Periodica Polytechnica Mechanical Engineering, 70(1), pp. 1–16, 2026

Data-driven Decision Support in Custom Manufacturing Planning

Richárd Szabó[1*], László Gönczy[1,2], Balázs Mikó[3], Dániel Nagy[4]

[1] Department of Artificial Intelligence and Systems Engineering, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary
[2] Quanopt Ltd., Bocskai út 77–79., H-1113 Budapest, Hungary
[3] Bánki Donát Faculty of Mechanical and Safety Engineering, Óbuda University, József körút 6., H-1081 Budapest, Hungary
[4] Component Ltd., Bácsalmás utca 1–3., H-1119 Budapest, Hungary
[*] Corresponding author, e-mail: szabor@mit.bme.hu

Received: 19 September 2024, Accepted: 06 March 2026, Published online: 17 March 2026

Abstract
In the design for manufacturing, the choice of the best processing activities and production scheduling plays an important role. The aim of the research is to create an automated process estimation system based on the processing plan of previously manufactured artifacts, which supports the scheduling software in the case of a custom manufacturing environment. We present a method and corresponding similarity metrics and evaluate the performance of our method on a set of real-life manufacturing plans and design data.

Keywords
manufacturing workflow planning, intelligent design, shape distance metrics, custom manufacturing

1 Introduction

1.1 Context

There are many possible ways to increase the efficiency and productivity of the production process. In addition to design for manufacturing and the choice of the best processing elements and parameters, production scheduling also plays an important role. The task of production scheduling is to distribute each individual production task among the available resources, i.e., to schedule the use of resources. In the case of component production, resources are machining tools (including the possibility of manual steps as well).

The basic information for scheduling is the manufacturing workflow plan. The manufacturing workflow is a sequence of manufacturing steps, including the necessary execution time. A manufacturing step is a unit of design that can be carried out on a given piece of equipment without interrupting the process.

1.2 Challenges

Effective production schedules require the existence of a manufacturing workflow plan, which is a challenge in the case of custom manufacturing. Planning a manufacturing workflow with the depth, detail, and accuracy of data to implement the schedule requires tedious work from the engineers. The task of manufacturing workflow planning is to compare machining needs and conceptual possibilities, to determine (select) the manufacturing methods, machine tools, and equipment that can be used, and to determine the sequence of manufacturing steps.

1.3 Method

For more than 30 years, researchers have been developing automated tools for part classification, indexing, and analysis, spanning a wide range of research areas (CAD/CAM, engineering design, knowledge representation, case-based inference, computer vision, pattern recognition, body modeling, shape modeling, and computer graphics). The results of shape analysis show that it may be a suitable basis for the creation of new technologies. The application of shape analysis allows us to transform geometry-centric engineering data into a knowledge base that can significantly improve the product realization process [1]. This knowledge base can be used to look for similarities between designs.

There are three reasons for looking for similarities. The first is cost estimation, which can (also) be done based on similarity. The second is the desire to reuse parts in the

Cite this article as: Szabó, R., Gönczy, L., Mikó, B., Nagy, D. "Data-driven Decision Support in Custom Manufacturing Planning", Periodica Polytechnica Mechanical Engineering, 70(1), pp. 1–16, 2026. https://doi.org/10.3311/PPme.38356

design process, and the third is the development of subpart families for more efficient process design [2].

1.4 Contributions

The aim of our research is to design an automated manufacturing workflow estimation process, based on the manufacturing workflows of historical designs. The manufacturing workflow plans can be used by the engineers to estimate the cost of machining new products and can be used to support the scheduling software in the case of a custom manufacturing environment. The main contributions of our paper are the following:

1. We propose an automated model processing and workflow estimation process. The process is based on historical data, thus it can be tailored to the given manufacturing company's workflows. The estimation process is based on a white-box machine learning algorithm, the so-called decision trees. This means that the decisions made by the estimation process can be easily interpreted.
2. We present topology-based similarity metrics, which are adapted from the neighbourhood graph shapes method [3].
3. We evaluated the presented topology-based similarity metrics and well-known histogram-based similarity metrics from the literature on 3300 industrial models. Based on the results of our evaluation, we propose a method to combine different similarity metrics to cover different aspects of the models and to get more accurate estimations.

The rest of the paper is structured as follows. Section 2 presents the overall concept that is used to process 3D models and estimate the workflows required to manufacture them. The inputs of the process and their evaluation are presented in Section 3. Similarity metrics and their usage to compare 3D models are presented in Section 4. Section 5 presents the training and the usage of the estimator models. We evaluate the presented approach in Section 6. Section 7 presents the literature related to our work. Concussions and future work are discussed in Section 8.

2 Overview of the approach

In Section 2, we introduce our innovative approach to processing historical information and train our models for accurate prediction. Later, the trained models are used to predict the necessary production steps for the new products.

2.1 Model processing and workflow estimation process

In Section 2.1, we present our model processing and workflow estimation process, depicted in Fig. 1. There are two main tasks that our process solves:

1. processing 3D models for the training of the predictive models;
2. processing the next order and predicting the manufacturing steps.

The inputs of the process:

1. Product data: the product data consists of the 3D models of the workpieces, the manufacturing workflows assigned to these models, and the product

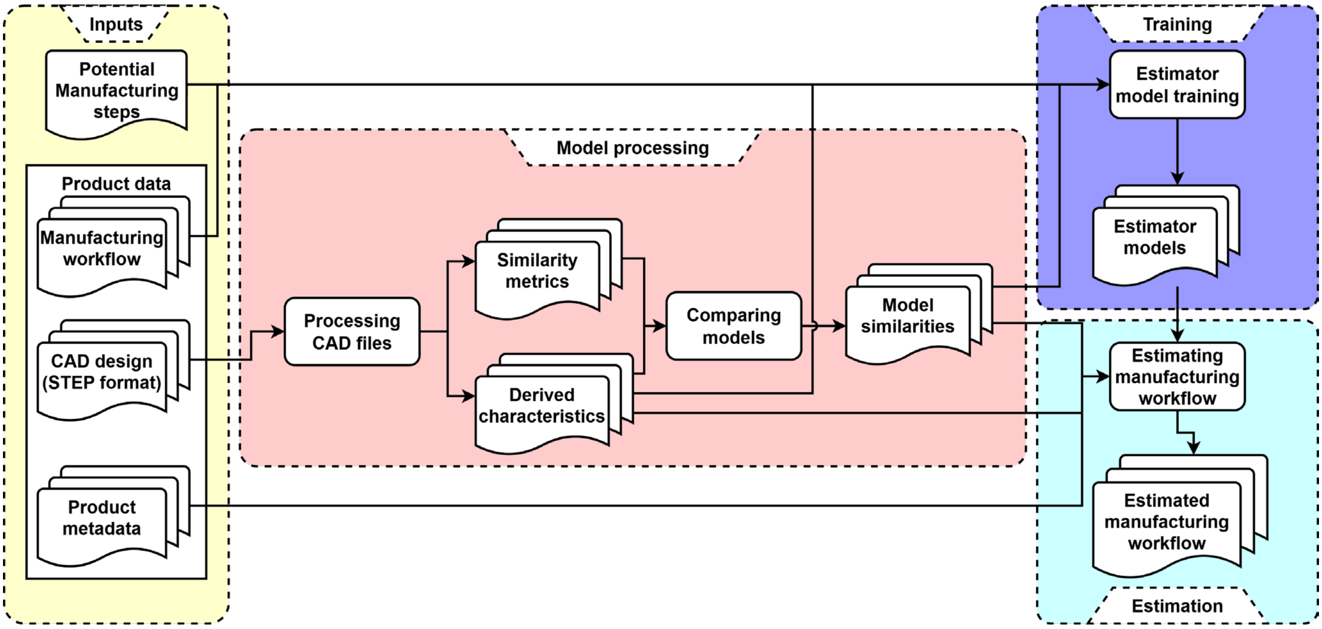

Fig. 1 Model processing and workflow estimation process

metadata. The product metadata contains valuable information regarding the ordered and manufactured workpieces, e.g., the material of the products and required surface treatment methods, which have to be considered during the planning of the manufacturing workflow:

- Training: in the case of building and training estimator models we use historical product data, provided by Component Ltd. These historical data consist of:
  - industrial CAD models;
  - the respective manufacturing workflows of the workpieces represented by the CAD models;
  - the product metadata of the workpiece.
- Estimation: in case of new orders, CAD models and product metadata (e.g., the material of the products and required surface treatment methods) are provided by the customers, these models must be processed and extended with derived metrics to serve as a basis for prediction of the manufacturing workflow of the order.

