Results in Engineering 17 (2023) 100999

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

Modelling of thermal shrinkage of seamless steel pipes using artificial neural networks (ANN) focussing on the influence of the ANN architecture

Raphael Langbauer [a,*], Georg Nunner [b], Thomas Zmek [b], Jürgen Klarner [b], René Prieler [a], Christoph Hochenauer [a]

[a] Graz University of Technology, Institute of Thermal Engineering, Inffeldgasse 25/B, 8010, Graz, Austria
[b] voestalpine Tubulars GmbH & Co KG, Alpinestraße 17, 8652, Kindberg, Austria

ARTICLE INFO

Keywords:
Diameter prediction
Artificial neural network
Model response
Steel pipe
Hot rolling

ABSTRACT

This paper presents two main novel findings. (1) The first finding is the development of an artificial neural network (ANN) model for thermal shrinkage of seamless steel pipes, which represents a new application for ANNs. Mill operators need such fast and accurate models to predict the final pipe outer diameter at ambient temperature based on the hot state immediately after rolling. The goal of this work was to lower the reject rate. However, small relative changes in the diameter are currently difficult to predict by using conventional ANNs. Therefore, a more sensitive target variable and a modified ANN architecture were applied to solve this problem. Data for training and validation were obtained from measurements on a hot-rolling mill. (2) The second finding is based on an investigation performed to determine the number of hidden neurons affected the model response, considering the data used. The knowledge obtained helps to determine the most suitable number of hidden neurons and to prevent overfitting. No generally accepted solution to these problems had previously been proposed in the literature. Consequently, this paper significantly supplements current research studies that describe applications of ANNs.

1. Introduction

The range of application for steel pipes is large. In mechanical engineering, these are mostly used to transport fluids and to transfer or absorb mechanical loads. Pipes are preferred to bar stock for steel constructions in practically all cases, because both have a similar mechanical load capacity when equal outer diameters are used although pipes are significantly lighter. In addition, less material is needed for manufacturing. The main types of pipes, i.e. welded pipes and seamless pipes, are classified by their production process [1].

Welded pipes have low wall thicknesses and their production is more cost-efficient than seamless pipes because welded pipes are made from thin sheet metal. However, the welding seam and the encircling heat-affected zone represent weak points [1,2]. For this reason, welded pipes are used for low-load applications, such as when transporting fluids under low pressure. Examples include pipelines [3–5] or different types of heat exchangers [6,7]. Welded pipes are also well-suited for steel constructions due to their low wall thickness and, thus, low weight. Thermal treatment can improve the mechanical properties of welded pipes. The recrystallisation of the microstructure that occurs along the welding seam and in the heat-affected zone improves the characteristics [1,2].

The production of seamless steel pipes is significantly more expensive than that of welded pipes. Many process steps are required to manufacture a pipe with precise geometry from a solid cylindrical block [8]. However, it is possible to manufacture thick wall thicknesses by using this procedure. In addition, welding is avoided, thus also avoiding the creation of a weak point on the pipe. Due to these mentioned properties, seamless pipes are highly suitable for high-load applications. These pipes are used to transport fluids under high pressure in pressure boilers [9] or to transfer torque in machinery. The petroleum industry has also a high demand for seamless pipes due to their use in drilling operations [10]. A further advantage of seamless pipes is that specific mechanical properties can, if required, already be created during the finishing hot-rolling process by normalised rolling [11] or thermo-mechanical processing [12,13]. Therefore, thermal treatment is only necessary in certain cases.

* Corresponding author.
E-mail addresses: raphael.langbauer@tugraz.at (R. Langbauer), georg.nunner@vatubulars.com (G. Nunner), thomas.zmek@vatubulars.com (T. Zmek), juergen.klarner@vatubulars.com (J. Klarner), rene.prieler@tugraz.at (R. Prieler), christoph.hochenauer@tugraz.at (C. Hochenauer).

https://doi.org/10.1016/j.rineng.2023.100999
Received 9 January 2023; Received in revised form 28 February 2023; Accepted 2 March 2023
Available online 5 March 2023

Nomenclature

R	Correlation coefficient [−]
R^2	Coefficient of determination [−]
T	Temperature [K]
σ	Standard deviation [−]

Abbreviations

ANN	Artificial neural network
CFD	Computational fluid dynamics
CU	Cooling unit
DMS	Diameter measuring system
DSU	Descaling unit
ELU	Exponential Linear Unit
FDM	Finite difference method
FEM	Finite element method
LReLU	Leaky Rectified Linear Unit
MAPE	Mean absolute percentage error
MP	Measuring point
MSE	Mean squared error
QCU	Quick cooling unit
ReLU	Rectified Linear Unit
RHF	Reheating furnace
RMSE	Root mean squared error
RS	Roll stand
SELU	Scaled Exponential Linear Unit
SRRM	Stretch-reducing rolling mill

Pipes also have to be able to withstand corrosion [14,15] and thermal loads in addition to mechanical loads. In such certain cases, steel with a higher amount of suitable alloying elements is used [16]. The chemical elements chromium (Cr) and nickel (Ni) significantly improve the corrosion resistance [17]. These elements also have a positive influence on the thermal load capacity of steel, which decreases as the temperatures increase [18].

The main objective of this work was to develop a numerical model that could be used to estimate the thermal shrinkage of seamless steel pipes, starting from the hot state immediately after the pipe has been rolled. The required variable was the outer diameter of the pipe in its cold state after being rolled and completely cooled to ambient temperature. The permitted deviation from the specified diameter is defined in the material standards (e.g. ASTM A53) and must not be exceeded. ASTM A53, for example, permits a maximum deviation in the outer diameter of seamless steel pipes of 1%. Pipes that do not meet this deviation limit are marked as rejects and can no longer be sold. Producing rejects places undesirable cost burdens on the manufacturer and should be avoided. Therefore, the mill operator should change the process parameters quickly to prevent rejects from being produced. One challenge that the operator faces, however, is that the cold pipe diameter can only be measured for quality control purposes after it has completely cooled down, and the thermal shrinkage that many pipes undergo can result in their becoming rejects in the interim. These observations indicated that a numerical model is needed that can be used to predict the diameter of pipes immediately, i.e. to estimate their thermal shrinkage immediately, enabling the operator to perform interventions early on.

Basically, thermal expansion or shrinkage is a linear mechanism which can be described simply by applying mathematical methods appropriate for straight bodies. Low computation times are needed to perform calculations of such cases, one example of which is the elongation of pipes in the axial direction. However, predicting the thermal expansion of the annulus of a pipe cross section is significantly more difficult. The outer boundary layers expand more than the inner boundary layers, and this difference is a function of the wall thickness of the pipe, the temperature difference, and the steel composition. The inner layers of the pipe restrain the elongation of the outer layer; therefore, residual stresses are induced within the pipe's wall [19]. This elongation process differs clearly from that seen in a straight body. The residual stresses can be resolved by plastic deformation, if the temperature is high enough to reduce the yield stress of the steel [18]. This means that the temperature is also a parameter that influences the elongation of the pipe cross section in the circumferential direction. Additionally, alloying elements can affect the thermal expansion coefficient of steel and how strongly the yield stress depends on the temperature. Currently, this problem cannot be addressed by using numerically efficient mathematical methods for the reasons described above.

Soft computing techniques [20] (e.g. genetic algorithms, fuzzy logic, machine learning), however, are highly suitable for solving this problem due to the low computation times required. In recent years, these techniques have been applied more frequently and provided satisfying results. Artificial neural networks (ANNs) are a sub-group of machine learning algorithms and are the most frequently used out of all the soft computing techniques mentioned. ANNs have diverse range of application, being used in such areas as mining blasting [21], petroleum production [22], and the production of sustainable fuels [23–25]. Numerical problems for fluid flow [26], thermal [27], and structural analysis [28] are also addressed. However, the alternative soft computing techniques previously mentioned are also currently in use. Some researchers have compared these techniques with ANNs for some specific use cases [29–31].

ANNs are being increasingly used for modelling industrial [8] and thermal processes [32] and employed instead of elaborate physical models, because they are highly suitable for solving complex problems that have many influencing parameters. The works of Lim and Kang [33], Son et al. [34], and Zhang et al. [35] show that physics-based modelling is often not possible due to the lack of sufficient computation time or accuracy. These authors also suggested replacing the currently used mathematical models by an ANN. ANNs can even be used to represent the nonlinear trends resulting from interactions among many influencing variables. Therefore, ANN-based models provide more accurate predictions for regression problems than regression models or polynomial functions, if nonlinear system behaviour is observed, as demonstrated by Esendağ et al. [36]. The computation time required when applying the former models is extremely low due to the simple mathematics used; therefore, ANN are highly suitable for performing calculations online as part of the running production process, enabling results to be obtained rapidly. When predictions can be made quickly enough, the production operator can intervene earlier and in more purposeful ways. Physics-based models often are not suitable for these applications due to the high amount of computational effort needed to obtain results. In addition, these models require the use of powerful hardware to achieve adequate computational times, unlike ANNs.

In this paper, the development and validation of a new ANN model is described that can be used to accurately predict the thermal shrinkage in the outer diameter of seamless steel pipes. The required variable is the outer diameter in the cold state at ambient temperature. The initial state for thermal shrinkage is the hot state immediately after hot-rolling. Two types of models were developed to make this prediction. The first model type was used to predict the cold diameter directly as target variable, and the second type, to predict the temperature difference between the cold and hot states. The prediction from the second model type is added to the hot diameter to obtain the cold diameter. Subsequently, the outcomes of a study on different architectures carried out by applying both model types are reported. Three promising architectures with a low

error margin were selected and applied to calculate the results for the training and validation data set. Subsequently, surface plots are described that were generated as a model response, and the results of an evaluation of the trends for these surfaces regarding their correct physical behaviour are reported. Finally, a discussion is presented of how the architecture and the data point coverage of the parameter space affect the model responses.

2. Review of literature

2.1. Hot-rolling process from the present work

Some scientists working in different research fields have also reported findings on the hot-rolling process studied in the present work (see Section 3). Landfahrer et al. [37] studied the reheating furnace upstream of the rolling mill and developed a numerically efficient model by using a CFD model that could be used to predict the transient temperature distribution on the tube blanks in the furnace while considering their translative and rotative movement. In a subsequent work, Landfahrer et al. [38] also studied different operating conditions and their influence on the temperature distribution. Landfahrer et al. [39] then investigated and modelled the formation of scale in the furnace by applying the existing CFD model, which characterised the gas-phase combustion. Finally, Landfahrer et al. [40] presented a model that could be used to apply different fuels and oxidisers and to measure the effect on the temperature distribution of the tube blanks.

