Athens Journal of Sciences 2024, 11: 1-21
https://doi.org/10.30958/ajs.X-Y-Z

Quality Testing in Aluminum Die-Casting – A Novel Approach Using Acoustic Data in Neural Networks

By Manfred Rössle[*] & Stefan Pohl[±]

In quality control of aluminum die casting various processes are used. For example, the density of the parts can be measured, X-ray images or images from the computed tomography are analyzed. All common processes lead to practically usable results. However, the problem arises that none of the processes is suitable for inline quality control due to their time duration and to their costs of hardware. Therefore, a concept for a fast and low-cost quality control process using sound samples is presented here. Sound samples of 240 aluminum castings are recorded and checked for their quality using X-ray images. All parts are divided into the categories "good" without defects, "medium" with air inclusions ("blowholes") and "poor" with cold flow marks. For the processing of the generated sound samples, a Convolutional Neuronal Network was developed. The training of the neural network was performed with both complete and segmented sound samples ("windowing"). The generated models have been evaluated with a test data set consisting of 120 sound samples. The results are very promising. Both models show an accuracy of 95% and 87% percent, respectively. The results show that a new process of acoustic quality control can be realized using a neural network. The model classifies most of the aluminum castings into the correct categories.

Keywords: *acoustic quality control, aluminum die casting, convolutional neural networks, sound data*

Introduction

A fast and cost-efficient quality control plays a central role in manufacturing companies. Modern methods open completely new possibilities for designing such processes. Recording a wide variety of data and processing it with innovative technologies helps to gain new insights. These include technologies such as neural networks, which belong to the wide field of artificial intelligence.

Frequently used methods of quality assurance for aluminum castings are computed tomography and X-ray of the parts. This involves taking images of the parts to be inspected to detect any defects, such as air pockets ("blowholes") or cracks. Taking a computed tomography scan is lengthy compared to process times. With an average process time of about 30 seconds per piece, a recording time of 20-30 minutes (!) per piece is clearly too long so that an inline process control is not feasible in a meaningful way.

[*]Professor, Department of Business Information Systems, Aalen University of Applied Sciences, Germany.
[±]Carl-Zeiss AG, Germany.

To create a new inline-capable process of quality assurance, it will be examined whether the use of sound data processing with neural networks is a viable way. Based on the idea that bodies with different densities produce different sounds and frequencies, it is assumed that manufacturing defects, such as air pockets or cracks, change the density of the parts and this can be identified by the neural network. The resulting process could be integrated into an existing manufacturing process at low cost.

Related Work

In recent years, there has been great progress in the field of artificial intelligence. One of these fields is audio data processing in neural networks. Examples are speech, music and pattern recognition in audio files, as well as audio classification. Many application examples can already be found in practice today. For the processing of sound data there are a variety of possibilities, which differ depending on the problem. For audio classification, image representation and processing of audio data, among others, have shown promise (Boddapati et al. 2017, Khamparia et al. 2019, Piczak 2015, Salamon and Bello 2017). Other approaches investigate the processing of raw audio data, without the prior extraction of imaging or manually created features (Abdoli et al. 2019, Yuji Tokozume 2017). The raw audio data is directly provided as input to the neural network. Thus, the processing is not exclusively limited to audio signals, but also applicable to other digital signals like vibration.

When processing the imaged audio signals, the spectrogram, Mel spectrogram or Mel Frequency Cepstral Coefficient (MFCC) are often used. The resulting image representations can be further processed like conventional images in neural networks. Good results in image recognition have been obtained mainly with Convolutional Neural Networks (CNNs) (Krizhevsky et al. 2017). However, CNNs are also used in the processing of raw audio signals.

In most measurement methods, the audio data is processed as a complete block. Another possibility is to divide the audio file into several segments and make them available to the neural network. In this case, the individual segments are classified and later merged for the overall result (Hassan et al. 2019).

Specific approaches for quality assurance of aluminum die-castings using audio data in neural networks cannot be found in the literature. However, other interesting methods for quality control using neural networks are available. Examples are automated quality control of aluminum castings (Mery 2020), (Nguyen et al. 2020) or automated localization of casting defects (Nie et al. 2017) based on X-ray images and their processing in CNNs.

There are concrete approaches for acoustic quality and condition control. For example, the quality of welds (Lv et al. 2017), ceramic tiles (Cunha et al. 2018), the condition of gearboxes (Jing et al. 2017), machines (Kothuru et al. 2019), wind turbines (Kong et al. 2020) or hydropower plants (Voith 2020) can be tested acoustically. A very exotic approach is acoustic quality testing of dried strawberries, to distinguish ripe from overripe fruit (Przybył et al. 2020).

Design and Execution of the Research Process

This chapter describes the design and execution of the research process. After a description of the used aluminum parts and the process of obtaining the sound samples, a derivation of the architecture of the neural network and a description the execution of the experiments follows.

Design of Experiments

Description of the Aluminum Parts

The aluminum castings are provided by the foundry of Aalen University. They are manufactured using the aluminum die-casting process. A total of 240 parts are available for creating sound samples. Each part measures 19.8 x 14.8 x 0.4 cm. The parts are each casted with a defined set of parameters. These parameters are chosen very "extreme", so that the desired properties of the categories described below are achieved in every case.

All castings are checked for quality by means of X-ray images and labeled accordingly by die-casting experts. This labeling allows each part to be assigned without doubt to one of the three categories. The number of parts is the same for each category. There are 80 parts assigned to each category.

- Category "good"

The parts in the "good" category have optimum die-casting parameters. No defects in the form of white spots are visible on the respective X-ray image.

- Category "medium"

Parts in the "medium" category have a changeover point that is too early. This leads to blow holes in the material. Defects in this category cannot be detected visually in the X-ray image or are very difficult to detect. They manifest themselves in barely visible white spots.

- Category "poor"

Parts in the "poor" category have too low gating speed. Typical defects are cold flow marks, which can be perceived as bright spots in the X-ray image. Unlike aluminum parts in the "good" and "medium" categories, these parts can be visually distinguished from the others because they have an uneven surface.

