A vibration-based method for contact pattern assessment in straight bevel gears

Marco Buzzoni[a], Gianluca D'Elia[a,*], Emiliano Mucchi[a], Giorgio Dalpiaz[a]

[a]*University of Ferrara, Department of Engineering, via Saragat 1, 44122 Ferrara, Italy*

Abstract

So far, the study of gear contacts in lightly loaded gears by means of vibration analysis has not been sufficiently addressed in the literature. Indeed, the complex nature of the physical phenomena involved makes the vibration analysis extremely challenging. This paper deals with the development and the validation of an approach for the contact pattern assessment in straight bevel gears within a pass/fail decision process. The proposed methodology is based on blending vibration-based condition indicators with classification algorithms in order to discriminate proper contact patterns from improper ones. Specifically, three different classification algorithms have been investigated: the Naive Bayes classifier, the weighted k-Nearest Neighbors classifier and a novel classifier proposed by the authors. The classifier accuracies are evaluated with a MC cross-validation that includes an extended experimental campaign consisting of more than one hundred different straight bevel gear pairs. The results show that the proposed classifier is superior to the other considered classifiers in terms of average accuracy. Finally, this manuscript proposes an original methodology that provides a reliable and quick assessment of the contact pattern in straight bevel gears considering different speeds, gear parameters and surface finish.

Keywords: Straight bevel gears, Contact pattern, Vibration analysis, Naive Bayes, k-Nearest Neighbors

1. Introduction

Straight Bevel Gears (SBGs) play a crucial role in the field of mechanical power transmissions, with particular regard to vehicle transmissions [1]. Two common examples of SBG applications are differential drives or anytime power must be transmitted between incident axes.

One of the most important parameters about SBG performances in terms of vibrations and durability is the contact pattern whereby forces and motion are transmitted. In the industrial context, it is a matter of fact that the contact pattern tests are still widely used for the final state assessment of SBGs [2]. Briefly, these tests consist in the inspection of the traces left on the tooth faces by two meshing gears mounted on a dedicated test-rig where the teeth are coated with a marking compound. The contact pattern evaluation allows to detect manufacturing errors depending on the trace characteristics, particularly shape and position. Moreover, the contact pattern tests can be used to detect assembling errors. In fact, the bevel gears can be assembled with different mounting distances but only the design mounting distance guarantees the correct meshing that implies quiet functioning, low vibration levels and endurance. The manufacturing errors can cause deviations on the correct mounting distances and, as a consequence, produce endurance issues due to

*Corresponding author

Email addresses: marco.buzzoni1@unife.it (Marco Buzzoni), gianluca.delia@unife.it (Gianluca D'Elia), emiliano.mucchi@unife.it (Emiliano Mucchi), giorgio.dalpiaz@unife.it (Giorgio Dalpiaz)

wear phenomena and uneven distribution of the forces among teeth. Despite the contact pattern test is quick and does not require particular measurement instruments, it is almost completely subjective and strongly relies on the tester experience and on the test conditions [3]. Hence, there is an urgent need to update the contact pattern inspections with objective and advanced tools.

In this scenario, vibration analysis is a powerful approach in order to detect anomalies in gears. The academic interest about gear fault identification by means of vibration analysis is evinced by the large number of research works about this topic gathered in more than five decades [4]. Nowadays, several well-established signal processing techniques are available for the detection and identification of localized and distributed gear faults. It is worth to mention (second-order) cyclostationary analysis [5, 6], phase-amplitude demodulation [7], time synchronous averaging [8] (TSA), cepstral analysis [9], blind deconvolution methods [10], auto-regressive models [11], spectral kurtosis [12] and Empirical Mode Decomposition algorithms [13]. Recently, the combination of pattern recognition techniques and statistical indicators has become a convenient methodology for the diagnosis of gears [14–17]. This strategy answers the need for characterizing complex physical phenomena (e.g. wear in gears) through vibration analysis without an explicit knowledge or explanations of how these phenomena manifest within the vibration signature.

The specialized literature offers a number of papers focusing on the development of contact pattern models of SBG [18–20]. The state of the art reported in Refs. [18–20] shows that great attention has been devoted on modeling teeth contacts in SBGs. In general, the contact pattern can be predicted through cutting simulations and analytical models Kolivand et al. [20]. Generally, these approaches do not include the possible deviations from the ideal tooth flank geometry - due to manufacturing errors or assembling errors, for instance - which usually occur in real scenarios. Moreover, these models need a number of geometrical parameters as input and are fit for setup and tune the cutting process rather than to verify contact patterns after the cutting process. In this scenario, vibration analysis is a good candidate for the verification of contact patterns in SBGs since vibration-based approaches are generally more flexible and quick than model-based approaches.

However, as the authors are aware, the investigations on contact pattern assessment in SBGs by means of vibration analysis are limited. De facto, a first – and unique – interest on this topic has been shown by Jedliński and Jonak [21, 22]. In their first exploratory work [21], they pointed out that there would be a link between the contact area and vibrations. Moreover, they realized that it is not trivial to evaluate contact areas through vibration analysis, in particular by using basic signal processing techniques. The same authors then proposed a method based on artificial neural networks for the evaluation of spiral bevel gear assembly in terms of relative contact pattern lengths with respect to the entire tooth face width. Under the hypothesis that changes in gear assembling are reflected into the vibration signature, Jedliński and Jonak proposed a method for the prediction of the relative contact pattern lengths taking into account thirteen spiral bevel gears and three different network types: multilayer perceptron, radial basis function and support vector machine. Their research aims at verifying the position of spiral bevel gears by exploiting the relative contact pattern lengths. Thanks to their promising results, their research work proposes a pioneering methodology for the gear assembly assessment.

As remarked in Refs. [3, 22], the tooth contact inspections on bevel gears are pivotal in order to detect both incorrect mounting and manufacturing errors. Despite this subject is of remarkable interest, the lack of research works in this topic is likely due to two aspects. The first one regards the complexity of characterizing the contact pattern from the vibrational standpoint, which shares many aspects with vibration-based wear analysis [23]. In fact, for each type of contact pattern corresponds a characteristic sliding contact between the gear tooth faces. The second one mainly concerns the effort of conducing an extensive experimental campaign involving different types of bevel gears with a significant number of observations and, when possible, considering natural manufacturing errors.

According to the paper of Jedliński and Jonak [22], the proposed research work focuses on the vibration analysis of contact patterns in bevel gears due to manufacturing errors from a different perspective. The main goal is developing and validating a pass/fail diagnostic tool for the contact pattern assessment of SBGs as an alternative to the standard contact pattern test, that is completely subjective. Hence, this research work is focused on discriminate the proper contacts from the improper ones through objective tools with limited user interactions.

Specifically, this paper proposes a strategy for the assessment of contact patterns in real-time by using a combination of vibration analysis and classification algorithms. Three different classification algorithms are considered: the weighted k-Nearest Neighbors (wk-NN) algorithm [14], that is one of the simplest non-parametric classification algorithm; the (weighted) Naive Bayes algorithm (NB) that is a simple parametric classification algorithm [24]; and, finally, an original combination of the two previous methods, called modified wk-NN (mwk-NN).

The considered classifiers have been investigated taking into account an extended experimental campaign that involves six sets of SBGs consisting in healthy gears and gears with natural manufacturing errors (that lead to improper tooth contacts). The experimental test conditions make the vibration analysis even more complicated since the SBGs considered in this research work are obtained by milling process – therefore with a limited accuracy – and tested under light loads. These test conditions are needed since the contact pattern is usually evaluated before the heat treatment and thus the gears are prone to failures with full load tests. This paper reports an extended experimental investigation together with a comprehensive comparison of the considered classifiers in terms of classification accuracy. This work is an original contribute about bevel gear contacts taking into account a relevant number of SBGs and considering also different kinds of contacts. In particular, a remarkable effort has been made on verifying the proposed method by means of an extensive experimental campaign that accounts hundreds of measurements. This research engages the topic of the vibration analysis of bevel gear contacts from a different standpoint with respect to Ref. [22]. Specifically, different classifiers (parametric and non parametric) has been compared a novel one has been proposed also. The proposed classifier, that is a combination of wk-NN and NB, proved to be superior and returns good accuracy level requiring a limited number of test to be trained with respect to the other classifiers. Furthermore, the focus of this paper is on designing a pass/fail procedure than predicting the contact line length [22].

This paper is organized as follows: the classifiers and the methodology are described in Section 2; Section 3 outlines the experimental campaign; the results are presented and discussed in Section 4; the final remarks are given in Section 5.

2. Method

2.1. Naive Bayes classifier

The NB classifier is a popular and simple parametric classifier rooted on the NB conditional probability model. The simplicity of this method lies on its key hypothesis: given a prior distribution, the probabilities are conditionally independent. This classification method is nothing but a combination of the NB model, that is an application of the Bayes' theorem, with a decision rule based on a maximum a posteriori criterion.

