Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 99 (2021) 301–306

www.elsevier.com/locate/procedia

14th CIRP Conference on Intelligent Computation in Manufacturing Engineering, CIRP ICME ´20

Adaptive quality prediction in injection molding based on ensemble learning

Alexander Schulze Struchtrup[a], Dimitri Kvaktun[a,*], Reinhard Schiffers[a]

[a]*Universität Duisburg-Essen, Lotharstraße 1, Duisburg 47057, Germany*

* Corresponding author. Tel.: +49-203-379-1655 ; fax: +49-203-379-4379. *E-mail address:* dimitri.kvaktun@uni-due.de

Abstract

Despite the central role of quality, industry applications of data-based quality prediction for thermoplastics injection molding are rare, because of a suboptimal cost-benefit ratio. Therefore, we present a holistic approach for seamless part quality prediction, which automates the necessary data processing steps. Since the performance of the seven supervised learning algorithms applied with Bayesian hyperparameter-optimization depends on aspects such as process state, etc., we combine the learnt models using an ensemble-method to ensure good results under varying conditions. The results show that the ensemble's performance significantly depends on the chosen ensemble-hyperparameters, so future research should focus on their automatic identification.

Keywords: "Machine learning; Smart manufacturing; Process monitoring; Optimization; Injection molding"

1. Introduction and state of the art

Injection Molding offers the production of complex geometries in a single, discontinuous production step, which is highly reproducible. The process consists of the following steps [1]: Initially, the plastics material is plasticized by the rotational movement of the screw and additional heat input using heating elements. Thereupon, the melt is injected in the cavity of the mold, where the final shape of the product is formed. To compensate the shrinkage of the material due to the cooling, the screw applies further pressure during the packing pressure phase. Finally, the mold opens and the cooled part is ejected.

The injection molding process can be described via three control loops [2]. The first loop includes machine control, which adjusts machine parameters like injection speed, barrel temperature or holding pressure. The optimization potential here is considered to be largely exhausted [3]. The second loop uses process variables such as the melt pressure during injection or cavity pressure to control the machine actuators. This is the second most researched field with products like *KraussMaffei APC+* [4] or *Engel iQWeightControl* [5] available. Recent research shows that even further improvement is possible [6]. An overview of the various mathematical and statistical descriptions and optimizations of the injection molding process and the individual steps is provided by Fernandes, Pontes et al. [7]. Despite the high effort taken in product and process optimization by the plastics processing companies, internal and external perturbations are still negatively affecting the quality of the molded part. Sample based quality inspection however leads to delayed detection of scrap parts or in worst cases may fail to achieve even this.

Therefore, the focus today lies on the third loop, which uses the part quality for controlling the process. It should be noted here that this process control has been pursued for several decades now [8–10]. Regardless of good research results, these methods have not yet established themselves in the industry. Nevertheless, this type of quality control is back in the focus of research for several reasons.

Some of the most important reasons are listed below [11]:

- greater computational power at lower cost,
- improved and easily accessible algorithms,

2212-8271
10.1016/j.procir.2021.03.045

- a large amount of data in higher quality is available and easily accessible through improved sensor and machine technology,
- social focus, reinforced by political strategies like “Industry 4.0” (Germany) [12] or „Advanced Manufacturing“ (USA) [13].

Still, the use of direct part quality for machine control in an industrial context is often not possible because of the additional equipment-cost and time involved. Thus, the part quality must be predicted from available machine and process parameters. Different research studies show good results using machine learning for quality prediction. Tercan, Guajardo et al. [14] used transfer learning for improving the performance of neural networks and reducing the learning phase by using simulation data. Ogorodnyk, Lyngstad et al. [15] showed that artificial neural networks and decision trees can be used for the distinction of high- and low-quality injection molding parts. They further improved the prediction accuracy using two different feature selection methods to eliminate irrelevant process parameters [16]. In the field of machine learning for regression problems, it has been shown that high predictive performance for the part width is possible [17]. The authors used various regression methods including artificial neural networks and support vector machine on raw signal und image data. Building on this work, they have applied generative adversarial networks on thermographic images to predict the final injection molded part geometry [18].

