Rapport de Projet de fin d'étude

Détection préventive des états de stress

DESIX Thibaut, MORON Benjamin, PALMILHA Jeremy

Université de Nice Sophia Antipolis Département de Mathématiques - Valrose Master d'Ingénierie Mathématique

Table des matières

1	Intr	roduction	2
2	Exp	ploration des fichiers	2
	2.1	État des lieux du fichier WESAD	2
	2.2	Profil des sujets de l'étude	3
	2.3	Extraction des données	4
3	Tra	vail préliminaire sur les données	4
	3.1	Étiquettes de label	4
	3.2	Capteur à la poitrine	5
	3.3	Capteur au poignet	6
	3.4	Observation des corrélations entre les variables physiologiques	
		et le stress	8
4	Pré	traitement des données	9
	4.1	Rééchantillonnage des données du poignet	9
	4.2	Normalisation des données	10
	4.3	Binarisation des labels	10
5	Mo	dèles de prédictions	11
	5.1	Régression logistique	12
	5.2	Random forest	15
	5.3	Méthode de validation	17
	5.4	Modèle avec poignée et poitrine en commun	18
	5.5	Model commun testé sur un sujet	19
6	Cor	nclusion	21
7	Bib	liographie	22

1 Introduction

Le stress touche tous les individus dans la vie quotidienne, influence notre bien-être et peut être nocif lorsque sa durée est prolongée, pouvant entraîner un risque de problèmes de santé comme des maladies cardiaques, de l'hypertension artérielle ou du diabète, ainsi que des problèmes psychologiques comme de l'anxiété et une dépression. Pour le prévenir, il est nécessaire de le détecter à l'aide de facteurs physiologiques comme la fréquence cardiaque, la variabilité de la fréquence cardiaque, la pression artérielle, la température corporelle, etc . . . L'objectif de notre projet est de détecter ce stress en utilisant des algorithmes d'apprentissage automatique afin de l'anticiper et de le traiter.

2 Exploration des fichiers

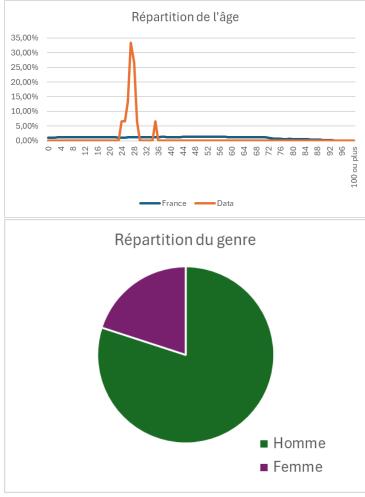
2.1 État des lieux du fichier WESAD

Avant de commencer à construire des modèles, il est essentiel de comprendre la nature des données avec lesquelles on travaille. L'exploration permet de visualiser les données et d'identifier les types de variables. Cela donne un aperçu clair de ce à quoi ressemblent les données et des défis potentiels à relever.

Dans cette étude, 17 sujets ont participé. Cependant, en raison d'un dysfonctionnement du capteur, les données de deux sujets (S1 et S12) ont dû être écartées. Par conséquent, les dossiers respectifs sont manquants dans le jeu de données. Des informations sur chaque sujet sont disponibles dans les fichiers SXreadme.txt, (avec X le numéro du sujet.

Les modalités de capteurs suivantes sont incluses : pouls du volume sanguin, électrocardiogramme, activité électrodermale, électromyogramme, respiration, température corporelle et accélération sur trois axes. En parallèle, les sujets de l'expérience ont exprimé leur ressenti continuellement (notamment quand ils étaient stressés) lors de la capture. Pendant l'expérience, ils ont été soumis à différentes situations considérées comme neutres, relaxantes ou stressantes afin de provoquer d'éventuelles évolutions sur leurs mesures corporelles.

2.2 Profil des sujets de l'étude



La répartition des âges sur les sujets étudiés est différente de celle de la population totale.

Ici aussi, on voit que les sujets de test sont 80% d'hommes, ce dataset ne représente pas la répartition de genre de population Française (48% d'hommes pour 52% de femmes).

