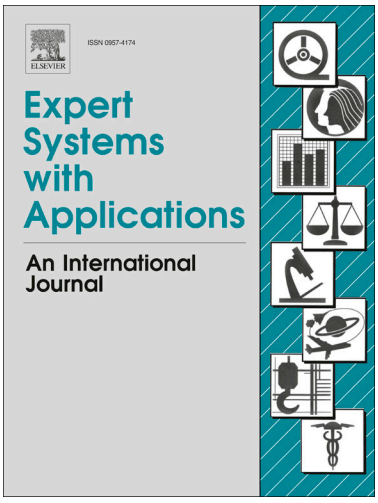
Journal Pre-proofs



One-shot learning for acoustic diagnosis of industrial machines

Stavros Ntalampiras

PII: S0957-4174(21)00425-5
DOI: https://doi.org/10.1016/j.eswa.2021.114984
Reference: ESWA 114984

To appear in: *Expert Systems with Applications*

Received Date: 17 July 2020
Accepted Date: 30 March 2021

Please cite this article as: Ntalampiras, S., One-shot learning for acoustic diagnosis of industrial machines, *Expert Systems with Applications* (2021), doi: https://doi.org/10.1016/j.eswa.2021.114984

One-shot learning for acoustic diagnosis of industrial machines

Stavros Ntalampiras

University of Milan, via Celoria 18, 20133, Milan, Italy, e-mail: stavros.ntalampiras@unimi.it

Abstract

Automatic acoustic monitoring of machine health comprises a relevant field as, unfortunately, such equipment often suffers from faults, malfunctions, aging effects, etc. However, it is still an unexplored domain of research where the majority of existing works relies on traditional machine learning based approaches. After providing a critical survey of the available methods, this work highlights the most relevant limitations and designs a solution specifically addressing them. We introduce the one-shot learning paradigm into the specific domain and suitably extent it to a) classify machine states, b) detect novel ones, and c) incorporate them in the class dictionary online. The backbone of the present system is a Siamese Neural Network (SNN) composed of convolutional layers. Conveniently, every processing stage depends on a standardized feature set free of domain knowledge, i.e. spectrograms. Interestingly, we enhance SNN's classification ability by an appropriately designed data selection scheme. The proposed solution is applied on a publicly available dataset of vibration signals representing four states of a drill bit, i.e. healthy state, chisel wear, flank wear, and outer corner wear. After extensive experiments thoroughly examining every aspect of the proposed solution, it is shown to achieve state of the art results while using limited amount of training data. Importantly, at the same time it is able to operate under evolving environments. Last but not least, we show that the obtained predictions are interpretable, a property which is rapidly becoming a requirement in modern machine learning based technologies.

Keywords: Machine acoustics, machine health condition monitoring, one-shot learning, fault diagnosis, deep learning, online learning.

1. Introduction

The field of machine acoustics aims at assessing the health of generic machine equipment based on its acoustic and/or vibration emissions (Nandi and Ahmed, 2019; Yang et al., 2019). Unfortunately, industrial (and not only) equipment suffers from aging effects, faults, malfunctions, etc. which might have serious consequences not only in terms of production but also w.r.t human lives and/or property loss (Gurina et al., 2020; Lasisi et al., 2019; Xue et al., 2019; Pontoppidan et al., 2019) and other interconnected devices in the IoT context (Zanella et al., 2014; Liu et al., 2020; Alippi et al., 2016; Ntalampiras, 2018). Thus, machine health assessment technologies are becoming attractive since they can provide accurate and real-time diagnostics regarding the machine of interest without relying on human supervision (Nasir et al., 2019). Following the principles of generalized sound recognition technology (Ntalampiras, 2019), machine acoustics pipeline typically encompasses a signal processing and a pattern recognition phase (Wunderlich et al., 2018). The first is responsible for capturing characteristic properties of the involved signals, while the second aims at identifying patterns of normal and abnormal operation.

Interestingly, such solutions have provided reliable performance including advanced diagnostics in a series of automated machine condition monitoring applications, e.g. drill bit (Rafezi and Hassani, 2018), bearings (Zhou et al., 2007; Wang et al., 2017; Fu et al., 2018), wind turbine gearbox (Yang et al., 2018), petrochemical unit (Xiong et al., 2018), etc. As a thorough overview of such solutions is out of the scope of this work, we point out here the typical line of thought. The signal processing phase is based on handcrafted features, which heavily depends of domain knowledge. Such features may belong to time, frequency, or wavelet domains, while they could be also employed simultaneously (Dai et al., 2014; Lashari et al., 2019). Subsequently, a variety of discriminative and non-discriminative pattern recognition techniques have been used to model the distribution exhibited by the extracted features. Discriminative ones seek boundaries separating the categories existing in the training data (Nandi and Ahmed, 2019); these primarily embrace decision trees, artificial neural networks including deep learning (Shevchik et al., 2019), and support vector machines. Non-discriminative ones process the data of each class independently with respect to the rest while most techniques are based on probabilistic theory, e.g. hidden Markov models (Ntalampiras, 2015).

