White Paper

Cavity pressure-based machine learning service for advanced injection molding processes

Abstract / motivation

Injection molding is the state-of-the-art process used to produce plastic parts for various applications. To increase productivity, highly standardized and automated processes are established with quality control and reproducibility as the main keys to success. Measurement of cavity pressure has proven to be a powerful parameter for characterizing the quality of the resultant parts. On this basis, cavity pressure measurement systems such as ComoNeo from Kistler allow real-time evaluation of the process and automated separation of parts that fail to meet the requirements. This is achieved by comparing the actual readings from cavity pressure measurements with a previously stored reference curve.

This highly effective method has played a major part in improving quality, reducing waste, and increasing the stability of the injection molding production process. Parts are evaluated and rejects are separated with a high degree of reliability – but even so, the operator's intuition and know-how are still critical factors in defining corrective actions.

Thus far, operators have been required to set the rejection limits for quality control manually in advance of production – a time-consuming task that can only be performed by experienced personnel. We now propose a new approach based on artificial intelligence and data science that aims to achieve two goals: first, to improve the user experience by eliminating the need to define limits; and second, to improve the performance of quality prediction by reducing the number of false positives (parts of bad quality which the system erroneously classifies as good).

To achieve these objectives, we make use of an advanced machine learning model based largely on cavity pressure measurements together with some additional machine data. The model is trained to detect anomalies in the parameter space spanned by individually engineered mathematical features. By combining domain expertise in injection molding with the mathematical technique of time series analysis, this approach allows in-depth understanding and interpretation of the model's predictions. The results are used to detect anomalies (i.e. suspected deviations) in the process. As a further refinement, we have incorporated a model explanation approach based on a feature importance algorithm into a newly developed second algorithm. This method enables operators to determine possible causes of quality deviations so they can initiate adjustments of the relevant process parameters.

Authors

Dr. Patrick Waibel

Sandra Haltmeier

Dr. Robert Vaculik

Thomas Wuhrmann

Dr. Nikola Pascher

Contents

1. Injection molding – process and challenges

The injection molding process is the standard method for mass production of plastic products. The brief explanation below is offered as an introduction to the process. For a thorough explanation of the injection molding process and a general layout of the injection molding machine, please refer to Johannaber and Michaeli [1].

The plastic material is melted in the plastification or injection unit and is then injected into a mold containing a cavity which replicates the negative form of the desired part. Once the melt has cooled down in the temperature-controlled mold and the plastic has solidified, the ready-shaped part is ejected. This may appear to be a simple process, but the detailed configuration and setup are highly complex operations that require expert know-how – due in particular to the large number of process parameters that influence the required quality features.

In order to monitor and control the process, the relevant process values must be measured. It has been shown (e.g. by Bichler [2]) that cavity pressure is a powerful parameter for describing and subsequently monitoring the process [2]. The measured pressure curve in the cavity can generally be segmented into different process phases:

1. The **filling phase**, when the cavity is filled at a controlled flow rate. This process is controlled by the movement of the transversal screw in the injection unit. As soon as the cavity is 100% filled, the velocity control flow front becomes irrelevant; at this point, the control parameter is switched from "screw position" to "pressure" (this is known as the switchover point).
2. The **compression phase**, when the prevailing pressure is adjusted to the desired holding pressure; this should compensate for material shrinkage in the subsequent holding phase. As a result of this adjustment, the pressure in the cavity increases to its maximum value.
3. The **holding phase**, when the adjusted pressure level presses a certain amount of fresh melt into the cavity to compensate for the volume shrinkage of the plastic material as it solidifies. This can continue until the open area in the "gating zone" freezes and closes the open melt channel. The material in the cavity then cools down and shrinks. This is clearly visible as a downgrading of the pressure profile until atmospheric pressure is reached.

The progression of the pressure reading over time (see Figure 1) during the production cycle has proven to be a very good indicator of possible deviations in the process.

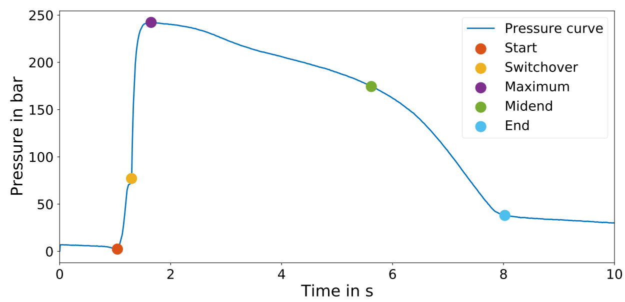

Fig. 1: Typical pressure curve in injection molding processes including characteristic points.