2. Potential manufacturing steps: the list of steps that can be performed to manufacture the workpieces, see Section 3.1 for details. Note that in our case, these steps were applied at the premises of Component Ltd., however, the method is agnostic to the concrete company.
3. Model processing phase:
   - Automated processing of CAD files: calculating the similarity metrics and the derived measures presented in Section 4. The similarity metrics will be used to find the best match across the historical models, while the derived measures will support the estimation of workflow steps.
   - Comparing models: comparison of the models resulting in their pairwise similarity metrics. In the case of processing historical data, the model similarity is used to build the estimator model. In the case of new orders, the model similarity is used to predict the steps required to manufacture the workpiece.
4. Training phase:
   - Estimator model training: training of the workflow estimator models using the calculated similarity and the manufacturing workflows of the historical CAD models, and the applicable manufacturing steps, using the decision support algorithms presented in Section 5.1.
5. Estimation phase:
   - Estimating manufacturing steps: predicting the steps required to manufacture a new order using the method presented later in Section 5.2.

3 Evaluation of historical data

Before we could train the estimator models, we had to process the historical data. During the project, we analyzed about 3300 historical manufacturing workflows and their respective 3D models, produced by Component Ltd. in recent years. In Section 3, we first analyze the potential manufacturing steps of Component Ltd., then we briefly evaluate the manufacturing workflows compiled from the potential manufacturing steps.

3.1 Potential manufacturing steps

The Component Ltd. keeps records of eighteen different manufacturing steps, including industry standard steps, e.g., turning, milling, sawing, etc., surface treatment methods, and the following other steps: manual work, outsourced work, and digital measuring. The complete list of potential manufacturing steps can be seen in Table 1.

We divided the eighteen manufacturing steps into two categories according to their role in the manufacturing plan estimation process:

- Optional steps: optional steps can be included in the manufacturing process. Our task is to estimate

Table 1 List of the potential manufacturing steps

Manufacturing step	Code	Estimation
CAD-CAM design	A	Not estimated
Cutting	B	Not estimated
Turning	C	Estimated
CNC machining	CNC	Not estimated
2D milling	D1	Estimated
5D milling	D5	Estimated
Hard milling	E	Estimated
Grinding	F	Not estimated
Drilling	FM	Not estimated
Wire EDM	G	Estimated
Heat treatment	H1	Not estimated
Eloxation / Anodization	H2	Not estimated
Manual work	X_hand	Estimated
Radial drill	I	Not estimated
Work done by a subcontractor	X_sub	Estimated
Other	X	Not estimated
3D measurement	3D	Not estimated
Validation	ZV	Not estimated

whether these steps are needed for a given workpiece, and if they are needed, we compute the time required to execute them. Such steps are turning, milling, sawing, etc. The time required for these steps depends on the dimensions and material of the workpiece.

- Obligatory steps: these steps are ordered by the customer and included in the product metadata, e.g., surface treatment methods, or need expert review, e.g., CAD-CAM design, digital measuring. These steps are omitted from the estimation process as they depend on external factors.

3.2 Historical data of manufacturing workflows

We analyzed about 3300 historical manufacturing workflows compiled from 10000 manufacturing steps. These manufacturing workflows were designed by experts. Our goal is to support the decision-making process of the experts who design the manufacturing workflows, thus we analyzed only the planned manufacturing workflows and not the manufacturing workflows executed in the factories.

The histogram of the lengths of the manufacturing workflows can be seen in Fig. 2. In most cases, the manufacturing workflows contain three different manufacturing steps, with 2D milling (D1) being the most frequently used manufacturing step.

4 Similarity between models

Finding similarities between models has a number of applications, one of which is estimating the manufacturing workflow. The estimated workflow then can be used to estimate the cost of the production or optimize the production line.

The main similarity assessment algorithms for component and technology design can be classified into the following seven categories [2–6]:

1. non-geometry property-based similarity;
2. geometry property-based similarity;
3. histogram-based similarity;
4. image comparison of details;
5. similarity based on a function of space;
6. graph-based similarity;
7. similarity based on features.

The density function-based metrics measure the distances between various points of the surface of the model, which can be used to describe the dimensions of the model. In the case of two similar models, small differences, e.g., a different number of holes or different positions of holes can lead to significant differences in the density functions of models, while their production workflow could be similar. In the case of topology-based similarity metrics, the dimensions of the models are ignored, this means that models with the same number of holes, but in different positions are very similar to each other, however, models with significant size differences are also very similar to each other, even though their production workflow can differ, e.g., because of the limitations of the machines.

In order to give an accurate estimation for the manufacturing workflow of a new order, we have to cover different aspects of the 3D models to find the most similar historical model to the new order. Thus, we apply both density function-based and topology-based metrics.

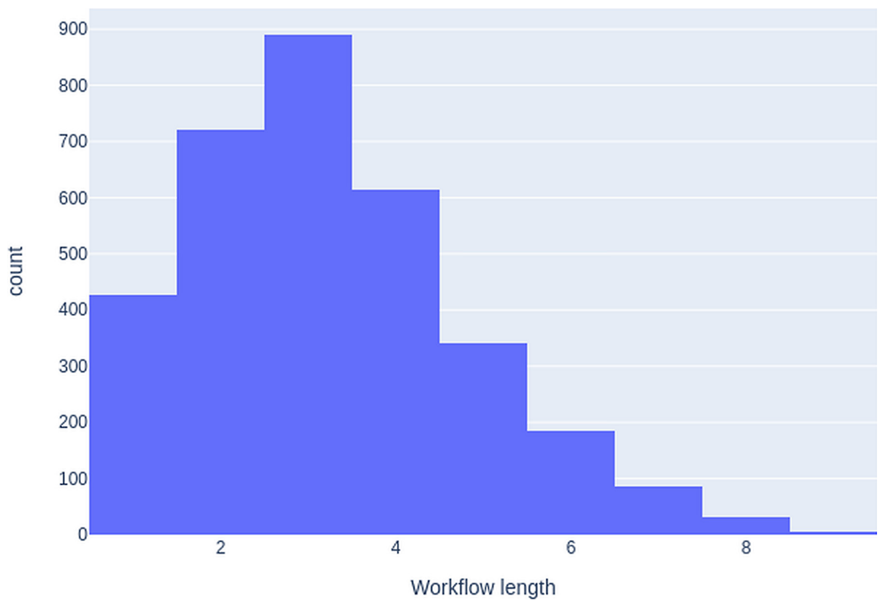

Fig. 2 Histogram of the length of the analyzed workflows

In Section 4, we first present the selected density function-based similarity metrics and their usage in the model comparison. Next, as a new contribution, we present topology-based similarity metrics and their usage in the model comparison. Finally, we describe the derived measures, which are later used in the estimator training and the workflow estimation steps.

4.1 Running example

Consider the base example model shown in Fig. 3 (a). We see a cube with four holes going through from top to bottom. This is modelled with six plane surfaces (S_1, S_2, S_3, S_4, S_5 and S_6 respectively) and four cylinder surfaces (S_7, S_8, S_9 and S_{10} respectively) connecting S_1 and S_6.

Parallel hole example model (Fig. 3. (b)) is mostly the same as the base example model (Fig. 3 (a)) with a fifth hole going through from top to bottom. The manufacturing steps of parallel hole example model (Fig. 3. (b)) have the same complexity as the manufacturing steps of base example model (Fig. 3 (a)).

Orthogonal hole example model (Fig. 3 (c)) is mostly the same as base example model (Fig. 3 (a)) with a fifth hole going through from side to side. The manufacturing steps of orthogonal hole example model (Fig. 3 (c)) are more complex than the manufacturing steps of base example model (Fig. 3 (a)). The manufacturing of orthogonal hole example model (Fig. 3 (c)) requires an additional drilling step, because the fifth hole is on a different axis than the original four (considering a traditional drill).

4.2 Application of selected density function-based metrics

According to Osada et al. [7], we have chosen density function-based metrics to compare the similarity of two objects given in a 3D model. Osada et al. [7] investigated five different distance interpretations for density function-based similarity, these are shown in Fig. 4:

- A3: measures the angle between three random points on the surface of a 3D model.
- D1: measures the distance between a fixed point and one random point on the surface. We use the centroid of the boundary of the model as the fixed point.
- D2: measures the distance between two random points on the surface.
- D3: measures the square root of the area of the triangle between three random points on the surface.
- D4: measures the cube root of the volume of the tetrahedron between four random points on the surface.

To calculate the metric value for a given model, we sample the surface of the model and calculate the distances between the sampled points based on the metric. This sampling corresponds to the sampling of the Probability Distribution Function (PDF) of the metric, interpreted by its histogram.

Fig. 5 shows the histograms of the D2 metric for the example models. In this case, it shows the observed PDF of D2 distance values calculated in this case on 20.000 point pairs.

