Engineering Applications of Artificial Intelligence 122 (2023) 106149

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

A novel approach for quality control of automated production lines working under highly inconsistent conditions

F.M. Bono *, L. Radicioni, S. Cinquemani

Mechanical department, Politecnico di Milano, Via La Masa 1, Milano, 20156, Italy

ARTICLE INFO

Keywords:
Artificial Intelligence
Machine learning
Neural networks
Quality control
Process control
Automation industry

ABSTRACT

When addressing product quality standards in manufacturing lines, a critical issue is the identification of the parameters that define the quality of the final product and their tracking. The problem of process control under inconsistent working condition of an automatic machinery, i.e. when some parameters are highly variable, is still quite unexplored in literature. This objective becomes even more challenging when the most important process variables are not directly measurable. This paper demonstrates that it is possible to achieve quality control by coupling a soft sensor, whose predictive model is a neural network, with an anomaly detector. The methodology has been applied to automatic machinery placed in a manufacturing line, where high variability in production rate has an important effect on the measured physical variables. This makes automated and accurate quality control difficult, due to the fact that in this test case the data collected are accelerometers signals, extremely sensible to variation in machine productivity by definition. It is shown that this approach outperforms many other classification methods (Support Vector Machines, Ensemble Bagged Tree, Discriminant Analysis, K-nearest neighbours and the direct application of a Neural Network) proposed in the past, achieving satisfactory results evaluated on the basis of four metrics (Accuracy, precision, recall and F_1-score), even if anomalous data have been collected in a limited number of machine's working points. In particular, an accuracy over 92% has been reached also for production rates where only nominal conditions are collected. This procedure exceeds the direct training of a neural network (accuracy of 57.6% at new production rates), as well as the application of shallow methods based on the extraction of dimensionless features (around 35% in accuracy at new production rates).

1. Introduction

The attention to quality control strategies has seen a substantial rise in the past few decades (Mitra, 2016). Many factors have contributed to this phenomenon: the elevated production rates of modern manufacturing lines, the tense competition in the global market, the need to guarantee a sustainable process, the high price of energy and raw materials are only few examples which explain why industries must cut waste in production more than ever (Albers et al., 2016). Hence, the urgency in the search of innovative and efficient systems capable of detection in anomalous working conditions which can lead to production of low quality outcomes. Ideally, those systems should work in real time (so that they can immediately warn when the line is producing poor quality outputs), while being cheap and easily replaceable.

The identification of the *critical process variables* (CPVs), also called *key quality variables* (Yuan et al., 2020), i.e. the group of physical quantities that must be tracked to assure the compliance of the final output, is, in many cases, everything but a banal task. This is because in the manufacturing industry it is not always clear which are the reasons behind defects and dedicated investigations are necessary most of the time (Murua et al., 2020). Consider Fig. 1. If the reasons behind the output products' defects are clear, or the prior knowledge over the process is sufficient to define the CPVs, it is quite easy to take action at the correct workstation to adjust the product quality. If that is not the case, further investigation must be performed. Subsequently, the most important CPVs to assure product quality should be identified.

When they have been finally assessed, a sufficient number of sensors must be employed to make the tracking of the CPVs possible. Not all of the CPVs are always accessible for measurement. Sometimes the machineries do not offer enough space to fit a sensor in, or the environment might be particularly harsh, as it might happen in chemical plants. In all those cases, estimating the CPV in other ways could be helpful (Zhu et al., 2020; Rege et al., 2002).

1.1. Problem statement

In an industrial context, a system able to compute a variable by means of the available measurements and a predictive model is called

* Corresponding author.
E-mail address: francescomorgan.bono@polimi.it (F.M. Bono).

https://doi.org/10.1016/j.engappai.2023.106149
Received 18 July 2022; Received in revised form 17 January 2023; Accepted 10 March 2023
Available online 16 March 2023

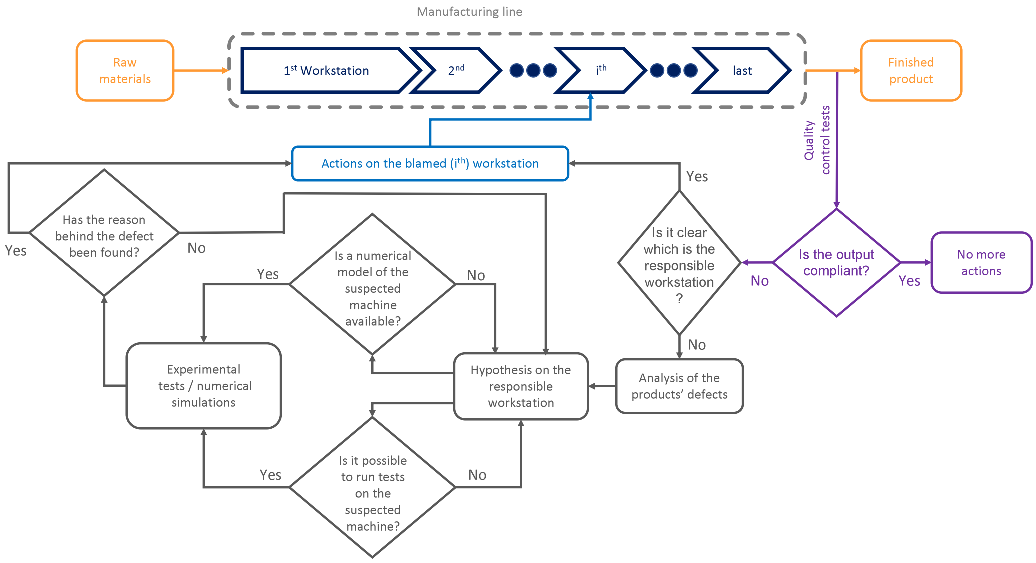

Fig. 1. Recommended procedure for the identification of the critical process parameters, to correct the responsible workstation.

soft sensor. This expression is derived from the words *software* and *sensor*. After all, these systems are nothing more than algorithms that yield data same as their physical equivalents (Kadlec et al., 2009).

To select the most suitable predictive model, it is crucial to distinguish between two main classes of soft-sensors: *model-driven* and *data-driven*. The model-driven ones rely on first-principles models (Hsiao et al., 2021), hence assuming that the phenomena underling the process are stable and known, which is usually not the case. In all those situations where the relationships among variables are complex or unclear, data-driven models are a valuable option. As a drawback, they require the collection of information from the machinery during production (Kano and Nakagawa, 2008). Furthermore, it is worth mentioning that examples of soft sensors based on numerical models do exist in literature (Guo et al., 2019).

In general, the low-cost associated to soft sensors compared to their hardware counterparts, the simplicity of employment on top of the already existing systems and the capacity of estimating values in real time, whereas some physical sensors would require long time before yielding results, make their usage appealing in many fields (Fortuna et al., 2005).

The presented research deals with data-driven soft sensors, based on *machine learning* (ML) algorithms. A ML model can be trained to recognize peculiar characteristics in the data, such as the most frequent non-nominal working conditions or, in general, situations deviating from standards. In industrial field, gathering a training set can be particularly troublesome when the working conditions of the process are highly variable, which means that many working points should be explored. In this context, a *working point* can be defined as a unique combination of all the parameters that affect the system. Clearly, the latter model behaves differently in distinct working points, determining significant changes in the overall response. The associated risk is that the system could not be able to generalize for new working points not present in the training set.

1.2. State of the art

In data-driven approaches, this problem is usually addressed as the duality between interpolation and extrapolation. The first refers to the prediction of values that fall inside the training set, whereas the second concerns the predictions outside the boundary of the dataset (McCartney et al., 2020).

Regression and interpolation approaches in a changing environment are first addressed by Worden in Worden et al. (2002). The topic of systems under inconsistent working conditions has been addressed in literature mainly in the problem of fault diagnosis of bearing and wind turbines. In fact, the vibration introduced at different rotating speeds of those systems is highly unpredictable, and the modelling of such a phenomenon has been addressed by an active group of researchers. However, none of them focused on the possible application on quality control. In McBain and Timusk (2009) showed how is possible to interpolate among statistical distribution parameters to adapt the decision boundaries at different rotational speeds. In Sohaib and Kim (2020) proposed an approach based on a bispectrum analysis followed by a convolutional Neural Network (CNN). In Hasan et al. (2020) further investigated this approach by implementing a multitask-aided transfer learning, to improve the robustness of the model. In Chen et al. (2019), a recurrent neural network combined with an attention mechanism is proposed by Chen et al. to detect and classify a fault in wind turbine operating at variable speed. In Li et al. (2020) a model-based approach coupled with a power spectral density allowed to extract faulty characteristics for variable speed wind turbines. Looking at industrial applications, it is worth mentioning some methods based on the extraction of dimensionless parameters, which are claimed to be rather robust against speed variation, as in Lei et al. (2009), Zhang et al. (2013), Qin et al. (2018). In particular, it will be shown in Section 4 that the use of dimensionless parameters lead to poor results in the case under study.