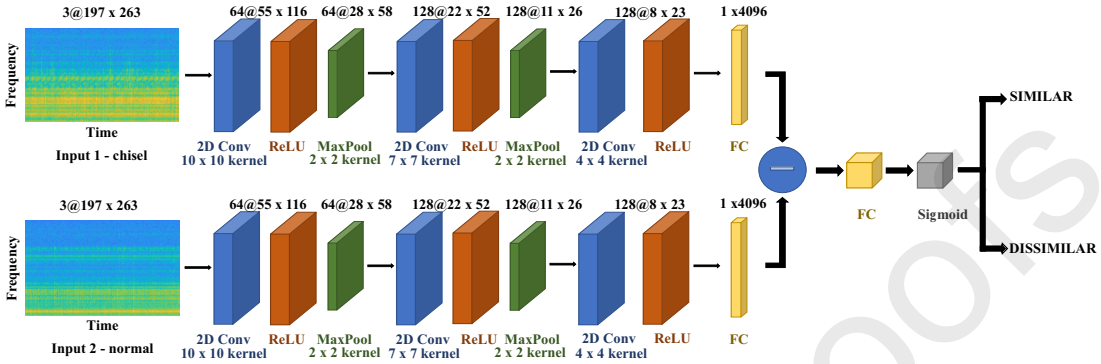


Figure 1: The pipeline of the proposed one-shot learning scheme using Siamese neural networks. Each input is passed though a series of convolutional, ReLU and max-pooling layers completed by a common end based on binary cross-entropy loss.

This paper focuses on drill bit monitoring which is an interesting problem of increased relevance as regards to automation in manufacturing industries, where excellent quality and efficiency requirements need to be met (Vununu et al., 2018; Sangeetha B. and S., 2019; Xu et al., 2014). As such, the present problem has attracted the interest of the scientific community. The authors of (Verma et al., 2015) present a drill bit monitoring system based on a support vector machine with a radial basis kernel function elaborating on carefully designed features coming from the frequency domain. Interestingly, in (Rafezi and Hassani, 2018) a tricone bit health monitoring system is presented. Vibration signal analysis is carried out via wavelet packet decomposition feeding an artificial neural network. Moving on, in (Dai et al., 2014) the authors present a condition monitoring system for bone drilling by elaborating the standard deviation of wavelet coefficients. Their case study includes evaluation through the drilling operation in *in vitro* porcine spines. Last but not least, a paper on drill bit monitoring using deep learning and spectral analysis of acoustic emissions is presented in (Vununu et al., 2018). The following gaps are observed in the related literature:

- *poor data availability*: most existing solutions elaborate on proprietary datasets not available to the research community limiting reproducibility and comparability. At the same time, obtaining data representative of faults, malfunctions, aging effects, etc. presents increased difficulty,

- *mandatory domain knowledge*: the majority of existing works is based

on profound knowledge of the problem specifications and the available dataset to properly design the employed features,

- *inability to operate in the concept drift environment*: the related literature assumes complete knowledge of a stationary class dictionary during both training and testing (Ditzler et al., 2015).

The solution proposed in this work closes the above mentioned gaps by suitably modifying the one-shot learning paradigm. *One-shot learning* is defined as the classification task strictly bounded by the condition that we may observe only a single sample belonging to each class in order to make inference(s) regarding test samples. In essence, the problem is solved by training a mechanism able to make predictions on the similarity of the test samples to those *a-priori* available. Such a line of thought has been explored in image recognition (e.g. handwritten character recognition (Lake et al., 2015), feature learning (Zhu et al., 2020), etc.) reaching state of the art results. In acoustic signal processing, one-shot learning is still unexplored with the exception of generative speech concepts (Lake et al., 2014).

We demonstrate the efficacy of one-shot learning via exhaustive experiments including the case of non-stationary environments. More specifically, we consider a drill bit monitoring application including four states (one healthy and three faulty) using a publicly available dataset. The main novel points of this work are:

- removes the need of handcrafted features,
- reaches state of the art accuracy using a small amount of training data,
- designs a reliable mechanism to detect and react to changes in the environment,
- incrementally updates the class dictionary in an online manner, and
- provides an interpretation of the predictions made by the proposed machine learning based solution.

In the following, we a) formalize the problem, b) delineate the proposed solution, c) describe the experimental protocol along with a detailed analysis of the obtained results, and d) draw conclusions and briefly discuss potential extensions.

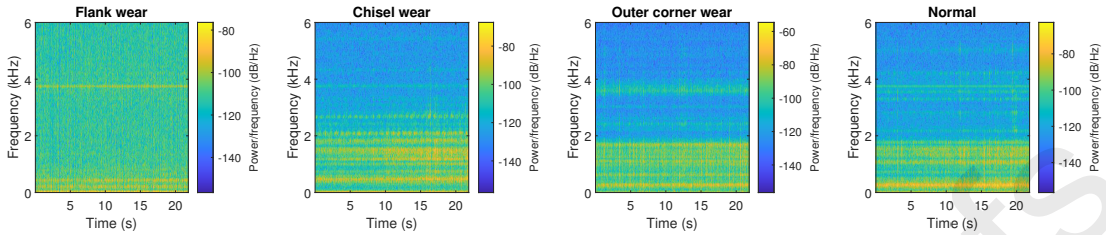


Figure 2: Representative spectrograms extracted out of the four types of machine conditions considered in this work, i.e. a) flank wear, b) chisel wear, c) outer corner wear, and d) normal.

2. Problem formulation

We suppose a single-channel vibration signal y containing machine emissions is available, while dictionary $\mathcal{D}$ comprises the set of machine health states $\mathcal{D} = \{S_1, \ldots, S_m\}$, where m is the number of a-priori known states. No assumption is made regarding composition and size of $\mathcal{D}$ which is unbounded and may alter at any point in time. Moreover, we suppose that emissions representative of a specific state follow a consistent, yet unknown probability density function $P_i, i \in [1, m]$ (Ntalampiras and Potamitis, 2019; Umapathy et al., 2005; Stowell et al., 2015; Ntalampiras, 2016).