Recording of Sound Samples

The sound samples were recorded in the soundproof room of the Faculty of Optometry and Hearing Acoustics at Aalen University. This offers optimal conditions for the recordings without interfering noise. The recording equipment was also provided by Aalen University and consisted of a professional recording device (Zoom Handy Recorder H4n) and an ECM8000 measurement microphone from Behringer. The recordings were made in WAV format with 96 kHz and a depth of 24 bits.

Figure 1. *Design of Experiment*

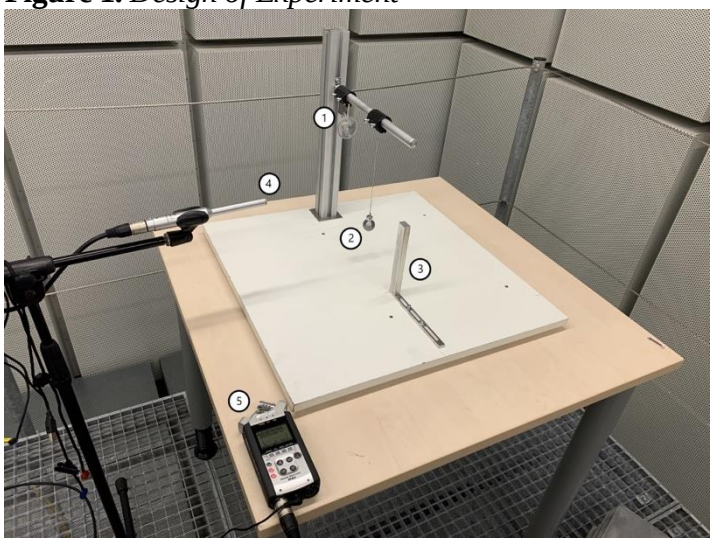


Figure 1 shows the setup with the holding device and the equipment used. The following elements can be seen in the image:

1) Suction cup for fixing the aluminum castings
2) Pendulum with aluminum ball
3) Wedge for constant force application of the pendulum
4) Microphone
5) Recording device

The holding fixture was specially designed and manufactured to record the sound samples. This guarantees a consistent environment for holding each aluminum casting.

The suction cup (1) ensures that the damping of the vibrations on the aluminum casting after the pendulum (2) has bounced is as low as possible. This allows the sound to propagate in the best possible way. The pendulum ball is an aluminum ball, held by a cord on a crossbar to the aluminum casting. A wedge (3) ensures that the acting force of the pendulum on the aluminum casting remains as constant as possible. The resulting sound, which is transmitted through the air, is then recorded in mono via a microphone (4) and stored as a WAV file on the recording device (5).

Recording continues for a few seconds to capture any after-oscillations. The resulting "silence" at the beginning and end of each recording, must be removed during data pre-processing. Since the recording device must be operated manually, the actual length of each sound sample varies approximately between five and seven seconds.

Due to the small number of pieces and the resulting relatively small data set, two sound samples are taken from each aluminum casting. No changes are made to the recording parameters. The resulting data set contains a total of 480 sound samples in the form of digital audio files. However, these cannot be used immediately for the analysis, as they have to be processed beforehand.

Determination of a Suitable Architecture of the Neural Network

There are many different types of neural networks to choose from. Not every type is equally well suited for the problem under investigation. Suitable types can be identified by analyzing previous investigations. The most common types of neural networks for audio data processing are (Purwins et al. 2019, p. 10):

- Convolutional Neural Networks
- Recurrent Neural Networks
- Convolutional Recurrent Neural Networks

As the literature review shows, there are no significant differences between the types in terms of their results in audio data classification. However, there are differences in performance of data processing and model evaluation. For example, Convolutional Neural Networks have an advantage in this area (Purwins et al. 2019, p. 10).

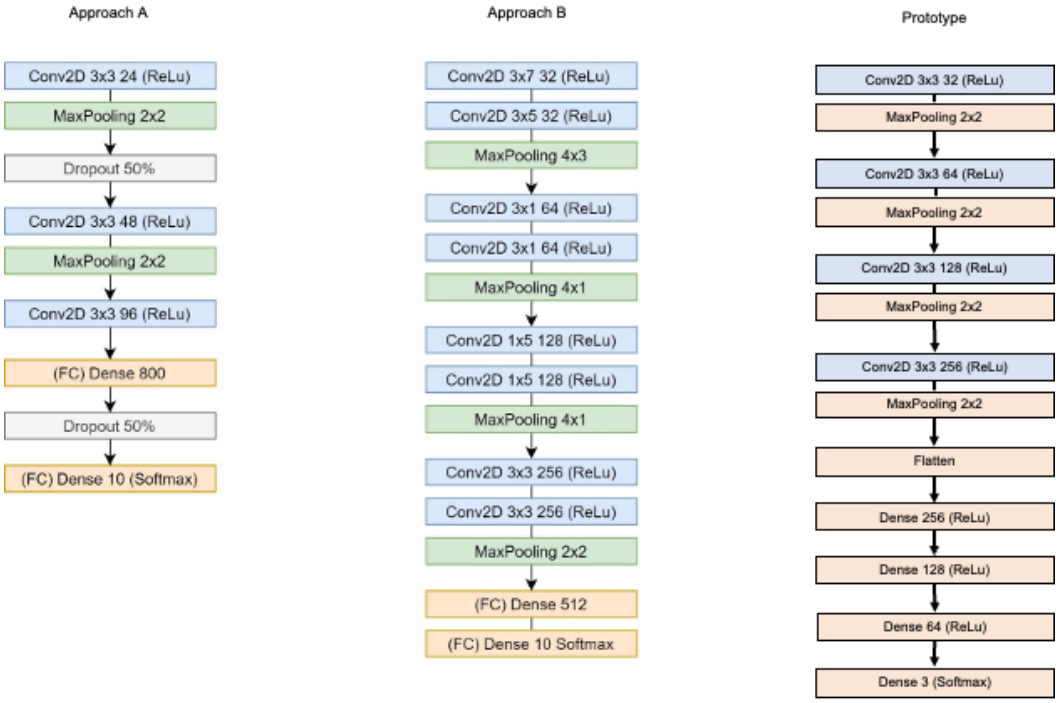
After the selection of the type has been made, the identification of the architecture of the network can be carried out. Unlike the previously mentioned points, the structure of the network architecture is subject to the circumstances of the investigation. Depending on the data set, the structure of the network may vary. Using the same data set different architectures can lead to different results.

