.

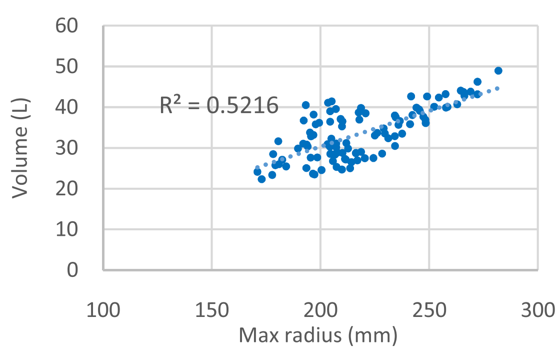

Fig. 16. Correlation between maximum radius and volume.

features. The reason for using this method is that multivariate analysis is one of the simplest machine learning algorithms available. If sleeping parameters with such a simple model result in good accuracy, then it proves and validates the hypothesis about how these features are better inputs than direct CAD parameters in regression-based machine learning for envisioned prediction models. Another advantage with choosing this method is that a simple regression can be easily implemented by any CAD designer or industrial designer utilizing ready-made libraries and modules. The applicability serves the aim of being utilized as a tool for the early phases of the design process.

Two models are trained through multivariant linear regression. One model with 14 CAD parameters is introduced in Table 1 and the other model with 3 of the extracted sleeping parameters from the previous section (the selection was based on better correlation criteria) namely: Area (Fig. 10 top), Length of the medial axis (Fig. 15), and Sum of circumferences of all circles inscribed (Fig. 18). In the training process, for minimizing the cost function, gradient descent is chosen because of its performance in convergence. The cost function (J) is used with both models, where it deducts y (the original output) from the hypothesis h_θ (the predicted output) in Equation 4. In the $h_\theta(x)$, the weights of the prediction model are denoted by $\theta_0. ..\theta_n$ and $x_0. ..x_n$ are the response variables. n is the number of variables, m is the number of data points and α is the learning rate.

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + ...+ \theta_n x_n \tag{4.1}$$

$$J(\theta_1, \theta_2, \theta_3, ..., \theta_n) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i)^2 \tag{4.2}$$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \tag{4.3}$$

For both problems, the training was performed in 5 folds. In this way, data for 100 design samples are normalized and then divided into 5 sections. Minimum (worst) accuracy among all sections is reported as the output precision. Each section with a total of 20 samples is used as a training set and one section with 80 as a testing set. The low number of training samples was chosen to make the training a little challenging and show the performance difference between the two problems clearly. As for learning rate, 5 different learning rates were chosen based on best practices in literature and subsequently were tried out and one was chosen as $\alpha = 0.2$, based on its performance in minimizing the cost function for both problems. The convergence rate of the algorithm using three sleeping parameters is compared to its rate when using 14 CAD parameters and the result shows slightly better convergence for the sleeping parameters as depicted in Fig. 19.

For the regression model trained by sleeping parameters, final values for θwere extracted according to Eq. (5) where y implies the output volume and the x_1 is the area, x_2 is the length of the medial axis and x_3 is the sum of circumferences of all circles inscribed.

