

RAPPORT

Afin d'expérimenter les différents modèles, il était d'abord plus simple de formater toutes les données.

On se rend compte en analysant les données que les classes sont déséquilibrées : il y a plus de cas normaux que de cas anormaux, ce qui est naturel.

Il y a quasiment 2 fois plus de cas normaux que d'anormaux. On peut alors déjà suspecter tous les modèles de plus prédire la classe des cas normaux précisément car celle-ci est surreprésentée.

Pour palier ce problème, il aurait été possible d'effectuer de l'augmentation de données, soit pour augmenter globalement tout le jeu de données et ainsi avoir un échantillon de données plus grand, soit juste augmenter les cas anormaux.

Mais finalement, aucune augmentation de données n'a été faite car nous n'avons pas trouvé des manières « simples » d'augmenter des données ECG. Dans le cas de données images, l'augmentation des données est plus naturelle, mais cela n'est pas le cas pour des ECG.

Plusieurs modèles sont entraînés. Les différents modèles sont des MLP, des CNN ou des RNN, et le nombre de couches et de neurones varient entre chaque modèle. On a d'abord commencé par des modèles plus « simples », avec moins de couches et de neurones, et on augmente ces 2 facteurs au fil des modèles.

La piste de l'apprentissage par transfert a été abordée, mais abandonnée car cela semble trop compliqué pour des données ECG, et cela fait aussi l'objet de papiers (dont un papier publié par Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar and Pierre-Alain Muller).

La piste de l'apprentissage par transfert a été abordée car le jeu de données est assez faible, et il peut être difficile d'apprendre « from scratch ». Il n'est pas non plus très courant de ce que nous avons compris d'entraîner des modèles from scratch, l'apprentissage par transfert est grandement utilisé pour entraîner des modèles, cela facilite grandement la tâche.

Les modèles sont évalués à l'aide des métriques « accuracy » et « f1-score », mais nous privilégions la métrique « F1-Score » car le jeu de données est déséquilibré. Le score « accuracy » n'est pas pertinent dans ce cas car nous pourrions avoir un score d'accuracy élevé sans pour autant avoir un modèle qui généralise (on peut prendre comme exemple la détection de fraude dans le milieu bancaire).

Enfin, en plus des métriques d'évaluation, on évalue aussi le temps d'inférence des modèles, ce qui implique aussi leur consommation énergétique.

Les modèles plus complexes consomment évidemment plus, mais ne sont pas forcément « meilleurs ».

Dans le cas d'analyse d'ECG, il faut regarder au-delà des chiffres de métrique. Il est par exemple moins grave de prédire un cas normal comme étant anormal plutôt que l'inverse.

Dans ce cas, ces modèles pourraient pousser le docteur à effectuer plus d'analyses sur le patient pour se rendre compte que le patient n'a rien, alors que dans le cas inverse, cela pourrait entraîner des répercussions grave sur le patient.