Applied Computer Science, vol. 19, no. 4, pp. 136–150
doi: 10.35784/acs-2023-40

Submitted: 2023-11-01 | Revised: 2023-12-01 | Accepted: 2023-12-13

Keywords: Vibroacoustic signal, VAG, Osteoarthritis, knee joint, kinetic chain, Artificial Neural Network, Multilevel Perceptron, MLP, Radial Basis Function, RBF, SVM, naive Bayes classifier

Robert KARPIŃSKI [0000-0003-4063-8503]*,
Przemysław KRAKOWSKI [0000-0001-7137-7145]**, ***,
Józef JONAK [0000-0003-4658-4569]*, *Anna MACHROWSKA* [0000-0003-3289-2421]*,
Marcin MACIEJEWSKI [0000-0001-9116-5481]****

COMPARISON OF SELECTED CLASSIFICATION METHODS BASED ON MACHINE LEARNING AS A DIAGNOSTIC TOOL FOR KNEE JOINT CARTILAGE DAMAGE BASED ON GENERATED VIBROACOUSTIC PROCESSES

Abstract

Osteoarthritis is one of the most common cause of disability among elderly. It can affect every joint in human body, however, it is most prevalent in hip, knee, and hand joints. Early diagnosis of cartilage lesions is essential for fast and accurate treatment, which can prolong joint function. Available diagnostic methods include conventional X-ray, ultrasound and magnetic resonance imaging. However, those diagnostic modalities are not suitable for screening purposes. Vibroarthrography is proposed in literature as a screening method for cartilage lesions. However, exact method of signal acquisition as well as classification method is still not well established in literature. In this study, 84 patients were assessed, of whom 40 were in the control group and 44 in the study group. Cartilage status in the study group was evaluated during surgical treatment. Multilayer perceptron - MLP, radial basis function - RBF, support vector method - SVM and naive classifier – NBC were introduced in this study as classification protocols. Highest accuracy (0.893) was found when MLP was introduced, also RBF classification showed high sensitivity (0.822) and specificity (0.821). On the other hand, NBC showed lowest diagnostic accuracy reaching 0.702. In conclusion vibroarthrography presents a promising diagnostic modality for cartilage evaluation in clinical setting with the use of MLP and RBF classification methods.

* Lublin University of Technology, Faculty of Mechanical Engineering, Department of Machine Design and Mechatronics, Poland, r.karpinski@pollub.pl
** Medical University of Lublin, Chair and Department of Traumatology and Emergency Medicine, Poland, przemyslawkrakowski@umlub.pl
*** Orthopaedic and Sports Traumatology Department, Carolina Medical Center, Poland
**** Lublin University of Technology, Faculty of Electrical Engineering and Computer Science, Institute of Electronics and Information Technology, Poland, m.maciejewski@pollub.pl

1. INTRODUCTION

Degenerative joint disease, known as osteoarthritis, is a chronic joint disorder characterized by the gradual deterioration of joint cartilage and damage to other joint structures. Symptoms include pain, stiffness, limited joint mobility, and cracking. Causes include factors such as age, genetics, obesity, injuries, and excessive joint stress (Williams & Pierre-Louis, 2024). Joint replacements are one of the most commonly performed surgeries around the world with increasing numbers of patients receiving total (TKA) or unicompartmental (UKA) knee arthroplasty. It is well known that knee replacement surgery reduces pain, improves function and increases quality of life (W-Dahl et al., 2022), but some studies show that up to a quarter of patients suffer from surgery-related pain after UKA or TKA (Ashoorion et al., 2023). Moreover, there is a risk of endoprosthesis loosening over time and 15 year survivorship is estimated at 90% (Barnett & Toms, 2012). This means that with increasing number of joint replacements which is even estimated to rise 400% by the 2040 year (Singh et al., 2019) more and more patients will be subjected to revision TKA in which results are not as favorable as in primary joint replacement. Therefore early stage osteoarthritis (OA) treatment is gaining more and more interest in research and clinical field in order to prolong joint preservation time. To be able to treat OA at early stages fast and accurate diagnosis have to be established. Currently, the gold standard for the diagnosis of OA is conventional X-ray with a Kellgren-Lawrence score of 0-4 for disease progression. Conventional X-rays are also used for planning of the procedures and following up the results of the treatment. However, this radiological modality have no capacity to actually visualize hyaline cartilage, and is based on bone remodeling due to its overstrain after cartilage damage. Therefore, the conventional X-rays have little potential for early OA diagnosis. Other modalities that are implemented in the diagnostic protocol are ultrasound (US) and magnetic resonance imaging (MRI). Ultrasound stands out from the other methods mentioned above in that it is the only dynamic diagnostic method, but the bone is impenetrable to ultrasound waves, so a large area of the knee joint is inaccessible to this examination and the ultrasound results are not significantly different from those obtained with an AP radiograph in the standing position (Riecke et al., 2014). Therefore, for cartilage lesions most commonly MRI is utilized. MRI is by far the most sensitive and specific diagnostic tool, however, literature shows that its sensitivity is widely spread ranging 45% up to 94% (Figueroa et al., 2007). Moreover, MRI was shown to grossly underestimate the cartilage lesion grade especially in early stages of OA (Krakowski et al., 2021a). Moreover, evaluation of cartilage in magnetic resonance imaging requires specialized radiologists, its costly and time consuming. Therefore, examination awaiting time can be prolonged. Solivetti et al. (2016) showed that up to 20% of patients referred for MRI did not receive proper physical examination prior to MRI. Those referrals were mainly made by general practitioners, who are not trained in orthopaedic examination (Krakowski et al., 2021b; Krakowski et al., 2021c). Given the information above it seems that orthopaedic community requires additional examination tools that can be implemented in clinical practice. Vibroacoustic diagnostics is a scientific field that deals with the study of vibrations and sounds in various structures, such as machinery, vehicles, buildings or electronic devices. It is used to determine the technical condition of these structures and identify possible problems or damage. Vibroacoustic machine diagnostics is a method used to assess the technical condition of machinery and equipment. It relies on the analysis of vibrations and sounds emitted by machines during their operation. By monitoring and analyzing these signals, it is possible to detect problems early, such as uneven wear, malfunctioning components, or the need for lubrication. Vibroacoustic diagnostic methods are widely used in industry, especially in areas related to safety, reliability and operational efficiency of machinery and equipment, and in recent years it has also been applied in medicine (Jonak et al., 2019).

