Vibroarthrography using Convolutional Neural Networks

Dimitri Kraft
University of Rostock
Rostock, Germany
dimitri.kraft@vcric.igd-r.fraunhofer.de

Gerald Bieber
Fraunhofer-Institut fuer Graphische Datenverarbeitung
Rostock, Germany
gerald.bieber@igd-r.fraunhofer.de

ABSTRACT

Knees, hip, and other human joints generate noise and vibration while they move. The vibration and sound pattern is characteristic not only for the type of joint but also for the condition. The pattern vary due to abrasion, damage, injury, and other causes. Therefore, the vibration and sound analysis, also known as vibroarthrography (VAG), provides information and possible conclusions about the joint condition, age and health state. The analysis of the pattern is very sophisticated and complex and so approaches of machine learning techniques were applied before. In this paper, we are using convolutional neural networks for the analysis of vibroarthrographic signals and compare the results with already known machine learning techniques.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing design and evaluation methods**; • **Ubiquitous and mobile computing systems and tools**; • **Applied computing** → Life and medical sciences; • **Computing methodologies** → Machine learning;

KEYWORDS

Vibroarthrography, deep learning, deep neural networks, convolutional neural networks, spectrogram

ACM Reference Format:
Dimitri Kraft and Gerald Bieber. 2020. Vibroarthrography using Convolutional Neural Networks. In *The 13th PErvasive Technologies Related to Assistive Environments Conference (PETRA '20), June 30-July 3, 2020, Corfu, Greece.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3389189.3397993

1 INTRODUCTION

A joint is the connection between bones in the body, which link the skeletal system into a functional whole. They are constructed to allow for different degrees and types of movement [12]. Knee joints in particular are one of the most vulnerable joints in the human body and are more prone to various injuries and disorders. Joint disorder and diseases are impacting the life of people drastically and are the main cause for disability of elderly people. Arthritis and other joint disorders are causing huge economic costs for nations. In the US, $303.5 billion or 1% of the 2013 US Gross Domestic Product (GDP) were spend to treat arthritis. One of the most common disease of joints is osteoarthritis (OA), a degenerative disease where changes in bones, articular cartilages and soft tissues occur. OA affecting nearly 10% of the population worldwide and is frequently observed in elderly people (44% in people > 80 years). OA was the second most costly health condition treated at US hospitals in 2013. In that year, OA accounted for $16.5 billion, or 4.3%, of the combined costs for all hospitalizations [19]. Tools and technologies for quantify the health status of human joints outside of the clinical setting are investigated by researchers throughout the past decade including the analysis of vibration and sound caused by human joints, the range of motion during extension and flexion and gait analysis in general. The analysis of vibration caused by human joints is called Vibroarthrography (VAG) [18] or Vibration arthrometry [10]. VAG is a non-invasive screening tool for vibration and sound analysis of natural joints using accelerometers which was first introduced by McCoy et al. in 1985 [18].