The Kolmogorov–Smirnov test [8] (KS test) is a non-parametric statistical test to decide whether two samples

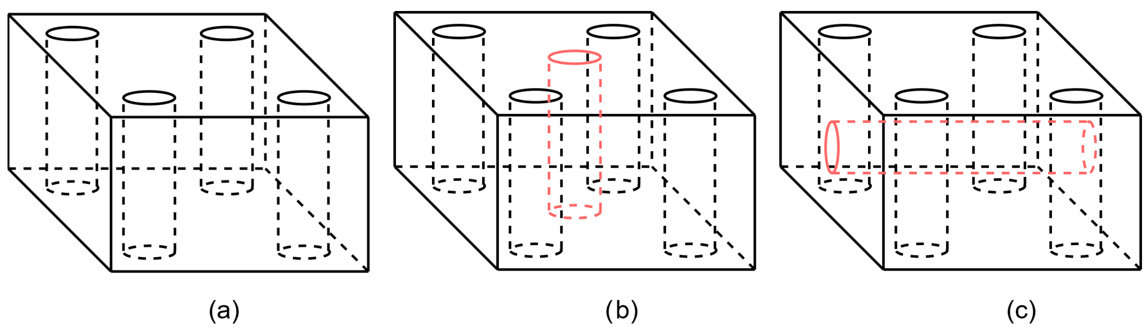

Fig. 3 Example models: (a) base; (b) parallel hole; (c) orthogonal hole

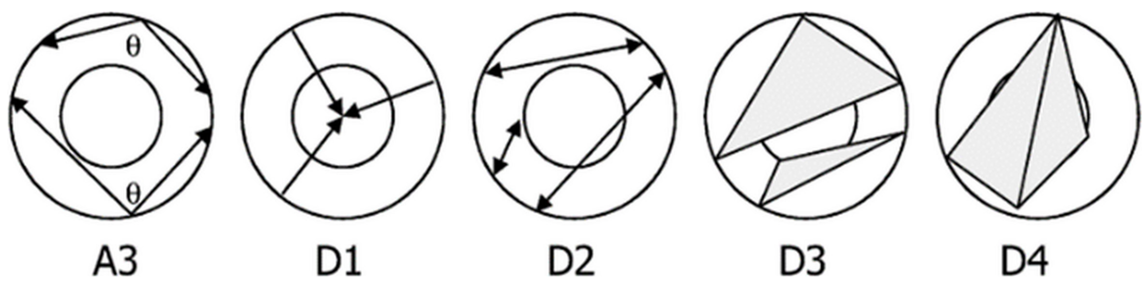

Fig. 4 Density function-based metrics [7]

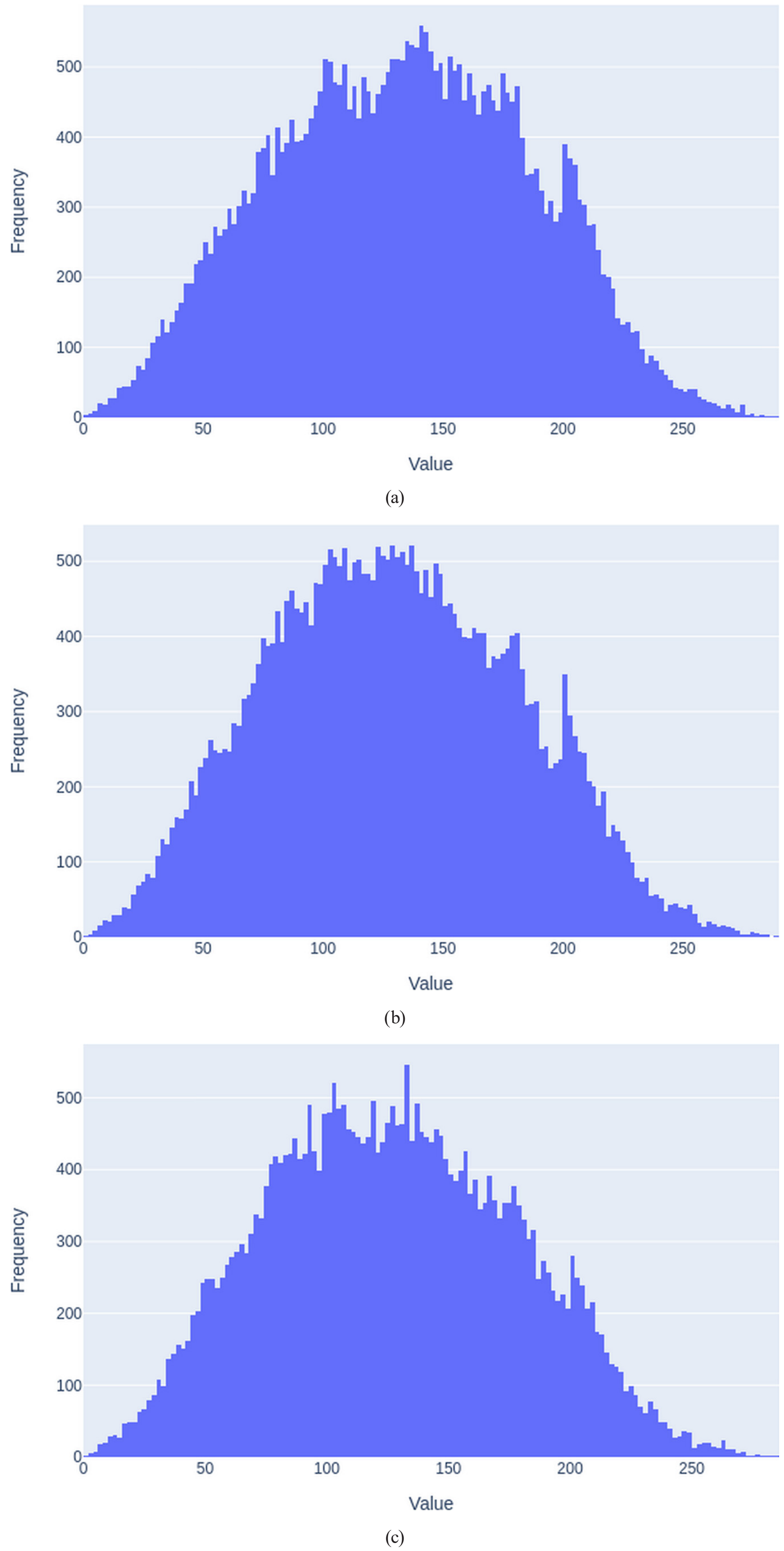

Fig. 5 Histograms of the D2 metric of the example models: (a) base; (b) parallel hole; (c) orthogonal hole

came from the same distribution. To determine the similarity of two 3D models, we perform a two-sample on the samples of the two models. The P-value of the KS test tells us how likely is that the samples came from the same distribution, which we can interpret as how likely is that samples came from the same model.

4.3 Definition and application of topology-based metrics

We adapted the neighborhood graph shapes method [3] to define similarity metrics based on the adjacent faces of the surface of the 3D models. The similarity metrics are derived from 11 types of surfaces:

$$\mathcal{T} = \begin{Bmatrix} \text{plane, cylinder, cone, sphere, torus,} \\ \text{Bézier surface, B-spline surface,} \\ \text{surface of extrusion, offset surface,} \\ \text{custom type of surface} \end{Bmatrix}.$$

For a given surface S_i in model M_i, we collect the types of those surfaces that have at least one common edge with S_i. These collections are called *adjacent surface sets*, and an *adjacent surface set* is a function mapping the types of the adjacent surfaces into the number of occurrences, represented as a multiset [9]. A multiset m over a set X is a function:

$$m : X \to \mathbb{N},$$

where $m(X)$ is the number of occurrences for all elements $x \in X$ in m.

Formally an adjacent surface set $(\mathcal{A})$ is a multiset of surface types $\mathcal{T}$. For example, an adjacent surface set:

$$a \in \mathcal{A},$$

where a plane has four adjacent plane surfaces, and four adjacent cylinder surfaces are the following:

$$a = \text{plane} \mapsto 5,\ \text{cylinder} \mapsto 4.$$

The topology of model M_i is a function mapping all the *adjacent surface sets* observed in M_i into the number of occurrences, represented as a multiset. Formally, the topology (T) is a multiset of *adjacent surface sets* $(\mathcal{A})$. For example, a topology $t \in T$ is the following:

$$t = (\text{plane} \mapsto 5,\ \text{cylinder} \mapsto 4) \mapsto 2,$$
$$(\text{plane} \mapsto 5) \mapsto 4,\ (\text{cylinder} \mapsto 1,\ \text{plane} \mapsto 2) \mapsto 4.$$

To calculate the similarity of the models, we transform the multisets describing the models into vectors and calculate the *cosine distance* of these vectors. To transform the *topology* multisets into vectors, we have to define an ordering of the elements of the multisets. To achieve this, we order the universe of the *topology* multisets in the order of the first observation of its elements. The $\boldsymbol{v}$ vector transformed from a multiset M_i is defined as follows in Eq. (1):

$$\boldsymbol{v}_{M_i}[j] = \begin{cases} m_{M_i}(a_j), & \text{if } a_j \in M_i \\ 0, & \text{if } a_j \notin M_i \end{cases}, \forall a_j \in \Omega. \quad (1)$$

4.3.1 Topology-based similarity metrics

We defined five topology-based similarity metrics. They only differ in how the adjacent surface sets are defined, and how the number of occurrences of the adjacent surface sets are registered in the models. Each metric is an equivalence relation on the set of all possible 3D models. Metric1 is finer than Metric2, Metric3, and Metric4. Metric2 is finer than Metric4. Metric3 is finer than Metric4.