Every pressure curve provides a fingerprint that reflects the quality of the process and the produced part. Various factors may influence this pressure curve, so these same factors can also affect quality. They include environmental conditions (such as temperature), machine aging, mold configuration, material or process deviations and – of course – the machine settings (see: Wortberg [3]).

ComoNeo from Kistler [4] allows real-time sorting of produced parts by qualifying the cavity pressure curve during the injection process with certain boundary conditions. Before the process can begin, this approach requires manual calibration which can only be performed by an experienced operator. Due to the vast parameter space and the countless dependencies involved, plastics injection molding may be regarded as an error-prone process requiring intensive user control [1]. Although the majority of applications operate at high levels of quality and automation, additional quality outcome and process stability are also influenced greatly by the operator's skill and intuition.

In the Industry 4.0 context, we aim to make the process easier to use for operators – while improving quality and process reliability – by using advanced machine learning (ML) algorithms based on cavity pressure data. By combining the information output from the algorithm with process knowledge and additional data from various sensors inside the machine, we are able to deduce additional information about the reasons for quality deviations. This development is an important step towards complete automation of the process and will greatly enhance ease of use for operators.

Recent years have seen increasing efforts to improve the process using data-based machine learning methods. Schiffers et al. [5] offer a comprehensive overview of unsupervised anomaly detection in injection molding. The authors make use of an ensemble method to cluster their input features, enabling rough classification of the clusters and, hence, identification of failure causes. They also stress the importance of the feature engineering step in the overall pipeline. Jung et al. [6] compare several ML models for quality prediction in injection molding processes. They use feature importance tests to identify the most relevant features. Chihun Lee et al. [7] use a weight-prediction model to develop a recommender system. The authors' approach is based on simulation data and transfer learning. They also make use of geometric features to capture the variance of multiple parts in their model input data.

2. Experimental setup and data exploration

Training and evaluation of a machine learning model require the acquisition of a suitable dataset. A comprehensive representation of the underlying distribution of all good cases is also useful when training an anomaly detection model. In the experiment conducted for this White Paper, a typical plastic object was selected as an example: the produced part was approx. 43.5 mm long and 22 mm high, with wall thicknesses of approx. 2 mm. The part is used to mount small signs on boxes used in sensor production at Kistler. The most important section is the snap connection, which should neither be too small nor too large (see Figure 2).

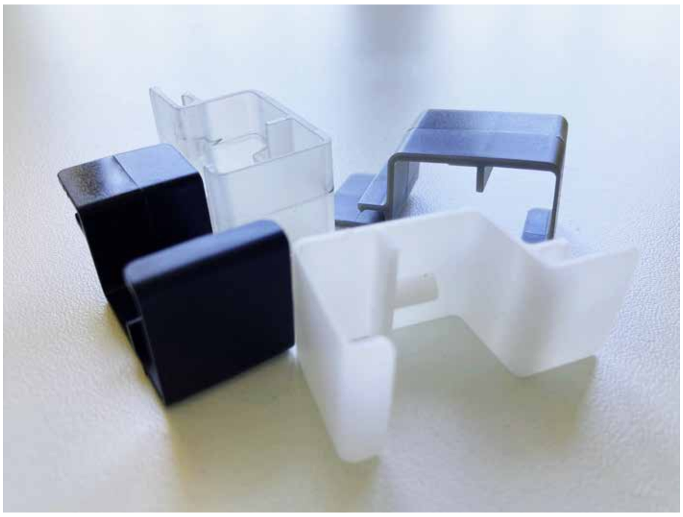

Fig. 2: Typical plastic part used in this study. The plastic clip is used to mount signs on boxes.

This part has been produced more than one thousand times. 15 different machine settings were selected during the experiment, not only to increase the variance of good part distribution but also to induce bad parts (=anomalies) for the model evaluation. A new batch number was set for each machine setting. Each batch began and ended at a specific cycle number. The machine produced a few parts between one machine setting (batch) and the next in order to reach a stable point in production.

The cavity pressure, screw position, hydraulic pressure, nozzle temperature and clamping force were recorded during the experiment, matched with the akvisIO software from Kistler [8] and then exported. Data analysis was performed offline. The cavity pressure sensor was placed at the end of the cavity – the last point to be reached by the melt.