PETRA '20, June 30-July 3, 2020, Corfu, Greece

ACM ISBN 978-1-4503-7773-7/20/06…$15.00
https://doi.org/10.1145/3389189.3397993

2 RELATED WORK

The analysis of knee joint sound experiences a long history of investigation since 1976. In that year Chu et al. published a series of papers investigating the acoustic pattern of knee joints. The researchers found that cartilage damage around the knee may be classified with acoustic sensors [1]. Further they indicated that pattern recognition techniques in acoustic signals may be used to distinguish between normal, rheumatoid and degenerative knees [2]. Later in 1985, a research team evaluated the diagnostic potential of vibration arthrography by examining 250 subjects who were undergoing diagnostic arthroscopy and suggested that VAG will become a significant diagnostic aid in the clinical evaluation of the locomotive system. In 86% of these cases, a characteristic signal was obtained. Further, it was possible to identify the affected side and determine how far posteriorly the meniscal injury lay [11]. In 1990 Frank et al. [6] concluded that different states of pathology can be seen in different frequency bands, but also noted that further sub-classification requires a more sophisticated feature set. They proposed that those features can be used to train a machine learning classifier to develop procedures for categorization of the signals with respect to the degree of a joint disorder. Later, in 1992, Tavathia et al. [31] proposed a linear prediction modeling technique to segment and analyze VAG signals. Segmentation, as argued in their paper, is needed due to the non-stationary nature of VAG signals. By segmenting the signal into quasi-stationary segments, traditional spectral estimation methods, e.g. Fourier Transformation, can be applied to extract useful features from the frequency domain. Reddy et al. [27] showed that significantly differences in VAG signals can be found above 100Hz. After applying a discrete Fourier transformation on the VAG signals they observed that the mean power over the frequency range of 100 to 500 Hz statistically separated all the four groups in their study and provided quantitative measures of the knee damage. In 1997, Krishnan et al. [13] published a technique for filtering, modeling and classifying VAG signals. After segmenting and labeling the VAG signals, an autoregressive (AR) model provide useful coefficients to train a logistic regression classifier. They achieved a reasonable accuracy of 68.9% by using the first six AR coefficients as features. In the same year, Rangayyan et al. [23] obtained an accuracy of 75.6% via leave-one-out (LOO) cross validation (CV) with the same segmentation technique by using AR and cepstral coefficients with a logistic regression classifier. Krishnan et al. [14] argued in 2000 that the segmentation of VAG signals can be avoided by using a joint time-frequency distribution technique. This enables the analysis of VAG signals without a preliminary segmentation and therefore overcomes the drawbacks in associating the clinical information obtained during arthroscopy or auscultation with the segments of the corresponding VAG signal. By extracting spectral features from the time-frequency distributions (TFD) they achieved an accuracy of 68.9% with a logistic regression classifier. In 2007, Rangayyan et al. [25] proposed a less sophisticated technique to classify abnormal and normal VAG signals by extracting statistical features of the first and second half of a vibroarthrographic signal. They obtained a reasonable area under Receiver Operating Characteristics (ROC) curve of 0.78, considering that only two simple statistical features were used. For classification, an artificial neural network based on radial basis function was employed. The classification of VAG signals via spectral features like energy, energy spread, frequency, and frequency spread was performed by using a single-layer artificial neural network with an accuracy of 91.4% by combining the knee angle over time and VAG signal [12]. Wu and Krishnan [33] used a multiple classifier system (MCS) based on a recurrent neural network (RNN) and a least-squares support vector machine (LS-SVM) to classify a dataset [17], consisting of 89 knee-joint VAG signals with an accuracy of 80.9% by applying a Parzen window [21] to estimate the probability density function (PDF) of the signals. After estimation, entropy-based and statistical features are extracted from the PDF. This dataset is to date the only dataset, which is publicly available. It was used in multiple studies [13, 20, 23–25, 29, 32] since then. The aforementioned work only focuses on the assessment of the knee state by extracting features based on a priori knowledge of the signal characteristics. The recent advances in machine learning e.g. convolutional neural networks, may leverage the automatic feature extraction process and find richer representations of vibroarthrographic signals for classification.

2.1 Convolutional Neural Networks

Convolutional neural networks (CNN) were first introduced by Krizhevsky et al. [15] and gained attention by winning the ImageNet challenge [3] with a large margin in 2012. Since then, new CNN Architectures, e.g. ResNet [9] were introduced and enhanced the performance of CNN significantly. CNN's were mostly used for the image domain. 1D and 2D Convolutional Neural Networks are now widely used for the classification of speech, sound and vibration by either operating on the raw time domain signal data or converting the time-domain signals into the time-frequency domain via spectrograms or scalograms and then classified on pixel value level. Spectrograms are 2D images representing sequences of spectra with time along one axis, frequency along the other, and brightness or color representing the strength of a frequency component at each time frame. This representation is thus at least suggestive that some of the convolutional neural network architectures for images could be applied directly to sound [34]. Convolutional Neural Networks are exploiting the convolution operation to extract meaningful low and high level features. This may be seen as learning various filters which extract basic (e.g. vertical, horizontal edges) and high level features (e.g. mouth, nose or eyes of a face). A Convolutional Neural Network typically consists of several convolutional layers. Often each convolutional layer is followed by a non-linear activation function to induce non-linearity, as the convolution operation itself is linear. After the activation function, a dimension reduction operation, referred as pooling is applied to reduce overfitting and further provide translation invariance.

