www.czasopisma.pan.pl PAN www.journals.pan.pl

ARCHIVES of FOUNDRY ENGINEERING

ISSN (2299-2944)
Volume 2021
Issue 4/2021

103 – 109

10.24425/afe.2021.138687

15/4

Published quarterly as the organ of the Foundry Commission of the Polish Academy of Sciences

Methodology for Diagnosing the Causes of Die-Casting Defects, Based on Advanced Big Data Modelling

A. Okuniewska *, M.A. Perzyk, J. Kozłowski
Institute of Manufacturing Technologies, Warsaw University of Technology,
Narbutta 85, 02-524 Warsaw, Poland
* Corresponding author. E-mail address: aliceokuniewska@gmail.com

Received 24.09.2021; accepted in revised form 23.11.2021

Abstract

The purpose of this paper was to develop a methodology for diagnosing the causes of die-casting defects based on advanced modelling, to correctly diagnose and identify process parameters that have a significant impact on product defect generation, optimize the process parameters and rise the products' quality, thereby improving the manufacturing process efficiency. The industrial data used for modelling came from foundry being a leading manufacturer of the high-pressure die-casting production process of aluminum cylinder blocks for the world's leading automotive brands. The paper presents some aspects related to data analytics in the era of Industry 4.0. and Smart Factory concepts. The methodology includes computation tools for advanced data analysis and modelling, such as ANOVA (analysis of variance), ANN (artificial neural networks) both applied on the Statistica platform, then gradient and evolutionary optimization methods applied in MS Excel program's Solver add-in. The main features of the presented methodology are explained and presented in tables and illustrated with appropriate graphs. All opportunities and risks of implementing data-driven modelling systems in high-pressure die-casting processes have been considered.

Keywords: Fault diagnosis, Die casting, Process control, Data analytics, Application of information technology to the foundry industry

1. Introduction

Nowadays, the majority of manufacturing industries aim to increase their competitiveness by providing products in accordance with customer requirements, so in the highest quality, at the right price and available in the timeframe desired by the customer. Similarly, with foundries, where the main aspects influencing their competitiveness are the quantity and quality of the produced castings [1]. Thus, the objective is to produce castings that have no manufacturing defects, such as leakage. Previously, there was no concentration on the analysis of the pressure tightness of castings, as it was believed that the decrease in pressure tightness was directly related to casting porosity. However, from the whole list of specific properties that a casting must present, such as strength, plasticity, fatigue resistance, chemical compatibility and others, pressure tightness has been found to be probably the most common and the most important [2]. The presence of defects in castings, detected in the production process, is one of the main reasons for increasing the cost of production and therefore worsening the competitiveness of foundries [3]. For a medium-sized foundry, reducing the number of defects by only 1% leads to savings of several million PLN per year [4]. In order to maintain competitiveness and meet the increasing quality requirements, foundries have started to develop industrial data analysis, which also has a significant impact on process efficiency and optimization [5]. Additionally, in the beginning of the second decade of the XXI century, the interest in processing large data sets and the development of applications of data mining and machine learning methods in the production and manufacturing sector increased significantly. This was caused by the new concept of Industry 4.0, which also concerns the optimization of ways of working, new technologies and ways of functioning of a given traditional manufacturing company in the modern digital reality. This concept became already a reality in many fields of industry, including foundries [6]. In the foundry industry, the Smart Factory, according to its main objectives, may be able to efficiently control the process, so as to diagnose defective products assessed in real-time, thus becoming more reactive. It could therefore take action before a potential hazard occurs, in the example, attempt to diagnose a product defect even before it occurs, by new predicting ability [7]. On the basis of a diagnosed casting defect, the machine could correct certain parameters and, through this self-adjustment, prevent its occurrence. However, this task is complicated by the fact that the foundry process is widely recognized as one of the most complex in the manufacturing industry, because of many diversified processes related to the preparation of primary products and processing of final products [8].