From this selection of current applications of machine learning in injection molding, it is clear that machine learning can be used for quality prediction. However, in the presented cases, a model is often chosen and retained for the prediction. Yet it has been shown that the injection molding process is dynamic and changes over time. To our best knowledge, the adaption of the machine learning models based on new, unseen data in injection molding was not studied.

In this work, we are proposing a holistic quality prediction framework, which automates the necessary data processing steps. Furthermore, we evaluate the question whether the adaptive selection between different machine learning algorithm can improve the prediction performance of the overall framework. For this purpose, we use an ensemble-like strategy. The strategy is to some extent similar to the switching strategy developed by Bayrak, Wang et al. [19] for the prediction of product concentration in mammalian cell culture bioreactors.

The paper is organized as follows: Section 2 describes the methodology we used in this work. This includes the different machine learning models used in the holistic framework, the different data processing steps, which are necessary to get high prediction performances and the adaptive model selection based on ensemble learning. Additionally, the experimental data is presented; the results from the different experiments are provided in section 3. Finally, section 4 summarizes this work and gives an outlook for future research.

2. Methodology

2.1. Machine learning algorithms

The adaptive model selection/weighting algorithm utilizes artificial neural networks (ANN), support vector machines (SVM), binary decision trees (DT), k-nearest neighbors (kNN), ensemble methods (EM) (Bagging and Boosting) based on DTs and Gaussian process regression (GP). Furthermore, normal multiple linear regression (MLR) [20] is added to the analysis to compare classical statistical methods with machine learning.

Artificial neural network. ANN is one of the most popular learning algorithms, used in many various applications. They consist of one input layer, one or more hidden layers and one output layer. Each layer consists of interconnected neurons, which are processing the input sum using an activation function, while the connection weights are adjusted in the learning process [21].

Support vector machine. SVM are initially developed for classification problems. The goal is to create an optimal hyperplane in n-dimensional space with the maximum margin between the classes [22]. The algorithm was adapted to regression problems using a margin of tolerance [23].

Binary decision tree. DTs are simple and easy to understand and interpret using different split metrics for tree construction. The models are created by recursively partitioning the data space by minimizing the root mean squared error [24].

k-nearest neighbors. kNN (cf. Fig. 1) is a “lazy learning” algorithm which output depends on the nearest neighbors in the data space. For regression problems the output value is determined through the (weighted) average value of the neighbors [25].

Ensemble method. Ensemble methods combine different learning algorithms to enhance the predictive performance, from a single algorithm. In this work we focus on LSBoost, which is a Boosting method [26], and Bagging, which is a weighted combination of multiple decision trees [27].

Gaussian process regression. GPR is a probability distribution over possible function. Using the training data as evidence, the prior is updated by use of Bayes´ rule [28].

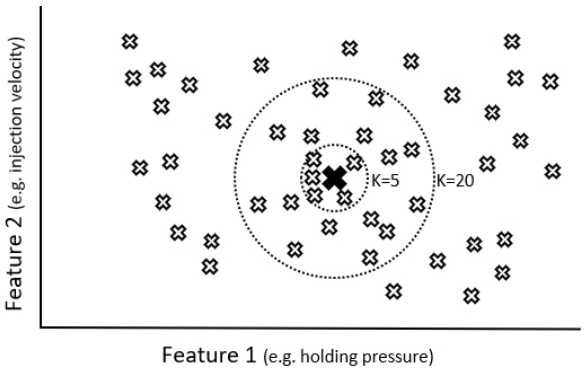

Fig. 1. k-nearest neighbor search, exemplary display for two features

2.2. Adaptive model selection

The general framework used in this work is displayed in Fig. 2. It mainly consists of four superordinate steps. The first step is data pre-processing. The complete data set is split into a test and learning data set using the holdout-method. While the learning data set is required for training the models and hyperparameter optimization, the test data set is used to provide unseen data for the model evaluation [29]. We call this type of data split "horizontal data split", because each injection molding cycle is written row by row in the data log during production. In this research, we use 80 % data for learning and 20 % data for testing.