Let $\mathbf{t}$ be a test observation of J features in a J-dimensional space $\mathcal{S}$ whose label b is known. Now, suppose that the observations lying in $\mathcal{S}$ are divided into L classes $C = \{1, \cdots, l, \cdots, L\}$. From the definition of conditional probability:

$$P(C_l|\mathbf{t}) = \frac{P(C_l)p(\mathbf{t}|C_l)}{p(\mathbf{t})} \tag{1}$$

where p refers to a probability density function, P refers to a probability mass function and C_l is the l^{th} class, the NB model can be deducted by neglecting the denominator – since it is constant being not dependent on C_l – and by applying the chain rule to Equation (1):

$$P(C_l|\mathbf{t}) \; \alpha \; P(C_l) \prod_{j=1}^{J} p(\mathbf{t}_j|C_l) \tag{2}$$

where $\mathbf{t}_j$ is the j^{th} feature of $\mathbf{t}$. The right-hand side of Equation (2), by assuming that the classes are equally probable and that the data follow a Gaussian distribution, can be computed as follows:

$$P(C_l) \prod_{j=1}^{J} p(\mathbf{t}_j|C_l) = \frac{1}{L} \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi\sigma_{j,l}^2}} e^{-\frac{(\mathbf{t}_j-\mu_{j,l})^2}{2\sigma_{j,l}^2}} \tag{3}$$

where $\mu_{j,l}$ and $\sigma_{j,l}^2$ are the sample mean and the sample variance, respectively, of the j^{th} feature estimated by means of training observations of class l. Clearly, Equation (3) does not return a probability since the scaling factor $p(\mathbf{t})$ has been neglected (see Equation (1)). However, if the features of $p(\mathbf{t})$ are known, $p(\mathbf{t})$ is constant for any C_l. Thus, the right-hand side of Equation (2) can be treated as a probability for any C_l. These simplifications are allowed since the aim of the NB classifier is to predict an unknown label from an observation $p(\mathbf{t})$ rather than to estimate the specific probabilities.

The NB classifier can be thus formalized combining the NB model in Equation (2) with a maximum a posteriori decision rule that classifies the test observations on the basis of the most probable hypothesis. This decision rule can be formalized as follows:

$$\hat{b} = \underset{l=1,\cdots,L}{\operatorname{argmax}} \, P(C_l) \prod_{j=1}^{J} p(\mathbf{t}_j|C_l). \tag{4}$$

where $\hat{b}$ is the predicted label of $\mathbf{t}$.

In general, the goal of a classifier is to compute $\hat{b}$ from observations whose labels are unknown. In order to establish the classifier reliability, it is mandatory to validate the classifier by assessing the model accuracy by means of a set of M test observations with known labels. Let us assume that the test observations are arranged as a matrix $\mathbf{T}$ of dimension $M \times J$. The classification accuracy λ can be estimated through:

$$\lambda = \frac{\sum_{m=1}^{M} \delta_{\hat{b_m}, b_m}}{M} \tag{5}$$

where δ is the Kronecker delta while $\hat{b_m}$ and b_m are the estimated label (Equation (3)) and the actual label of the m^{th} observation, respectively.

On these grounds, some considerations should be made. This classifier is based on two very strong assumptions that are seldom met in real data: conditional independence and Gaussian prior distribution. Notwithstanding these drawbacks, the naive Bayes classifier is frequently used due to its extremely low computational effort and its overall good classification performance [24, p. 380-381].

2.2. The k-Nearest Neighbor classifier

The k-NN method is a non-parametric probability density function (pdf) estimator which can be extended to classification problems. The following explanation is a concise version of the one given in Ref.

[24] that provides an explanation of this method in a Bayesian context, according to the explanations given in Subsection 2.1. The Bayesian standpoint turns out to be useful for justifying also the method proposed in this work and for giving a common framework to this theoretical section.

Bearing in mind the nomenclature used previously, the probability of the observation **t** to reside into a small region $\mathcal{S}^* \in \mathcal{S}$ is:

$$P = \int_{\mathcal{S}^*} p(\mathbf{t}) d\mathbf{t}. \tag{6}$$

Then, k observations, where $k < N$, residing within $\mathcal{S}^*$ can be assumed as distributed according to a binomial distribution, since the observations do or do not reside in $\mathcal{S}^*$. If the binomial distribution is sharply peaked around the mean – that happens for large N – and if $\mathcal{S}^*$ is small enough such that $p(\mathbf{t})$ can be considered as constant, it can be demonstrated that an estimate of $p(\mathbf{t})$ is given by:

$$p(\mathbf{t}) = \frac{k}{NV} \tag{7}$$

where k is the number of points falling into $\mathcal{S}^*$ and V is the related volume. Equation (7) shows that the k-NN method gives an estimation of the pdf where k is fixed and can be seen as a smoothing parameter (with large value of k, overfitting problems may be encountered). Given k and a new observation, one may calculate the radius of the hypersphere – if the distance is Euclidean – containing exactly k observations and then estimate the pdf through Equation (7). From this standpoint, k-NN is exploited for the pdf estimation from a set of observations without prior assumptions of the data distribution. Applying this method to classification problems, a new observation can be classified by estimating the distances between the test observation and the training observations residing into the hyper-sphere of volume V that contains exactly k samples and then selecting the most likely class. Recalling the Bayes' theorem and Equation (6), it can be demonstrated [24] that the posterior probability of **t** being part of C_l is:

$$P(C_l|\mathbf{t}) = \frac{p(\mathbf{t}|C_l)P(C_l)}{p(\mathbf{t})} = \frac{k_l}{k} \tag{8}$$

where k_l refers to the number of the neighbor samples of class C_l. As done previously in Equation (4), the maximum a posteriori decision rule for k-NN can be defined as:

$$\hat{b} = \operatorname*{argmax}_{l=1,\cdots,L} P(C_l|\mathbf{t}) = \operatorname*{argmax}_{l=1,\cdots,L} \frac{k_l}{k}. \tag{9}$$

Finally, the classifier accuracy can be estimated through Equation (5).

The nonparametric methods, by definition, are not limited by the assumption of a prior distribution, as in the case of the NB classifier. However, the nonparametric methods are less computationally efficient than parametric ones since they need larger training sets that imply greater computational effort. Hence, the training set size, N, plays a crucial rules. It should be small enough to keeping the computational effort low and it should be large enough to avoid misclassification. In fact, Cover and Hart [25] demonstrated that the error rate in a two class case is *"not more than twice the Bayesian error rate"*, i.e. the irreducible error:

$$\hat{R} \leq R \leq \hat{R}\left(2 - \frac{N}{N-1}\hat{R}\right) \tag{10}$$

where R is the probability of error of the NN rule and $\hat{R}$ is the Bayes' error. Intuitively, $\mathcal{S}$ will be densely populated by the training samples when N is large. Therefore, for fixed k, $\mathcal{S}$ will be finely scanned through volumes V arbitrarily small according to the considered N.

2.3. Dimension reduction

The classifiers previously described may be used with dataset with large dimensions. With specific reference to vibration-based diagnosis through condition indicators [14, 22, 26], it is not unusual to deal with many condition indicators that imply an high-dimensional dataset. The problem of dealing with high-dimensional spaces is also known as "the curse of dimensionality" and affects the classifier performance in different ways.

In general, two general issues can be encountered in classification problems in high-dimensional space: the classifier performance that does not increase according to the number of considered dimensions and the computational effort. Specifically, the k-NN classifier is affected by the curse of dimensionality if the training samples are not clustered in well defined classes [27]. The effects of the curse of dimensionality in NB classifiers leads to a different issue: the computation of Equation (3) can lack of precision when one of the product members reach (approximately) nil values.

For these reasons, the k-NN and NB classifiers should be trained after a dimension reduction of the dataset. In this work, the dimension reduction is performed by means of the feature selection method proposed in [14]. Briefly, this feature selection method is based on selecting only the features whose Euclidean distances among different classes are above a given threshold, that in this work is set to 0.5. Further details about this dimension reduction method can be found in Refs. [14, 22]. Note that the combination of the k-NN classifier and the feature selection method gives birth to the weighted k-NN classifier (wk-NN). In this work, the dimension reduction has been applied to the NB classifier as well.

2.4. Proposed classifier

In many real scenarios, it is often burdensome or even unfeasible to collect many vibration signals in order to build an extended database. In light of the previous considerations, a limited database may affect the accuracy of the wk-NN classifier and NB classifier even performing a dimension reduction for improving the classification accuracy.