The downstream rolling process was investigated by Langbauer et al. [13]. This research group developed a mathematical model to calculate the transient temperature distribution within the pipe cross section both during and after rolling. The mechanical properties of the pipe were determined based on results obtained during the running production. In a subsequent study, Langbauer et al. [8] were able to predict the outer surface temperature of the finish rolled pipes by using an ANN model that indicated whether the post-process temperature limits had been met. These findings were particularly valuable, because the mill operator can change process parameters to meet these limits if a prediction is available early enough and the pipes are still in the reheating furnace.

After the pipes are produced via the process shown in Fig. 1, they are transported to a cooling bed where they can cool down to ambient temperature. Raič et al. [41] investigated the heat transfer that occurred for an array of horizontal pipes due to natural convection and radiation processes. The authors established a correction function for the existing correlation equations that takes into account the distance between the pipes. Raič et al. [42] also investigated the pipe cooling process to determine the optimal position for mixed convection in order to avoid thermal strains. For this determination, the researchers performed a full-scale CFD simulation of the cooling bed, including the pipes and their translative and rotative movement.

2.2. Determination of the optimal number of hidden neurons

In recent years, ANNs have been applied with increasing frequency to solve a wide variety of problems. In the following paragraphs, some publications are cited with respect to the method applied to determine the optimal number of hidden layers and neurons.

The heat transfer characteristics of natural convection processes for a sinusoidal cylinder in a long rectangular enclosure were examined by Seo et al. [26] to improve the total heat transfer performance of this cylinder as compared to a circular cylinder. Three-dimensional direct numerical simulations were carried out to generate the data for this study, and an ANN was trained using these data. The simulation results could be reproduced successfully. Only one year earlier, Huang et al. [27] had presented a method that could be used to predict the transient surface heat flux of a nonlinear heat conduction system with complex geometry online. An ANN was employed to calculate temperatures to solve the forward problem, and the training data were obtained from the offline FEM simulations. The surface heat flux was computed by applying an adaptive, sequential Tikhonov regularisation method and by considering the temperature measurements as the inverse problem. In both of these studies, the architecture of the ANN was determined by varying the numbers of hidden layers and neurons, whereby the architecture with the least error was selected. However, some disadvantages are associated with using this method, namely, the effort required to use it is high because every possible combination of hidden layer and neuron numbers up to an upper limit has to be applied, training every model is time-consuming, and the upper limit is not defined in the literature. The advantage of using it is that every option (combinations of particular number of hidden layers and number of neurons) can be checked and no option can be ignored.

Some researchers have also investigated similar topics using ANNs to model the non-linear behaviour of steel under thermal stresses or thermal load. Toparli et al. [19] analysed what happens when cylindrical steel bars at a temperature of 600 °C are quenched in water. They determined the temperature distribution and the residual stresses resulting from the strong temperature gradients in the bars after cooling by applying the finite element method (FEM). Elastic-plastic deformations were considered in the simulation for pipes of various diameters, and an ANN was trained on the obtained results. Wu et al. [18] also carried out an experimental study that employed steady state tensile tests to analyse two different structural steels and to determine their high-temperature mechanical properties over a temperature range of 20–800 °C. The authors showed how the steel strength decreased as the temperatures increased and observed that plastic deformation occurred. The experimental results were then used to develop an ANN model that could reproduce the data accurately. Rubio-Ramirez et al. [43] predicted that angular distortions that resulted from welding two thin plates of high-strength steel with a backpropagation ANN. The input

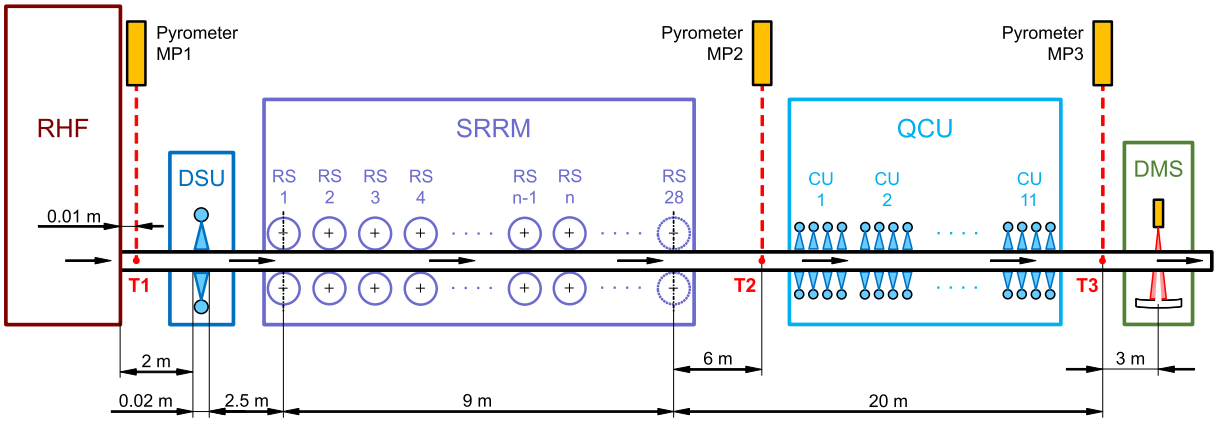

Fig. 1. Specific section of the process used to produce seamless steel pipes examined in this work.

data were obtained from an analytical formulation and an FEM simulation using a thermo-elastic-plastic model. These authors then carried out a sequence of experiments to validate their analytical and numerical results. Liu et al. [44] also investigated a similar welding application, predicting the angular deformation and the transverse, longitudinal, and maximum residual stresses in the welded thin plates by using FEM data to train their model. These authors also detected trends in the mechanical and thermal properties of the investigated steel as a function of temperature. In these four studies, the researchers determined which (optimal) ANN architecture deviated the least from the target values by trial and error, a common method used for this purpose. Only a limited number of arbitrary combinations of particular number of hidden layers and number of neurons is applied. The advantage of using this method is that samples from a large parameter range can be applied with a low amount of effort. Finally, the best model is chosen, i.e. the model which exhibits the lowest error. The disadvantage of using this method is that not all options can be tested, and better solutions remain undiscovered.

ANNs are frequently applied in the steel industry to predict process parameters or evaluate their influence. Xia et al. [45] investigated the influence of rolling parameters on the predictions for rolling power, force, and slip produced by an ANN model. Production data were applied for training, and the ANN architecture was determined by studying a few neurons in the single hidden layer. Zhang et al. [35] also developed an integrated model that could be used to predict the rolling force needed to roll extra-thick plates and that should replace the existing theoretical model. The integrated model consists of an ANN combined with the theoretical model that can be applied to solve the black box problem. The authors used a genetic algorithm to enhance the accuracy of the ANN and obtained training data from the production. The ANN architecture was determined by trial and error, but the component method was applied to limit the range of the number of neurons in the hidden layers. This allowed all combinations of neuronal numbers to be applied on the ANN, and the architecture with the highest accuracy was finally selected. The advantage of using this method is the low effort required due to the limited range of parameters although every option within this range can be checked. The limitation of the parameter range is based on empirical equations. Therefore, one possible disadvantage is that these equations are not suitable for every use case and data set. The literature review revealed that the component method is rarely used.

Many studies on ANN applications have been performed, but it is still difficult to identify the best way to determine the optimal architecture, i.e. the number of hidden layers and the number of neurons in each hidden layer. In 2013, Sheela and Deepa [46] reviewed the different methods used at the time to determine the number of neurons in hidden layers, and no other comprehensive review has been provided since then. No uniform procedure has gained widespread acceptance in the literature, although these parameters are highly relevant for the quality of the ANN results.

Single statistical values like the correlation coefficient (R), the coefficient of determination (R^2), the mean squared error (MSE), the root mean squared error (RMSE), or the mean absolute percentage error (MAPE) are usually used to describe the quality of a model. These statistical values are calculated based on data points in the validation data set to avoid overfitting. The architecture with the lowest error or the highest accuracy is selected. Because the deviations of all data points are specified by a single value, however, information about the location and the corresponding amount of the deviations is not available. Consequently, overfitting cannot be distinguished from a constant offset. Overfitting can also occur between data points as short-wave oscillations, and it is not currently possible to detect this undesirable effect.

Some researchers have gone one step further and plotted the model response surfaces resulting from a variation in two input variables, just as in the present study. In these studies, the plots were used to evaluate trends in the target variable, but not to determine and validate the ANN architecture. A model approach for predicting the tensile strength of a hot-rolled strip was proposed by Kim et al. [47] using experimental data from a strip mill for training. The distribution of the available data points and the model response surface were shown. Bagheripoor and Bisadi [48] built an ANN model that could be used to calculate the rolling torque and rolling force for a strip-rolling process with simulation data. Surface plots revealed the model behaviour depending on these two variables. An ANN model to predict the hot-rolled strip crown was later developed by Deng et al. [49]. These authors studied selected parameters used for the training process as well as the architecture and illustrated the distribution of data points for important variables in the production data used for training. They also provided surface plots of the model response to evaluate trends for the target variable.

The present work differs from the other works cited mainly in terms of the method used to make a final decision about the ANN architecture. Other works selected the architecture solely based on one of the statistical values mentioned before. In the present work, the model response was also checked for overfitting in addition to making a preceding selection based on the lowest error. This procedure enables overfitting to be detected that is not indicated by a statistical value, because the oscillations are not located nearby to the data points. For this detection, the effect of the ANN architecture on the model response was analysed in-depth, considering the distribution of the data points in the multidimensional parameter space. In addition, the findings of this analysis provide a way to establish the limits of the range of application for the parameters. It is highly relevant for ANN models to be trained on production data, because the distribution of the data points may not be uniform, and the risk that the application range may be exceeded is high. This work extends that of many other researchers by providing another procedure that can be used to effectively detect overfitting and, consequently, to determine the optimal ANN architecture.

Moreover, ANNs were applied for the first time in this work to predict the thermal shrinkage of annular pipe cross sections while considering the influence of alloying elements. The change in the pipe diameter that occurs due to thermal shrinkage is low compared to the diameter size itself. Therefore, it is difficult to accurately predict this small relative change by using conventional ANNs. In this work, a novel approach was taken to solve this problem. A more sensitive target variable as opposed to the required variable and a modified ANN architecture were applied to achieve more accurate predictions.