$$y = 33.36 + 1.87x_1 - 1.26x_2 + 3.35x_3 \tag{5}$$

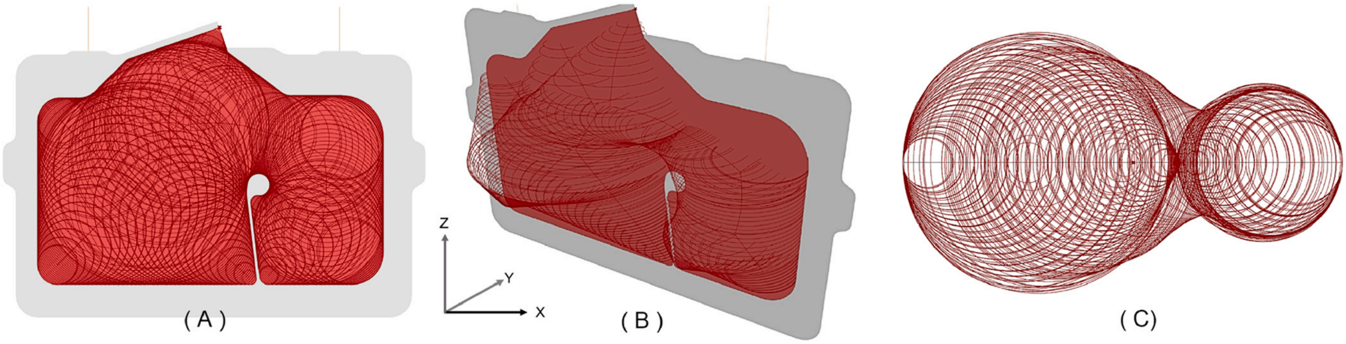

Fig. 17. Study of the circles inscribed in geometry that are used to generate medial axis.

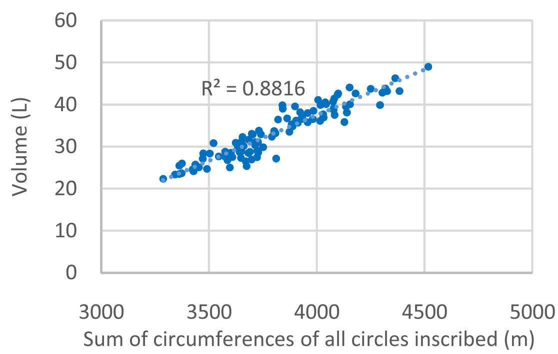

Fig. 18. Correlation between the sum of circumferences of all circles inscribed and volume.

Table 2
Comparison of the correlation between CAD and Sleeping parameters.

Name of the parameter (refer)	R^2 Correlation with the output (Volume)
Offset1 (Fig. 8)	0.037
Island Length (Fig. 8)	0.0183
Island Angle (Fig. 8)	0.0441
Area (Fig. 9 top)	0.829
Length of the medial axis (Fig. 14)	0.752
Sum of circumferences of all circles inscribed (Fig. 17)	0.8816

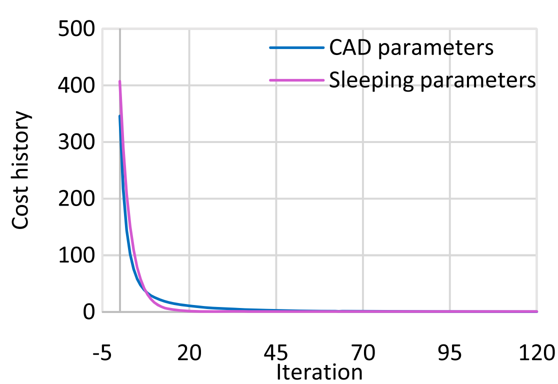

Fig. 19. Convergence of gradient descent with learning rate = 0.15 using sleeping parameters.

To maintain a high level of interpretability and to further demonstrate the effectiveness of the proposed approach a Support Vector Regression (SVR) is applied to the same training data. In contrast to MLR that fits a line, SVR tries to fit a street of lines (hyperplanes), a characteristic that equally penalizes high and low misestimates and makes the algorithm robust to outliers. It also enables SVR to free computational complexity from the dimensionality of the input space (Awad and Khanna, 2015). Even though the underlying optimization is complex, but the application is easy to employ if one uses ready-made implementations. The python module 'scikit-learn' (Pedregosa et al., 2011) implementation of SVR was used throughout this study since it is a free and open-source library and has good interoperability with other python libraries. Three different kernels namely, polynomial, radial basis, and linear are tried out and the best performance that belonged to polynomial was chosen with commonly used hyperparameters from literature. Other model hyperparameters are also selected based on best practice from literature and trial and error on the dataset.