Regarding the basis of the Industry 4.0 concept, considering the level of advancement of foundry process automatization and their data collection capabilities, it can be noted that high-pressure die-casting foundries (HPDC) probably represent the highest level [7]. Additionally, die casting is one of the main manufacturing processes in the automotive industry [9], as it requires high precision in the production of geometrically complex non-ferrous castings such as aluminum [10], [11]. It is characterized by the highest quality part production with the highest dimensional accuracy and reduced cost per part [12]. Selected castings, such as safety-critical automotive components, are 100% quality controlled in the further stages of production, as in their case the consequences of failure are severe, and the additional cost associated with quality control is high but also justified [2]. However, this process is highly dependent on individual process parameters affecting casting quality, so it is necessary to obtain their optimal combinations to minimize the formation of casting defects [10]. It should also be taken into account that the die casting process is considered to be inflexible, with low adaptability to changes introduced in the technological properties of the product designs implemented to evolve and progress the process [15]. Therefore, it is important to constantly monitor process parameters and adapt them to changing requirements based on expert knowledge, which can significantly affect product quality [13], [14].

Due to their casting properties [17], aluminum alloys are widely used in the production of components [16] such as engine blocks, structural components of vehicles: stringers, shock absorbers, tank lid frames, in the automotive industry. Foundries have to meet the increasing demands of customers in terms of safety, so they have to constantly improve the quality of their products [7]. This can be achieved by discovering the relation between process parameters and the appearance of a defect in the casting, by advanced modelling and analyzing this limited part of industrial data.

The amount of data available that describes the production process determines the possibility of discovering hidden relationships in the data. The cooperating HPDC foundry database currently collects most of the parameters of the die casting process, storing approximately sixty thousand new data sets each month. The data consists of parameters, most of which can be manipulated from the operator panels and are the most important parameters describing the die casting process. In order to improve processes, the company is working on the development of IT systems capable of registering signals coming from machines and process parameters, as well as on the creation of algorithms for castings quality prediction, castings defect detection for process control and selection of proper parameters for its optimization. This paper describes a methodology for data analysis as part of the company's objectives mentioned above.

2. Research methodology

2.1. Characteristic and preparation of high-pressure die-casting data sets

Collecting industrial data from the foundry process, i.e., the stage of creating a representation of a given phenomenon is very difficult because a standard casting production process consists of approximately a hundred parameters that can have a significant influence on the product. Discovering the dependencies between these parameters is highly complicated and almost impossible, especially in the case of parameters coming from different stages of casting production. Data related to the casting from the AlSi9Cu3(Fe) aluminum alloy cylinder blocks included various process parameters related to temperatures, velocities, pressures, alloy temperature and chemical composition etc. The first selection of relevant parameters was made by process engineers on the basis of expert knowledge and included a wide range of variables. In the real process data, there were collected over 10000 samples of examined castings and the total number of variables was 59. Another important element was to decide which values treat as dependent variables (resultant or output signals) and which as independent variables (input signals). The dependent variable in the collected data was leakage detected by the high-pressure testing.

After the initial selection of significant parameters, a data analysis methodology was developed and was divided into four stages. The first stage is data preprocessing, that is, data cleaning, covering filling in missing values, improving accuracy by detecting outliers, removing redundant data and repetitions. Created graphs of the variables flow showed a break of the curve at the boundary value, indicating two quantitatively and qualitatively different leakage ranges. It is important to note that the data has a low representation of critical values, as out of 10094 samples, only 70 samples report a casting defect, which is 0.7% of the data results and makes further analysis very difficult. This presents the important research problem of strong data imbalance, as on the one hand a large amount of data is available, but on the other hand, there is a small representation of some critical values. These rare cases can be ignored by some data-driven models. Based on graphs of the variables flow, five data sets were created, each contained a different number of observations and with a different proportion of records with smaller values of the dependent variable:

1) the first set contained all observations,
2) the second set contained 70 observations with extreme leakage values, i.e., greater than or equal to 7.5,
3) the third set contained observations of the output close to normal, with increased occurrence of elevated leakage values,
4) the fourth set containing 140 observations, half of them obtained from set 2 and the other half from set 3, from its upper range (selected to determine what influences the increase in leakage values to an undesirable level),
5) the fifth set containing 140 observations, half of them obtained from set 2 and the other half from set 3, selected randomly.

