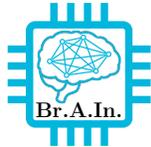


Thèse: Applications de l'apprentissage profond pour l'analyse de la respiration pour l'aide au diagnostic



IMT Atlantique, Brest, France
OSO AI, Brest, France

1 Résumé

Objectif Aide au diagnostic de pathologies respiratoires par deep learning sur des enregistrements audio

Domaine de recherche Deep Learning, Audio, aide au diagnostic

Encadrement académique Dr. Nicolas Farrugia, IMT Atlantique

Partenaire industriel Olivier Menut, OSO AI.

Partenaire clinique CHRU, Brest.

Début Fin 2021

2 Description du sujet

De nouvelles applications de l'apprentissage profond émergent, notamment dans le domaine de la Santé, telles que l'aide au diagnostic par le son. En particulier, le son produit par la respiration humaine comporte une remarquable diversité, ainsi que des caractéristiques temporelles, régulières et rythmiques. L'altération de ces caractéristiques peut indiquer la présence de pathologies, telles que l'apnée du sommeil, ou des difficultés respiratoires concomitantes à des maladies cardiovasculaires ou pulmonaires. L'objectif de cette thèse est d'analyser la respiration à partir du son, dans un but d'aide au diagnostic. Alors que l'oreille experte d'un médecin sait identifier les premiers signes pathologiques, un système basé sur l'IA permettrait la détection automatique de tels signes, par exemple dans un milieu médicalisé, dans un EPHAD, ou à domicile. Nous nous baserons sur les avancées récentes en apprentissage profond, notamment l'apprentissage auto-supervisé [1, 2, 3, 4], afin de rechercher des représentations efficaces du signal audio de respiration. A terme, de telles représentations ouvriront de nouvelles possibilités d'aide au diagnostic par le son.

Les travaux de recherche initiés dans cette thèse exploreront la réalisation d’algorithmes pour l’aide au diagnostic de pathologies à partir d’enregistrements sonores de respiration de patients et de personnes fragiles. Plusieurs problématiques spécifiques apparaissent dans ce contexte :

- La quantité de données annotées, par patient, est typiquement très réduite.
- Pour un même patient, une variabilité importante peut-être constatée en fonction du moment où les enregistrements ont été effectués.
- Pour une pathologie ou un même groupe de symptômes donnés, la variabilité inter-individus peut-être importante, notamment compte tenues de différences morphologiques.
- Un enregistrement sonore de respiration dispose de caractéristiques temporelles et spectrales particulières, mais peut également aussi comporter des aspects complexes et non stationnaires, compte tenu du contexte non contraint d’enregistrement. De tels enregistrements comporteraient typiquement d’autres types de sons, ainsi que des sons intrinsèques au sujet mais considérés comme artefacts : toux, parole, bruits du lit ou de la chambre, etc.

Les enjeux spécifiques sont donc la faible annotation, la variance des données intra-individus et inter-individus, et la présence de nombreux artefacts. Une approche «traditionnelle» d’apprentissage automatique aurait taché de rejeter manuellement tous les artefacts, puis de sélectionner manuellement des caractéristiques pertinentes pour ce type de signaux, à partir de la connaissance médicale / experte, par exemple en se focalisant uniquement sur certaines fréquences ou certains rythmes typiques de la respiration. Cependant, de telles approches disposent d’une faible capacité de généralisation, étant souvent très sensibles à la qualité des données disponibles pour l’apprentissage. Le coût de mise en œuvre d’une telle solution est très important, car il nécessite non seulement une annotation robuste des données, mais également des prétraitements manuels très chronophages.