In the next step, we use the learning data set for feature selection. Feature selection is an important step of selecting a feature set that is most relevant for modeling. Furthermore, the reduced feature space decreases the computational effort to build a model [30]. In this study we use a filter approach with sequential forward selection (SFS) [31] as the search strategy, correlation-based feature selection (CFS) [32] as the performance evaluation metric and Pearson's correlation coefficient (PCC) [33] as the relevance criterion. This combination has proved to be very effective in previous investigations [34] for quality prediction with regression machine learning models. After the feature subset is selected, the learning and test data sets are adjusted accordingly. We call this type of data selection "vertical data selection", since the features of the injection molding cycle (e. g. holding time, injection speed, etc.) are written column by column in the data log.

After the feature selection the actual modeling takes place. First, a horizontal data split is performed to create a neighbor data set required for the subsequent ensemble learning. 50 % of the learning data set is used for modeling. Here, the six different machine learning models and linear regression mentioned earlier are trained. Hyperparameters are optimized via Bayesian hyperparameter optimization [35] using 5-fold cross-validation [36] for evaluation of parameter sets. An overview of the hyperparameters can be seen in [37]. In this research, the hyperparameters are divided into categorical and numerical according to Bermúdez-Chacón, Gonnet et al. [38]. This leads to a tree-like model structure. The prediction performance of each branch model (here base model) is evaluated using the test data set and the best models of the different machine learning algorithms are selected as best base models. The coefficient of determination (R^2) [26] is used as goodness-of-fit criterion.

The ensemble learning takes place after the best base models are selected. In this study we evaluated three different strategies, whereby only two are shown in the workflow.

The first strategy simply averages the outputs of the seven base models. This case, which we call "unweighted average ensemble" is unshown in the workflow. Obviously, we do not have to make use of the neighborhood information and there are no hyperparameters in this strategy.

The second strategy, which we call "single model selection ensemble", is the opposite of the first. Here kNN-search with Euclidean distance metric (cf. Fig. 1) is applied to each data point of the test data set to determine the k-nearest in the neighbor data set. This results in a data set of test point neighbors, which have the most similarity (shortest distance in space) to the test data. The corresponding process and quality data can be selected from the neighbor data set through horizontal data selection.