The last assumption is the availability of an initial training dataset $TS = y_t, t \in [1, T_0]$ with labeled pairs (y_t, S_i), where t is the time instant and S_i the machine state with $i \in [1, m]$. On the contrary, we make no assumption regarding as to if/when an unknown machine state might occur. The ultimate goal is to identify the machine state accurately and update dictionary $\mathcal{D}$ appropriately.

3. One-shot learning for acoustic change detection, dictionary learning and machine state identification

The backbone of the proposed system is a Siamese Neural Network (SNN) suitably learning *similar* and *dissimilar* relationships using training data in TS. Interestingly, when there functionalities are appropriately exploited, they can serve classification purposes. In addition, such a method can be followed in classification tasks involving data of different modalities (e.g. text, infrared, video, etc.) with only slight alterations. The following subsections describe the SNN architecture, feature extraction module, as well as the processes for change detection and machine state identification.

3.1. Siamese Neural Networks

SNNs were first presented in (Bromley et al., 1994) with application onto the signature verification problem. An SNN is composed by a twin network connected to a shared ending point (Tian et al., 2020). However, each twin processes diverse inputs as illustrated in Fig. 1. The ending point includes calculation of a predetermined metric based on the highest-level representation learned by each twin. Thus, the networks aim at satisfying the same goal with a unique optimization function leading to tied weights and assuring that analogous inputs are going to be closely located in the feature space. It is worth noting that the twins have symmetric topologies meaning that they are interchangeable. In other words, the same metric value will be produced even if we reverse (top/bottom) networks' positions and/or inputs. As regards to the metric function, we employed binary cross entropy loss followed by a sigmoid activation, conveniently normalizing the output into the [0,1] interval.

Following the advances in literature related to audio signal processing (Purwins et al., 2019), each twin in SNN is composed of convolutional layers. Convolutional neural networks (CNNs) have been very successful in generalized audio classification, including speech and music, and combined with their implementation simplicity, have become quite popular with the specific research community. Motivated by these observations, we decided to employ them for vibration signal processing due to the existing similarities.

CNNs include simple alterations in the traditional multilayer perceptron model, i.e. a) their topology includes several stacked layers, while b) each convolutional layer is succeeded by a max-pooling one. The interesting property of CNNs is that stacked convolutional layers customize the neurons so that locally limited structures are emphasized in the 2D plane. That said, each hidden unit 'sees' only a small part of the input instead of its totality. This part is typically referred to as unit's *receptive field*. Weights characterizing the hidden units are learned based on the presented set of inputs and construct a feature map capturing their properties. Given that the dimensionality of such feature maps could be excessive, max-pooling layers are inserted where the maximum value of neighboring units is retained to represent the entire neighbor. Interestingly, this operation is proven to render the network robust to translational shifts (Piczak, 2015). Finally, we considered rectified linear units (ReLu), i.e. the activation function is $f(x) = max(0, x)$.

1. Input: test vibration signal y^t, trained SNN $\mathcal{N}$, dictionary $\mathcal{S} = \{S_1, \ldots, S_m\}$, while each class is represented by extracted spectrograms $\langle \mathcal{F}_S^i \rangle_{i=1}^{i=|S|}$;
2. Extract spectrogram s of y^t ;
3. Initialize similarity vector $V = []$;
4. **for** *j=1:m* **do**
   5. **for** *i=1:|S|* **do**
      6. Query $\mathcal{N}$ with the pair $\{s, \mathcal{F}_j^i\}$ and get similarity score $V(j, i)$;
   end
end
7. Predict the class maximizing the similarity score $S^* = \arg\max_{\mathcal{S}} \{V(:, i)\}$ and assign it to y^t;

Algorithm 1: The proposed machine state identification algorithm using vibration signals based on one-shot learning ($|\bullet|$ denotes the cardinality operator).

3.2. Model architecture

The present SNN is composed of 3 layers as depicted in Fig. 1 (experimenting with more layers did not provide improved performance). The first two convolutional layers are followed by ReLu and max-pooling ones; the last one however, is succeeded by a fully-connected one. A distance operation concludes the SNN along with a fully-connected layer and a sigmoid function deciding on the inputs' relationship (similar/dissimilar) via thresholding.

The filters composing convolutional layers have varying size with a stationary stride equal to 1. At the same time, max-pooling layers have 2×2 kernels with $stride = 2$. SNN is concluded by a flattening layer collecting the entirety of units included in the last convolutional twin layers and the distance calculation follows.

During the training process, binary cross-entropy loss among network's prediction and ground truth is computed towards updating the SNN. The process is based on the standard version of backpropagation algorithm where the gradient sums the weights of each twin network. Minibatch size is suitably selected based on the training set, while learning rate is 6e−5. Weights and biases are initialized using narrow normal distributions with zero-mean and 0.01 standard deviation, while the maximum number of permitted epochs is

Table 1: Confusion matrix (in %) obtained while considering 75% of training data.

Presented \ **Predicted**	*Similar*	*Dissimilar*
Similar	**95.9**	4.1
Dissimilar	2.5	**97.5**

2000.

3.3. Feature extraction

In the related literature, vibration emissions of machines are characterized by features extracted from the frequency domain. Towards eliminating the need of any feature engineering process, we make use of the entire spectrogram without any type of post-processing. To this end, we employ short time Fourier transform with FTT size equal to 1024. Spectrograms representative of the four considered machine states are shown in Fig. 2.

3.4. Change detection and machine state identification

The above described SNN learns the similar/dissimilar relationship between pairs of spectrograms extracted from the available vibration emissions. Based on this property, we propose a direct extension of the one-shot learning paradigm facilitating change detection and classification applications. Under such a learning scheme, a change is detected in case a novel spectrogram is recognized as dissimilar to every state existing in $\mathcal{D}$. Subsequently, the specific sample forms an additional class which increments $\mathcal{D}$. On the opposite case, the class with the maximum similarity score is predicted to characterize the novel spectrogram. The SNN can efficiently address classification tasks in poor data availability environments.

