

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

DÉVELOPPEMENT D'UN MODÈLE DE PRÉDICTION DU
COMPORTEMENT DES PILOTES À PARTIR DE DONNÉES
MULTIMODALES

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN INFORMATIQUE - CONCENTRATION EN
INTELLIGENCE ARTIFICIELLE

PAR

ARMAND BRIÈRE

JUILLET 2023

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.04-2020). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

TABLE DES MATIÈRES

LISTE DES TABLEAUX	v
LISTE DES FIGURES	vi
RÉSUMÉ	viii
INTRODUCTION	1
0.1 Contexte général de la recherche	1
0.2 Questions de recherche	2
0.3 Méthodologie	3
0.4 Objectifs de la recherche	4
0.5 Les acteurs du projet	4
0.6 Organisation de ce mémoire	4
CHAPITRE I REVUE DE LA LITTÉRATURE PERTINENTE	7
1.1 Analyse du comportement des pilotes dans l'aéronautique	7
1.1.1 Comportement des pilotes	8
1.1.2 Respiration et Rythme cardiaque	9
1.2 Prédiction du rythme cardiaque	11
1.3 Prédiction de la dilatation de la pupille	12
1.4 Architectures neuronales pertinentes : RNN et LSTM	13
1.4.1 RNN	13
1.4.2 LSTM	15
1.5 L'apprentissage par transfert	16
CHAPITRE II MODÈLE MULTIMODAL POUR LA PRÉDICTION DU COMPORTEMENT	18
2.1 La théorie	18
2.1.1 Analyse neuronale	20

2.2	Notre méthodologie	23
2.2.1	Modèle multimodal pour la prédiction du comportement . . .	23
2.2.2	Sous-modèles de prédiction cognitive	26
2.2.3	Prédiction du rythme cardiaque	27
2.2.4	Prédiction de la dilatation de la pupille	30
2.2.5	Prédiction de la charge mentale	31
2.2.6	Sous modèle du contexte	32
2.2.7	Sous modèle de l’environnement de simulation	33
2.3	Conclusion	34
CHAPITRE III EXPÉRIMENTATION		36
3.1	Mise en contexte	36
3.1.1	Partenaires du projet PilotAI	37
3.1.2	Objectif de l’expérience	38
3.1.3	Participants de l’expérience	39
3.2	Outils de prises de mesures	40
3.2.1	Polar H10	40
3.2.2	Gazepoint GP3	41
3.2.3	NCO EEG Headset	42
3.2.4	<i>X-Plane</i>	43
3.3	Analyse des données	44
3.3.1	Données du transfert d’apprentissage	44
3.3.2	Traitement des données	46
3.3.3	Uniformisation du format de données	50
3.3.4	Analyse du volume de données	53
CHAPITRE IV ENTRAÎNEMENT DES MODÈLES		56
4.1	Hyperparamètres	57
4.2	Métriques de validation	59

4.2.1	<i>Mean Squared Error</i> (MSE)	59
4.2.2	<i>Mean Absolute Error</i> (MAE)	60
4.3	Sous modèles	61
4.3.1	Prédiction du rythme cardiaque	61
4.3.2	Prédiction de la dilatation de la pupille	62
4.3.3	Prédiction de la charge mentale	63
4.3.4	Prédiction des données extraites du simulateur	65
4.4	Modèle multimodal pour la prédiction du comportement	66
4.5	Absence d'apprentissage	67
4.5.1	Validation du modèle multimodal	68
	CONCLUSION	71
4.6	Limitations	72
4.7	Perspectives futures	72
	RÉFÉRENCES	83

LISTE DES TABLEAUX

Tableau		Page
3.1	Scénarios des expériences	39
3.2	Rythme cardiaque du pilote 2 lors de l'expérience 2-5	48
3.3	Comparaison du nombre de lignes de données	54

LISTE DES FIGURES

Figure	Page
1.1 Réseaux de neurones classiques et RNN.	14
2.1 Modèle multimodal théorique.	22
2.2 Modèle multimodal hybride pour la prédiction du comportement.	24
2.3 Architecture des sous-modèles de prédiction cognitive.	27
2.4 Fenêtre coulissante sur un jeu de données.	29
2.5 Couche neuronale du modèle de prédiction du rythme cardiaque.	30
2.6 Couche neuronale du modèle de prédiction de la dilatation de la pupille.	31
2.7 Couche neuronale du modèle de prédiction de la charge mentale.	32
2.8 Modèle multimodal détaillé.	35
3.1 Capteur fréquence cardiaque polar H10.	40
3.2 Un système de suivi oculaire.	41
3.3 Répartition du rythme cardiaque des pilotes de l'expérience.	47
3.4 Répartition du rythme cardiaque des pilotes de l'expérience et du jeu de donnée publique.	49
3.5 Répartition de la dilatation des pupilles des pilotes de l'expérience.	50
3.6 Temps d'attente entre chaque prise de mesure.	52
3.7 Volume de données avant et après nettoyage (Rythme cardiaque).	54
3.8 Volume de données avant et après nettoyage (Dilatation de la pupille).	55
4.1 Connexion entre les WP du projet PilotAI.	75