Each regression model (MLR and SVR) is fed one time by all 14 CAD model parameters and the second time by the selected 3 sleeping parameters and the results are compared. The error between the result of the volume from the prediction model with the result acquired from FEA models (i.e testing data) is studied with common accuracy matrices in machine learning, the MSE, and R^2 coefficient and the result is shown in Table 3.

Mean Squared Error (MSE) is taken into account to make sure the difference between two compared sets is significant. The MSE metric measures the squared and averaged number of differences between predicted and expected sets. Since this metric is on the scale of the data point, its high number shows higher error or lower accuracy (Rad and Khalkhali, 2018). In this criteria, since there is no optimal range, the lower the error rate the better, and 0 means the model is perfect.

6. Discussion

In this section, the results from two previous sections will be discussed, and some important design keynotes will be highlighted. The correlation coefficient for area and volume is $R^2 = 0.829$ as is depicted in Fig. 10 (top). It can be inferred that with calculating area from CAD software (without any FEA simulation) using two equations for lower and higher bounds shown on the figure, designers will be able to have a rough estimation on volume, however, the error margin will be high according to error criteria. Moreover, Fig. 10 (bottom) reveals that without islands or inner sewing lines the 'area' would be sufficient in predicting the 'volume' since there is a highly linear correlation between them. And this is expected behavior since without the inner sewing lines (the ones that create the inner island) the bag shape will become similar to a box shape with round edges. Moreover, it can be inferred that the deviation from the trend line in Fig. 10 (top) is the result of a change in these island-associated parameters. This information can enable designers to have a better understanding of the decisions that they are making in the concept phase when meeting requirements over coverage and volume. Overall, independence from other CAD parameters when measuring the area and high correlation with the volume makes the

Table 3
Comparison of the accuracy between the regression model trained by two sets of parameters.

	Multivariant Linear Regression		Support Vector Regression	
Accuracy of the regression model among predicted and expected sets	All 14 CAD model parameters	Selected 3 Sleeping parameters	All 14 CAD model parameters	Selected 3 Sleeping parameters
R^2	0.6318	0.9505	0.8027	0.9544
MSE	14.7304	1.8802	14.3419	1.7784

area an interesting sleeping parameter to be used in regression analysis.

As it can be inferred from Fig. 11 (top), circumference with volume shows a weak correlation $R^2 = 0.37$ but the study is interesting from the design perspective. This parameter gives intuition on the behavior of the island and can help the designers have a better estimation when working with island design. As shown in this picture, if the change in bag dimensions results in increasing circumference, the volume increases rapidly which verifies previous findings. Yet from Fig. 11 (bottom) it is inferred that with increasing the circumference in the island's dimensions the volume is decreasing. This can be explained since the bigger size of the island limits how much the bag can get larger which in turn means less thickness for the bag.

The medial axis length also shows a good correlation with volume with $R^2 = 0.75$ as depicted in Fig. 15. This parameter also is interesting because it is possible to calculate it like area with the help of a raw geometry independent from the parameters that are used to generate the shape. And therefore, has been considered as another sleeping parameter to build regression models for predicting volume in the previous section. The high correlation can be explained since longer the medial axis naturally means that the geometry has narrow and twisted chambers and in other words, smaller chambers will lead to having a lower thickness. However, if the geometry has a big open area that can be inflated and become thickened, the medial axis length will become shorter.

The correlation between maximum radius and volume is depicted in Figure16. However, the figure shows a lack of precision for design samples that have a maximum radius between 180 and 230 mm. This can be inferred from $R^2 = 0.52$ which is due to a clear higher deviation from the trend line over the mentioned region. To fully understand the relation of circles that are used to create medial axis their circumferences are added up as described The sum of all the circles inscribed in the geometry and the volume demonstrate an excellent correlation with $R^2 = 0.88$ and thus it can be argued that it is a better feature and can increase the precision in regression analysis.