When the change detection test does not signal a change, the proposed algorithm continues to machine state identification following Alg. 1 which requires fours arguments (Alg. 1, line 1)

- an vibration signal to test, denoted as y^t,
- the trained SNN $\mathcal{N}$,
- the dictionary $\mathcal{D} = \{S_1, \ldots, S_m\}$, while each class is represented by extracted spectrograms $\langle \mathcal{F}_{\mathcal{S}}^i \rangle_{i=1}^{i=|S|}$, and

Table 2: $\mathcal{M}^s$ (in %) achieved by while employing 75% of training data in stationary conditions (maximum rates are emboldened).

Input 2 / ***Input 1***	*Flank*	*Chisel*	*Outer corner*	*Normal*
Flank	**88.5**	5.1	6.4	-
Chisel	4	**92.5**	-	3.5
Outer corner	5.9	-	**90.1**	4
Normal	-	5.1	5.4	**89.5**

- the available spectrograms $\langle \mathcal{F}_{\mathcal{S}}^i \rangle_{i=1}^{i=|S|}$.

The proposed algorithm extracts the spectrogram s of y^t (Alg. 1, line 2) and initializes similarity vector V (Alg. 1, line 3). Subsequently, we query $\mathcal{N}$ using the existing pair combinations which outputs the corresponding similarity scores, thus updating V (Alg. 1, line 4-6). The last step of the algorithm assigns to y^t the label of the class using the maximum similarity score existing in V (Alg. 1, line 7).

4. Experimental set-up and results

This section describes the a) employed dataset, b) suitably formed figures of merit, c) contrasted methods, d) obtained results in both stationary and evolving environments and e) interpretation of SNN's operation towards class assignment.

4.1. Dataset

In order assess extensively the performance achieved by the proposed method we used a dataset specifically designed for drill bit monitoring. Interestingly, the dataset is available to the research community facilitating reproducibility (Verma et al., 2015). There, an experimental protocol is proposed as well for comparability purposes.

During a drilling process, the following parts are required: a) a drilling machine, b) a work piece, c) a fixture, and d) a cutting tool. A drill bit is characterized by a cone structure including chisel edge, cutting lips, web, flute, heels, body and shank. The drill bit contacts and crack the work piece by means of chisel edge. After the *penetration* stage, the drill bit enters the material, the so-called *steady* stage. Naturally, drill bit suffers from

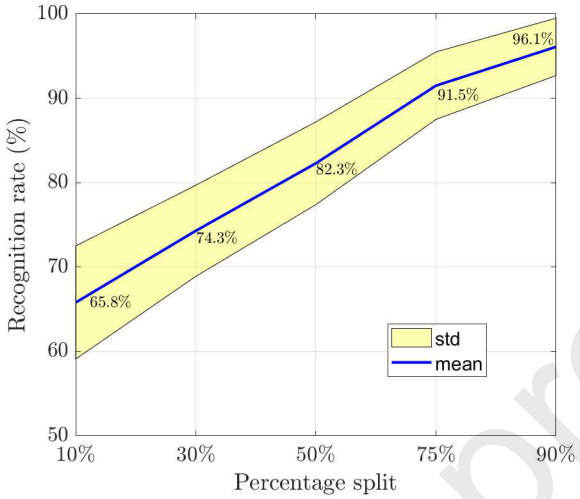


Figure 3: Recognition rate vs. size of training set under stationary conditions.

aging effects and malfunctions when used for large periods of time, hence deformations appear. The present dataset includes four states, i.e. a healthy one and three faulty:

a) *Chisel wear*: due to extreme stress, the temperature at chisel point reaches very high levels and its edge debilitates,

a) *Flank Wear*: the friction between the work piece and flank of drill bit causes erosions, which worsen as cutting speed escalates, and

a) *Outer corner wear*: the enormous impact forces and friction existing among drill bit and hole's inner head affect the outer corner which abrades and, over time, gets destroyed.

More information on the dataset is available in (Verma et al., 2015). Conveniently, the dataset is perfectly balanced across classes with 30 samples per class of equal duration. As the size is not particularity large, applying the proposed one-shot learning algorithm could be beneficial as very deep networks may suffer from overfitting. We used the preprocessing conducted in (Verma et al., 2015). The sampling frequency is 12kHz, while we used frames of 0.02ms overlapping by 50% to compute STFT and extract the spectrograms depicted in Fig. 2.

Table 3: $\mathcal{M}^d$ (in %) achieved by while employing 75% of training data in stationary conditions (minimum rates are emboldened).

Input 2 / ***Input 1***	*Flank*	*Chisel*	*Outer corner*	*Normal*
Flank	**11.5**	94.9	93.6	100
Chisel	96	**7.5**	100	96.5
Outer corner	94.1	100	**9.9**	96
Normal	100	94.9	94.6	**10.5**

4.2. Figures of merit

We employed effective and widely-used figures of merit assessing the performance of every method thoroughly. One interesting detail for the case of one-shot learning is that we can additionally employ confusion matrices at the entire dataset level demonstrating the efficacy of the method in recognizing similarities and dissimilarities. To this end, the following matrix was defined:

$$\mathcal{M}^s = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}, \tag{1}$$

where

- s_{11} (in %) denotes the number of times that samples fed in the first input of SNN were identified as similar to samples coming from the same class,
- s_{12} (in %) denotes the number of times that samples fed in the first input of SNN were identified as dissimilar to samples coming from the same class,
- s_{22} (in %) denotes the number of times that samples fed in the second input of SNN were identified as similar to samples coming from the same class,
- s_{21} (in %) denotes the number of times that samples fed in the second input of SNN were identified as dissimilar to samples coming from the same class.