Quality labels were assigned in a manual process, and every second part was marked with a consecutive number. An optical setup in combination with Kistler's KiVision software [9] was then used to measure characteristic lengths and acquire information about the surface quality of the parts and their general appearance. More than 420 parts were labeled in the course of the experiment.

Figure 3a) shows the 15 batches with the corresponding labels for each part. For each batch, one parameter adjustment was varied systematically. Many batches consisted entirely of either good or bad parts, but only a few of them included both good and bad parts.

Figure 3b) shows three different pressure curves from three different batches. All of them have similar characteristics, but there are also significant differences between them. Although curves 1 and 4 seem closer to each other, it is curves 1 and 11 which correspond to good quality, whereas curve 4 corresponds to a part of bad quality. To capture the differences as well as the similarities between the pressure curves, domain expertise in plastics was integrated into the subsequent feature engineering step.

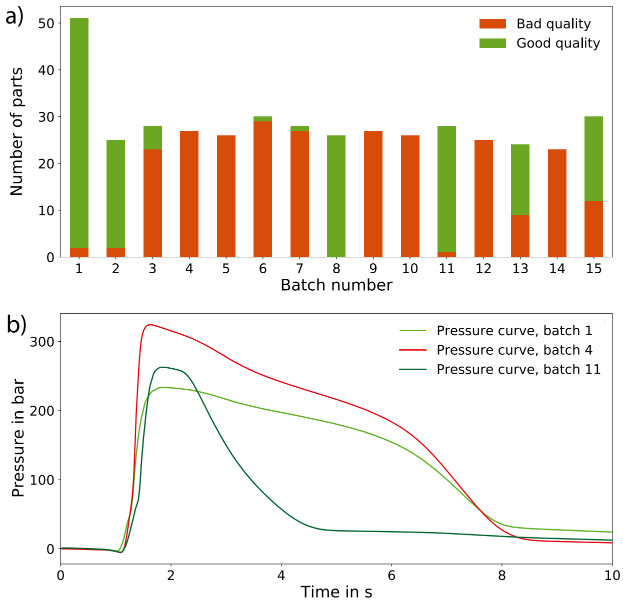

Fig. 3: Overview of the 15 different batches. One parameter was varied per batch. a) Quality outcome of the different batches. b) Examples of corresponding pressure curves.

3. Feature engineering

The feature engineering step was the stage of the data science pipeline where the domain knowledge of the plastics experts came into play. This helped to reduce the very high-dimensional time-series feature space into a lower-dimensional space of interpretable features.

Figure 1 highlights the most relevant points in the pressure curve as seen from the domain expert's perspective. The first of these was the start point of the injection phase: this is not directly visible in the cavity pressure curve, but it must be close to the point where the melt reaches the sensor for the first time. The start point is where the first increase in the measured pressure occurs. However, the measurements also included low-level noise which led to unsmooth curves; this was particularly noticeable in the first and second gradients. Rather than on the original curve, therefore, the start point was set at the point where the curve gradient of a fitted spline with smoother attributes rose above 100 bar/sec for the first time.

The second relevant point was the switchover point from the injection to the compression phase. This point is visible in the cavity pressure curve, and it could be calculated with the help of the second gradient. More specifically, the switchover point is defined by the maximum in the second gradient between the start and the maximum point.

The third relevant point was the maximum point, which also marks the boundary between the compression and holding pressure phases. ComoNeo calculates the maximum pressure point directly, so it was already available and could be used directly.

The fourth relevant point from the domain expert's perspective was the midend point. This point is not directly linked to the process phases but is adjacent to the point in time when the inflow has solidified. The midend point is defined as the first point where the first gradient of the smooth spline is less than –40 bar/sec, starting at the end point and moving towards the maximum point.

The fifth and final relevant point was the end point, where the pressure has largely decreased after the maximum point and is stable at a lower level. Viewed in terms of the process, this is the point when the holding pressure is removed. The algorithm defines the end point as the first point where the first gradient of a smooth fitted spline is less than –25 bar/sec, starting at the end of the curve and moving towards the maximum point.

It is very important that the point detection algorithms are sufficiently robust to cope with multiple curve shapes; otherwise, anomaly detection would be greatly influenced by the points themselves rather than the resulting features. Figure 4 shows a variety of different curve shapes and proves the robustness of the algorithms.

These points could be taken as the basis for calculating several features, including the x and y values as well as the gradients at these points.