Metric1

The first similarity metric registers the types of all the adjacent surfaces to define the adjacent surface sets and counts all the occurrences of a given adjacent surface set to determine the multiset describing the model.

For base example model (Fig. 3 (a)), S_1 is a plane, it has four adjacent plane surfaces (S_2, S_3, S_4, S_5) and four adjacent cylinder surfaces (S_7, S_8, S_9, S_{10}). S_2 is a plane and has four adjacent plane surfaces (S_1, S_3, S_4, S_6). S_7 is a cylinder and has two adjacent plane surfaces (S_1, S_6). S_6 has the same adjacent surface types as S_1. S_3, S_4 and S_5 have the same adjacent surface types as S_2. S_8, S_9 and S_{10} have the same surface types as S_7. Thus, the registered adjacent surface sets are the following (Eq. (2)):

$$\begin{aligned} a_{S_1} &= a_{S_6} = \text{plane} \mapsto 5,\ \text{cylinder} \mapsto 4 \\ a_{S_2} &= a_{S_3} = a_{S_4} = a_{S_5} = \text{plane} \mapsto 5 \\ a_{S_7} &= a_{S_8} = a_{S_9} = a_{S_{10}} = \text{cylinder} \mapsto 1,\ \text{plane} \mapsto 2 \end{aligned} . \quad (2)$$

The multiset describing the topology of the example model is:

$$t_a = a_{S_1} \mapsto 2,\ a_{S_2} \mapsto 4,\ a_{S_7} \mapsto 4.$$

Note that the adjacent surface sets are in the order of their first observation. This property is used during the calculation of the similarity of different models.

Metric2

The second similarity metric registers the types of all the adjacent surfaces to define the adjacent surface sets. Metric2 registers whether a given adjacent surface set is observed in the given model. Thus, the registered adjacent surface sets are the same as in the case of Metric1, but the multiset representing the base example model (Fig. 3 (a)) is:

$$t_a = a_{S_1} \mapsto 1,\ a_{S_2} \mapsto 1,\ a_{S_7} \mapsto 1.$$

In the case of parallel hole example model (Fig. 3 (b)), a model with a fifth hole going through S_1 and S_6 would be considered the same as the base example model (Fig. 3 (a)), however, it slightly differs in the case of Metric1. Furthermore, in the case of orthogonal hole example model (Fig. 3 (c)), where the fifth hole goes through S_2 and S_4, they cannot be distinguished by using Metric2, even though there might be differences in the sequence of the operations.

Metric3
The third similarity metric registers only the set of the adjacent surfaces to define the adjacent surface sets and counts all the occurrences of a given adjacent surface set. For base example model (Fig. 3 (a)) the registered adjacent surface sets are the following (Eq. (3)):

$$\begin{aligned} &a_{S_1} = a_{S_6} = \text{plane} \mapsto 2,\ \text{cylinder} \mapsto 1 \\ &a_{S_2} = a_{S_3} = a_{S_4} = a_{S_5} = \text{plane} \mapsto 5 \\ &a_{S_7} = a_{S_8} = a_{S_9} = a_{S_{10}} = \text{cylinder} \mapsto 1,\ \text{plane} \mapsto 1 \end{aligned} \quad . \qquad (3)$$

The multiset representing the topology of base example model (Fig. 3 (a)) is:

$$t_a = a_{S_1} \mapsto 2,\ a_{S_2} \mapsto 4,\ a_{S_7} \mapsto 4.$$

Metric4
The fourth similarity metric registers only the set of the adjacent surfaces to define the adjacent surface sets and only registers whether a given adjacent surface set can be observed in the given model. Thus, the registered adjacent surface sets are the same as in the case of Metric3, but the multiset representing the base example model (Fig. 3 (a)) is:

$$t_a = a_{S_1} \mapsto 1,\ a_{S_2} \mapsto 1,\ a_{S_7} \mapsto 1.$$

Metric5
Collecting all the possible adjacent surface sets for the universe of the multisets can be computationally heavy in the case of large, diverse model sets. Runtime registration of an observed adjacent surface set can require iterating through a large multiset, where the inclusion can only be decided after iterating through the whole multiset since the adjacent surface sets are registered in the order of observation. We defined the fifth metric to solve the computational problem with a predefined, fixed-length universe of the *adjacent surface sets*. Before the processing of the models we define the universe of the *adjacent surface sets* as the possible pairs of the surface types:

$$\Omega = \left\{ \begin{matrix} (\text{plane} \mapsto 2),\ (\text{plane} \mapsto 1,\ \text{cylinder} \mapsto 1), \\ (\text{plane} \mapsto 1,\ \text{cone} \mapsto 1), \ldots, \\ (\text{extrusion} \mapsto 1,\ \text{other} \mapsto 1),\ (\text{offset} \mapsto 2), \\ (\text{offset} \mapsto 1, \text{other} \mapsto 1), (\text{other} \mapsto 2) \end{matrix} \right\}.$$

In base example model (Fig. 3 (a)) there are six plane surfaces, and each plane surface has four adjacent plane surfaces, thus the number of plane surface pairs is twenty-four. In base example model (Fig. 3 (a)) there are four cylinder surfaces, and each cylinder surface has two adjacent plane surfaces, thus the number of adjacent plane-cylinder pairs is eight. The multiset representing base example model (Fig. 3 (a)) is described as follows:

$$t_a = (\text{plane} \mapsto 2) \mapsto 24,\ (\text{plane} \mapsto 1,\ \text{cylinder} \mapsto 1) \mapsto 8,$$
$$(\text{plane} \mapsto 1,\ \text{cone} \mapsto 1) \mapsto 0,\ \ldots,\ (\text{other} \mapsto 2) \mapsto 0.$$

4.4 Augmentation with derived measures

We describe additional features of the models which are later used by the workflow estimator models.

We define the bounding box of the 3D models to approximate the volume of the initial workpieces from which the final workpieces are machined. The ratio between the volume of the final workpiece and the volume of the bounding box gives a sense of how much material must be removed from the initial workpiece to manufacture the final workpiece. The ratio between the surface area of the final workpiece and the surface area of the bounding box gives a sense of the complexity of machining the final workpiece because the larger change in the surface area leads to more complex manufacturing steps; thus we can argue about the complexity of the steps that are required to manufacture the final workpiece.

The necessity of some manufacturing steps is decidable directly from the 3D model. If the initial workpiece is cylindrical, we can be almost certain that turning is needed to manufacture the workpiece. We also search for holes in the model, which tells us if drilling is needed to manufacture the workpiece. Finally, if the model contains surfaces defined by splines there must be expert supervision in the manufacturing workflow.

5 Estimation process and analysis

In Section 5, we present the machine learning algorithms used for decision support. After that, we present the application of the decision support algorithms to estimate the manufacturing workflows of the workpieces.

5.1 Machine learning algorithms for decision support

5.1.1 Decision Trees

The Decision Tree learning algorithm [10] is a supervised learning approach. The leaves of the Decision Tree are the possible outcomes of the target variable and the inner nodes are the decisions based on the input features. Decision trees with discrete target variables are called classification trees (Decision Tree Classifiers), whereas decision trees where the target variable can take continuous values are called regression trees (Decision Tree Regressors).

Decision trees can handle both numerical and categorical data. Unlike most machine learning algorithms, the decision trees are white-box models, thus the importance of each input feature is known. The decisions made by the decision tree can be supported by the structure of the decision tree, which represents the training data.

During our work to estimate the time of the steps of the manufacturing workflow, we applied Decision Tree Regressors.

5.1.2 Random forests

The random forest is an ensemble learning method [11], where several smaller decision trees are used together. During the training phase, both the training data and the training variables are split randomly, so there is a low probability that any of the resulting trees were trained under the same condition. This made the Random Forests method more robust than a simple Decision Tree and made it more immune to overfitting, which is a general problem of machine learning algorithms.

During our work to predict the steps of the manufacturing workflow, we applied Random Forests due to their robustness to faults and overfitting; in addition to the advantages of the Decision Trees.