The input process parameters were discretized on the basis of visual frequency of occurrence in characteristic ranges of values (typical, increased, decreased). The second stage is the main selection of the relevant variables that are most relevant for diagnosing the formation of casting defects. This stage is important for the effective control of casting quality, especially when some parameters are forced, and also for the selection of variables that will be used in the following stages of the research. The third stage included the construction of an advanced model driven by data with the output variable containing information about a possible defect in the casting, and parameters of the casting process as input variables. This stage was important because it can be used to support the diagnosis of castings defect, and it was also the input to the next stage of research. The fourth stage consisted of multivariate optimization of process parameters for a maximum and minimum value of casting defect, to determine what exactly determines the formation of a casting defect, or lack thereof.

3. Results of data modelling and analysis

3.1. Data pre-processing

Data preprocessing is performed to improve data quality and prepare data for further analysis. This is a necessary step and requires process knowledge to correctly classify the sample as a measurement error, an outlier or a valid process-relevant value. In this step, focus was placed on suspicious data that could be erroneous values caused by sensor faults, machine error codes, miscalculated data, format errors or outliers that could provide a view of the problem. During this step, we confirmed characteristics of industrial data such as, imperfect data quality, variety of types of variable distributions, imbalance in value representation, correlations between different process parameters.

The first of the mentioned characteristics, imperfect data quality, was related to the presence of missing, outliers, duplicated, imprecise, or incorrect values, visible in the graphs as well as and hidden incorrectness such as values reasonable from the point of view of the variable value range, but which were found to be error codes of the measuring equipment. The second feature of industrial data is the variety of types of distributions of the variables such as near normal (Gauss) distribution, near Gamma distribution and others. The third feature of industrial data is the imbalance in the representation of values, described earlier. The fourth feature is the correlations between process parameters, which were found by analyzing the correlation coefficients: linear Pearson and non-parametric Spearman, for all input variables. The main finding was the large number of highly and very highly correlated input variables, which amounted to 25, or about 44% of the total number of input variables. Through the correlation analysis, the following main types of correlation sources were noticed [7]: natural correlations, for example between water temperature and water flow, these types of correlations can be replaced by a single variable, more important from the point of view of casting quality, intentional correlations, which means dependent on the human factor, should be completely eliminated, and random correlations, which means caused by the occurrence of certain values at the same time, they should be avoided or their analysis should be deeper. In the case of modelling the relations between input and output variables, dealing with data where the number of records representing critical process output values is small, some correlations between input and output variables may be coincidental. These correlations may obscure important physical dependencies in the process, due to their weakness and complexity. Values of some input variables may be intentionally entered by staff as a reaction to values of other variables or simply based on their individual experiences, leading to 'local' correlations with outputs appearing in the data. Ultimately, the input-output process model can thus easily reflect non-existent correlations. This type of analysis has led to the concept of real relationships in the process being obscured by accidental or artificially introduced relations in the data, which the models will indicate as equally valid. Identifying such "parasitic" variables is difficult. In this study, their effect was reduced by dividing the process data into five test sets with different ranges of the output variable (see 2.1).

The input and output data were then discretized for each of the five data sets. The interval ranges were determined by visually assessing, sorted according to the values, assumed by a given parameter (typical, increased and decreased values), the flow diagrams for each of the process parameters and adjusted on the basis of the histogram analysis and the graph of the number of observations in each interval, in order to detect the influence of the process parameters on the leakage value.

