VPS VIBRATIONS IN PHYSICAL SYSTEMS

DOI: 10.21008/j.0860-6897.2021.1.08
Vibrations in Physical Systems, 32(1):2021108, 2021

Voice Pathology Assessment Using X-Vectors Approach

Katarzyna KOTARBA[1], Michał KOTARBA[2]

Corresponding author: Katarzyna KOTARBA, urbaniec@agh.edu.pl

[1] AGH University of Science and Technology, Department of Mechanics and Vibroacoustics, al. Mickiewicza 30, 30-059 Kraków
[2] TECHMO Voice Technologies, ul. Torfowa 1/5, 30-384 Kraków

Abstract Voice pathology assessment using sustained vowels has proven to be effective and reliable. However, only a few studies regarding detection of pathological speech based on continuous speech are available. In this study we evaluate the usefulness of various regression models trained on continuous speech recordings from Saarbruecken Voice Database in the detection of voice pathologies. The recordings were used for extraction of speaker embeddings called x-vectors based on mel-frequency cepstral coefficients and gammatone frequency cepstral coefficients. Since the dataset used in this study is imbalanced, various over- and undersampling techniques were applied to the training set to ensure robustness of models' decision boundaries. The models were trained on both imbalanced and resampled training sets using 5-fold cross-validation. The best results were obtained for Multi Layer Perceptron trained on GFCC-based x-vectors, achieving accuracy of 0.8184, F1-score of 0.8212, and ROC AUC score of 0.8810 for the testing set.

Keywords: x-vectors, speaker embeddings, voice pathology, MFCC, GFCC

1. Introduction

Computer aided voice pathology detection is a promising tool for physicians and speech therapists. It has become an even more important field of study now, when the need for telemedicine technologies increased significantly due to the epidemiological situation and reduced access to traditional healthcare. Developing an efficient and stable algorithm for laryngeal pathology assessment based on the speech signal can not only help lessen the transmission of many infectious diseases, but also maintain the continuum of medical care by minimizing the exposure of medical staff [1]. What is more, it can help improve otolaryngology access in rural settings and significantly reduce the expenses incurred by both patient and peripheral medical centers [2].

Many voice pathology detection algorithms incorporate features extracted from sustained vowels. Sustained vowels are used during standard medical examination of vocal folds and are therefore a natural choice for development of voice pathology assessment methods. This approach proved to be extremely effective. Hemmerling et al. [3] obtained the accuracy rate of 100% in the classification of healthy and pathological voice using vowel */a/* and random forest classifier. Fang et al. [4] used deep neural network (DNN) and reached the accuracy rate of 99.14% based on 13-dimensional mel-frequency cepstral coefficients (MFCC) features extracted from recordings of vowel */a/*. Al-Nasheri et al. [5] investigated 22 acoustical parameters extracted from Multidimensional Voice Program [6] and their applicability in the pathological speech detection. The parameters used can be divided into several types: frequency related, intensity related, noise related and tremor related. The authors used a statistical approach, performing a *t*-test to verify if the mean values of the two classes (healthy and pathological voice) are significantly different. This method yielded the accuracy rate of 99.68%.

Despite its effectiveness, medical diagnosis based on sustained vowels has its drawbacks – it cannot be performed based on the conversation between the physician and the patient. Some works addressed this issue and proposed algorithms based on continuous speech. Vasilakis and Stylianou [7] performed a discrimination of pathological and healthy speech using short-term jitter estimations and reached the score of 87.8% in terms of area under ROC curve. Cordeiro et al. [8] yielded the accuracy rate of 74% in the detection of unilateral paralysis of vocal folds and vocal fold edema using 12 MFCC features and Gaussian

Mixture Models. Guedes et al. [9] attempted to diagnose three diseases: dysphonia, vocal cords paralysis and chronic laryngitis using transfer learning approach and neural network classifiers. The best results were obtained for the detection of vocal cords paralysis – F1-score reached 80%. However, the multinomial classification of all three diseases resulted in the F1-score of only 40%.

