SPIRA-BM: Biomarkers for Respiratory Conditions by Audio Analysis via Artificial Intelligence

Marcelo Finger[1], Larissa C. Berti[2], Elisa Y. Nakagawa[3], Celso R. F. de Carvalho[4], Beatriz R. de Medeiros[5], Flaviane R. F. Svartman[5], Marcelo G. Queiroz[1], Arnaldo Cândido Jr[6], Marcelo M. Gauy[7], Murilo G. Gazzola[8], Marcus V. M. Martins [5], Jaqueline Scholz[9], Sara Ziotti[9]

[1]Departamento de Ciência da Computação – Instituto de Matemática e Estatística
Universidade de São Paulo (USP) – São Paulo – SP – Brazil
mfinger@ime.usp.br -- corresponding author

[2]Departamento de Fonoaudiologia
Universidade Estadual Paulista (UNESP)/FFC/Marília – SP – Brazil

[3]Departamento de Sistemas de Computação – ICMC
Universidade de São Paulo (USP)

[4]Departamento de Fisioterapia, Fonoaudiologia e Terapia Ocupacional
Faculdade de Medicina da Universidade de São Paulo – SP – Brazil

[5]Departamento de Letras Clássicas e Vernáculas – FFLCH
Universidade de São Paulo (USP)

[6]Departamento de Ciência da Computação e Estatística – IBILCE
Universidade Estadual Paulista (UNESP)/São José do Rio Preto – SP – Brazil

[7]Departamento de Ciências e Tecnologia — ICT
Universidade Estadual Paulista (UNESP)/Itapeva – SP – Brazil

[8]Faculdade de Computação e Informática
Universidade Presbiteriana Mackenzie – SP – Brazil

[9]InCor/Hospital das Clínvaicas
Faculdade de Medicina da Universidade de São Paulo – SP – Brazil

Abstract. *This paper presents an ongoing project that aims to investigate audio biomarkers for respiratory conditions such as Respiratory Insufficiency, severe asthma, and smoking. The SPIRA-BM project is currently developing inexpensive detectors for these biomarkers, using audio collection and processing on mobile computing devices, and using machine learning, artificial intelligence, and signal analysis techniques.*

Resumo. *Este artigo apresenta um projeto em andamento com o objetivo de investigar biomarcadores de áudio para condições respiratórias, como insuficiência respiratória, asma severa e tabagismo. O projeto SPIRA-BM está desenvolvendo detectores baratos para estes biomarcadores, utilizando de coleta e processamento de áudio em dispositivos de computação móvel e lançando mão de técnicas de aprendizado automático, inteligência artificial e análise de sinais.*

1. Introduction

During the COVID-19 pandemic, a multidisciplinary group came together to present a scientific contribution to face the pandemic. The SPIRA Study [Aluisio et al. 2022] investigated the possibility of detecting Respiratory Insufficiency through audio analysis, from the perspective of patient triage. The focus of the project was the detection of respiratory insufficiency (not the detection of COVID-19), seeking to select patients who would require hospitalization, with data collected by cell phones in the COVID-19 wards of several hospitals in São Paulo. The main result of that study was to demonstrate the feasibility of detecting respiratory insufficiency with a high degree of accuracy, 96.5% [Casanova et al. 2021b, Gauy and Finger 2021a]. It also showed that multidisciplinary efforts are crucial and need to focus on several tasks: automatic data analysis (Big Data), contrasted with classic analysis (small data) in line with the development of software for cell phones (capture and collection) and back-end (signal processing and neural networks).

The current SPIRA-BM project is based on the results obtained in the early SPIRA study and the team formed to study and develop low-cost biomarkers related to the respiratory system from the processing of voice and speech audio on mobile devices, using artificial intelligence and machine learning. The project goes beyond COVID-19 concerns, addressing biomarkers for respiratory insufficiency due to various causes, and no longer just those generated by COVID-19, as well as seeking to develop biomarkers related to severe asthma and indicators associated with smoking.

According to [Biomarkers Definitions Working Group 2001], a *biomarker* is a property that is objectively measured and evaluated as an indicator of biological processes or responses to therapeutic interventions. In this context, by *development of a biomarker* we refer to the development of methods to measure characteristics related to the respiratory tract from voice and speech signals. In the context of audio processing, we separate *voice* properties, which refer to the analysis of time series of signals from human voice emissions, from *speech* properties, which are related to the analysis of human language. These audio measurements are developed both by artificial intelligence methods and by methods associated with the areas of phonetics, phonology, and speech therapy.