It makes sense to follow proven architectures and adapt them to the needs of the project. The choice of architecture depends, among other things, on the type of data and the size of the data set. Since the data to be processed changes with the selection of the audio feature, the appropriate choice of it should also be considered.

Again, previous research in audio classification provides clues for an architecture. Costa et al. (2017, p. 34) as well as Huzaifah (2017, p. 3), show an approach with convolutional layers followed by a max-pooling layer. With each Convolutional Layer, the number of filters increases. Several Fully Connected Layers are used for classification. Figure 2 Approach A shows the described structure.

Another interesting approach, illustrated in Figure 2 Approach B, is shown by Lai et al. (2018, p. 359) with two consecutive convolutional layers, each followed by a max-pooling layer. Again, the number of filters increases with each Convolutional Layer. The end is again formed by several Fully Connected Layers. This approach is inspired by the well-known architecture VGG Net (Simonyan and Zisserman 2014, p. 3). Both approaches follow the structure presented by Salamon and Bello (2017, p. 280) and Piczak (2015, p. 3).

Figure 2. *Architectures of CNNs*



Source: Approach A: Huzaifah (2017, p. 3), Approach B: Lai et al. (2018, p. 359). Prototype: Authors.

Based on the findings reported above, a prototype of a CNN was developed as shown in Figure 2 (Prototype). The different areas are divided into several blocks for more clarity.

The first layer of the neural network is the input layer. It receives as input the image representation of the sound samples in the form of the calculated Mel-Frequency Cepstral Coefficients. This is followed by several identical blocks each consisting of a Convolutional Layer and a MaxPooling Layer. A total of four of these blocks are built into the neural network. Each convolutional layer of these blocks uses a different number of filters. This doubles with each subsequent layer. The first block starts with 32 filters and the last block ends with 256 filters. Their task is to recognize features and patterns from the input data.

Once the four blocks have been run through, an attempt is made to perform a classification based on the information learned from the input data. To do this, the data must first be unrolled ("flattened") and transformed from multidimensional to one-dimensional data structures. The connection of all neurons of the input and output layers takes place in the so-called "Dense Layer". A total of four of these are used in the neural network. Unlike in the blocks before, the number of units decreases with each layer. The first Dense Layer has 256 units, the last three. As can be seen, the last Dense Layer has as many units as there are possible classes.

Selection of suitable audio features

Certain specific audio features have also shown promise in the past. The most common audio features are raw audio data, spectrograms, Mel-Frequency Cepstral

Coefficients, and Log-Mel Spectrograms, whereby Mel-Frequency Cepstral Coefficients and Log-Mel Spectrograms are the most commonly used features in audio data processing (Purwins et al. 2019, p. 10). They produce, in contrast to the raw audio data, a more compact representation of the information. This leads to better performance in training and processing the data by the neural network. However, these features need to be computed by defined functions, which may lead to a loss of information (Purwins et al. 2019, p. 10).

Based on the explained points, the choice of the network type falls on the Convolutional Neural Network and the choice of the audio feature on the Mel-Frequency Cepstral Coefficients. The combination of CNN and MFCC has proven to be a promising basis in the past, provided very good results.

Execution of the Research Process

Preprocessing of Audio Data

The generated sound samples must be preprocessed before being given as input to the neural network. The sound samples used for training the neural network as well as for evaluation and testing must always have the same length. If all audio files are considered, the file with the shortest recording duration is 5.8 seconds and the file with the longest recording duration is 9.6 seconds.

To ensure that all sound samples have the same recording duration, they are cut both at the beginning and at the end. A self-developed function is used to cut the audio files. It removes areas that are below a specified threshold. This threshold applies only to values located at the beginning and end of the file, but not to values located between relevant information of the audio file.

Since the function distinguishes relevant from irrelevant information based on the amplitude values, an additional parameter must be passed for a constant length of each file. This parameter defines a fixed length for each audio file, even if the amplitude value was already undercut before this value.

Specifically this means that if, for example, an audio file falls below the amplitude value after three seconds, but the parameter sets a length of five seconds, the audio file will not be truncated until that later point. This guarantees a constant length of five seconds of each audio file.

Maintaining a constant length is essential for training the neural network. For an error-free training process, each file must have the same input format. All audio files that are passed to the generated model for prediction must correspond to this input format.

The sound samples are recorded with a sampling rate of 96 kHz. The transformation into the time-frequency spectrum reveals in which frequency range the relevant information is located. In our case, most of the information is found in the range between 0 Hz and 8,000 Hz.

The sampling rate is reduced from 96 kHz to 16 kHz. This results in a reduction of the amount of data, which leads to a faster processing of the sound samples. The sampling rate reduction is done with the help of a so-called resampling function. This takes over the reduction of the sampling rate after the cutting process and saves the files as new audio files.

Since the files before and after the sampling rate reduction differ widely in their properties, they are compared in Table 1. Besides the changed parameters, such as sampling rate and fixed length, especially the size of the individual audio data has decreased considerably.

Table 1. *Properties of Raw and Preprocessed Samples*

	Before Reduction	**After Reduction**
Sampling rate	96 kHz	16 kHz
Length	between 5.8 and 9.6 seconds	5 seconds
Size	between 3.2 and 5.3 Megabyte	0.313 Megabyte

Training of the Neural Network

The neural network is trained using two different methods. At first, the neural network processes the audio files from the training data set without modification and in full length. In the second method, random regions (segments) of equal length are taken from the sound samples and passed to the neural network for training. The relatively small data set can be artificially enriched using this method.

- Complete sound samples

With this method, the sound samples are used as a complete block. For this purpose, the MFCC coefficient of the entire sound sample is calculated. The 360 samples thus obtained are passed to the neural network for training.

- Segmented sound samples

To artificially increase the size of the data set for training, random segments of equal size are taken from each sound sample. Both the placement of the section within the sound sample and the choice of the sound sample itself, happens randomly. A two-second segment is then taken from each sound sample. The MFCC coefficient is then determined from this region and passed to the neural network as input.

To obtain the highest possible number of samples of each class, this process is performed 20,000 times. This allows the model to work with 20,000 training data sets.