3. Hot-rolling process used to finish seamless steel pipes

In the present work, a focus was placed on a specific section of the production process. This process section is schematically illustrated in Fig. 1. The specific steps required to transform the pipe material from its initial state to an intermediate product, i.e. the tube blank, are not shown in this figure, but are described briefly in Langbauer et al. [8]. The process chain shown in Fig. 1 begins with the natural gas-fired walking beam reheating furnace (RHF). The tube blanks enter this process chain at a temperature of about 500 °C as a result of the previous manufacturing steps before being conveyed into the furnace. In this furnace, the products are reheated to about 950 °C; thus, the tube temperatures are high enough to ensure low deformation forces during the subsequent rolling process. This reheating process takes approximately 7–50 min, whereby the exact time is determined by the geometry and mass of the tube blank, the alloying elements in the steel, and the required mechanical properties of the finished product. Afterwards, the tube blanks are conveyed from the reheating furnace into the rolling mill on a roller conveyor. The surface temperature of the blanks is measured by using a pyrometer located immediately after the outlet at measuring point 1 (MP 1). The furnace control system then evaluates whether the target temperature has been achieved on the basis of these measured values. The tube blanks are illustrated in Fig. 1 as moving from the left to the right, i.e. in the direction of the narrow arrows.

After leaving the RHF, the tube blanks are conveyed through the descaling unit (DSU). At this process step, scale is removed from the

surface of the blanks. This is a vital process step, because the scale would otherwise be pressed onto the surface during rolling. Afterwards, this scale flakes off, violating the size tolerance requirements and resulting in an insufficient final surface finish. In the DSU, this scale is removed by water that jets from nozzles under high pressure and is directed toward the surfaces of the tube blanks. The roller conveyor after the DSU leads to the stretch-reducing rolling mill (SRRM). At this process step, the pipes obtain their final shape, i.e. the final outer diameter and wall thickness are defined. A stepwise reduction in the outer diameter is performed on individual roll stands (RSs). This is necessary, because the degree of reduction is limited to a single RS; this limitation should not be exceeded or profile defects can result. The outer diameter of the tube blank is clearly larger than the outer diameter of the small finish pipes. To achieve these dimensions, a high total reduction in size from the first RS to the last occurs, requiring many RSs, and the reverse is true for large pipes, i.e. a low number of RSs are needed to achieve a low reduction in size. Consequently, the number of RSs required in the SRRM is determined by the pipe dimension to be produced, where the maximum number is 28.

The RSs consist of a casing with three powered rolls that have an angular offset of 120°. The tube blank is pulled through the rolls, and the outer diameter is reduced. The deformation is plastic in nature for the most part, but also partially elastic. The elastic rebound is dictated by the steel alloy used and the temperature of the material during rolling. This is a particularly important aspect to note with regard to the last RS, since the finish diameter in the hot state must fall within a specific tolerance range to meet the target value for the final outer diameter in the cold state.

The powered rolls of the individual RSs rotate at different speeds. Longitudinal stress forces are induced in the tube wall as the tube is pulled apart or compressed, allowing the operator to control the wall thickness of the tubes. Due to the reduction in the tube diameter that occurs during the rolling process and the conservation of mass, the speed of the RSs at the SRRM outlet and the tube velocity are higher than the speed at the inlet. The product is finished after being rolled in the SRRM; at this point, the tube is called a pipe. A pyrometer located at MP2 after the SRRM enables the operator to measure the pipe's surface temperature. This value indicates the product's mechanical properties and shows whether the required quality standard has been met.

After being rolled in the SRRM, the pipe is conveyed to the quick cooling unit (QCU), which is the last process step in the process chain. Eleven (11) individual cooling units (CU) are available for water cooling in the QCU. Water jets are produced by the nozzles on the CUs, and these are switched on to cool the finished pipes abruptly. This cooling process naturally influences the mechanical properties of the material, and no cooling is performed if no specific quality requirements need to be met. The pipe processing is finished after the QCU. The pipe finally passes MP3 on the roller conveyer, where its hot surface temperature is measured; this measured value is also used to evaluate the pipe's quality.

Afterwards, the cross-sectional shape of the pipe's outer surface is determined in the diameter measurement system (DMS). This outer surface is not exactly cylindrical; therefore, the smallest and largest diameters as well as the mean diameter can be determined from the measured shape. The limits for the minimal and maximal diameters are specified in the cold state, but the measurement is carried out in hot state by applying the light-section method, a contactless method in which a ribbon of light is projected onto the pipe surface by a laser. A high-speed camera is used to detect this line by taking images at a frame rate of 500 pictures per second. The pipe's cross-sectional shape can then be determined by evaluating the image data, and the diameter values are provided with a measurement accuracy of ± 0.05 mm. The pipes are conveyed to a cooling bed on a roller conveyer after passing through the DMS, where they cool down to ambient temperature. Finally, the size tolerance requirements must be met in the cold state, i.e. the deviations of the smallest and largest diameters from the nominal diameter have to be less than 1% to meet the requirements of the ASTM A53 quality standard.

Raytek MR1SA pyrometers are used at MP1, MP2, and MP3. These measurement devices exhibit a measurement accuracy of ± 0.5% based on the measured temperature +2 K. An additional error margin of ± 3 K has to be considered, which is caused by the remaining system when performing temperature measurements.

4. Data representation

The data used in the present work were obtained from an industrial hot-rolling process described in Section 3 of this paper. The measurements were performed during the running production, and the data points obtained depended on the product line of the hot-rolling facility. Consequently, the parameter space may have been poorly covered, and the distribution of the data points was of interest. Fig. 2 illustrates the distribution of the data points in three-dimensional parameter space. The three parameters chosen are the input variables 'outer diameter in the hot state', 'temperature', and 'mass fraction of chromium', as explained in the next section. The data points appear as coloured dots in Fig. 2. Their colour shows the diameter difference between the cold and hot states due to thermal shrinkage as a result from experiments. The locations of single data points illustrate the corresponding values of the parameters. The distribution based on only two parameters is drawn across the planes of the parameter space as a projection of the data points. This view is easier to interpret than the three-dimensional representation.

The largest number of data points in Fig. 2 throughout the diameter range of 33.7 mm–202.2 mm are available for objects at high temperatures and with low mass fractions of chromium up to 1%. This area is satisfactorily covered for this data set. Higher values than zero for the mass fraction of chromium are mainly available at temperatures above 800 °C, but for many different diameters. Data points with temperatures lower than 700 °C are only available for individual diameters. In general, the parameter space shown is poorly covered by these data points. Only one area has a dense and thus satisfactory distribution of data points. In the remaining areas of Fig. 2, only a few isolated data points are visible. No data are available for objects with a high mass fraction of chromium and at low temperatures.

Fig. 2 also shows that the highest diameter differences due to thermal

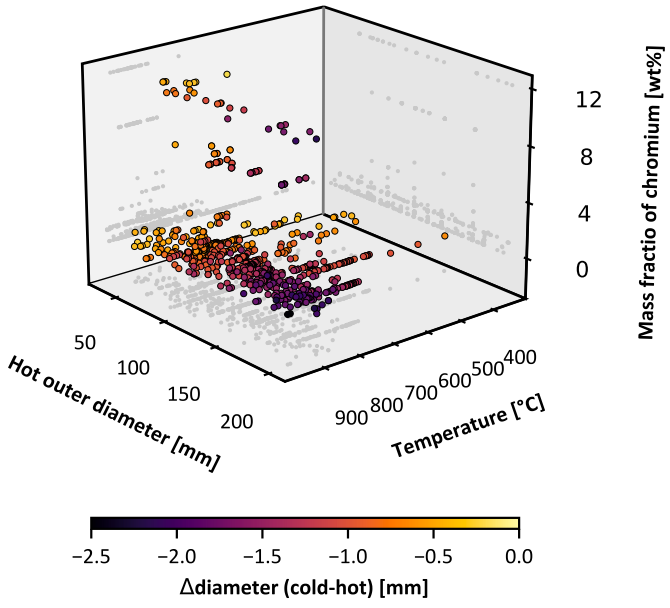

Fig. 2. Distribution of all data points used in the three-dimensional parameter space and diameter difference.

shrinkage occur for high outer diameters and high temperatures, as shown by the experimental results. The change in diameter decreases when diameters are smaller and temperatures are lower. The influence of chromium is not easily visible in Fig. 2 because only a few data points are available for objects with higher mass fractions of chromium.

The entire data set consists of 723 data points. Of these, 80% formed the training dataset subsequently used to train the models. The remaining 20% of the data points represent the validation data set. This validation was performed to indicate whether the data points were overfitted during training by detecting deviations between the predicted values and the target values in this data set. A random algorithm was applied to assign the data points to the data sets for training and validation purposes.

5. Artificial neural network models

5.1. Input and target variables

In seamless steel pipe production, the size tolerance requirement of 1% needs to be met in the cold state or the pipes will be rejected. A value should be calculated that enables the user to compare this value with the limits; therefore, the outer diameter of the pipes in the cold state was defined as a target variable in the ANN models. The target values used to train the models are the measurement values from the DMS, which were obtained by moving the finished pipes in the cold state through the DMS for a second time. This approach was taken, because the measurement method applied by the DMS (i.e. applying light) provides accurate results and provides data on the local cross-sectional shape of the outer surface along the pipe's length. The minimal, maximal, and mean diameters could be determined, and, in this work, the mean diameter was used as the target variable.

The selection of the input variables has to be performed carefully. If important variables are not selected, the process will be described incompletely, and the results for the target variable will deviate strongly from the measured values. Wrong trends and bad results are to be expected when using unimportant input variables [8]. For this reason, only three variables were selected as input.

The first input variable selected is the outer diameter in the hot state as measured in the DMS in mm. The minimum value in the data set is 33.7 mm, and the maximum value is 202.2 mm. This value has the highest influence on the results, because the diameter in the cold state is only slightly smaller than it is in the hot state, i.e. the change in the diameter due to thermal shrinkage is low as compared to the overall outer diameter. Precisely this difficulty is encountered when attempting to accurately predict this small relative change.

The next input variable selected is the temperature at MP 3 in Fig. 1 in °C. The measured surface temperature of the pipe is approximately the same as its temperature during the diameter measurement (i.e. previously measured temperature); therefore, this indicates the temperature in the hot state before thermal shrinkage occurs. The values used a temperature range of 344–965 °C. These provide information about the temperature difference between the two states of interest in the pipe and are related to the difference in diameters due to thermal shrinkage.

The last input variable selected is the mass fraction of the element chromium (Cr) in wt% in the steel alloy used. By choosing this variable, it is possible to take into account the specific expansion coefficient of the material. This is considered to be an important aspect, because the shrinkage behaviour of steel alloys with chromium has been observed in operations to differ from the behaviour of steel alloys without chromium. The minimum and maximum values in the data set for the mass fraction are 0 wt% and 12 wt%, respectively. The advantages of selecting and applying fewer input variables are that the data analysis becomes much easier and the amount of effort that must be invested is reduced. By using only three input variables, the model behaviour can even be plotted.