Above-mentioned works used open-access speech corpora in German and English. Nonetheless, studies regarding diagnosis of vocal tract pathologies in Polish speaking patients are also available, e.g. the work published by Wszołek et al. [10] or the paper published by Engel et al. [11].

The main objective of this study is to evaluate the usefulness of various regression models in the detection of voice pathologies based on continuous speech. It is worth noting that even in text-dependent scenario, continuous speech utterances may vary in length significantly. As most of the classification algorithms require fixed size of the input data, the signals which are too long are often clipped – it allows to overcome the problem with the input size, but leads to information loss. Another solution is zero-padding of the signals that are too short – it preserves all the information, but introduces redundancy. In this study, the approach based on speaker embeddings called x-vectors [12] is used. The main advantage of the x-vectors is the fact that they can be extracted from the signals differing in length, while the obtained embeddings are fixed sized and can therefore be used as the input data for any classification algorithm. What is more, the x-vectors approach has proven to be effective in other medical applications, i.e. Parkinson's disease detection [13, 14]. Importantly, the studies regarding usage of x-vectors in medical diagnosis used only MFCC features. Moreover, to the best of our knowledge, no studies available in literature carried out the comparative analysis of results obtained using different regressive models or attempted to establish the most suitable model for voice pathology detection.

The rest of this paper is organized as follows. In section 2 the speech corpus, data preprocessing, feature extraction and models' training are described. In section 3 obtained results are presented and discussed. In section 4 the work is concluded and future perspectives are discussed.

2. Proposed method

2.1. Saarbruecken Voice Database

German speech corpus Saarbruecken Voice Database (SVD) [15] was used to train and evaluate models detecting pathological speech. The SVD contains sustained vowels and the phrase *Guten Morgen, wie geht es Ihnen?* uttered by 628 healthy people and 1269 people suffering from different diseases, e.g. various types of dysphonia, chronic laryngitis, vocal cords paralysis. In this study we do not differentiate the diseases and only binary classification aimed to detect any pathology is performed.

The SVD subset consisting of uttered phrase is divided into two stratified sets: training set consisting of 80% of the data and testing set consisting of 20% of the data. The details of the sets are listed in Tab. 1.

Tab. 1. Details of the two subsets of the speech corpus used for models' training and evaluation: training set and testing set.

	Number of utterances	
Set	Sick	Healthy
Training	1010	507
Testing	259	121

2.2. Data preprocessing and feature extraction

Firstly, all the recordings were downsampled from 50 kHz to 16 kHz. They were used to extract so-called x-vectors – fixed-dimensional speaker embeddings designed for text independent speaker recognition [12, 16]. The x-vectors are meant to capture speaker characteristics over the entire utterance and are suitable for analysis of utterances differing in length.

Sampling frequency of 16 kHz was chosen for two reasons: 1. it is the sampling frequency of the Voxceleb2 corpus used for training of x-vector extractors used in this study and SVD database was down sampled in order to match this sampling rate 2. even though it is recommended to use higher sampling

frequencies for speech analysis [17], some of the medical speech corpora (i.e. LANNA speech corpus containing recordings of children suffering from specific language impairment [18]) are recorded with a sampling rate of only 16 kHz and the proposed method is meant to be suitable for their analysis as well.

Recently, a study on the effect of feature sets on speaker identification was presented by Farooq et al. [19]. According to the authors, x-vectors based on gammatone frequency cepstral coefficients (GFCC) [20] provide better results than standard x-vectors based on mel-frequency cepstral coefficients (MFCC) (see Fig. 1). We decided to verify if they would also be more accurate for voice pathology detection task.

The embeddings used in this study were extracted using Kaldi speech recognition toolkit [21] – the details of the extraction procedure depending on the feature set used are described below.