3.2. Significance analysis

Usually, the selection of input variables for data-driven modelling is supported by an analysis of their significance (importance), using some statistical tools such as ANOVA or Kruskal-Wallis test and advanced machine learning models [18]. The idea of the statistical approach is that if in groups of data records containing different levels of an input variable also the output values are significantly different, then this input variable should be considered significant in terms of its influence on the output data. The reverse reasoning can also be applied: if for groups of records containing different levels of an output variable, we also observe significantly different values of that input variable, then it should be considered as significant. In both cases, it is possible to identify only potentially significant variables. Furthermore, only one input variable at a particular moment is considered, and simultaneous actions of several process parameters (including synergistic or competitive ones) are not taken into account. Therefore, the initial reduction of the dimensionality of the model by removing insignificant variables should be carried out with caution. As a consequence, insignificant variables may be present in the training data rather than those that actually affect the process performance. It should be noted that if insignificant variables are discovered, their possible change could be carried out without quality consequences of the produced castings in order to reduce quality control costs.

In the present paper, significance analysis of the input variables was performed to identify the variables that are most significant and, based on these, to create a plan for advanced data-driven modelling to diagnose the causes of casting defects. The analysis was implemented on the Statistica platform. ANOVA analysis of variance in 4 variants was used to select significant variables. The first variant is a classical one-way ANOVA, used to determine the impact of characteristic levels of process parameters, i.e., inputs in the data. A significant problem was the application of this type of analysis for data sets that do not have distributions resembling normal distributions, however, according to [19] with a larger number of points it is possible to apply ANOVA. The expression for F-statistic, calculated by comparing two variances, s_1 and s_2, by dividing them (1), can be considered as a certain measure of the significance of a variable, especially when comparing the significance of different input variables.

$$F = s_1^2/s_2^2 \tag{1}$$

The second variant is the Kruskal-Wallis Test, or one-way ANOVA in the rank version, used when the dependent variable has a distribution other than the normal distribution. The third variant is the classical reversed ANOVA, which allows, in the present paper, to determine whether there is a strong variation in any of the process parameters in the high- and low-leakage groups, which may suggest its relations with the casting defect formation. The fourth variant is reversed rank ANOVA, also known as reversed Kruskal-Wallis test, which can be used regardless of the lack of distributions close to normal distributions.

The analysis of 1680 calculations resulted in two statistical parameters: in the case of ANOVA (direct and reversed) the F and p statistics were obtained, and in the case of the Kruskal-Wallis test (direct and reversed) the H and p statistics were obtained. Based on these values, three criteria of variables were created for each data set by selecting variables where the p-value was less than 0.05. The ‘basic’ criterion included variables determined according to the results of the Kruskal-Wallis test (results for set 3 are presented in Fig.1.), the ‘extended’ criterion according to the results of the ANOVA rank test, and the ‘maximum’ criterion according to the results of all the above analyses. The numbers of qualified variables are presented in Fig.2. The analysis confirmed the effectiveness of the reliability of the method, as for example for the fifth set the number of significant variables was reduced by 81% compared to the initial number.

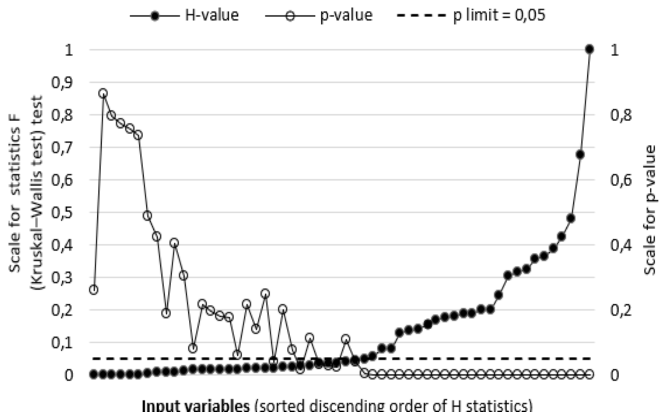

Fig. 1. Criteria of selection the statistically significant variables for diagnosing the causes of die-casting defects, based on the Kruskal–Wallis test critical p-value for the 3rd dataset