1.3. Paper contributions

In this paper, five main contributions can be identified compared to state of the art:

- The research focuses on quality control of industrial machinery. In literature, most of the papers are focused on fault detection. Therefore, it is essential to fit these techniques in the field of quality control.
- The paper addresses a relevant lack in literature, that is, quality control under highly inconsistent working conditions.
- The time domain series are directly fed into the convolutional and recurrent neural network, without any specific feature extraction. However, to so, a general approach based on resizing is defined, to account for variations in the observations size due to the inconsistent working conditions.
- To make possible the application of this technique in highly inconsistent working conditions, the definition of general rules

to collect the training set and to develop algorithm architecture are crucial. The proposed method for the experimental campaign, conducted at meaningful working points, allows to limit the overall number of experiments conducted.

- The proposed methodology can work for whichever combination of input data, regardless of the data are one or more among aggregated variables, signals, images per manufacturing cycle.

This paper copes with this issue in manufacturing by introducing a novel methodology for the collection of the training set and the update of the soft sensor, as well as a predefined structure for a *neural network*. This methodology has proven to be more robust compared to similar systems found in literature and it opens new scenarios related to the use of deep learning in process monitoring for quality control purposes. This novel approach is applied to an automatic machinery test case, to compare the performances with respect to the other methods in literature.

The paper is structured as follows. The methodology suggested for the collection of the training set (by monitoring of the machinery in action at specific *working points*), is presented in Section 2, where a description of the recommended neural network's structure and the process for uploading of the soft sensor are included. Section 3 introduces a machinery which the methodology has been applied to. This should allow the reader to implement the same methodology in other contexts. In Section 4, the results obtained are presented and evaluated. The paper concludes with a few comments in Section 5.

2. Proposed methodology

This paper proposes a methodology for the building of a soft sensor for quality control purposes, in case the production machinery to be monitored is characterized by highly inconsistent working conditions.

Consider the common situation where an automated quality control system is required to evaluate the quality of each product so to reduce the volume of non-compliant ones manufactured. Eventually, the *defect analysis* showed that there is a correlation between the issues and a specific CPV, that cannot be directly measured and the construction of a dedicated soft sensor is required.

This task becomes rather complex in case the machinery works in highly *inconsistent* conditions, for instance when there are sensible variations in a machine's production rates. As highlighted in Section 1, scientific research that dealt with these problems in an automated production line were based on the extraction of some dimensionless parameters (Lei et al., 2009; Zhang et al., 2013; Qin et al., 2018). Those should be suitable for being applied in this context. However, in Section 4, the results obtained after their evaluation showed they are not applicable in the case study.

The construction of a soft sensor starts with the selection of available measurement sensors that will produce the input data. Sometimes machinery come with sensors already mounted. If the information already available are evidently not sufficient, the adoption of more sensors should be considered. In general, accelerometers or other sensors able to acquire the dynamics of a system represent a good solution. It has been proven that vibrational signals collected during a production process carry relevant information about the process itself (Bruwer et al., 2007).

The selection of the sensors to use is not a straightforward task. This could be modelled as an optimization problem; however, the enormous number of variables (the possible sensors, their position, their number etc.) to consider and their effect on machine learning models make a rigorous approach to the problem impossible. Therefore, to accomplish this task, one must rely on previous experience, or work on a numerical model of the machine, if available (Bono et al., 2022a).

2.1. Training set collection

Since the value of the CPVs must be tracked, it is required to collect dataset from a machine to make the training of the soft sensor possible. However, considering the entire *working region* of the machine is in most of the cases unfeasible. One should then contemplate a sub-group of working points to evaluate. For the sake of clarity, a working point is a unique combination of *input parameters*, for example the system's speed, pressure, products' weight etc. Consider Fig. 2, where the working region of a generic manufacturing machine regulated by two input parameters is shown. x_i is the ith input parameter, whereas x_{iL} and X_{iU} are the lower and the upper boundary values for that variable, respectively. If the system counts N parameters, it is recommended to consider all the vertices of the N-dimensional hyper-cube defined by all the upper and lower limits, hence a number of $W = 2^N$ working points. This is a considerable reduction in the number of different working conditions that need to be acquired.

In this work, a design of experiments approach (for example the one proposed in Zăvoianu et al., 2021) is not useful. In fact, in industrial cases where complex measurements (such as accelerations) are collected from a machinery, with the objective of evaluating its behaviour in producing compliant outputs, different unpredictable factors can modify the system response. Therefore, implementing a design of experiment approach would require the evaluation of a higher number of working points. However, evaluating only the vertices of the hypercube is effective in constructing the quality control system, as shown in Section 4.

Before training the soft sensor, it is important to assign target values of the CPV to each working point of the machine. During this procedure, collecting data related to anomalous working conditions might be non trivial, but it is crucial for the training of a supervised mechanism. In the case of a classification problem, these values can be assigned through manual labelling by an expert that can evaluate the quality with a visual inspection, instead, in case of a regression problem, the CPV can be measured with specific equipment. For this reason, this operation can become very costly, but it is fundamental for the construction of the entire quality control system. In most manufacturing lines, the production of a non-compliant output is a time-dependent phenomenon that arises after a limited number of production cycles. In this case, the collection of anomalous working conditions becomes easier.

2.2. Model for the soft sensor

The quality control system that this paper proposes is suitable when a soft sensor for the CPV estimation is required. In an industrial context, the physical modelling of machinery can be relatively time consuming, whereas they might not be always able to interpret the reasons behind a defect. Besides, data-driven models are appealing for their deployment readiness, although the collection of the training set is needed, with all the problems associated to produce experiments and data collection. Deep learning models, in the form of *neural networks*, are among the most flexible data-driven approaches. They can take, as input, objects from different feature spaces, elaborate separately the information and then merge the knowledge so as to get a unique output. Because of this, the soft sensor recommended in this proposed methodology for the estimation of the CPV is a neural network. In the most general case, this neural network must feature three main input branches (Fig. 3), according to the type of the available data. Moreover, different inputs require a specific type of preprocessing, i.e. all the steps necessary to bring the raw data in a desired form. Two types of preprocessing are essential here:

- **Scaling**: this process is mainly described by an operation that transforms all the values according to a defined rule, so that all scaled data have the same degree of influence on the training process. In this way the method is immune to the choice of

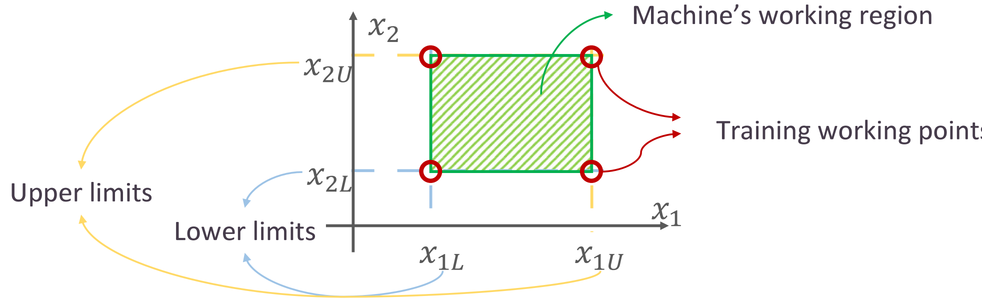

Fig. 2. Example of the machine's working region and the suggested working points for the training set.

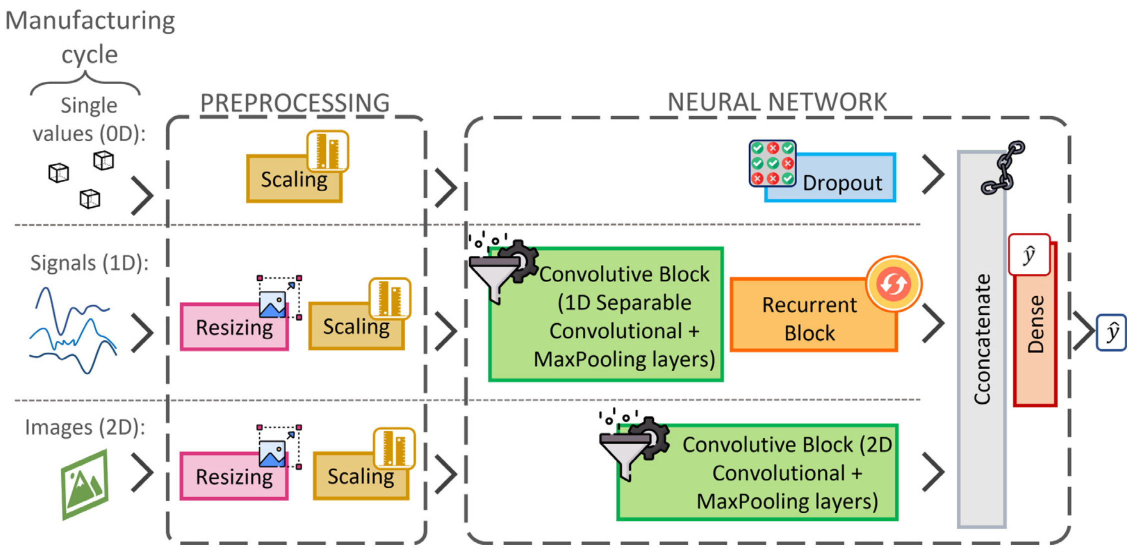

Fig. 3. Layout of a soft sensor: preprocessing for the possible input data and structure of the neural network.

measurement unit (Kosti et al., 2012; Pal and Sudeep, 2016). Many approaches to perform a scaling operation can be found in literature.