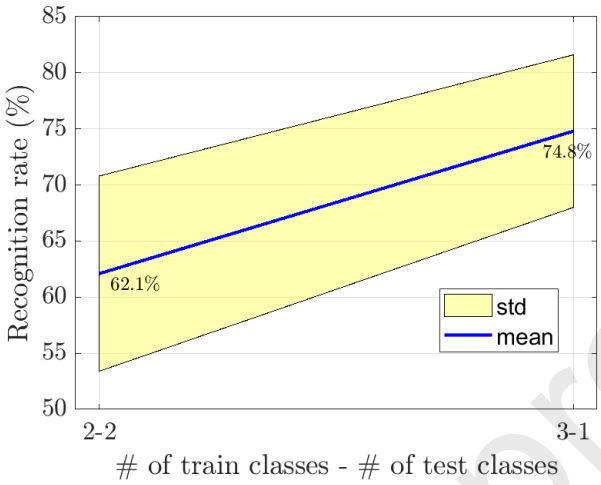


Figure 4: Recognition rate vs. number of classes in $\mathcal{D}$ under non-stationary conditions.

In this case, the objective it to maximize the values in the diagonal. A matrix assessing the dissimilarities $\mathcal{M}^d$ can be defined in an analogous way with the difference being that we are aiming at minimizing its diagonal. Interestingly, the sum of similarity and dissimilarity matrices characterizing the accuracy of a given method is 100%, i.e. $\mathcal{M}^s + \mathcal{M}^d = 100$ for every element.

4.3. Contrasted method

The proposed method is contrasted against the one presented in (Verma et al., 2015) which represents the state of the art in vibration based machine health diagnosis. There, the authors, after exploring time, frequency and wavelet domain features, they show that frequency based ones modeled by an SVM with RBF kernel offer the best performance reaching 95.5% in a 4-fold cross validation scheme. Towards a reliable comparison, we followed the specific experimental protocol unless specified otherwise. Furthermore, we do not make any differentiation between *steady* and *penetration* states aiming at a generically applicable system. In addition, we applied a Gaussian mixture model (GMM) based classification scheme, where the number of components was selected from the set $\{2, 4, 8, 16, 32, 64, 128\}$ following the maximum recognition rate criterion. As regards to the respective learning process, GMMs with diagonal covariance matrices were constructed, while the number of k-means iterations for cluster initialization was set to 100. The

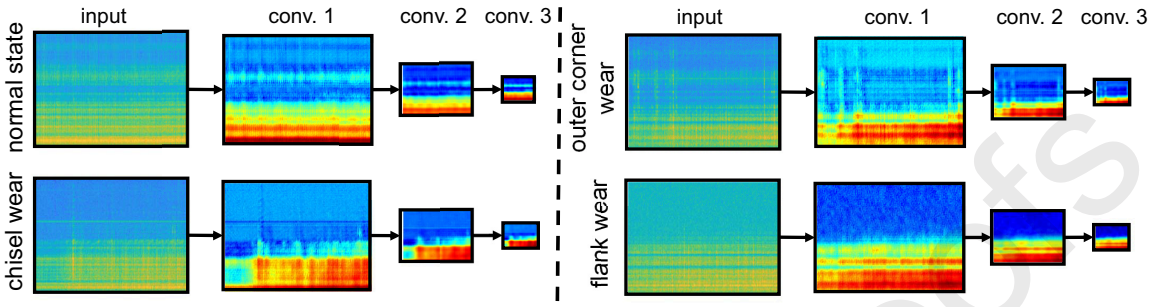


Figure 5: The three convolutional layers as they are activated by spectrograms representative of every considered class, i.e. *normal state*, *flank*, *chisel*, and *outer corner wear*.

highest recognition rate was achieved by the GMM formed by 8 components and was equal to 85.8%.

4.4. Experimental design under stationary conditions

During the specific experimental phase, we assume knowledge of every class, i.e. $\mathcal{D}$ includes all states described in section 4.1. We focus on the impact that the amount of training data has on the recognition rate, thus we employ the following splits $split = \{10\%, 30\%, 50\%, 75\%, 90\%\}$. The case of 75% follows the splitting used in the contrasted method. The experiment associated with each split was iterated 50 times with randomly chosen data among iterations. Fig. 3 reports average and standard deviation of the recognition rate w.r.t each split setting. As regards to the SNN, the maximum number of permitted epochs is 2000 with early stopping, the mini bath size 50, test batch 200 while the number of similarity/dissimilarity tests is 20. It should be noted that similar and dissimilar input pairs were produced randomly and have equal sizes.

4.5. Performance under stationary conditions

Fig. 3 demonstrates the recognition rates achieved w.r.t various percentage splits along with the corresponding standard deviations. At this stage, complete knowledge of $\mathcal{D}$ is assumed. We observe that the proposed scheme is able to provide rates ranging from 65%-96.1% as the amount of training data increases. Interestingly, even at the 10% setting, i.e. only 3 training files per class, the average rate is substantially higher that chance. At the same time, it should be taken into account that the SNN is not trained specifically

for classification but only for similarity assessment. In the 75% percentage split, SNN provides 91.5±4% which is at similar levels with the state of art classification method. At the same time, it surpasses the GMM-based classification scheme.