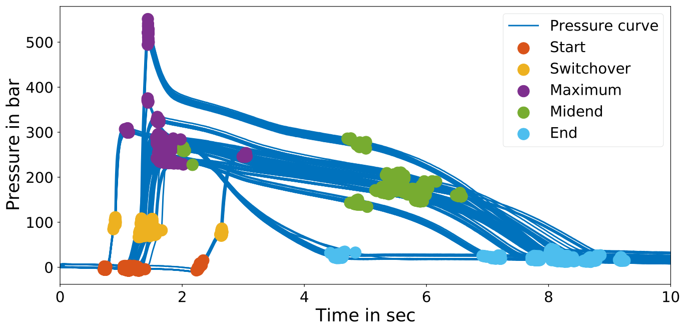

Fig. 4: The algorithms are robust and allow reliable deduction of the relevant points in the pressure curves.

Features based on the Dynamic Time Warping (DTW) method were engineered in order to capture the similarity between a specific pressure curve and an average pressure curve for a good part. As opposed to the simple Euclidean distance between two curves (Batista et al.[10]), DTW makes it possible to compare time series in respect of temporal shifts, as shown in Figure 5 (Salvador and Chan [11], Ding et al. [12]).

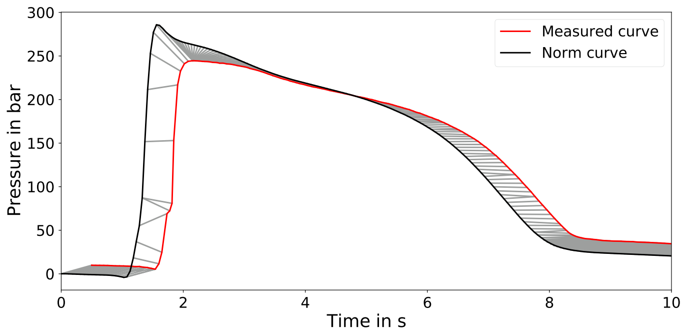

Fig. 5: Example of DTW for typical pressure curves.

In addition to the cavity pressure-based features, ten further features based on machine data were calculated. These features depend on the clamping force, hydraulic pressure, screw position, and nozzle temperature. Of the total of 25 features, 15 (including the DTW features) are based on cavity pressure; the remaining ten features based on other sensor data were manually engineered and fed into the machine learning model. Table 1 shows an overview of all the features.

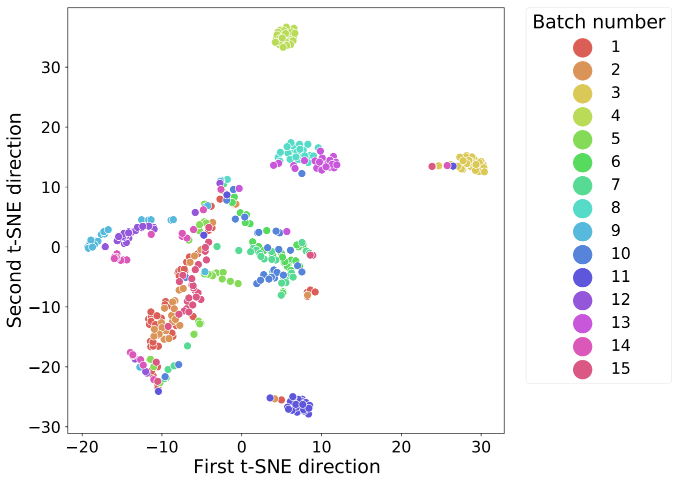

Fig. 6: A two-dimensional representation of the feature space with the help of the t-SNE method.

Figure 6 shows a two-dimensional representation of the feature space with the help of the t-distributed stochastic neighbor embedding (t-SNE) method (see: van der Maaten [13]). The presence of several batch clusters implies that the pressure curves were influenced by certain variations in the machine settings. This effect is also captured by the lower-dimensional, manually engineered features.