The wall thickness of the pipes was initially considered as an input variable due to its inferred important influence on thermal shrinkage; however, no significant enhancement of the results or the validation loss was observed. Therefore, the variable of the wall thickness can be neglected without suffering a loss in terms of accuracy. The choice of three variables ensures a meaningful graphical illustration of the data and the model response in three-dimensional plots. Displaying more than three input variables is significantly more difficult.

5.2. Artificial neural network models

In order to achieve accurate results for the outer diameter of the finished pipes in the cold state, two different types of models were developed. These two types of models respectively consist of a single shallow feed-forward ANN with a linear activation function in the output layer. The three input variables 'outer diameter in the hot state', 'temperature', and 'mass fraction of chromium' mentioned previously are applied. The number of neurons and the activation function used in the single hidden layer define the specific models used for prediction. These two parameters were determined as the result of the parameter study described in Section 6.1. The best setup was selected on the basis of the least validation loss during the model training; however, the two types of models basically differ in terms of the target variable applied.

The first model type is used to directly predict the outer diameter of the cold pipe as the target variable. The applied specific model of this type is referred to as 'model 1', and its architecture is shown in Fig. 3. Model 1 has 110 neurons in its hidden layer, and the sigmoid activation function was selected. This model is a conventional ANN.

The second model type is used to calculate the difference in the pipe diameters between the cold and the hot state based on the neuron in the output layer, i.e. the target variable of the ANN is the difference in the diameters. This predicted value is added with the input value for the outer diameter in the hot state, and the diameter in the cold state is calculated as a result. The design and method of calculation for this model type is shown in Fig. 4. The training process is performed only for the ANN by using the difference in the diameters as a target variable. The addition step is not included in this process and is only performed for calculations. In this work, two specific models were developed based on the second model type, which differ only in terms of the number of their neurons and the activation function. The first new model based on the second model type is referred to as 'model 2', and its architecture is shown in Fig. 4. This has 40 neurons in its hidden layer, and the ReLU (Rectified Linear Unit) activation function is used. The second new model based on the second model type is referred to as 'model 3'. This

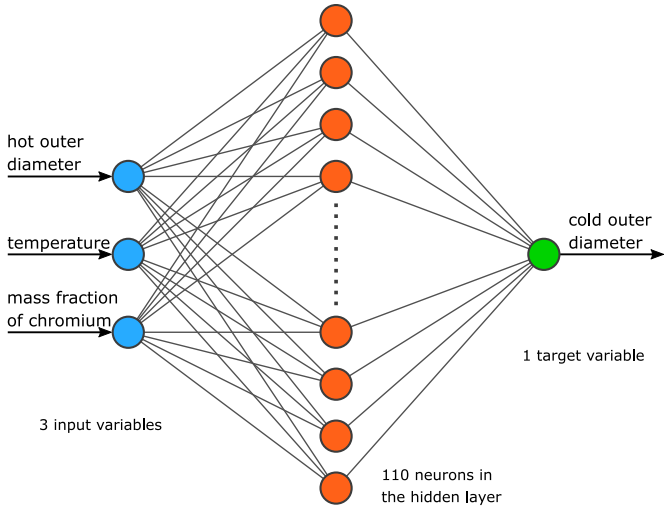

Fig. 3. Design of model type 1 and architecture of 'model 1'.

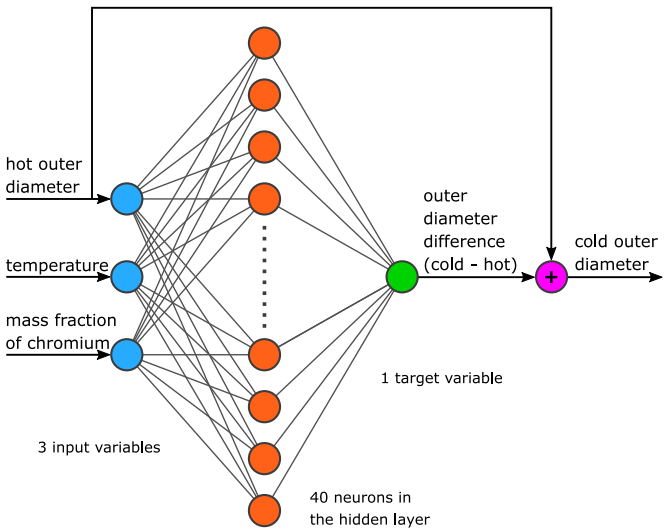

Fig. 4. Design of model type 2 and architecture of 'model 2' and 'model 3' with only 10 neurons.

has 10 neurons, and the sigmoid activation function is used. The architecture of model 3 appears to be equivalent to that of model 2 as shown in Fig. 4, but it differs in terms of its number of neurons.

The Adaptive Moment Estimation (Adam) [50,51] was employed as an optimiser for model training. The mean squared error (MSE) was used as a loss function, which has to be minimised to achieve a low loss after training. The MSE is calculated according Eq. (1) [52]. y_i is a target value from the measurement, and f_i is the corresponding predicted value of a particular data point i. n is the number of all data points in the data set used. The applied training algorithms and modules for constructing the ANNs are available in the end-to-end, open-source machine learning platform 'TensorFlow'. The source code was developed in the program language 'Python' for this work.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2 \tag{1}$$

6. Results and discussion

First, a study was performed to determine the number of neurons and the activation function in the hidden layer in order to identify the best setups for the two model types. A setup is defined as a combination of the number of neurons and the activation function in the hidden layer. Once these setups had been identified, the models described in Section 5.2 could be defined by applying the best setups for the specific model type. These models were then employed to predict the cold outer diameter of pipes after they were hot-rolled and cooled to ambient temperature (cold state). Training data and validation data were used as input data for the prediction, and model response surfaces were generated. A model and, therefore, a particular setup could finally be selected after the model response had been checked for overfitting. This is an important step, because overfitting may be not evident from the statistical values (e.g. MSE, R^2), if no data points are located nearby. The results of this study enabled the identification of the best model for the present purpose. These are compared and discussed below.

6.1. Study of different ANN architectures

The input and output variables of the ANNs were defined by considering the present problem, i.e. to accurately predict the cold outer diameter of steel pipes. The number of neurons and the activation function in the hidden layer can be selected freely. Therefore, a study of these two parameters was performed to identify the best setup, i.e. the one that resulted in the least loss, and use this setup to perform further calculations. The validation loss was chosen as the indicator for the evaluation, because this also provided information about the model performance in terms of predictions based on data that were not included in the training data set. The validation loss is the mean squared error (MSE) of the results for the validation data set. The training loss is not shown. The values for the training loss are always lower than those for the validation loss, but the order of magnitude is the same.

In this study, several models with different setups were trained to obtain the validation loss. The number of neurons in the hidden layer varied from 10 to 200 neurons in steps of 10 neurons, respectively. Basically, the step size needed to be equal to one in order to consider all possible numbers within a specified range. However, the step size was not set to one due to the high amount of effort that would need to be invested due to the high number of neurons needed to achieve a low loss; however, the step size of 10 is sufficient for illustrative purposes. The sigmoid, tanh (hyperbolic tangent), ReLU (Rectified Linear Unit), LReLU (Leaky Rectified Linear Unit), ELU (Exponential Linear Unit), and SELU (Scaled Exponential Linear Unit) activation functions are available in TensorFlow. One of these was applied in combination with a number of neurons as a single setup to train the models based on model types 1 and 2, see Figs. 3 and 4. This training process was executed for 2000 epochs. The epoch with the least validation loss was subsequently identified, and the training run was repeated up to that epoch. This iteration method ensured that training would stop at the optimal epoch. The final validation losses for the variations in terms of the number of neurons when applying one specific activation function are plotted in Fig. 5 as lines. The least MSE of a line is indicated by a circle and a horizontal and a vertical dotted line. However, this figure only shows data from the study that were relevant for the final selection of the best model setups. In general, the MSE of the points shown in Fig. 5 is low, as is revealed by the ordinate scaling. Low training and validation losses indicate the high quality of the training dataset used, and these are obtained from measurements. This means that the measurement inaccuracy is low because an accurate optical measurement system was used. Langbauer et al. [8] clearly showed a higher MSE in another study, and this was significantly affected by the measurement inaccuracy due to the high temperatures.

The least validation loss for model type 1 was observed for 110 neurons when using the sigmoid activation function. The validation loss fluctuated when a lower number of neurons was included, but an

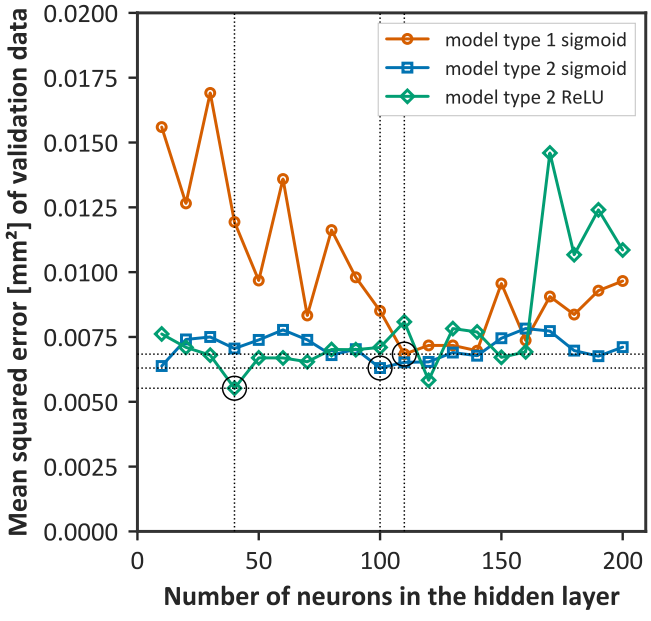

Fig. 5. Validation loss of different ANN architectures.

examination of the curve reveals an upward trend. Including a higher number of neurons also led to an increase in the validation loss. Consequently, the best setup for model type 1 was detected as having 110 neurons and applying the sigmoid activation function in the hidden layer; this version of model type 1 is referred to as 'model 1' (Fig. 3).

