



A time series classification approach to non-destructive hardness testing using magnetic Barkhausen noise emission

Martin Unterberg¹ · Joachim Stanke¹ · Daniel Trauth¹ · Thomas Bergs^{1,2}

Received: 26 October 2020 / Accepted: 10 February 2021 / Published online: 16 March 2021
© The Author(s) 2021

Abstract

The process setup of manufacturing processes is generally knowledge-based and carried out once for a material batch. Industry experts observe fluctuations in product quality and tool life, albeit the process setup remains unchanged. These fluctuations are mainly attributed to fluctuations in material parameters. An in-situ detection of changes in material parameters would enable manufacturers to adapt process parameters like forces or lubrication before turbulences like unexpectedly high tool wear or degradation in product quality occurs. This contribution shows the applicability of a deep learning time series classification architecture that does not rely on handcrafted feature engineering for the classification of hardness fluctuations in a sheet-metal coil using magnetic Barkhausen noise emission. This methodology is not limited to the detection of hardness fluctuations in sheet-metal coils and can potentially be applied for the in-situ material property classification in different manufacturing processes and for different material parameters.

Keywords Deep learning · Machine learning

1 Introduction

Manufacturing processes are generally parametrized according to worker experience and specifications that the material supplier provides. These specifications are ensured by material sampling and followed destructive testing, e. g. through tensile tests or hardness testing. Only when turbulences like excess tool wear or a significant degradation in part quality occurs during the manufacturing process, process parametrization is changed. This approach assumes that microstructural parameters of supplied materials remain constant for one batch. Industry experts e. g. in fineblanking, however, observe fluctuations in tool wear progression and part quality, both on a sheet-metal coil and batch level, although the process setup remains unchanged [1].

Past research has shown that fluctuations in material quality and microstructural properties occur e. g. in sheet-metal

forming processes such as deep drawing [2] and bending [3]. These fluctuations have proven to have a negative influence on product quality. In the case of bending it was found that even batch fluctuations have a considerable influence on the required bending force, the shape of the workpiece, and the springback of the sheet-metal, thus leading to significant angular deviations. The material characteristics partly fluctuate so strongly, that it is considered sensible to determine the material parameters separately for each batch in separate tests. A common method is the flat tensile test, which is a destructive material testing method. In practice, however, this is often omitted for reasons of time and cost [3].

In order to counteract material fluctuations and to enhance process resilience, a methodology is required that non-destructively detects material properties and deviations in real time either before or during the manufacturing process. This enables a proactive regulatory intervention through an adaptive process parametrization. A promising approach for the non-destructive determination of material properties of ferromagnetic materials is the measurement and analysis of magnetic Barkhausen noise (MBN).

The challenge in analyzing MBN to infer material parameters lies in the stochastic nature of the signal. Past approaches mainly used hand crafted feature engineering to transform the signal as an input for regressors and classifiers.

✉ Martin Unterberg
M.Unterberg@wzl.rwth-aachen.de

¹ Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University, Campus-Boulevard 30, 52074 Aachen, Germany

² Fraunhofer Institute for Production Technology (IPT), Steinbachstraße 17, 52074 Aachen, Germany

However, this approach reduces the information contained in the signal through assumptions made in the time intensive feature engineering process. Potentially important features that correlate to certain material parameters get lost through this transformation. This contribution utilizes MBN to detect hardness fluctuations in a sheet-metal material using segmented raw MBN time series as an input for a deep convolutional neural network architecture. The proposed approach does not rely on the common transformation of the MBN signal to a handcrafted feature space and thus is potentially more flexible than traditional approaches to MBN analysis by learning meaningful representations of the material state. However, the proposed methodology is not limited to hardness or material classification and can potentially be applied to other industrial time series classification (TSC) tasks.

2 Background and state of the art

This section gives a brief overview of the phenomenon of MBN and its applicability to non-destructive material testing. Furthermore, state of the art approaches to the task of TSC are reviewed and presented.

2.1 Magnetic Barkhausen noise

The class of non-destructive material testing methods is subdivided according to operating principles and includes ultrasonic testing, x-ray testing, as well as electrical, mechanical and magnetic methods [4]. The selection of the optimal method is done individually, since each method is characterized by different advantages and disadvantages. With regard to the selection of a suitable method for the microstructural evaluation of ferromagnetic steels, it was found that an analysis of MBN is preferable to ultrasonic testing. For experimental proof, the microstructure of different samples was changed by means of heat treatment in a first step. Subsequently, the samples were measured in the second step with both non-destructive testing methods and the results obtained were compared. The results showed that MBN reacted much more sensitively to a change in the microstructure [5].