Training Parameters

Regardless the type of sound sample, the models are trained with 10-fold cross validation. In cross-validation, the entire data set used for training the neural network is divided into k equally sized subsets, where k is the number of subsets (10 in this case). Compared to manually splitting training and validation data, cross-validation is less likely to have an unfavorable distribution of possible classes within the subsets. For the model's overall performance, the average is taken from all obtained metrics (Olson and Delen 2008, p. 141).

The total data set, which consists of 480 sound samples, is manually divided into 75% training and 25% test data. The remaining 360 sound samples in the training data set are subdivided again using 10-fold cross validation. As usual, the subsets are divided into non-overlapping, equal-sized sets of 90% training and 10% validation data each, resulting in 10 training and 10 validation data sets.

The neural network is then trained with the generated segmented data sets. The resulting models can be compared with each other and allow an easier selection of the best model. Table 2 shows the training parameters used for both methods.

Table 2. *Important Parameters of the Models*

Parameter	**Complete sound samples**	**Segmented sound samples**
Loss function	Categorical Crossentropy	Categorical Crossentropy
Optimization function	Adam	Adam
Metrics	Accuracy	Accuracy
Number of epochs	30	60
Batch size	32	512

The loss and optimization functions used are the same for both methods. "Categorical Crossentropy" is selected as loss function and "Adam" (Adaptive Moment Estimation) as optimizer. Also identical is the metric "Accuracy" for both methods. These parameters are chosen based on the research shown in Chapter 2. These have led to be promising results. The specified number of epochs and the batch size have been proven to be optimal by several training runs.

Results

This section shows the results of the training process and the application of the generated models to the test data set. The first subsection contains a description of how to interpret the results from the training. The necessary steps to generate the predicted values are explained in the second subsection. The results obtained are explained in the third subsection.

Evaluation of the Models

After the successful training of the models, they have to be checked for their performance. The key figures collected during training provide an indication of the expected performance of the model. The two key figures "Accuracy" and "Loss" are decisive for this. They can be used to identify problems such as overfitting or underfitting of the model. Overfitting occurs when the model delivers good results on the training data, but poor results on the test data set. Underfitting occurs when the model delivers poor results on the training data (Wani et al. 2020, pp. 47–48).

For better clarity, these key figures are shown in diagrams. The number of learning cycles is shown on the x-axis and the accuracy and loss values of the training and validation data are shown on the y-axis. In this way, the change in both values over the entire course of the training can be displayed and evaluated.

If the accuracy is considered, both values should ideally rise in a curve and approach the value "1" with increasing number of learning cycles. The curve of the validation data set should run parallel to the curve of the training data set. An emerging gap occurring between training and validation data indicates the overfitting of the model (Moolayil 2019, p. 134).

If, on the other hand, the Loss value is considered, it should decrease with increasing number training cycles. It thus runs in the opposite direction to the accuracy value. The loss values of training and validation data should approach "0" with an increasing number of learning cycles. Here, too, a widening gap between the two curves indicates a problem (Gulli 2017, p. 38).

However, the decisive factor for the performance of the model is the generalization. This tells how good the applicability is to data that the neural network has not yet processed. This is tested with the test data set taken before. As described above, the distribution is 75% training and 25% test data.

Since the membership of each sound sample to a category is known, an accurate evaluation of the results can be performed. Both models, from complete and segmented sound samples, are applied to the test data set and the results are compared.

For this purpose, the probability values of each model are determined for each individual sound sample of the test data set. The exact determination of the values is explained in the following chapter. The higher the determined values match the actual values, the better the performance of the model used.

For a better overview of the results from the test data set, the actual and determined class memberships can be compared in a cross table ("confusion matrix"). For this purpose, the determined classes are listed on the X-axis and the actual classes on the Y-axis in a confusion matrix. Accordingly, an ideal line runs from the top left to the bottom right. Values that are outside this ideal line represent incorrectly classified values. If a model achieves 100% accuracy on the test data set, all values lie on the ideal line.

Determination of the Probability Values

The probability values are the results of the model that makes a prediction of class membership. The results are presented in the form of percentages for each possible class. The higher the percentage for a single class, the higher the probability that the sound sample can be assigned to this class. The different classes represent the possible grades (good / medium / poor) of the aluminum castings. To obtain the probability values, the audio data must be processed in advance and the corresponding features extracted. Basically, this is done with the same principle as used for training the model. Depending on whether whole or segmented sound samples are considered, the determination of the probability values looks different.

- Complete sound samples

To determine the probability values of a complete sound sample, the sound sample must be processed as a whole block. For this purpose, the audio file is read in full length and the MFCC coefficient is determined from it. The obtained data is passed to the model for classification. This will lead to the prediction. No further steps need to be performed.

- Segmented sound samples

To classify a segmented sound sample, it is divided into segments. To do this, four segments of equal size, consisting of a two-second window, are taken from the sample.

The distribution of the ranges is not random in this case. Starting with the first range, at the beginning of the sound sample, the remaining ranges follow, each with an offset of one second, until the end of the sound sample. For each of the four areas, the MFCC coefficient is transmitted and the probability values are calculated by the model. The final result of the classification then is the average of all values.

Results from Training and Test Data

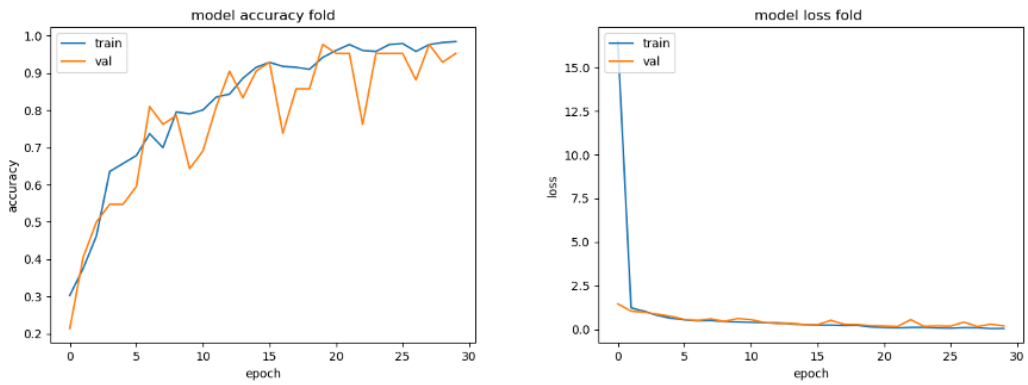
Below are the results from the training phase and the application to the test data set. The separation is done according to complete and segmented sound samples. Thereby, the sections follow the same scheme. At the beginning, the results of the training phase are presented. From all generated models of the cross validation, the diagrams show the model, with the best performance in each case. For a combined overview, a diagram with the average values of the ten models is shown. Overall training cycles this forms a smoothed representation of the results.