Commentaire général:

Les données WESAD sur lesquelles nous avons travaillé ont des particularités qu'il faut garder en tête pour avoir un regard critique sur les résultats. Premièrement, dans les sujets il y a seulement un fumeur et un sportif sur quinze patients. De plus, il y a une écrasante majorité d'hommes (seulement 2 femmes) et ces sujets sont tous de jeunes adultes. Les modèles seront fortement biaisés à cause du manque de diversité des données. Par conséquent, si on souhaite les utiliser pour faire des conclusions sur l'intégralité de la population ou sur des populations différentes, nous n'avons aucune garantie d'obtenir de bonnes prédictions.

2.3 Extraction des données

Les données nous ont été fournies sous la forme d'un fichier pickle. Ceci a permis aux créateurs du jeu de données de nous les transmettre tout en gardant leur structure particulière, ici un dictionnaire dans un fichier compact. Le procédé qui permet ceci est la sérialisation et le module pickle est celui qui effectue ceci le mieux pour des données manipulées sur Python. Pour y avoir accès nous avons dû procéder à la désérialisation du fichier autrement nommé "unpickling". Ensuite, nous l'avons stocké et manipulé à l'aide de dataframes Pandas qui sont les classes standard de data pour l'apprentissage machine sur Python.

3 Travail préliminaire sur les données

3.1 Étiquettes de label

Le champ 'label' fait référence à l'identifiant de la condition du protocole d'étude respective, échantillonné à 700 Hz. Les identifiants suivants sont fournis avec leurs significations :

- 0 : État non-défini / État transitoire
- 1 : État de base
- 2 : État de stress
- 3 : État d'amusement
- 4 : État de méditation

Ces identifiants permettent de catégoriser les données selon les différentes phases ou conditions du protocole d'étude, offrant ainsi une contextualisation importante pour l'analyse des données.

Chacun des fichiers qui contiennent les données que nous exploiteront offrent des informations précieuses sur différentes variables physiologiques, avec des échantillonnages et des unités de mesure spécifiques, permettant une analyse approfondie des données physiologiques captées.

Les données brutes des capteurs ont été enregistrées avec deux appareils : un dispositif porté sur la poitrine (RespiBAN) et un dispositif porté au poignet de la main non dominante (Empatica E4). Les étiquettes du protocole d'étude sont synchronisées avec les données brutes de RespiBAN.

3.2 Capteur à la poitrine

Toutes les variables issues du capteur RespiBAN (sur la poitrine) sont échantillonnées à 700 Hz. Nous disposons de différentes variables de capteurs physiologiques, chacune avec sa propre formule de conversion en unités de mesure pertinentes.

L'électrocardiogramme (ECG), l'activité électrique du cœur est mesurée en millivolts, avec une formule de conversion qui ajuste le signal en fonction de la tension d'alimentation.

L'activité électrodermale (EDA), mesurée en microsiemens, la formule prend également en compte la tension d'alimentation pour convertir le signal brut en unités de mesure.

L'électromyographie (EMG) mesure l'activité électrique des muscles, également en millivolts, avec une formule de conversion similaire à celle de l'ECG.

La température (TEMP), mesurée en degrés Celsius, utilise une formule complexe prenant en compte la tension de sortie du capteur et la résistance thermique de la thermistance.

Les accéléromètres XYZ (ACC) mesurent les accélérations linéaires dans les trois axes, converties en unités de gravité en fonction de valeurs de calibrage spécifiques.

Enfin, la respiration (Resp) est mesurée en pourcentage et est calculée en fonction du signal brut et de la résolution du convertisseur analogique-numérique.

Chaque formule est spécifique au capteur correspondant, offrant ainsi une interprétation précise des données physiologiques captées.