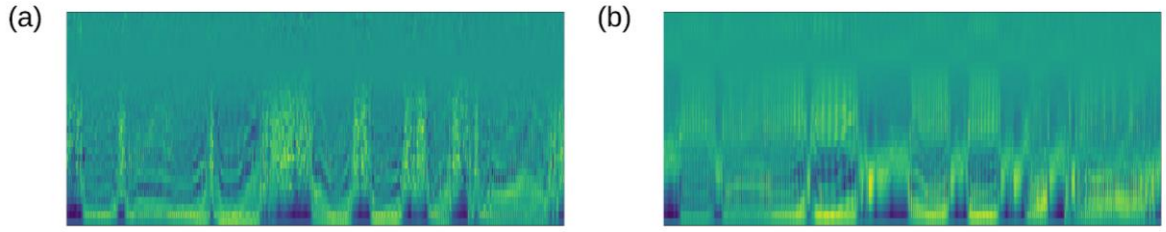

Fig. 1. Sample features extracted from the phrase *Guten Morgen, wie geht es Ihnen?* uttered by the same person: a) MFCC features, b) GFCC features.

2.2.1. MFCC-based x-vectors

The standard MFCC-based x-vectors were extracted using a pretrained model provided by Kumar et al. [22]. Firsty, the 30-dimensional MFCC features were extracted from each signal using a frame width of 25 ms and an overlap of 15 ms, following the recipe provided by the authors. Then the default energy-based Kaldi voice activity detection (VAD) algorithm was performed and the frames containing non-speech were removed. The obtained features were fed to the pretrained model and 512-dimensional x-vectors were extracted.

The analysis performed by Chaudhari and Dhonde [23] showed that 30-dimensional MFCC features provide better results in speaker recognition task than smaller number of features, while maintaining the acceptable computational time. What is more, 30-dimensional MFCC features were used by Kumar et al., whose pretrained model is used in this study. We decided to match the dimensionality of MFCC and GFCC feature sets to ensure that any differences in performance metrics values would not be caused simply by the different feature size but only by the filterbank (i.e. mel and gammatone) used.

2.2.2. GFCC-based x-vectors

The VoxCeleb2 corpus [24] containing over a million utterances from approximately 7300 speakers was used to train a deep neural network (DNN) performing a speaker identification task. The architecture of the DNN is based on the pretrained model by Kumar et al. – for details, see Ref. [22].

Firstly, the 30-dimensional gammatone frequency representation (GTF) of signals and the cepstral coefficients were calculated using a frame width of 20 ms and overlap of 10 ms. Then autoregressive moving average (ARMA) filtering was applied and long-term signal variability (LTSV) [25] was calculated. Finally, VAD was performed based on voicing and LTSV probability. All the procedures were performed using Featxtra Toolbox for Kaldi [26].

The extracted GFCC features were used to train the abovementioned DNN implemented in PyTorch [27]. The DNN was then used for 512-dimensional x-vectors extraction from signals with a length in the range of 32-500 frames.

2.3. Training set resampling

Most of the machine learning algorithms perform poorly, when the dataset used for their training is highly imbalanced – the classifiers tend to be biased towards the majority class, as it still allows them to obtain relatively high accuracy rates [28]. To overcome this problem, data resampling techniques are used,

providing the model with a more balanced dataset. In some cases these techniques may result in obtaining more robust decision boundaries of the model and, hence, better results of minority class classification.

Seven oversampling techniques:

- SMOTE [29],
- SMOTEN [29],
- SVM-SMOTE [30],
- Borderline SMOTE [31],
- KMeans-SMOTE [32],
- ADASYN [33],
- random minority oversampling with replacement,

and ten undersampling techniques:

- One-Sided Selection [34],
- Neighbourhood Cleaning Rule [35],
- Condensed Nearest Neighbour [36],
- Edited Nearest Neighbours [37],
- Repeated Edited Nearest Neighbours [38],
- NearMiss [39],
- AllKNN [40],
- extraction of majority-minority Tomek links [40],
- undersampling with Cluster Centroids [41],
- random majority undersampling with replacement

implemented in Python's imbalanced-learn package [42] were applied to the training sets. The testing set was not resampled to ensure reliability of models evaluation results.