Feature name	Feature description
Starttime	Time of first (steep) increase of pressure
Gradient Start Switch	Gradient from start point to switchover point
Integral Start Switch	Integral from start point to switchover point
Switchover Pressure	Pressure at switchover point
Integral Switch Max	Integral from switchover point to maximum point
Max Pressure	Maximum pressure
Max Time	Time at maximum pressure
2nd Gradient at Max	Second gradient at maximum pressure
Integral Max End	Integral from maximum point to end point
Gradient Max Midend	Gradient from maximum point to midend point
Gradient Midend End	Gradient from midend point to end point
Endtime	Time at end point (curve is mostly flat afterwards)
DTW Start Max	DTW from start point to maximum point
DTW Max End	DTW from maximum point to end point
DTW	DTW overall
Clamping Force Max	Maximum clamping force
Clamping Time Max	Time at maximum clamping force
Clamping Force Integral	Overall integral of clamping force
Hydraulic Pressure Max	Maximum hydraulic pressure
Hydraulic Time Max	Time at maximum hydraulic pressure
Hydraulic Integral	Overall integral of hydraulic pressure
Screw Integral	Overall integral of screw position
Screw Switchover	Switchover position of screw
Screw Gradient	Gradient from beginning to switchover point
Nozzle Temperature Mean	Mean value of nozzle temperature curve

Table 1: Overview of all calculated features. The first section lists all features based on cavity pressure; the second section summarizes all features based on similarity and DTW; and the last section lists all features based on different machine parameters.

4. Model and cause-of-failure qualifier

There were several preconditions for the development of the quality prediction model. First: the number of anomalies in injection molding processes is usually very low, so data from bad parts cannot be utilized for model development and training. This means that direct allocation to different defect classes is not a feasible approach in this use case. To address this issue, we selected an approach based on anomaly detection which aims to identify parts whose behavior deviates from that of good parts.

Second: from the domain expert's perspective, a purely black-box machine learning model would not provide adequate trust. Decisions reached by the model would not be comprehensible to an operator. This created the requirement for some form of model explanation.

The training data for the experiment was based on 80 parts from batches which only included good parts. Data from all batches was used as test data.

Several state-of-the-art machine learning models were analyzed (local outlier factor (Breunig et al. [14]), one-class support vector machine (Schölkopf et al. [15]), and robust covariance (Rousseeuw and Van Driessen [16]). An isolation forest delivered the best performance for the relevant dataset. Isolation forests work on the principle of the decision tree algorithm (Liu et al.[17]). They isolate the outliers by randomly selecting one feature from the given set of features, and then randomly selecting a split value between the maximum and minimum values of the selected feature. This random partitioning of the features creates smaller paths in the trees for the anomalous data values, thus making it possible to distinguish them from normal data values.

With the help of confusion matrices (Figure 7), we compared the performance of this approach with the ground-truth data, which was acquired with the help of an optical setup in combination with Kistler's KiVision software [9]. The accuracy score of the model is 85%: in other words, 85% of the quality labels – either bad or good – were correctly predicted by the model. Depending on the application – for example, production of medical devices – the primary goal is to minimize the number of false positive predictions. There is usually a trade-off between the number of false positives and false negatives; in this case, however – as shown in Figure 7 – the number of false negatives was still at an acceptable level, taking account of the wide process range.

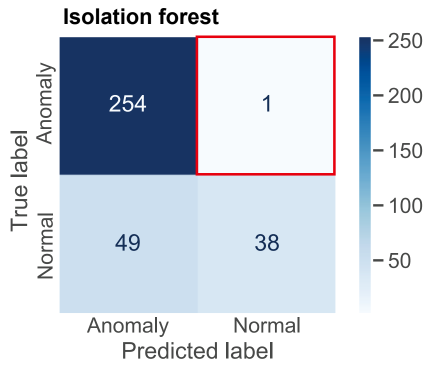

Fig. 7: Convolution matrix to show the results from the isolation forest model.

We observed that the isolation forest model had some difficulty in distinguishing normal parts from those with anomalies, especially in batches that included both good and bad parts. (These good parts were not used to train the model because the numbers of good parts in such cases were too low to achieve a reasonable train-test split.) We assume that the number of training samples used was not high enough to capture the entire distribution (variety) of good parts, and that more training data would help to continue improving the model's performance.

As mentioned, one of the development preconditions was some form of model explanation that would enhance user acceptance. A black-box model that allows no insights into the reasons behind a prediction does not create adequate trust on the part of the users.

One example of an appropriate approach that allows deeper insights into tree-based models is the extraction of feature importances. This approach opens up the possibility of identifying features that are relevant to the model outcome. Unfortunately, however, most feature importance algorithms suffer from a common disadvantage because they do not clearly show whether a particular feature produced a moderate effect across all predictions or a high effect for only a few samples.