Crystallite ferromagnetic materials are characterized by the fact that they are divided into crystal regions of different sizes, which exhibit magnetic moments in the same direction in groups. The elementary magnetic moments are aligned by coupling forces of adjacent atoms without the influence of an external magnetic field. The crystal regions are called domains and are of great importance for magnetization and are separated by so-called domain walls [6]. Figure 1 shows several domains with different orientations of the magnetic moments. If an external increasing magnetic field is now

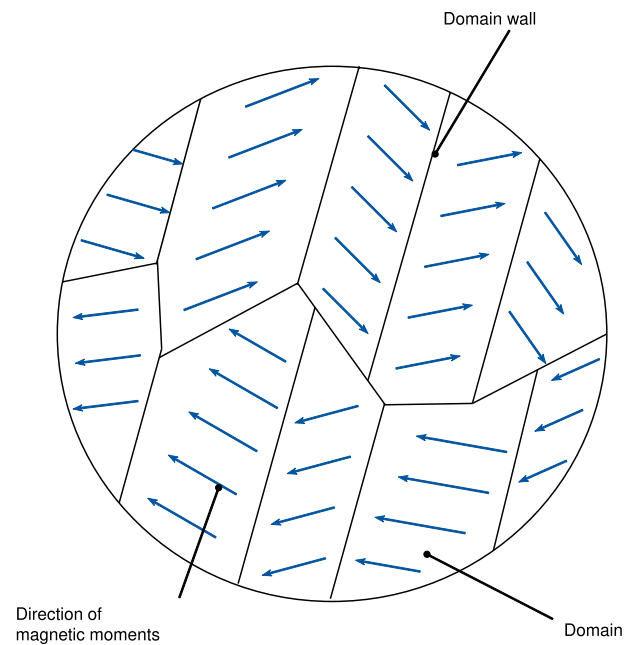


Fig. 1 Representation of the distribution of domains separated by domain walls in a ferromagnetic material

applied near to the material, the magnetization of the microstructure changes and domain walls move depending on the field direction and field strength.

At a low field strength of the applied magnetic field, the movements of the domain walls are reversible. If the field strength exceeds a certain threshold value, the reversibility is lost. In this case, lattice disturbances restrict the process to the extent that there is a delay in the movement or a temporary standstill of the domain wall. If the field strength is further increased to an individual threshold value, the domain wall overcomes the defect and a sudden movement of the domain wall occurs, which causes a jerky change in the flux density. This phenomenon is called Barkhausen jump and is characterized in the magnetization curve by a course that resembles a staircase function [7], as visualized in Fig. 2.

The shifts and changes of the materials magnetization induce current pulses which are measurable and acoustically perceptible via an amplifier and loudspeaker and measurable via a sensitive sensor coil. Figure 3 shows an excerpt of one second of MBN as a result of an external magnetic field with an excitation frequency of $f_M = 300$ Hz with an amplitude of $U = 0.573$ V.

The audible and measurable noise due to the jerky increase in field density is referred to as MBN [8]. The resulting time series has been subject to past research with different approaches. SORSA et al. used 72 features generated from the raw time series, the raw root mean square (RMS) signal, and from a filtered moving window signal as an input for a multivariate linear regression model for

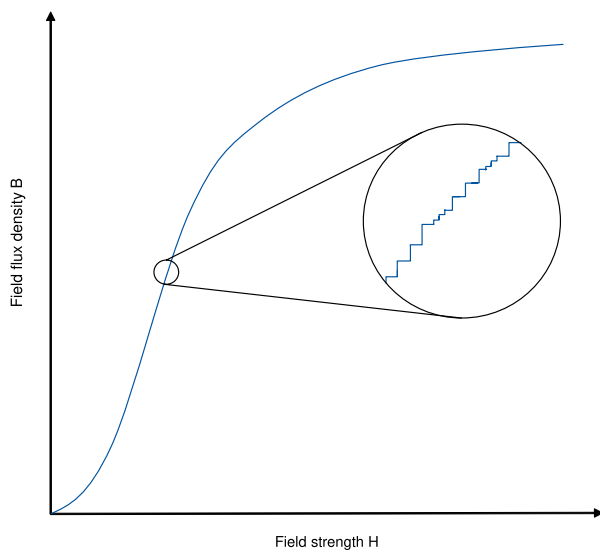


Fig. 2 Visualization of the discontinuous transitions in the magnetization curve

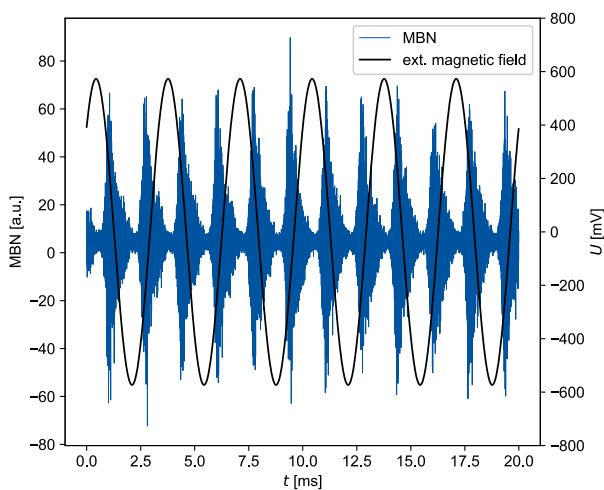


Fig. 3 Excerpt of 20 ms from one second of MBN plotted against the exciting field strength

hardness and residual stress detection [9]. TAN et al. combined ultrasonic measurements and MBN to determine the hardness of 45 steels with different heat treatments [10]. From the RMS signal the inherent features distance of two peaks, half peak width, and peak position were extracted. LUO et al. measured the MBN on hot formed steel of 36.2 to 62.4 HRC. Their method considered the MBN peaks in the raw signal of each period of the alternating field and averages them over the entire measurement [11]. XIUCHENG et al. used the RMS signal to extract the features peak height, peak position and peak width at 50 percent and 75 percent of the peak height, respectively. Predictions of the hardness were performed using both multivariate linear regression

and a fully connected neural network (FCNN). The FCNN outperformed the multivariate linear regression [12]. These approaches have in common that they first transform the time series to a feature space, e. g. through extraction of spectral features, extraction of features from the time domain like RMS or through an extraction of both spectral and time domain features.