All extracted features can be obtained independently from other CAD parameters and this allows designers to have freedom when designing. The reason is when creating a parameterized CAD model, the designers need to always follow a unique convention and use the same features (such as curves, constraints, etc.) in the geometry. Obtaining these independent features can be done very fast and in an automatic way within the CAD environment. Moreover, correlation with simulation output will increase the accuracy and efficiency of the regression models in the machine learning process and will facilitate a live prediction model concerning decision-making in the design process.

As discussed in the introduction, the literature conveys that correlation can be a good criterion for selection features in machine learning. Therefore, Table 2 can be used to argue that the 3 extracted features in this table, with 80 + correlation, are superior in building a regression model than using direct CAD parameters. Another comparison between the two regression models is presented in Table 3. As it can be seen from the table, the accuracy for the model trained by sleeping parameters is $R^2 = 0.95$ and for the model trained by 14 CAD parameters is $R^2 = 0.63$. The maximum range for this error criteria is $R^2 = 1$, so it can be interpreted that the use of these parameters helps to get to an excellent regression precision over the estimated values. To make sure of the performed R^2 error comparison results, the MSE of the two sets is also performed and depicted in Table 3 which confirms the findings. For the model trained by 3 sleeping parameters, MSE is almost 7 times smaller than the one for the model that is trained by CAD parameters. Since smaller values for MSE shows better performance, we can once again confirm that the model trained by sleeping parameters is more accurate than the one trained by usual CAD parameters. Thus performed analysis proves the benefit of using sleeping parameters and its ability to perform accurate regressions with a small number of samples and with no need for complex machine learning algorithms.

The goal of the presented feature extraction framework is to simplify the problem so the regression can be done with any simple and easy-to-handle estimation model. To ensure that a more advanced algorithm can not perform better with the 14 CAD parameters, and also to ensure how good a job is MRL is doing on the sleeping parameters, the Support Vector Regression model is applied on both datasets. As it is shown in Table 3 the CAD parameters are giving $R^2 = 0.8$ when applied on 14 parameters, which is considered relatively underperformed. The model however reports the accuracy $R^2 = 0.95$ as seen from the same table when it is trained with sleeping parameters. The MSE error criterion is also showing an improvement from 14.34 to 1.77 which is considered substantial.

The fact that MLR and SVR have very close accuracy rates when trained with sleeping parameters, proves that these new features have successfully reduced the model order so it is now actually possible to regression the problem with MLR as good as SVR. This shows that sleeping parameters by increasing the quality of training features, have made it possible to get an acceptable result with MLR and there is no need for advanced and complicated regression algorithms such as SVR. Which was the aim from the beginning and is indeed an advantage and justification for pursuing feature extraction on CAD using the proposed framework.

The calculated simple regression will empower designers in the early stage of airbag design to have a real-time prediction model and therefore potentially will reduce the development lead time. This model can be added to a CAD environment so when designers change a length and/or a radius and/or an offset they can quickly see the impact of their decisions on the volume of the bag without any need to perform a complex finite element analysis. Moreover, independence from conventional parametrization in CAD will provide the flexibility for being creative with new solutions since they will not be forced to follow one standard parameterization in complex geometries which is very much needed in today's industry.

The proposed methodology is transferable to all volume simulations in airbag models that are using 2D geometries as inputs such as knee and side airbags that deploy from the passenger seat. Additionally, this methodology can be utilized by other simulations that use 2D shapes as inputs such as the design of wire patterns for seat heaters in the automotive industry. Other inflatable structures that require volume simulation can benefit from the finding of this paper, such as high-pressure vessels, inflatable tunnel plugs, inflatable rubber dams, different kinds of inflatable boats, etc. The methodology is also scalable to any performance evaluation that requires good enough accuracy but fast evaluation for decision-making in the early stages of the design process.