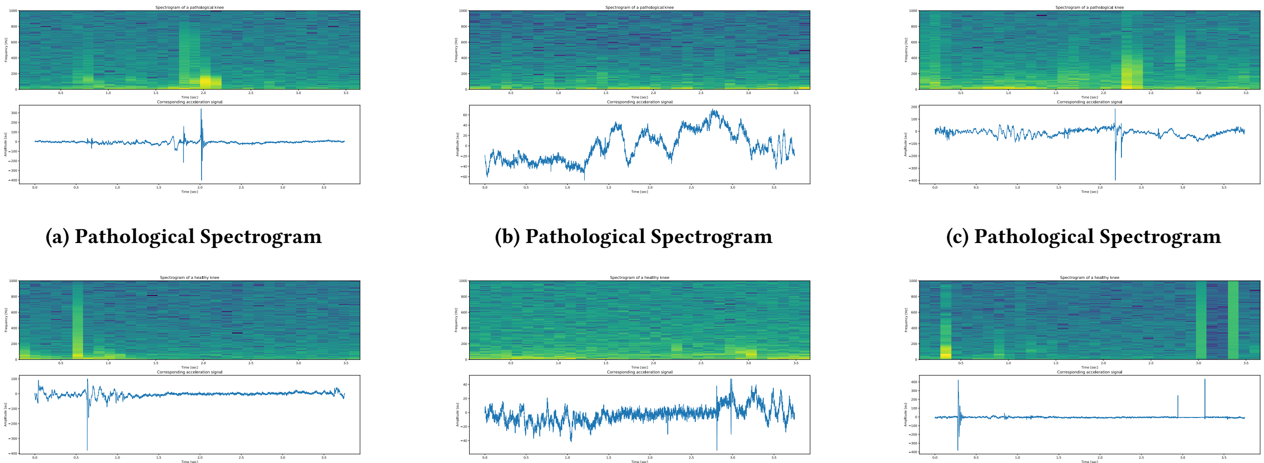
3 EXPERIMENTAL SETUP

3.1 Data Description

As the current availability of VAG signals is insufficient and the need for fair comparison is unavoidable, we use the widely used and only publicly available dataset for classification of vibroarthrographic signals published by Rangayyan et al. in 1997. This dataset contains 38 abnormal (pathological signals) and 51 normal signals (no identified pathology). No further sub-classes exists, although the abnormal class consists of at least five different pathologies. During the data acquisition process, Rangayyan et al. reported that each subject sat on rigid table in a relaxed position with the leg being tested freely suspended in air. The VAG signal was recorded by placing an accelerometer (model 3115a, Dytran, Chatsworth, CA) at the mid-patella position of the knee as the subject swung the leg over an approximate angle range of 135 degrees (approximately full flexion) to 0 degrees (full extension) and back to 135 degrees in 4 seconds. The transducer has a nominal sensitivity of 10 mV/G at 100 Hz, and a 3-dB bandwidth of 0.66 to 12,000 Hz. The subjects were provided verbal directions to complete the flexion to extension and back to flexion cycle in a period as close to 4 seconds as possible, with equal duration of 2 seconds for each of flexion and extension. The first half (approximately) of each VAG signal corresponds to extension, and the second half to flexion of the leg [26]. The 89 vibroarthrographic samples consists of either 7500 timestamps or 8000 timestamps each associated with an acceleration value of the used single axis accelerometer. To ensure same length across the samples, we cropped all signals to 7500 timestamps, by removing the last 500 timestamps.