2.2 Time series classification

TSC is a task where a classifier has to assign a label Y to a timely ordered sequence of values

$$X = [x_0, x_1, \dots, x_t] \tag{1}$$

with $x_i \in \mathbb{R}^m$ and $m = 1$ in case of a univariate and $m > 1$ in case of a multivariate time series of length t .

TSC is subject to intensive research since its applicability to a plethora of domains, reaching from machine failure detection in an industrial setting [13] to stock market data [14] or speech recognition [15]. Hence, manifold approaches to TSC exist, each with their own up- and downsides. A valuable benchmarking tool for the performance of a TSC model is provided through the University of California Riverside (UCR) Time Series Archive [16]. TSC approaches can be roughly divided into state of the art TSC models that require a transformation of the raw time series to a feature space (e. g. ensemble classifiers) and deep learning models that use the raw time series as an input and generate features on their own, e. g. through convolution and pooling operations. The Collective Of Transformation-based Ensemble (COTE) classifier is based on 35 classifiers and is extended through a hierarchical voting system of the COTE classifiers to HIVE-COTE. HIVE-COTE is widely recognized as a state of the art TSC model [17]. HIVE-COTE has a computational complexity of $O(n^2 \cdot t^4)$ for a dataset of size n and time series length t and took more than 72,000 s to train on a dataset of $n = 700$ time series of length $t = 46$ on a high end device at the time of publication [18]. It is thus not suited for model updates in an industrial setting, where sensoric data with high sampling rates like MBN or acoustic emission (AE) is collected. Recent research has shown that while being much faster since leveraging parallel GPU computations, deep learning models like convolutional neural networks (CNN) perform equally good or, in the case of InceptionTime, even outperform HIVE-COTE on the UCR dataset both in training and prediction time [18]. The CNN architecture InceptionTime by FAWAZ et al. was released in 2019 and is able to outperform ResNet, which was proposed as a baseline model for TSC by WANG et al. in 2017 [19], while scaling better [18]. InceptionTime consists of an ensemble of five deep learning models for TSC, where each classifier consist of a cascade of multiple so called Inception

modules, which were initially proposed by SZEGEDY et al. for computer vision purposes and lead to the rise of CNN in computer vision tasks [20]. Especially for computer vision tasks like image recognition, deep CNN architectures are state of the art. While CNN learn spatial information and features from images, they have shown to be able to learn temporal information and features from time series. The developers of InceptionTime state that “put simply, the time series problem is essentially the same class of problem” as image classification, “just with one less dimension”. CNN are a promising and scalable approach to TSC and hence suited for industrial purposes for time critical TSC tasks like material property or tool health classification.

3 Methodology

This section describes the experimental setup of this contribution. This includes the hardness and MBN measurements, preprocessing of the resulting MBN signals, and training and test set configurations for the deep learning model.

3.1 Dataset generation: hardness measurements

The analyzed specimen stem from a 16MnCr5 (AISI: 5115) sheet-metal coil which is a generic fineblanking steel [21]. The sheet-metal coil has been cold rolled section-wise to specific thicknesses representing industrial standard tolerance boundaries. The aim was to generate sections on the sheet-metal coil with varying hardness properties that represent industrial standards. For this purpose, the sections varied between thicknesses of 3.95 mm and 4.05 mm. In total, 22 specimen of dimensions 16500 mm × 6800 mm have been analyzed for their hardness properties. Since even specimen-wise material inhomogeneities that lead to hardness fluctuations are to be expected and to counteract inherent uncertainties through hardness measurements, every specimen has been divided into 8 cells of dimensions 3000 mm × 3000 mm. The Brinell hardness test was chosen to determine the hardness of the samples. Brinell testing is a destructive method widely used in industry to determine the hardness of metallic materials. The Brinell test is suitable for all steels with smooth surfaces of up to 650 Brinell hardness (HBW) [22]. The case hardening steel 16MnCr5 can reach up to 207 HBW, depending on the exact composition and finishing [23]. The *Indentec* device from *Zwick Roell Testing Systems GmbH* was used to perform the Brinell test. The Brinell test was carried out according to DIN EN ISO 6506 [22]. A ball diameter of $D = 2.5$ mm, a test kilogram-force of $F = 306.5$ N and an exposure time of 10 to 15 s were selected as test specifications.