- **Resizing**: since neural networks accept as input only objects of pre-defined shape, in many cases an adjustment of the input size is required. Resizing is one way to carry out this operation. It consists of a linear interpolation in case of 1D data, a bi-linear interpolation in case of 2D objects, and so on.

The input data are treated separately, according to their dimension, in both preprocessing and in the types of layers dedicated to their study. Evidently, the network might count only two or even one single branch, if, for instance, there are only 1 dimensional data (i.e. signals). The three branches are:

1. The *first branch* takes as inputs all the *single values*, that is, all the values that are considered constant for a single manufacturing cycle, as well as all the machinery's *input parameters* $(x_1, \ldots, x_N)$ that define the working points, as described in Section 2.1. All these instances are all 0-dimensional by definition. Every 0-D input must be then independently scaled. Eventually, a dropout layer acts as a *regularizer*. It helps the network in finding general patterns, by randomly dropping a portion of the input features. (Srivastava et al., 2014)
2. The *second branch* elaborates the signals if they are present. Since at different working points the number of samples might vary, it is suggested to resize them to a common length (Hashemi, 2019) before applying a scaling method. Then, the signals are elaborated by a combination of convolutional and recurrent layers, a structure firstly proposed for natural language processing. Recurrent layers are heavily prone to noise, while convolutional layers present de-noising properties, so that the unique local patterns in the inputs have less influence on the network's performances. (Sainath et al., 2015). The convolutive block can be obtained by alternating 1D (Separable) Convolutional layers and maxpooling layers. The recurrent block can be built by using one (or more) *Gated Recurrent Unit* (GRU) or *Long short-term memory* (LSTM) layers. (Yu et al., 2019; Cho et al., 2014)
3. In the case images or other 2D objects are available, a *third branch* made of 2D-convolutional layers should be implemented. Resizing and scaling must be applied here as well.

The output $\hat{y}$ represents the CPV to predict. If its value is a single or more continuous numbers, the problem will be a *regression*. If the CPV can assume only a defined set of classes, the problem will be a *classification*. Regardless of the CPV's nature, the structure of the neural network does not change, which is a significant advantage compared to other machine learning methods. It is also evident that the nature of the network can be adjusted according to the data available and the task required, without modifying the described methodology.

2.3. Training of the soft sensor

Once the model has been designed, data corresponding to the manufacturing cycles must be collected as explained in Section 2.1. Each manufacturing cycle, namely the measurements coming from the production of one single piece manufactured, is associated to an observation. Then the data must be partitioned into *training*, *validation* and *test* sets. Finally, the model is fitted onto the training set and its performance are evaluated through the validation set at every training epoch. The model performance can be further enhanced by

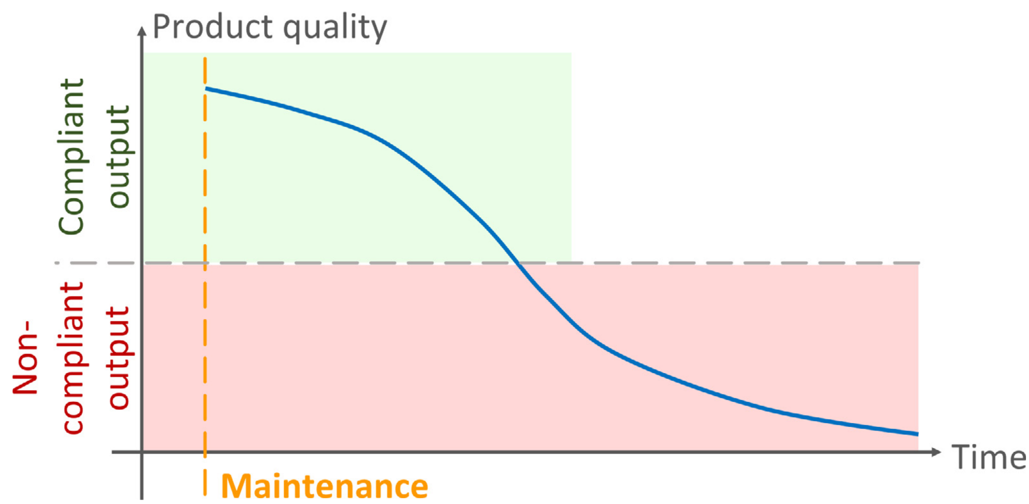

Fig. 4. Graphical representation of the product quality Vs time elapsed after maintenance.

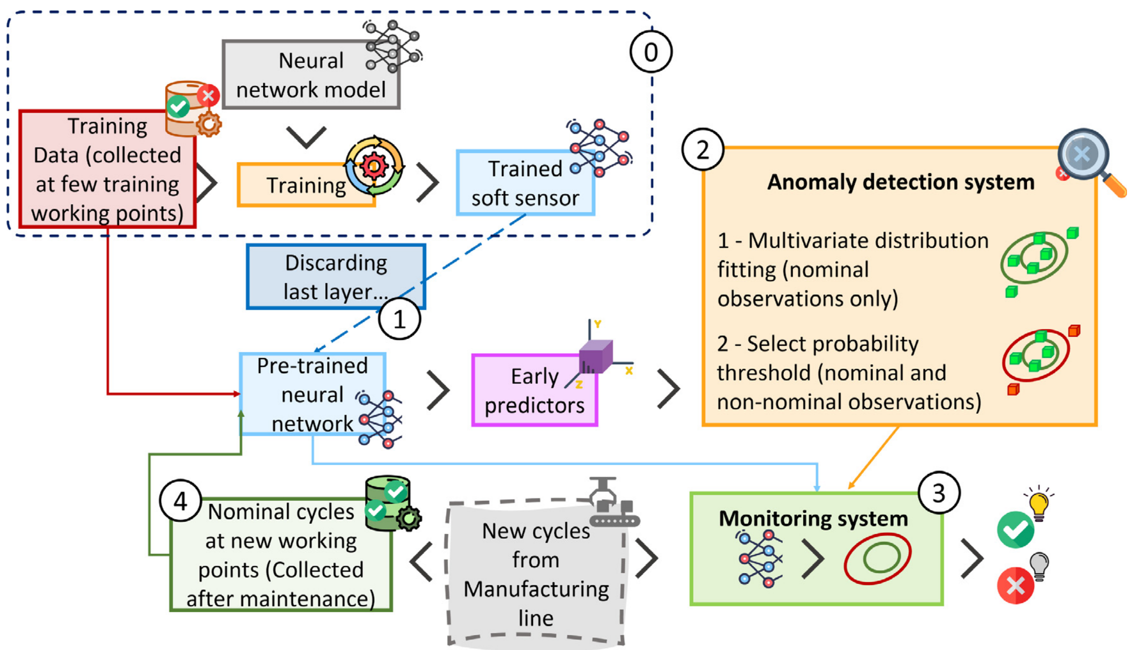

Fig. 5. Arrangement of the quality control system. Starting from the soft sensor, the monitoring system is created by coupling the pre-trained neural network and the anomaly detection system.

tuning some of the model's hyper-parameters, by using, for example, a Bayesian optimization, where the objective function is the validation loss. Finally, the evaluation of the loss on the *test* set is crucial to avoid *overfitting* on both training and validation sets.

Eventually, with reference to Fig. 2, one can empirically verify that the performance of the soft sensor worsens rapidly by operating far from the training points, which is a clear indication that the classifier struggles in extrapolation. To cope with this drawback, it is recommended to abandon the idea of achieving an accurate prediction of the CPV in the new working points and to construct a quality control system on top of the soft sensor, as described in the upcoming sections.

2.4. Acquisition of new nominal points

While it is evidently not possible to collect anomalies at every possible working point, during the regular functioning of the machine one can expect the production of compliant outputs for a certain number of manufacturing cycles. Only after a time the output will become non-compliant, due to the wear of components, loosening of tolerances etc.

This condition is synthetically explained in Fig. 4. This means that, after maintenance, it should in theory be possible to collect a certain number of *nominal* manufacturing cycles in totally new working points, which are the ones of the machine during the regular production. If this assumption is accepted, the intelligent quality control system may be updated with new data after every successful maintenance operation. This second acquisition of data is considered as not mandatory and it is made to strengthen the algorithm with new nominal data that comes from normal functioning of the machine, helping the algorithm in generalizing the quality control for more working points.

2.5. Quality control set up

To achieve quality control for new working points as well, the capabilities of neural networks and the flexibility of a multivariate Gaussian distribution are coupled by means of transfer learning. The method is summarized in Fig. 5:

1. **Transfer learning:** The last layer of the trained neural network is removed. Thus, once a new manufacturing cycle enters the network, it is not converted into a number (regression) or a category (classification), but it exits as an object characterized by meaningful features. The observations mapped into the reduced space generated by the neural network are called "early predictors" from now on. The dimensions of this new reduced space depend on the size and the type of the neural network's penultimate layer.
2. **Anomaly detection system:** All the manufacturing cycles available (nominal and non-nominal) pass through the pre-trained neural network and exit as early predictors. By applying the anomaly detection method, a multivariate Gaussian distribution is fitted onto the nominal early predictors, while the non-nominal early predictors (anomalies) are used to select a proper *threshold probability*.
3. **Monitoring procedure:** The coupling of the pre-trained neural network and the anomaly detector make up the final *quality control system*. In this way, every new manufacturing cycle will be flagged as nominal or non-nominal.
4. **System updating:** The collection of a certain amount of manufacturing cycles after every maintenance of the machine (the exact number will depend on the typology of the process to be monitored), regarded as *nominal*, will be used for the update of the multivariate Gaussian distribution, whereas the weights of the pre-trained neural network are frozen. In this way, an increasing number of the machine's working points will be accounted for.