Vibroarthrography (VAG) seems to fulfill this examination gap (Nalband et al., 2018). Healthy cartilage is capable of almost frictionless motion, therefore it generates little vibration, with the advance of surface fibrillation and cartilage loss more and more acoustic signals can be obtained from affected joint (Prior et al., 2010; Schlüter et al., 2019). Those acoustic signals generated by change in stress distribution can be recorded from knee joint surface (Nevalainen et al., 2021). Vibroarthrography is inexpensive, reproducible, fast and does not require use of radiation, therefore it can be a proposed method for patient screening in suspicion of OA (Karpiński et al., 2021b; Karpiński et al., 2021a). The first to introduce examination of acoustic changes in the knee joint was Blodgett in 1902, when he showed changes in generated knee sounds depending on cartilage status. Over the course of last century, VAG showed in preclinical evaluation to be effective in determining cartilage loss with accuracy surpassing 90 percent (Andersen et al., 2016). Both acoustic and vibration emission have been proposed and evaluated in the research (Emadi Andani & Salehi, 2024; Łysiak et al., 2020). However, up to this date there is no specific examination nor acquisition protocols that are widely accepted by the research society. What is also important is the fact that most of published studies used imaging modalities or a simple physical examination as a reference for cartilage examinations, which obviously poses a high risk of bias in these studies. Computer-Aided Diagnosis (CAD) plays a pivotal role in contemporary medicine, providing significant support to physicians in the process of diagnosing various ailments. It contributes substantially to the advancement of medical practices, offering assistance in the analysis of medical images from various modalities, such as MRI, CT, and mammography. CAD is exceptionally beneficial in the early detection of diseases, including cancer, due to its ability to identify subtle abnormalities that may escape the human eye. It aids doctors by offering a second opinion, assists in the quantification of changes, which is crucial for monitoring disease progression, and utilizes advanced algorithms for pattern recognition indicative of specific illnesses. Furthermore, it is applied in personalized medicine and serves as an educational tool.

Artificial intelligence methods, such as artificial neural networks, are now one of the most important and widely used tools in classification and regression tasks. They are used in various fields of life and industry due to their ability to learn and adapt to complex patterns in data (Machrowska et al., 2020a; Machrowska et al., 2020b; Rogala et al., 2019; Szabelski et al., 2022). Due to their ability to automatically learn and recognize complex patterns in data, they are widely used in various medical fields including medical diagnosis in cases, where typical statistical methods do not work. These methods allow the discovery of complex relationships, the explanation of observed trends and the possibility of prospective use seem to be extremely helpful in clinical practice.

Therefore, a study was carried out in which the cartilage status is visually verified during surgical treatment of the knee joint. The purpose of this study was to create a testing and acquisition protocol for vibroacoustic joint testing, and then to compare selected classification methods using machine learning, and to evaluate the usefulness of the various tools in diagnosing articular cartilage damage of the knee based on recorded vibroacoustic signals.

This article consists of the following four parts: an introduction, which describes the theoretical basis of VAG and possible clinical applications, the second paragraph describes the study and control group in detail, as well as the signal acquisition and extraction protocol. The third paragraph presents the results with a discussion and summary of the results of other study groups. The fourth paragraph provides conclusions based on the results.

2. MATERIALS AND METHODS

2.1. Participants

The study involved 84 participants, of whom 40 were allocated to the control group and 44 to the study group. Healthy control group was enrolled from volunteers without any previous history of knee trauma or pain and proved negative during physical examination. Any positive finding during physical examination e.g. positive meniscal tests have been considered as an exclusion rule. The study group have been enrolled out of patients scheduled for surgical treatment of knee joint due to intraarticular lesions. If only ligamentous or meniscal lesions, without cartilage chondromalacia was found during the surgery, the patient was excluded from the study group. Each patient signed written consent for participation in this study. Control group consisted of 13 males and 27 females and OA group 21 and 23 respectively. OA group showed higher base BMI (30.72) than control group (22.98). Also, in this study there was a discrepancy in age between HC with mean age 24.63 and 57.23 in OA group. This discrepancy is justified by the fact that osteoarthritis affects mostly elderly. A detailed description of the group of participants in the study, including mean values and standard deviation (SD) values, is presented in Table 1. The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of the Medical University of Lublin consent number KE-0254/261/2019 on 26/09/2019.

Tab. 1. Characteristics of study participants.

Study group	N	Males/ Females	Age (years ± SD)	Heigh (m ± SD)	Weight (kg ± SD)	BMI	Tegner-Lyshom score
Healthy control (HC)	40	13/27	24.63 ± 5.52	1,72 ± 0.08	69,15 ± 15.58	22,98 ± 3.57	98,55 ± 3.95
Osteoarthrisis (OA)	44	21/23	57,23 ± 13.24	1,70 ± 0.09	88,39 ± 14.06	30,72 ± 4.71	43,50 ± 15.14

2.2. Study protocol and signal acquisition

Recording of vibroacoustic signals generated by the moving knee joint was performed using the author's measurement system built on a typical commercially available knee orthosis, the block diagram of which is presented in Figure 1. The measurement system was based on the Arduino Mega2560 R3 module. A CM01B piezoelectric contact microphone connected to an analog input located on the patella was used for signal acquisition. The sampling rate used was 1400 Hz with 10-bit resolution. In addition, an EMS22A50-D28-LT6 Bourns digital encoder built into the axis of rotation of the orthosis was used to measure the position of the knee, which made it possible to determine the angular position of the joint. A galvanic barrier was used on the USB connector to ensure patient safety, and the device itself was powered by an 11.1V lithium-ion battery. Data were recorded using RealTerm software and transferred to a computer in ASCII format.