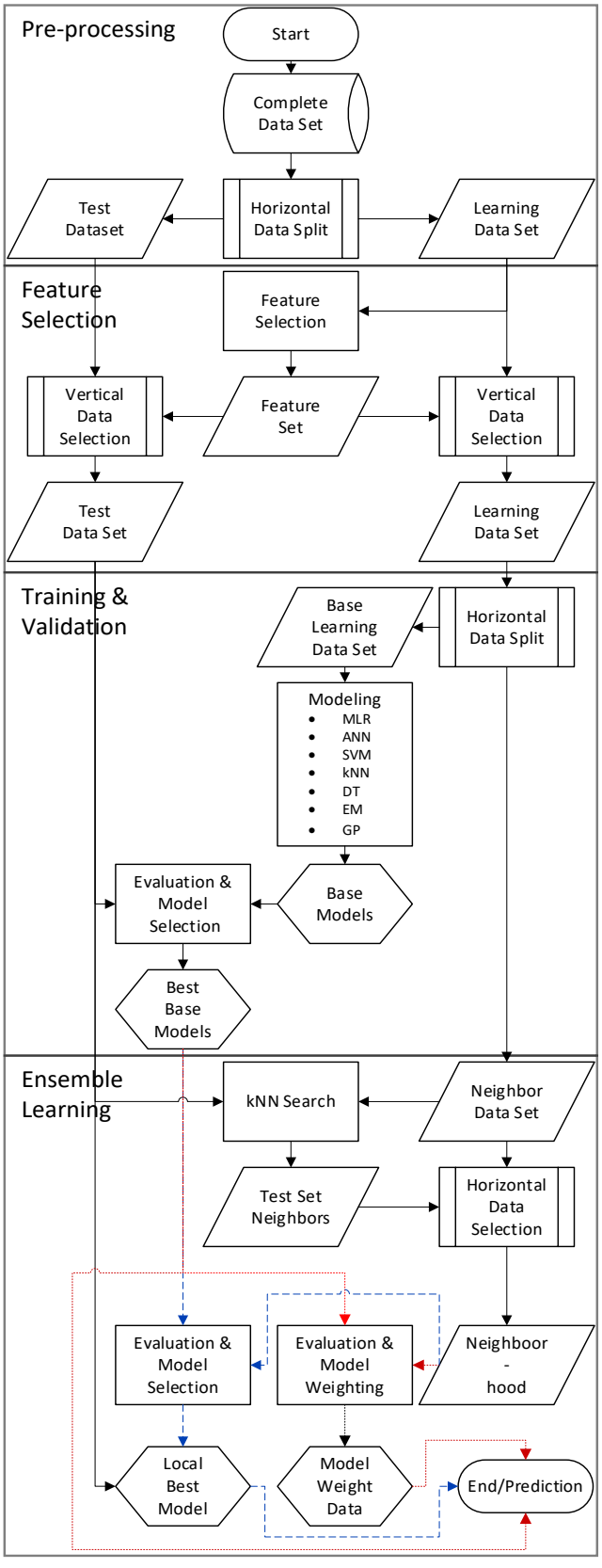

Fig. 2. algorithm workflow

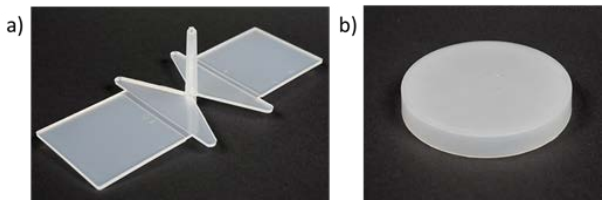

Fig. 3. (a) plate specimen; (b) cover specimen

This neighborhood data set is used to evaluate every best base model. The model which best predicts the neighborhood is selected as local best model and taken for prediction of the corresponding test data point. The workflow links, exclusively used for this strategy, are dashed blue in Fig. 2.

The last strategy is called “weighted average ensemble” and is a combination of the first and the second strategy. In this case all seven base models are incorporated and their output is combined. However, unlike in the first strategy, the outputs are weighted using the base models’ coefficients of determination achieved on the test point’s neighborhood (cf. formula 1).

$$w_j = \frac{R_j^2}{\sum_i^N R_i^2} \quad (1)$$

Here, w_j is the weight of the model j and N the total number of models incorporated. The dotted red lines in the workflow chart are exclusively used for strategy 3.

2.3. Experimental study

In this paper, practical experiments of plate and cover specimens (cf. Fig. 3) provide the data for the holistic quality prediction framework including the ensemble learning. The experiments were carried out on a *KraussMaffei 120-380 PX* fully-electric injection molding machine (IMM). We use the weight and length of the plate specimens and the weight and diameter of the cover specimens as quality criteria. The weight is measured with a *Sartorius Entris 153I-1S* lab balance, upon which they are automatically placed by the IMM’s linear robot *LRX50*. The length and the diameter are automatically photographed with a *Canon Eos 5D Mark III DSLR-camera with EF 70-200 mm f/4L USM lens*. The available process features are directly taken from the machine’s actual value protocol. For data generation, six different process states are induced, to mirror real-world injection molding production situations: start-up, stable process, downtimes, regrind material, regrind material + adaptive process control (APC) and a central composite design (DOE) At each process state, 1000 injection molding cycles are carried out, except for the DOE with 860 cycles due to its predefined structure. More information about the various experiments are reported in [31]. In total, 48 machine and process parameters and two times two quality criteria were logged during each cycle. All data processing steps are carried out in *MATLAB2019b* using the statistics and machine learning toolbox as well as additional, own implementations of algorithms, where necessary.