Subsequently, the results of the best model from the application to the test data set are shown. For this purpose, both a confusion matrix and the actually determined probability values for each sound sample are clearly presented in a table.

- Complete sound samples

The course of the accuracy over the 30 training cycles of training and validation data can be seen in Figure 3. Due to the relatively small size of the data set, the course of the validation data is erratic, whereas the curve of the training data runs without major jumps. In the course of the investigation, the maximum number of 30 training cycles, with a given amount of data and structure of the neural network, turns out to be optimal. Further training cycles do not improve the results.

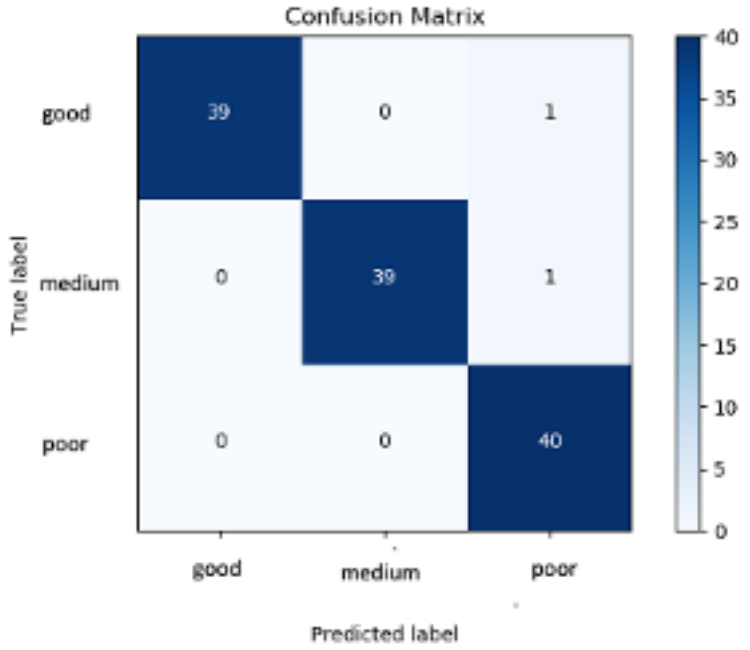
Figure 3. *Model Accuracy and Loss for Complete Sound Samples*



The curves of the loss values in Figure 3 run very flat and parallel to each other without major fluctuations. The fact that the loss value of the validation data does not increase towards the end excludes an overfitting of the model.

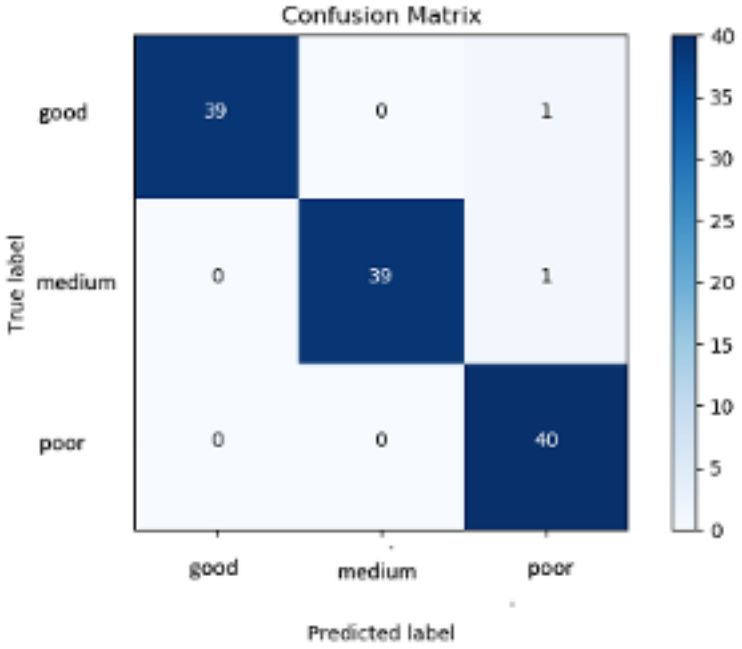
For a smoothed course of both curves, the average of all values is shown. This figure contains the values of all ten created models from the cross-validation. Here it can be seen clearly that training and validation accuracy run simultaneously in a curve towards "1". This is also the case with the loss curve in the same figure. However, here the values run toward "0".

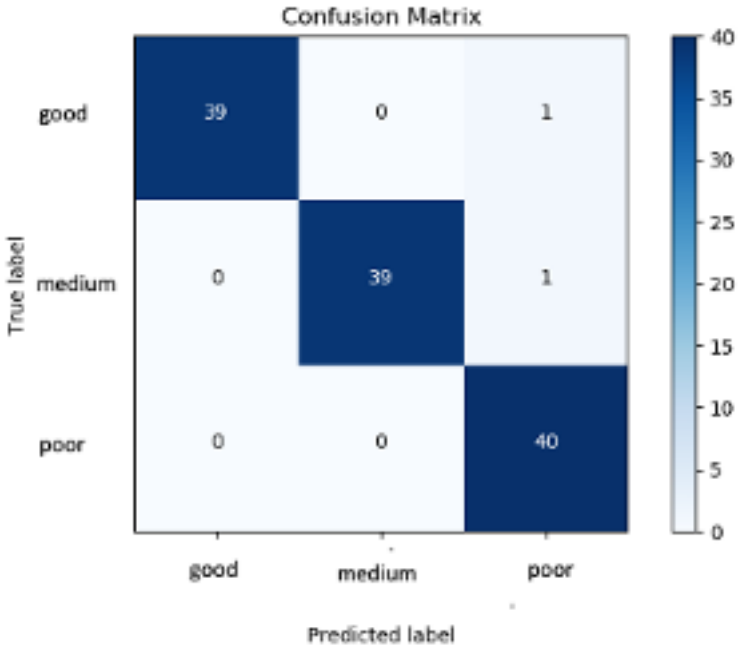
The confusion matrix in



illustrates the good accuracy value on the test data set. Almost all sound samples are classified correctly. However, one sound sample of the "good" class and one sound sample of the "medium" class were each incorrectly assigned to the "poor" class. On the positive side, no sound sample belonging to the "medium" and "poor" classes was assigned to the "good" class. That would be the worst case in practice, but it does not occur here. In so far, the model is working very well.