Model type 2 showed the best performance when 40 neurons were included and the ReLU activation function was used. This setup is referred to as 'model 2' (Fig. 4) based on model type 2. The validation loss fluctuates within the investigated range, but is approximately on the same low level; the loss is only clearly higher when 170 or more neurons are included. The second variation based on model type 2, i. e model 3, using the sigmoid activation function is also shown in Fig. 5. The curve fluctuates slightly on the same low level. Low losses are already visible when a low number of neurons is included; therefore, this curve was further investigated. The findings indicate that model type 2, i.e. models 2 and 3, is more suitable for addressing the problem addressed in this work, because the average MSE of the two curves for the validation loss is lower and the losses appear to be approximately on the same low level. These curves also appear to be smoother than the curve seen for model type 1. The lowest validation loss for model type 2 using the sigmoid activation function was obtained when including 100 neurons, but the MSE when including 10 neurons is almost equal. Including a lower number of neurons has an advantage when predicting new data, an aspect which is discussed below. On the basis of these study findings, the setup with 10 neurons and the sigmoid activation function was applied in model 3.

Model 2 displays the lowest minimal validation loss (0.0055227 mm^2, see Fig. 5), model 1 displays the highest minimal loss (0.0068377 mm^2, see Fig. 5), and model 3 shows a loss of 0.0063016 mm^2, i.e. a value between the losses displayed by models 1 and 2.

6.2. Prediction results

Predictions were performed with models 1–3 using the training and validation data set as input. Fig. 6 shows the results of reproducing the target values from the training data set in a) and from the validation data set in b) when applying model 1. The results for models 2 and 3 appear identical to those shown in Fig. 6; therefore, they are not displayed. In Fig. 6, the single predicted values for the cold pipe diameter are drawn as a function of the corresponding measured values. The diagonal line indicates the ideal case, i.e. if the predicted value is equal to the measured value. The predicted values near this line represent accurate predictions. The 1% error margin is also drawn in these figures, and this is located near the diagonal due to the large range of pipe diameter values. The points between the two lines deviate by less than ± 1% from the measured value. A deviation of 1% from the specified target value is also the quality control limit set for seamless steel pipes according to the ASTM A53 material standard. In the model to be applied, the measured values should deviate by significantly less than 1% from the predicted values for a meaningful prediction to be ensured. A linear regression analysis was also performed for all results to evaluate their quality. The regression line is obtained using the linear least-squares approximation, shown as a dashed line in Fig. 6. The coefficient of determination (R^2) is calculated by Eq. (2). y_i is a target value from the measurement, and f_i is the corresponding predicted value of a particular data point *i*. n is the number of all data points in the data set used. $\bar{y}$ is the average value of the n target values. The value of R^2 is shown in the upper left-hand corner of Fig. 6.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - f_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{2}$$

The results from all models are highly accurate, as shown in Fig. 6. The data points are located near the diagonal in the plot, and no scatter is visible; no outliers are detected. All points fall within the 1% error margin, even for pipes with small diameters. Consequently, all three models are deemed suitable for the application to meet the required quality standards. The regression line is almost congruent with the diagonal line, and the coefficient of determination (R^2) is equal to 1.

Because the predicted data points are in excellent agreement with the measured data points, deviations are difficult to see in Fig. 6. The relative deviations, however, are shown in Fig. 7. These are calculated as the absolute values of the difference between the predicted and the measured cold pipe diameter in mm divided by the measured cold pipe diameter in mm. The relative deviations were assigned to the most appropriate of the 10 equal bins by setting a range of 0.05%, e.g. the values of 0.17 and 0.2 were assigned to the bin '0.15–0.20'. Values equal to the upper limit of the bin were also included in this bin. All values higher than the last specified value of 0.5% were allocated to the bin '> 0.5'. These bins are drawn along the horizontal axis shown in Fig. 7. The relative frequency (rel. freq.) of the values in the bins is plotted along the left-hand vertical axis, and the corresponding distribution function (distr. func.), on the right-hand side; these represent the cumulative sum of the relative bin frequencies. Fig. 7 is similar to a histogram in that it displays the absolute values for the deviations. No information is lost when using the absolute values, because the distribution of the frequency in the positive and negative directions for the relative deviations

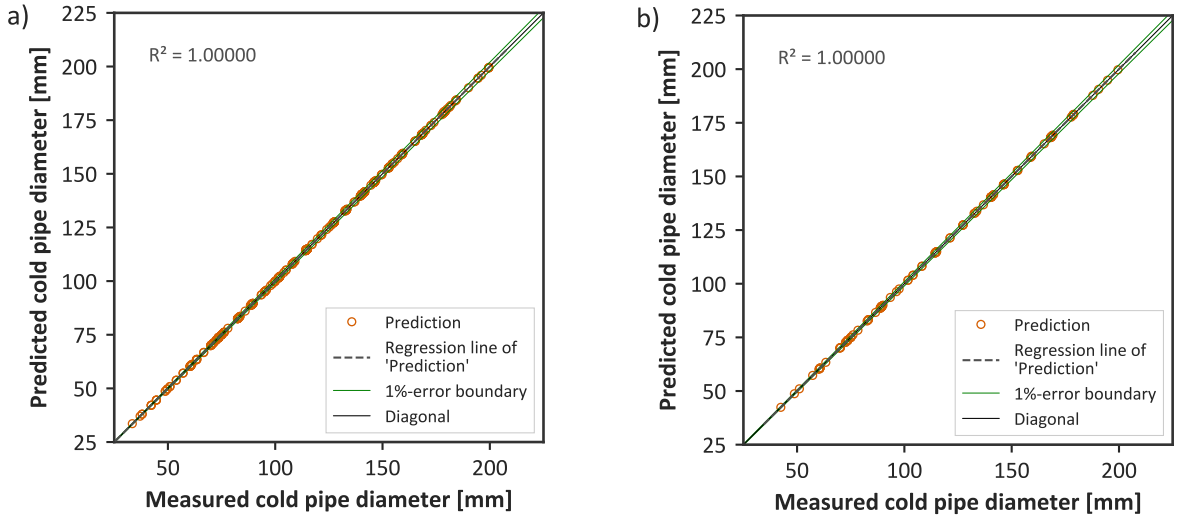

Fig. 6. Model 1 results: predictions based on a) training data and b) validation data.

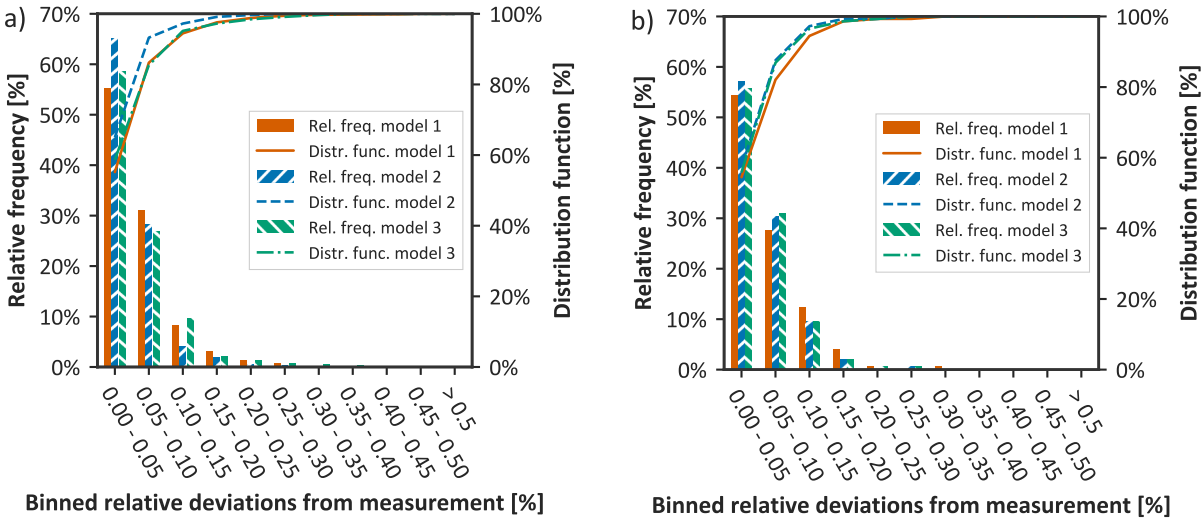

Fig. 7. Comparison of models 1–3 in terms of the relative deviation using the a) training data and b) validation data as input data.

is symmetric.

The good performance is seen in Fig. 7, illustrating how accurately the target values for the training and validation data are calculated by using the three models. The significant points are discussed in detail by referring to the example of the results from model 1 and using the training data set, as shown in Fig. 7 a). The values for the other models and data sets are presented in Table 1.

Of all data points predicted by model 1, 86.159% deviated by less than 0.1% from the measured values available in the training data set. This is indicated by the relative frequency of the first two bins and the distribution function shown in Fig. 7 a). Deviation rates less than and equal to 0.1% are considered particularly, since these are significantly lower than the acceptable deviation of 1% listed in to the ASTM A53 material standard. The relative frequency of data points below and equal to the limit of 0.1% shows how suitable a model is for industrial application. Predicted values with a relative deviation of at most 0.154% are included in the double standard deviation 2σ ($2\sigma = 95\%$). This can be seen at the point along the horizontal axis in Fig. 7 a) where the distribution function exceeds the value of 95%. Most of the data points do not deviate by more than 0.154% from the measured values. The highest deviation observed is 0.48243%, but the relative frequency for deviation rates higher than 0.3% is almost zero (Fig. 7 a). These values are also provided in Table 1.

All three models provided accurate results, and no relative deviation rate higher than 0.5% was detected, indicating the high quality of the data. Applying model 2 resulted in the lowest deviation rates from the measured values for both data sets. In Fig. 7 a), the curve of the distribution function for the training data was clearly higher than the others from the beginning (values shown in Table 1), and the point of the double standard deviation 2σ is at a lower deviation rate of 0.109%. The other two curves for model 1 and model 3 are almost congruent. The function values differ only insignificantly, but model 3 has a lower maximal deviation rate of 0.3842%. Using this model allows the user to achieve at least the same accuracy as model 1 when using only 10 neurons in the hidden layer. The two curves are similar to the curve seen for model 2 only from a relative deviation rate of 0.25% and above.

The difference among the three models is smaller when analysing the validation data, as seen in Fig. 7 b). Model 2 is still the best, but the distribution function for model 3 is converging. For deviation rates of up to 0.15%, the values are highly similar. Model 2 meets the double standard deviation 2σ earlier at a rate of 0.14%; however, model 3 has a lower maximum relative deviation rate of 0.25173%. Model 1 shows the poorest performance. The curve obtained by applying this model is similar to the other curves only at the point where the deviation rates are less than or equal to 0.2%. The final value of 100% for the distribution function is also met at higher maximum deviation rates of 0.32075%. The value of the maximum deviation seen when applying model 1 is the highest out of all three models.