The quality monitoring system built in this way can manage a wide variety of input data, while being robust in therms of significant variation of the operating conditions. To evaluate the performance of this arrangement, it is applied on a real test case described in the next section.

3. Test case description

Since the objective of the work is to realize a quality control system that can be applied in an automated production line, the approach is tested on a real case, where the estimation of products' quality is needed. In this context, before implementing an automatic control system, the output quality was evaluated with spot-checks three times per day, by means of a destructive process. When a product is found non-compliant, all the products manufactured in two successive spot-checks are discarded. This procedure was time-consuming and led to the waste of many manufactured pieces, generating relevant economical losses.

The machine considered is a stamping system for thin metal sheets. It consists of a rotating machine (from now on "Wheel"), with an horizontal axis of rotation, equipped with ten identical subsystems named "stamping units". For the sake of clarity, a CAD model of the entire system is reported in Fig. 6.

The stamping unit is characterized by four bodies (Ⓐ,Ⓑ,Ⓒ and Ⓓ) that move across a plane parallel to the wheel's axis of rotation, and a "fixed" body (called *die*) (Ⓔ), rigidly linked to the wheel. The *Punch* Ⓓ presses the metal sheet against body Ⓔ to shape the metal sheet into the desired form.

The four moving bodies are connected in the following way: body Ⓐ is connected with the wheel by means of bearing ①; bearing ② connects Ⓑ to body Ⓐ; body Ⓒ is an hydraulic cylinder mounted on body Ⓐ and it acts a force on body Ⓑ; the punch Ⓓ is linked with body Ⓑ by means of a spherical joint.

The punch can perform little rotations in all the directions during machine operation, in order to prevent damage of the entire structure in case of irregular alignments. However, during maintenance activities, its position is tightly regulated by means of six bolts with rubber tips ③.

Finally, the system is driven by a cam which imposes a motion law to the entire machine, allowing the stamping unit to pass from *open* configuration (represented in Fig. 7), where the metal sheet is loaded by means of a conveyor belt, to *close* configuration (Fig. 8), that corresponds to the condition in which the punch presses the material over the die. Once the operation has been carried out, the unit returns to the open position to let the unloading of the finished product.

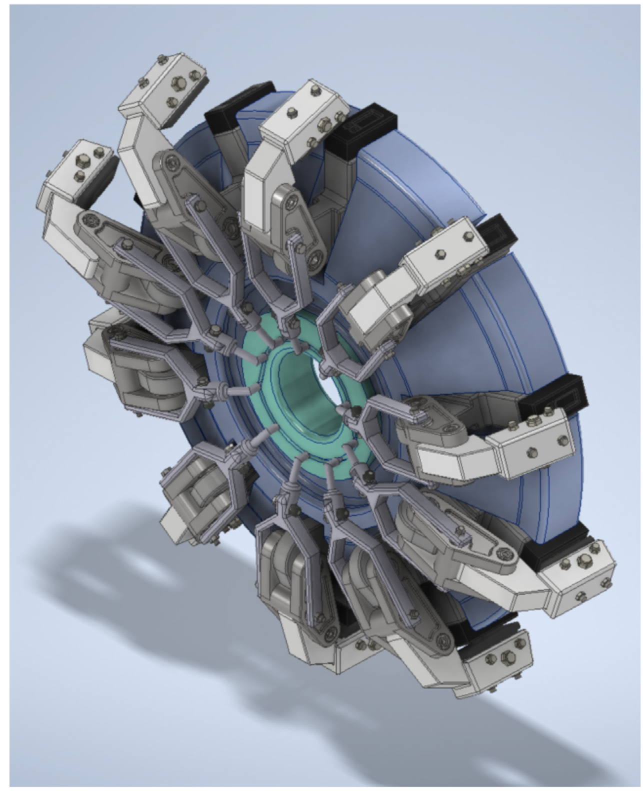

Fig. 6. Overview of machine considered in the test case.

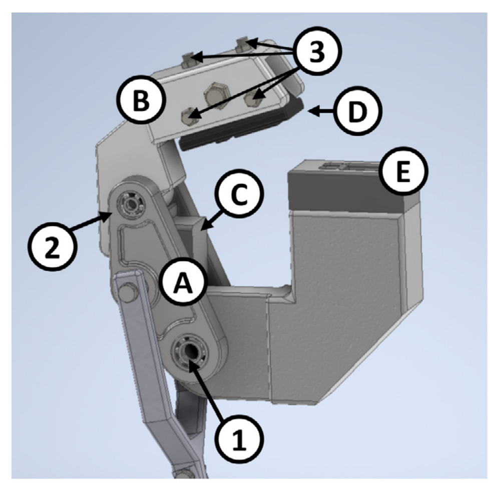

Fig. 7. Stamping unit in open configuration.

From previous analysis (both numerical and experimental) on this machinery, it was understood that the primary cause of non-compliant products is the alignment between the two surfaces of the punch Ⓓ and the die Ⓔ. This quantity represents the CPV of the process, and it was noted that a direct measurement through sensors was not feasible. A soft sensor is consequently needed to detect the alignment condition of the bodies, so as to verify the presence of non-nominal working conditions.

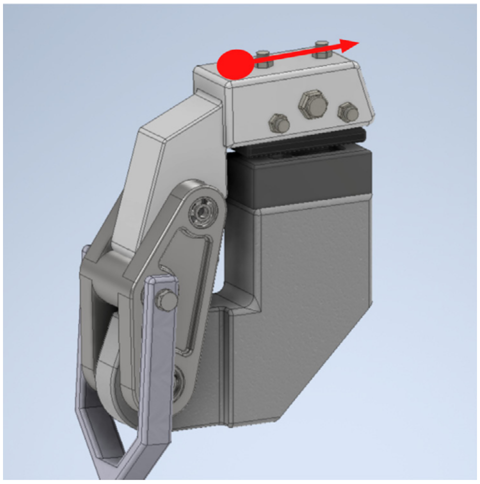

Fig. 8. Stamping unit in closed configuration.

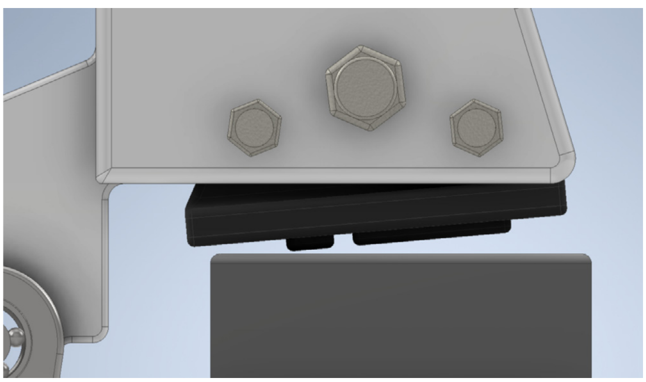

Fig. 9. Front-rear misalignment.

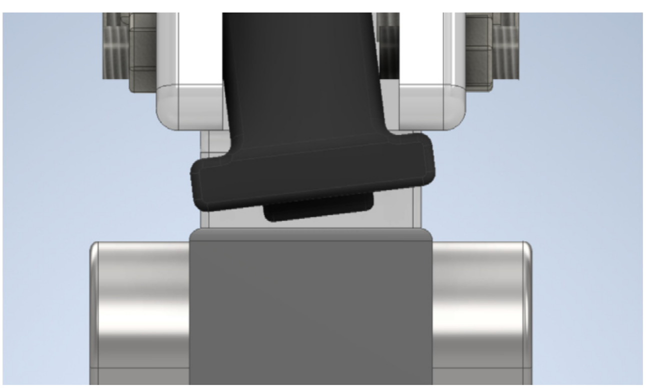

Fig. 10. Left–right misalignment.

Two different types of misalignments were observed; they have been named in accordance with the operator's position during the maintenance of the unit: in Fig. 9 the *front-rear* misalignment (mis_{FR}), and in Fig. 10 the *left–right* misalignment (mis_{LR}).

An accelerometer was chosen to be used on body Ⓑ of the stamping unit, mounted as shown in Fig. 8 with a red arrow, to acquire real-time information from the machine. The accelerometer signal will be employed as input for the soft sensor. Consistently with what it is explained in Section 2, the choice of type and location of this sensor has been carried out by means of a numerical approach, by running simulations on a virtual model of the stamping unit.

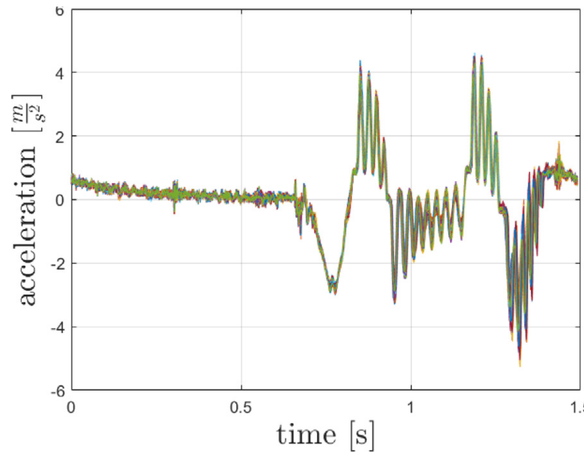

Fig. 11. Example of cycles collected at 400 ppm.