3.2 Data Preparation

The data preparation process consists of converting the 89 one-dimensional vibroarthrographic data samples into two-dimensional image data in form of spectrograms. This renders the time series classification task into a image classification task. Spectrograms can be used as a way of visualizing the change of a non-stationary signal's frequency content over time by applying the Fast Fourier Transformation on overlapping windows of the signal with size L.



(a) Pathological Spectrogram **(b) Pathological Spectrogram** **(c) Pathological Spectrogram**

(d) Healthy Spectrogram **(e) Healthy Spectrogram** **(f) Healthy Spectrogram**

Figure 1: Spectrograms with corresponding vibration signals

The STFT is applying the following equation to each window $w(n)$ where $n \in [0 : N-1]$ with window length of N and a hop size H determining the overlapping:

$$X(m,k) = \sum_{n=0}^{N-1} X[n + mH]w[n]e^{-2\pi jkn/N}$$

with $m \in [0 : M]$ where $M = [\frac{L-N}{H}]$ and $k \in [0 : K]$ where $K = N/2$. To convert the times series data into a spectrogram we use the scipy 1.4.1 python package. We use a Tukey window for the STFT (Short Time Fourier Transformation) and set the window size to 128. After calculating the spectrogram, we scale it to decibel, as the scaling in amplitude is not logarithmic and thus the values cover a wide range of amplitude. By using 128 samples per window we have a resolution of about 15Hz and 64ms of time per bin. This results in an input shape of 65×66. We use the raw logarithmic scaled values of the resulting spectrogram and do not scale the spectrogram in any other way. When using the STFT one should notice that a wide window gives better frequency resolution but poor time resolution. A narrower window achieves a finer time resolution but poor frequency resolution. Examples for Spectrograms of the knee joint vibrations are denoted in figure 1. Note that these are stretched for visualization purposes.

3.2.1 Preprocessing. Preprocessing in Vibroarthrography mostly consists of applying moving average filters to eliminate baseline wandering [32] and filter other artifacts by levering band pass filters [29] or other high pass filter. We do not apply any preprocessing on vibroarthrographic data (e.g. filtering, scaling, standardization), besides the aforementioned calculation of spectrograms and do not think that any preprocessing in form of filtering is required for the usage of vibroarthrography with convolutional neural networks when using spectrograms.

3.2.2 Augmentation. Data Augmentation can be viewed as an injection of prior knowledge about the invariant properties of the data against certain transformations. Augmented data can cover unexplored input space, prevent overfitting, and improve the generalization ability of a deep learning model [7]. Before augmentation, one requires knowledge about the semantic characteristics of the input data. As the typically applied techniques like rotation, mirroring or masking on image data are not useful for spectrograms, we only applied shifting and flipping with a probability $p = 0.75$ and random shift $[-3500..3500]$ in samples along the time axis before converting into spectrograms. Other data augmentation techniques for time series data e.g. injecting Gaussian noise or scaling in amplitude are not applicable, as its not clear, if such transformations would distort the data in an unsafe way, thus decreasing the performance of the neural network. We suspect, that shifting or flipping the vibroarthrographic signal in the time dimension does not contribute to the performance of an CNN, as the learned filters are translation-invariant by definition. However, by shifting (rolling) the vector in the time dimension, elements that roll beyond the last position of the vector are re-introduced at the first position. This induce some additional variance to data and may affect the performance of the neural networks.