The recording of vibroacoustic signals was conducted for repetitive sequences of knee joint straightening and bending movements in the range of 90° - 0° - 90°. The study was carried out for sequences performed in a closed kinetic chain (Machrowska et al., 2019). The procedure adopted in recording the signals involved rising from a seated position in a chair with knee flexion to 90° until reaching a fully upright standing position (knee flexion angle was 0°), and then descending to return to a seated position (90° knee flexion) in about 2s. Signals were recorded for 10 complete repetitions of the procedure described above.

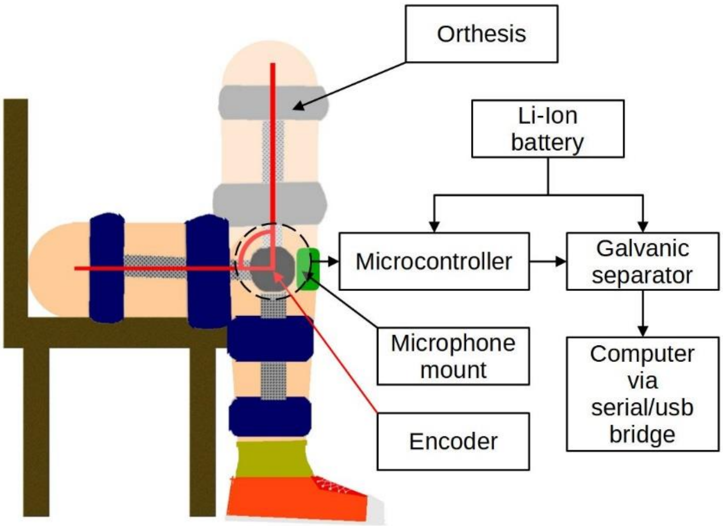

Fig. 1. Block diagram of the measurement system

2.3. Signal Preprocessing

Traditional signal analysis methods, such as the Fourier transform, assume that a signal is stationary, that is, has constant characteristics over time. However, many real signals, such as biomedical signals or vibroacoustic signals, change over time, making traditional analysis methods less effective. A solution to this problem may be to use the Ensemble Empirical Mode Decomposition (EEMD) procedure. This procedure is one form of adaptive modal decomposition EMD (Empirical Mode Decomposition). The fundamental idea of techniques using EMD-derived algorithms is to decompose complex signals into a finite number of components representing the signal in a selected frequency row. The process of such filtering is called "sifting," and the result is the obtaining of a set of IMFs (Intrinsic Mode Functions). A variation of EMD used in the presented research is EEMD (Wu & Huang, 2009). The EEMD variant makes it possible to reduce the occurrence of the unfavorable phenomenon known as mode mixing, which manifests itself in the leakage of the spectrum between separate components. An important difference between the base algorithm and EEMD is the addition of Gauss white noise, which allows to reduce the mentioned undesirable effect. In each of the techniques that fall into the "xEMD" group of methods, an iteration is carried out to extract individual IMFs based on envelopes built on the occurrence of local extremes. The result of EEMD, unlike the classical procedure, is to obtain a frequency-filtered signal with the addition of white noise of finite amplitude. The included component allows to fill the time-frequency space thus helping to properly separate the individual frequency components and thus reduce the phenomenon of spectrum leakage. The result of the presented research was a set of frequency components of the signal and a trend signal, which is the remainder of the waveform after the extraction of IMFs. For further analysis, a signal stripped of the monotonic trend was used so that it would not affect further results. The procedure used in this study was described in detail in the authors' previous studies (Jonak et al., 2019; Karpiński et al., 2022a; Karpiński et al., 2022b).

2.4. Feature extraction

A number of signal features (parameters), referred to as state indicators, are used in vibroacoustic diagnostics. The following is a summary of the parameters determined for the discrete signals used in this study. These parameters were chosen on the basis of literature data and previous studies by the authors, who selected the optimal signal features. Based on the results of preliminary research and literature analysis, eight indicators were determined for the recorded acoustic signals (Karpiński, 2022; Karpiński et al., 2019; Karpiński et al., 2022a; Karpiński et al., 2022b):

- The root mean square (RMS) value is one of the most widely used parameters determined in machine diagnostics. The RMS value takes into account the time history of the waveform and contains information about the magnitude of the amplitude, but is not sensitive to singular impulse in the signal.

$$x_{\mathrm{RMS}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i{}^2} \tag{1}$$

where: x_i – is the value of the discrete signal at the nth point, n = 1, ..., N;,
N – is the number of samples in the signal.

- Peak value (PV), also known as maximum signal value, unlike RMS, is an indicator that is very sensitive to rapid changes in the state of the objects under study.

$$\hat{x} = max|x_i| \tag{2}$$

- The crest factor (CF) is a measure that gives the ratio of the peak value (PV) to the RMS value of a signal. In other words, this factor indicates how extreme the peaks are in a waveform.

$$x_{\mathrm{CF}} = \frac{\hat{x}}{x_{\mathrm{RMS}}} \tag{3}$$

- The form factor (FF) is a measure that gives the ratio of the RMS value to the straightened average value (SA).

$$x_{\mathrm{SF}} = \frac{x_{\mathrm{RMS}}}{\bar{x}} \tag{4}$$

– Variance (VAR) is a measure of the dispersion of sample results around the center of the distribution; it is the expected value of the square of the variance of a random variable minus its population mean.

$$VAR = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{5}$$

– Skewness (SKW) is an asymmetry coefficient used to determine what the distribution looks like, i.e. whether the data is distributed on both sides of the mean. It is expressed by the formula:

$$SKW = \frac{\bar{x}-d}{s} \tag{6}$$

where: d – is the dominant, or value with the highest probability of occurrence, or the value most frequently occurring in the sample, while s is the value of the standard deviation,
$\bar{x}$ – average value.