3. Results

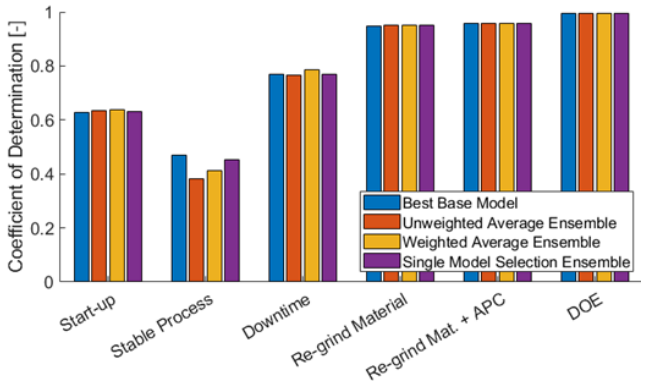

Fig. 4. Prediction performance for the plate specimen weight

The achieved results of the best base models as well as the three ensemble variants are depicted in Fig. 4 to Fig. 7. Already the comparison of the coefficients of determination of the base models (blue bars) illustrates that the model quality strongly depends on the examined molded part, the quality criterion and the process state in which process and quality data are monitored.

Regarding the plate weight (cf. Fig. 4), very good model can be achieved when using re-grind material with or without adaptive process control or when conducting a DOE, in which case an excellent coefficient of determination of 99.5 % is achieved. In comparison, the results achieved with the startup and downtime data sets are rather inferior and the stable process result is insufficient. For the plate length (cf. Fig. 5), the re-grind material data set yields the best performance with a very good coefficient of determination of 92.0 %. In combination with the adaptive process control the re-grind performs acceptably, which also applies to the DOE. The model qualities achieved for startup, stable process and process with downtimes are insufficient.

Regarding the cover weight (cf. Fig. 6), a very good result is achieved with the DOE with a coefficient of determination of 96.5 %. Startup, stable process and the process with downtimes also yield good results around 80 %. Both re-grind data sets yield only mediocre results. For the prediction of the cover diameter (cf. Fig. 7), only the DOE yields a good model with a coefficient of determination of 92.1 %.

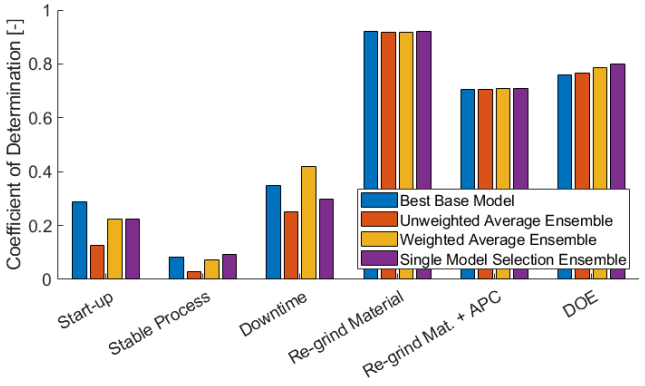

Fig. 5. Prediction performance for the plate specimen length

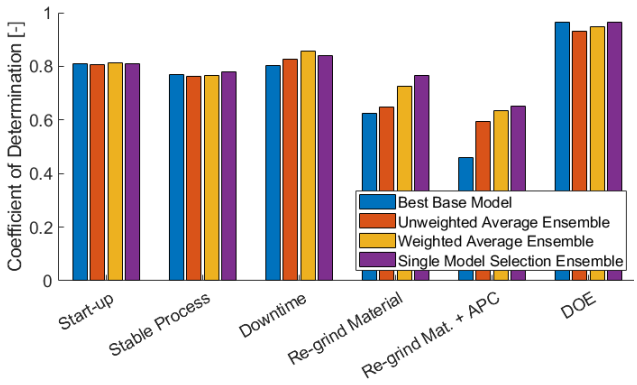

Fig. 6. Prediction performance for the cover specimen weight

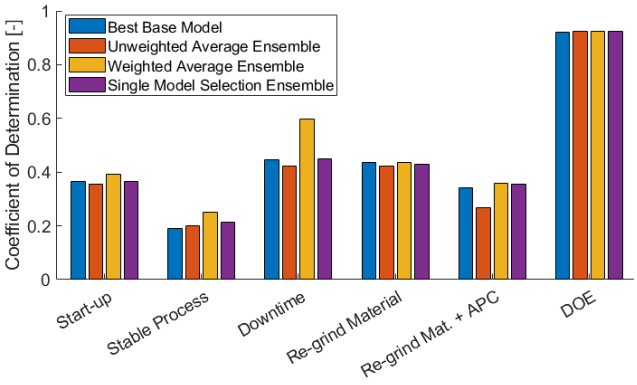

Fig. 7. Prediction performance for the cover specimen diameter

Further evaluations not displayed in this paper have shown, that the partly insufficient model qualities are mainly due to a too small variance in quality data compared to the measurement precision, i.e. the concerned process states are too stable. Consequently, to achieve higher coefficients of determination, one must choose either another process state with higher variance (in most cases the DOE works best here) or increase the precision of the measuring equipment.