3.3 Model Architecture

The baseline CNN for the classification of spectrograms is a two-layer convolutional neural network architecture. Our approach to classification of the spectrograms consists of greatly reducing the number of layer and filter to prevent overfitting. Overfitting occur fairly fast on small datasets and a common strategy is reduce the number of learnable parameters. By removing the flattening layer right before the fully connected layer and replace them with a global average pooling (GAP) layer followed by a fully-connected layer we further reduced the number of trainable parameters significantly. After employing a grid search, we settle with the following structure:

$$C(32) - C(64) - FC(2)$$

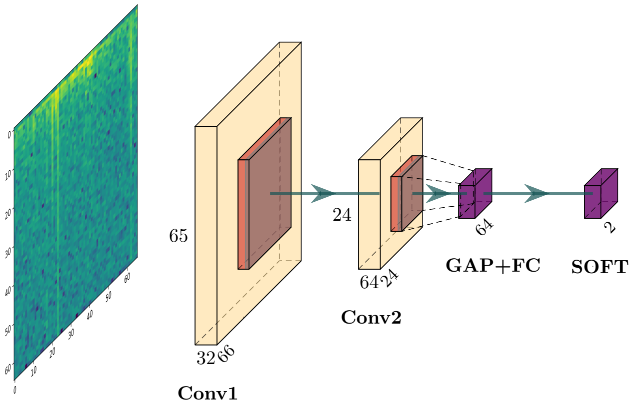


Figure 2: Architecture of our convolutional neural network

The model architecture is depicted in figure 2 Each convolutional layer $C(N)$, with N channels and a kernel size of 5 with padding of 2 on each side and stride of 3. The final fully-connected layer maps 64 neurons to two neurons for classification. After the first and second convolutional layer we use Batch Normalization followed by a ReLU activation function. The ReLU is a non-linear activation function which is defined as $\sigma(x) = max(0, \mathbf{x})$. We utilized the PyTorch 1.2.0 Framework [22] for training and testing our approach. Our model was trained 100 epochs with the AdamW Optimizer [16] with initial learning rate set to 0.0001, batch size of 32, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and no weight decay using the Cross Entropy loss function. All convolutional layers were initialized via He initialization [8].

3.3.1 Global Average Pooling Layer. Global Average Pooling (GAP) allow the Convolutional Neural Network to process data with different dimensions and reducing the number of learned parameters significantly [5]. Further a GAP layer enable one to use Class Activation Maps for visualization. Global Average Pooling calculates the mean for each channel (activation map) resulting in an vector of averaged activation values.

4 EVALUATION

4.1 Evaluation Method

In order to employ a robust evaluation process, our approach consists of applying a 5-Fold stratified cross validation on the 89 samples resulting in training five convolutional neural network models with different learned weights. A stratified Cross validation ensures, if possible, that a fold is evenly balanced with respect to labels. In our approach a single fold consists of 63 samples for training the neural network, 8 samples for check-pointing the best performing model on unseen validation data to avoid overfitting and 18 samples for final testing after each fold. Only the training data is seen by the network during training a single fold. With each fold, we reset the weights of the convolutional neural network to ensure, that each performance measure after a fold is not biased. We repeated the 5-Fold stratified cross validation five times and averaged the results for each fold to overcome initialization bias of the convolutional neural network and ensure a robust evaluation in terms of variance. This results in training the neural network 25 times to estimate the generalization performance. Figure 3 shows how the cross validation is performed for one trial.

Figure 3: Cross validation procedure

4.1.1 Evaluation metric. We assess the classification performance of our deep learning model by using the F_1 score, precision, recall, the Receiver Operating Characteristics (ROC) curve and area under curve (AUC) value. A ROC curve show the performance of a binary classifier system by varying the discrimination threshold, i.e. at which probability to classify a sample as positive or negative. A AUC value of 1 indicates perfect discrimination capabilities, where a AUC of 0.5 is the value to classify correctly by chance.

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

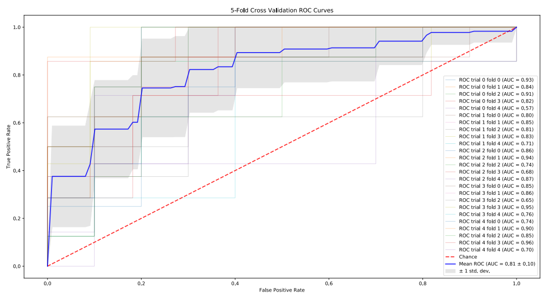
$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