This set of low-cost biomarkers relies on the development of mobile computing techniques on mobile devices, both in the data collection phase and in the practical use of the biomarker. Mobile devices will be used both in the data collection phase during the study and in the health assessment phase of patients in the field.

1.1. Project Goals

The SPIRA-BM project aims to develop theoretical, technical, and technological knowledge about biomarkers of respiratory conditions, whose application is practical, inexpensive, and viable for the Brazilian population. These goals shall be achieved by interdisciplinary interaction between the fields of Computer Science and Artificial Intelligence, Medical Sciences, Linguistics, Speech Therapy and Analysis, and Physical Therapy.

2. A Brief Literature Review

Although there is a large literature on the applications of artificial intelligence in healthcare, the detection of respiratory insufficiency through audio has so far been restricted

to the work of the SPIRA project [Aluisio et al. 2022], with initial accuracy results of around 91% [Casanova et al. 2021b] using a convolutional neural network (CNN); later [Gauy and Finger 2021a] achieved 96.5% accuracy using a pretrained Transformer network. Recent preliminary results indicate that it is possible to distinguish different causes of respiratory insufficiency by audio analysis [Gauy et al. 2023].

Instead of exploring respiratory insufficiency, several groups have sought to diagnose COVID-19 through audio recordings of coughing, breathing, and speech [Imran et al. 2020, Brown et al. 2020, Han et al. 2022], as part of efforts during the pandemic. A study by [Coppock et al. 2022] indicates that to diagnose COVID, voice techniques obtain similar accuracy to that of a simple list of symptoms, indicating that the potential use of these techniques for respiratory insufficiency may bring more benefits than the detection of SARS-CoV-2 contamination. The works of [Wynants et al. 2020] and [Roberts et al. 2021] show that methodological concerns with avoiding biases in the collection and elaboration of datasets is fundamental to obtain reliable results using machine learning, a concern present in the SPIRA project and in the current research methodology.

It should be noted that, in the first international competition for the detection of COVID-19 through coughing, the SPIRA group was awarded the first place [Casanova et al. 2021a]. This work was recently extended with additional data and other techniques, reaching 87% accuracy [Atmaja et al. 2025], apparently leading COVID-19 voice diagnostic to a new level.

The original view aiming at treating voice and speech as a biomarker predates the pandemic and was developed within the Speech and Audio Processing group at IST, Lisbon [Botelho et al. 2019, Trancoso et al. 2019, Nevler et al. 2019]. It is in this scope that the research lines of this project intend to fit in.

3. Methodology

The SPIRA-BM Project goal to develop low-cost biomarkers based on audio analysis, covers three major research lines of research in the respiratory systems, each with a *computational challenge* and a *respiratory problem* to be addressed:

Classification: Development and validation of tools for detecting ***respiratory insufficiency*** in general.

Estimation: Assessment of indicators of ***smoking activity*** by voice, aiming at their use in health programs aimed at tobacco smoking cessation.

Event prediction: Development of a tool to predict ***asthma exacerbation in severe chronic patients.***

The data collected in each of the above lines can also allow processing in other activities, such as estimating O_2 saturation in the blood from respiratory insufficiency data, and investigating the possibility of audio classification of asthma patients with symptoms under/out of control.

For each research line above, previous experience showed that the process of collecting, analyzing and processing audio data requires the focus on several complementary approaches, incorporated into the project methodology as four distinct activity axes:

a. Data collection and creation of a database and audio excerpts, with data collected from patients and controls for each research line. Data collection is currently

underway in the following institutions: Hospital das Clínicas, São Paulo; InCor; Hospital do Servidor Público Estadual, São Paulo; Beneficência Portuguesa; Santa Casa de Marília; Hospital da Unimar.

b. Software development, both for data collection and processing on mobile devices, as well as for audio signal processing and machine learning activities on servers on fixed platforms.
c. Audio analysis – Voice and speech signal analysis techniques, usually with *small data* and *white box*, investigating various aspects and properties of audio signals that can serve as indicators of the health status of the emitters.
d. Machine learning – Artificial intelligence techniques, usually with *big data* and *black box*, aiming to develop machine learning programs with different methods and neural network architectures, both in study and in the tool validation phases.