Figure 4. *Confusion Matrix for Complete Sound Samples*





The probability values determined by the model are listed in

Table 4 in the Appendix. Each row in the three tables represents one sound sample. For a more compact overview, the 120 sound samples of the three categories are listed side by side. This results in a total of 40 sound samples per class. The left column indicates the actual category. The three following columns each show the value of the prediction by the model. The values of each category are given in percent.

With this overview, the values of the two incorrectly classified samples can be analyzed. The sample with the actual category "good", is assigned by the model with 85.13% to the category "poor". Likewise, the sample with the actual group "medium", is assigned by the model with 99.97% to the category "poor". The respective rows are highlighted in the tables.

It is recommended to set a minimum value of at least 80% to ensure an unambiguous classification for all sound samples. Considering this threshold, the result changes from two to six misclassified sound samples. The remaining sound samples are assigned to the correct category by a clear margin.

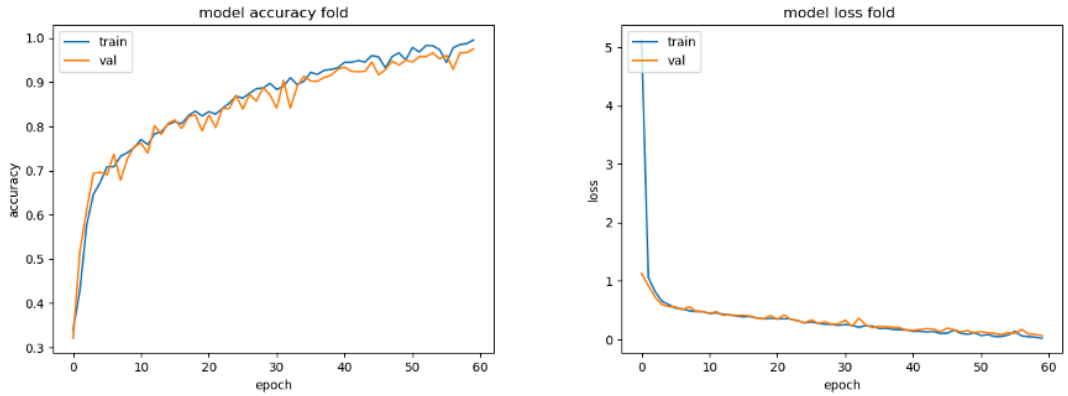
- Segmented sound samples

The data set is much larger than in the previous method due to the artificial enrichment. This also allows a larger number of learning cycles. As can be seen in Figure 5, a maximum of 60 training cycles are possible with this method until both curves run permanently against a value of "1". This turns out to be optimal for the given test parameters.

Both curves of training and validation run roughly parallel to each other. The validation accuracy curve is slightly worse than the training accuracy curve, which corresponds to a normal curve.

The curve of the loss values, shown in Figure 5, also shows an optimal course of training and validation data. The curves are relatively flat and run almost identically, which does not indicate overfitting or underfitting.

Figure 5. *Model Accuracy and Loss for Semented Sound Samples*



For a smoothed course of the learning cycles of training and validation data, the average values of all models are shown. Here, too, one can see the almost identical course of both values of Accuracy and Loss. The figure shows that there are no major fluctuations among the different models.

Table 3 shows a detailed overview of the individual values of Accuracy and Loss. Each of the ten generated models is listed here.

The achieved accuracy values of the best and worst model hardly deviate from the calculated average of all models. The best model, model no. 5, achieves a value of 90% and the worst model, model no. 7, a value of 86%. The average of all models is 87% accuracy.

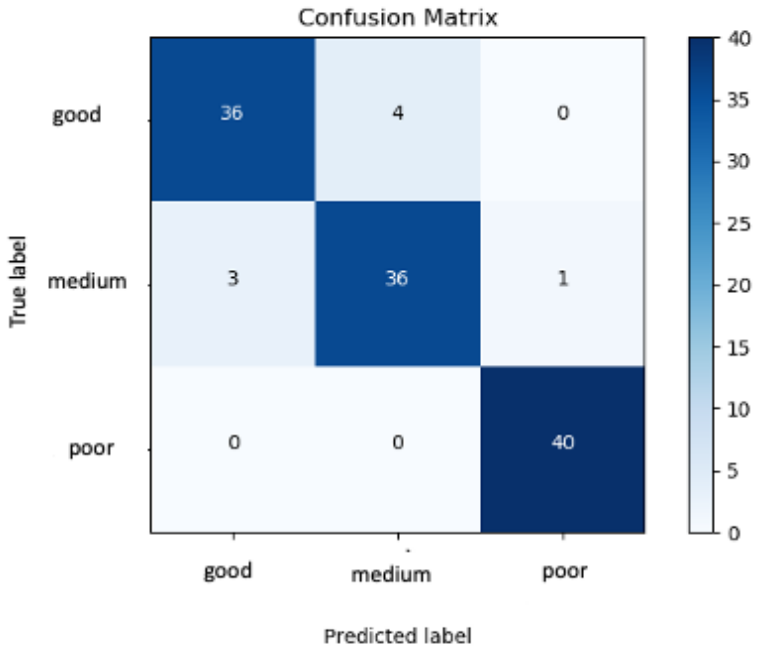
As clearly presented in the previous chapter, the probability values in Table 3 are given as percentages. Here, too, each line represents one of the 120 sound samples, which are arranged next to each other by category.