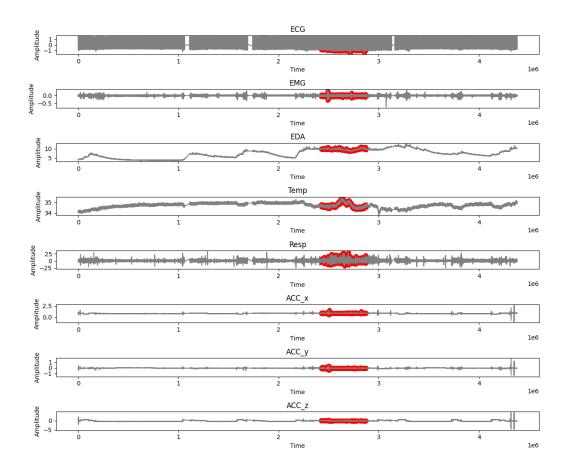


FIGURE 1 – Tracés des données de la poitrine avec le vrai stress

3.3 Capteur au poignet

Les variables issues du capteur Empatica E4 au poignet ne sont pas échantillonnées à la même fréquence. Les trois colonnes de données XYZ correspondent aux trois canaux de l'accéléromètre (ACC) et fournissent des données échantillonnées à 32 Hz. Les données sont fournies en unités de 1/64g. Le Capteur Biométrique de Volume de Sang (BVP) est un capteur optique non invasif qui mesure la dynamique cardiovasculaire en détectant les modifications de la translucidité artérielle, les données correspondantes sont échantillonnées à 64 Hz et proviennent du photopléthysmographe.

Ces données ne nécessitent pas de conversion supplémentaire et peuvent être utilisées telles quelles.

Les données correspondantes à l'activité électrodermale (EDA) et à la température (TEMP) sont échantillonnées à 4 Hz.

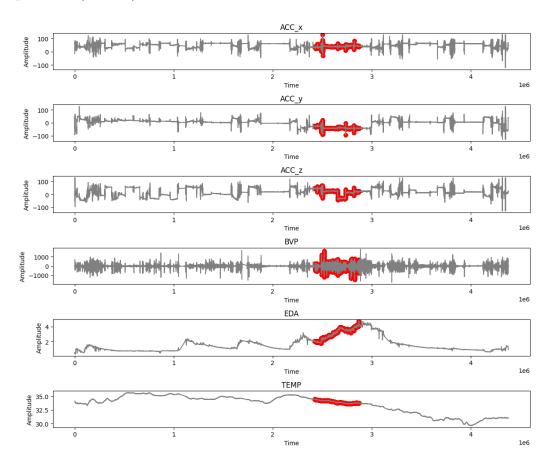


FIGURE 2 – Tracés des données du poignet avec le vrai stress.

Les figures représentant les variables issues de la poitrine et du poignet sont tracées en fonction du temps de l'étude, l'intervalle en rouge correspond à la période à laquelle le sujet a été soumis à une situation de stress.

3.4 Observation des corrélations entre les variables physiologiques et le stress

La corrélation est une mesure statistique qui évalue la force et la direction de la relation linéaire entre deux variables. Elle aide à déterminer dans quelle mesure les valeurs d'une variable varient quand il y a des changements observés dans une autre variable. Cette étape implique l'analyse des corrélations entre les différentes variables physiologiques et l'état de stress rapporté par le sujet. Cela permet de comprendre les relations entre ces variables pour identifier celles qui génèrent une redondance d'information. Mais cela permet surtout de trouver les variables qui sont le plus corrélées au label à prédire.

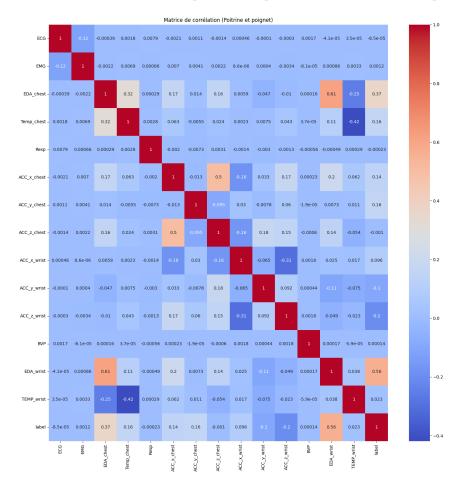


FIGURE 3 – Matrice de corrélation.

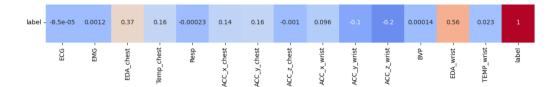


FIGURE 4 – Corrélation des variables avec label

On remarque que les corrélations avec le stress les plus fortes sont obtenues par les mesures effectuées par le capteur au poignet.

4 Prétraitement des données

Le prétraitement des données constitue l'étape fondamentale pour purifier, normaliser et structurer les informations, garantissant ainsi la fiabilité des résultats analytiques et la robustesse des modèles prédictifs. Nous allons détailler les étapes du prétraitement, telles que le rééchantillonnage et l'analyse des données ainsi que la transformation du problème multi-classes en un problème binaire.