The objectives set out above suggest a matrix organization of the project activities, with one entry per line of research, and another entry per approach. The project matrix organization is illustrated in Figure 1, where we can see the three research lines and the activity axes that cross all those lines.

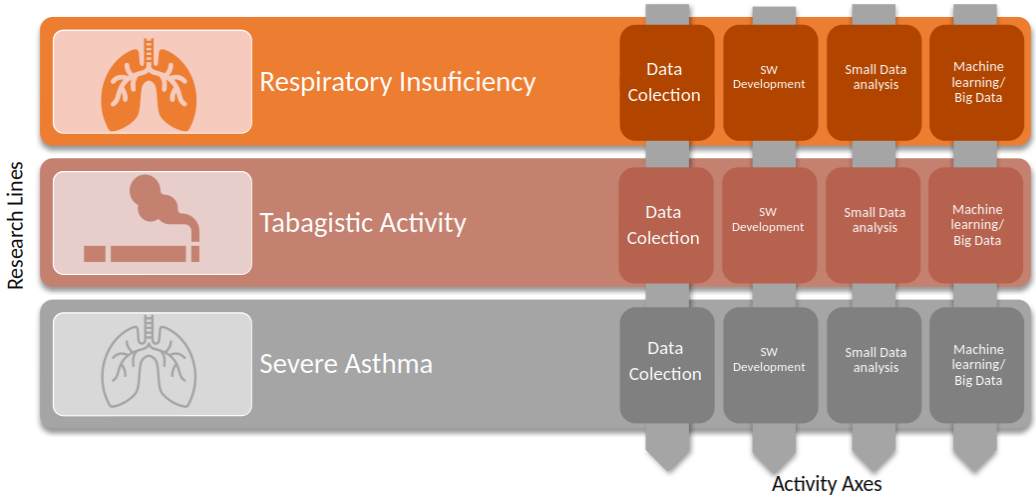

Figure 1. SPIRA-BM Project Matrix

Thus, the respiratory insufficiency biomarker will deal with the classification of audios according to the presence of small data and big data audio patterns, to be studied in the project. Similarly, the smoking biomarker will estimate the value of COex (exhaled carbon monoxide). And the severe asthma biomarker will predict the probability of an exacerbation occurring in the next two days.

It is important to note that the computational problems are not restricted to a single biomarker. For example, audio recordings of respiratory failure will be analyzed to estimate the patient's blood oxygen saturation. Audio recordings of severe asthma patients will be used to classify the asthma patient's voice as indicating symptoms under or out of control, etc. The voice of smoking patients can allow us to classify the patient as a non-smoker, a light smoker, or a heavy smoker, which is sometimes called the patient's tabagistic load. This shows that, with the data collected, we will have a wealth of possible analyses to perform, allowing for the study of very detailed biomarkers.

Detailing the project's methodology, the research structure for each of the three types of biomarkers follows a similar pattern, as illustrated in Figure 2. There are two phases in the development of biomarker detection techniques, which is common in all three projects. The first of these, illustrated in the upper part of Figure 2, involves data collection, preprocessing, and generation of a dataset, which can be used for both machine learning (*Big Data*) illustrated in the figure, and for audio analysis (*Small Data*), which is absent from the illustration. In the first case, a learning model is generated that will be used in the next phase, which applies the model trained, as illustrated in the lower part of the figure. In the application phase, patient audio is also captured using a mobile device, which will be analyzed by the trained model and an assessment of the biomarker is emitted; the latter task can be performed by remote processing in the cloud or by local processing on the mobile device itself.