2.4. Models training

Sixteen different regression models (listed in Tab. 2) implemented in Python's scikit-learn library [43] were trained on MFCC-based and GFCC-based embeddings using a 5-fold cross-validation algorithm. Feature selection technique based on ANOVA F-value, namely scikit-learn's *SelectKBest*, was applied to reduce feature dimensionality, leading to reduction of training time and models' overfitting. Models' hyperparameters and number of features were optimized using Optuna framework [44]. The best model (i.e. the model yielding the highest ROC AUC score on the testing set) was additionally trained and optimized on resampled training sets.

3. Experimental results

Pathological speech detection is a binary task. Since the SVD contains only one continuous speech utterance per speaker, the two subsets used for models training and evaluation do not overlap in terms of the speakers involved and the classification may be considered to be speaker independent.

The following metrics were used to evaluate proposed models' performance [45, 46]:

$$Recall = \frac{TP}{TP+FN} \cdot 100\%, \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \cdot 100\%, \tag{2}$$

$$F1\ score = \frac{2 \cdot Precision \cdot Recall}{Precision+Recall} \cdot 100\%, \tag{3}$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \cdot 100\%, \tag{4}$$

$$ROCAUC = \int_0^1 ROC(t)dt, \tag{5}$$

where TN stands for true negative, TP stands for true positive, FN stands for false negative, FP stands for false positive and ROC stands for receiver operating characteristic curve constructed as a plot of recall versus false positive rate. The area under a ROC curve (ROC AUC) can be interpreted as the probability that

the binary classifier will yield a higher value for a randomly chosen positive instance than for a randomly chosen negative instance [45].

The ROC AUC score yielded for the testing set was used to choose the best model. For the MFCC-based x-vectors, the best ROC AUC was obtained by the NuSVR classifier (see Tab. 2). On the other hand, the best results for the GFCC-based x-vectors were obtained by Multi Layer Perceptron (MLP). Moreover, the GFCC-based models proved to be superior to the models based on MFCC features, yielding higher ROC AUC scores in almost all cases while using fewer features: MFCC-based NuSVR reached reported results using 511 features, while GFCC-based MLP used only 480 features. GFCC-based MLP was therefore chosen for further evaluation using resampled training sets.

Tab. 2. ROC AUC scores yielded for testing set by evaluated regression models. The best results obtained by MFCC-based and GFCC-based models are **boldface**.

	ROC AUC	
Regression model	MFCC-based	GFCC-based
Logistic	0.6652	0.7203
Ridge	0.8242	0.8627
SGD	0.6308	0.8434
Elastic Net	0.8231	0.7050
ARD	0.8029	0.8762
Bayesian Ridge	0.8457	0.8749
Huber	0.7596	0.8493
Radius Neighbors	0.6780	0.8084
Poisson	0.8402	0.8082
AdaBoost	0.7751	0.8382
Extra Trees	0.7644	0.8294
Random Forest	0.7692	0.8321
LDA	0.7261	0.7025
KNeighbors	0.8075	0.8590
MLP	0.7306	**0.8810**
NuSVR	**0.8490**	0.8611

Tab. 3. Classification performance metrics obtained for validation folds and testing set by GFCC-based MLP. The standard deviation of each metric yielded for validation folds is reported in the brackets.

	Original data		Data oversampled using SVM-SMOTE technique	Data undersampled using TomekLinks technique
Metric	Validation folds	Testing set	Testing set	Testing set
Accuracy	0.8062 (±0.0193)	0.8184	0.8105	0.8105
Precision	0.7068 (±0.0441)	0.8282	0.8240	0.8240
Recall	0.7258 (±0.0617)	0.8184	0.8105	0.8105
F1-score	0.7139 (±0.0333)	0.8212	0.8140	0.8140
ROC AUC	0.8713 (±0.0193)	0.8810	0.8807	0.8639

The values of performance metrics obtained for testing sets by MLPs trained on original (non-resampled), oversampled, and undersampled GFCC-based training sets are reported in Tab. 3. Additionally, the mean values of the results yielded for validation folds by the model trained on the original training set are provided. The thresholds used during the binary classification process were determined based on the ROC curve analysis. To keep the results clear, only the values obtained using the best over- and undersampling techniques, namely SVM-SMOTE and extraction of majority-minority Tomek links, are provided.