7. Conclusion

This paper applies correlation-based feature extraction on CAD model parameters. The aim is to have better features as input in machine learning which will in turn help to build more accurate regression-based models. Performed literature study, confirms the identified gap in the literature for lack of novel ways of utilizing CAD as input for data-based tools. It also confirms that potentially such tools could add to the efficiency of design processes by removing iterative loops (such as metamodels) in early development phases. First, the inefficiency of using CAD parameters alone, as input for estimation tools was investigated by showing how these parameters lack correlation with volume. Finite element simulation was used to study the effect of each parameter alone on the volume output through a parametric study. Using the concept of fundamental geometrical entities such as area, circumference, and medial axis, a group of parameters that are referred to as *sleeping parameters* are extracted. Rhino/Grasshopper was employed for creating the medial axis and measuring the parameters. Utilizing a correlation coefficient, it was shown that these parameters have a better correlation with volume as simulation. Multivariate Linear Regression as an example of a simple, and Support Vector Regression as an example of sophisticated machine learning algorithms are used on sleeping parameters and the usual CAD geometry parameters. The comparison between the two trained models proves that these extracted parameters are superior to be used in regression models. In future work, the generalization of this method on other case products will be carried out to show that it is possible to extract such sleeping parameters from other products' CAD models. A framework to extract and rank features from CAD models can be developed for designers in the next study.

CRediT authorship contribution statement

Mohammad Arjomandi Rad: Conceptualization, Methodology, Study design, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Kent Salomonsson:** Methodology, Study design, Investigation, Data curation, Supervision. **Mirza Cenanovic:** Software resources, Resources, Writing – review & editing, Supervision. **Henrik Balague:** Validation, Resources, Writing – review and editing, Supervision, Visualization. **Dag Raudberget:** Resources, Writing – review and editing, Supervision, Project administration, Funding acquisition, **Roland Stolt:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work has been carried out within the project Butterfly Effect in the school of engineering, Jönköping University. The authors would like to acknowledge the staff in Autoliv company in Sweden who were involved in the research projects and the Swedish Knowledge Foundation ('KK-Stiftelsen' with grant number 20180189) for the financial support as well as our former colleagues Dr. Joel Johansson and Dr. Tim Heikkinen who made this work possible. Finally, we also thank and acknowledge the editor and reviewers of the journal for valuable comments, which indeed has helped to improve the quality of this manuscript.