In general, models 2 and 3, i.e. models based on model type 2, produced more accurate results than model 1 when using training and validation data. Therefore, model type 2 is more suitable for solving the current problem, as indicated by Fig. 5. These findings indicate that the optimal approach is to apply the difference in the pipe diameters (cold and hot states) as the target variable, because this is the more sensitive variable. The target values are two orders of magnitude lower than the cold diameter of the pipes predicted when applying model type 1. Therefore, the relative change in the output values when different input values are used is higher; thus, it may be easier for the optimiser to train the model. The change in diameter due to thermal shrinkage relative to the cold diameter is small; therefore, the relative change in the output values when using different inputs is also small.

The highest level of agreement between modelled and measured values is usually achieved when using training data, because the models are trained with exactly these data. The accuracy decreases when data sets include data points for unknown values or unknown combinations of values that are not included in the training data, i.e. the validation data set. This behaviour was observed for models 1 and 2. Model 3 provides about the same accuracy for both data sets; hence, its performance is higher than that of the other models based on validation data.

Table 1
Values for significant points shown in Fig. 7

Model	Dataset	Value of the distribution function for relative deviations ≤0.1%	Relative deviations within the double standard deviation 2σ	Maximum relative deviation
model 1	training	86.159%	≤0.154%	0.48243%
model 2	training	93.253%	≤0.109%	0.29617%
model 3	training	85.467%	≤0.148%	0.3842%
model 1	validation	82.069%	≤0.152%	0.32075%
model 2	validation	87.586%	≤0.14%	0.26592%
model 3	validation	86.897%	≤0.147%	0.25173%

The lower number of neurons in the hidden layer may explain this. The number of neurons serves as a measure of the degrees of freedom which still allow the ANN to achieve the target values in the training process. If the number of neurons is too high, a behaviour approaching overfitting occurs. The error for the predicted data point values in the training data consequentially decreases. The error for the unknown data points increases, and the model progressively loses its correct deep knowledge about the system. This results in high deviations at points within the multidimensional parameter space where no data points are available for training. These findings support the hypothesis that ANNs with lower numbers of neurons are preferred, as long as the loss is sufficiently low, to prevent overfitting and incorrect behaviour at poorly or uncovered locations within the multidimensional parameter space. This hypothesis is also further supported by the model response surfaces resulting from the model applications, as described in the next section.

This work resulted in the development of a highly accurate ANN model that can be used to predict the cold diameter of seamless steel pipes after thermal shrinkage has occurred, based on measurements of pipes in their hot state after the hot-rolling process. This model represents a novel finding in the research on industrial processes. The outcome was made possible by using a more sensitive variable as a target variable rather than directly using the required variable. Performing a subsequent arithmetic operation enables the user to obtain the required variable. An ANN model with only 10 neurons in the hidden layer that can be used to provide accurate results for both training and unknown data can be developed by using the approach described in this work.

6.3. Model responses

ANNs are highly complex constructs, and the multiple connections between the neurons in the layers make it difficult to understand their behaviour. A small change in a single input variable can lead to significant changes in the output, since ANNs can be used to represent strong nonlinear trends. However, a method has been used to plot the model behaviour. Two input variables are varied within a specified range represented by a grid defined on a plane by data points for the two input variables. A constant value is assigned to the remaining input variables. The resulting data set is used as input data for the ANN. The output is the model response, which can be plotted as a surface plot. The two varied input variables are displayed along the *x*- and *y*-axes across the base

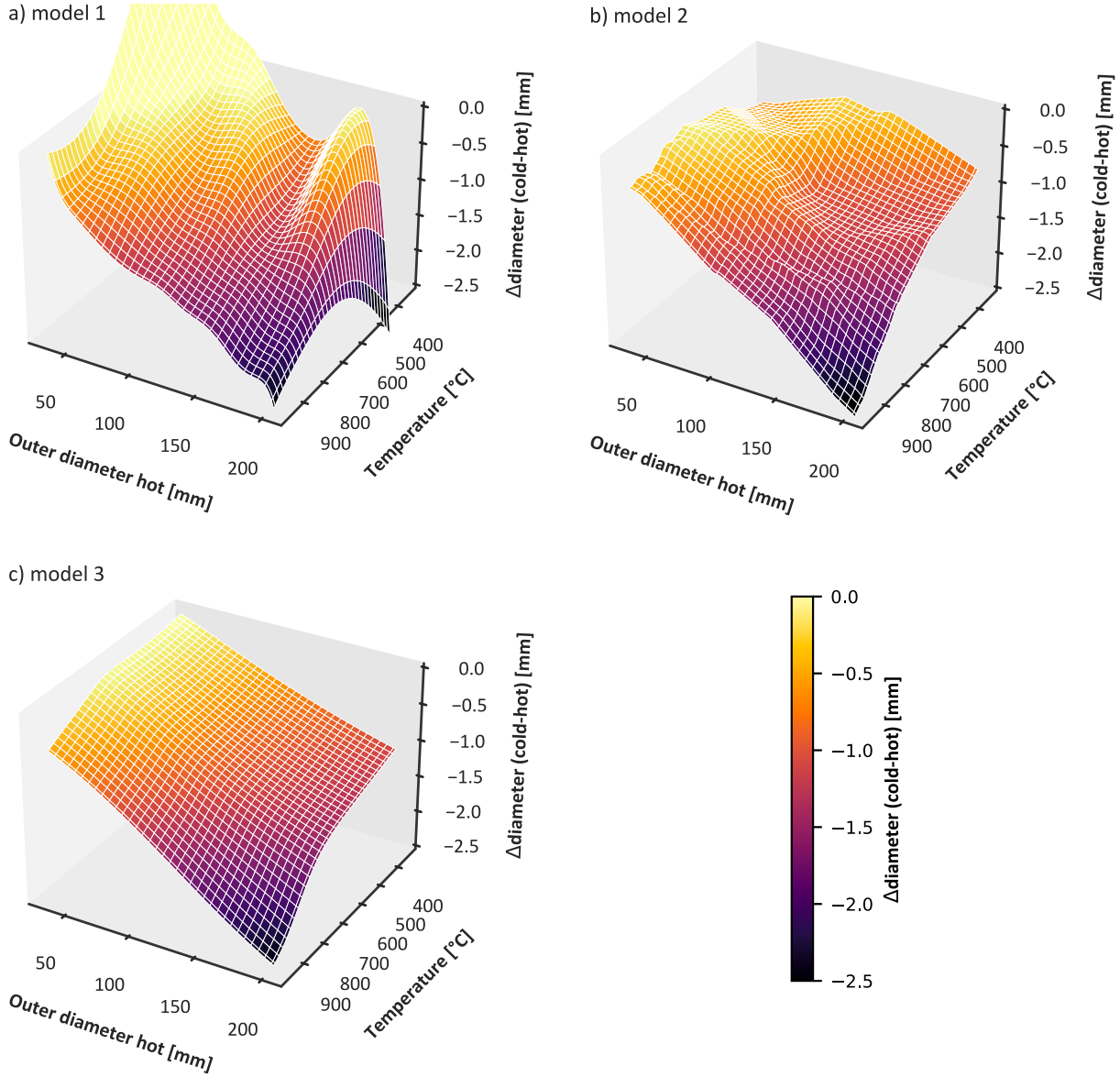

Fig. 8. Model responses for 1 wt% Cr when applying a) model 1, b) model 2, and c) model 3.

area. The values for the target variable of the ANN are plotted along the vertical *z*-axis as a surface. Providing such model responses in the form of surface plots can help users to gain insight into the behaviour of complex ANNs.

This insight can only be gained, however, for ANN models with more than three input variables, because the variables with an assigned constant value that are not used for the variation also influence the output. As all input variables are related to each other, it becomes steadily more difficult to display and understand the model responses. Validly interpreting these responses is only possible for a small, restricted area of the multidimensional parameter space. In the present case, only three input variables were used, and, thus, a meaningful analysis of the model response was possible.

The input variables 'hot outer diameter' and 'temperature' were varied against each other within a specified range. The model responses from the three models using these input data are shown in Fig. 8 for objects with a mass fraction of chromium of 1 wt%. The hot outer diameter and the temperature are visible along the two axes in the lower part of the figure. The difference in diameters between pipes in the cold and the hot state represents the model response; this is drawn along the vertical axis as a surface. The surface colours reveal the functional values. Originally, the models predicted the cold outer diameter as a target variable, but the difference in diameters was selected for these plots in order to more effectively visualise the effect.

In Fig. 8 a), the model response of model 1 (architecture shown in Fig. 3) is displayed. The functional values for temperatures above 700 °C appear plausible, but the highly uneven nature of the surface at lower temperatures is not expected. The increase in the difference in the diameter (cold vs. hot) and the sharp drop in this difference at low temperatures and with large diameters are not realistic. The values for the difference in diameter become positive at low temperatures and with low diameters, as shown in Fig. 8 a). This is also realistically impossible, because steel always shrinks during the cooling process, and the difference values are negative. The missing data points at these locations within the three-dimensional parameter space are responsible for the improbably physical trends shown for the surface. Fig. 2 provides information about the local coverage of the three-dimensional parameter space. No data points for any fraction of chromium are available for training in these areas; thus, the model is unable to learn about specific behaviour there. Predictions performed at these locations are extrapolations, and these should be avoided when applying ANNs, as they may lead to completely wrong results. The trends observed at the well-covered locations are acceptable.

The model response for model 2 (for architecture see Fig. 4) shown in Fig. 8 b) appears much more planar, although small waves are still visible. The maximum shrinkage, i.e. the minimum values, is located at large diameters and high temperatures. This result is plausible, because the difference between the cold and hot diameters increases as the pipe diameter and temperature in the hot state increase. The surface gradient becomes almost zero within the area showing low diameters and temperatures. This behaviour is not realistic, however, because thermal shrinkage always occurs. The model response seen for model 2 in Fig. 8 b) displays mostly correct trends, but the small oscillations seen have no basis in physical reality.

The surface of the model response seen for model 3 (for architecture, see Fig. 4) in Fig. 8 c) is smooth, as would be expected for the present case. The lowest difference is located correctly, i.e. at high diameters and temperatures. Starting from this point, the surface shows a monotonically increasing trend in every direction. The maximum value appears in the opposite corner. The values for the difference between the cold and hot diameter decrease approximately linearly as the hot outer diameter increases. This result makes sense, because thermal shrinkage is a linear process for planar bodies. Deviations from the linear trend observed may be due to the curved shape of the pipe. The difference decreases also approximately linearly when temperatures are increased up to 700 °C; above this temperature, a buckle and a subsequent steeper drop can be seen. The model response displayed for model 3 in Fig. 8 c) clearly shows correct trends based on the physical mechanism of thermal shrinkage. Consequently, model 3 is the most suitable model of all presented models for the present case and provides the best model response.