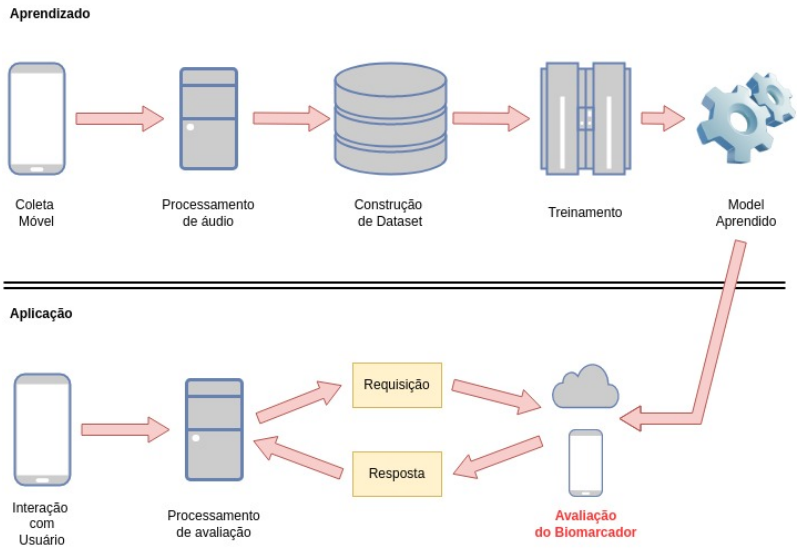

Figure 2. SPIRA-BM Development of Biomarkers

3.1. Ethical considerations.

We have submitted the three lines of research to the Ethics Committee of the Hospital das Clínicas de São Paulo and Plataforma Brasil. We already have two approved authorizations to collect audio from patients hospitalized for respiratory insufficiency and severe asthma, with CAAE codes: 30918120.0.0000.0068 and 85388624.0.0000.0068.

Due to ethical considerations, data is totally anonymized, to the point that we do not collect any personal identification data. Instead, we use the hospital's ID number, and we do not have access to the hospital's registry, which allows us to identify patients in the study but not to relate that identification to any individual.

4. Ongoing Work

Several project activities are ongoing at the moment, among which:

- Audio data collection for: the data collection for severe asthma has already started and is well advances, with more than 200 patients collected; data collection at this point aims at developing audio biomarkers for patients with symptoms under/out

of control, and the detection of vocal cord dysfunction. Data collection for respiratory insufficiency has also started, aiming at developing a biomarker that can both detect its occurrence and propose a most-likely cause, such as lung infection, heart condition or COPD. In the research line for cessation of smoking, the audio data collection is about to start.

- Audio analysis (small data) for voice and speech characteristics in patients with severe asthma. Phonological analysis of patient speech impaired due to respiratory problems is under analysis.
- A big data study on the possibility of inferring blood oxygen saturation (SpO2) from audio analysis has been submitted and is under review.

Here are some technical details of the current technical developments.

Signal types. There are three kinds of audio signal types to be used for machine learning: data in the *temporal domain*, which is a direct recording of the audio signal; *spectrogram*, which performs a Discrete Fourier Transform on temporal slices of the data, typically at 10 ms intervals; and *cepstrogram*, which performs a cepstral transformation at the same time slices, also called MFCC-transform [Casanova et al. 2021b]. Previous results for human voice classification have showed a considerable advantage in using cepstrograms, which is the preferred approach for our initial analyses.

Algorithms considered. According to previous experience [Casanova et al. 2021b, Gauy and Finger 2021b, Gauy and Finger 2023], we should start our developments by using pretrained audio (convolutional) neural networks (PANNs) [Kong et al. 2020], which we should consider in the formats PANN6, PANN10 and PANN14. Other neural architectures will later be considered, in particular pretrained audio-transformer and Masked Auto-Encoders (MAEs) [Huang et al. 2022].

Preprocessing steps. We do not employ any kind of audio filters. Previous work has demonstrated that the injection of noise, to uniformize signals yields best results, and we plan to employ this technique.

5. Conclusions

The SPIRA-BM project is an ongoing research effort to study and develop audio biomarkers based on a multidisciplinary approach to the study of respiratory conditions. Several areas are jointly contributing to this goal: Computer Science (AI and Software Engineering), Medicine (Pneumatology), Speech Therapy, Physical Therapy and Linguistics.

With the progression of data collection, several voice and speech data studies and big-data models will become available. Thus, the development of practical tools for respiratory insufficiency diagnostics with an indication of its probable cause will become available. A *clinical validation* of this tool is a planned activity as soon as the trained models are stable, which should test it in direct use by medical crews; contact with healthcare is ongoing throughout the project.

Acknowledgements

This work was supported by FAPESP under grant 2023/00488-5 (SPIRA-BM) and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES) — Finance Code 001. It was partially carried out at the Center for Artificial Intelligence (C4AI-USP), supported by FAPESP grant 2019/07665-4 and by the IBM Corporation. Marcelo Finger was partly supported by CNPq grant PQ 302963/2022-7. The sponsors' involvement in the work was limited to the provided funding.