Particularly, k-NN offers good accuracy levels only when the number of samples are large enough. This is due to the relationship between the NN rule error and the number of samples (see Equation (10)). Thus, in the case of a limited sample set, it can be augmented artificially in order to reduce the theoretical error of the NN rule. This task can be accomplished by assuming that the data follows an arbitrary distribution, whose parameters can be estimated from the measured data, and by improving the sample set with extrapolated samples. In the one hand, data extrapolation breaks the nonparametric nature of the k-NN method which may results in a loss of prediction accuracy due to a poor distribution assumption. Actually, this aspect can have limited consequences since a loss of accuracy with respect to the mk-NN would suggest an improper choice of the prior distribution. In the other hand, the number of training data N, can be arbitrarily large implying a reduction of the NN probability error (see Equation (10)).

In this scenario, a novel methodology rooted on the concept of k-NN classifier can be introduced in order to overcome the issue of limited training sets. Dropping the non-parametric nature of the k-NN classifier, the following assumptions can be made: (i) the features follow a Gaussian distribution and (ii) their distributions are mutually independent, i.e. the features are not assumed as a multivariate Gaussian distribution. Thus, the k-NN classifier can be trained through a set populated with observations extrapolated by assuming a Gaussian prior distribution with parameters, i.e. mean and variance, estimated from the experimental data. Assumption (i) allows for generate an arbitrarily large set of extrapolates features just by estimating sample mean μ and sample variance σ^2 from experimental data. Assumption (ii) allows to generate these sets independently, without the estimation of the covariance matrix. The extrapolation of samples from a Gaussian distribution can be easily implemented in Matlab environment by means of the function "randn".

The decision rule of the proposed classifier is therefore based on the k-NN rule (see Equation (8)) conditioned to a set of independent Gaussian prior distributions:

$$P(C_l|\mathbf{t}, \mu^*, \sigma^*) = \frac{p(\mathbf{t}, \mu^*, \sigma^*|C_l)P(C_l)}{p(\mathbf{t}, \mu^*, \sigma^*)} = \frac{k_l}{k} \tag{11}$$

where μ^* and σ^* are the prior distribution parameters estimated through the experimental training data. The main difference between the wk-NN and the mwk-NN resides on the definition of the training set. In fact, the training set used in the wk-NN is composed only by experimental samples whereas the mwk-NN exploits an extrapolated training set that can be arbitrarily large.

The mwk-NN has the same computational complexity of the k-NN, that is $O(JNk)$ for the basic algorithm version:

- $O(JN)$ for the computation of all the Euclidean distances in the case of one nearest neighbor;
- $O(JNk)$ for the computation of all the Euclidean distances in the case of k nearest neighbors.

In fact, the proposed modification of the k-NN does not involve changes in the prediction algorithm but regards the artificial augment of the sample size. This strategy allows for reducing the theoretical minimum error rate reported in Equation (10) by increasing the number of samples N.

The comparison of the proposed method, NB and wk-NN in terms of classification accuracy is faced following the scheme reported in Figure 1. It is expected that, even with a poor distribution assumption, the performance of the proposed method should be superior or at least equal to the NB classifier performance while should be superior to wk-NN due to the larger training set.

2.5. Assessment of classifier accuracy

In general, the assessment of classifier accuracy is carried out by using the cross-validation. The aim of cross-validation on predictive models is to estimate the model accuracy by means of a set of labeled observations. In simple words, the training set is partitioned into two subsets, one of actual training and one of test. The classifier accuracy is assessed evaluating how many test observations are correctly classified (see Equation (5)).

In this context, data partitioning plays a crucial role: in fact, the training set and the test set can be selected by taking into account different combinations that can lead to different classifier accuracies particularly for small sets of observations. Thus, when the training set is not numerous, data partitioning influences the results and it is therefore necessary to assess the model accuracies by considering the average accuracy with different data combinations. Frequently, exploring all the possible combinations can be unfeasible, even taking into account small amount of data. For instance, the possible combinations for a dataset of $N = 24$ observations equally divided into two classes ($L = 2$) and partitioned in sets of the same size is

$$c = \binom{N/2}{N/4}^L = 8.5 \cdot 10^5. \tag{12}$$

In this work, the Monte-Carlo (MC) approach fits the need to estimate the classifier accuracy without accounting all the available combinations and therefore reducing the computational effort. Indeed, the MC cross-validation estimates the model accuracy by using training sets and test sets randomly arranged (using an uniform distribution) without replacement. This process is repeated with c^* different (random) combinations, where $c^* < c$ and the classifier accuracy is estimated by averaging all the trial accuracies.

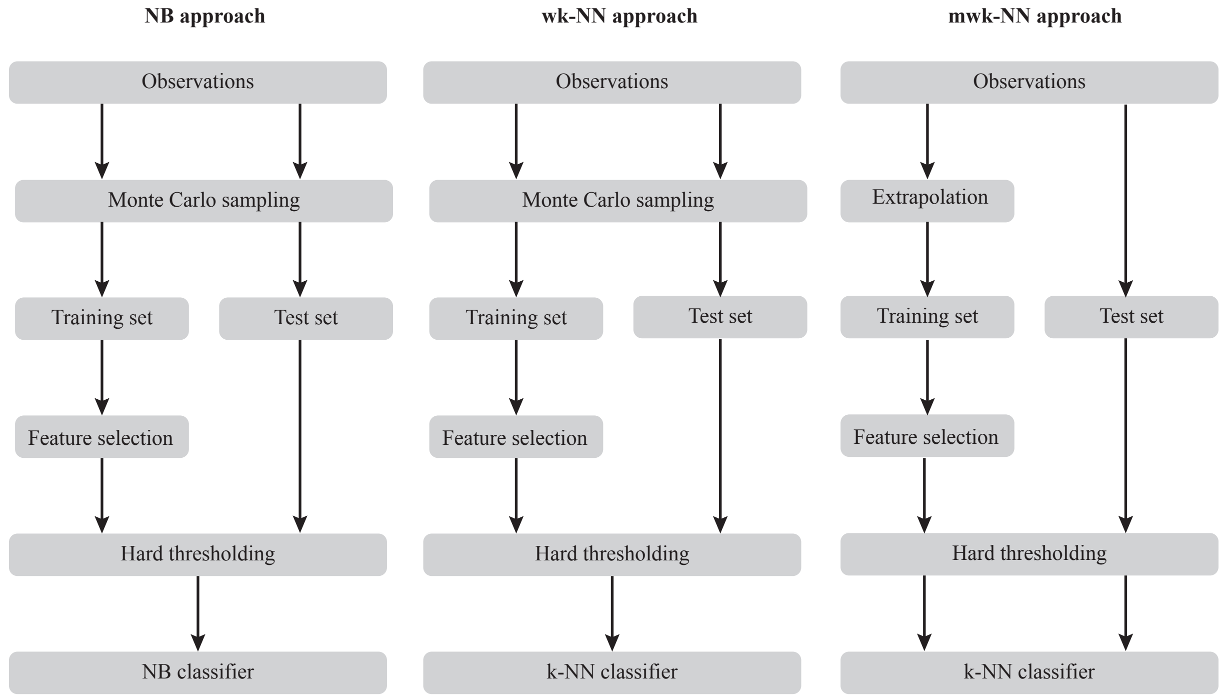

Figure 1: Flow charts of the assessment of the classifier accuracy.

2.6. Feature parameters

The specialized literature counts a considerable number of scalar indicators that can be used to detect changes in the gear vibration signature due to gear faults as tooth cracks and pitting. De facto, how wear phenomena related to different contact typologies appear in the vibration signals is still a matter of discussion [22, 23]. Therefore, the characterization of different contact patterns in SBGs through vibration analysis can be carried out by using a large set of scalar indicators finding out hidden correlations among the indicators by using pattern recognition.

In this work, thirty three features have been taken into account. According to the scheme reported in Figure 2, these features can be divided into six families: time-domain features, features based on the TSA, features based on the residual signal, features based on the regular signal, features based on the difference signal and features based on the pitch-averaged signal. For the sake of brevity, the features described in Figure 2 can be estimated through the formulae reported in a very well written table in Ref. [22].