References

Arjomandi Rad, M., 2020. Reducing development lead time in iterative, simulation-driven design processes. Jönköping Univ.
Awad, M., Khanna, R., 2015. Support vector regression. Efficient Learning Machines. Springer, pp. 67–80.
Bertoni, A., Larsson, T., Larsson, J., & Elfsberg, J., 2017. Mining data to design value: A demonstrator in early design. Paper presented at the DS 87–7 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 7: Design Theory and Research Methodology, Vancouver, Canada, 21–25.08. 2017.
Blum, H., 1967. A Transformation for Extracting New Descriptors of Shape. 4 MIT Press, Cambridge.
Bouhlel, M.A., Bartoli, N., Otsmane, A., Morlier, J., 2016. Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction. Struct. Multidiscip. Optim. 53 (5), 935–952.
Caesarendra, W., Tjahjowidodo, T., 2017. A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. Machines 5 (4), 21.
Cai, J., Luo, J., Wang, S., Yang, S., 2018. Feature selection in machine learning: a new perspective. Neurocomputing 300, 70–79.
Chakrabarti, A., Blessing, L., 2016. A review of theories and models of design. J. Indian Inst. Sci. 95 (4), 325–340.
Chang, K.-H., 2016. e-Design: Computer-Aided Engineering Design. Academic Press.
Chavare, A.P., Khare, P., Bhise, A., Belanke, P., 2013. Performance Driven Package Feasibility of Side Restraints Using KBE Tools. SAE Technical Paper (0148–7191).
Chiarello, F., Belingheri, P., Fantoni, G., 2021. Data science for engineering design: State of the art and future directions. Comput. Ind. 129, 103447. https://doi.org/10.1016/j.compind.2021.103447
Dix, J., Hammoud, S., Cardinali, A., Mitchell, A., Fulk, D., 2012. Influence of complying with FMVSS 226 (ejection mitigation) on side airbag occupant out of position injury assessment. SAE Int. J. Passeng. Cars Mech. Syst. 5 (2012-01–0466), 358–368.
Du, X., Zhu, F., 2018. A new data-driven design methodology for mechanical systems with high dimensional design variables. Adv. Eng. Softw. 117, 18–28.
Emert, J., Nelson, R., 1997. Volume and surface area for polyhedra and polytopes. Math. Mag. 70 (5), 365–371.
Evans, N.C., & Leigh, M.J., 2013. FMVSS 226 ejection mitigation: A review.
Fabbri, R., Estrozi, L.F., Costa, L.D.F., 2002. On Voronoi diagrams and medial axes. J. Math. Imaging Vis. 17 (1), 27–40.
Fan, K.-K., Chiu, C.-H., Yang, C.-C., 2014. Green technology automotive shape design based on neural networks and support vector regression. Eng. Comput. 31 (8), 1732–1745. https://doi.org/10.1108/EC-11-2012-0294
Fatfouta, N., Le-Cardinal, J.S., 2021. An ontology-based knowledge management approach supporting simulation-aided design for car crash simulation in the development phase. Comput. Ind. 125, 103344.
Federal motor vehicle safety standards , 2011, Ejection mitigation; phase-in reporting requirements; Incorporation by Reference, 76C.F.R.
Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. Deep Learning. Deep learning 1 MIT Press, Cambridge.
Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.
Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L.A., 2008. Feature Extraction: Foundations and Applications. 207 Springer.
Hall, M.A., 1999. Correlation-based feature selection for machine learning.
Heikkinen, T., 2021. Transparency of Design Automation Systems Using Visual Programming–Within the Mechanical Manufacturing Industry.
Huang, G., Hu, J., He, Y., Liu, J., Ma, M., Shen, Z.,. Zhong, K., 2021. Machine learning for electronic design automation: A survey. arXiv preprint arXiv:2102.03357.
Ilies, H.T., 2006. Parametric solid modeling. Paper presented at the ASME 2006 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers.
Khalkhali, A., Noraie, H., Sarmadi, M., 2017. Sensitivity analysis and optimization of hot-stamping process of automotive components using analysis of variance and Taguchi technique. Proc. Inst. Mech. Eng. Part E J. Process Mech. Eng. 231 (4), 732–746.
Kondo, M., Bezemer, C.-P., Kamei, Y., Hassan, A.E., Mizuno, O., 2019. The impact of feature reduction techniques on defect prediction models. Empir. Softw. Eng. 24 (4), 1925–1963.
Kumar, G., & Bhatia, P.K., 2014. A detailed review of feature extraction in image processing systems. Paper presented at the 2014 Fourth international conference on advanced computing & communication technologies.
La Rocca, G., 2012. Knowledge based engineering: between AI and CAD. Review of a language based technology to support engineering design. Adv. Eng. Inform. 26 (2), 159–179.
Li, Y., Wang, H., & Liu, J., 2017. Can CNN Construct Highly Accurate Models Efficiently for High-Dimensional Problems in Complex Product Designs? arXiv preprint arXiv:1712.01639.
Pahl, G., Beitz, W., 2013. Engineering Design: A Systematic Approach (2nd, illustrated ed.). Springer Science & Business Media, London.
Park, W., 2017. A study of airbag design and optimization methodology. Paper presented at the Proc. 25th Int. Technical Conf. Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Dubourg, V., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

Rad, M.A., Khalkhali, A., 2018. Crashworthiness multi-objective optimization of the thin-walled tubes under probabilistic 3D oblique load. Mater. Des. 156, 538–557.