Table 1
Overall number of cycles for production rate, pressure and alignment condition.

	Nominal				Mis_{FR}				Mis_{LR}			
	Pressure [bar]											
	1.5	2	2.5	3	1.5	2	2.5	3	1.5	2	2.5	3
Rate [ppm] 400	283	260	258	221	269	262	254	239	249	251	245	253
500	294	235	265	243	246	257	245	253	251	232	234	264
600	145	166	173	196	184	186	186	179	190	191	193	183
800	201	231	239	264	261	246	239	246	247	255	247	252

A summary of the manufacturing cycles collected during some experiments, to simulate different working and alignment conditions for the stamping unit, is reported in Table 1. The machine counts two main parameters that have a relevant impact on the machine behaviour (thus influencing the vibration measurement of the accelerometer): the hydraulic circuit's *pressure* and the *production rate*. The first parameter determines the force applied to body Ⓑ, in turn influencing the vertical force that pushes on the sheet, whereas the second one regulates the wheel's velocity. Generally, the pressure might vary from 1.5 to 3 bar, while the production rate ranges between 400 and 800 parts per minute (ppm). The subgroups of parameters considered for the experiments are 1.5, 2, 2.5, and 3 bar for the pressures and 400, 500, 600 and 800 ppm for the production rates.

Considering that the sampling frequency of the accelerometer is constant for different tests, the production rate influences the cycles' length, i.e. the number of samples acquired from the accelerometer in a single manufacturing cycle. The presence of different shapes for the input signals represents a major issue for a neural network, since it must always be fed with tensors of fixed size. An example of the time histories acquired at different production rates are shown in Figs. 11 and 12. To simulate the two misalignment conditions, the bolts with rubber tips ③ are loosened, a feeler gauge 0.05 mm thick is placed between the punch and the die, and then the bolts are tightened again. In case of mis_{DO} misalignment, the gauge is positioned on the operator side (as shown in Fig. 9), instead, for mis_{LR} condition, it is placed on the right-hand part of the die (consistently with Fig. 10).

4. Test case results

In the considered case, the machine can work in two different faulty conditions determined by the two possible misalignments between punch and die. The objective is to understand when the product results non-nominal, without being interested in distinguishing the type of misalignment. Hence, an anomaly detector is necessary to discriminate when the machine deviates from the nominal condition.

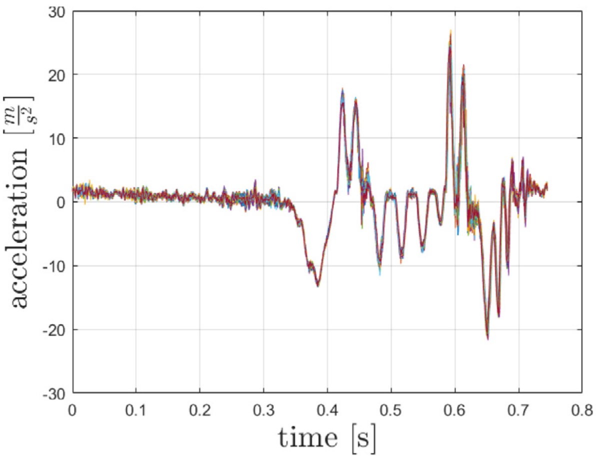

Fig. 12. Example of cycles collected at 800 ppm.

From previous experience in the development of a quality control system for this machine with different approaches (Bono et al., 2022b), it has been understood that the production rate has a big impact on the performance of the algorithms, since it is difficult to generalize the machine's behaviour for different values of this parameter. On the contrary, the pressure has shown to have a negligible effect on the classifiers' performances. For this reason, the proposed approach is applied with the objective of generalizing the monitoring at different production rates.

In the following paragraphs, the results obtained with this novel approach are compared with the algorithms and methodologies found in literature (Lei et al., 2009; Zhang et al., 2013; Qin et al., 2018), relying on the use of dimensionless features in order to generalize results for a mechanical system working at different rotating speeds. In particular, the parameters adopted are: *Skewness coefficient*, *Kurtosis coefficient*, *clearance factor*, *shape factor*, *impulse factor* and *crest factor*. The information coming from these features are put together to develop shallow learning models, as described in the related works. Instead for deep models, the input are the raw acceleration measurement and process variables (i.e. pressure and production rate), as highlighted in Fig. 3. The output for all the models is the condition which the system is working at (i.e. *Nominal*, Mis_{FR}, Mis_{LR}). The dataset adopted for the following results is the one described in Section 3.

In this case, to scale data, *z-score* transformation is applied by using the following formula:

$$z = \frac{x-\mu}{\sigma} \tag{1}$$

In this way the signals' distribution and each dimensionless feature will display a mean of zero and a standard deviation of one. This procedure is applied for all adopted algorithms; furthermore, for supervised ones, the label (*Nominal*, Mis_{FR}, Mis_{LR}) corresponding to the working condition is associated to each time signal.

These three categories may be further classified as compliant class (*Nominal*) and non-compliant class (abnormal, namely Mis_{FR}, Mis_{LR}), so that four possible outcomes can occur for every observation. Eventually, four values are derived by testing every classifier:

- *True nominal* (TN): number of nominal observations correctly predicted as nominal.
- *True abnormal* (TA): number of abnormal observations correctly predicted as abnormal.
- *False nominal* (FN): number of abnormal observations wrongly predicted as nominal.
- *False abnormal* (FA): number of nominal observations wrongly predicted as abnormal.

And from these values four *metrics* are employed, to compare the performances of the classifiers (Hossin and Sulaiman, 2015):

$$accuracy = \frac{TN+TA}{TN+TA+FN+FA} \tag{2}$$

$$recall = \frac{TA}{TA+FN} \tag{3}$$

$$precision = \frac{TA}{TA+FA} \tag{4}$$

$$F_1 score = \frac{2 * precision * recall}{precision + recall} \tag{5}$$

In this test case, it is important to evaluate these metrics at the different production rates of the machine, since the challenge of the task is in obtaining high classification performances at production rates not used for training the algorithm.

4.1. Shallow learning

By following the procedure described at the beginning of this section, four different shallow models have been built on top of dimensionless features:

- **Discriminant analysis (DA)**: Discriminant analysis (DA) is a method based on the probabilistic assumption that all the observations for a specific class are realization of a normal probability distribution. If a classification problem is considered with a number of classes equal to g, the idea is to fit a multivariate Gaussian distribution (with dimension m) for each class in the training set; then these distributions are used to determine separation boundaries between classes in the feature space.
- **K-nearest neighbours (KNN)**: This method is based on the idea to classify a new observation by looking at the K neighbours, i.e. the K nearest data-points in the dataset. The algorithm works by computing the distance between each data-point and the new observation, and it finds the probability of the points being similar to the new data.
- **Support-vector machine (SVM)**: Support Vector Machine (SVM) is a method used for binary classification problems, but it can be extended also to multi-class problems by considering one binary classification at a time (i.e. by applying an SVM while considering only two classes at every step). The core idea of this algorithm is to search for a classification boundary, able to separate two classes of objects. This is done in an iterative fashion, by using the observations closer to the decision boundary, that take the name of "support vectors".
- **Ensemble bagged tree (EBT)**: Decision trees are classifiers which operate by recursively partitioning the feature space. A decision tree is made up of nodes and branches. Except for the initial node and the test nodes (i.e. nodes without outgoing branches), every node counts one incoming branch and two (or sometimes more) outgoing branches. This means that, in decision trees, each decision node splits the feature space into two or more subspaces. In binary decision trees, the discrimination takes the form of a yes/no question. A new observation is classified by navigating it from the root to an ending leave, so that its attributes are iteratively tested along the procedure (Rokach and Maimon, 2009). More decision trees are usually collected in a form of ensemble learning, as it happens for boosting aggregation (Bagging). In Bagging, more weak classifiers are trained on different subgroups of data, created by taking samples uniformly and with replacement from the original dataset. (Sexton and Laake, 2009)

Each observation is characterized by six features and the models are trained by considering just observation at lower and upper limit (respectively 400 and 800 ppm) for the production. During training, each model undergoes a Bayesian optimization procedure (implemented