4.2 Results

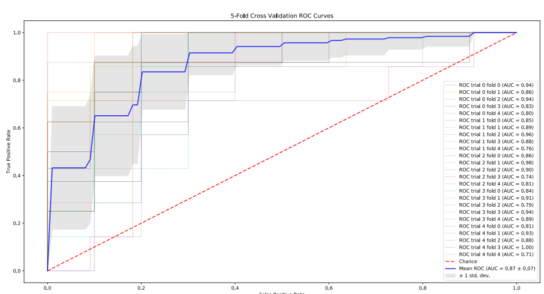
Our convolutional neural network performs comparably to other proposed work on vibroarthrography. We achieve an average F_1 score of 74% using a stratified 5-fold cross validation (see figure 1). We repeated the process five times to assess the variability of the performance measure and found that the performance of our convolutional neural network is varying with a standard deviation of 9% averaged over all folds with respect to the F_1 score when no data augmentation is used. While using Data Augmentation we observed an increase in F_1 score to 80% and approximately the same standard deviation. Precision, recall and AUC are denoted in table 1. We achieve comparable results with previous research with respect to the averaged AUC value of 0.87 and 0.81 with and without data augmentation, respectively (see figure 4).

Table 1: Averaged results for cross validation

	F_1		Precision		Recall		AUC	
	μ	σ	μ	σ	μ	σ	μ	σ
W/o Augmentation	0.74	0.09	0.76	0.07	0.77	0.10	0.81	0.10
W/ Augmentation	0.80	0.08	0.80	0.07	0.82	0.06	0.87	0.07



(a) ROC curves without data augmentation during cross validation



(b) ROC curves with data augmentation during cross validation

Figure 4: ROC curves for cross validation

5 DISCUSSION

Past Research in Vibroarthrography is often carried out by first applying several preprocessing steps and afterwards extracting hand-crafted features in vibroarthrographic signals, both requiring expert domain knowledge. Additionally, a specific classifier needs to be identified to distinguishes between symptomatic and non-symptomatic knees based on the extracted features. This work propose a novel approach to assessing the state of the knee condition by applying small-scale two-layer convolutional neural networks on spectrograms obtained from acceleration signals and achieve comparable performance to previous research. We convert the raw accelerometer signals into spectrograms and feed those into the convolutional network for classification. The hyperparameters e.g. number of layers, kernel size and number of filters for the CNN are easily found by a simple grid search or randomized search. During cross validation we observed that the performance of our convolutional neural network increases while using data augmentation. This indicates that mild augmentation in form of shifting and flipping in the time domain may be applicable to vibroarthrographic signals. However, the dataset used in the present study still suffer from several problems. First, the small number of subjects is often insufficient to train a robust machine learning model, especially when deep neural network architectures are used. It may be possible that former research on classifying vibroarthrographic signals using this dataset is strongly over fitting on exactly this dataset, besides using cross validation techniques. Second, the dataset only distinguish between abnormal and normal knee conditions but contain data with several different pathologies. The differences between those pathologies might not be covered sufficiently by this small number of samples. Using spectrograms for classification of vibration and sound via convolutional neural networks is a common practice now days. One of the problems with using spectrograms for classification of sound and vibration is the trade-off between frequency and time resolution when using the STFT. This may be overcome, by using multi-scale spectrograms and feeding multiple spectrograms into the network [4] or using scalograms [28], which represent the time-frequency domain of a signal by applying the continuous wavelet transformation instead of a STFT. Further, the usage of convolutional neural networks for feature extraction and classification of spectrograms underlie various problems. First, the axes of a spectrogram (time, frequency) differ from images (x, y pixel position) in its content. A learned filter, which is translation invariant, extract translation invariant features in the image domain. This may work for typical images, as classification should not depend on the exact position of a feature or object (e.g. the position of a person in an image) but is not applicable to spectrograms where the exact position across the frequency axis is crucial, as the information along the y-axis represents the amplitude corresponding to a certain frequency. Although one may argue that the exact position in time is not important, as occurring vibration events represents the same, regardless of the time. In Vibroarthrography this may not hold, as it is imaginable to use the corresponding angle of the knee during examination as an axis. Second, frequencies are not locally grouped in spectrograms, which may complicate the process of finding local features in spectrograms by using convolutional neural networks, as those features in spectrograms are spaced apart. Locally connected convolutional layers may overcome this problem by learning a filter for each pixel position of an image or activation map and ensures that the weights of the kernel remain unshared [30]. This may increase the performance of our convolutional neural network approach in the future.