– The M6A parameter is the sixth central moment normalized by the variance raised to the third power. It is defined as:

$$M6A = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^6}{\left[\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2\right]^3} \tag{7}$$

– The parameter M8A is called the eighth central moment, normalized by the variance to the fourth power. It is defined as:

$$M8A = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^8}{\left[\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2\right]^4} \tag{8}$$

2.5. Classification tools

In machine learning, there are many types of classifiers that are used to predict categories or labels for input data based on its features. Below are some common types of classifiers/classification methods in machine learning:

– Logistic regression - a basic classifier used for binary classification problems. It predicts the probability of belonging to one of two classes (Lemon et al., 2003).
– Decision trees - a graphical way of representing a set of decision rules. It works by dividing the data into subgroups until an explicit classification is obtained (Kotsiantis, 2013).
– Random forest - a collection of decision trees. Each tree in the forest is trained on a different subset of the data, and the result is averaged, leading to more stable and accurate predictions (Shaik & Srinivasan, 2019).

- Support Vector Machines (SVM) - an algorithm that finds optimal hyperplanes separating different classes in the feature space. It is used for both classification and regression problems (Zhang, 2012).
- K-NN (k nearest neighbors) - a method which assigns a new data point to a class based on most of the k nearest neighbors of that point in the feature space (Zhang et al., 2017).
- Naive Bayes classifier - based on a probability theory and assumes independence between features. Despite its simplicity, it often gives good results, especially for small data sets (Huang & Li, 2011).
- Gradient Boosting Machines (GBM) - a technique that builds multiple weak models (usually decision trees) and combines them into a strong model by iteratively correcting the errors of previous models (Aziz et al., 2020).
- Deep learning and Neural Networks - deep learning, uses neural networks to solve complex classification problems. These networks consist of multiple layers of neurons that learn data representations at different levels of abstraction (Goodfellow et al., 2016).
- Cascade classifier - a classifier that consists of a series of simple classifiers, each of which is responsible for a specific part of the feature space (Bennasar et al., 2014).

In this study, multilayer perceptron - MLP, radial basis function - RBF, support vector method - SVM and naive classifier - NBC were used to solve classification issues. All calculations were carried out using the Statistica 13.3 package. These classifiers differ in their operating principle, learning method, as well as layout structure. For this reason, when making assignments of cases to different classes, they make errors corresponding to different positions of the data (Ghahramani & Kim, 2003).

Multilayer Perceptron (MLP) is one of the basic models used in machine learning. It is a type of artificial neural network that consists of at least three layers: an input layer, one or more hidden layers, and an output layer. MLP is an example of a supervised deep learning model that can be applied to both classification and regression problems. This type of neural network has been widely and extensively described in the literature (Bose, 2007; Rogala et al., 2021) .

A radial basis function (RBF) network is a type of artificial neural network that, like an MLP-type network, consists of three layers: an input layer, a hidden layer and an output layer. It is a machine learning model used for both classification and regression. RBF networks have the ability to model complex relationships in data, especially when those relationships are nonlinear and local. They are used in various fields such as pattern recognition, signal processing, and forecasting (Meng Joo Er et al., 2002).

Support Vector Machines (SVM) is a supervised machine learning algorithm applied to both classification and regression problems. SVM is popular for its effectiveness in the machine learning field and its ability to handle both linear and non-linear problems. A key advantage of SVM is its ability to deal with complex relationships in data, even for high-dimensional data (Chih-Wei & Chih-Jen, 2002).

Naive Bayes classifier is a probabilistic classification algorithm that is based on Bayes' theorem and the assumption of mutual independence between features (hence the name "naive"). Despite its simplicity, it often produces satisfactory results, especially for large data sets and text classification problems. It is often used in email spam classification tasks, sentiment analysis in texts or document categorization (R. Liu et al., 2018).

In summary, the main differences between these classifiers lie in their structure, mode of operation, application scope, and approach to solving classification problems. MLP and RBF are neural networks that differ in structure and functions, SVM is a technique based on hyperplanes and margins, while the naive Bayes classifier is a simple probabilistic approach based on feature independence.

2.6. Evaluation of classification accuracy

Typical indices such as ROC (Receiver Operating Characteristic) curves and F1 and MCC (Matthews correlation coefficient) were used to assess the accuracy of the selected classification methods. These are measures that allow comparison of classification quality also in groups where unbalanced data sets are analyzed.

ROC curves are a popular way to assess the quality of classification models. They are a tool for assessing the correctness of a classifier's performance, especially for assessing the quality of a model that predicts membership in two classes. An ROC curve is a graph that shows the relationship between the degree of true positives and the degree of false positives for different decision thresholds. The closer the ROC curve is to the upper left corner, the better the model is, as it has a lower value of false positives and a higher value of true positives. It is a graph showing sensitivity versus the proportion of false positives (1-specificity) for each cutoff point. The ROC curve is a commonly used tool for evaluating and cross-comparing classification models. A very popular approach is to determine the area under the graph of the ROC curve, denoted as AUC (area under curve), and treat it as a measure of the classification accuracy of a given model. The value of the AUC index takes values in the range [0,1], the larger it is, the better the model.

