Preview

More than just sound: Harnessing metadata to improve neural network classifiers for medical auscultation

Christian Matek[1,2,*]
[1]Institute of Pathology, University Hospital Erlangen, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany
[2]Institute of Computational Biology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany
*Correspondence: christian.matek@helmholtz-muenchen.de
https://doi.org/10.1016/j.patter.2021.100426

Label-efficient algorithms are of central importance for machine learning applications in many medical fields, where obtaining expert annotations is often expensive and time-consuming. Soni et al. show how contrastive learning can help build classifiers for one of the oldest and most revered methods of clinical medicine: auscultation of heart and lung sounds.

When René Théophile Hyacinthe Laënnec published his book on the use of "mediate auscultation" in 1819,[1] he provided physicians around the world with one of the most powerful tools in the history of medicine: the stethoscope. Together with the instrument, which 200 years on has become one of the emblems of the medical profession, he also provided a classification of normal and abnormal heart and lung sounds and introduced terms like egophony, bronchophony, or vesicular breathing,[2] which to this day feature prominently in the clinical examination practices of physicians around the globe. With tools for easy and high-quality digital sound recording now at hand, these rich and easily obtainable data represent a treasure trove of clinical knowledge and lend themselves to analysis with modern, data-driven algorithms. In this issue of *Patterns*, Soni et al.[3] show how heart and lung sounds can be evaluated efficiently using a contrastive learning scheme.

One of the main challenges in the use of machine learning methods on many types of medical data is that while a large enough number of high-quality data is the key ingredient for training modern algorithms, most conventional, supervised schemes also require classification labels that can be used as ground truth at training time. In contrast to data from the everyday domain (such as the classical problem of telling an image of a cat from that of a dog), where ground-truth labels can be obtained at scale from a large number of annotators, labeling of medical data often requires expert knowledge. Furthermore, even experts can show considerable variability in inter- and intra-rater agreement. For those reasons, compiling datasets large enough to successfully apply supervised machine learning techniques is often prohibitively expensive, a phenomenon noted across many medical data domains, including digital pathology,[4] dermatology,[5] radiology,[6] and ophthalmology.[7]

Motivated by these limitations concerning the availability of labeled data, a number of strategies have been pursued in recent years that allow training on data either lacking labels completely or possessing only weak labels, i.e., labels that carry only part of the relevant descriptive information on a given data point. In the context of medical data, this might mean that some background information on a patient is known, while the ground truth, in this case the expert diagnosis, is not. Within the framework of neural networks, the common aim of these methods is to optimize the structure of hidden representations of the data, thus optimizing feature extraction that forms the basis for downstream tasks such as classification into diagnostic groups. One way of implementing this optimization task is contrastive learning, which considers pairs of training data that share or do not share certain features (so-called positive and negative pairs). By using a contrastive loss function, the representations of positive pairs are mapped to close points in latent space, while dissimilar, negative pairs are being pushed apart. In the context of image classification for example, a common strategy is the SimCLR scheme published by Chen and co-authors in 2020,[8] which uses image augmentation strategies to generate positive pairs of data.

Soni et al. build on this contrastive learning strategy and extend it from the imaging into the sound analysis domain. For selecting positive and negative sample pairs in the sound recording databases, they use clinical patient-related data, such as patient age, sex, and anatomic recording location of the given heart of lung sound recording. Compared to costly expert annotations, this type of information is much more readily accessible also in the medical domain. The authors show that using this clinical meta-information allows for significant improvements in the development of classifiers for heart and lung sounds. Interestingly, for the task of recognizing abnormal lung sounds, the results show that harnessing the age and sex category of patients for negative pair generation leads to the highest improvement in the classification scheme, which lines up with clinical experience showing that those two parameters are correlated with lung disease. Hence, demographic knowledge can be of help in improving recognition of abnormal cases, to an algorithm as much as to a human diagnostician considering these factors when performing auscultation on a patient.

The work by Soni and co-authors shows an innovative way of harnessing patient metadata as a source of weak

labels and using it as a means to improve neural network training in the label-expensive setting characteristic of machine learning tasks on medical data. Consequently, it will be interesting to systematically explore which labels can serve as helpful weak labels for both heart and lung sounds and other data sources, such as imaging or lab data. Will metadata found useful for improving the algorithm be related to known clinical parameters in the respective task? If they are not, might this observation point us toward a "Clever Hans"-type behavior,[9] indicating the algorithm only appears to have learned the relevant parameters of a task? As the distribution of data and metadata may vary between different patient cohorts, generalization of the results of classifier algorithms to other diagnostic settings has to be critically evaluated, in order to avoid biased prediction when using the classifiers in a real-world setting. This includes the effect of confounding factors that might not always be reflected in the labels used at training time, such as background noise level, or chestpiece

used when recording the sound samples. These components may also provide useful metadata for the improvement of algorithms that provide the practicing physician with a diagnostic aid.

After the construction of his first stethoscopes, Dr Laënnec used his invention to study various conditions known at the time, from lung emphysema to liver abscesses and bone fractures. His situation at the time is not dissimilar to the current exploration of a growing number of medical fields by machine learning methods. For these rich novel methods to fulfill their full promise, methods making efficient use of easily obtainable labels as the one presented by Soni et al. will be of central importance.

REFERENCES

1. René Théophile Hyacinthe Laënnec (1819). Traité de l'auscultation médiate ou Traité du diagnostic des maladies des poumons et du coeur, fondé principalement sur ce nouveau moyen d'exploration (Brosson & Chaudé).

2. Roguin, A. (2006). Rene Theophile Hyacinthe Laënnec (1781-1826): the man behind the stethoscope. Clin. Med. Res. *4*, 230–235.

3. Soni, P.N., Shi, S., Sriram, P.R., Ng, A.Y., and Rajpurkar, P. (2021). Contrastive learning of heart and lung sounds for label-efficient diagnosis. Patterns *3*, 100400.

4. Tizhoosh, H.R., and Pantanowitz, L. (2018). Artificial Intelligence and Digital Pathology: Challenges and Opportunities. J. Pathol. Inform. *9*, 38.

5. Chan, S., Reddy, V., Myers, B., Thibodeaux, Q., Brownstone, N., and Liao, W. (2020). Machine Learning in Dermatology: Current Applications, Opportunities, and Limitations. Dermatol. Ther. (Heidelb.) *10*, 365–386.

6. Willemink, M.J., Koszek, W.A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., Folio, L.R., Summers, R.M., Rubin, D.L., and Lungren, M.P. (2020). Preparing Medical Imaging Data for Machine Learning. Radiology *295*, 4–15.

7. Dubis, A.M., Arikan, M., Sallo, F., Montesel, A., Hagag, A.H., Ahmed, H.M., Book, M., Faatz, H., Cicinelli, M., Ongun, S., et al. (2021). Democratizing Deep Learning Research Through Large Publicly Available Datasets and Tools. Invest. Ophthalmol. Vis. Sci. *62*, 1809.

8. Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations. H.D. Iii and A. Singh, eds. *119*, 1597–1607.

9. Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., and Müller, K.R. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. Nat. Commun. *10*, 1096.