One excellent tool which is highly suitable for identifying feature importances – and which can even circumvent the drawback just mentioned – is the SHapley Additive exPlanations (SHAP) method (Lundberg [18]). The SHAP method is a game theoretic approach that can explain the output of any tree-based machine learning model. Individual feature importances are supplied for each sample, making it possible to discover exactly which features contributed most to the model outcome for one single part. This method can be compared to a linear regression model which clearly identifies the feature/weight combination that contributed most to the model outcome. With the SHAP method, however, this information can be extracted from very complex tree-based models such as XGBoost (Chen and Guestrin [19]) or isolation forests.

Figure 8 shows one possible result of the SHAP algorithm. The top ten features of the model are listed vertically. The color in Figure 8a) indicates whether a feature value was high or low; the horizontal location shows whether the effect of that value caused a positive or negative impact on the model output, and also the strength of the impact.

A calculation of the mean value of these ratings over all data points yields the overall feature importance across the entire dataset (Figure 8b). When the SHAP-based results were compared with domain expertise, many similarities were found: in both cases, maximum pressure was rated as an important feature. Also, as expected, the gradient between the start and switchover points appeared to exert a major influence.

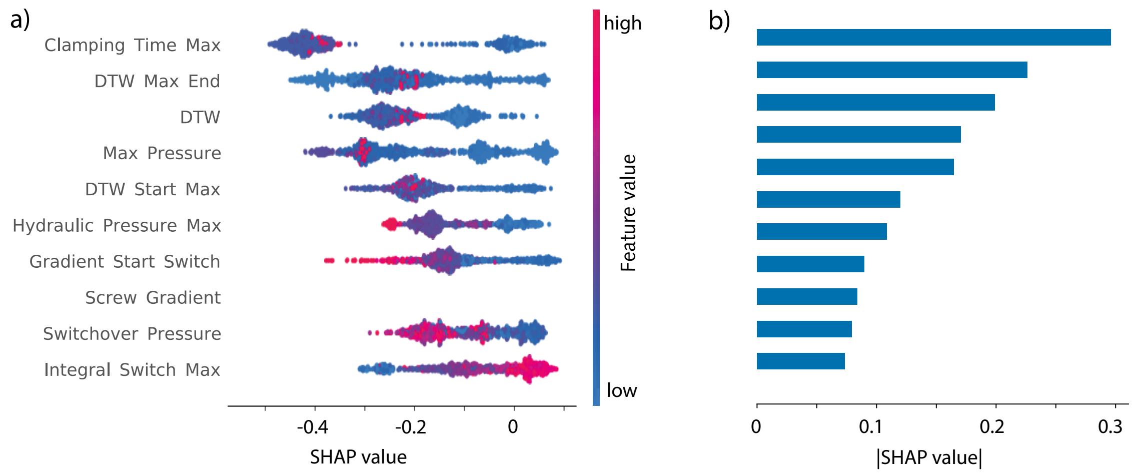

Fig. 8: Results of the SHAP algorithm for the first ten features.

By applying the SHAP method, we can now identify those features which most probably caused an anomaly for a specific sample. However, the features are still somewhat abstract, and it is not clear how a normal operator could take advantage of this information. These issues prompted the development of a "cause-of-failure qualifier" which relates anomalies to real machine settings. This is a basic requirement so that an operator can initiate appropriate countermeasures.

The method consists of two parts. First, the relationships between the extracted features and the machine settings are analyzed. One option that yields these relationships is the use of correlation coefficients between all features and all machine settings in the dataset. The calculation of all possible correlation coefficients provides a correlation table that reflects the relationships between machine settings and features (Figure 9). Blue indicates a negative correlation; red, a positive one. The more saturated the colors are, the higher the absolute correlation.

The correlation table for the machine settings was calculated with the help of data from the experiment. We used the machine settings and the features of all available data to calculate the normalized correlation factors for each feature with each machine setting. It can be seen that one or two machine settings are slightly under-represented in the features, and they do not have any strong correlations; this reduces the performance of the qualifier.

The second stage of the SHAP method yields the relationships between the anomalies and the causative features for each part. Both stages of the method were included and combined as appropriate in our cause-of-failure qualifier. In the event of an anomaly, this qualifier identifies the machine settings that are most likely related to its cause (Figure 8).