4.1 Rééchantillonnage des données du poignet

Pour harmoniser la fréquence des données extraites du capteur Empatica E4 avec la fréquence des étiquettes, nous avons opté pour une technique d'échantillonnage connue sous le nom de "upsampling". Cette méthode consiste à augmenter la densité des échantillons en interpolant les données existantes pour créer de nouveaux points de données à une fréquence plus élevée. Dans notre cas, cela signifie que les variables provenant du capteur du poignet sont désormais échantillonnées à une fréquence uniforme de 700 Hz, ce qui correspond à la fréquence des étiquettes. Ce processus garantit une cohérence temporelle entre les données physiologiques capturées par le capteur et les étiquettes correspondantes, facilitant ainsi une analyse précise et synchronisée des données.

4.2 Normalisation des données

Une fois que les données ont été collectées, nous avons entrepris une analyse descriptive approfondie pour mieux comprendre leur distribution et leurs caractéristiques. Cette analyse a impliqué le calcul de diverses statistiques descriptives, telles que la moyenne, l'écart type, le minimum et le maximum, afin d'obtenir un aperçu complet de la répartition des valeurs dans l'ensemble des données. Après avoir examiné ces statistiques, nous avons identifié la nécessité potentielle de normaliser les données. La normalisation vise à mettre à l'échelle les valeurs des différentes variables dans un intervalle spécifique, ce qui facilite la comparaison entre elles et garantit que chaque variable contribue de manière égale à l'analyse. Cette étape est particulièrement importante lorsque les variables ont des échelles différentes ou des distributions asymétriques, permettant ainsi une interprétation plus fiable et une meilleure performance des modèles de prédiction qui seront appliqués par la suite.

4.3 Binarisation des labels

Dans le cadre de la résolution d'un problème comportant cinq labels distincts, dont l'un est spécifiquement associé à des situations de stress, une stratégie pertinente consiste à simplifier le problème en le transformant en un problème binaire. Cette conversion implique de considérer les instances étiquetées avec le label associé au stress comme des cas positifs, tandis que les autres labels sont regroupés comme des cas négatifs. Cette approche se justifie par la focalisation sur l'identification des moments de stress, une préoccupation majeure dans notre problème de détection du stress où nous observons des réactions physiologiques face à des stimuli stressants. En réduisant la complexité du problème et en le concentrant sur la détection du stress, cette approche binaire simplifie l'analyse des données et permet de concevoir des modèles plus ciblés et efficaces pour répondre aux besoins spécifiques de détection du stress.

Après le prétraitement, les sujets contiennent 8 variables pour le jeu de données sur la poitrine et 6 variables pour le jeu de données sur le poignet avec chacune entre 3,528,700 et 4,825,799 données. La dimension importante de ces ensembles de données nous conduit à privilégier l'utilisation de modèles de prédiction conçus pour traiter efficacement de grandes quantités de données.

5 Modèles de prédictions

Nous entrons dans la phase de modélisation prédictive, où nous utilisons des modèles d'apprentissage automatique pour classifier l'état de stress chez un sujet à partir des données physiologiques collectées. Compte tenu de notre analyse descriptive, nous formulons l'hypothèse que les données du capteur du poignet offriront de meilleures performances prédictives que celles du capteur à la poitrine.

En effet, nos observations ont révélé que les variables physiologiques mesurées par le capteur du poignet présentaient des corrélations plus significatives avec l'étiquette de stress que celles mesurées par le capteur à la poitrine.