In the following, the relationships for the estimation of the TSA, the residual signal, the regular signal and the difference signal are given. The discrete angular resampled signal, x, is assumed to have L samples corresponding to N shaft revolutions (where is N integer) and a fixed angular resolution $\Delta\theta$ equal to $2\pi N/L$. The residual signal $r_x[\theta]$, the regular signal $g_x[\theta]$ and the difference signal $d_x[\theta]$ are derived from the TSA, $m_x[\theta]$, referenced to the shaft revolution [28]. In order to avoid a burdensome nomenclature, the discrete angular variable has been expressed as θ instead of $l\Delta\theta$, where $0 \leq l < \frac{L}{N}$ is the sample index. The TSA signal is then defined as:

$$m_x[\theta] = \frac{1}{N}\sum_{n=0}^{N-1} x[\theta + n\frac{L}{N}\Delta\theta]. \tag{13}$$

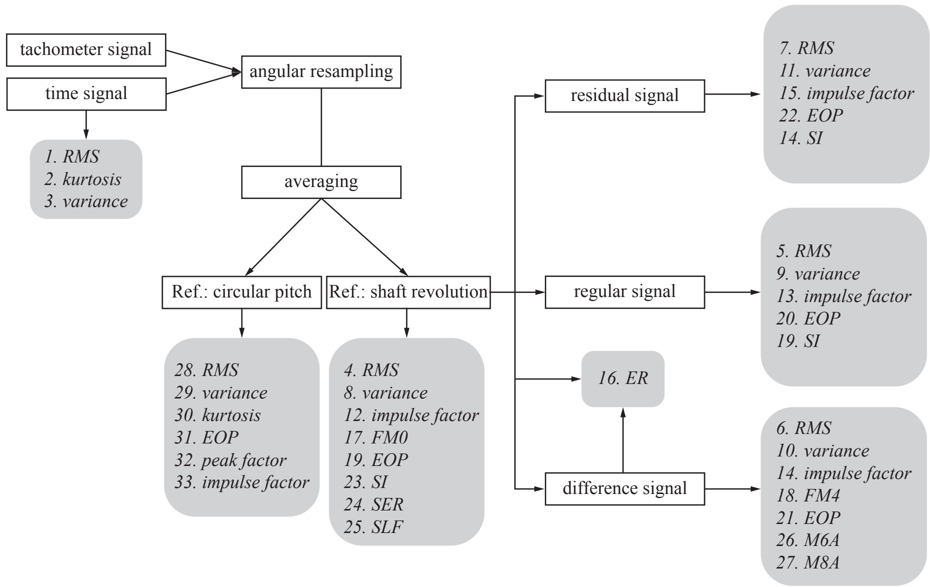

Figure 2: Schematic of the features.

Analogously, the signals derived from the TSA are reported hereafter:

- the regular signal is obtained by keeping only the gearmesh harmonics from the $m_x[\theta]$

$$g_x[\theta] = \sum_{k=1}^{N_{gm}} c_k e^{jzk\theta} \quad with \quad c_k = \frac{1}{\Theta} \sum_{\theta=0}^{\Theta} m_x[\theta] e^{-jk\theta} \tag{14}$$

  where $\Theta = \frac{\Delta\theta L}{N}$, j is the imaginary unit, c is the Fourier coefficient and N_{gm} is the integer number of the gearmesh harmonics;

- the residual signal is defined as $m_x[\theta]$ filtered from the gearmesh components and the first two shaft rotational harmonics

$$r_x[\theta] = m_x[\theta] - g_x[\theta] - \sum_{p=1}^{N_{rh}} c_p e^{jp\theta} \quad with \quad c_p = \frac{1}{\Theta} \sum_{\theta=0}^{\Theta} m_x[\theta] e^{-jp\theta} \tag{15}$$

  where $N_{rh} = 2$ is the even integer number of the considered rotational harmonics;

- the difference signal is constituted of the residual signal purified from the first-order gearmesh sidebands

$$d_x[\theta] = m_x - \sum_{p=1}^{N_{rh}} c_p e^{jp\theta} - \sum_{i=N_{sb}/2}^{N_{sb}/2} \sum_{k=1}^{N_{gm}} c_{i,k} e^{j\theta(kz+i)} \quad with \quad c_{i,k} = \frac{1}{\Theta} \sum_{\theta=0}^{\Theta} r_x[\theta] e^{-j\theta(kz+i)} \tag{16}$$

where $N_{sb} = 2$ is the even integer number of the gearmesh sidebands. Besides, the TSA referenced to the mesh period has been considered as well. The resulting averaged signal is the first-order cyclostationary

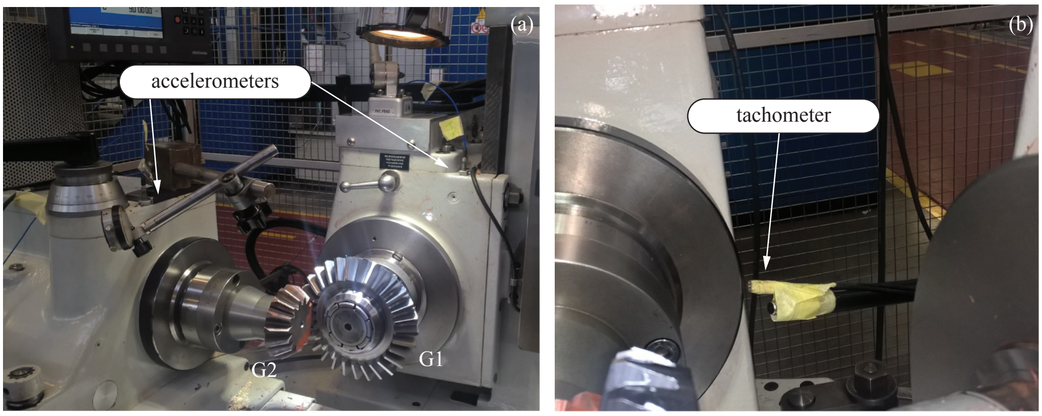

Figure 3: Experimental setup: (a) accelerometer layout and (b) tachometer.

[29] part of the signal related to vibration phenomena synchronized with the mesh period rather than the shaft revolution period. The resulting signal is constituted of all the contributions that are periodically repeated at every circular pitch. Therefore, this signal represents the vibration signature referenced to all the periodic contributions inherent to the circular pitch. From Equation (13), the time synchronous averaged signal referenced to the mesh period can be defined as:

$$m_x{}^p[\hat{\theta}] = \frac{1}{zN} \sum_{\hat{n}=0}^{zN-1} x[\hat{\theta} + \hat{n} \frac{L}{zN} \Delta\theta] \tag{17}$$

where $0 \leq \hat{\theta} < \frac{L}{zN}$ is the discrete angular variable within the mesh period.

3. Experimental campaign

The experimental campaign has been conducted on six sets of SBGs by means of a Gleason tester for bevel gears. According to Figure 3, the test bench acts as a multiplier since the driven gear is always the pinion. From now, G1 refers to the driving gear with z_1 teeth and G2 refers to the driven gear with z_2 teeth. All the gears have been tested with a shaft angle of 90 degrees under light load 8 Nm and without lubrication. These test conditions are needed since the contact pattern is usually evaluated before the heat treatment and therefore the gears are prone to failures with full load tests. Moreover, lightly loaded gears are very challenging to investigate from the vibrational standpoint. This aspect will be discussed later.

The vibrations signals have been collected by two piezoelectric accelerometers type PCB 352C18 placed on both the tester sides in radial direction. Concurrently, the instantaneous speed of the fastest shaft (driven gear) has been measured with an optical tachometer type Sick WLL170T-P135. The sampling frequency has been set to 25.6 kHz and the measurement duration has been fixed to 10 s. The acquisition system is constituted of a National Instruments cDAQ-9191 CompactDAQ Chassis equipped with a NI-9234 module and driven by a dedicated LabVIEW software. An example of the measured signals is given in Figure 4.

Each gear pair has been mounted by a specialized operator who verified the correct gear positioning in terms of: mounting distances, shaft alignment and backlash. Moreover, the test gear mates with a master gear that never changes within the same dataset. The master gear is a gear having superior manufacturing quality. This test procedure ensures that the gear pairs are mounted following the design specifications and

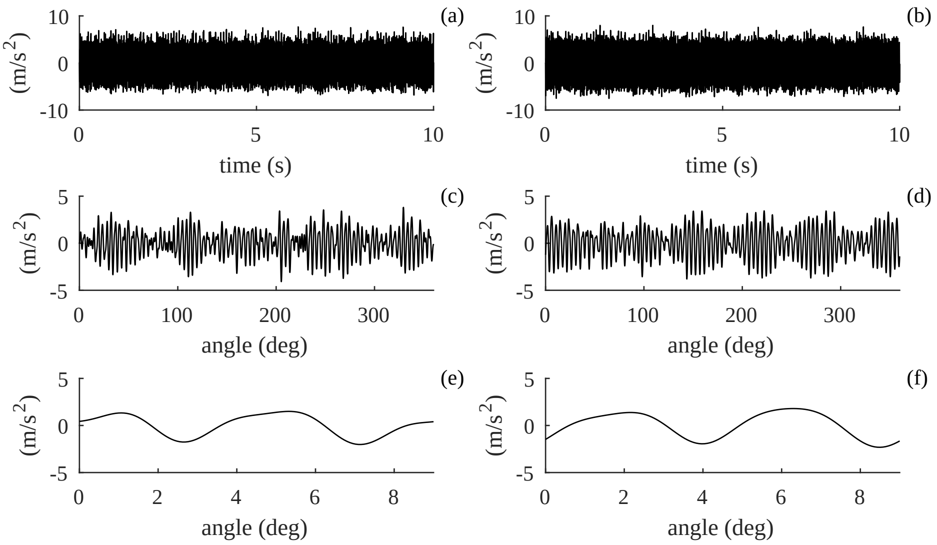

Figure 4: Examples of measured signals: (a) raw vibration signal in clockwise direction, (b) raw vibration signal in counterclockwise direction, (c) TSA referenced to the driven shaft period in clockwise direction, (d) TSA referenced to the driven shaft period in counterclockwise direction, (e) TSA referenced to the gearmesh period in clockwise direction, (f) TSA referenced to the gearmesh period in counterclockwise direction.