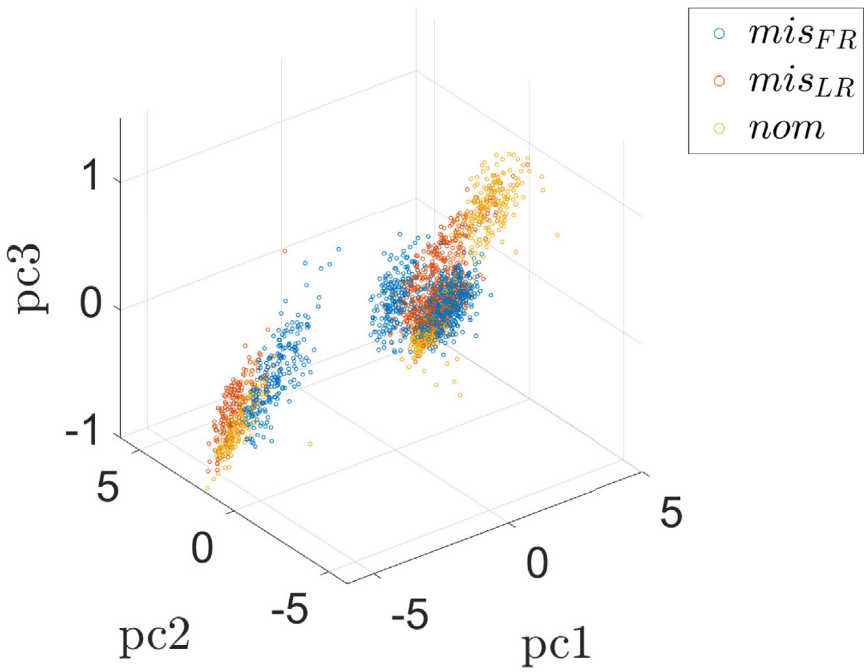

Fig. 13. Data-points at 500 and 600 ppm. First three principal components.

in python Scikit-learn package) to find the best hyperparameter values (Yang and Shami, 2020) that permit to obtain a higher overall accuracy level. Depending on the considered model, different hyperparameters are contemplated. To run this optimization technique, a maximum number of iterations of 30 (for each model) and a callback to stop the procedure if two successive iterations show a difference in accuracy lower than 0.001. For DA, the solver is set to svd, the shrinkage parameter is optimized; for KNN, the number of neighbours, the algorithm used to compute the nearest neighbours and the weight function are considered; for SVM, the kernel function is the main hyperparameter (linear kernel seems to be the best one in this case); for DT, Gini function is adopted to measure the quality of a split, instead different hyperparameters are optimized: the minimum number of data points to split a decision node or to obtain a leaf node and the maximum number of leaf nodes.

The models are trained considering 80% of the data available at the two boundary production rates (400 and 800 ppm), and then, they are evaluated for the remaining partition at these rates. Moreover, the metrics are also computed for the other two production rates (500 and 600 ppm). The performances of shallow methods are reported in Table 2. By looking at this table, it is possible to infer that the SVM model obtained the best performances at 400 and 800 ppm, instead when dealing with production rates not present in the training set, the four models are not able to work properly showing high classification error. This behaviour can be interpreted with the plot in Fig. 13, where the three principal components calculated on the dataset at 500 and 600 ppm are represented: the different working conditions are not well separated by just considering dimensionless features. Therefore, a different approach should be adopted to cope with this issue.

The training times, considering the Bayesian optimization procedure, in seconds for the different shallow methods are: 19 s for DA, 16 s for KNN, 6290 s for SVM and 874 s for EBT. instead, the prediction times for the entire test set: 41 ms for DA, 37 ms for KNN, 12 ms for SVM and 799 ms for EBT.

4.2. Neural network

To apply the proposed methodology (Section 2), it is necessary to train a neural network by considering data referred to the two boundary working points, which are in this case the upper and lower limit for the production rates (respectively 400 and 800). To cope with the irregular number of samples at distinct production rates, the "resizing approach" is applied, so to always have the same number of samples, corresponding to the dimension of longest ones (i.e. data at 400 ppm). The architecture of the ANN adopted for this soft sensor is represented in Fig. 14. Consistently with the structure prescribed in Fig. 3 the network is characterized by two different branches: the first one is devoted to the analysis of accelerometer signals by means of two 1D convolutional blocks (convolutional layer followed by a pooling layer), whereas the second carries out the analysis of pressures and production rates, by applying a dropout layer to reduce the effect of their particular relationship in the training set. Then a concatenate layer merges together the tensors coming from the two branches, before the final classification is realized by means of a dense layer.

This network is the result of the Bayesian hyperparameter procedure. In neural networks, two families of hyperparameters are present: *structural hyperparameters*, that define the overall architecture of the model, and *optimizer hyperparameters*, that influence the quality and the speed of the training procedure. The optimizer hyperparameters are chosen by adopting an ADAM optimizer, a batch size of 128, a maximum number of epochs of 100 and a callback that stops the training if the accuracy does not improve in 10 epochs. Instead the structural ones are optimized to obtain the previously described architecture. These hyperparameters are the dimensions of convolutive layers, the dropout rate (considered a unique value for all the dropout layers) and the recurrent dropout rate in the GRU layer.

Eventually, the neural network achieves an accuracy of 99.8% in training at (400 and 800 ppm), hence it is possible to conclude that not only the resizing looks suitable for the analysis of periodic signals, but also that neural network in this case outperforms shallow methods. On the other hand, when the model is tested on the data at 500 and 600 ppm the accuracy drastically reduces respectively to 57.4% and 57.8%. The detailed metrics for the application of the neural network are described in Table 2. Considering that this is a triple classification, these results in test are not dramatic nor sensational, but they are certainly far from an acceptable outcome that could be used for an automatic control quality system. The training time (including the overall optimization procedure) is 39420 s, whereas the times elapsed in prediction for the entire test set is 824 ms.

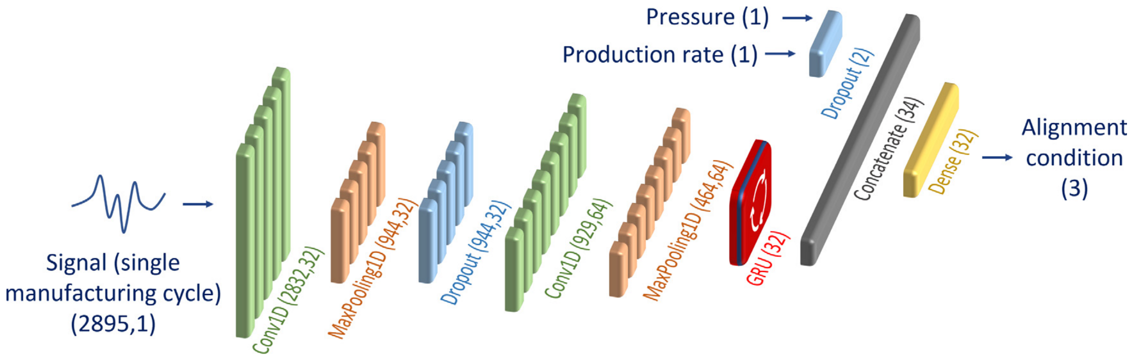

Fig. 14. Neural network's structure for the soft sensor.

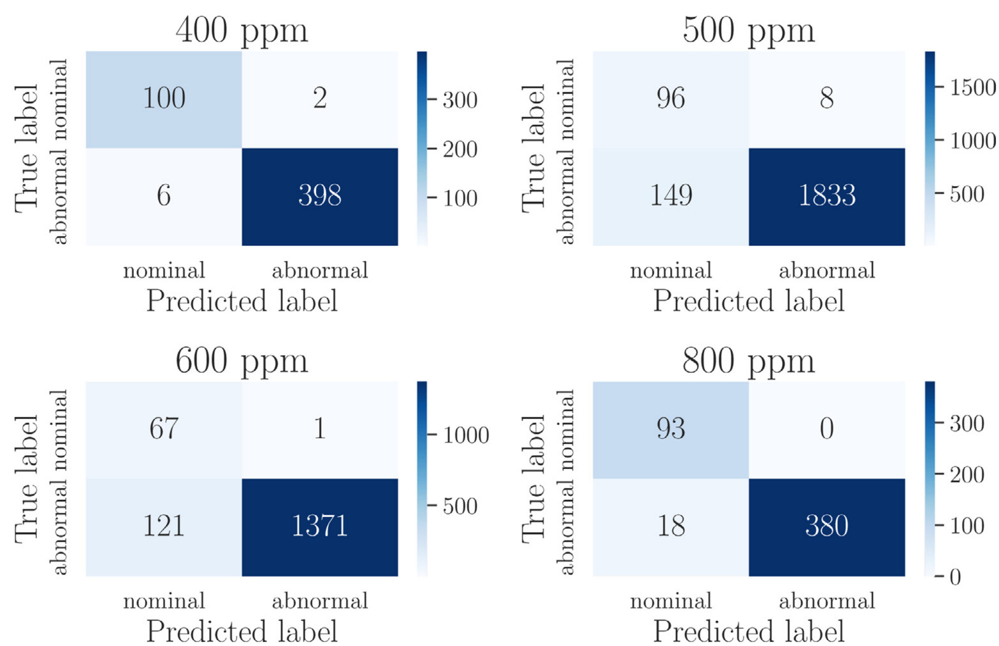

Fig. 15. Confusion matrices for observation in the test set at different production rate.

4.3. Application of the proposed methodology

Nominal points at the four rates can be acquired during normal operations of the plant, just after maintenance to be sure they correspond to nominal conditions. To develop the anomaly detector, it is necessary to get access to some anomalies at two "boundary" production rates (to train a neural network), but only then many nominal points at intermediate rate are used to fit the anomaly detector. However, to verify the effectiveness of the approach, the latter is also tested with abnormal data at intermediate speeds.