5.2 Workflow estimation process

Estimating complete manufacturing workflows is a challenging task. Instead of estimating the complete workflow, we are focusing on estimating individual steps one by one. The workflow estimation process contains three steps and can be seen in Fig. 6:

- Predict first step: prediction of the first manufacturing step;
- Predict next step: prediction of the next manufacturing step using Random Forests;
- Calculate time for steps: calculating the time of the whole manufacturing process, from the time required by the individual steps, using Decision Tree Regressors, including the additional steps added by the customer.

5.2.1 Prediction of the next manufacturing step

To estimate the next manufacturing step, we trained a Random Forest for each manufacturing step that must be estimated (see Table 1) and an additional one that represents the end of the workflow. The inputs of each Random Forest are the same, namely: the current step, the

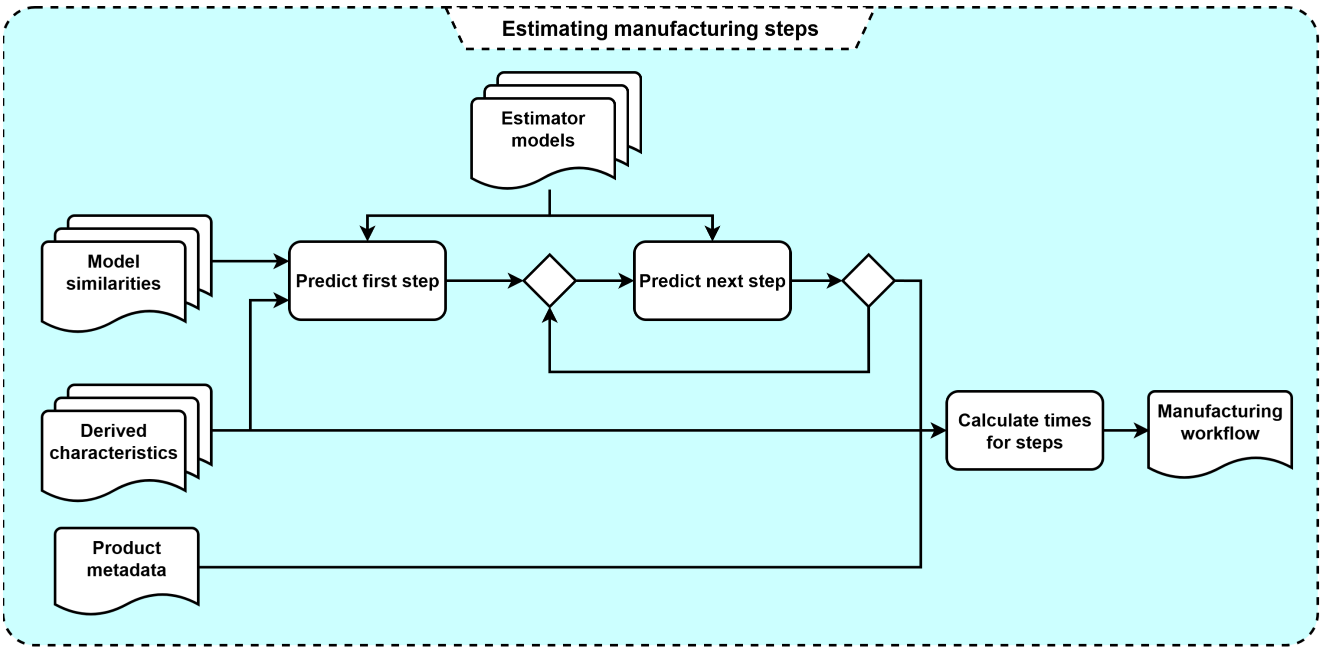

Fig. 6 Estimation process of the manufacturing workflow

position of the step to be estimated, the derived measures presented in Section 4.3, the material of the workpiece, the manufacturing steps that are not estimated, and the type of the previous steps. The output of each Random Forest is whether the given manufacturing step should be included as the next step and its confidence. This prediction process of the next manufacturing steps continues until the output of the Random Forest representing the end of the workflow has the highest confidence.

Fig. 7 shows the importance of the input variables as observed during the training of the Random Forest method which selects the next step of the production workflow. The importance of the variables shows that the most important variable of the Random Forests is the current manufacturing step. This means that each step is highly dependent on the previous step, thus the estimation of the first step is critical since an incorrectly chosen first step can lead to incorrect manufacturing workflows.

We evaluated several methods to predict the first step of the workflow, including additional Random Forests trained specifically to predict the first step, and similarity-based methods based on the metrics presented in Section 4.

5.2.2 Estimating the time of the manufacturing workflow

After predicting the manufacturing workflow, we extend the workflow with the manufacturing steps ordered by the customer, e.g., surface treatment methods. Finally, our algorithms calculate the time required to complete each step on the workpiece.

To estimate the time of the manufacturing workflow, we trained Decision Tree Regressors for cutting (B), turning (C), 2D milling (D1), 5D milling (D5), hard milling (E), grinding (F), and wire EDM (G) manufacturing steps. We assigned fixed manufacturing times for the rest of the possible manufacturing steps based on the experts' decisions. The inputs of each Decision Tree Regressor are the same, namely: the derived measures presented in Section 4.3 and the material of the workpiece.

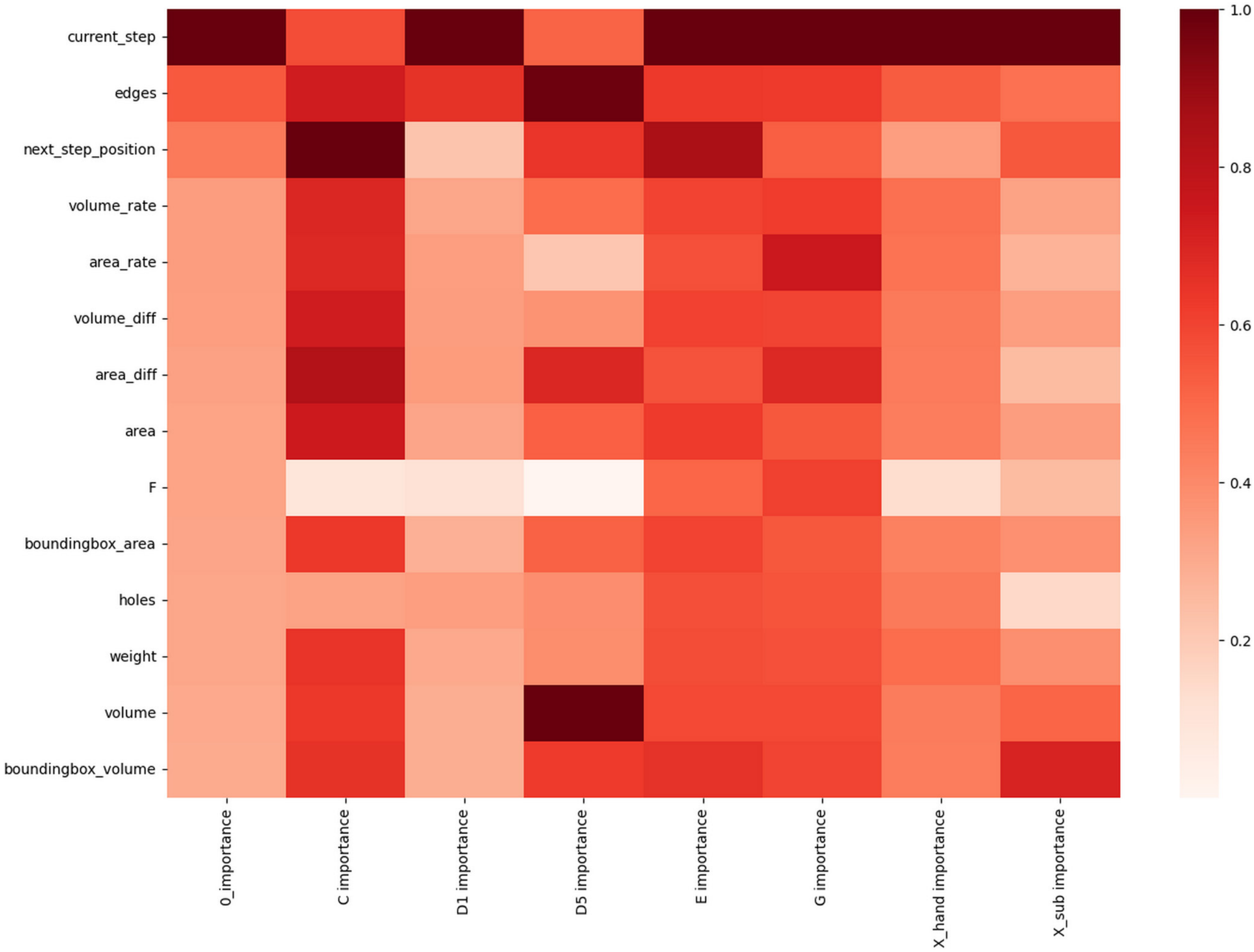

Fig. 7 Importance of the input features of the Random Forests

6 Results

We evaluate the presented decision support method on 1000 randomly selected historical models and their manufacturing workflows. We measured:

1. the number of cases when the estimated first step matches the real first step;
2. the number of cases when the estimated workflow (excluding the not predictable steps and the manufacturing times) matches the real workflows;
3. how much the predicted workflow matches the real workflow calculated with the normalized Hamming distance, which is interpreted as a percentage.