Table 1: Details of the test conditions (the speed is referred to the driving shaft speed).

no. dataset	heat treatment	speed (rpm)	z1	z2	no. of tests	test gear
1	no	542	21	14	20	G2
2	no	811	21	14	22	G1
3	no	520	40	13	30	G1
4	yes	512	40	13	30	G1
5	no	180	40	13	30	G2
6	yes	512	40	13	30	G2

that any change in the vibration signal is due to just the tested gear. It should be mentioned also that both the directions of rotation have been tested, i.e. clockwise and counterclockwise.

Different types of incorrect contact pattern, obtained by natural manufacturing errors, have been investigated considering also possible additional effects of surface distortions due to the heat treatment. Table 1 summarizes some details about the test conditions while Figure 5 shows some examples of different contact patterns investigated in this work.

Eventually, it should be stressed that in this research work a remarkable number of tests have been carried out and investigated: 324 runs, considering both the direction of rotation and involving 162 different gear pairs.

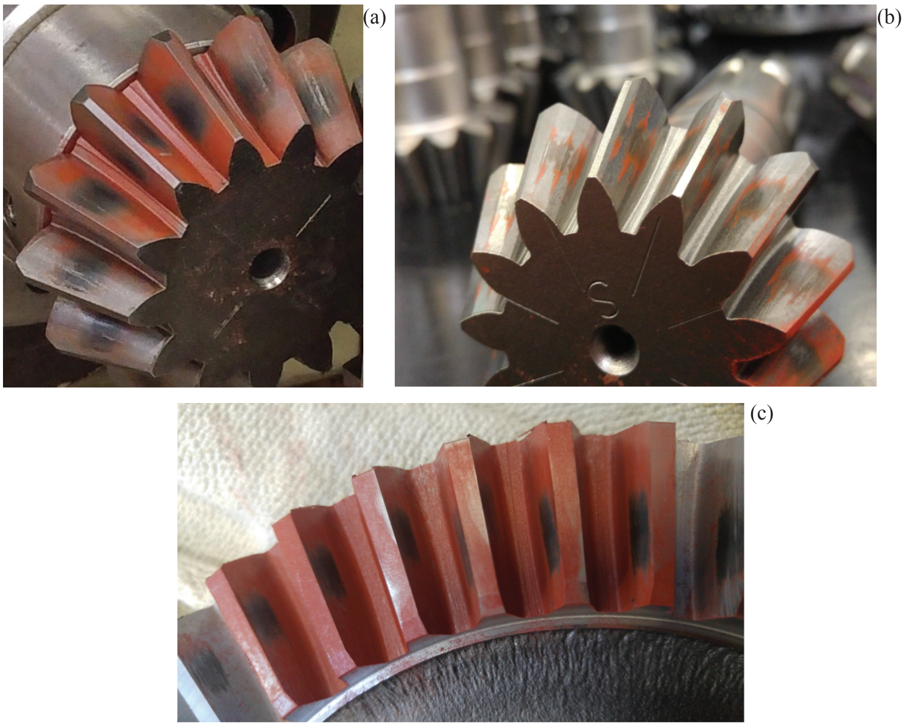

Figure 5: Examples of traces due to different contact types: (a) desired contact under light load, (b) bridged contact and (c) crossed contact.

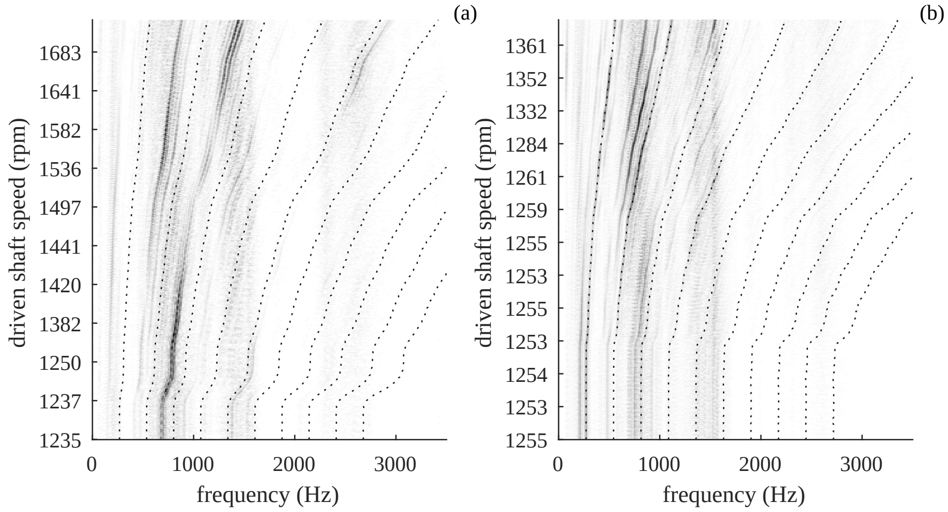

Figure 6: Short-time Fourier Transform of the vibration signal in (a) the clockwise run-up test and (b) the counterclockwise run-up test. The dotted curves refer to the theoretical gearmesh harmonics.

4. Analysis of results

This section focuses on two pivotal aspects of this research: establishing if the considered methods allow for discriminating proper contacts from improper contacts and establishing which classifier is the best one.

It is worthwhile to spend a few words about the challenges of vibration analysis of lightly loaded gears. The main problem in lightly loaded gears is that, generally, they exhibit a high transmission error, that reaches (theoretically) its minimum value at the design transmission error [30, p. 22]. The combination of light load and high transmission error leads to non-linear phenomena related to the contact loss among teeth. As reported in Ref. [30, p. 185-187], many unusual contributions appear into the vibration signature that, together with the measurement noise and other possible interferences, mask the low amplitude meshing contributions. This behavior is clearly shown in Figure 6: the time-frequency representations of both the rotation directions highlight that the meshing harmonics (marked with dotted lines) are not visible at all, as a result of the previous considerations. Moreover, some unexpected periodic components can be found between the first and the second gearmesh harmonics as well as the second and the third gearmesh harmonics.

In this investigation, the following classifier parameters have been selected: 80000 MC iterations for the wk-NN classifier, 80000 MC iterations for the NB classifier and 80000 extrapolated observations for the mwk-NN classifier. The observations have been equally split into the training set and the test set, with the exception of the mwk-NN classifier that uses all the experimental observations for the test and only extrapolated observations for the training. As reported in Subsection 2.4, the proposed method uses extrapolated data for the training and the experimental data for the test. This clearly means that the MC cross-validation is not needed in this case. Furthermore, all the considered classifiers have been performed before the feature weighting illustrated in Subsection 2.3. Since the weights estimated in wk-NN and NB have not been reported in the following since they change depending on the selected combination of training set and test set. Note that the algorithms used in the following analysis have been coded in Matlab environment by the authors.

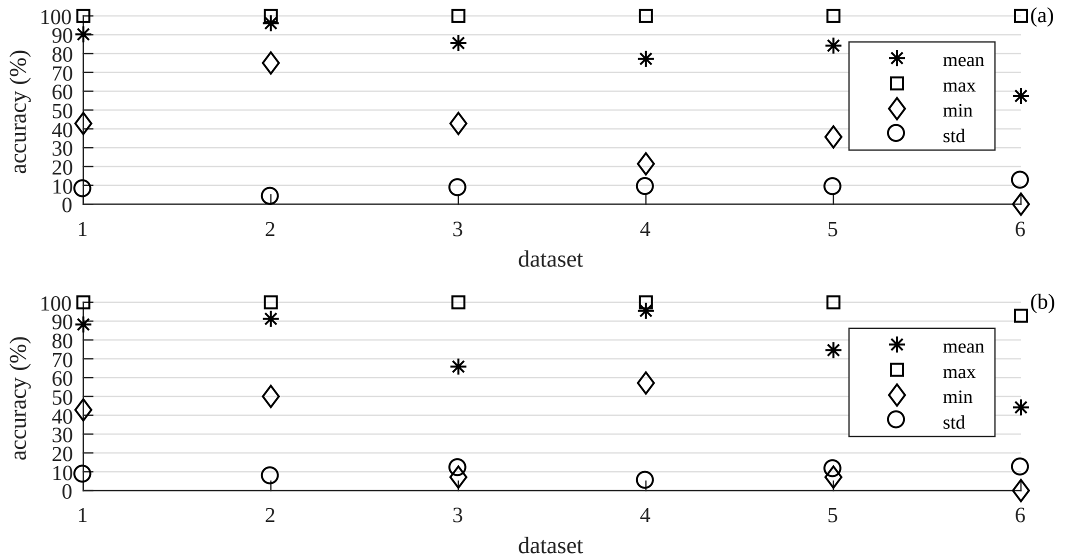

Figure 7: Statistical parameters related to the classification accuracy of the NB classifier for (a) the clockwise direction and (b) for the counterclockwise direction.