Then, the dataset (Table 1) is partitioned in the following way:

- *Fitting set*: 70% of the nominal observations at each speed.
- *Threshold set*: composed of 20% of the nominal observations at each speed and 80% of abnormal observations at the two limit velocities.
- *Test set*: used to verify the hypothesis that this algorithm can generalize results at different speed. It is composed by 10% of nominal observations at each speed, 20% of abnormal observation at the two limit velocities and all data related to anomalous conditions at 500 and 600 ppm.

The process pipeline consists in the following steps:

1. The neural network described in Fig. 14 is trained (nominal and abnormal data at 400 and 800 ppm).
2. The last layer of the neural network is removed, so that the outputs are the early predictors.
3. The data-points at 400 and 800 ppm, plus all the cycles regarded as nominal from the production line (at every available intermediate rate, then 500 and 600 ppm) pass through the pre-trained neural network, so that to extract the early predictors, that are now vectors of 32 meaningful features.
4. The fitting set is used for the definition of the multivariate Gaussian distribution.
5. The threshold set is used to define the threshold probability, accordingly with the procedure proposed in An and Liu (2019).
6. The system could now be deployed on the production line for making predictions on new cycles. Its performance is verified on the test set.
7. Once the machinery is stopped for maintenance and activated again, the system saves a certain number of new cycles as *nominal* and the process could start again from step 3.

Eventually, after the application of the anomaly detector fed with early predictors, this method leads to satisfactory results. The performances in terms of accuracy evaluated on the test set are: 98.4% at 400 ppm, 92.4% at 500 ppm, 92.1% at 600 ppm and 96.3% at 800 ppm. The detailed metrics for the application of the proposed methodology are described in Table 2 and confusion matrices for the different speeds are reported in Fig. 15. The training time to fit the Gaussian distribution

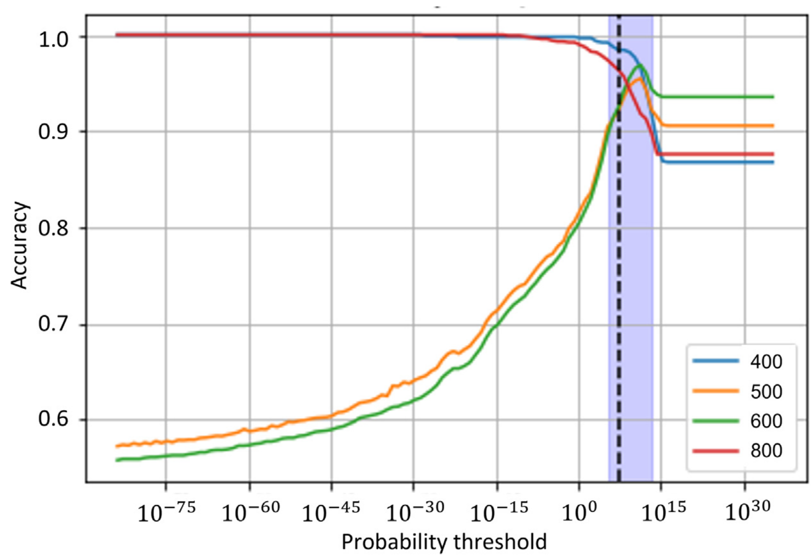

Fig. 16. Accuracies at the different speeds as function of the probability threshold.

Table 2
Metrics for shallow learning methods (DA, SVM, KNN, EBT), neural network (NN) and proposed methodology (PM).

		DA	SVM	KNN	EBT	NN	PM
400	Accuracy	81,8%	97,5%	95,9%	94,8%	99,8%	98,4%
	Precision	87,1%	98,2%	96,9%	96,2%	100,0%	98,5%
	Recall	88,5%	98,4%	97,6%	96,9%	99,8%	99,5%
	F1score	87,8%	98,3%	97,3%	96,5%	99,9%	99,0%
500	Accuracy	39,7%	39,9%	40,2%	41,1%	57,4%	92,5%
	Precision	40,7%	41,1%	41,1%	41,5%	57,1%	92,5%
	Recall	89,9%	89,7%	90,2%	91,2%	96,8%	99,6%
	F1score	56,0%	56,3%	56,5%	57,1%	71,8%	95,9%
600	Accuracy	31,9%	31,4%	33,5%	35,1%	57,8%	92,2%
	Precision	34,8%	35,0%	36,9%	37,9%	56,4%	91,9%
	Recall	70,8%	69,8%	71,7%	73,4%	99,1%	99,9%
	F1score	46,7%	46,7%	48,7%	50,0%	71,9%	95,7%
800	Accuracy	82,9%	97,8%	95,3%	95,1%	99,8%	96,3%
	Precision	83,5%	98,4%	96,6%	97,1%	99,7%	95,5%
	Recall	95,9%	98,7%	97,1%	96,4%	100,0%	100,0%
	F1score	89,3%	98,6%	96,9%	96,8%	99,9%	97,7%

and then set the threshold is 1.1 s, while the time to deploy it on the test set is 5 ms.

Some considerations can then be inferred:

- This arrangement leaves the performances at the UR (800) and LR (400) untouched, with respect to the neural network developed, but overcomes the problem of considering intermediate production rates.
- At 500 and 600 ppm the only limit is that the system misclassifies the mis_{lr} for nominal points; since the accelerometer is placed in driver-operator direction (and in fact all misdo are correctly labelled), this might be due to a lack of information rather than a limit of this approach.
- The probability threshold chosen in this way is suitable to obtain good performances at each speed, but evidently intermediate rates (500 and 600 ppm), for which the nominal observations are only theoretically available, seem penalized with respect to the two boundary speeds (400 and 800 ppm).

Regarding the last point, in the dataset, abnormal observations at intermediate rates are present, and then it is possible to evaluate and visualize how the performance, at different speeds, changes for different threshold values of the multivariate Gaussian (Fig. 16). Evidently, there is a range of probability thresholds (the region highlighted in blue colour), which minimizes the number of misclassifications at the different speed; however, the threshold chosen with the presented algorithm (step List 5 of the procedure pipeline), represented by the dashed vertical line in Fig. 16, belongs to this region, showing that the proposed methodology is robust indeed. In particular, the curve at 500 and 600 ppm in Fig. 16 have been realized by having anomalies as well, which is obviously not the case when the quality control system has been just developed (as stated in Section 2.5). At the beginning, only nominal points are collected in different working points. However, when maintenance is required, the system can flag the last manufacturing cycles as anomalies, collecting them at different production rates of interest. In this way, the curves in Fig. 16 can be effectively drawn, and the threshold can be set to a value which maximizes the overall accuracy (intended as mean value of the accuracy at each production rate).

5. Conclusions

In this paper, a novel approach to address the problem of quality control for machines working under highly inconsistent condition is proposed. The system relies on the application of a soft sensor and the use of its pre-trained neural network to extract meaningful features for feeding an anomaly detection system. The procedure to collect a limited, but meaningful dataset, is reported as well. The methodology has been applied to a real test case, i.e a workstation of a manufacturing line characterized by significantly variable working conditions. This made it possible to prove how this method has shown to be effective in a situation where other methods based on simple feature extraction failed.

With this work, it was highlighted how some approaches developed for fault detection in machines cannot be applied in a quality control context. This is proven in Table 2, where the metrics for the different algorithms are compared. In the reported test case, the solution allows to move from a spot-check quality control (performed every 4 h) to real time monitoring. The economic impact of the monitoring system can be quantified in terms of the number of non-compliant outputs, which passes from around 18% to less than 1%. This surely demonstrates that investment related to implementation of the acquisition system, the collection of the training set and development of the algorithm are justified. Moreover, the approach looks easily adaptable on a wide variety of quality control problems where the collection of a training set appears troublesome and finally, it adequately allows to overcome the limits underlined in previous works on similar topics (Bono et al., 2022b).

A future development of this work might address the application of the proposed methodology on other case studies, to validate or improve the procedure itself. Moreover, it might be stated that having a multivariate Gaussian distribution is not the best possible choice for the anomaly detector. The time required to fit the Gaussian is strictly related to the number of features. In this work, the neural network allows to generate a low dimensional feature space by means of the early predictors (transfer learning) and, for this reason, the computational time related to fitting the Gaussian is competitive. However, future works must involve the evaluation of different models as anomaly detectors, possibly to increase the extrapolation capabilities of the quality control systems.

Moreover, the use of Generative Adversarial Networks (GANs) and Autoencoders look promising in other research fields, and therefore those methods should necessarily addressed by future works.

CRediT authorship contribution statement

F.M. Bono: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **L. Radicioni:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **S. Cinquemani:** Conceptualization, Methodology, Resources, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

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

Data availability

The data that has been used is confidential.

References

Albers, A., Gladysz, B., Pinner, T., Butenko, V., Stürmlinger, T., 2016. Procedure for defining the system of objectives in the initial phase of an industry 4.0 project focusing on intelligent quality control systems. Procedia CIRP 52, 262–267. http://dx.doi.org/10.1016/j.procir.2016.07.067, URL https://www.sciencedirect.com/science/article/pii/S2212827116308666, The Sixth International Conference on Changeable, Agile, Reconfigurable and Virtual Production (CARV2016), Bath, United Kingdom.

An, Y., Liu, D., 2019. Multivariate Gaussian-based false data detection against cyber-attacks. IEEE Access 7, 119804–119812. http://dx.doi.org/10.1109/ACCESS.2019.2936816.