Pour mesurer l'efficacité de nos modèles, nous utilisons les métriques suivantes :

La précision (precision) mesure le nombre de vrais positifs divisé par le nombre total de positifs prédits par le modèle. Elle indique la proportion d'observations positives prédites par le modèle qui sont réellement positives.

$$Pr\'{e}cision = \frac{Vrais\ positifs}{Vrais\ positifs + Faux\ positifs}$$

Le rappel (recall) mesure le nombre de vrais positifs divisé par le nombre total de vrais positifs et de faux négatifs. Il indique la proportion d'observations positives réellement prédites par le modèle parmi toutes les observations positives réelles.

$$\text{Rappel} = \tfrac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux n\'egatifs}}$$

Le score F1 (F1-score) est la moyenne harmonique de la précision et du rappel. Il est utile lorsque les classes sont déséquilibrées.

$$F1\text{-}score = 2 \times \tfrac{Pr\acute{e}cision \times Rappel}{Pr\acute{e}cision + Rappel}$$

L'exactitude (accuracy) mesure la proportion d'observations correctement classées parmi toutes les observations. C'est le rapport entre le nombre total de prédictions correctes et le nombre total d'observations.

$$\text{Exactitude} = \frac{\text{Nombre total de prédictions correctes}}{\text{Nombre total d'observations}}$$

5.1 Régression logistique

Nous avons choisi d'utiliser une régression logistique dans un premier temps pour sa simplicité et son court temps d'exécution.

La régression logistique est une technique d'analyse statistique utilisée pour prédire la probabilité qu'une variable dépendante catégorielle ait une certaine valeur en fonction d'un ensemble de variables indépendantes. Cette méthode est utilisée lorsque la variable dépendante est binaire. Elle est donc adaptée à notre problème après la binarisation des labels. Ici, 0 indique 'non-stressé' et 2 'stressé'.

Rapport de cl	assification precision	pour l'en		test (poitrine) : support
0 2	0.90 0.52	0.98 0.20	0.94 0.29	15644495 2093155
accuracy macro avg weighted avg	0.71	0.59	0.88 0.61	17737650 17737650

FIGURE 5 – Rapport de classification poitrine

Rapport de	classi	Fication	pour	l'en	semble	e de	test	(poignet)	:
	pred	cision	reca	11	f1-sc	ore	supp	ort	
		0.94	0.	.97	0	.96	15644	1495	
		0.72	0.	. 54		.61	2093	3155	
accura	cv				0.	. 92	17737	7650	
macro a	vg	0.83	Θ.	.75		. 78	17737	650	
weighted a	vg	0.91	0.	.92		.92	17737	7650	

FIGURE 6 – Rapport de classification poignet

Nos premières prédictions avec la régression logistique ont montré une précision moyenne pour le label de stress, ce qui suggère que le modèle a tendance à classer les négatifs en positifs. Cependant, un recall faible indique que le modèle a du mal à capturer les exemples positifs, tandis qu'un F1-score faible traduit une mauvaise performance globale du modèle. Ces résultats soulignent l'importance du déséquilibre des classes dans l'évaluation de l'accuracy, qui reste élevée en raison de ce déséquilibre. Ainsi, la prochaine étape consiste à équilibrer les classes. Par ailleurs, il est intéressant de noter que les données provenant du poignet semblent offrir de meilleures prédictions que celles provenant de la poitrine.

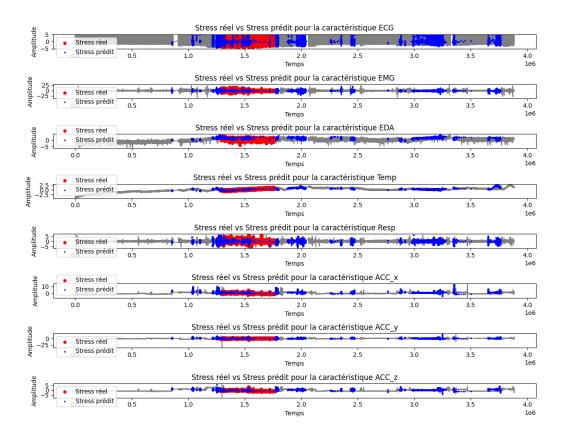


FIGURE 7 – Tracés des données de la poitrine avec le vrai stress et le stress prédit.

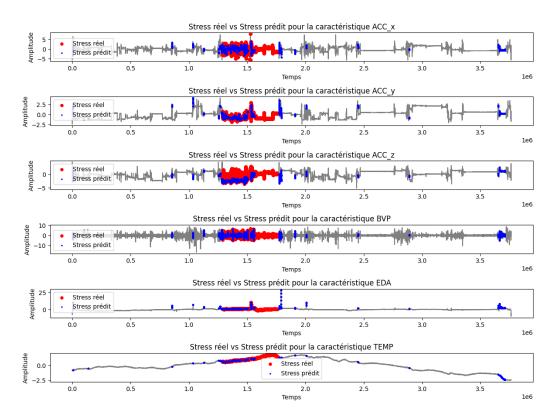


FIGURE 8 – Tracés des données de la poignet avec le vrai stress et le stress prédit.