The standard deviation of performance metrics values obtained for validation folds are small, implying model's stability. The small difference between ROC AUC yielded for validation folds and testing set shows model's good generalization abilities. The difference between binary metrics (i.e. precision, recall, F1-score)

obtained for validation folds and testing set might be attributed to smaller amount of data in training folds than in the set used for training of the final model.

Applying various resampling techniques did not improve the performance of the model. Both undersampling and oversampling of the training set led to a small decrease of the performance for testing set, suggesting that models trained on the non-resampled data may be a better choice if they were to be used in real-life scenario.

Unfortunately, other studies regarding pathological speech detection based on continuous speech provide only the results obtained during the cross-validation process. According to Tabe-Bordbar et al. [47] and Rao et al. [48], the classifier should not be evaluated only based on cross-validation results, as this approach does not provide the full insight into the classifier's generalization abilities and the results may be misleading. Moreover, the studies providing classification metrics values obtained for isolated testing set used sustained vowels instead of continuous speech and therefore could not be directly compared with the method proposed in this paper.

4. Conclusions

In this study an x-vector approach to speaker-independent voice pathology assessment based on continuous speech is proposed. X-vectors based on MFCC and GFCC features were extracted using pretrained model and DNN trained to perform speaker verification task, respectively. Various regression models and resampling techniques were tested to ensure choosing the most suitable classifier. The results are promising – mean accuracy of 0.8062, mean F1-score of 0.7139, and mean ROC AUC score of 0.8713 on validation folds and accuracy of 0.8184, F1-score of 0.8212, and ROC AUC score of 0.8810 on the testing set were obtained by the best model, namely GFCC-based MLP. Even though the results are worse than the best results reported for classifiers trained on sustained vowels, the accuracy of the proposed method yielded for testing set exceeds the highest accuracy rates reported for continuous speech-based classifiers described in literature. What is more, the x-vector approach enables using signals with different lengths, overcoming one of the biggest problems with using continuous speech signals with machine learning algorithms requiring fixed size of the input data. Finally, the study shows superiority of GFCC-based x-vectors over MFCC-based x-vectors in the detection of voice pathology.

The x-vector extractors were trained on Voxceleb2 speech corpus which consists of utterances in English and used for embedding extraction from the SVD corpus which consists of German speech. Despite the differences in languages of both corpora, the obtained results of voice pathology assessment are good. It may suggest that the proposed method is language-independent and should be suitable also for tests in other languages. Nevertheless, both English and German are Germanic languages, so further evaluation using languages from other language groups, e.g. Slavic languages, should be performed.

In the future, separate models for males and females may be trained and evaluated, as this approach has been found to provide better results than the gender-independent approach presented in this study [3, 49]. The proposed method should also be evaluated on other datasets containing continuous speech, e.g. Massachusetts Eye and Ear Infirmary Database (MEEI) [50].

References

1. B. Anthony Jnr. Use of telemedicine and virtual care for remote treatment in response to COVID-19 pandemic. Journal of Medical Systems, 44(7):132, 2020.
2. R. Philips, N. Seim, L. Matrka, B. Locklear et al. Cost savings associated with an outpatient otolaryngology telemedicine clinic. Laryngoscope Investigative Otolaryngology, 4(2):234-240, 2019.
3. D. Hemmerling, A. Skalski, J. Gajda. Voice data mining for laryngeal pathology assessment. Computers in Biology and Medicine, 69:270-276, 2016.
4. S.-H. Fang, Y. Tsao, M.-J. Hsiao, J.-Y. Chen et al. Detection of pathological voice using cepstrum vectors: A deep learning approach. Journal of Voice, 33(5):634–641, 2019.
5. A. Al-Nasheri, G. Muhammad, M. Alsulaiman, Z. Ali et al. An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification. Journal of Voice, 31(1):113.e9-113.e18, 2017.