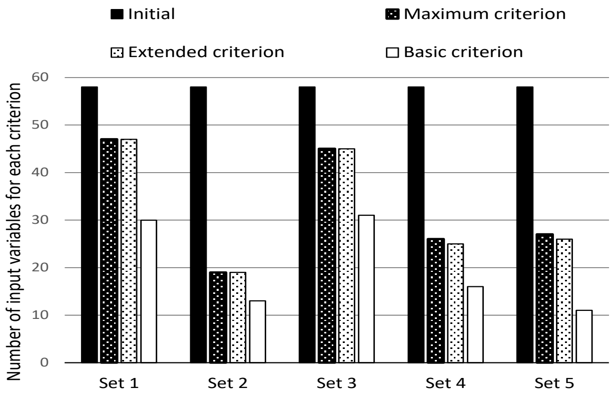

Fig. 2. Number of significant input variables for each criterion related to the initial number of variables based on the Kruskal–Wallis test (direct and reversed) and on the ANOVA analysis (direct and reversed)

3.3. Advanced data modelling

In this case, the application of artificial neural networks able to illustrate hidden and complex dependencies found in production data is reasonable [20]. When creating a modelling plan using artificial neural networks, it should be kept under consideration that determining the right number of hidden layers and the number of neurons hidden in each layer is a general problem and a kind of challenge for the artificial neural network designer.

In the present work, 590 neural models were created from MLP-type one-way networks with one hidden layer and with the number of neurons in the hidden layer varying from 7 to 23 for large datasets and from 2 to 5 for small datasets. In [21], it was concluded that there is no reason to use more than one hidden layer as it does not increase the quality of the result but only complicates the model. It is important to be aware of the main risk to the generalization ability of the network, resulting from its overfitting to less important details that are irrelevant to the solution of the problem. Testing sets are created to control this problem [22]. However, it is important to remember that process data (especially from foundries) can have data deficiencies. It is also difficult to select the best quality data set from them in order to select data for testing and validation (completely independent) sets. Therefore, these sets are often omitted from the network learning process if there is no need to create them. However, in the presented paper the quality of predictions for models using a testing set was also checked. A very important point was to test the stopping of learning by specifying a testing set equal to 0%, 10%, 15%, or 20%. Two output activation functions were used - tangent and linear. All calculations were performed using Statistica software.

Example results of modelling of the first set according to the basic criterion with the highest number of observations are presented in Fig. 3. and Fig. 4. The research leads to the following conclusions: in very few cases the network is able to learn if we separate a testing set, but sometimes the network is able to do it, as we can see in Fig.5. This raises the question which models are the best to use for further analysis, whether to use those that have better generalization ability because the network was stopped for an error growth for new data, or to use those models that gave better results but their generalization ability was poor because stopping of the learning was not applied. The most promising results were obtained for modelling in the big datasets for the third set, according to the extended criterion. In this set the root mean square error (RMSE) reached even 0.85 for tests without learning stop, the model was containing 23 neurons and a tangent and linear activation functions at the output. The most promising results were obtained for modelling in the small datasets for the fourth set, according to the maximum criterion. In this set the RMSE reached even 0.4 for tests without learning stop, the model was containing 4 neurons and tangent and tangent activation functions at the output. In the cases of a model with 0% of the values in the test set, the best results were obtained with the smallest RMSE value (so probably with over-fitting).

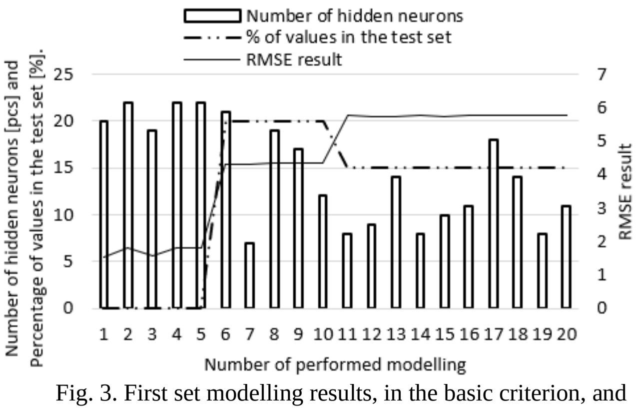

Fig. 3. First set modelling results, in the basic criterion, and tanh-tanh activation function