References

Aluisio, S. et al. (2022). Detecting respiratory insufficiency via voice analysis: The Spira Project. In *Practical Machine Learning for Developing Countries*. ICLR.

Atmaja, B. T., Asmoro, W. A., Sasou, A., et al. (2025). Cross-dataset COVID-19 transfer learning with data augmentation. *Int. J. of Information Technology*, pages 1–14.

Biomarkers Definitions Working Group (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, 69(3):89–95.

Botelho, M. C., Trancoso, I., Abad, A., and Paiva, T. (2019). Speech as a biomarker for obstructive sleep apnea detection. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5851–5855. IEEE.

Brown, C. et al. (2020). Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, page 3474–3484. ACM.

Casanova, E., Candido Jr., A., Fernandes Jr., R. C., Finger, M., Gris, L. R. S., Ponti, M. A., and Pinto da Silva, D. P. (2021a). Transfer Learning and Data Augmentation Techniques to the COVID-19 Identification Tasks in ComParE 2021. In *Proc. Interspeech 2021*, pages 446–450. Stefan Steidl Computational Paralinguistics Award, COVID-19 Cough Sub-Challenge Prize.

Casanova, E. et al. (2021b). Deep learning against COVID-19: Respiratory insufficiency detection in brazilian portuguese speech. In *Findings of the ACL*. ACL.

Coppock, H. et al. (2022). Audio-based ai classifiers show no evidence of improved covid-19 screening over simple symptoms checkers. *arXiv:2212.08570*.

Gauy, M. and Finger, M. (2021a). Audio mfcc-gram transformers for respiratory insufficiency detection in covid-19. In *STIL2021*, pages 143–152, Porto Alegre, RS, Brazil.

Gauy, M. M. et al. (2023). Discriminant audio properties in deep learning based respiratory insufficiency detection in brazilian portuguese. In *Artificial Intelligence in Medicine, AIME 2023*, page 271–275, Berlin, Heidelberg. Springer-Verlag.

Gauy, M. M. and Finger, M. (2021b). Audio mfcc-gram transformers for respiratory insufficiency detection in covid-19. In *STIL 2021 ()*.

Gauy, M. M. and Finger, M. (2023). Acoustic models of brazilian portuguese speech based on neural transformers. *arXiv preprint arXiv:2312.09265*.

Han, J., Xia, T., Spathis, D., Bondareva, E., Brown, C., Chauhan, J., Dang, T., Grammenos, A., Hasthanasombat, A., Floto, A., et al. (2022). Sounds of covid-19: exploring realistic performance of audio-based digital testing. *NPJ digital medicine*, 5(1):1–9.

Huang, P.-Y., Xu, H., Li, J., Baevski, A., Auli, M., Galuba, W., Metze, F., and Feichtenhofer, C. (2022). Masked autoencoders that listen. *Advances in Neural Information Processing Systems*, 35:28708–28720.

Imran, A., Posokhova, I., Qureshi, H. N., Masood, U., Riaz, M. S., Ali, K., John, C. N., Hussain, M. I., and Nabeel, M. (2020). Ai4covid-19: Ai enabled preliminary diagnosis for covid-19 from cough samples via an app. *Informatics in Medicine Unlocked*, 20:100378.

Kong, Q., Cao, Y., Iqbal, T., Wang, Y., Wang, W., and Plumbley, M. D. (2020). Panns: Large-scale pretrained audio neural networks for audio pattern recognition.

Nevler, N., Ash, S., Irwin, D. J., Liberman, M., and Grossman, M. (2019). Validated automatic speech biomarkers in primary progressive aphasia. *Annals of Clinical and Translational Neurology*, 6(1):4–14.

Roberts, M. et al. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for covid-19 using chest radiographs and ct scans. *Nature Machine Intelligence*, 3:199–217.

Trancoso, I., Correia, M. J. R. F., Teixeira, F., Abad, A., Botelho, M. C. T., and Raj, B. (2019). Speech as a (private?) biomarker for speech affecting diseases. In *In ICIEA 2019 - The 14th IEEE Conference on Industrial Electronics and Applications*, Xi'an, China. IEEE. Keynote paper.

Wynants, L. et al. (2020). Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*, 369.