The normalized Hamming distance is defined as follows:

- σ is a manufacturing step from the set of possible manufacturing steps;
- σ^* is the sequence of manufacturing steps;
- $w \in \sigma^*$ is a manufacturing workflow;
- $w^i \in \sigma$ is the i^{th} manufacturing step of workflow w;
- $|w|$ is the length of workflow w;
- match: $\sigma \times \sigma \rightarrow \{0, 1\}$ is a function, where:

$$\operatorname{match}(\sigma_1, \sigma_2) = \begin{cases} 1, & \text{if } \sigma_1 = \sigma_2 \\ 0, & \text{otherwise} \end{cases}.$$

The following equation:

$$H = \frac{\sum_{i=1}^{\min(|p|,|r|)} \operatorname{match}(p^i, r^i)}{\max(|p|,|r|)},$$

is the normalized Hamming distance, where $p, r \in \sigma^*$ are the predicted and the real manufacturing workflows, respectively.

The most critical step in the workflow estimation process is the selection of the first step of the workflow. We compared 13 different methods to estimate the first step of the workflows, summarized in Table 2.

First, we initiated the workflow estimation process with the first step of the manufacturing workflow defined by the experts. This resulted in 87% fully matching workflows and 89% average normalized Hamming distance. Because the rest of the workflow estimation process is the same for each compared method, this is the limitation of our current decision support method.

Next, we used random forests trained for the estimation of the first manufacturing step. This method resulted in 92% matching first steps, 79% fully matching workflows and 87% average normalized Hamming distance.

Table 2 Results of the evaluation

First step selection method		First step matching	Total workflow matching	Average workflow matching
Real first step		100%	87%	89%
Estimation based	Random forests	92%	79%	87%
Density function-based similarity metrics	A3	69%	60%	73%
	D1	77%	67%	78%
	D2	70%	61%	74%
	D3	69%	60%	73%
	D4	68%	59%	73%
Topology-based similarity metrics	Metric1	85%	74%	82%
	Metric2	78%	70%	80%
	Metric3	68%	59%	73%
	Metric4	85%	75%	83%
	Metric5	85%	76%	84%
Combination of selection methods		93%	81%	88%

Based on the similarity metrics presented in Section 4 we selected the most similar model to each test model and chose their first manufacturing steps to the workflow estimation process. In the case of the density function-based similarity metrics, the best-performing metric was the D1 metric in which case the first steps matched in 77% of the test models, there were 67% fully matching workflows and 78% average workflow matching. In the case of the topology-based similarity metrics both Metric1, Metric4, and Metric5 performed similarly, resulting in 85% matching first steps, ~75% fully matching workflows, and ~83% average normalized Hamming distance.

After analyzing these results, we found that when the Random Forests method predicts turning (C), 2D milling (D1), or wire EDM (G) as the first step, it is correct in 90% of the cases. We also found that when the first step is hard milling (E), manual work (X_hand), or work done by a subcontractor (X_sub), Topology Metric1 is the most precise. The combination of the Random Forests and the Topology Metric1 methods resulted in 93% matching first steps, 81% fully matching workflows, and 88% average normalized Hamming distance, which is more accurate than using them separately (see the results in the line of Table 2 called *"Combination of selection methods"*).

The results of the evaluation are summarized in Table 2 and the true positive rate of the predicted steps by each method is shown in Figs. 8 to 13 and the number of manufacturing steps selected as the first steps by each method is shown in Fig. 14.

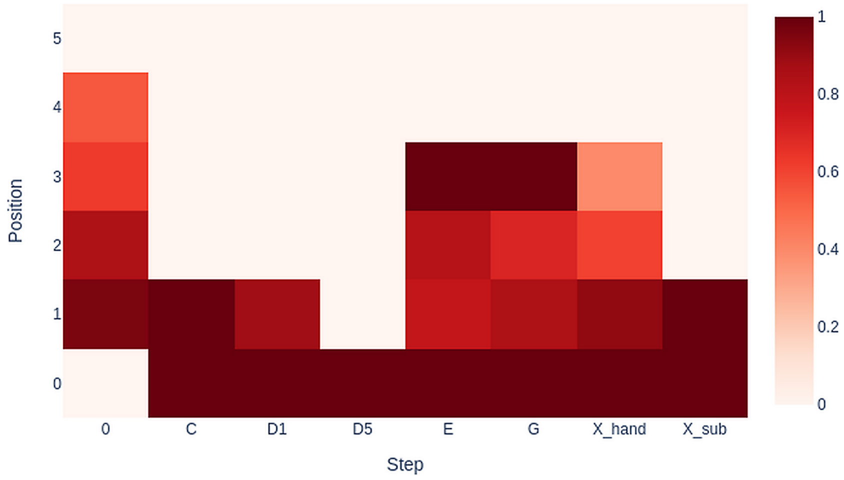

Fig. 8 True positive rate of estimating manufacturing steps using the real first step

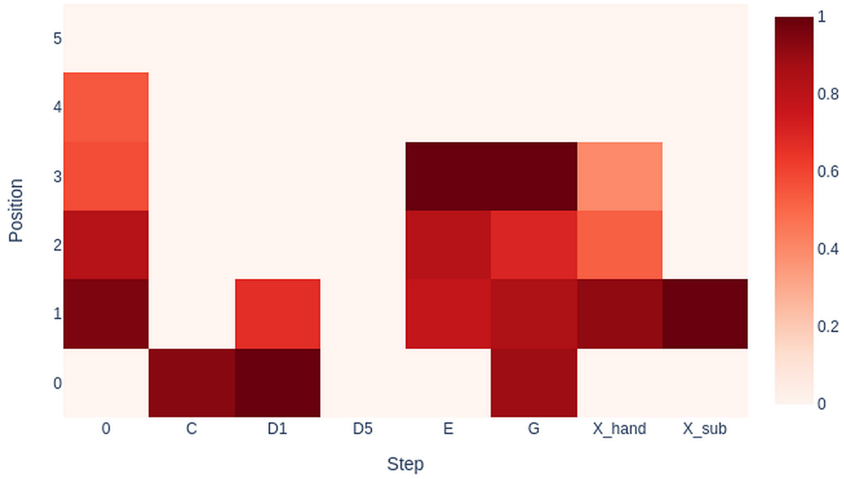

Fig. 9 True positive rate of estimating the manufacturing steps using Random Forests based first step

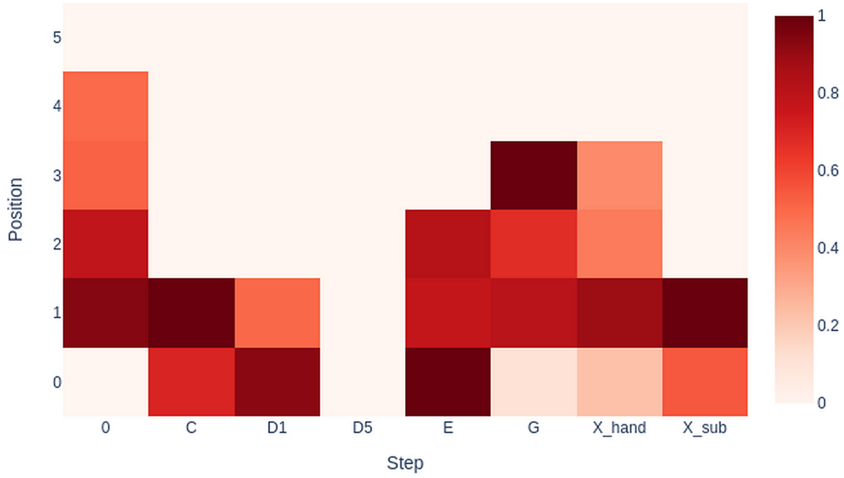

Fig. 10 True positive rate of estimating the manufacturing steps using Topology Metric1 based first step

With the presented results, we showed that our model processing and workflow estimation process can be used to support the decision-making process of the experts working in custom manufacturing design. We showed that in 81% of the cases, we can fully estimate the required manufacturing workflow. We must emphasize that the presented process is a decision-support process, which always requires expert review. The presented approach is generic and can be used in other companies as well, but it must always be tailored to the individual company because of the differences in the possible manufacturing steps and the manufacturing workflows in general.