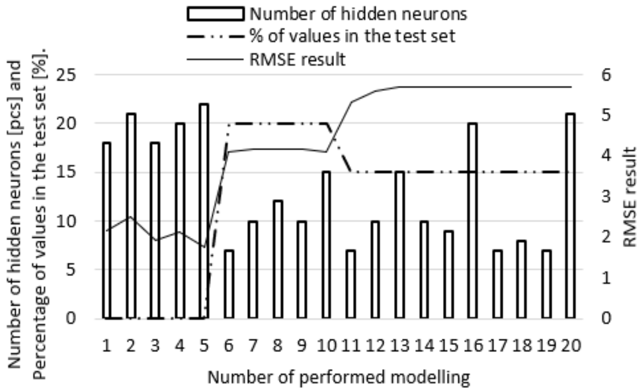

Fig. 4. First set modelling results, in the basic criterion, and tanh-lin activation function

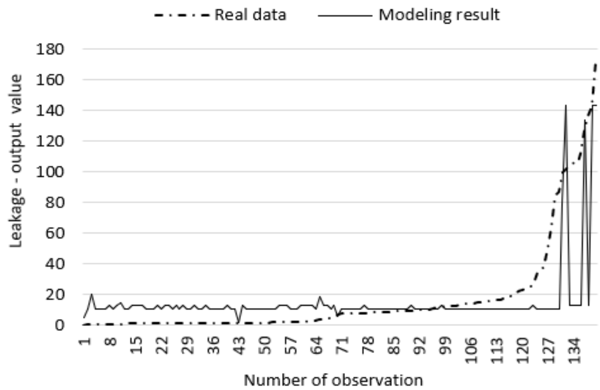

Fig. 5. Fifth set modelling results, in the basic criterion, with 2 hidden neurons and 20% values in the test set

3.4. Multidimensional optimization of process parameters

The developed strategy for querying the best models to obtain information on the causes of defects in order to determine what exactly influences the formation of a defect in casting, or its lack, included multivariate optimization of process parameters for maximum and minimum defect values, using a gradient (with multi-start) and evolutionary methods.

The optimalisation was started by repeating the modelling of artificial neural networks for the best models, with simultaneous storage of weights and programming of the model response. The optimization of the process parameters for the minimum value of leakage, so a casting without a defect, and the maximum value of leakage, so casting with a defect, was possible thanks to this. The optimalisation included both: the absolute best models but characterized by a lack of the ability to generalize (especially obtained without a test set), as well as models characterized by a higher mean square error of prediction but with the ability to generalize (containing test sets), in order to determine what determines the formation of a product defect. All optimization calculations were performed using MS Excel program's Solver add-in.

The results of the analysis of the five data sets according to the basic criterion indicate that in most cases the multidimensional optimization of process parameters is not able to illustrate exactly what influences the formation of the defect in casting or, more precisely, which values of process parameters influence the production of defective castings. However, in several cases, it was possible to obtain such an answer (fig. 6). In the graph, it can be observed that higher values of the independent variable (process parameter "multiplication delay", expressed in milliseconds) favor the leakage. Multiplication is a crucial action in the process, aimed at reduction of the castings shrinkage porosity through forced feeding of the liquid alloy into solidifying casting. This parameter defines the moment when the multiplication starts and its importance obtained from the modelling is not surprising. Most probably, if the feeding starts too late it may appear ineffective due to high fraction of the solidified metal in the casting. In [7] the same parameter was found as a statistically significant influence on the material's properties. This result implies that this process parameter certainly has an influence on the leakage occurrence and should be treated by foundry employees as critical.