The measured hardness of the samples was in the range 118–138 HBW. The interval is not considered to be very

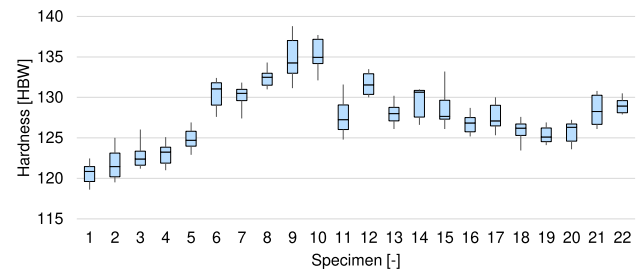


Fig. 4 Boxplot of the resulting intra- and interspecimen hardness distribution

large in materials science, but corresponds to a realistic deviation of the hardness on a sheet-metal coil in the fineblanking industry and likely leads to deviations in the process result with a static process setting.

Figure 4 shows a boxplot diagram of the measured hardness values in HBW of the 22 samples. Each specimen is represented in the boxplot through the observed hardness quartiles of 8 measurements.

3.2 Dataset generation: magnetic Barkhausen noise measurement and labeling

A sensor of the type *μmagnetic* of the company *QASS GmbH* was used to measure the MBN. The sensor consists of three coils, two excitation coils for generating the magnetic field and the sensor coil for measuring the MBN. At a given excitation frequency, the excitation coils carry out a continuous alternating magnetization of the sample. A power amplifier is connected behind the sensor to amplify the measured MBN. The measured data is recorded on a measuring computer.

In the experimental setup of this contribution the MBN measurements were performed with an excitation frequency of $f_M = 300$ Hz with an amplitude of $U = 0.573$ V. The research question whether this configuration leads to an optimal correlation between the resulting MBN signal of this specific material and the hardness of the material is, albeit important for future research, out of the scope of this contribution. The depth of the MBN analysis is especially dependent on the used excitation frequency due to the skin effect [24]. Thus, higher excitation frequencies lead to analysis closer to the material surface. The parameter configuration has been chosen according to recommendations of the company *QASS GmbH* for the given material and has proven to lead to meaningful results with this material and material thickness in past works [25]. Sensor and specimen were separated by galvanic isolation through a 0.05 mm thick polypropylene layer during measurements. Since MBN is a stochastic signal and every measurement contains potentially different information, each cell has been measured in total 5

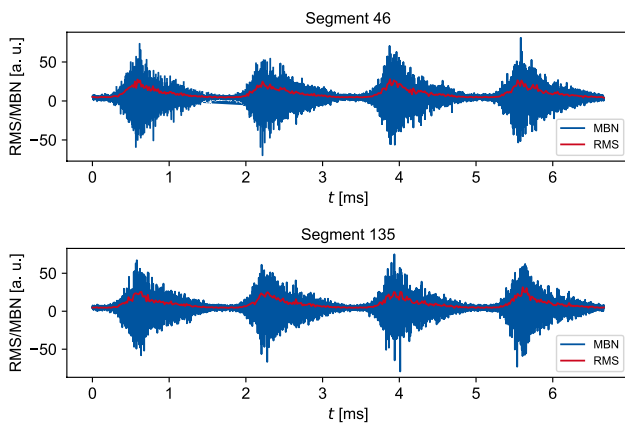


Fig. 5 Visualization of 2 segments with each 2 excitation cycles from a MBN measurement (blue) and the corresponding RMS signal used for segmentation (red)

times for 1 s with a sensor sampling rate of $f_s = 4$ MHz. In total, 880 MBN measurements have been performed, each time series containing approximately 4 million datapoints. The resulting signal is measured in arbitrary units (a. u.) with the device, which, albeit not a SI-unit, can be used for comparisons between measurements executed by this device with the same settings.

3.3 Preprocessing

Each sample contained approximately 4 million datapoints and, because of the excitation frequency of $f_M = 300$ Hz, approximately 600 Barkhausen jumps. Thus, each time series has been segmented into subsegments containing 2 excitation cycles and thus 4 Barkhausen jumps. To prevent phase shifts in the segments, the first incomplete excitation period was cut off. When measuring the MBN with the device from *QASS GmbH*, no information on the phase of the external magnetic field was recorded. Therefore, the local minimum of the RMS signal within the length of half a period of excitation of the external magnetic field is determined as the cut-off point. The local minimum is a reasonable value, since by squaring when generating the RMS signal, this point can be considered as the point with the lowest activity or oscillation of the MBN. Figure 5 shows two resulting segments of a MBN measurement and the corresponding RMS signal.

To generate samples with the same length, the last incomplete segment was removed from the sample pool. The resulting segments had a length of $t = 26,666$ each. Preprocessing has been implemented with Python 3.7 using the libraries *SciPy*, *numpy* and *pandas*.

Table 1 Overview of tested segment configurations

	C_1	C_2	C_3	C_4
Train	1, 2, 9, and 10	1, 2, 4, 5, 6, 7, 9, and 10	1, 2, 3, 4, 7, 8, 9, and 10	1, 2, 4, 5, 6, 7, 9, 10, 13, 14, 15, and 17
Test	3 and 8	3 and 8	5 and 6	3, 8, and 16
Classes	2	2	2	3

3.4 Train and test set generation

The models have been trained with the preprocessed segments of different specimen in 4 different training and test set configurations C_1, C_2, C_3, C_4 . Configurations C_1, C_2 , and C_3 have been put together as a binary classification problem and segments have been divided into two classes around the median hardness of the dataset 124.7 HBW. Configuration C_4 was set up as a classification task with 3 different hardness classes with class boundaries at 125.46 HBW and at 129.91 HBW.