A comparison of the performance of the seven applied supervised learning algorithms reveals that the artificial neural network (fully-connected feedforward with one hidden layer) yields the best results on 10 of 24 data sets, followed by Gaussian process regression performing best on seven data sets. Support vector regression is third with three top positions, followed by the decision tree-ensembles (two), binary decision trees as well as multiple linear regression (both one). The “lazy learner” k-nearest neighbor regression performs under average in all cases. Obviously, different learning algorithms perform best on different conditions, which motivates our ensemble-approach that incorporates multiple learners.

For ensemble learning, we applied a grid search for the hyperparameter k, starting with five neighbors going up to 400 (size of neighbor data set) with a step size of five. Fig. 8 shows exemplarily the model performances depending on the neighborhood size for one of the 24 data sets. As standard of comparison for the three ensemble models, we use the best base model, since the aim of an ensemble model is to perform better than this. In the depicted case, we can see that the unweighted average ensemble is slightly better than the best base model. Obviously, both are not dependent on the chosen neighborhood size. For the other two ensembles, there is a running-in characteristic until k = 80, after which both curves reach a plateau until approx. k = 320. After leaving the plateau, the single model selection ensemble performance converges to the best base model performance and the performance of the weighted average ensemble converges to the unweighted average ensemble performance.

Looking at the overall performances of the ensemble models on all 24 data sets (cf. Fig. 4 to Fig. 7), we find that the unweighted average ensemble surpasses the best base model’s performance in 12 of 24 cases. As this is exactly 50 %, we conclude that the question whether the ensemble is better than the best base model, is a question of coincidence, so the unweighted average ensemble does not create measurable added value. Still, it is interesting, that it performs basically comparable as the best base model, taking into account that all models, also the below-average ones, are equally used for calculating the prediction.

The weighted average ensemble performs significantly better, surpassing the best base model in 19 data sets and the single model selection ensemble performs best with 20 of 24 data sets. Since these two variants outperform the best base model on the great majority of data sets, it seems reasonably to use them for prediction. Looking at Fig. 8 again, one can see that the fluctuations of the single model selection ensemble are significantly larger than those of the weighted average ensemble. This is not surprising since the latter averages the prediction of all included base models which is expected to have a smoothing effect, however, when it comes to automatic hyperparameter optimization (which we excluded in this paper) it seems a little bit easier to find a good number of neighbors for the weighted average ensemble than for the single model selection ensemble.

4. Conclusion and Outlook

Ensemble learning is often used to improve the prediction performance by combining different single or base models to one new model. In this study, we evaluated whether ensemble learning can enhance the quality prediction in injection molding. We tested three different strategies using an unweighted average ensemble, a weighted average ensemble and a single model selection ensemble.

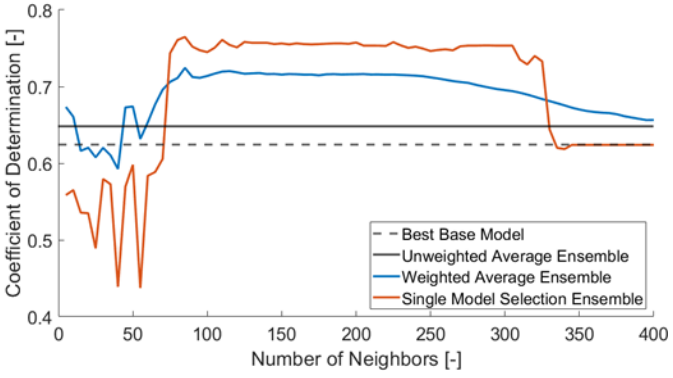

Fig. 8. Model performances on cover weight with re-grind material data set

All variants are integrated in a quality prediction framework, which automatically performs data pre-processing, feature selection, training and validation of the models including hyperparameter optimization.