Table 3. *Accuracy and Loss of Each Model*

Model	**Accuracy**	**Loss**
Model 1	0.8833	0.4938
Model 2	0.8750	0.5660
Model 3	0.8750	0.7467
Model 4	0.8667	0.5775
Model 5	0.9000	0.7267
Model 6	0.9000	0.7494
Model 7	0.8625	0.8475
Model 8	0.8708	0.7148
Model 8	0.8708	0.4842
Model 10	0.8708	0.3785
Ø	0.8775	0.6285

With reference to the results shown in the confusion matrix of Figure 6 the model has difficulties classifying samples of the "good" and "medium" class. Thus, the model, four sound samples of the category "good" with a value of 69.99%, 63.88%, 64.36% and 62.39% are incorrectly assigned to the category "medium". Similarly, three sound samples of the class "good" and one sound sample of the class "poor", are incorrectly assigned to the category "medium". The values for the three sound samples of the category "good" are 67.27%, 74.94% and 68.04% and the value for the single sound sample of the category "poor" is 56.75%.

Figure 6. *Confusion Matrix for Segmented Sound Samples*



As before with complete sound samples, there are cases that are only categorized with a deviation of 2 to 3% to another class. This should be remedied by a fixed threshold of the probability value of at least 80%. Considering the threshold, the number of misclassified sound samples increases from 8 to 44.

Discussion

The conducted research shows that quality assurance by means of sound data processing in neural networks leads to very good and usable results, at least in our experimental setup. Regardless the promising results, some crucial points have to be critically pointed out.

The size of 360 sound samples is very small for training neural networks. As shown, the significantly larger number of segmented sound samples - contrary to all expectations - does not lead to better results. This fact requires a more detailed investigation. Regardless of whether "complete" or "segmented" sound samples are used, to substantiate the results obtained so far, the approach should be validated with a significantly larger number of complete sound samples.

Further need for research arises from the fact that the casting parameter were extremely set during the casting process. As already shown, the idea of parameterization was to achieve a defined result. This was undoubtedly achieved in the given laboratory situation: Three disjoint groups of part qualities emerged, which were clearly separable at the data level.

In practice, however, the situation is completely different. The production process is set with the presumably perfect parameters. Over time, it will happen that individual parameters change, for example due to environmental influences or

due to variances in the material properties. Defects are thus created insidiously by a minimal variation of several parameters or environmental influences. Accordingly, there are also parts that are "more or less" good. This practically very relevant grey area between the respective classes was not represented by our experiments. In this respect there is a need for further research with significantly less extreme parameter sets.

Furthermore, it must be pointed out that the parts examined here had a very simple geometry, which seems quite appropriate for experimental purposes. In practice, however, the parts are likely to have a much more complex geometry. In this respect, it seems urgent to perform comparable experiments to investigate the effectiveness of the approach for more complex parts.

Regarding the production process in practice, an "inline" solution is conceivable. This would allow the finished parts to be checked for quality within a short time after the casting process. For this purpose, a corresponding testing device should be designed which allows sound samples and their classification to be carried out on a kind of assembly line. Since the sound samples used in this work were generated in the laboratory, further investigations should be carried out in a manufacturing environment. Here, possible interfering noises can occur, which must either be learned beforehand by the model or removed during the preprocessing process.

Conclusions

The obtained results show that a new inline quality assurance process using sound data processing in neural networks is basically possible. Against the background of the limitations discussed above, it is necessary to conduct further investigations.

References

Abdoli S, Cardinal P, Lameiras Koerich A (2019) End-to-end environmental sound classification using a 1D convolutional neural network. *Expert Systems with Applications* 136(Dec): 252–263.

Boddapati V, Petef A, Rasmusson J, Lundberg L (2017) Classifying environmental sounds using image recognition networks. *Procedia Computer Science* 112: 2048–2056.

Costa YMG, Oliveira LS, Silla CN (2017) An evaluation of Convolutional Neural Networks for music classification using spectrograms. *Applied Soft Computing* 52(C): 28–38.

Cunha R, Medeiros De Araujo G, Maciel R, Nandi GS, Da-Ros MR, et al. (2018) Applying non-destructive testing and machine learning to ceramic tile quality control. In *SBESC 2018. 2018 VIII Brazilian Symposium on Computing Systems Engineering: Proceedings*, 54–61. Salvador, Brazil, November 6-9, 2018. Los Alamitos, CA: Conference Publishing Services, IEEE Computer Society.

Gulli A (2017) *Deep learning with Keras. Implement neural networks with Keras on Theano and TensorFlow*. Birmingham, UK: Packt Publishing.

Hassan SU, Zeeshan Khan M, Ghani Khan MU, Saleem S (2019) Robust sound classification for surveillance using time frequency audio features. In *2019 International*

Conference on Communication Technologies (ComTech), 13–18. 20-21 March, 2019, Military College of Signals, National University of Sciences & Technology. Piscataway, New Jersey: IEEE.

Huzaifah M (2017) *Comparison of time-frequency representations for environmental sound classification using convolutional neural networks*. Available at: https://arxiv.org/pdf/1706.07156.

Jing L, Zhao M, Li P, Xu X (2017) A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement* 111(Dec): 1–10.

Khamparia A, Gupta D, Nguyen NG, Khanna A, Pandey B, Tiwari P (2019) Sound classification using convolutional neural network and tensor deep stacking network. *IEEE Access* 7(Jan): 7717–7727.

Kong Z, Tang B, Deng L, Liu W, Han Y (2020) Condition monitoring of wind turbines based on spatiotemporal fusion of SCADA data by convolutional neural networks and gated recurrent units. *Renewable Energy* 146(Feb): 760–768.

Kothuru A, Nooka SP, Liu R (2019) Application of deep visualization in CNN-based tool condition monitoring for end milling. *Procedia Manufacturing* 34: 995–1004.

Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. *Communications of the ACM* 60(6): 84–90.

Lai J-H, Liu C-L, Chen X, Zhou J, Tan T, Zheng N, et al. (2018) *Pattern Recognition and Computer Vision*. Cham: Springer International Publishing.

Lv N, Xu Y, Li S, Yu X, Chen S (2017) Automated control of welding penetration based on audio sensing technology. *Journal of Materials Processing Technology* 250: 81–98.

Mery D (2020) Aluminum casting inspection using deep learning: a method based on convolutional neural networks. *Journal of Nondestructive Evaluation* 39(1): 1–13.

Moolayil J (2019) *Learn keras for deep neural networks*. Berkeley, CA: Apress.

Nguyen TP, Choi S, Park S-J, Park SH, Yoon J (2020) Inspecting method for defective casting products with convolutional neural network (CNN). *International Journal of Precision Engineering and Manufacturing-Green Technology* 8(Feb): 583–594.