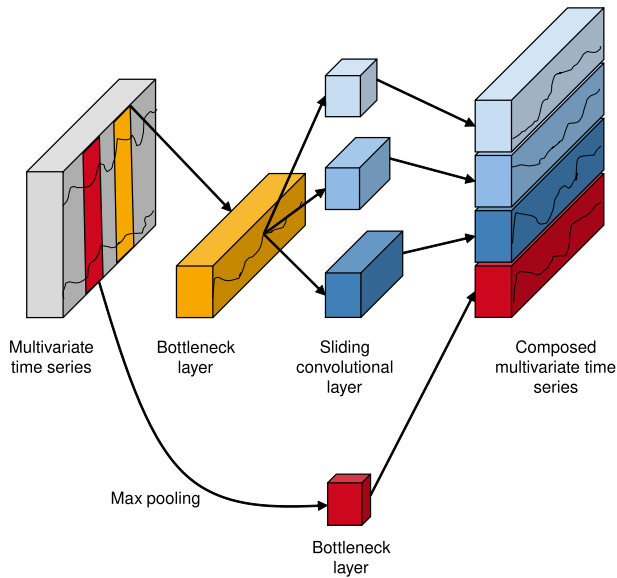
Table 1 gives an overview of the configurations that have been used for training and testing the model. Numbers indicate which specimen the utilized segments stem from.

The configurations were put together under certain assumptions. The first assumption was that segments from the same measurement contain inherent biases, since consecutive Barkhausen jumps influence each other. In a configuration that contains different segments from the same measurement in both the training and test set, the model would potentially be able to easily identify patterns from this exact measurement. For this reason, all test sets contain only segments from unknown specimen and measurements across all configurations, while maintaining equally distributed class sizes. Furthermore, it was assumed that a classification of segments belonging to specimen that showed an average hardness closer to the class boundaries is a more difficult task for the model than a classification of segments with a larger hardness difference to the class boundaries. Finally, it was assumed that a low amount of variability in measurements in the training set (e. g. many segments of fewer measurements) hinder a generalization on the test set.

Configuration C_1 has been chosen to contain segments from measurements that showed the largest hardness difference to the class boundary, but not a high amount of variability of specimen for the training and test set. Therefore, the training set consisted of all segments from all measurements of specimen 1, 2, 9, and 10. The segments for the test set have been randomly sampled as 30 % of all segments from specimen 3 and 8. C_2 was chosen to contain more variability in the training set and assigned the same test set as C_1 . The training set was put together through randomly sampling 30 % of all segments from all measurements of specimen 1,

Table 2 Overview of trainings and test set sizes for the different configurations

	C_1	C_2	C_3	C_4
Training set size	23,840	14,304	14,304	16,450
Test set size	3,576	3,576	3,576	5,364

**Fig. 6** Schematic of an Inception module in InceptionTime

2, 4, 5, 6, 7, 9, and 10. Configuration C_3 has been set up with a similar training set as C_2 , except for specimen 5 and 6 that have been used as test segments and 3 and 4 that have been used as training segments. Thus, the test set of configuration C_3 contained specimen that were closer to the class boundary than configurations C_1 and C_2 . Finally, configuration C_4 showed the highest amount of variability in terms of measurements in the training set with segments from 12 specimen in total. The training set has been put together by randomly sampling 23 % of all segments from all measurements of specimen 1, 2, 4, 5, 6, 7, 9, 10, 13, 14, 15, and 17. For the test set 30 % of all segments from all measurements of specimen 3, 8, and 16 have been randomly sampled. Table 2 gives an overview of the resulting trainings and test set sizes.

3.5 Model training and testing

The training and testing in this contribution have been done with the originally proposed InceptionTime architecture. This architecture consists of so called Inception modules. An Inception module in InceptionTime consists of filters of varying lengths that allow them to extract relevant and, through cascading multiple Inception modules, also hierarchical

features from time series. Figure 6 shows the schematics of one Inception module.

In the schematic, both the input and the output time series are labelled as multivariate time series. The application of m filters of length 1 to a univariate time series of length l results in a multivariate time series with m channels of length l . Thus, although the initial input in the InceptionTime architecture can be a univariate time series, the representation of the time series that is propagated through the InceptionTime layers can be considered multivariate [17]. The bottleneck layer consists of 32 channels in the originally proposed architecture and is followed by a combination of three convolutional layers. The convolutional layers are, unlike in the Inception architecture proposed for image classification tasks, not of length l like the input, but rather perform a sliding window operation different lengths ($d \in \{10, 20, 40\}$). The sliding window operation is useful to consider the sequential behavior of time series. A max pooling layer with a bottleneck is utilized for dimension reduction. All convolutional layers and pooling layers contain 32 units, so that the final output of an InceptionTime module has $4 \cdot 32 = 128$ channels. Three InceptionTime modules form one InceptionTime block. These blocks are connected through skip connections in order to mitigate the vanishing gradient problem [18]. The last InceptionTime module is followed by a global average pooling and a fully connected layer.