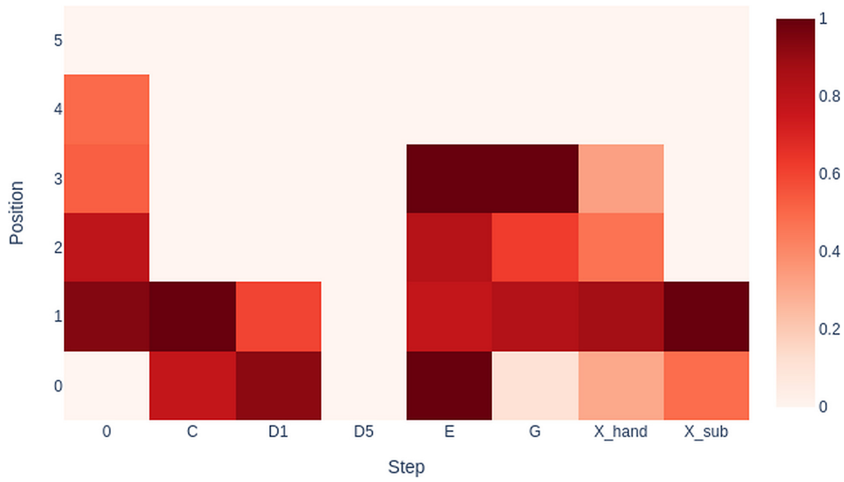

Fig. 11 True positive rate of estimating the manufacturing steps using Topology Metric5 based first step

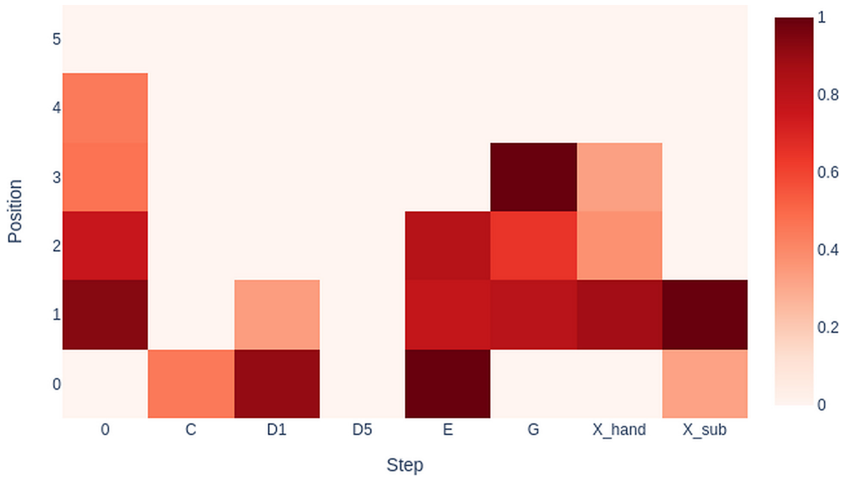

Fig. 12 True positive rate of estimating the manufacturing steps using Density metric D1 based first step

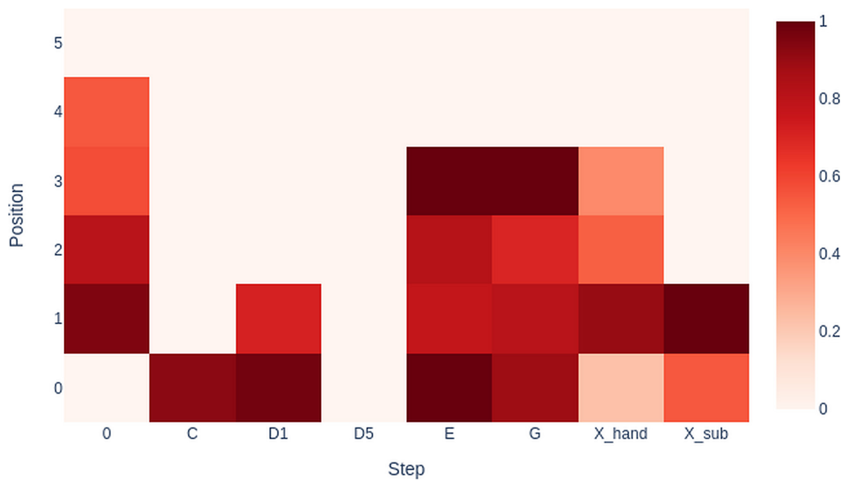

Fig. 13 True positive rate of estimating the manufacturing steps using the Combination-based first step

7 Related work

7.1 Similarity between models

Similarity finding is a key issue of the analogy-based design principle. There are several ways to determine the similarity of two models introduced in the literature, each having its advantages. We showed how to combine different similarity metrics to predict manufacturing workflows more accurately.

Cardone et al. [12] investigate the similarity of rectangular parts produced by milling based on the similarity of the individual geometric elements. They decompose the geometry and compare the individual elements. This

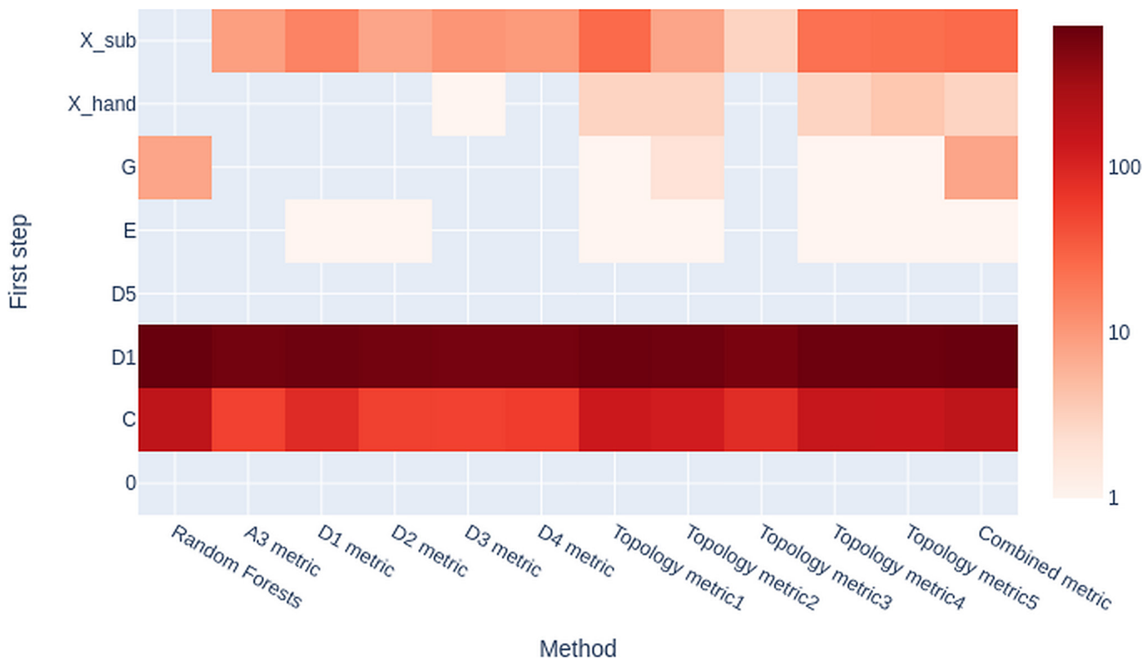

Fig. 14 Number of manufacturing steps selected as first step by each metric

analytical solution requires that the parts to be compared are built from predefined geometric elements, but this approach is less sensitive to the location of the elements.

Zhang and Zhou [6] use an integrated product information model to determine the similarity of parts. First, they compare non-geometric data with reusable design documents or available resources. Then, geometric data and shape similarities are compared using a two-step evaluation algorithm in order to rank the similarity of reusable models. In the first step, the shapes of the models are classified by absolute angle and distance (AAD), and a vector support machine (SVM) excludes models that do not belong to the class of the model being searched. In the second step, a genetic algorithm is used to estimate the similarity of shape features based on shape, process, direction, and positional distribution. Geometric similarity is determined using the AAD algorithm. This involves randomly selecting points on the surface of the model using the Monte Carlo method and then determining the distance of each possible pair of points, which are represented by a histogram.

The interpretations of the information contained in the density functions describing 3D models are the key question of density-function-based similarities. Osada et al. [7] investigated five different distance interpretations for density function-based similarity, represented by histograms.