Nie J-Y, Obradovic Z, Suzumura T, Ghosh R, Nambiar R, Wang C (Eds.) (2017) *2017 IEEE International Conference on Big Data. Dec 11-14, 2017, Boston, MA, USA: Proceedings*. Piscataway, NJ: IEEE.

Olson DL, Delen D (2008) *advanced data mining techniques*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.

Piczak KJ (2015) Environmental sound classification with convolutional neural networks. In D Erdoğmuş (ed.), *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, 1–6. 17-20 Sept. 2015, Boston, USA. Piscataway, NJ: IEEE.

Przybył K, Duda A, Koszela K, Stangierski J, Polarczyk M, Gierz L (2020) Classification of Dried Strawberry by the Analysis of the Acoustic Sound with Artificial Neural Networks. *Sensors (Basel, Switzerland)* 20(2): 499.

Purwins H, Li B, Virtanen T, Schluter J, Chang S-Y, Sainath T (2019) Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing* 13(2): 206–219.

Salamon J, Bello JP (2017) Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters* 24(3): 279–283.

Simonyan K, Zisserman A (2014) *Very deep convolutional networks for large-scale image recognition. https://arxiv*.org/pdf/1409.1556

Voith (2020) *OnCare.* Available at: http://voith.com/corp-de/products-services/automation-digit al-solutions/oncare.html?

Wani MA, Bhat FA, Afzal S, Khan AI (2020) *Advances in deep learning.* Singapore: Springer.

Yuji Tokozume TH (2017) *Learning environmental sounds with end-to-end convolutional neural network*. IEEE.

Appendix

Table 4. *Probability Values for All Complete Sound Samples*

	class		
	good	**medium**	**poor**
good	99,97	0,03	0,00
good	92,62	7,38	0,00
good	91,79	7,95	0,26
good	83,11	16,87	0,02
good	67,24	32,73	0,03
good	94,76	4,94	0,31
good	**14,79**	**0,09**	**85,13**
good	79,21	20,77	0,02
good	97,58	2,40	0,02
good	97,25	2,72	0,03
good	99,84	0,11	0,05
good	99,13	0,86	0,01
good	99,77	0,19	0,03
good	99,71	0,29	0,01
good	99,68	0,30	0,02
good	99,99	0,01	0,00
good	99,90	0,10	0,00
good	98,99	1,00	0,00
good	99,95	0,05	0,00
good	99,88	0,12	0,00
good	99,72	0,28	0,00
good	85,79	1,53	12,69
good	100	0,00	0,00
good	98,00	1,99	0,01
good	99,85	0,10	0,05
good	81,28	18,69	0,03
good	99,99	0,01	0,01
good	99,90	0,06	0,04
good	95,47	4,43	0,10
good	99,31	0,68	0,01
good	91,44	8,54	0,02
good	82,90	16,99	0,11
good	99,90	0,10	0,00
good	99,85	0,09	0,06
good	99,91	0,08	0,01
good	99,90	0,08	0,01
good	99,96	0,03	0,01
good	99,82	0,14	0,04
good	98,12	1,87	0,01
good	97,21	2,78	0,01

	class		
	good	**medium**	**poor**
medium	37,96	56,25	5,80
medium	7,10	90,67	2,23
medium	32,53	67,35	0,12
medium	3,71	96,28	0,01
medium	0,03	99,97	0,00
medium	0,01	99,99	0,00
medium	0,00	100,00	0,00
medium	0,00	100,00	0,00
medium	0,13	99,86	0,01
medium	0,07	99,93	0,01
medium	0,27	99,72	0,01
medium	0,00	100,00	0,00
medium	0,00	100,00	0,00
medium	0,00	100,00	0,00
medium	0,00	100,00	0,00
medium	0,00	100,00	0,00
medium	0,00	100,00	0,00
medium	0,11	99,89	0,01
medium	0,79	98,94	0,28
medium	0,06	99,94	0,00
medium l	0,41	99,57	0,02
medium	0,02	99,98	0,00
medium	7,75	92,22	0,04
medium	**0,03**	**0,00**	**99,97**
medium	0,44	99,56	0,00
medium	0,00	100,00	0,00
medium	1,73	97,94	0,33
medium	0,00	100,00	0,00
medium	0,00	100,00	0,00
medium	0,00	100,00	0,00
medium	0,00	100,00	0,00
medium	0,03	99,94	0,03
medium	0,00	99,99	0,00
medium	0,00	100,00	0,00
medium	0,00	100,00	0,00
medium	0,00	100,00	0,00
medium	7,71	90,98	1,32
medium	0,00	100,00	0,00
medium	0,01	99,99	0,00
medium20	0,02	99,95	0,04

	class		
	good	**medium**	**poor**
poor	0,27	0,00	99,73
poor	0,07	0,00	99,93
poor	0,25	0,01	99,75
poor	0,36	0,01	99,63
poor	0,10	0,00	99,90
poor	0,51	0,00	99,49
poor	0,50	0,00	99,50
poor	1,71	0,01	98,28
poor	4,07	0,08	95,86
poor	0,14	0,00	99,86
poor	0,06	0,00	99,94
poor	0,00	0,00	100,00
poor	0,01	0,00	99,99
poor	0,00	0,00	100,00
poor	0,00	0,00	100,00
poor	0,05	0,00	99,95
poor	0,25	0,00	99,75
poor	1,30	0,03	98,67
poor	0,00	0,00	100,00
poor	0,03	0,00	99,97
poor	0,03	0,00	99,97
poor	0,04	0,00	99,96
poor	0,00	0,00	100,00
poor	0,00	0,00	100,00
poor	0,05	0,00	99,95
poor	0,00	0,00	100,00
poor	0,02	0,00	99,98
poor	0,01	0,00	99,99
poor	0,00	0,00	100,00
poor	0,01	0,00	100,00
poor	0,02	0,00	99,98
poor	0,02	0,00	99,98
poor	0,04	0,00	99,96
poor	0,00	0,00	100,00
poor	0,14	0,00	99,86
poor	0,09	0,00	99,91
poor	0,10	0,00	99,90
poor	0,02	0,00	99,98
poor	0,40	0,10	99,51
poor	0,39	0,00	99,61