Table 1 tabulates the confusion matrix including similarities and dissimilarities w.r.t the 75% split. We see that the SNN is better at recognizing dissimilarities (97.5%) than similarities (95.9%). Furthermore, M^s and M^s are presented in Tables 2 and 3 respectively. We observe that *chisel* are correctly classified as similar with the highest rate and dissimilar with the lowest one. On the opposite side, *flank* is classified as similar with the lowest rate and dissimilar with the highest one. Overall, we observe analogous rates among the considered classes.

4.6. Experimental design under non-stationary conditions

This phase assumes no knowledge regarding composition and cardinality of dictionary $\mathcal{D}$; the only assumption requests that TS includes data coming from at least 2 classes allowing the SNN to learn similar and dissimilar relationships. It comes out that SNN will be tested on data belonging to classes unseen during training. To the best of our knowledge, this is the first time that the specific scenario is considered in the related literature. Such a necessity arises from the fact that it is unreasonable to assume complete knowledge of every fault, malfunction, aging effect, etc. the drill will undergo in the future. The performance of the proposed solution was evaluated with a varying number of unknown classes. Each experimental configuration was carried out 50 times, and here we present average and standard deviation value of the recognition rate. It should be mentioned that classes in TS were selected randomly w.r.t each iteration.

This experimental phase was carried out by considering every possible pair of classes in TS, while the rest comprised the testing classes. As regards to the SNN settings, the number of epochs had an upper limit of 2000 (early stopping was included), the minibatch size is 50, test batch 100, while and the number of tests per class quantifying similarity/dissimilarity was 10. Following the previous experimental phase, random generation of similar and dissimilar input pairs is balanced.

4.7. Performance under non-stationary conditions

Fig. 4 shows the recognition accuracy (average and standard deviation) as the number of known/unknown classes varies. As expected, the rate in-

creases as population in $\mathcal{D}$ surges. At the same time, the rates are lower with higher standard deviation values than those achieved under stationary conditions. The highest rate concerns having knowledge of 3 classes (74.8±6.8). Nonetheless, the performance depends significantly on the composition of the test and train class sets. More specifically, performance is boosted when TS includes data satisfying two conditions a) high intraclass similarities and b) high interclass dissimilarities. During this experimental phase, sample selection was random resulting to high standard deviation values. Motivated by this observation, section 4.9 presents a simple technique maximizing the potential of the proposed one-shot learning based solution.

4.8. Convolutional filters

This section investigates in detail the exact way SNN processes the input spectrograms via its convolutional filters localizing the parts mostly facilitating machine state identification. Fig. 5 how the three convolutional layers are activated by spectrograms representative of every considered class, i.e. *normal state*, *flank*, *chisel*, and *outer corner wear*.

We can see that each layer provides a simplified view of the obtained input, while concentrating on a distinctive part able to discriminate the considered classes. From the spectrograms, we observe that not every part of the spectrum is equally descriptive for all classes, hence SNN focuses on different parts depending on the given class. More specifically,

- for *normal state* spectrograms, SNN gives high importance to two discrete low-frequency bands and low values to a specific high-frequency band,
- for *chisel wear* spectrograms, SNN focuses on late appearing low-frequency bands without any type of discretization, and considers equally unimportant most high-frequency bands,
- for *outer corner wear* spectrograms, SNN concentrates on low-frequency bands in a relatively discrete way, while specific high-frequency are considered important, and
- for *flank wear* spectrograms, SNN gives significantly more importance on low-frequency bands w.r.t the high ones; interestingly, its focus is uniform.

A careful examination of the convolutional filters provides a meaningful interpretation of SNN's operation, which is a highly desired property towards reliable and trustworthy machine learning based solutions leading to wider acceptance of such automated systems (Garbuk, 2018; European Commission, 2020).

4.9. Distance-based selection of training data

Motivated by the findings in section 4.7 we designed an algorithm facilitating appropriate selection of training data. Given TS and $\mathcal{D}$, i.e. training samples, composition and number of known classes, the objective is to discover the most central w.r.t each class distribution samples and, at the same time, the most distant samples to the rest of the classes. Fig. 6 illustrates the available feature space using t-SNE operating on features of reduced dimensionality thanks to PCA. To this end, we compute the within-class sum of squared distances for every available sample and the corresponding interclass sum as follows

$$w = \sum_{i \in S, i=1}^{i=|S|} ||x - x_i||^2, n = \sum_{i \notin S, i=1}^{i=|\mathcal{D}|} ||x - x_i||^2.$$

Finally, we select the samples minimizing the quantity $(w - n)$ to populate TS. Interestingly, such an approach was applied under stationary conditions and improved the performance. More specifically, we obtained 94.2±0.3% after following a 4-fold cross validation experimental protocol as done in (Verma et al., 2015). The present technique is rather simple, computationally inexpensive for small-sized datasets (which is the typical application scenario for one-shot learning based solutions) and generically applicable to other datasets of homogeneous or heterogeneous modalities.