Bono, F.M., Cinquemani, S., Radicioni, L., Conese, C., Tarabini, M., 2022a. An approach for fault detection based on multibody simulations and feature selection algorithm. In: Meyendorf, N.G., Farhangdoust, S., Niezrecki, C. (Eds.), NDE 4.0, Predictive Maintenance, and Communication and Energy Systems in a Globally Networked World, Vol. 12049. SPIE, International Society for Optics and Photonics, pp. 117–125. http://dx.doi.org/10.1117/12.2607055.

Bono, F.M., Radicioni, L., Cinquemani, S., Conese, C., Tarabini, M., 2022b. Development of soft sensors based on neural networks for detection of anomaly working condition in automated machinery. In: Meyendorf, N.G., Farhangdoust, S., Niezrecki, C. (Eds.), NDE 4.0, Predictive Maintenance, and Communication and Energy Systems in a Globally Networked World, Vol. 12049. SPIE, International Society for Optics and Photonics, pp. 56–70. http://dx.doi.org/10.1117/12.2607072.

Bruwer, M.-J., MacGregor, J.F., Bourg, W.M., 2007. Soft sensor for snack food textural properties using on-line vibrational measurements. Ind. Eng. Chem. Res. 46 (3), 864–870.

Chen, J., Hu, W., Cao, D., Zhang, B., Huang, Q., Chen, Z., Blaabjerg, F., 2019. An imbalance fault detection algorithm for variable-speed wind turbines: A deep learning approach. Energies 12 (14), http://dx.doi.org/10.3390/en12142764, URL https://www.mdpi.com/1996-1073/12/14/2764.

Cho, K., van Merrienboer, B., Gülçehre, Ç., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation, CoRR abs/1406.1078. arXiv:1406.1078, URL http://arxiv.org/abs/1406.1078.

Fortuna, L., Graziani, S., Xibilia, M.G., 2005. Soft sensors for product quality monitoring in debutanizer distillation columns. Control Eng. Pract. 13, 499–508. http://dx.doi.org/10.1016/J.CONENGPRAC.2004.04.013.

Guo, W., Pan, T., Li, Z., Chen, S., 2019. Model calibration method for soft sensors using adaptive Gaussian process regression. IEEE Access 7, 168436–168443. http://dx.doi.org/10.1109/ACCESS.2019.2954158.

Hasan, M.J., Sohaib, M., Kim, J.-M., 2020. A multitask-aided transfer learning-based diagnostic framework for bearings under inconsistent working conditions. Sensors 20 (24), http://dx.doi.org/10.3390/s20247205, URL https://www.mdpi.com/1424-8220/20/24/7205.

Hashemi, M., 2019. Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation. J. Big Data 6 (1), 1–13.

Hossin, M., Sulaiman, M.N., 2015. A review on evaluation metrics for data classification evaluations. Int. J. Data Min. Knowl. Manag. Process. 5 (2), 1.

Hsiao, Y.-D., Kang, J.-L., Wong, D.S.-H., 2021. Development of robust and physically interpretable soft sensor for industrial distillation column using transfer learning with small datasets. Processes 9 (4), http://dx.doi.org/10.3390/pr9040667, URL https://www.mdpi.com/2227-9717/9/4/667.

Kadlec, P., Gabrys, B., Strandt, S., 2009. Data-driven soft sensors in the process industry. Comput. Chem. Eng. 33 (4), 795–814.

Kano, M., Nakagawa, Y., 2008. Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry. Comput. Chem. Eng. 32 (1–2), 12–24.

Kosti, M.V., Mittas, N., Angelis, L., 2012. Alternative methods using similarities in software effort estimation. In: Proceedings of the 8th International Conference on Predictive Models in Software Engineering. PROMISE '12, Association for Computing Machinery, New York, NY, USA, pp. 59–68. http://dx.doi.org/10.1145/2365324.2365333.

Lei, Y., He, Z., Zi, Y., 2009. Application of an intelligent classification method to mechanical fault diagnosis. Expert Syst. Appl. 36 (6), 9941–9948. http://dx.doi.org/10.1016/j.eswa.2009.01.065, URL https://www.sciencedirect.com/science/article/pii/S0957417409001109.

Li, P., Hu, W., Hu, R., Chen, Z., 2020. Imbalance fault detection based on the integrated analysis strategy for variable-speed wind turbines. Int. J. Electr. Power Energy Syst. 116, 105570. http://dx.doi.org/10.1016/j.ijepes.2019.105570, URL https://www.sciencedirect.com/science/article/pii/S0142061518339267.

McBain, J., Timusk, M., 2009. Fault detection in variable speed machinery: Statistical parameterization. J. Sound Vib. 327 (3), 623–646. http://dx.doi.org/10.1016/j.jsv.2009.07.025, URL https://www.sciencedirect.com/science/article/pii/S0022460X09005896.

McCartney, M., Haeringer, M., Polifke, W., 2020. Comparison of machine learning algorithms in the interpolation and extrapolation of flame describing functions. J. Eng. Gas Turbines Power 142 (6), http://dx.doi.org/10.1115/1.4045516, arXiv:https://asmedigitalcollection.asme.org/gasturbinespower/article-pdf/142/6/061009/6538416/gtp_142_06_061009.pdf, 061009.

Mitra, A., 2016. Fundamentals of Quality Control and Improvement. John Wiley & Sons.

Murua, M., Suarez, F.V., Lalmolda, J.A.O., Oscoz, M.P., Olivan, A.D., 2020. Machine learning-based analysis engine to identify critical variables in multi-stage processes: Application to the installation of blind fasteners. DYNA 95, 534–540. http://dx.doi.org/10.6036/9403, URL http://www.revistadyna.com/Articulos/Ficha.aspx?IdMenu=33094c58-1bb5-4163-8382-a0d0942862d2&Cod=9403&Idioma=en-GB.

Pal, K.K., Sudeep, K.S., 2016. Preprocessing for image classification by convolutional neural networks. In: 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology. RTEICT, Bangalore, India, pp. 1778–1781. http://dx.doi.org/10.1109/RTEICT.2016.7808140.

Qin, A., Hu, Q., Lv, Y., Zhang, Q., 2018. Concurrent fault diagnosis based on Bayesian discriminating analysis and time series analysis with dimensionless parameters. IEEE Sens. J. 19 (6), 2254–2265.

Rege, B.D., Gawel, J., Kou, J.H., 2002. Identification of critical process variables for coating actives onto tablets via statistically designed experiments. Int. J. Pharm. 237 (1), 87–94. http://dx.doi.org/10.1016/S0378-5173(02)00037-6, URL https://www.sciencedirect.com/science/article/pii/S0378517302000376.

Rokach, L., Maimon, O., 2009. Classification trees. In: Data Mining and Knowledge Discovery Handbook. Springer, pp. 149–174.

Sainath, T.N., Vinyals, O., Senior, A., Sak, H., 2015. Convolutional, long short-term memory, fully connected deep neural networks. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, Brisbane, Australia, IEEE, pp. 4580–4584.

Sexton, J., Laake, P., 2009. Standard errors for bagged and random forest estimators. Comput. Statist. Data Anal. 53 (3), 801–811.

Sohaib, M., Kim, J.-M., 2020. Fault diagnosis of rotary machine bearings under inconsistent working conditions. IEEE Trans. Instrum. Meas. 69 (6), 3334–3347. http://dx.doi.org/10.1109/TIM.2019.2933342.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Worden, K., Sohn, H., Farrar, C., 2002. Novelty detection in a changing environment: Regression and interpolation approaches. J. Sound Vib. 258 (4), 741–761. http://dx.doi.org/10.1006/jsvi.2002.5148, URL https://www.sciencedirect.com/science/article/pii/S0022460X02951485.

Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 415, 295–316. http://dx.doi.org/10.1016/j.neucom.2020.07.061.

Yu, Y., Si, X., Hu, C., Zhang, J., 2019. A review of recurrent neural networks: LSTM cells and network architectures. Neural Comput. 31 (7), 1235–1270. http://dx.doi.org/10.1162/neco_a_01199.

Yuan, X., Qi, S., Shardt, Y.A., Wang, Y., Yang, C., Gui, W., 2020. Soft sensor model for dynamic processes based on multichannel convolutional neural network. Chemometr. Intell. Lab. Syst. 203, 104050.

Zăvoianu, A.-C., Lughofer, E., Pollak, R., Eitzinger, C., Radauer, T., 2021. A soft-computing framework for automated optimization of multiple product quality criteria with application to micro-fluidic chip production. Appl. Soft Comput. 98, 106827. http://dx.doi.org/10.1016/j.asoc.2020.106827, URL https://www.sciencedirect.com/science/article/pii/S1568494620307651.

Zhang, Q.-H., Hu, Q., Sun, G., Si, X., Qin, A., 2013. Concurrent fault diagnosis for rotating machinery based on vibration sensors. Int. J. Distrib. Sens. Netw. 9 (4), 472675.

Zhu, X., Rehman, K.U., Wang, B., Shahzad, M., 2020. Modern soft-sensing modeling methods for fermentation processes. Sensors 20 (6), http://dx.doi.org/10.3390/s20061771, URL https://www.mdpi.com/1424-8220/20/6/1771.