6 CONCLUSION

In this paper, we presented a novel approach to vibroarthrography by applying convolutional neural networks on vibroarthrographic data transformed into spectrograms, rendering the time series classification task into an image classification problem. The evaluation is carried out on a publicly available dataset and should encourage other researcher to leverage deep neural network architectures for vibroarthrographic data analysis. We proposed a two-layer convolutional neural network for classification without the need for preprocessing in terms of filtering or scaling the accelerometer signal and achieve comparable results with an average F_1 Score of 80% and AUC of 0.87 with data augmentation. With the absence of augmentation we achieve a lower F_1 Score and AUC of 74% and 0.81 respectively. Those results are still lower than previous reported performance measures by other researcher on this dataset. Nevertheless, we suspect that deep learning techniques may be efficiently applied on the task of vibroarthrographic signal classification, if more data is available in the future.

ACKNOWLEDGMENTS

This work receives funding from the German Federal Ministry for Economic Affairs and Energy by ZIM-16KN04913, related to the project MOREBA.

REFERENCES

[1] Chu, M., Gradisar, I., and Mostardi, R. (1978). A noninvasive electroacoustical evaluation technique of cartilage damage in pathological knee joint. *Electronics Letters*, 16(4):437–442.
[2] Chu, M. L., Gradisar, I. A., Railey, M. R., and Bowling, G. F. (1976). Detection of knee joint diseases using acoustical pattern recognition technique. *Journal of Biomechanics*, 9(3):111–114.
[3] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*.
[4] Espi, M., Fujimoto, M., Kinoshita, K., and Nakatani, T. (2015). Exploiting spectro-temporal locality in deep learning based acoustic event detection. *EURASIP Journal on Audio, Speech, and Music Processing*, 2015:1.
[5] Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., and Muller, P. (2018). Deep learning for time series classification: a review. *CoRR*, abs/1809.04356.
[6] Frank, C. B., Rangayyan, R. M., and Bell, G. D. (1990). Analysis of knee joint sound signals for non-invasive diagnosis of cartilage pathology. *IEEE Engineering in Medicine and Biology Magazine*, 9(1):65–68.
[7] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*.
[8] He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034.
[9] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
[10] Kernohan, W. G., Beverland, D. E., McCoy, G. F., Hamilton, A., Watson, P., and Mollan, R. (1990). Vibration arthrometry. a preview. *Acta orthopaedica Scandinavica*, 61(1):70–79.
[11] Kernohan, W. G., Beverland, D. E., McCoy, G. F., Shaw, S. N., Wallace, R. G., McCullagh, G. C., and Mollan, R. A. (1986). The diagnostic potential of vibration arthrography. *Clinical orthopaedics and related research*, (210):106–112.
[12] Kim, K. S., Seo, J. H., Kang, J. U., and Song, C. G. (2009). An enhanced algorithm for knee joint sound classification using feature extraction based on time-frequency analysis. *Computer methods and programs in biomedicine*, 94(2):198–206.
[13] Krishnan, S., Rangayyan, R., Bell, G., Frank, C., and Ladly, K. (1997). Adaptive filtering, modelling and classification of knee joint vibroarthrographic signals for non-invasive diagnosis of articular cartilage pathology. *Medical and Biological Engineering and Computing*, 35(6):677–684.
[14] Krishnan, S., Rangayyan, R. M., Bell, G. D., and Frank, C. B. (2000). Adaptive time-frequency analysis of knee joint vibroarthrographic signals for noninvasive screening of articular cartilage pathology. *IEEE Transactions on Biomedical Engineering*, 47(6):773–783.