5.2 Random forest

Nous nous sommes ensuite orientés vers un algorithme d'apprentissage automatique plus complexe. Le Random Forest fonctionne en construisant un grand nombre d'arbres de décision pendant l'apprentissage, où chaque arbre est formé sur un sous-ensemble aléatoire des données d'entraînement et utilise un sous-ensemble aléatoire des caractéristiques disponibles. Lorsque des prédictions sont nécessaires, chaque arbre vote pour la classe, et la prédiction finale est déterminée par un vote majoritaire ou la moyenne des prédictions des arbres individuels. Ce modèle de prédiction est apprécié pour sa robustesse, sa capacité à gérer des ensembles de données de grande dimension avec des caractéristiques hétérogènes, et leur résistance au surapprentissage. Dans notre étude, nous avons varié le nombre d'estimateurs et la profondeur maximale, deux hyperparamètres qui influent sur les performances de ce modèle, le premier contrôle le nombre d'arbres de décisions construits dans le modèle, le deuxième contrôle la profondeur maximale de chaque arbre.

Dans un premier temps, nous avons exécuté notre algorithme avec une séparation des données de 70% pour l'ensemble d'apprentissage et de 30% pour l'ensemble de test et de faibles valeurs de paramètres, 20 estimateurs et une profondeur maximale de 10:

	precision	recall	f1-score	support	
ø	1.00	0.87	0.93	15644495	
2	0.50	0.97	0.66	2093155	
accuracy			0.88	17737650	
accui acy					
macro avg	0.75	0.92	0.79	17737650	
weighted avg	0.94	0.88	0.90	17737650	

FIGURE 9 – Rapport de classification poitrine

	precision	recall	f1-score	support
0	1.00	0.97	0.98	15644495
2	0.81	0.98	0.89	2093155
accuracy			0.97	17737650
macro avg	0.90	0.97	0.94	17737650
weighted avg	0.98	0.97	0.97	17737650

FIGURE 10 – Rapport de classification poignet

Les prédictions sont meilleures que celles obtenues par régression logistique, nous constatons une augmentation du recall indiquant que le modèle capture mieux les exemples positifs. Pour la poitrine, la precision sur le label de stress est moyenne et un f1-score insuffisant, tandis que le modèle pour le poignet donne des performances très satisfaisantes avec une accuracy de 0.97.

Nous avons maintenant séparé les données en 80% pour l'ensemble d'apprentissage et en 20% pour l'ensemble de test et augmenté les paramètres du modèle pour améliorer ses performances en fixant le nombre de d'estima-

teurs à 35 et aucune condition sur la profondeur maximale (ainsi les arbres de décision se termineront tous par une feuille pure) :

	precision	recall	f1-score	support
0	1.00	1.00	1.00	10430529
2	0.98	0.98	0.98	1394571
accuracy			1.00	11825100
macro avg	0.99	0.99	0.99	11825100
weighted avg	1.00	1.00	1.00	11825100

precision recall f1-score support

0 1.00 1.00 1.00 10430529
2 1.00 1.00 1.00 1394571

accuracy 1.00 1.00 11825100
macro avg 1.00 1.00 1.00 11825100
weighted avg 1.00 1.00 1.00 11825100

FIGURE 11 – Rapport de classification poitrine

FIGURE 12 – Rapport de classification poignet

Nous obtenons une prédiction parfaite, cela suggère que le modèle a surajusté les données dû à l'absence de condition sur la profondeur maximale. Ainsi, en ayant appris le bruit et les détails spécifiques aux données d'entraînement au lieu de capturer les caractéristiques générales, le modèle donne d'excellents résultats sur notre jeu de données mais en donnera de mauvaises sur un nouveau.