6. M. Nicastri, G. Chiarella, L. Gallo, M. Catalano et al. Multidimensional voice program (MDVP) and amplitude variation parameters in euphonic adult subjects. Normative study. Acta otorhinolaryngologica Italica: organo ufficiale della Società italiana di otorinolaringologia e chirurgia cervico-facciale, 24(6):337–341, 2004.
7. M. Vasilakis, Y. Stylianou. Voice pathology detection based on short-term jitter estimations in running speech. Folia phoniatrica et logopaedica: official organ of the International Association of Logopedics and Phoniatrics (IALP), 61(3):153–170, 2009.
8. H. Cordeiro, C. Meneses, J. Fonseca. Continuous speech classification systems for voice pathologies identification. IFIP Advances in Information and Communication Technology, 450:217–224, 2015.
9. V. Guedes, F. Teixeira, A. Oliveira, J. Fernandes et al. Transfer learning with audioset to voice pathologies identification in continuous speech. Procedia Computer Science, 164:662–669, 2019.
10. W. Wszołek, A. Izworski, G. Izworski. Signal processing and analysis of pathological speech using artificial intelligence and learning systems methods. Acta Physica Polonica. A, 123(6):995-1000, 2013.
11. Z. W. Engel, M. Kłaczyński, W. Wszołek. A vibroacoustic model of selected human larynx diseases. International Journal of Occupational Safety and Ergonomics, 13(4):367–379, 2007.
12. D. Snyder, D. Garcia-Romero, D. Povey, S. Khudanpur. Deep neural network embeddings for text-independent speaker verification. INTERSPEECH, 999–1003, 2017.
13. L. Jeancolas, D. Petrovska-Delacrétaz, G. Mangone, B.-E. Benkelfat et al. X-vectors: New quantitative biomarkers for early Parkinson's disease detection from speech. Front.Neuroinform, 15, 2021.
14. C. Botelho, F. Teixeira, T. Rolland, A. Abad et al. Pathological speech detection using x-vector embeddings, arXiv, 2003.00864, 2020.
15. W. Barry, M. Pützer. Saarbrücken voice databse. Institute of Phonetics, Univ. of Saarland. [Online]. Available: http://www.stimmdatenbank.coli.uni-saarland.de
16. D. Snyder, D. Garcia-Romero, G. Sell, D. Povey et al. X-vectors: Robust DNN embeddings for speaker recognition. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP):5329–5333, 2018.
17. K. Wu, D. Zhang, G. Lu, Zh. Guo. Influence of sampling rate on voice analysis for assessment of Parkinson's disease. The Journal of the Acoustical Society of America, 144:1416, 2018.
18. P. Grill, J. Tučková. Speech databases of typical children and children with SLI. PLoS ONE, 11(3): e0150365, 2016.
19. M. Farooq, F. Adeeba, S. Hussain. X-vectors based Urdu speaker identification for short utterances. Conference of the Oriental COCOSDA International Committee for the Coordination and Standardisation of Speech Databases and Assessment Techniques, 2019.
20. Y. Shao, Z. Jin, D. Wang, S. Srinivasan. An auditory-based feature for robust speech recognition. IEEE International Conference on Acoustics, Speech and Signal Processing: 4625–4628, 2009.
21. D. Povey, A. Ghoshal, G. Boulianne, L. Burget et al. The Kaldi speech recognition toolkit. IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, 2011.
22. M. Kumar, T. Jin-Park, S. Bishop, C. Lord et al. Designing neural speaker embeddings with meta learning. 2020.
23. A. A. Chaudhari, S. B. Dhonde. Effect of varying MFCC filters for speaker recognition. International Journal of Computer Applications, 128(14), 2015.
24. J. S. Chung, A. Nagrani, A. Zisserman. Voxceleb2: Deep speaker recognition. Proc. Interspeech 2018, 1086-1090, 2018.
25. P. K. Ghosh, A. Tsiartas, S. Narayanan. Robust voice activity detection using long-term signal variability. IEEE Transactions on Audio, Speech, and Language Processing 19(3):600–613, 2011.
26. Featxtra toolbox for Kaldi [Online]. Available: https://github.com/mvansegbroeck-zz/featxtra
27. A. Paszke, S. Gross, F. Massa, A. Lerer et al. Pytorch: An imperative style, high-performance deep learning library," in Advances in Neural Information Processing Systems 32:8024–8035, 2019.
28. B. Krawczyk. Learning from imbalanced data: Open challenges and future directions. Progress in Artificial Intelligence 5, 2016.
29. N. Chawla, K. Bowyer, L. Hall, W. Kegelmeyer,. Smote: Synthetic minority over-sampling technique. J. Artif. Intell. Res. (JAIR), 16:321–357, 2002.