An effective solution to the class imbalance problem is the Matthews correlation coefficient, which is a special case of the ϕ - phi coefficient (Chicco & Jurman, 2020). It is one measure of the quality of binary classification that takes into account both the predictive accuracy of positive and negative class labels. MCC is a linear measure, taking values from -1 to 1. A value of +1 indicates excellent agreement between actual and predicted classification results, a value of 0 indicates no relationship, a value of -1 indicates full disagreement (Luque et al., 2019).

Another measure used to assess the accuracy of binary classification is the F1 score. This measure combines precision and sensitivity into a single value. F1 is the harmonic mean between precision and sensitivity. Precision measures how many of the positive cases detected are true positives, and sensitivity measures how many of the true positives are detected. The F1 value ranges from 0 to 1, where 1 represents perfect classification accuracy. In practice, as with the value of the area under the ROC curve, the higher the F1 score, the better the binary classifier performs (Chicco & Jurman, 2020; Luque et al., 2019).

3. RESULTS AND DISCUSSION

Details of the classification results obtained in the analyzed problem using multilayer perceptron (MLP) neural networks, radial basis function (RBF) networks, support vector machine (SVM) method and classification using naive Bayes classifier (NBC) are presented below. Table 2 summarizes the details of classification accuracy for each method. The highest overall accuracy of almost 90% was achieved using MPL-type networks with 85.5%

and 95.45% correct classifications in the HC and OA groups, respectively. Classification accuracy for RBF networks was 82.14%. The lowest accuracy of 70.24 was obtained for the naive Bayes classifier, with only 50% accuracy in the OA group. To the best of the authors' knowledge, the MLP classification method was only tested by their group (Karpiński et al., 2022a; Karpiński et al., 2022b). Therefore, the authors are not able to compare their results with other researchers. On the other hand, RBF classification have been studied in detail by Rangayyan et al. (2013). In their research high values of AUC ranging 0.711up to 1 could be found. In the present study, the AUC using RBF was 0.821, which is lower than that reported in most of Rangayyan's work. This high accuracy can be attributed to fractal data analysis, which was implemented in those studies. The authors state that their results also differ from those published by Shidore et al. (2021) in regard to SVM classification. In their study, the reported AUC reached 0.926, while in the present study it was only 0.756 and the lowest accuracy was found using NBC. This classification method showed highest sensitivity reaching 0.88, however, its specificity was only 0.627. Different results were presented by Wu et al. (2013) and Yang et al. (2014). Both authors presented that, using NBC, the specificity was 0.936 and 0.979, respectively. The authors' results, unlike those of other researchers, were validated using intraoperative cartilage visualisation. The authors believe that such differences in their and other researchers' results are due to different methods of cartilage validation.

Tab. 2. Summary of classification accuracy for all analyzed methods.

Classification method		**HC**	**OA**	**Total**
MLP	Total	40.00	44.00	84.00
	Correct	33.00	42.00	75.00
	Incorrect	7.00	2.00	9.00
	Correct (%)	82.50	95.45	89.29
	Incorrect (%)	17.50	4.55	10.71
RBF	Total	40.00	44.00	84.00
	Correct	32.00	37.00	69.00
	Incorrect	8.00	7.00	15.00
	Correct (%)	80.00	84.09	82.14
	Incorrect (%)	20.00	15.91	17.86
SVM	Total	40.00	44.00	84.00
	Correct	33.00	30.00	63.00
	Incorrect	7.00	14.00	21.00
	Correct (%)	82.50	68.18	75.00
	Incorrect (%)	17.50	31.82	25.00
NBC	Total	40.00	44.00	84.00
	Correct	37.00	22.00	59.00
	Incorrect	3.00	22.00	25.00
	Correct (%)	92.50	50.00	70.24
	Incorrect (%)	7.50	50.00	29.76

Figure 2 presents a summary of the ROC curves for the classifiers analyzed. The highest value of the area under the curve was obtained for the MLP-type network (0.875), while the lowest value was obtained for the naive Bayes classifier (0.754).

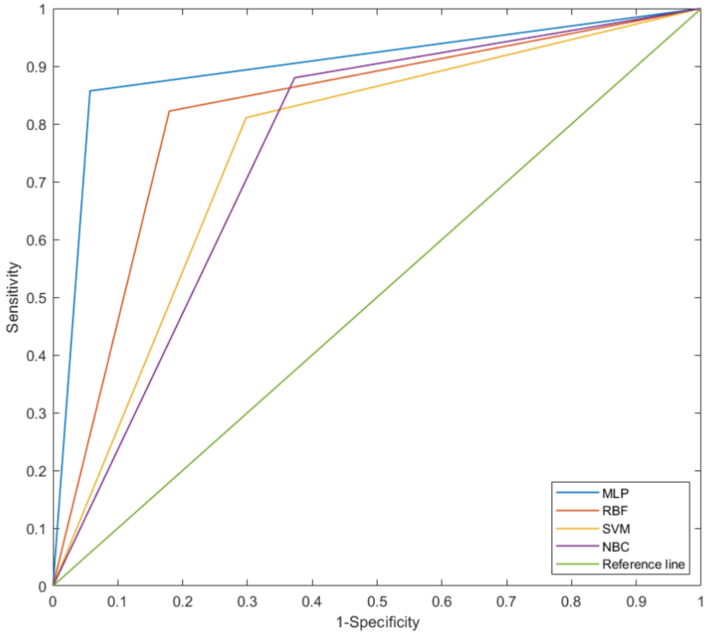

Fig. 2. Comparison of ROC curves for all classification methods.

A detailed analysis of the obtained classification parameters is presented in Table 3. The obtained results show that the MLP-type neural classifier coped with the analyzed classification problem best, obtaining MCC- 0.790 and F1 - 0.903, respectively. In the case of RBF-type networks, the results were slightly lower and amounted to MCC - 0.642 and 0.831, respectively. On the other hand, the naive Bayes classifier coped with the analyzed problem weakest, obtaining MCC - 0.464 and F1 0.638, respectively.