InceptionTime is an ensemble of 5 classifiers with the same structure. The classifiers are randomly initiated with different weights. The probability of a time series x_i belonging to a class c from a class space $[1, C]$ is then calculated as the average of the sum of the logistic outputs σ_c

$$\hat{y}_{i,c} = \frac{1}{n} \sum_{j=1}^n \sigma_c(x_i, \theta_j) \quad (2)$$

of each classifier θ_j [18]. This work utilized an architecture with 6 InceptionTime modules in each InceptionTime classifier and ReLU as an activation function. The training and testing of the InceptionTime model has been done with the Python library *tsai*¹, that provides an InceptionTime implementation for the *PyTorch* and the *fastai v2* API [26]. The models have been trained for 50 epochs with a batch size of 4 on a Tesla P100 GPU. As an optimizer the algorithm *Adam* has been used [27]. The training has been done with the *fit_one_cycle* function of the *fastai v2* library with a maximum learning rate of 10^{-5} . This function starts with a lower learning rate and increases the learning rate until hitting the maximum learning rate and decreases the learning rate again, while doing the inverse for momentum [26].

¹ <https://github.com/timeseriesAI>.

This approach is called *Icycle policy* and was originally proposed by SMITH in 2018 [28]. According to the author, this approach leads to faster model convergence.

4 Results and discussion

Table 3 shows the achieved test set accuracies of the classifier after training the models for 50 epochs. Furthermore the highest achieved accuracy and the corresponding epoch is presented. The test accuracy denotes the proportion of correctly classified samples on the test set.

The highest accuracy both overall and for the binary classification task was achieved with configuration C_2 after 41 epochs with 0.995. This confirms the assumption presented in 3.4 that the binary classification task with the highest hardness differences contained in the test set leads to the best classifier performance and thus strengthens the applicability of this TSC approach to hardness classification. It is, however, remarkable, that configuration C_2 outperforms configuration C_1 accuracy wise in the test set although the training set is significantly smaller. The overall second highest accuracy was achieved with

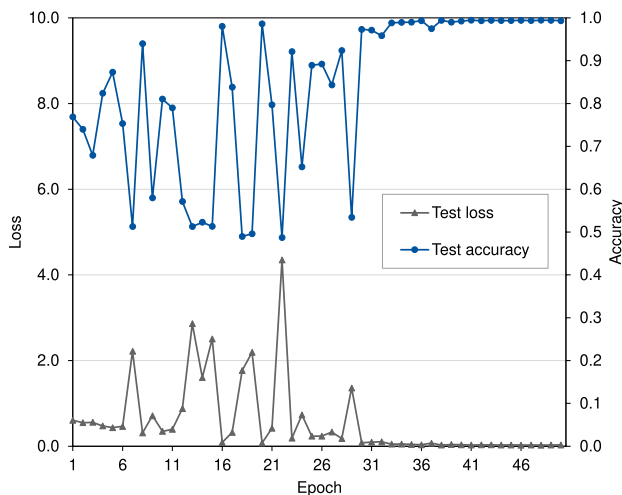


Fig. 7 Test set accuracy and loss of segment configuration C_2 improved over 50 epochs

Table 3 Achieved classification accuracies for the different configurations after 50th epoch and top accuracy

	C_1	C_2	C_3	C_4
Test accuracy after epoch 50	0.788	0.993	0.874	0.979
Highest accuracy (epoch)	0.882 (45)	0.995 (41)	0.925 (20)	0.982 (48)
Classes	2	2	2	3

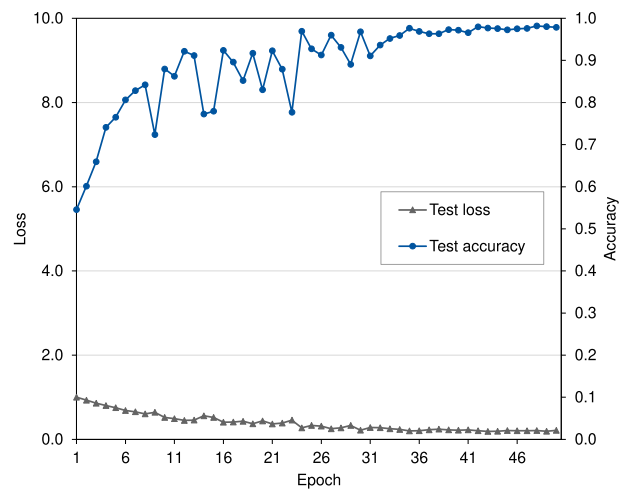


Fig. 8 Test set accuracy and loss of segment configuration C_4 improved over 50 epochs

configuration C_4 after 48 epochs with 0.982. This is remarkable, since the training set consists of the hardness wise most heterogeneous samples of all configurations utilized. The higher variability in sample hardness in the training set may be an important aspect for the classifier to learn more meaningful features and to avoid overfitting and thus to enhance the generalization capability of the model. These findings suggest further research with samples from different material charges and even models trained with different materials.