Model 1 shows a highly uneven surface in Fig. 8 a) due to the poor or missing coverage of the three-dimensional parameter space. Significantly smaller waves can be seen in Fig. 8 b), and applying model 3 provides the expected smooth surface with the correct physical trends, but the same data were used to train all of the models. In addition to the poor coverage of the parameter space mentioned above, a second influencing factor was identified, which is also discussed in Section 6.2: the number of neurons in the hidden layer. This number is a measure for the adaptability of the ANN. A high number enables the ANN to provide a good approximation of the target values based on the training data, because the optimiser has many options to adjust weights and bias. Using a high-order polynomial function for curve fitting is a comparable example. The increasing probability of overfitting is one disadvantage of including a high number of neurons. This overfitting does not depend on the number of epochs used in the training process. Using a high number of neurons enables excessive flexibility, which can result in oscillations in the model response and high amounts of deviation from the physical trends for poorly or uncovered locations within the parameter space, since no data points are available to determine the model behaviour. These considerations are also valid regarding extrapolations. Only 10 neurons were used in model 3: This led to a smooth surface, and model 3 seems to also be suitable for extrapolating values over a limited range. The target values from the training data cannot be achieved exactly when using such a low number of neurons, but the physical trends can be represented more accurately. Thus, the findings of the present work show that model 3 provides more accurate results for unknown data, as shown for the validation data set described in Section 6.2, because this model does not tend to predict physically impossible behaviour at poorly covered locations, and no oscillations occur due to overfitting. Consequently, model 3 provides better results for the model response than model 2, although the validation loss after training was slightly higher. The results at well-covered locations are acceptable for all models, expect for the oscillations.

The applied data described in Section 4 were obtained from the rolling process presented in Section 3. All input and target variables, except the mass fraction of chromium, were determined by performing measurements. These measured values contain different kinds of measurement uncertainty; therefore, it is not mandatory to exactly meet the target values associated with the training data. A lower number of neurons in the hidden layer should be preferred if the results for the validation data are almost the same. The validation data represent unknown data that are not applied in the training process. Obtaining a good approximation of these data points is more important to ensuring high model quality. For this reason, the validation loss after training and the predicted results using the validation data as input should be used to evaluate the models. In this case, deviations from the physical trend occur at some locations of the model response. Many data points were generated during the variation in the two input variables to accurately calculate the model response. These data points are also located in the poorly covered or uncovered space, but no data points found in the validation data set are located in this space. Therefore, deviations from the physical trend are not visible in the validation loss or in the results predicting the validation data set.

The poor or missing coverage at some locations within the multidimensional parameter space is another difficulty associated with using measured data from an industrial process. Poor coverage results from the product line of the hot-rolling facility. Some combinations of parameter values are not applied, because the resulting product is not demanded by the customer. Some other combinations are technically not possible with the given process. Therefore, some locations within the parameter space cannot be assigned data points. The poorly covered

locations are not a problem as long as no prediction is performed for this location. Problems can be avoided by systematically generating data points by performing specific experiments, but this procedure incurs high costs and requires a large amount of effort.

In this work, the effects of the number of neurons in the hidden layer and the data used on the model response were analysed. The study findings make an important contribution to the scientific community, because they provide researchers and operators with a method for determining the most suitable ANN architecture and preventing overfitting. Until now, no generally accepted solution had been found for this problem. In addition, the distribution of the data points used and the effect of this distribution on the model response were considered, aspects that had rarely been considered before.

7. Conclusion

Manufacturers of seamless steel pipes make every effort to fulfil the requirements regarding size tolerance in material standards and, thus, to avoid rejects. Fast models are needed to provide accurate predictions already during the production phase. In this work, an artificial neural network (ANN) model is proposed to achieve this purpose. The model can be applied to predict the final pipe outer diameter in the cod state at ambient temperature based on the hot state immediately after rolling. However, small relative changes in the diameter are currently difficult to predict when using conventional ANNs. Therefore, the diameter difference (cold-hot) was used as target variable instead of the cold diameter to solve this problem, because the diameter difference is more sensitive with respect to small relative changes. By taking this approach, it was possible to achieve a lower validation loss during the training process and to obtain improved results based on calculations using the training and validation data set as input. Out of all calculated data points, 95% show relative deviations from the measured values by at most 0.148% for both data sets. The development of a new ANN model that uses a more sensitive target variable to accurately predict the cold diameter of seamless steel pipes is the first novel finding of this work.

Moreover, to obtain accurate results while preventing overfitting, the most suitable number of hidden neurons in the model had to be determined. No generally accepted solution to these problems had previously been proposed in the literature. Therefore, the effect of the number of hidden neurons on the model response was investigated while also considering the distribution of the data points used. This investigation allowed the optimal number of hidden neurons to be determined, which is 10 (model 3) in the present case. The presented procedure is the second novel finding of this work, as it can be applied by other researchers to determine the most suitable number of hidden neurons in other models. This procedure is recommended for validating ANN models because it provides more comprehensive information about the model behaviour than a single statistical value (e.g. MSE, R^2). Consequently, this paper significantly supplements other current research studies on the applications of ANNs.

Credit authors statement

Raphael Langbauer: Methodology, Software, Validation, Formal analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Georg Nunner:** Resources, Writing - Review & Editing, Project administration. **Thomas Zmek:** Data Curation. **Jürgen Klarner:** Supervision. **René Prieler:** Writing - Review & Editing. **Christoph Hochenauer:** Resources, Writing - Review & Editing, Supervision, 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 authors do not have permission to share data.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] V. Gopinathan, O. Pawelski, V.C. Venkatesh, Effect of cold and hot rolling and normalising on the structure and properties of welded joints, J. Mech. Work. Technol. 1 (1978) 361–370, https://doi.org/10.1016/0378-3804(78)90038-4.
[2] E. Korkmaz, C. Meran, Mechanical properties and microstructure characterization of GTAW of micro-alloyed hot rolled ferritic XPF800 steel, Engineering Science and Technology, Int. J. 24 (2021) 503–513, https://doi.org/10.1016/j.jestch.2020.04.006.
[3] O.M. Awe, S.T.A. Okolie, O.S.I. Fayomi, Analysis and optimization of water distribution systems: a case study of Kurudu post service housing estate, Abuja, Nigeria, Results in Engineering 5 (2020), 100100, https://doi.org/10.1016/j.rineng.2020.100100.
[4] M. Barnetche, L.F. González-Portillo, J. Muñoz-Antón, R. Abbas, M. Ibarra, R. Barbero, A. Rovira, Analysis of the thermal inertia of pipelines in SHIP, Results in Engineering 17 (2023), 100908, https://doi.org/10.1016/j.rineng.2023.100908.
[5] T. Joshi, O. Parkash, A.A. Murthy, G. Krishan, Numerical investigation of Bi-model slurry transportation in a straight pipe, Results in Engineering 17 (2023), 100858, https://doi.org/10.1016/j.rineng.2022.100858.
[6] A.K. Rostami, D.D. Ganji, Selecting superior fin geometry among four suggested geometries for shell and helically coiled finned tube heat exchangers with numerical simulation and experimental validation, Results in Engineering 17 (2023), 100867, https://doi.org/10.1016/j.rineng.2022.100867.
[7] S. Erne, G. Scheger, W. Wiedemair, Numerical and experimental investigation of surface-stabilized combustion in a gas-fired condensing boiler, Results in Engineering 17 (2023), 100738, https://doi.org/10.1016/j.rineng.2022.100738.
[8] R. Langbauer, G. Nunner, T. Zmek, J. Klarner, R. Prieler, C. Hochenauer, Development of an artificial neural network (ANN) model to predict the temperature of hot-rolled steel pipes, Advances in Industrial and Manufacturing Engineering 5 (2022), 100090, https://doi.org/10.1016/j.aime.2022.100090.
[9] T. Świątkowski, S. Kalisz, J. Wnorowska, An innovative conversion of a district heating water-tube boiler for cogeneration of electricity and heat, Results in Engineering 13 (2022), 100350, https://doi.org/10.1016/j.rineng.2022.100350.
[10] Z. Li, C. Zhang, G. Song, Research advances and debates on tubular mechanics in oil and gas wells, J. Petrol. Sci. Eng. 151 (2017) 194–212, https://doi.org/10.1016/j.petrol.2016.10.025.
[11] R. Subbiah, M. Rahel, A. Sravika, R. Ambika, A. Srujana, E. Navya, Investigation on microstructure and mechanical properties of P91 alloy steel treated with normalizing process - a review, Mater. Today Proc. 18 (2019) 2265–2269, https://doi.org/10.1016/j.matpr.2019.07.008.
[12] P. Singh, U. Batra, S. Sangal, Fracture toughness behavior of 38MnSiVS5 microalloyed steel after isothermal transformation and thermomechanical processing, Mater. Today Proc. 4 (2017) 8528–8537, https://doi.org/10.1016/j.matpr.2017.07.199.
[13] R. Langbauer, G. Nunner, T. Zmek, J. Klarner, R. Prieler, C. Hochenauer, Investigation of the temperature distribution in seamless low-alloy steel pipes during the hot rolling process, Advances in Industrial and Manufacturing Engineering 2 (2021), 100038, https://doi.org/10.1016/j.aime.2021.100038.
[14] J.Z. He, J.N. Lu, X.Y. Deng, X.Q. Xing, Z.C. Luo, Premature fracture of high-strength suspension springs caused by corrosion fatigue cracking, Results in Engineering 16 (2022), 100749, https://doi.org/10.1016/j.rineng.2022.100749.
[15] S. Huzni, I.B.M. Ibrahim, S. Fonna, R. Pidaparti, Physics-based surrogate model for reinforced concrete corrosion simulation, Results in Engineering 16 (2022), 100659, https://doi.org/10.1016/j.rineng.2022.100659.
[16] B. Harish, D. Satish Kumar, A novel statistical investigation on tensile property of high alloy steel (EN31) and carbon steel (EN09), Mater. Today Proc. (2022), https://doi.org/10.1016/j.matpr.2022.08.286.
[17] D. Wang, Q. Zhong, J. Yang, S. Zhang, Effects of Cr and Ni on the microstructure and corrosion resistance of high-strength low alloy steel, J. Mater. Res. Technol. 23 (2023) 36–52, https://doi.org/10.1016/j.jmrt.2022.12.191.
[18] Z. Wu, Z. Liu, L. Li, Z. Lu, Experimental and neural networks analysis on elevated-temperature mechanical properties of structural steels, Mater. Today Commun. 32 (2022), 104092, https://doi.org/10.1016/j.mtcomm.2022.104092.
[19] M. Toparli, S. Sahin, E. Ozkaya, S. Sasaki, Residual thermal stress analysis in cylindrical steel bars using finite element method and artificial neural networks, Comput. Struct. 80 (2002) 1763–1770, https://doi.org/10.1016/S0045-7949(02)00215-8.
[20] D. Ibrahim, An overview of soft computing, Procedia Comput. Sci. 102 (2016) 34–38, https://doi.org/10.1016/j.procs.2016.09.366.
[21] J.O. Gidiagba, L. Tartibu, M.O. Okwu, International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), in: Application of Soft Computing Technique Based on ANN Model Prediction in