Dans cette thèse, nous proposons de tirer parti des avancées récentes en apprentissage de réseaux de neurones profonds, qui peuvent être entraînés de bout en bout avec très peu de prétraitements sur les données. Ce type de modèle permet d’utiliser toute la variabilité des données dans l’apprentissage, en minimisant le temps passé au prétraitement des données. L’enjeu de ce type d’apprentissage est de trouver des représentations dans des espaces latents, de haute dimension, permettant d’extraire l’information pertinente. Des approches récentes d’auto-supervision [1, 2] permettent d’envisager cet apprentissage de représentation sans nécessiter d’annotations, mais en créant des tâches «prétextes» d’apprentissage supervisé à partir des données elles-mêmes, en considérant par exemple les relations temporelles entre des segments audio échantillonnés à des instants différents. La première étape de ces travaux consistera donc à utiliser l’apprentissage auto-supervisé sur les enregistrements de respiration afin d’apprendre des représentations latentes [3, 4]. Une fois les bonnes représentations apprises, nous utiliserons le peu d’annotations disponibles dans un contexte de «few-shot learning». Nous comparerons différentes familles de méthodes, telles que le meta-apprentissage et les méthodes métriques, par une approche similaire à nos travaux précédents sur des signaux de neuroimagerie [5] et sur l’apprentissage efficace basé sur les formes d’ondes audio [6] ainsi que l’apprentissage avec peu d’exemples [7, 8].

3 Contexte

Ce travail sera effectué entre l’équipe BRAIn team du département Mathematical and Electrical Engineering à IMT Atlantique sur le campus de Brest, supervisé par Nicolas Farrugia, ainsi que dans l’en-

treprise OSO AI company, supervisé par Oliver Menut. Ce projet est financé par le programme d'excellence AI@IMT, et par OSO AI. Les données médicales sont déjà disponibles et continueront à être fournies via des collaborations dédiées entre OSO AI et des CHRU en France.

4 Profil souhaité

- Master ou Ecole d'Ingénieur en Data Science ou Informatique ou Acoustique ou Traitement du Signal.
- Connaissances générales en apprentissage automatique
- Maîtrise de la programmation
- Souhaitable : expérience avec des signaux audio, deep learning / machine learning sur de grandes bases de données.

5 Comment candidater

Par email à nicolas.farrugia@imt-atlantique.fr avec les pièces suivantes :

- CV, incluant des éléments de bibliographie ou de réalisations personnelles.
- lettres de recommandation ou emails de contacts de précédents encadrants ou professeurs
- Une lettre de motivation expliquant votre intérêt spécifique pour ce projet
- Relevé(s) de notes (au moins du Master 1 et Master 2 ou équivalent)
- Optionnel : film de 2 à 3 minutes réalisé par le candidat, pour motiver sa candidature.

Références

- [1] J. Bai, W. Wang, Y. Zhou, and C. Xiong, "Representation learning for sequence data with deep autoencoding predictive components," *arXiv preprint arXiv :2010.03135*, 2020.
- [2] S. Schneider, A. Baevski, R. Collobert, and M. Auli, "wav2vec : Unsupervised pre-training for speech recognition," *arXiv preprint arXiv :1904.05862*, 2019.
- [3] P.-H. Chi, P.-H. Chung, T.-H. Wu, C.-C. Hsieh, Y.-H. Chen, S.-W. Li, and H.-y. Lee, "Audio albert : A lite bert for self-supervised learning of audio representation," in *2021 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 2021, pp. 344–350.
- [4] L. Wang and A. v. d. Oord, "Multi-format contrastive learning of audio representations," *arXiv preprint arXiv :2103.06508*, 2021.
- [5] M. Bontonou, N. Farrugia, and V. Gripon, "Few-shot learning for decoding brain signals," *arXiv preprint arXiv :2010.12500*, 2020.
- [6] N. Pajusco, R. Huang, and N. Farrugia, "Lightweight convolutional neural networks on binaural waveforms for low complexity acoustic scene classification," in *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2020 Workshop (DCASE2020)*, Tokyo, Japan, November 2020, pp. 135–139.
- [7] Y. Hu, V. Gripon, and S. Pateux, "Exploiting unsupervised inputs for accurate few-shot classification," *arXiv preprint arXiv :2001.09849*, 2020.

- [8] —, “Leveraging the feature distribution in transfer-based few-shot learning,” *arXiv preprint arXiv :2006.03806*, 2020.