30. H. Nguyen, E. Cooper, K. Kamei. Borderline over-sampling for imbalanced data classification. International Journal of Knowledge Engineering and Soft Data Paradigm, 3:4–21, 2011.
31. H. Han, W.-Y. Wang, B.-H. Mao. Borderline-smote: A new over-sampling method in imbalanced data sets learning. Adv. Intell. Comput., 3644:878–887, 2005.
32. F. Last, G. Douzas, F. Bação. Oversampling for imbalanced learning based on k-means and smote. arXiv, 2017.
33. H. He, Y. Bai, E. Garcia, S. Li, ADASYN: Adaptive synthetic sampling approach for imbalanced learning. Proceedings of the International Joint Conference on Neural Networks,1322 – 1328, 2008.
34. M. Kubat. Addressing the curse of imbalanced training sets: One-sided selection. Fourteenth International Conference on Machine Learning, 2000.
35. J. Laurikkala. Improving identification of difficult small classes by balancing class distribution. Proc. 8th Conf AI Med Eur Artif Intell Med:63–66, 2001.
36. P. Hart. The condensed nearest neighbor rule. IEEE Transactions on Information Theory, 14(3):515–516, 1968.
37. D. L. Wilson. Asymptotic properties of nearest neighbor rules using edited data. IEEE Transactions on Systems, Man, and Cybernetics, SMC-2(3):408–421, 1972.
38. I. Tomek. An experiment with the edited nearest-neighbor rule. IEEE Transactions on Systems, Man, and Cybernetics, SMC-6(6):448–452, 1976.
39. J. Zhang, I. Mani. KNN Approach to Unbalanced Data Distributions: A Case Study Involving Information Extraction. Proceedings of the ICML'2003 Workshop on Learning from Imbalanced Datasets, 2003.
40. I. Tomek. Two modifications of CNN. IEEE Transactions on Systems, Man, and Cybernetics, SMC-6(11): 769–772, 1976.
41. Y.-P. Zhang, L.-N. Zhang, Y.-C. Wang. Cluster-based majority under-sampling approaches for class imbalance learning. IEEE International Conference on Information and Financial Engineering, 2010.
42. G. Lemaître, F. Nogueira, C. K. Aridas. Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning. Journal of Machine Learning Research, 18(17):1–5, 2017.
43. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel et al. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.
44. T. Akiba, S. Sano, T. Yanase, T. Ohta et al. Optuna: A next-generation hyperparameter optimization framework. 25rd International Conference on Knowledge Discovery and Data Mining, 2019.
45. T. Fawcett. An introduction to ROC analysis. Pattern Recognition Letters, 27:861–874, 2006.
46. C. Calì, M. Longobardi. Some mathematical properties of the ROC curve and their applications. Ricerche di Matematica, 64:391-402, 2015.
47. S. Tabe-Bordbar, A. Emad, S. Zhao, S. Sinha. A closer look at cross-validation for assessing the accuracy of gene regulatory networks and models. Scientific Reports, 8:6620, 2018.
48. R. Rao, G. Fung. On the dangers of cross-validation. An experimental evaluation. SIAM International Conference on Data Mining, 588–596, 2008.
49. J. P. Teixeira, P. O. Fernandes, N. Alves. Vocal acoustic analysis – classification of dysphonic voices with artificial neural networks. Procedia Computer Science, 121:19–26, 2017.
50. Massachusetts Eye and Ear Infirmary. Voice disorders database, ver. 1.03. Kay Elemetrics Corp., Lincoln Park, NJ, 1994.