Tab. 3. Comparison of classification parameters for the analyzed methods.

Classification method	Accuracy (%)	Sensitivity	Specificity	AUC	Precision	Recall	F1 score	MCC
MLP	0.893	0.857	0.943	0.875	0.955	0.857	0.903	0.790
RBF	0.821	0.822	0.821	0.837	0.841	0.822	0.831	0.642
SVM	0.750	0.811	0.702	0.756	0.682	0.811	0.741	0.510
NBC	0.702	0.880	0.627	0.754	0.500	0.880	0.638	0.464

Vibroarthrography (VAG) as a method for joint assessment has certain limitations, including subjective result interpretation dependent on the operator, lack of visual imaging of joint structures, absence of standardized assessment criteria, variable sensitivity and specificity, constraints in cases of obese or elderly patients, and the need for specialized equipment. Despite these limitations, VAG can be a useful tool in joint evaluation, especially when combined with other diagnostic methods. However, consulting with a physician for an accurate diagnosis and treatment plan is always recommended.

4. CONCLUSIONS

The study showed high diagnostic accuracy for MLP and RBF classification methods. Specificity of 0.943 and sensitivity of 0.857 provides sound grounds for implementing this examination protocol in clinical practice. NBC proved to have little value in diagnostics, therefore its use in clinical setting is highly questionable. The results confirm that vibroarthrography can be an inexpensive, non-invasive, safe, and, most importantly, effective tool for detecting degenerative changes occurring in the knee joints. Further studies are necessary to establish golden standard in VAG examination, as well as signal extraction and acquisition. The authors plan to conduct extensive research to determine the best method of signal processing, as well as the appropriate clinical indications for the use of VAG in patients.

Author Contributions

Conceptualization, R.K., P.K. and J.J.; methodology, R.K.; software, R.K. and A.M.; validation, R.K.; formal analysis, R.K., and J.J.; investigation, R.K., P.K. and J.J.; resources, R.K., and M.M.; data curation, A.M.; writing—original draft preparation, R.K., P.K. and A.M.; writing—review and editing, R.K..; visualisation, R.K. and M.M.; supervision, J.J.; project administration, R.K.; funding acquisition, R.K. and J.J. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the Scientific Fund of the Lublin University of Technology: FD/IM/052.

Institutional Review Board Statement:

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of the Medical University of Lublin consent number KE-0254/261/2019 on 26/09/2019.

Conflicts of Interest

The authors declare no conflicts of interest.

REFERENCES

Andersen, R. E., Arendt-Nielsen, L., & Madeleine, P. (2016). A review of engineering aspects of vibroarthography of the knee joint. *Critical Reviews in Physical and Rehabilitation Medicine*, *28*(1–2), 13–32. https://doi.org/10.1615/CritRevPhysRehabilMed.2016017185

Ashoorion, V., Sadeghirad, B., Wang, L., Noori, A., Abdar, M., Kim, Y., Chang, Y., Rehman, N., Lopes, L. C., Couban, R. J., Aminilari, M., Malektojari, A., Ghazizadeh, S., Rehman, Y., Ghasemi, M., Adili, A., Guyatt, G. H., & Busse, J. W. (2023). Predictors of persistent post-surgical pain following total knee arthroplasty: A systematic review and meta-analysis of observational studies. *Pain Medicine*, *24*(4), 369–381. https://doi.org/10.1093/pm/pnac154

Aziz, N., Akhir, E. A. P., Aziz, I. A., Jaafar, J., Hasan, M. H., & Abas, A. N. C. (2020). A study on gradient boosting algorithms for development of AI monitoring and prediction systems. *2020 International Conference on Computational Intelligence (ICCI)* (pp. 11–16). IEEE. https://doi.org/10.1109/ICCI51257.2020.9247843

Barnett, A. J., & Toms, A. D. (2012). Revision total hip and knee replacement. *Clinics in Geriatric Medicine*, *28*(3), 431-446. https://doi.org/10.1016/j.cger.2012.05.008

Bennasar, M., Setchi, R., Hicks, Y., & Bayer, A. (2014). Cascade classification for diagnosing dementia. *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 2535–2540). IEEE. https://doi.org/10.1109/SMC.2014.6974308

Bose, B. K. (2007). Neural network applications in power electronics and motor drives - An introduction and perspective. *IEEE Transactions on Industrial Electronics*, *54*(1), 14–33. https://doi.org/10.1109/TIE.2006.888683

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*, 6. https://doi.org/10.1186/s12864-019-6413-7

Chih-Wei, H., & Chih-Jen, L. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, *13*(2), 415–425. https://doi.org/10.1109/72.991427

Emadi Andani, M., & Salehi, Z. (2024). An affordable and easy-to-use tool to diagnose knee arthritis using knee sound. *Biomedical Signal Processing and Control*, *88*, 105685. https://doi.org/10.1016/j.bspc.2023.105685

Figueroa, D., Calvo, R., Vaisman, A., Carrasco, M. A., Moraga, C., & Delgado, I. (2007). Knee chondral lesions: incidence and correlation between arthroscopic and magnetic resonance findings. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, *23*(3), 312-315. https://doi.org/10.1016/j.arthro.2006.11.015

Ghahramani, Z., & Kim, H. C. (2003). Bayesian classifier combination. *Gatsby Computational Neuroscience Unit University College London*.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Huang, Y., & Li, L. (2011). Naive Bayes classification algorithm based on small sample set. *2011 IEEE International Conference on Cloud Computing and Intelligence Systems* (pp. 34–39). IEEE. https://doi.org/10.1109/CCIS.2011.6045027