Rahman, M.H., Xie, C., & Sha, Z., 2019. A deep learning based approach to predict sequential design decisions. Paper presented at the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.

Ramamurthy, R., Farouki, R.T., 1999. Voronoi diagram and medial axis algorithm for planar domains with curved boundaries I. Theoretical foundations. J. Comput. Appl. Math. 102 (1), 119–141.

Ramnath, S., Haghighi, P., Kim, J.H., Detwiler, D., Berry, M., Shah, J.J.,. Menzel, S., 2019. Automatically generating 60,000 CAD variants for big data applications. Paper presented at the ASME 2019 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, IDETC-CIE 2019.

Ramnath, S., Haghighi, P., Ma, J., Kim, J.H., & Shah, J.J., 2020. Design Science Meets Data Science: Curating Large Design Datasets for Engineered Artifacts. Paper presented at the International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.

Secco, N.R., de Mattos, B.S., 2017. Artificial neural networks to predict aerodynamic coefficients of transport airplanes. Aircr. Eng. Aerosp. Technol.

Shan, S., Wang, G.G., 2010. Metamodeling for high dimensional simulation-based design problems. J. Mech. Des. 132 (5).

Simpson, T.W., Poplinski, J., Koch, P.N., Allen, J.K., 2001. Metamodels for computer-based engineering design: survey and recommendations. Eng. Comput. 17 (2), 129–150.

Song, J.-Y., Lee,K., & Ahn, B.-J., 2011. Simplifiedmodels for optimization of curtain air bag for US NCAP (0148-7191).Retrieved from.

Suh, N.P., 1998. Axiomatic design as a basis for universal design theory. Universal Design Theory, 3–24.

Sun, G., Wang, S., 2019. A review of the artificial neural network surrogate modeling in aerodynamic design. Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng. 233 (16), 5863–5872.

Suresh, K., 2003. Generalization of the mid-element based dimensional reduction. J. Comput. Inf. Sci. Eng. 3 (4), 308–314.

Wang, C.-C., Yang, C.-H., Wang, C.-S., Chang, T.-R., Yang, K.-J., 2016. Feature recognition and shape design in sneakers. Comput. Ind. Eng. 102, 408–422. https://doi.org/10.1016/j.cie.2016.05.003

Wang, G.G., Shan, S., 2007. Review of metamodeling techniques in support of engineering design optimization. J. Mech. Des. 129 (4), 370–380. https://doi.org/10.1115/1.2429697

Wang, J., Das, S., Rai, R., Zhou, C., 2018. Data-driven simulation for fast prediction of pull-up process in bottom-up stereo-lithography. Comput. Aided Des. 99, 29–42.

Wessing, S., 2018. diversipy (Version 0.8). Retrieved from https://pypi.org/project/diversipy/.

Yan, D.-M., Wang, W., Lévy, B., Liu, Y., 2013. Efficient computation of clipped Voronoi diagram for mesh generation. Comput. Aided Des. 45 (4), 843–852.

Yoo, S., Lee, S., Kim, S., Hwang, K.H., Park, J.H., & Kang, N., 2020. Integrating Deep Learning into CAD/CAE System: Case Study on Road Wheel Design Automation. arXiv preprint arXiv:2006.02138.

Yuce, B., Mastrocinque, E., Packianather, M.S., Pham, D., Lambiase, A., Fruggiero, F., 2014. Neural network design and feature selection using principal component analysis and Taguchi method for identifying wood veneer defects. Prod. Manuf. Res. 2 (1), 291–308. https://doi.org/10.1080/21693277.2014.892442

Yun, Y.-W., Choi, J.-S., Park, G.-J., 2014. Optimization of an automobile curtain airbag using the design of experiments. Proc. Inst. Mech. Eng. Part D J. Automob. Eng. 228 (4), 370–380.

Zhang, H., Ma, D., Raman, S.V., 2004. CAE-based side curtain airbag design. SAE Trans. 488–494.