[15] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.
[16] Loshchilov, I. and Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.
[17] Maussavi, Z. M., Rangayyan, R. M., Bell, G. D., Frank, C. B., and Ladly, K. O. (1996). Screening of vibroarthrographic signals via adaptive segmentation and linear prediction modeling. *IEEE transactions on biomedical engineering*, 43(1):15.
[18] McCoy, G. F., McCrea, J. D., Beverland, D. E., Kernohan, W. G., and Mollan, R. (1987). Vibration arthrography as a diagnostic aid in diseases of the knee. a preliminary report. *The Journal of bone and joint surgery. British volume*, 69(2):288–293.
[19] Murphy, L., Cisternas, M., Pasta, D., Helmick, C., and Yelin, E. (2017). Medical expenditures and earnings losses among us adults with arthritis in 2013. *Arthritis Care Res.*
[20] Nalband, S., Prince, A., and Agrawal, A. (2017). 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.
[21] Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076.
[22] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. (2017). Automatic differentiation in pytorch.
[23] Rangayyan, R. M., Krishnan, S., Bell, G. D., Frank, C. B., and Ladly, K. O. (1997). Parametric representation and screening of knee joint vibroarthrographic signals. *IEEE Transactions on Biomedical Engineering*, 44(11):1068–1074.
[24] Rangayyan, R. M., Oloumi, F., Wu, Y., and Cai, S. (2013). Fractal analysis of knee-joint vibroarthrographic signals via power spectral analysis. *Biomedical Signal Processing and Control*, 8(1):23–29.
[25] Rangayyan, R. M. and Wu, Y. (2007). Analysis of knee-joint vibroarthrographic signals using statistical measures. In *Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)*, pages 377–382. IEEE.
[26] Rangayyan, R. M. and Wu, Y. (2008). Screening of knee-joint vibroarthrographic signals using statistical parameters and radial basis functions. *Medical & biological engineering & computing*, 46(3):223–232.
[27] Reddy, N. P., Rothschild, B. M., Mandal, M., Gupta, V., and Suryanarayanan, S. (1995). Noninvasive acceleration measurements to characterize knee arthritis and chondromalacia. *Annals of Biomedical Engineering*, 23(1):78–84.
[28] Ren, Z., Qian, K., Wang, Y., Zhang, Z., Pandit, V., Baird, A., and Schuller, B. (2018). Deep scalogram representations for acoustic scene classification. *IEEE/CAA Journal of Automatica Sinica*, 5(3):662–669.
[29] Sharma, M. and Acharya, U. R. (2018). Analysis of knee-joint vibroarthographic signals using bandwidth-duration localized three-channel filter bank. *Computers & Electrical Engineering*, 72:191–202.
[30] Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708.
[31] Tavathia, S., Rangayyan, R. M., Frank, C. B., Bell, G. D., Ladly, K. O., and Zhang, Y. . (1992). Analysis of knee vibration signals using linear prediction. *IEEE Transactions on Biomedical Engineering*, 39(9):959–970.
[32] Wu, Y., Chen, P., Luo, X., Huang, H., Liao, L., Yao, Y., Wu, M., and Rangayyan, R. (2016). Quantification of knee vibroarthrographic signal irregularity associated with patellofemoral joint cartilage pathology based on entropy and envelope amplitude measures. *Computer Methods and Programs in Biomedicine*, 130:1–12.
[33] Wu, Y. and Krishnan, S. (2011). Combining least-squares support vector machines for classification of biomedical signals: a case study with knee-joint vibroarthrographic signals. *Journal of Experimental & Theoretical Artificial Intelligence*, 23(1):63–77.
[34] Wyse, L. (2017). Audio spectrogram representations for processing with convolutional neural networks. In *Proceedings of the First International Conference on Deep Learning and Music, Anchorage, US, May, 2017., pp. 37-41*, pages 37–41.