Jonak, J., Karpinski, R., Machrowska, A., Krakowski, P., & Maciejewski, M. (2019). A preliminary study on the use of EEMD-RQA algorithms in the detection of degenerative changes in knee joints. *IOP Conference Series: Materials Science and Engineering*, *710*, 012037. https://doi.org/10.1088/1757-899X/710/1/012037

Karpiński, R. (2022). Knee joint osteoarthritis diagnosis based on selected acoustic signal discriminants using machine learning. *Applied Computer Science*, *18*(2), 71–85. https://doi.org/10.35784/acs-2022-14

Karpiński, R., Krakowski, P., Jonak, J., Machrowska, A., Maciejewski, M., & Nogalski, A. (2021a). Analysis of differences in vibroacoustic signals between healthy and osteoarthritic knees using EMD algorithm and statistical analysis. *Journal of Physics: Conference Series*, *2130*, 012010. https://doi.org/10.1088/1742-6596/2130/1/012010

Karpiński, R., Krakowski, P., Jonak, J., Machrowska, A., Maciejewski, M., & Nogalski, A. (2021b). Estimation of differences in selected indices of vibroacoustic signals between healthy and osteoarthritic patellofemoral joints as a potential non-invasive diagnostic tool. *Journal of Physics: Conference Series*, *2130*, 012009. https://doi.org/10.1088/1742-6596/2130/1/012009

Karpiński, R., Krakowski, P., Jonak, J., Machrowska, A., Maciejewski, M., & Nogalski, A. (2022a). Diagnostics of articular cartilage damage based on generated acoustic signals using ANN - Part I: Femoral-tibial joint. *Sensors*, *22*(6), 2176. https://doi.org/10.3390/s22062176

Karpiński, R., Krakowski, P., Jonak, J., Machrowska, A., Maciejewski, M., & Nogalski, A. (2022b). Diagnostics of articular cartilage damage based on generated acoustic signals using ANN - Part II: Patellofemoral joint. *Sensors*, *22*(10), 3765. https://doi.org/10.3390/s22103765

Karpiński, R., Machrowska, A., & Maciejewski, M. (2019). Application of acoustic signal processing methods in detecting differences between open and closed kinematic chain movement for the knee joint. *Applied Computer Science*, *15*(1), 36–48. https://doi.org/10.23743/acs-2019-03

Kotsiantis, S. B. (2013). Decision trees: A recent overview. *Artificial Intelligence Review*, *39*, 261–283. https://doi.org/10.1007/s10462-011-9272-4

Krakowski, P., Karpiński, R., Jojczuk, M., Nogalska, A., & Jonak, J. (2021a). Knee MRI underestimates the Grade of cartilage lesions. *Applied Sciences*, *11*(4), 1552. https://doi.org/10.3390/app11041552

Krakowski, P., Karpiński, R., Jonak, J., & Maciejewski, R. (2021b). Evaluation of diagnostic accuracy of physical examination and MRI for ligament and meniscus injuries. *Journal of Physics: Conference Series*, *1736*, 012027. https://doi.org/10.1088/1742-6596/1736/1/012027

Krakowski, P., Karpiński, R., Maciejewski, R., & Jonak, J. (2021c). Evaluation of the diagnostic accuracy of MRI in detection of knee cartilage lesions using Receiver Operating Characteristic curves. *Journal of Physics: Conference Series*, *1736*, 012028. https://doi.org/10.1088/1742-6596/1736/1/012028

Lemon, S. C., Roy, J., Clark, M. A., Friedmann, P. D., & Rakowski, W. (2003). Classification and regression tree analysis in public health: Methodological review and comparison with logistic regression. *Annals of Behavioral Medicine*, *26*(3), 172–181. https://doi.org/10.1207/S15324796ABM2603_02

Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, *108*, 33–47. https://doi.org/10.1016/j.ymssp.2018.02.016

Luque, A., Carrasco, A., Martín, A., & De Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, *91*, 216–231. https://doi.org/10.1016/j.patcog.2019.02.023

Łysiak, A., Froń, A., Bączkowicz, D., & Szmajda, M. (2020). Vibroarthrographic signal spectral features in 5-class knee joint classification. *Sensors*, *20*(17), 5015. https://doi.org/10.3390/s20175015

Machrowska, A., Karpiński, R., Jonak, J., Szabelski, J., & Krakowski, P. (2020a). Numerical prediction of the component-ratio-dependent compressive strength of bone cement. *Applied Computer Science*, *16*(3), 88-101. https://doi.org/10.23743/acs-2020-24

Machrowska, A., Karpiński, R., Krakowski, P., & Jonak, J. (2019). Diagnostic factors for opened and closed kinematic chain of vibroarthrography signals. *Applied Computer Science*, *15*(3), 34-44. https://doi.org/10.23743/acs-2019-19

Machrowska, A., Szabelski, J., Karpiński, R., Krakowski, P., Jonak, J., & Jonak, K. (2020b). Use of Deep Learning Networks and statistical modeling to predict changes in mechanical parameters of contaminated bone cements. *Materials*, *13*(23), 5419. https://doi.org/10.3390/ma13235419

Meng Joo Er, Shiqian Wu, Juwei Lu, & Hock Lye Toh. (2002). Face recognition with radial basis function (RBF) neural networks. *IEEE Transactions on Neural Networks*, *13*(3), 697–710. https://doi.org/10.1109/TNN.2002.1000134

Nalband, S., Prince, A., & Agrawal, A. (2018). Entropy-based feature extraction and classification of vibroarthographic signal using complete ensemble empirical mode decomposition with adaptive noise. *IET Science, Measurement & Technology*, *12*(3), 350–359. https://doi.org/10.1049/iet-smt.2017.0284