The conducted analysis shows that ensemble learning strategies can generally improve the prediction performance in comparison to a single prediction model. While in our study, the unweighted average ensemble yields result comparable to those of the best base model, the weighted average ensemble and the single model selection ensemble outperform the best base model on 19 and 20 of 24 data sets, respectively.

Such adaptive model selection or weighting can be used in a real-time production process to dynamically change between various machine learning algorithms. This has the advantage that no machine learning algorithm needs to be selected from the beginning. In future work, the number of neighbors for the neighborhood data set should be considered a hyperparameter which needs to be automatically optimized. Moreover, we only used the Euclidian distance as distance metric. Other metrics, like Chebychev or Mahalanobis which show good results in classification problems, may improve the ensemble prediction performance and should therefore also be considered.

References

[1] Osswald T, Turng L-S, Gramann P. Injection molding handbook. 2nd ed. Munich: Hanser; 2008.

[2] Karbasi H, Reiser H, editors. Smart Mold: Real-Time in-Cavity Data Acquisition. First Annual Technical Showcase & Third Annual Workshop, Canada; 2006.

[3] Schiffers R.Verbesserung der Prozessfähigkeit beim Spritzgießen durch Nutzung von Prozessdaten und eine neuartige Schneckenhubführung. PhD thesis, University of Duisburg-Essen; 2009.

[4] Schiffers R, Kruppa S, Moser S. Für jeden Schuss der richtige Umschaltpunkt. Kunststoffe 2014(11); p. 58-63.

[5] Pillwein G. Method for quantifying process fluctuations in an injection operation of an injection molding machine. Patent: AT511391B1; 2011.

[6] Hornberg K, Hopmann C. Die Umschaltproblematik und ihre Folgen: Phasenlose Prozessregelung soll den Umschaltvorgang revolutionieren. Kunststoffe 2020(3); p. 82-86.

[7] Fernandes C, Pontes A J, Viana J C, Gaspar-Cunha A. Modeling and Optimization of the Injection-Molding Process: A Review. Advances in Polymer Technology 2018(37); p. 429-449.

[8] Vaculik R. Regelung der Formteilqualität beim Spritzgießen auf Basis statistischer Prozessmodelle. PhD thesis, RWTH Aachen; 1996.

[9] Al-Haj Mustafa M. Modellbasierte Ansätze zur Qualitätsregelung beim Kunststoffspritzgießen. Aachen: Shaker; 2000.

[10] Häußler J. Eine Qualitätssicherungsstrategie für die Kunststoffverarbeitung auf der Basis künstlicher Neuronaler Netzwerke. PhD thesis, University of Duisburg-Essen; 1994.

[11] Wuest T, Weimer D, Irgens C, Thoben K-D. Machine learning in manufacturing: advantages, challenges, and applications. Production & Manufacturing Research 2016(4); p. 23-45.

[12] Bundesministerium für Bildung und Forschung. Industrie 4.0. Online: https://www.bmbf.de/de/zukunftsprojekt-industrie-4-0-848.html.

[13] Advanced Manufacturing National Program Office (AMNPO). Strategy for American Leadership in Advanced Manufacturing. Online: https://www.manufacturing.gov/news/announcements/2018/10/strategy-american-leadership-advanced-manufacturing; 2018.

[14] Tercan H, Guajardo A, Heinisch J, Thiele T, Hopmann C, Meisen T. Transfer-learning: Bridging the gap between real and simulation data for machine learning in injection molding. Procedia CIRP 2018(72); p. 185-190.

[15] Ogorodnyk O, Lyngstad O V, Larsen M, Wang K, Martinsen K. Application of Machine Learning Methods for Prediction of Parts Quality in Thermoplastics Injection Molding. In: Wang K, Wang Y, Strandhagen J O, Yu T, editors. Advanced Manufacturing and Automation VIII. Singapore: Springer Singapore; 2019. p. 237–244.

[16] Ogorodnyk O, Lyngstad O V, Larsen M, Martinsen K. Application of feature selection methods for defining critical parameters in thermoplastics injection molding. Procedia CIRP 2019(81); p. 110–114.