Nous avons alors décidé d'optimiser les hyperparamètres avec Random Search qui exécute le modèle de prédiction avec des combinaisons aléatoires de valeur d'hyperparamètre et renvoie la meilleure combinaison. Pour l'algorithme de RandomForest, nous trouvons que les meilleurs paramètres sont 30 estimateurs et 15 de profondeur maximale pour la poitrine et 35 estimateurs et 15 de profondeur maximale pour le poignet.

	precision	recall	f1-score	support
0	1.00	0.92	0.96	10430529
2	0.62	0.98	0.76	1394571
accuracy			0.93	11825100
macro avg	0.81	0.95	0.86	11825100
weighted avg	0.95	0.93	0.93	11825100

precision recall f1-score support

0 1.00 0.99 0.99 10430529
2 0.91 1.00 0.95 1394571

accuracy 0.99 11825100
macro avg 0.95 0.99 0.97 11825100
weighted avg 0.99 0.99 0.99 11825100

FIGURE 13 – Rapport de classification poitrine

FIGURE 14 – Rapport de classification poignet

On observe encore une fois une meilleure efficacité du modèle associé au poignet, la precision des labels est élevée avec une excellente capacité à capturer les exemples positifs et une très bonne performance.

Prédictions sur les variables des données poignet :

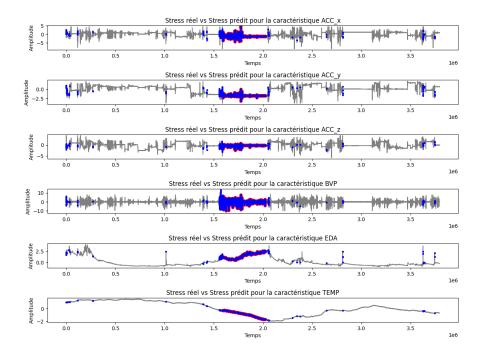


FIGURE 15 – Tracés des données du poignet avec le vrai stress

5.3 Méthode de validation

La validation croisée est une technique d'évaluation des performances d'un modèle prédictif. Elle consiste à diviser l'ensemble des données en plusieurs sous-ensembles (plis), puis à entraîner et évaluer le modèle sur différents sous-ensembles successivement.

La validation croisée sur le dernier modèle du poignet donne ces exactitudes pour 5 plis : [0.89831291, 0.96672891, 0.92628959, 0.90466787, 0.91276429]. On ne distingue pas une disparité importante entre l'accuracy obtenue sur l'ensemble de test et sur l'ensemble de validation, c'est une bonne chose : l'inverse aurait suggéré que le modèle aurait surajusté les données.

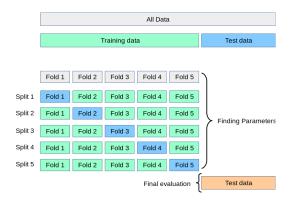


Figure 16 –

5.4 Modèle avec poignée et poitrine en commun

Dans cette partie, nous avons combiné les features récoltées avec l'appareil placé sur la poitrine et les données provenant du poignet. Grâce au ré échantillonnage effectué au préalable sur les données du poignet, les dimensions ont coïncidé et nous avons ainsi concaténé toutes ces colonnes dans un nouveau dataset.

Ensuite, nous avons sélectionné les variables les plus explicatives selon la matrice de corrélation, nous avons donc gardé la température de la poitrine, l'accéléromètre et l'activité électrodermale du poignet. Ce qui rend le modèle plus synthétique et les calculs plus rapides.

En se basant sur les modèles précédents, nous avons pris 30 estimateurs. On a néanmoins fait le choix d'un maximum de profondeur de 10 pour faire valoir la pertinence du pré-traitement supplémentaire en observant si les mesures de précision sont proches des résultats du modèle du poignet.

	precision	recall	f1-score	support
0	1.00	0.97	0.98	10430529
2	0.83	0.98	0.90	1394571
			2 27	44005400
accuracy			0.97	11825100
macro avg	0.91	0.98	0.94	11825100
weighted avg	0.98	0.97	0.97	11825100

FIGURE 17 – Rapport de classification du poignet

La conclusion est similaire à celle obtenue pour le modèle utilisant uniquement les données du poignet, mais avec une profondeur maximale réduite d'un tiers. Cette modification permet un traitement beaucoup plus rapide, bien que cela se traduise par une perte de précision de seulement 10%.