Nevalainen, M. T., Veikkola, O., Thevenot, J., Tiulpin, A., Hirvasniemi, J., Niinimäki, J., & Saarakkala, S. S. (2021). Acoustic emissions and kinematic instability of the osteoarthritic knee joint: Comparison with radiographic findings. *Scientific Reports*, *11*, 19558. https://doi.org/10.1038/s41598-021-98945-2

Prior, J., Mascaro, B., Shark, L. K., Stockdale, J., Selfe, J., Bury, R., Cole, P., & Goodacre, J. A. (2010). Analysis of high frequency acoustic emission signals as a new approach for assessing knee osteoarthritis. *Annals of the Rheumatic Diseases*, *69*, 929–930. https://doi.org/10.1136/ard.2009.112599

Rangayyan, R. M., Oloumi, F., Wu, Y., & Cai, S. (2013). Fractal analysis of knee-joint vibroarthrographic signals via power spectral analysis. *Biomedical Signal Processing and Control*, *8*(1), 23-29. https://doi.org/10.1016/j.bspc.2012.05.004

Riecke, B. F., Christensen, R., Torp-Pedersen, S., Boesen, M., Gudbergsen, H., & Bliddal, H. (2014). An ultrasound score for knee osteoarthritis: A cross-sectional validation study. *Osteoarthritis and Cartilage*, *22*(10), 1675–1691. https://doi.org/10.1016/j.joca.2014.06.020

Rogala, M., Gajewski, J., & Ferdynus, M. (2019). Numerical analysis of the thin-walled structure with different trigger locations under axial load. *IOP Conference Series: Materials Science and Engineering*, *710*, 012028. https://doi.org/10.1088/1757-899X/710/1/012028

Rogala, M., Gajewski, J., & Górecki, M. (2021). Study on the effect of geometrical parameters of a hexagonal trigger on energy absorber performance using ANN. *Materials*, *14*(20), 5981. https://doi.org/10.3390/ma14205981

Schlüter, D. K., Spain, L., Quan, W., Southworth, H., Platt, N., Mercer, J., Shark, L. K., Waterton, J. C., Bowes, M., Diggle, P. J., Dixon, M., Huddleston, J., & Goodacre, J. (2019). Use of acoustic emission to identify

novel candidate biomarkers for knee osteoarthritis (OA). *PLOS ONE*, *14*(10), e0223711. https://doi.org/10.1371/journal.pone.0223711
Shaik, A. B., & Srinivasan, S. (2019). A brief survey on random forest ensembles in classification model. In S. Bhattacharyya, A. E. Hassanien, D. Gupta, A. Khanna & I. Pan (Eds.), *International Conference on Innovative Computing and Communications* (Vol. 56, pp. 253–260). Springer Singapore. https://doi.org/10.1007/978-981-13-2354-6_27
Shidore, M. M., Athreya, S. S., Deshpande, S., & Jalnekar, R. (2021). Screening of knee-joint vibroarthrographic signals using time and spectral domain features. *Biomedical Signal Processing and Control*, *68*, 102808. https://doi.org/10.1016/j.bspc.2021.102808
Singh, J. A., Yu, S., Chen, L., & Cleveland, J. D. (2019). Rates of total joint replacement in the United States: future projections to 2020–2040 using the National Inpatient Sample. *The Journal of Rheumatology*, *46*(9), 1134–1140. https://doi.org/10.3899/jrheum.170990
Solivetti, F. M., Guerrisi, A., Salducca, N., Desiderio, F., Graceffa, D., Capodieci, G., Romeo, P., Sperduti, I., & Canitano, S. (2016). Appropriateness of knee MRI prescriptions: Clinical, economic and technical issues. *La Radiologia Medica*, *121*, 315-322. https://doi.org/10.1007/s11547-015-0606-1
Szabelski, J., Karpiński, R., & Machrowska, A. (2022). Application of an Artificial Neural Network in the modelling of heat curing effects on the strength of adhesive joints at elevated temperature with imprecise adhesive mix ratios. *Materials*, *15*(3), 721. https://doi.org/10.3390/ma15030721
W-Dahl, A., Kärrholm, J., Rogmark, C., Mohaddes, M., Carling, M., Sundberg, M., Bülow, E., Nåtman, J., Carlsen, H., Isaksson, R., & Rolfson, O. (2022). *Annual Report 2022*. Swedish Arthroplasty Register. https://registercentrum.blob.core.windows.net/refdocs/10.18158/BklrLg8NOo.pdf
Williams, J., & Pierre-Louis, K. (2024). Osteoarthritis of the Knee. *Physician Assistant Clinics*, *9*(1), 59–69. https://doi.org/10.1016/j.cpha.2023.08.003
Wu, Y., Cai, S., Yang, S., Zheng, F., & Xiang, N. (2013). Classification of knee joint vibration signals using bivariate feature distribution estimation and maximal posterior probability fecision criterion. *Entropy*, *15*(4), 1375-1387. https://doi.org/10.3390/e15041375
Wu, Z., & Huang, N. E. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, *01*(01), 1–41. https://doi.org/10.1142/S1793536909000047
Yang, S., Cai, S., Zheng, F., Wu, Y., Liu, K., Wu, M., Zou, Q., & Chen, J. (2014). Representation of fluctuation features in pathological knee joint vibroarthrographic signals using kernel density modeling method. *Medical Engineering & Physics*, *36*(10), 1305–1311. https://doi.org/10.1016/j.medengphy.2014.07.008
Zhang, S., Li, X., Zong, M., Zhu, X., & Cheng, D. (2017). Learning *k* for kNN classification. *ACM Transactions on Intelligent Systems and Technology*, *8*(3), 1–19. https://doi.org/10.1145/2990508
Zhang, Y. (2012). Support vector machine classification algorithm and its application. In C. Liu, L. Wang, & A. Yang (Eds.), *Information Computing and Applications* (Vol. 308, pp. 179–186). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-34041-3_27