Figure 7 reports the accuracies obtained by means of NB classifier in terms of mean, maximum, minimum and standard deviation. The NB classifier reaches an average accuracy above 70% in both the rotation directions for Dataset 1, Dataset 2, Dataset 4 and Dataset 5. According to Figure 6, the average accuracies change depending on the rotation direction and represent a further difficulty on classifying correctly the data. Moreover, it should be remarked that, according to these results, the minimum values can be far lower than a flip coin (50%) and sometimes below 10%. This is an evidence of the importance to take into account an average accuracy estimated from different random combinations of training samples and test samples. Indeed, in particular when the number of observations is low, the choice of the training set and the test set strongly influence the classifier accuracy. Nevertheless, the mean values are far closer to the maximum value than the minimum value suggesting that the possible accuracies are asymmetrically distributed in favor of the greater accuracies.

Similar results have been achieved by using the wk-NN. In this case, Figure 8 reports the results obtained by considering $k = 3$, $k = 5$ and $k = 7$. The even values of k are neglected in order to avoid ties, i.e. uncertainty about the class assignment. It is worth noting that different k lead to similar accuracies. Thus, in this case, k is not a parameter that strongly influence the accuracy in terms of mean and the other statistical parameters.

The accuracies estimated for different k by using the mwk-NN method are reported in Figure 9 and Figure 10 for the clockwise rotation and the counterclockwise rotation, respectively. In this case, the accuracies estimated for each dataset considering different values of k have been reported together with their respective weights. In particular, the feature weights reported in the diagrams on the left side are reported in terms of normalized amplitudes according to Ref. [14]. The dotted lines refers to the threshold which, in this work, has been set to 0.5. All the features below the threshold are not considered in the classification process. Concerning the feature weights, it can be noted that they change depending on the dataset and, as enlighten in Figure 6, on the rotation direction. This behavior explains different facets of analyzing the contact of lightly loaded gears. The first one is that single condition indicators are not

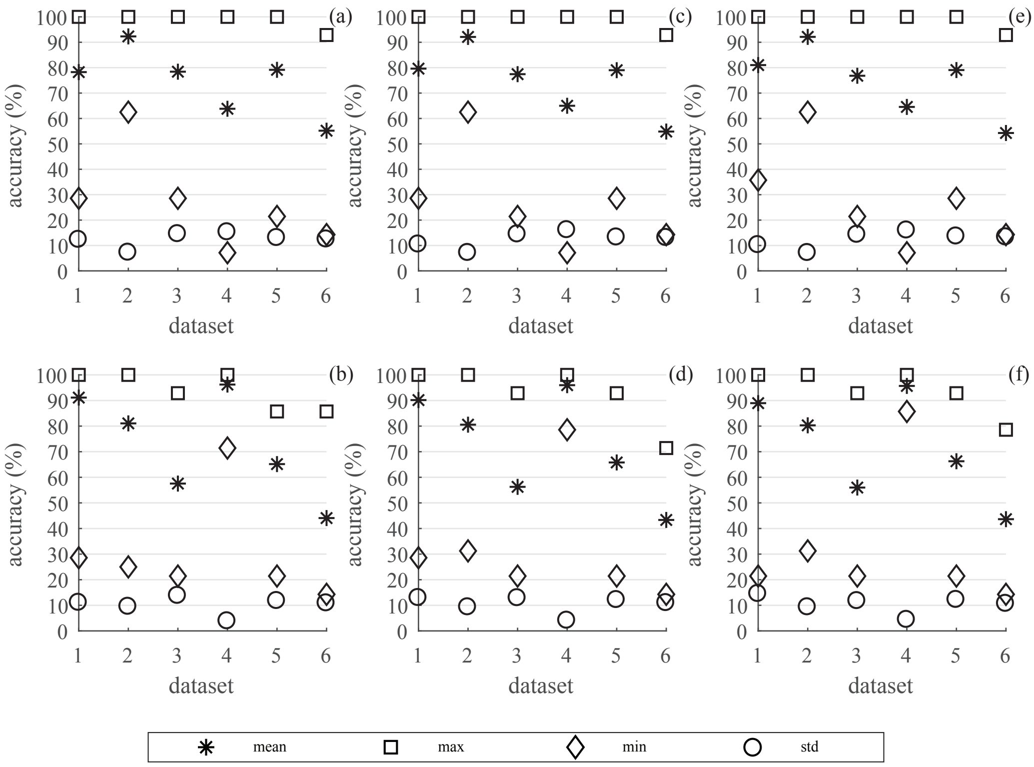

Figure 8: Statistical parameters related to the classification accuracy of the wk-NN classifier for the clockwise direction with (a) $k = 3$, (c) $k = 5$ and (e) $k = 7$ and for the counterclockwise direction with (b) $k = 3$, (d) $k = 5$ and (f) $k = 7$.

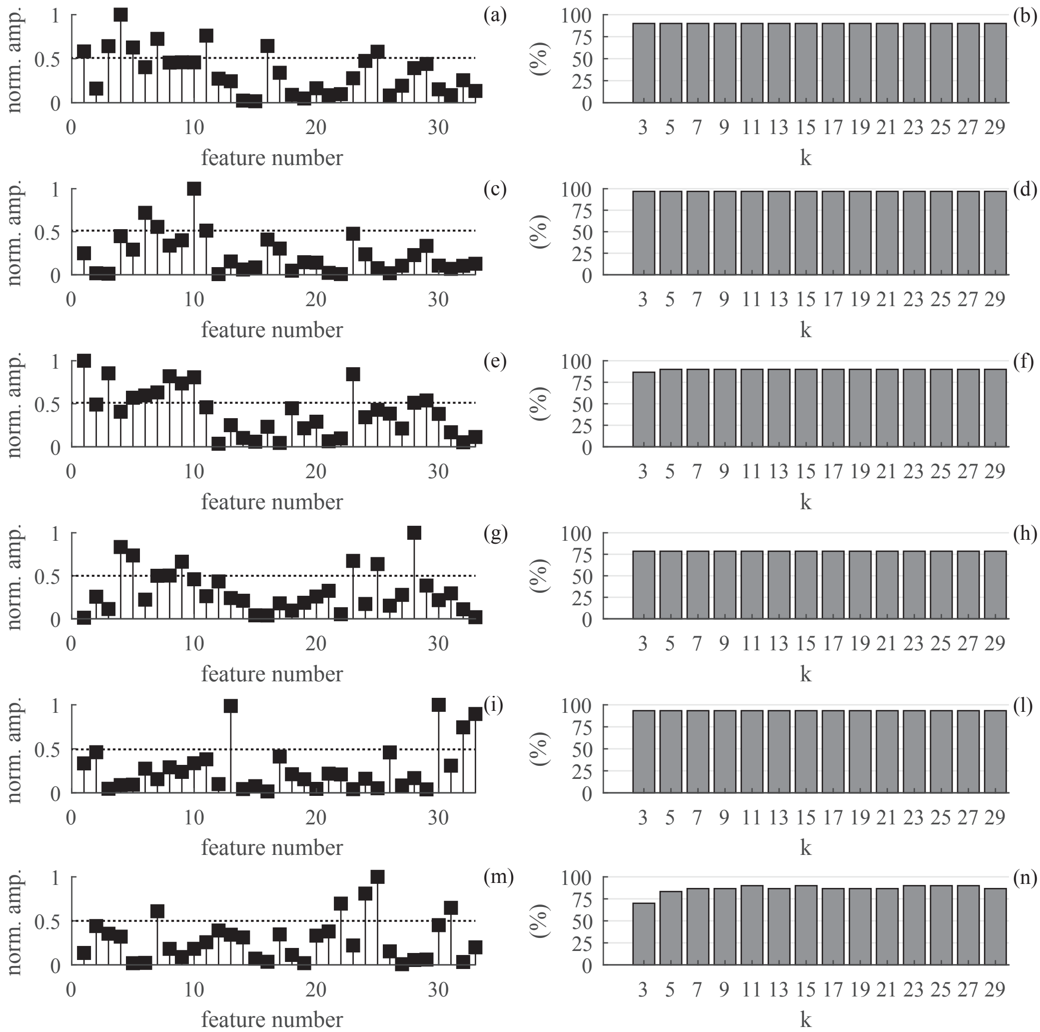

Figure 9: Feature weights (left column) with their classification accuracies (right column) of the mwk-NN classifier for the clockwise direction: (a-b) dataset 1, (c-d) dataset 2, (e-f) dataset 3, (g-h) dataset 4, (i-l) dataset 5 and (m-n) dataset 6.