5. Conclusion

This article described an one-shot learning based solution specifically designed to identify machine health state from vibration signals. Interestingly, the proposed solution is able to offer state of the art results without considering any type of feature engineering based on domain knowledge. At the same, it is able to operate satisfactorily in non stationary environments via a suitably designed change detection mechanism allowing the system to react to the appearance of new machine states and incorporate them in the class

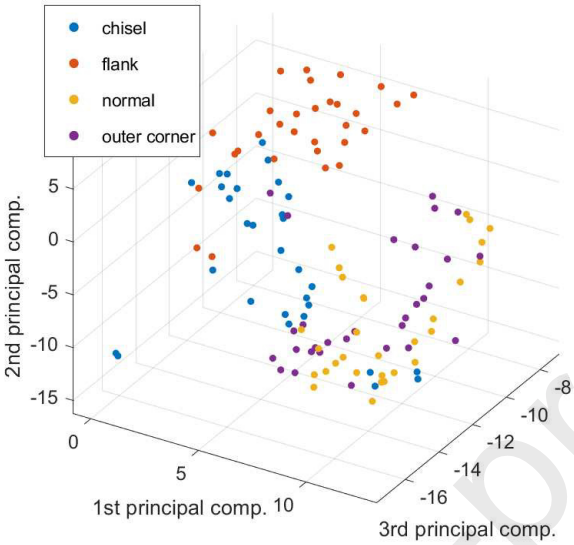


Figure 6: t-SNE plot of the feature space reduced to 50 PCA components including data coming from every class (*normal state*, *flank*, *chisel*, and *outer corner wear*).

dictionary on-the-fly. Importantly, the system is not specifically trained for classification but only learns the relationship similar/dissimilar between input pairs. A thorough experimental campaign was followed on a publicly available dataset including four machine health states. Towards assessing every aspect of the proposed system, apart from figures of merit widely used in the related literature, a novel one was designed able to demonstrate the capability in identifying similar and dissimilar input pairs. Last but not least, we analyzed the convolutional layers towards isolating spectrograms' regions relevant for identifying each class. It was shown that such regions are easily interpretable by humans contributing to wider acceptance of machine learning based solutions. Such a feature is rapidly becoming a standard requirement in machine learning based solutions (Gu et al., 2020). We argue that a relevant part contributing to the success of this solution is its ability to consider both similarities and dissimilarities to known classes at the same time.

In the future, we are going to work on the following extensions: a) given the flexibility of the proposed solution, we wish exploit it to solve different problems of similar requirements, b) investigate sufficient conditions w.r.t data composition and quantity towards improving the performance achieved

in non-stationary environments, c) develop an advanced method to optimally select data samples w.r.t every class available during training considering restrictions related to both intraclass similarities and interclass dissimilarities combined with poor data availability, and d) extent the current theoretical framework towards including data coming from heterogeneous modalities which could potentially offer improved performance.

Acknowledgments

We gratefully acknowledge the support of NVIDIA Corp. with the donation of the Titan V GPU used for this research. This work was carried out within the project entitled Advanced methods for sound and music computing funded by the Piano Sostegno alla Ricerca of University of Milan.

References

Alippi, C., Ntalampiras, S., Roveri, M., 2016. Online model-free sensor fault identification and dictionary learning in cyber-physical systems. In: 2016 International Joint Conference on Neural Networks (IJCNN). pp. 756–762.

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R., 1994. Signature verification using a "siamese" time delay neural network. In: Cowan, J. D., Tesauro, G., Alspector, J. (Eds.), Advances in Neural Information Processing Systems 6. Morgan-Kaufmann, pp. 737–744.

Dai, Y., Xue, Y., Zhang, J., 2014. Condition monitoring based on sound feature extraction during bone drilling process. In: Proceedings of the 33rd Chinese Control Conference. pp. 7317–7322.

Ditzler, G., Roveri, M., Alippi, C., Polikar, R., Nov 2015. Learning in nonstationary environments: A survey. IEEE Computational Intelligence Magazine 10 (4), 12–25.

European Commission, 19 February 2020. White paper on artificial intelligence: a european approach to excellence and trust. Tech. rep., Brussels.

Fu, Q., Jing, B., He, P., Si, S., Wang, Y., 2018. Fault feature selection and diagnosis of rolling bearings based on eemd and optimized elmanadaboost algorithm. IEEE Sensors Journal 18 (12), 5024–5034.

Garbuk, S. V., 2018. Intellimetry as a way to ensure ai trustworthiness. In: 2018 International Conference on Artificial Intelligence Applications and Innovations (IC-AIAI). pp. 27–30.

Gu, D., Li, Y., Jiang, F., Wen, Z., Liu, S., Shi, W., Lu, G., Zhou, C., 2020. Vinet: A visually interpretable image diagnosis network. IEEE Transactions on Multimedia, 1–1.

Gurina, E., Klyuchnikov, N., Zaytsev, A., Romanenkova, E., Antipova, K., Simon, I., Makarov, V., Koroteev, D., Jan. 2020. Application of machine learning to accidents detection at directional drilling. Journal of Petroleum Science and Engineering 184, 106519.
URL https://doi.org/10.1016/j.petrol.2019.106519

Lake, B. M., Salakhutdinov, R., Tenenbaum, J. B., Dec. 2015. Human-level concept learning through probabilistic program induction. Science 350 (6266), 1332–1338.
URL https://doi.org/10.1126/science.aab3050

Lake, B. M., ying Lee, C., Glass, J. R., Tenenbaum, J., 2014. One-shot learning of generative speech concepts. In: Bello, P., Guarini, M., McShane, M., Scassellati, B. (Eds.), CogSci. cognitivesciencesociety.org.

Lashari, S., Takbiri-Borujeni, A., Fathi, E., Sun, T., Rahmani, R., Khazaeli, M., Apr. 2019. Drilling performance monitoring and optimization: a data-driven approach. Journal of Petroleum Exploration and Production Technology 9 (4), 2747–2756.
URL https://doi.org/10.1007/s13202-019-0657-2

Lasisi, A., Sadiq, M. O., Balogun, I., Tunde-Lawal, A., Attoh-Okine, N., 2019. A boosted tree machine learning alternative to predictive evaluation of nondestructive concrete compressive strength. In: 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA). pp. 321–324.

Liu, P., Zhang, Y. F., Wu, H., Fu, T., 2020. Optimization of edge-plc based fault diagnosis with random forest in industrial internet of things. IEEE Internet of Things Journal, 1–1.