Çiçek [13] searches for similar components by topological analysis of STEP models. The topology is defined using a matrix describing the neighborhood of each geometric basis element. This is used to determine dimensional and non-dimensional similarity. Ma and Tian [14] investigate the similarity of CAD models both topologically by analyzing the constructive solid geometry (CSG) tree describing the relationship of each geometric primitive and by comparing the geometry using histograms generated from the D2 distance definition. Chu and Hsu [15] search for similar parts by comparing the topology of the CAD model. In their study, they generated a shape-feature adjacency graph and a topology graph based on the CAD model. The combined similarity of the graphs on the CAD models they studied proved to be more effective than comparing D2 distance histograms.

Harik and Barakat [16] investigate the similarity of the geometric elements that make up parts and their relationships. They first determine the relationship of each geometric element based on a STEP format model of the part and generate a relationship graph. Then, analysis of the relationship graph creates a machining graph to search for similar parts.

Kusiak and Cho [17] investigate the design of component groups using process data. They create a machine-part occurrence matrix and group parts that occur on the same machines into a group. The Medial Axis Transformation (MAT) method for the structural representation of models is presented by Renner and Stroud [18]. They replace pairs of related elements with vertices and connect them. The vertex can be a bisector point or a corner point. In the case of MAT, the bisector point of the segment connecting the two elements. In the global structure of a positive object, only the faces are represented by nodes, whereas in the global structure of a negative object (negative MAT), all edges and vertices are also represented as nodes.

7.2 Scheduling in manufacturing systems

Custom manufacturing design companies are constantly challenged to meet ever-changing needs and requirements. Industry 4.0 manufacturing systems enable engineers to estimate the cost of ordered products quickly and efficiently and optimize their production lines to serve multiple orders.

Simulation-based decision support systems are widely used to help engineers make decisions about whether to take new orders, optimize their production lines, or run "what if" analysis to find bottlenecks, or calculate production time with equipment downtime [19].

Simulation-based decision support systems however require the existence of manufacturing workflow plans. The presented model processing and workflow estimation process is suitable to provide the necessary manufacturing workflow plans as the input of simulation-based decision support systems. With the integration of the presented method and existing Simulation-based decision support systems a complete decision support ecosystem could be created, to aid the engineers from new orders to manufacturing steps allocated to manufacturing machines.

8 Conclusion and future research

In this paper, we presented a method to support the intelligent planning of manufacturing workflows, based on historical manufacturing data. The suggested method takes 3D models of shapes and product requirements as input and estimates the manufacturing steps required to manufacture the product defined by the model and the user requirements.

Our method is based on similarity metrics defined to calculate the similarity of 3D models in terms of their geometric and topological properties. Based on the combination of these metrics and decision support methods considering the derived characteristics of the 3D models we estimated the manufacturing workflows needed to manufacture the workpieces.

We evaluated our method on the dataset of Component Ltd. and found that a combination of the selected metrics can effectively support this decision support method. Our approach showed promising results in determining the steps that can help the creation of early manufacturing design plans and, enriched by operational information such as resource parameters and cost, a basis for a quote.

There are several future research steps that could improve the method and we can also benefit from the lessons learned here in other domains. Intelligent methods could be applied to develop different metrics combinations to find the best metric suite for a given question, i.e., create composite metrics to find the best similar shape from the past for a given type of manufacturing steps. We are also considering to introduce multi-aspect categorization for shapes and applying the similarity metrics to find the best match within a category/group of shapes.

For a more generic application of the suggested multi-set metrics, we are investigating the applicability of these in other domains, e.g., in the testing of self-driving vehicles.

Acknowledgements

This work has been partially supported by the GINOP-2.1.2-8-1-4-16-2018-00405 project.

The project supported by the Doctoral Excellence Fellowship Programme (DCEP) is funded by the National Research Development and Innovation Fund of the Ministry of Culture and Innovation and the Budapest University of Technology and Economics, under a grant agreement with the National Research, Development and Innovation Office.

The authors express their gratitude to the colleagues who helped them in the writing of this publication: Dr. András Vörös and Dr. Oszkár Semeráth from BME. The authors would also like to thank all former colleagues at BME, Quanopt, and Component Ltd. who helped with the implementation of the methods described.

References

[1] Regli, W. C., Spagnuolo, M. "Introduction to shape similarity detection and search for CAD/CAE applications", Computer-Aided Design, 38(9), pp. 937–938, 2006. https://doi.org/10.1016/j.cad.2006.08.002

[2] Cardone, A., Gupta, S. K., Karnik, M. "A Survey of Shape Similarity Assessment Algorithms for Product Design and Manufacturing Applications", Journal of Computing and Information Science in Engineering, 3(2), pp. 109–118, 2003. https://doi.org/10.1115/1.1577356

[3] Rensink, A., Distefano, D. "Abstract Graph Transformation", Electronic Notes in Theoretical Computer Science, 157(1), pp. 39–59, 2006. https://doi.org/10.1016/j.entcs.2006.01.022

[4] Francia, D., Ceruti, A., Caligiana, G., Liverani, A. "Shape Similarity: Methods and Industrial Applications", In: Proceedings of the International Conference on Innovative Methods in Product Desing (IMProVe 2011), Padova, Italy, 2011, pp. 289–298. ISBN 978-887784-333-3

[5] Leake, D. B., Kinley, A., Wilson, D. "Case-Based Similarity Assessment: Estimating Adaptability from Experience", In: Proceedings of the 14th National Conference on Artificial Intelligence, Menlo Park, PA, USA, 1997, pp. 674–679. ISBN 0262510952

[6] Zhang, R., Zhou, X. "Similarity Assessment of Mechanical Parts Based on Integrated Product Information Model", Journal of Computer and Information Science in Engineering, 11(1), 011001, 2011.
https://doi.org/10.1115/1.3563055

[7] Osada, R., Funkhouser, T., Chazelle, B., Dobkin, D. "Shape distributions", ACM Transactions on Graphics, 21(4), pp. 807–832, 2002.
https://doi.org/10.1145/571647.571648

[8] Berger, V. W., Zhou, Y. "Kolmogorov–Smirnov test: Overview", In: Balakrishnan, N., Colton, T., Everitt, B., Piegorsch, W., Ruggeri F., Teugels, J. L. (eds.) Wiley Statsref: Statistics Reference Online, 2014. ISBN 9781118445112
https://doi.org/10.1002/9781118445112.stat06558

[9] da Fontoura Costa, L. "Multisets", [preprint] ResearchGate, 15 October 2021.
https://doi.org/10.13140/RG.2.2.26106.21440/4

[10] Quinlan, J. R. "Learning decision tree classifiers", ACM Computing Surveys (CSUR), 28(1), pp. 71–72, 1996.
https://doi.org/10.1145/234313.234346

[11] Breiman, L. "Random Forests", Machine Learning, 45(1), pp. 5–32, 2001.
https://doi.org/10.1023/A:1010933404324

[12] Cardone, A., Gupta, S. K., Deshmukh, A., Karnik, M. "Machining feature-based similarity assessment algorithms for prismatic machined parts", Computer-Aided Design, 38(9), pp. 954–972, 2006.
https://doi.org/10.1016/j.cad.2006.08.001

[13] Çiçek, A. "Similarity and scaling assessments of mechanical parts using adjacency relation matrices", Journal of Materials Processing Technology, 206(1–3), pp. 106–119, 2008.
https://doi.org/10.1016/j.jmatprotec.2007.12.007

[14] Ma, S., Tian, L. "Hierarchical 3D mechanical parts matching based-on adjustable geometry and topology similarity measurements", Journal of Central South University, 21(1), pp. 89–99, 2014.
https://doi.org/10.1007/s11771-014-1920-9

[15] Chu, C.-H., Hsu, Y.-C. "Similarity assessment of 3D mechanical components for design reuse", Robotics and Computer-Integrated Manufacturing, 22(4), pp. 332–341, 2006.
https://doi.org/10.1016/j.rcim.2005.07.005

[16] Harik, R. F., Barakat, F. G. "Part Similarity Algorithm Based on Manufacturing Features Similarity", Computer-Aided Design and Application, 7(5), pp. 663–674, 2010.
https://doi.org/10.3722/cadaps.2010.663-674

[17] Kusiak, A., Cho, M. "Similarity coefficient algorithms for solving the group technology problem", International Journal of Production Research, 30(11), pp. 2633–2646, 1992.
https://doi.org/10.1080/00207549208948181

[18] Renner, G., Stroud, I. "Volumetric tools for machining planning", In: Kovács, G. L., Bertók, P., Haidegger, G. (eds.) Digital Enterprise Challenges: Life-Cycle Approach to Management and Production, Springer, Boston, MA, 2002, pp. 106–116. ISBN 978-0-387-35492-7
https://doi.org/10.1007/978-0-387-35492-7_11

[19] Kasie, F. M., Bright, G., Walker, A. "Decision support systems in manufacturing: a survey and future trends", Journal of Modelling in Management, 12(3), pp. 432–454, 2017.
https://doi.org/10.1108/JM2-02-2016-0015