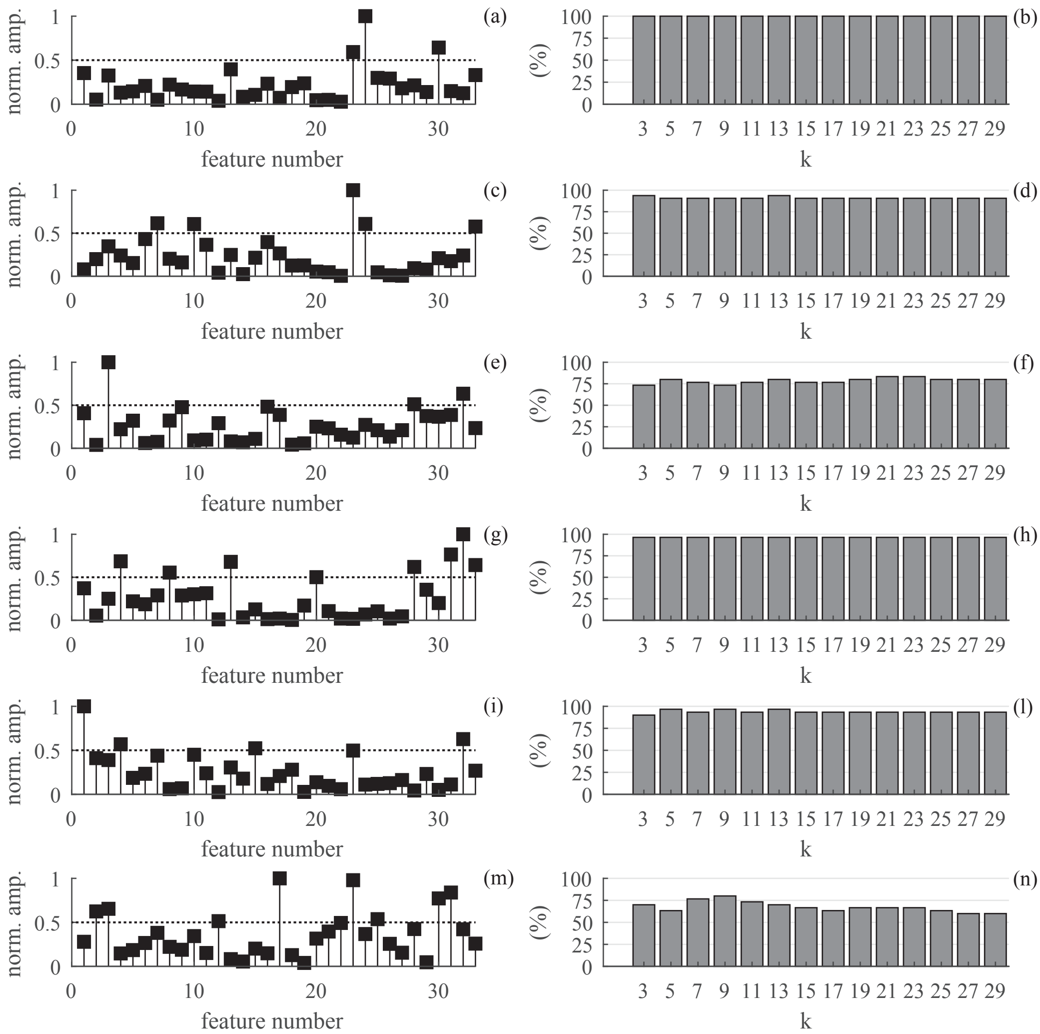

Figure 10: Feature weights (left column) with their classification accuracies (right column) of the mwk-NN classifier for the counterclockwise direction: (a-b) dataset 1, (c-d) dataset 2, (e-f) dataset 3, (g-h) dataset 4, (i-l) dataset 5 and (m-n) dataset 6.

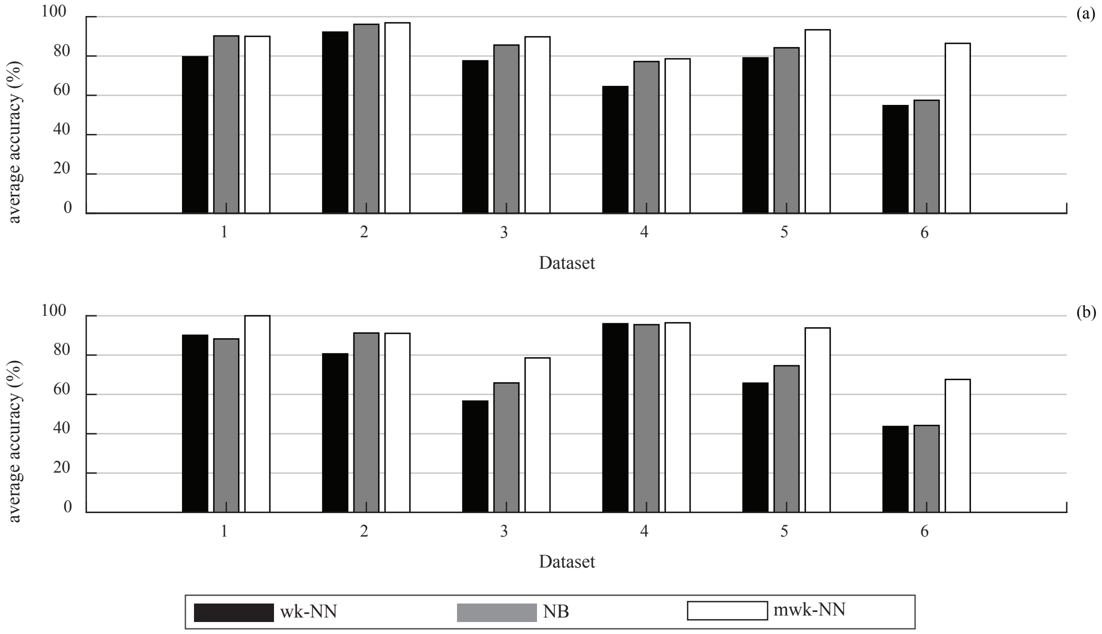

Figure 11: Comparison of the average accuracies for (a) the clockwise direction and (b) anticlockwise direction.

sufficient for the contact pattern evaluation. Then, the different trends of the weights reflect the fact that the vibration signatures appears very different even in the same dataset considering both the rotation directions. These differences are likely due to the cutting process. The SBGs under investigation are cut through milling process, where the grinding wheel interacts with one tooth space one by one. This process is cheap and fit for small series production but guarantees lower gear quality — thus heterogeneous surface finish — with respect to other kind of cut processes. Moreover, in our case the gears are cut by milling two flanks at a time by using two different grinding wheels. Using two different grinding wheels could lead to different quality levels of the process since they can have different tool wear. This is a likely reason of the differences, sometimes marked, for the clockwise and counter-clockwise configurations.

Finally, the variability of the feature weights depending on the dataset and on the rotation direction suggests that the data cannot be classified as a whole, at least with the considered classifiers. On these grounds, the marked variability of the weights is due to the fact that the indicators used are not fit for capture such complex micro-phenomena involved in the contact between teeth, especially in the case of light load. Moreover, the marked differences between the clockwise direction and the anticlockwise direction confirm that both the directions should be considered and classified separately.

Considering the classification accuracies, they have been reported in the diagrams on the right side of Figure 9 and Figure 10. In this case the maximum k has been limited to 29, neglecting the even k, as done before. The classification accuracy of all the datasets, with the only exception of dataset 6, are almost constant with respect to k. This behavior is desired for two reasons: reduce the importance to select the right k (i.e. the model complexity), at least for these data and for $k < 30$; allows for evaluating an average accuracy with respect to k.