Diverse Area of Mining Blasting Operations, IEEE, Durban, South Africa, 2022, pp. 1–7, 04.08.2022 - 05.08.2022.
[22] A. Nwachukwu, H.O. Omoregbee, M.O. Okwu, Lagouge K. Tartibu, D.R. Enarevba, A. Adesuji, A review of artificial neural network applications in petroleum exploration, production and distribution operations, International Journal of Scientific & Technology Research 11 (2022) 21–28.
[23] I. Veza, Irianto, H. Panchal, P.A. Paristiawan, M. Idris, I.R. Fattah, N.R. Putra, R. Silambarasan, Improved prediction accuracy of biomass heating value using proximate analysis with various ANN training algorithms, Results in Engineering 16 (2022), 100688, https://doi.org/10.1016/j.rineng.2022.100688.
[24] T.W.B. Riyadi, M. Spraggon, S.G. Herawan, M. Idris, P.A. Paristiawan, N.R. Putra, , M.F. R, R. Silambarasan, I. Veza, Biodiesel for HCCI engine: prospects and challenges of sustainability biodiesel for energy transition, Results in Engineering 17 (2023), 100916, https://doi.org/10.1016/j.rineng.2023.100916.
[25] C.P. Okonkwo, V.I.E. Ajiwe, M.C. Obiadi, M.O. Okwu, J.I. Ayogu, Production of biodiesel from the novel non-edible seed of Chrysobalanus icaco using natural heterogeneous catalyst: modeling and prediction using Artificial Neural Network, J. Clean. Prod. 385 (2023), 135631, https://doi.org/10.1016/j.jclepro.2022.135631.
[26] Y.M. Seo, K. Luo, M.Y. Ha, Y.G. Park, Direct numerical simulation and artificial neural network modeling of heat transfer characteristics on natural convection with a sinusoidal cylinder in a long rectangular enclosure, Int. J. Heat Mass Tran. 152 (2020), 119564, https://doi.org/10.1016/j.ijheatmasstransfer.2020.119564.
[27] S. Huang, B. Tao, J. Li, Z. Yin, On-line heat flux estimation of a nonlinear heat conduction system with complex geometry using a sequential inverse method and artificial neural network, Int. J. Heat Mass Tran. 143 (2019), 118491, https://doi.org/10.1016/j.ijheatmasstransfer.2019.118491.
[28] J.O. Gidiagba, L. Tartibu, M.O. Okwu, International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), 08.2022 - 05.08.2022, in: Crack Detection on a Structural Beam: A Simplified Analytical Method Based on Artificial Neural Network Model, IEEE, Durban, South Africa, 2022, pp. 1–7, 04.
[29] O.D. Samuel, M.O. Okwu, Comparison of Response Surface Methodology (RSM) and Artificial Neural Network (ANN) in modelling of waste coconut oil ethyl esters production, Energy Sources, Part A Recovery, Util. Environ. Eff. 41 (2019) 1049–1061, https://doi.org/10.1080/15567036.2018.1539138.
[30] M.O. Okwu, O.D. Samuel, O.B. Otanocha, L.K. Tartibu, H.O. Omoregbee, V. M. Mbachu, Development of ternary models for prediction of biogas yield in a novel modular biodigester: a case of fuzzy Mamdani model (FMM), artificial neural network (ANN), and response surface methodology (RSM), Biomass Conv. Bioref. 13 (2023) 917–926, https://doi.org/10.1007/s13399-020-01113-1.
[31] M.G.K. Machesa, L.K. Tartibu, M.O. Okwu, Performance analysis of stirling engine using computational intelligence techniques (ANN & Fuzzy Mamdani Model) and hybrid algorithms (ANN-PSO & ANFIS), Neural Comput. Appl. 35 (2023) 1225–1245, https://doi.org/10.1007/s00521-022-07385-0.
[32] D.R.E. Ewim, M.O. Okwu, E.J. Onyiriuka, A.S. Abiodun, S.M. Abolarin, A. Kaood, A quick review of the applications of artificial neural networks (ANN) in the modelling of thermal systems, Engineering and Applied Science Research 49 (2022) 444–458, https://doi.org/10.14456/EASR.2022.45.
[33] H.S. Lim, Y.T. Kang, Estimation of finish cooling temperature by artificial neural networks of backpropagation during accelerated control cooling process, Int. J. Heat Mass Tran. 126 (2018) 579–588, https://doi.org/10.1016/j.ijheatmasstransfer.2018.06.022.
[34] J.-S. Son, D.-M. Lee, I.-S. Kim, S.-G. Choi, A study on on-line learning neural network for prediction for rolling force in hot-rolling mill, J. Mater. Process. Technol. 164–165 (2005) 1612–1617, https://doi.org/10.1016/j.jmatprotec.2005.01.009.
[35] S.H. Zhang, L. Deng, L.Z. Che, An integrated model of rolling force for extra-thick plate by combining theoretical model and neural network model, J. Manuf. Process. 75 (2022) 100–109, https://doi.org/10.1016/j.jmapro.2021.12.063.
[36] K. Esendağ, A.H. Orta, İ. Kayabaşı, S. İlker, Prediction of reversible cold rolling process parameters with artificial neural network and regression models for industrial applications: a case study, Procedia CIRP 79 (2019) 644–648, https://doi.org/10.1016/j.procir.2019.02.061.
[37] M. Landfahrer, R. Prieler, B. Mayr, H. Gerhardter, R. Schöngrundner, J. Klarner, C. Hochenauer, Development of a numerically efficient CFD model to predict transient temperature distribution of mother tubes moving translative and rotative through a gas fired furnace, Appl. Therm. Eng. 123 (2017) 290–300, https://doi.org/10.1016/j.applthermaleng.2017.05.093.
[38] M. Landfahrer, R. Prieler, B. Mayr, H. Gerhardter, T. Zmek, J. Klarner, C. Hochenauer, Characterization of the temperature distribution on steel tubes for different operating conditions in a reheating furnace using CFD and three different measuring methods, Appl. Therm. Eng. 133 (2018) 39–48, https://doi.org/10.1016/j.applthermaleng.2017.12.098.
[39] M. Landfahrer, C. Schluckner, R. Prieler, H. Gerhardter, T. Zmek, J. Klarner, C. Hochenauer, Numerical and experimental investigation of scale formation on steel tubes in a real-size reheating furnace, Int. J. Heat Mass Tran. 129 (2019) 460–467, https://doi.org/10.1016/j.ijheatmasstransfer.2018.09.110.
[40] M. Landfahrer, C. Schluckner, H. Gerhardter, T. Zmek, J. Klarner, C. Hochenauer, Numerical model incorporating different oxidizer in a reheating furnace fired with natural gas, Fuel 268 (2020), 117185, https://doi.org/10.1016/j.fuel.2020.117185.
[41] J. Raič, M. Landfahrer, J. Klarner, T. Zmek, C. Hochenauer, Modelling of the cooling process of steel tubes in a rake type cooling bed, Appl. Therm. Eng. 169 (2020), 114895, https://doi.org/10.1016/j.applthermaleng.2019.114895.
[42] J. Raič, M. Landfahrer, J. Klarner, T. Zmek, C. Hochenauer, Coupled CFD simulation for the cooling process of steel tubes in a rake type cooling bed, Appl. Therm. Eng. 171 (2020), 115068, https://doi.org/10.1016/j.applthermaleng.2020.115068.
[43] C. Rubio-Ramirez, D.F. Giarollo, J.E. Mazzaferro, C.P. Mazzaferro, Prediction of angular distortion due GMAW process of thin-sheets Hardox 450® steel by numerical model and artificial neural network, J. Manuf. Process. 68 (2021) 1202–1213, https://doi.org/10.1016/j.jmapro.2021.06.045.
[44] F. Liu, C. Tao, Z. Dong, K. Jiang, S. Zhou, Z. Zhang, C. Shen, Prediction of welding residual stress and deformation in electro-gas welding using artificial neural network, Mater. Today Commun. 29 (2021), 102786, https://doi.org/10.1016/j.mtcomm.2021.102786.
[45] J.S. Xia, M. Khaje Khabaz, I. Patra, I. Khalid, J.R.N. Alvarez, A. Rahmanian, S. A. Eftekhari, D. Toghraie, Using feed-forward perceptron Artificial Neural Network (ANN) model to determine the rolling force, power and slip of the tandem cold rolling, ISA Trans. 132 (2023) 353–363, https://doi.org/10.1016/j.isatra.2022.06.009.
[46] K.G. Sheela, S.N. Deepa, Review on methods to fix number of hidden neurons in neural networks, Math. Probl Eng. 2013 (2013) 1–11, https://doi.org/10.1155/2013/425740.
[47] H.J. Kim, M. Mahfouf, Y.Y. Yang, Modelling of hot strip rolling process using a hybrid neural network approach, J. Mater. Process. Technol. 201 (2008) 101–105, https://doi.org/10.1016/j.jmatprotec.2007.11.293.
[48] M. Bagheripoor, H. Bisadi, Application of artificial neural networks for the prediction of roll force and roll torque in hot strip rolling process, Appl. Math. Model. 37 (2013) 4593–4607, https://doi.org/10.1016/j.apm.2012.09.070.
[49] J. Deng, J. Sun, W. Peng, Y. Hu, D. Zhang, Application of neural networks for predicting hot-rolled strip crown, Appl. Soft Comput. 78 (2019) 119–131, https://doi.org/10.1016/j.asoc.2019.02.030.
[50] D.P. Kingma, J.L. Ba, A Method for Stochastic Optimization, 2015. Adam.
[51] A. Tato, R. Nkambou, Improving Adam Optimizer, 2018.
[52] C. Sammut, G.I. Webb (Eds.), Encyclopedia of Machine Learning, Springer US, Boston, MA, 2010.