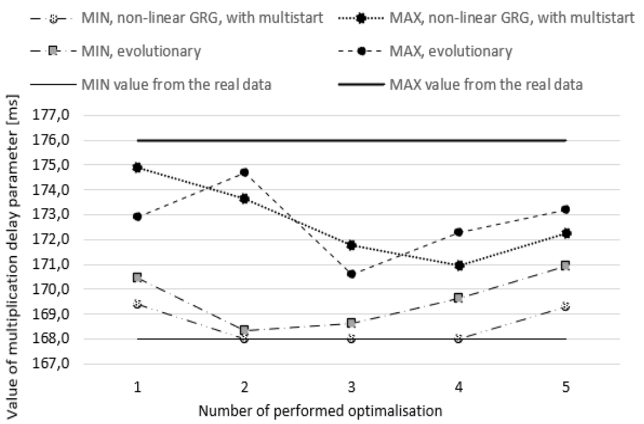

Fig. 6. Results of "multiplication delay" process variable optimization, based on results from advanced modeling (without test set, with 19 neurons in the hidden layer, tangent, and linear output activation functions) for the first set, according to the basic criterion

However, the observed dependencies must be further analyzed because the analysis was conducted for serial production, therefore not only the influence of specific parameters and their optimum values on the dependent variable should be taken into account, but also the variability of these process parameters. Fixed parameters are determined in the product development phase and change only slightly during the process, whereas dependent parameters are also determined in the product phase but are susceptible to change, for example by external conditions [7].

4. Conclusions and future work perspectives

The Key Performance Indicator (KPI) that can be found in the manufacturing industry developing according to Industry 4.0. the concept is to secure the company competitiveness through reducing number of defected products and rise the process of machine efficiency. Additionally, in manufacturing companies working according to lean methodology, emphasis is placed on the constraint triangle, which illustrates the main pillars important to the customer, that is, top product quality, short production time (lead time) and low price. A deterioration in the performance of any one pillar results in a reduction in performance in the other two. It is important, to focus on discovering the causes of product defects because an increase in production quality will have a positive impact on delivery times and on reducing costs for the final customer, thus increasing the competitiveness of the company.

Advanced data-driven soft modelling methods based on datasets from the die casting process were the basis of the proposed methodology. Data came from a real industrial process and was obtained in cooperation with the foundry. The main goal of this paper was to discover the causes of leakage in high pressure die casting process by using advanced data-driven modelling techniques to determine the relative significance of input variables.

The proposed methodology consisting of data preprocessing, significance analysis, advanced modelling, and multidimensional optimization of process parameters for minimum and maximum casting defect values, could be an element that influences the creation of an autonomous virtual operator [7]. Optionally, after further research, it may have an influence on developing the machine's ability to make decisions based on input parameters before the casting defect occurs. This would save material and reduce scrap production. Optionally, making decisions by the machine based on input parameters, after the defect has occurred. This would abandon costly quality tests in line with the Industry 4.0 concept.

In conclusion, the analyses carried out can be an important reference for high-pressure die-casting foundries. According to the article [7], the foundry practice is not only limited to a proper data analysis but in the first stage, to the description of all process parameters, which should be collected by the foundry. Therefore, the developed methodology can support the decision-making processes of determining the influence of certain process variables' values on the quality of cast parts and be a useful starting point for HPDC foundries. Further studies are still being carried out to assess more precisely the selection of relevant parameters. Further analyses are needed because generally recognized tools for handling imperfect data may be not satisfactory to apply in the foundry industry to predict the quality of the casting. Whereas the modelling results presented in this article may contribute to the next stages of research.

References

[1] Paturi, R.U.M., Cheruki S. (2020). Application, and performance of machine learning techniques in manufacturing sector from the past two decades: *A review. Materials Today: Proceedings.* 38(5), 2392-2401. DOI: https://doi.org/10.1016/j.matpr.2020.07.209

[2] Campbell, J. (2003). *Castings, the new metallurgy of cast materials, second edition*. Elsevier Science Ltd., ISBN: 9780750647908, 307-312.

[3] Kochański, A.W. & Perzyk, M. (2002). Identification of causes of porosity defects in steel castings with the use of artificial neural networks. *Archives of Foundry*. 2(5), 87-92. ISSN 1642-5308.

[4] Falęcki, Z. (1997). *Analysis of casting defects*. Kraków: AGH Publishers.