All models reached their top accuracy before the 50th epoch. Figure 7 shows the test set performance of configuration C_2 . It is apparent that the accuracy and the loss oscillate significantly until epoch 29. Towards the end of the training the accuracy and loss converge. A similar effect could be observed with configuration C_4 , although both accuracy and loss showed a less significant oscillation. This oscillation is most likely due to the higher learning rates through the *Icycle policy*, since it occurs in all configurations in the same epoch interval. Furthermore, the high amount of variability in the training set of configuration C_4 lead to a more steady improvement of the model (Fig. 8).

The originally proposed InceptionTime architecture has been used to train the models. A more knowledge-based approach to adapting the architecture based on the given data structure, e. g. through adaptations of the receptive field of the model, could lead to an even better model performance. Furthermore future research will have to test trained models for a specific sheet-metal coil on another sheet-metal coil of the same material and examine whether the performance is reproducible. Utilizing transfer learning could be a promising approach to both classifying different sheet-metal coil specimen and even different material specimen. Whether learned features of the models could be generalized is an

important question that has to be answered in order to use this methodology in an industrial setting.

A comparison of the deep learning approach to a feature engineering based approach of MBN analysis is out of the scope of this contribution, albeit being an interesting question for future works. Recent work focussing on TSC suggests that, given enough data, the deep learning approach will be superior. Moreover, the deep learning approach can be utilized without explicit knowledge of relevant features.

5 Summary and outlook

Uncertainties on a material level, namely fluctuations of microstructural parameters, lead to process uncertainties in industrial manufacturing processes. Manufacturing processes are generally parameterized once to the provided specifications of a material charge and only changed when turbulences already occurred. The specifications are chosen according to a sample from the a material charge that presumably resembles the specifications of the whole charge, e. g. in sheet-metal forming with samples from sheet-metal coils taken from the beginning and the end of a coil. Past research showed that even on a single sheet-metal coil level, this assumption does not hold. Hence, to increase process resilience, e. g. through anticipative adaption on a process level through process forces or lubrication, uncertainties on a material level have to be dissolved to reduce scrap and increase efficiency manufacturing processes in general.

MBN is a well established non-destructive testing method. In the past the MBN signal has been used to infer microstructural parameters such as hardness and residual stresses. Past research has shown that through feature extraction from the time and spectral domain and a following regression analysis (e. g. through linear regression or through fully connected neural networks), it is possible to successfully predict microstructural parameters. However, this approach requires feature engineering, which can be viewed as a transformation of the actual signal to a hand crafted feature space. Possibly important information about the raw signal is discarded. For industrial purposes, a flexible prediction approach that utilizes the raw MBN signal can potentially help to dissolve material uncertainties through using the raw MBN signal as a representation of the material state. By gathering labels through process and product feedback (e.g. through acoustic emission sensors to model the remaining useful tool life), influences of the MBN signal as a representation of the material state could be utilized to predict process configurations, e.g. forces or lubrication, that lead to the optimal process outcome. This contribution used the raw MBN signal to predict the hardness of sheet-metal material with a deep learning approach. A state of the art TSC architecture has been used to classify segments from MBN measurements regarding the corresponding

measured hardness of the material. The utilized model was able to reliably distinguish the varying hardness of the samples from a sheet-metal coil that has been cold rolled to fulfill the tolerance boundaries that sheet-metal manufacturers guarantee. The main findings can be summarized as follows:

1. InceptionTime with its parameters optimized for the UCR dataset can be utilized for MBN analysis
2. Thus, feature engineering is not mandatory for achieving exceptional results in MBN analysis
3. The utilized architecture performed best on samples with the highest differences in hardness
4. The second best performance was achieved with the highest sample- and hardness heterogeneity in the training set

Since the hardness variance that was observed in this contribution is rather low compared to past scientific approaches to non-destructive testing with MBN it is to be expected that this approach could be leveraged for different materials and different classes of manufacturing processes outside of sheet-metal forming.

This methodology is a promising approach to handling industrial sensory data, since TSC is a common task in an industrial setting. However, more work has to be done to adapt the model architecture to the given data. This contribution used the originally proposed architecture of InceptionTime, without changes in the receptive field of the model. It is reasonable to assume that this architectural choice is not optimal for the given data structure, and a model chosen optimally for the given data is not optimal for a classification of e.g. force measurements. Finally, a more fine-grained class distribution with a larger dataset is desirable for future works. Furthermore, approaches like Grad-CAM [29] that provide explainability for decisions of deep learning models are an interesting approach for future research in this area.

Acknowledgements The authors would like to thank the Federal Ministry of Economics and Technology of Germany for the kind support within the Project "SPAICER" (Project ID: 01MK20015A). Furthermore the authors would like to thank the German Research Foundation DFG for the kind support within the Cluster of Excellence Internet of Production (Project ID: 390621612).