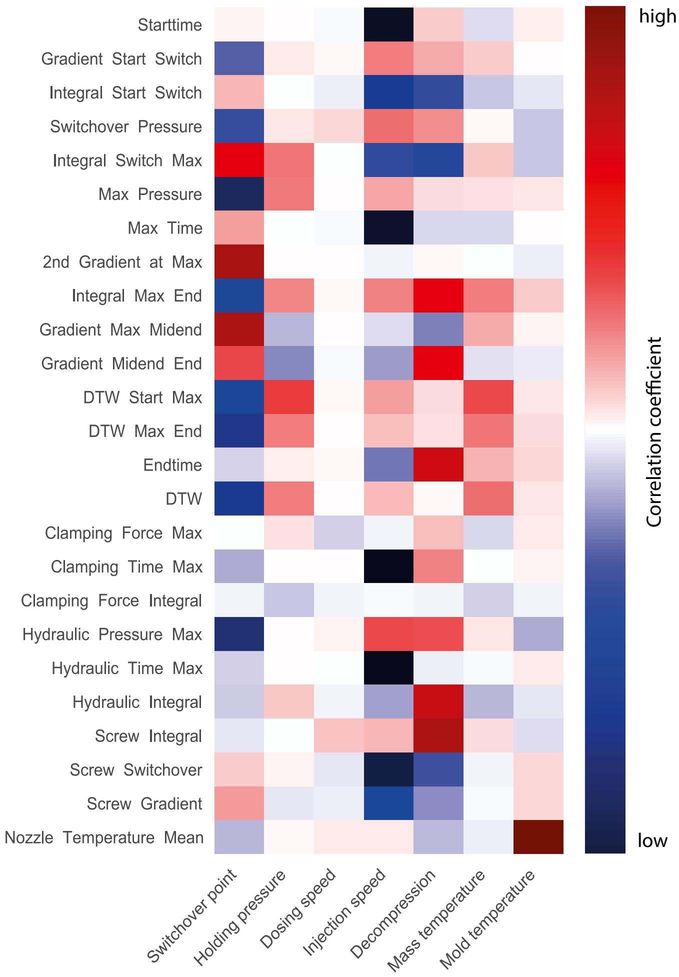

Fig. 9: Correlations between features and machine settings.

This cause-of-failure qualifier predicts the most important machine setting for each plastic part, based on the top five features. Since the actual machine settings used during production were available to us, we were able to evaluate the cause-of-failure qualifier. Figure 10 describes the working principle of the method. The results showed that the predictions were very accurate for many batches (Figure 11). In batch 3, the switchover point was reduced; in batch 9, the injection speed was reduced; and in batch 15, the mold temperature was increased. In the future, the predictions could be further improved by fine-tuning the correlation table and adding features that correlate with the under-represented machine settings. In addition, taking domain knowledge into account, the use of direct causal relationships instead of correlations is conceivable.

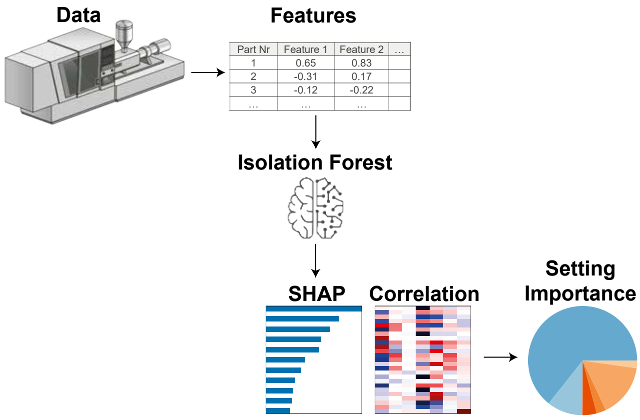

Fig. 10: Flowchart to describe the working principle of the cause-of-failure qualifier.