Nandi, A., Ahmed, H., Dec. 2019. Introduction to machine condition monitoring.
URL https://doi.org/10.1002/9781119544678.ch1

Nandi, A. K., Ahmed, H., 2019. Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines, 1st Edition. Wiley-IEEE Press.

Nasir, V., Cool, J., Sassani, F., Oct. 2019. Intelligent machining monitoring using sound signal processed with the wavelet method and a self-organizing neural network. IEEE Robotics and Automation Letters 4 (4), 3449–3456.
URL https://doi.org/10.1109/lra.2019.2926666

Ntalampiras, S., 2015. Fault identification in distributed sensor networks based on universal probabilistic modeling. IEEE Transactions on Neural Networks and Learning Systems 26 (9), 1939–1949.

Ntalampiras, S., 2016. Automatic analysis of audiostreams in the concept drift environment. In: 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP). pp. 1–6.

Ntalampiras, S., 2018. Fault diagnosis for smart grids in pragmatic conditions. IEEE Transactions on Smart Grid 9 (3), 1964–1971.

Ntalampiras, S., Oct. 2019. Generalized sound recognition in reverberant environments. JAES 67 (10), 772–781.

Ntalampiras, S., Potamitis, I., May 2019. A statistical inference framework for understanding music-related brain activity. IEEE Journal of Selected Topics in Signal Processing 13 (2), 275–284.

Piczak, K. J., Sep. 2015. Environmental sound classification with convolutional neural networks. In: 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP). pp. 1–6.

Pontoppidan, N. H., Lehn-Schiøler, T., Petersen, K. B., 2019. Machine learning for condition monitoring and innovation. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 8067–8071.

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

Rafezi, H., Hassani, F., 2018. Tricone bit health monitoring using wavelet packet decomposed vibration signal. In: 2018 5th International Conference on Control, Decision and Information Technologies (CoDIT). pp. 1012–1016.

Sangeetha B., P., S., H., 2019. Rational-dilation wavelet transform based torque estimation from acoustic signals for fault diagnosis in a three-phase induction motor. IEEE Transactions on Industrial Informatics 15 (6), 3492–3501.

Shevchik, S. A., Masinelli, G., Kenel, C., Leinenbach, C., Wasmer, K., Sep. 2019. Deep learning for in situ and real-time quality monitoring in additive manufacturing using acoustic emission. IEEE Transactions on Industrial Informatics 15 (9), 5194–5203.
URL https://doi.org/10.1109/tii.2019.2910524

Stowell, D., Giannoulis, D., Benetos, E., Lagrange, M., Plumbley, M. D., 2015. Detection and classification of acoustic scenes and events. IEEE Transactions on Multimedia 17 (10), 1733–1746.

Tian, S., Liu, X., Liu, M., Li, S., Yin, B., 2020. Siamese tracking network with informative enhanced loss. IEEE Transactions on Multimedia, 1–1.

Umapathy, K., Krishnan, S., Jimaa, S., 2005. Multigroup classification of audio signals using time-frequency parameters. IEEE Transactions on Multimedia 7 (2), 308–315.

Verma, N. K., Sevakula, R. K., Dixit, S., Salour, A., Feb 2015. Data driven approach for drill bit monitoring. pp. 19–26.

Vununu, C., Moon, K.-S., Lee, S.-H., Kwon, K.-R., Aug. 2018. A deep feature learning method for drill bits monitoring using the spectral analysis of the acoustic signals. Sensors 18 (8), 2634.
URL https://doi.org/10.3390/s18082634

Wang, Z., Zhang, Q., Xiong, J., Xiao, M., Sun, G., He, J., 2017. Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests. IEEE Sensors Journal 17 (17), 5581–5588.

Wunderlich, C., Tschöpe, C., Duckhorn, F., 2018. Advanced methods in NDE using machine learning approaches. Author(s).
URL https://doi.org/10.1063/1.5031519

Xiong, J., Liang, Q., Wan, J., Zhang, Q., Chen, X., Ma, R., 2018. The order statistics correlation coefficient and ppmcc fuse non-dimension in fault diagnosis of rotating petrochemical unit. IEEE Sensors Journal 18 (11), 4704–4714.

Xu, L. D., He, W., Li, S., 2014. Internet of things in industries: A survey. IEEE Transactions on Industrial Informatics 10 (4), 2233–2243.

Xue, J. Z., Ran Lin, T., Xing, J. P., Ni, C., 2019. Bearing fault diagnosis based on adaptive variational mode decomposition. In: 2019 Prognostics and System Health Management Conference (PHM-Qingdao). pp. 1–5.

Yang, H., Luo, T., Li, L., Rao, Y., Li, W., Liu, K., Qiu, Z., 2019. Research on drilling bit positioning strategy based on sins mwd system. IEEE Access 7, 109398–109410.

Yang, Q., Hu, C., Zheng, N., 2018. Data-driven diagnosis of nonlinearly mixed mechanical faults in wind turbine gearbox. IEEE Internet of Things Journal 5 (1), 466–467.

Zanella, A., Bui, N., Castellani, A., Vangelista, L., Zorzi, M., 2014. Internet of things for smart cities. IEEE Internet of Things Journal 1 (1), 22–32.

Zhou, W., Habetler, T. G., Harley, R. G., 2007. Bearing condition monitoring methods for electric machines: A general review. In: 2007 IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives. pp. 3–6.

Zhu, Y., Min, W., Jiang, S., 2020. Attribute-guided feature learning for few-shot image recognition. IEEE Transactions on Multimedia, 1–1.