[17] Nagorny P, Pillet M, Pairel E, Le Goff R, Loureaux J, Wali M, Kiener P. Quality prediction in injection molding. In: 2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA); 2017. p. 141–146.

[18] Nagorny P, Lacombe T, Favreliere H, Pillet M, Pairel E, Le Goff R, Wali M, Loureaux J, Kiener P. Generative adverserial networks for geometric surfaces prediction in injection molding: Performance analysis with Discrete Modal Decomposition. In: Proceedings 2018 IEEE International Conference on Industrial Technology (ICIT); 2018. p. 1514–1519.

[19] Bayrak E S, Wang T, Tulsyan A, Coufal M, Undey C. Product Attribute Forecast: Adaptive Model Selection Using Real-Time Machine Learning. IFAC-PapersOnLine 2018(51); p. 121-125.

[20] Mills T C. The Classical Linear Regression Model, in Analysing economic data: A Concise introduction. London: Palgrave Macmillan UK; 2014.

[21] Hagan M T, Demuth H B, Beale, M.H., Neural network design, 2nd ed. Boston: PWS Publ. Co.; 2014.

[22] Cortes C, Vapnik V. Support-vector networks. Machine Learning 1995(20); p. 273-297.

[23] Smola A J, Schölkopf B. A tutorial on support vector regression. Statistics and Computing 2004(14); p. 199-222.

[24] Breiman L, Friedman J, Stone C J, Olshen R A. Classification and regression trees. 1st ed. Boca Raton: CRC Press; 1984.

[25] Biau G, Devroye L, Dujmović V, Krzyżak A. An affine invariant k-nearest neighbor regression estimate. Journal of Multivariate Analysis 2012(112); p.24-34.

[26] Friedman J, Hastie T, Tibshirani R. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer; 2009.

[27] Breiman L. Random Forests. Machine Learning 2001(45); p. 5-32.

[28] Rasmussen C E, Williams C K I. Gaussian processes for machine learning. 3rd edn. Cambridge, Mass: MIT Press; 2008.

[29] Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Statistics Surveys 2010(4); p. 40-79.

[30] Jovic A, Brkic K, Bogunovic N. A review of feature selection methods with applications. In: Information and Communication Technology, Electronics and Microelectronics (MIPRO); 2015, p. 1200–1205.

[31] Chandrashekar G, Sahin F. A survey on feature selection methods. Computers & Electrical Engineering 2014(40); p. 16-28.

[32] Hall M A. Correlation-Based Feature Selection for Machine Learning. PhD thesis, University of Waikato; 1999.

[33] Asuero A G, Sayago A, González A G. The Correlation Coefficient: An Overview. Critical Reviews in Analytical Chemistry *(CRIT REV ANAL CHEM)* 2006(36); p. 41-59.

[34] Schulze Struchtrup A, Kvaktun D, Schiffers R. Comparison of Feature Selection Methods for Machine Learning based Injection Molding Quality Prediction. In: Proceedings of the Europe/Africa Conference Pretoria 2019-Polymer Processing Society PPS; 2019 submitted.

[35] Claesen M, De Moor B. Hyperparameter Search in Machine Learning. In: MIC 2015: The XI Metaheuristics International Conference; 2015, p. 1-5.

[36] Refaeilzadeh P, Tang L, Liu H. Cross-Validation. In: LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems. Boston, MA: Springer; 2009.

[37] Schulze Struchtrup A, Kvaktun D, Schiffers R. A Holistic Approach to Part Quality Prediction in Injection Molding Based on Machine Learning. In: Hopmann C, Dahlmann R, editors. Advances in Polymer Processing 2020: Proceedings of the International Symposium on Plastics Technology. Berlin, Heidelberg: Springer Berlin Heidelberg; 2020, p. 137–149.

[38] Bermúdez-Chacón R, Gonnet G H, Smith K. Automatic problem-specific hyperparameter optimization and model selection for supervised machine learning: Technical Report. ETH Zurich; 2015.

DOI: 10.1016/j.procir.2021.03.045
URN: urn:nbn:de:hbz:465-20220913-135720-9