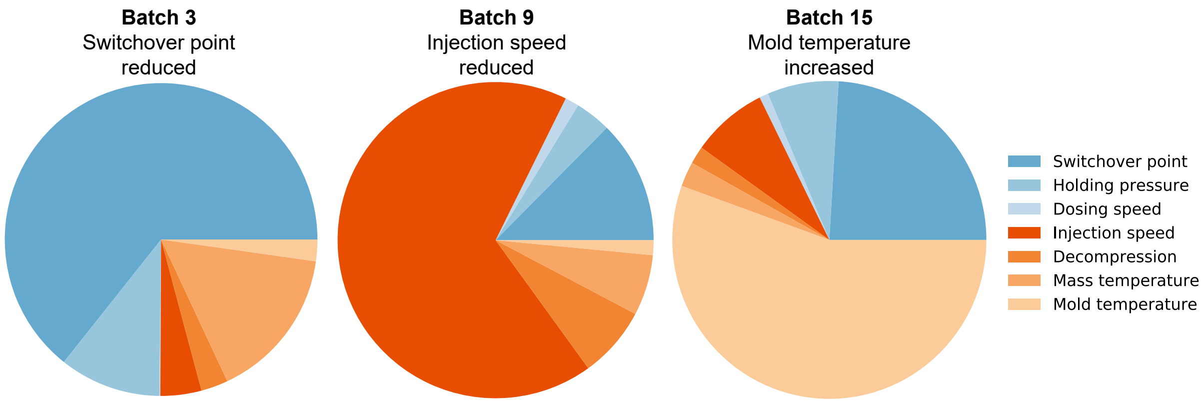

Fig. 11: The cause-of-failure qualifier outputs a pie chart showing the most likely machine setting that leads to a bad quality outcome.

5. Outlook

The objective going forward is to utilize this new technology to improve the setup of our systems. Work on improving and developing approaches such as the one described here will continue, with the goal of making the technology commercially available. Kistler is strongly committed to achieving the automation of monitoring and control processes in the injection molding world.

We believe that sensor data and intelligent use of data are the keys to the successful implementation of this approach.

References

[1] Friedrich Johannaber and Walter Michaeli: Handbuch Spritzgießen. Hanser, 2004.

[2] Martin Bichler: Qualitätssicherung beim Spritzgießen. Beuth, 2003.

[3] Johannes Wortberg: Qualitätssicherung in der Kunststoffverarbeitung. Carl Hanser, 1996.

[4] Kistler ComoNeo: www.kistler.com/comoneo

[5] R. Schiffers, K. Morik, A. Schulze Struchtrup, P.-J Honysz and J. Wortberg: Anomaly detection in injection molding process data based on unsupervised learning. Journal of Plastics Technology, 2018.

[6] H. Jung, J. Jeon, D. Choi and J.-Y. Park: Application of Machine Learning Techniques in Injection Molding Quality Prediction: Implications on Sustainable Manufacturing Industry. Sustainability, 2021.

[7] Chihun Lee, Juwon Na, Kyongho Park, Hyeonjae Yu, Jongsun Kim, Kwonil Choi, Dongyong Park, Seongjin Park, Junsuk Rho and Seungchul Lee: Development of Artificial Neural Network System to Recommend Process Conditions of Injection Molding for Various Geometries. Advanced Intelligent Systems, 2020.

[8] Kistler akvisIO. www.kistler.com

[9] Kistler KiVision. www.kistler.com

[10] Gustavo E. A. P. A. Batista, Eamonn J. Keogh, Oben M. Tataw and Vinicius M. A. Souza: CID: an efficient complexity-invariant distance for time series. Data Mining and Knowledge Discovery 28, 634–669, 2014.

[11] Stan Salvador and Philip Chan. FastDTW: Toward accurate dynamic time warping in linear time and space. Intelligent Data Analysis, 11.5, 561-580, 2007.

[12] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang and E. Keogh: Querying and mining of time series data: Experimental comparison of representations and distance measures. Proceedings of the VLDB Endowment, 1(2), 1542-1552, 2008.

[13] L.J.P. van der Maaten: Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research, 2579-2605, 2008.

[14] Markus Breunig, Hans-Peter Kriegel, Raymond T. Ng and Jörg Sander: Identifying density-based local outliers. Proc. ACM SIGMOD, 2000.

[15] B. Schölkopf, J.C. Platt, J. Shawe-Taylor, A. J. Smola and R.C. Williamson: Estimating the support of a high-dimensional distribution. Neural computation 13.7 1443-1471, 2001.

[16] P.J. Rousseeuw and K. A Van Driessen: Fast algorithm for the minimum covariance determinant estimator. Technometrics 41(3), 212, 1999.

[17] Fei Tony Liu , Kai Ming Ting and Zhi-hua Zhou: Isolation forest. Eighth IEEE International Conference on Data Mining, 2008.

[18] Scott M. Lundberg: A Unified Approach to Interpreting Model Predictions. 31st Conference on Neural Information Processing Systems, 2017.

[19] T. Chen and C. Guestrin. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 785–794, 2016.

Kistler Group
Eulachstrasse 22
8408 Winterthur
Switzerland

Tel. +41 52 224 11 11