5.5 Model commun testé sur un sujet

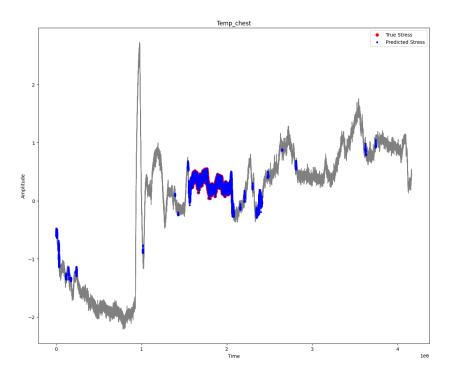


FIGURE 18 – Tracé de la température issue des données de la poitrine, vrai stress et le stress prédit

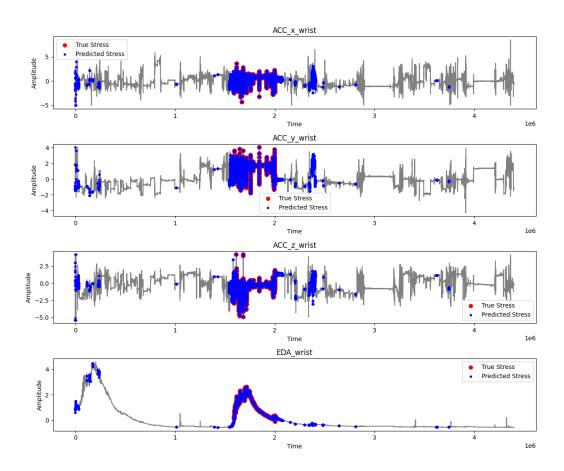


FIGURE 19 – Tracés des données du poignet, vrai stress et le stress prédit

6 Conclusion

Parmi les modèles testés, le Random Forest semble être le plus efficace, en effet, les métriques comme la precision, le recall, l'accuracy, et le F1-score obtenues avec cette méthode sont supérieures à celles d'une régression logistique au coût d'un temps de calcul plus long. Cependant, le temps de calcul est un non facteur dans le cadre de notre projet où le temps à notre disposition était amplement suffisant pour ces computations. Nous avons constaté que notre hypothèse selon laquelle les données provenant du capteur du poignet produisent de meilleurs résultats que celles issues du capteur thoracique s'est confirmée.

Les données ont été collectées sur un petit effectif (17 individus dont 2 aux données inexploitables). La méthode d'identification de l'émotion utilisée est discutable. En effet, la mesure de l'émotion du patient a été relevée en fonction du témoignage du patient lui-même. Les sujets de l'étude remplissaient des questionnaire au fur et à mesure de l'expérience pour évaluer leurs émotion à l'instant t. Cette méthode a été choisie car l'expérience originalement avait pour objectif d'identifier un panel d'émotions contrairement à nous où seul le stress importe. Une idée d'adaptation de la collecte la rendant adaptée aux besoins de l'étude serait comme suit : Des capteurs sont placés sur le patient ainsi qu'un analyseur sanguin en temps réel. Le test se déroule comme précédemment sauf que le taux de cortisol est aussi récolté. La cortisol est l'hormone sécrétée par le cerveau lors d'une situation de stress, donc évaluer son taux revient à évaluer le niveau de stress. Ceci permettrait non seulement d'avoir une mesure plus fiable du stress qu'un témoignage de ressenti, mais aussi d'observer les stades intermédiaires du stress de part le caractère continu de la donnée. Il serait possible de créer des classes correspondant à des encadrements de valeurs de label qui témoignerait de niveaux de stress: "faible", "moyen", "élevé" par exemple. On aurait aussi pu essayer d'aller plus loin dans le pré-processing en créant des données issues de celles existantes. Par exemple, quantifier la pente d'une variable en modélisant son gradient.

7 Bibliographie

Références

- [1] Pickle file: https://docs.python.org/3/library/pickle
- $[2] \ \ Pandas \ Structure: https://pandas.pydata.org/docs/user_guide/index.html$
- [3] Biosignal splux Documentation : https://support.pluxbiosignals.com/knowledge-base/biosignalsplux-documentation/
- [4] Scikit-learn Documentation Supervised learning : https ://scikit-learn.org/stable/supervised_learning.html#supervised-learning
- [5] Scikit-learn Documentation Model Selection and Evaluation : $https://scikit-learn.org/stable/model_selection.html$