Finally, the average accuracies estimated previously are compared in Figure 11. It should be noted that, since the influence of k is limited on the classification accuracy, the accuracies estimated through wk-NN and mwk-NN have been averaged also with respect to k. Concerning the detection and classification of improper

Table 2: CPU times and classification accuracies estimated using 80000 extrapolated samples.

	anticlockwise		clockwise	
	CPU time (s)	accuracy (%)	CPU time (s)	accuracy (%)
dataset 1	1.73	93.33	1.73	90.00
dataset 2	1.90	96.88	2.08	100.00
dataset 3	2.01	83.33	1.94	86.67
dataset 4	1.78	96.43	1.89	82.14
dataset 5	1.97	80.00	2.05	93.33
dataset 6	2.02	70.00	1.87	83.33
average	1.90	86.66	1.93	89.25

contacts, the considered classifiers are efficient, in different measures, despite their simplicity. Among all the datasets, dataset 6 seems the hardest to analyze, in particular in the counterclockwise direction where the average accuracy is even below a coin flip for wk-NN and NB classifiers.

The wk-NN classifier returns, in average, the worst accuracies while the NB classifier is placed between wk-NN and mwk-NN. The proposed method demonstrates a superior accuracy than the wk-NN classifier and the NB classifier, especially for datasets 4, 5 and 6. Finally, the global accuracies taking into account all the datasets and the rotation directions are the following: 79.2 % for the NB classifier, 73.4 % the wk-NN classifier and 87.9 % for the mwk-NN. Therefore, the mwk-NN classifier, an hybrid approach between wk-NN and NB, turning out to be best one in terms of average accuracy.

The computation time of the proposed methodology plays a pivotal role in real scenarios where the gear contact assessment has to be carried out quickly. Table 2 shows the computation times referenced to the results reported in Figure 10 and Figure 9. The CPU times have been estimated with a desktop computer Dell XPS 8700 equipped with a processor Intel CoreTM i7-4790 4th gen. 3.6 GHz. The average computation time, reported in the last row of Table 2 is about 1.9 s considering a sample size of 80000 samples. Note that the reported CPU times have involved the computation of a 14 different values of k, from 3 to 30 neglecting even values of k, that are needed in order to establish which is the optimal value of k for the classification. This means that the actual CPU times needed for the gear classification are below the ones reported in Table 2 since the classification of contacts must be performed by using a single value of k. Thus, this CPU time analysis shows that the proposed methodology can be used for the gear contact quality check in real time.

5. Final remarks

The present research work investigates the SBG contact pattern by developing a pass/fail procedure for the assessment of the proper tooth contact. In the wake of the pioneering work of Jedliński and Jonak [22], machine learning algorithms – both parametric and non-parametric – have been combined with vibration analysis tools. Particular care have been devoted to consider classifiers easy to implement and with reduced computational effort: the NB classifier and the wk-NN classifier. Furthermore, this paper proposes a novel hybrid classifier that is a combination of the NB and the wk-NN methods.

The hybrid methodology involving vibration analysis and machine learning has been tested with an extensive experimental campaign that consists of 324 experiments involving 162 different gear pairs subdivided into six datasets. This approach, with particular reference to the proposed classifier, proved to be effective for discriminating the SBGs exhibiting improper contacts patterns with a satisfying degree of accuracy. In fact, taking into account all the dataset and all the rotation directions the proposed methodology reaches a global average accuracy of 88.5 %. The results carried out in this research work are promising also for practical applications. Indeed, the validation involved a remarkable number of tests considering SBGs with different number of teeth, surface finishes and manufacturing errors. Moreover, the results achieved are particularly significant and novel since the vibration analysis of lightly loaded gears is itself challenging and, concurrently, how contact pattern can be interpreted through vibration signals is still object of discussion.

Conflicts of interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] S. P. Radzevich, Dudley's Handbook of Practical Gear Design and Manufacture, second edition, 2012.

[2] ANSI/AGMA 2005-D03, Design Manual for Bevel Gears, Technical Report, American Gear Manufacturers Association, Alexandria, VA, 2005.

[3] J. Klingelnberg, Bevel Gear, Springer Berlin Heidelberg, Berlin, Heidelberg, 2016.

[4] R. B. Randall, Vibration-based Condition Monitoring, John Wiley & Sons, Ltd, Chichester, UK, 2011.

[5] C. Capdessus, M. Sidahmed, J. Lacoume, Cyclostationary Processes: Application in Gear Faults Early Diagnosis, Mechanical Systems and Signal Processing 14 (2000) 371–385.

[6] A. Raad, J. Antoni, M. Sidahmed, Indicators of cyclostationarity: Theory and application to gear fault monitoring, Mechanical Systems and Signal Processing 22 (2008) 574–587.

[7] W. Wang, Early Detection of Gear Tooth Cracking Using the Resonance Demodulation Technique, Mechanical Systems and Signal Processing 15 (2001) 887–903.

[8] S. Braun, The synchronous (time domain) average revisited, Mechanical Systems and Signal Processing 25 (2011) 1087–1102.

[9] R. B. Randall, A History of Cepstrum Analysis and Its Application to Mechanical Problems, Mechanical Systems and Signal Processing 97 (2017) 1–16.

[10] M. Buzzoni, J. Antoni, G. D. Elia, Blind deconvolution based on cyclostationarity maximization and its application to fault identification, Journal of Sound and Vibration 432 (2018) 569–601.

[11] H. Endo, R. Randall, Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter, Mechanical Systems and Signal Processing 21 (2007) 906–919.

[12] J. Antoni, Fast computation of the kurtogram for the detection of transient faults, Mechanical Systems and Signal Processing 21 (2007) 108–124.

[13] M. Buzzoni, E. Mucchi, G. D'Elia, G. D. Elia, G. Dalpiaz, Diagnosis of Localized Faults in Multistage Gearboxes: A Vibrational Approach by Means of Automatic EMD-Based Algorithm, Shock and Vibration 2017 (2017) 1–22.

[14] Y. Lei, M. J. Zuo, Gear crack level identification based on weighted K nearest neighbor classification algorithm, Mechanical Systems and Signal Processing 23 (2009) 1535–1547.

[15] B. Samanta, Gear fault detection using artificial neural networks and support vector machines with genetic algorithms, Mechanical Systems and Signal Processing 18 (2004) 625–644.

[16] A. Rojas, A. K. Nandi, Practical scheme for fast detection and classification of rolling-element bearing faults using support vector machines, Mechanical Systems and Signal Processing 20 (2006) 1523–1536.

[17] J. Rafiee, F. Arvani, A. Harifi, M. Sadeghi, Intelligent condition monitoring of a gearbox using artificial neural network, Mechanical Systems and Signal Processing 21 (2007) 1746–1754.

[18] Y.-P. Shih, Mathematical Model for Face-Hobbed Straight Bevel Gears, Journal of Mechanical Design 134 (2012) 091006.

[19] I. Tsuji, K. Kawasaki, H. Gunbara, H. Houjoh, S. Matsumura, Tooth Contact Analysis and Manufacture on Multitasking Machine of Large-Sized Straight Bevel Gears With Equi-Depth Teeth, Journal of Mechanical Design 135 (2013) 034504.

[20] M. Kolivand, H. Ligata, G. Steyer, D. K. Benedict, J. Chen, Actual Tooth Contact Analysis of Straight Bevel Gears, Journal of Mechanical Design 137 (2015) 93302.

[21] Ł. Jedliński, J. Jonak, Quality evaluation of the bevel gear assembly based on analysis of the vibration signal, Diagnostyka 1 (2010) 23–26.

[22] Ł. Jedliński, J. Jonak, A disassembly-free method for evaluation of spiral bevel gear assembly, Mechanical Systems and Signal Processing 88 (2017) 399–412.
[23] C. Hu, W. A. Smith, R. B. Randall, Z. Peng, Development of a gear vibration indicator and its application in gear wear monitoring, Mechanical Systems and Signal Processing 76-77 (2016) 319–336.
[24] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
[25] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Transactions on Information Theory 13 (1967) 21–27.
[26] Y. Lei, J. Lin, Z. He, M. J. Zuo, A review on empirical mode decomposition in fault diagnosis of rotating machinery, Mechanical Systems and Signal Processing 35 (2013) 108–126.
[27] K. Beyer, J. Goldstein, R. Ramakrishnan, U. Shaft, When Is Nearest Neighbor Meaningful?, in: Database Theory ICDT'99, Lecture Notes in Computer Science, Springer Berlin Heidelberg, 1999, pp. 217–235.
[28] B. Assaad, M. Eltabach, J. Antoni, Vibration based condition monitoring of a multistage epicyclic gearbox in lifting cranes, Mechanical Systems and Signal Processing 42 (2014) 351–367.
[29] J. Antoni, Cyclostationarity by examples, Mechanical Systems and Signal Processing 23 (2009) 987–1036.
[30] J. D. Smith, Gear Noise and Vibration, CRC Press, Chichester, UK, 2011.