[5] Kim, J., Kim, J., Lee, J. (2020). Die-Casting defect prediction and diagnosis system using process condition data. *Procedia Manufacturing.* 51, 359-364. DOI: 10.1016/j.promfg.2020.10.051.

[6] Lewis, M. (2018). Seeing through the Cloud of Industry 4.0. In 73rd WFC, 23-27, (pp. 519-520). Krakow, Poland: Polish Foundrymen's Association.

[7] Perzyk, M., Dybowski, B. & Kozłowski, J. (2019). Introducing advanced data analytics in perspective of industry 4.0. in die casting foundry. *Archives of Foundry Engineering*. 19(1), 53-57.

[8] Perzyk, M., Kozłowski, J. & Wisłocki, M., (2013). Advanced methods of foundry processes control. *Archives of Metallurgy and Materials*. 58(3), 899-902. DOI: 10.2478/amm-2013-0096

[9] Makhlouf, M.M., Apelian, D. & Wang, L. (1998). Microstructures and properties of aluminum die casting alloys. *North American Die Casting.* https://doi.org/10.2172/751030

[10] Tariq, S., Tariq, A., Masud, M. & Rehman, Z. (2021). Minimizing the casting defects in high pressure die casting using taguchi analysis. *Scientia Iranica*. DOI: 10.24200/sci.2021.56545.4779.

[11] Fracchia, E., Lombardo, S., & Rosso, M. (2018). Case study of a functionally graded aluminum part. *Applied Sciences*. 8(7), 1113.

[12] Dargusch, M.S., Dour, G., Schauer, N., Dinnis, C.M. & Savage, G. (2006). The influence of pressure during solidification of high pressure die cast aluminium telecommunications components. *Journal of Materials Processing Technology*. 180(1-3), 37-43.

[13] Bonollo, F., Gramegna, N., Timelli, G. High pressure die-casting: contradictions and challenges. *JOM: the journal of the Minerals, Metals & Materials Society*. 67(5), 901-908. DOI: 10.1007/s11837-015-1333-8.

[14] Adamane, A.R., Arnberg, L., Fiorese, E., Timelli, G., Bonollo, F. (2015). Influence of injection parameters on the porosity and tensile properties of high-pressure die cast Al-Si alloys: A Review. *International Journal of Meterials.* 9(1), 43-53.

[15] Gramegna, N. & Bonollo, F. (2016). HPDC foundry competitiveness based on smart Control and Cognitive system in Al-alloy products. *La Metallurgia Italiana*. 6, 21-24.

[16] Łuszczak, M. & Dańko, R. (2013). State the issues in the casting of large structural castings in aluminium alloys. *Archives of Foundry Engineering*. 13(3), 113-116. ISSN (1897-3310).

[17] Davis, J.R. (1990). ASM handbook. ASM, Metals Park, OH. 123-151, 166-16.

[18] Perzyk, M., Biernacki, R. & Kozłowski, J. (2008). Data mining in manufacturing: significance analysis of process parameters. *Journal of Engineering Manufacture*. 222(11), 1503-1516. DOI: 10.1243/09544054JEM1182.

[19] Koronacki, J., Mielniczuk J. *Statistics for students of technical and natural sciences.* WNT (209-210, 458). (in Polish).

[20] Okuniewska, A., Methods review of advanced data analysis tools, in process control and diagnostics. Piech K. (red.) Issues Actually Addressed by Young Scientists, 17, 2020, Krakow, Poland, Creativetime, 328 p., ISBN 978-83-63058-97-5

[21] Lawrence, S., Giles, C.L., Tsoi, A.C. (1996). What size neural network gives optimal generalization? Convergence Properties of Backpropagation. Technical Report UMIACS-TR-96-22 and CS-TR-3617. Institute for Advanced Computer Studies, University of Maryland. College Park, MD 20742.

[22] Tadeusiewicz, R. (2005). First electronic brain model.

[23] https://natemat.pl/blogi/ryszardtadeusiewicz/129195,pierwszy-dzialajacy-techniczny-model-mozgu