Funding Open Access funding enabled and organized by Projekt DEAL.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long

as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bergs T, Niemietz P, Kaufmann T, Trauth D (2020) Punch-to-punch variations in stamping processes. In: 2020 IEEE 18th world symposium on applied machine intelligence and informatics (SAMI), IEEE, pp 000213–000218
- Harsch D, Fischer P, Berisha B, Heingärtner J, Renkci Y, Hora P (2018) Considering fluctuations of material properties, stainless steel 1.4301, on manufacturability of kitchen sinks. In: IOP conference series: materials science and engineering, Institute of Physics, vol 418, p 012113
- Siegert K (2015) Blechumformung: Verfahren, Werkzeuge und Maschinen, 1st edn. Springer, Berlin / Heidelberg
- Hull J, John V (2015) Non-destructive testing. Macmillan International Higher Education
- Gür CH, Cam I (2007) Comparison of magnetic barkhausen noise and ultrasonic velocity measurements for microstructure evaluation of sae 1040 and sae 4140 steels. *Materials Characterization* 58(5):447–454
- Zapperi S, Cizeau P, Durin G, Stanley HE (1998) Dynamics of a ferromagnetic domain wall: avalanches, depinning transition, and the barkhausen effect. *Phys Rev B* 58(10):6353
- Alessandro B, Beatrice C, Bertotti G, Montorsi A (1990) Domain-wall dynamics and barkhausen effect in metallic ferromagnetic materials. i. theory. *J Appl Phys* 68(6):2901–2907
- Jiles DC (2000) Dynamics of domain magnetization and the barkhausen effect. *Czechoslovak J Phys* 50(8):893–924
- Sorsa A, Leiviskä K, Santa-aho S, Lepistö T (2012) Quantitative prediction of residual stress and hardness in case-hardened steel based on the barkhausen noise measurement. *Ndt & E Int* 46:100–106
- Tan J, Xia D, Dong S, Xu B, Liang Y, Zhu H, Li E (2018) Research on joint nondestructive testing based on neural network. In: *Intelligent computing and internet of things*, Springer, pp 458–467
- Luo XY, Zhang Y, Wang ZJ, Zhang YS (2013) Non-destructive testing device for hot forming high strength steel parts based on barkhausen noise. *Appl Mech Materials Trans Tech Publ* 423:2555–2558
- Xiucheng L, Ruihuan Z, Bin W, Cunfu H (2018) Quantitative prediction of surface hardness in 12crmov steel plate based on magnetic barkhausen noise and tangential magnetic field measurements. *J Nondestructive Eval* 37(2):38
- Miao Q, Makis V (2007) Condition monitoring and classification of rotating machinery using wavelets and hidden markov models. *Mech Syst Signal Process* 21(2):840–855
- Seker SE, Mert C, Al-Naami K, Ayan U, Ozalp N (2013) Ensemble classification over stock market time series and economy news. In: 2013 IEEE international conference on intelligence and security informatics, IEEE, pp 272–273
- Abdel-Hamid O, Ar Mohamed, Jiang H, Deng L, Penn G, Yu D (2014) Convolutional neural networks for speech recognition. *IEEE/ACM Trans. Audio Speech Language Process* 22(10):1533–1545
- Dau HA, Bagnall A, Kamgar K, Yeh CCM, Zhu Y, Gharghabi S, Ratanamahatana CA, Keogh E (2019) The ucr time series archive. *IEEE/CAA J Automatica Sinica* 6(6):1293–1305
- Fawaz HI, Forestier G, Weber J, Idoumghar L, Muller PA (2019) Deep learning for time series classification: a review. *Data Min Knowl Discovery* 33(4):917–963
- Fawaz HI, Lucas B, Forestier G, Pelletier C, Schmidt DF, Weber J, Webb GI, Idoumghar L, Muller PA, Petitjean F (2019) Inception-time: finding alexnet for time series classification. [arXiv preprint arXiv:1909.04939](https://arxiv.org/abs/1909.04939)
- Wang Z, Yan W, Oates T (2017) Time series classification from scratch with deep neural networks: A strong baseline. In: 2017 international joint conference on neural networks (IJCNN), IEEE, pp 1578–1585
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1–9
- Klocke F (2014) *Manufacturing processes 4: forming*. Springer Science & Business Media
- ISO E, et al. (2014) *Metallic materials—brinell hardness test—part 1: Test method*
- DIN, EN (2000) *Cold-rolled narrow steel strip for heat-treatment - technical delivery conditions - part 2: Case hardening steels*
- Knyazeva M, Rozo Vasquez J, Gondecki L, Weibring M, Pöhl F, Kipp M, Tenberge P, Theisen W, Walther F, Biermann D (2018) Micro-magnetic and microstructural characterization of wear progress on case-hardened 16mncr5 gear wheels. *Materials* 11(11):2290
- Unterberg M, Niemietz P, Trauth D, Wehrle K, Bergs T (2019) In-situ material classification in sheet-metal blanking using deep convolutional neural networks. *Prod Eng* 13(6):743–749
- Howard J, Gugger S (2020) Fastai: a layered api for deep learning. *Information* 11(2):108
- Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. [arXiv preprint arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
- Smith LN (2018) A disciplined approach to neural network hyperparameters: Part 1—learning rate, batch size, momentum, and weight decay. [arXiv preprint arXiv:1803.09820](https://arxiv.org/abs/1803.09820)
- Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: Visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE international conference on computer vision*, pp 618–626