4.2	Fonction de perte durant l'entraînement du modèle de prédiction du rythme cardiaque avec les données publiques. (La <i>loss</i> est basé sur la MAE)	75
4.3	Fonction de perte durant l'entraînement du modèle de prédiction du rythme cardiaque avec les données de l'expérience. (La <i>loss</i> est basé sur la MAE)	76
4.4	Fonction de perte durant l'entraînement du modèle de prédiction de la dilatation de la pupille avec les données publiques. (La <i>loss</i> est basé sur la MAE)	77
4.5	Fonction de perte durant l'entraînement du modèle de prédiction de la dilatation de la pupille avec les données de l'expérience. (La <i>loss</i> est basé sur la MAE)	78
4.6	Fonction de perte durant l'entraînement du modèle de prédiction de la charge mentale avec les données de l'expérience. (La <i>loss</i> est basé sur la MAE)	79
4.7	Prédiction de la charge mentale avec les données de l'expérience. (charge mentale sur l'axe vertical et le temps sur l'axe horizontal)	79
4.8	Prédiction du rythme cardiaque avec les données de l'expérience. (nombre de battements cardiaque par minute sur l'axe vertical et le temps sur l'axe horizontal)	80
4.9	Prédiction de la dilatation de la pupille avec les données de l'expérience. (dilatation de la pupille gauche (LPD) sur l'axe vertical et le temps sur l'axe horizontal)	80
4.10	Fonction de perte durant l'entraînement du modèle de prédiction de l'environnement de simulation avec les données de l'expérience. (La <i>loss</i> est basé sur la MAE)	81
4.11	Fonction de perte durant l'entraînement du modèle multimodal avec les données de l'expérience. (La <i>loss</i> est basé sur la MAE) . .	81
4.12	Fonction de perte durant l'entraînement du modèle multimodal avec les données publiques. (La <i>loss</i> est basé sur la MAE)	82

RÉSUMÉ

Le comportement, que nous définissons comme étant les manifestations sensorielles du rythme cardiaque, dilatation de la pupille, charge mentale ainsi que les interactions du pilote avec l'avion, évoluent rapidement durant un vol. Cette évolution varie en fonction des pilotes et des phases de vols. Notre recherche se concentre sur la conception de modèle de prédiction du comportement qui sera utilisé par un modèle d'analyse de déviation du comportement par un autre étudiant du projet dit PilotAI. Une expérience a été effectuée afin de collecter des données de pilotes effectuant un décollage dans le simulateur X-Plane. Un total de 9 heures de données ont été collectés. Le rythme cardiaque, la dilatation de la pupille, la charge mentale et les données du simulateur ont été collectés en continu afin de construire un jeu de données sous forme de série chronologique. Ces données ont été analysées et traitées afin d'entraîner nos différents modèles de prédiction ainsi que le modèle multimodal de prédiction du comportement des pilotes. Le jeu de données initial de 9 heures a été réduit à 2.6 heures après son nettoyage. Le volume de données de l'expérience est donc plus faible que prévu, ce qui impacte la qualité de nos résultats pour certains modèles. De plus, l'apprentissage par transfert a été utilisé pour les modèles de prédiction du rythme cardiaque et de dilatation de la pupille afin de pallier ce manque de données. Les modèles de prédiction du rythme cardiaque et de la dilatation de la pupille sont capables d'effectuer des prédictions précises une seconde dans le futur avec une MAE de .1 chacun. Les modèles de prédiction de la charge mentale et des données du simulateur X-Plane souffrent d'un manque de données. Le modèle multimodal développé permet une structure modulaire où chaque modalité peut facilement être ajoutée ou supprimée. Le manque de données de l'expérience effectué dans le cadre du projet ne nous permet pas de valider son fonctionnement absolu ; cependant celui-ci a été testé sur un jeu de données publiques ce qui a permis de confirmer un apprentissage significatif du modèle.

Mots clés : Intelligence artificielle, apprentissage profond, analyse du comportement, modèle prédictif, apprentissage par transfert.

INTRODUCTION

0.1 Contexte général de la recherche

L'ensemble de la recherche est une partie d'un plus grand projet dit PilotAI qui sera détaillé dans les prochains chapitres et sections. Le projet PilotAI a pour objectif de concevoir une intelligence artificielle capable d'assister le pilote d'un avion. Le développement d'un tel modèle présente de nombreux avantages potentiels pour la communauté scientifique.

Premièrement, cela permettrait d'augmenter considérablement la sécurité générale du transport aérien. La majorité des accidents en vol sont d'origine humaine, comme la fatigue du pilote ou le manque d'attention lors d'une période critique. L'automatisation du processus de pilotage permettrait donc de réduire le taux d'erreurs en vol. Selon l'étude de Boeing, environ 80% des accidents sont d'origine humaine dans l'aviation commerciale (Rankin, 2007). Ainsi, un pilote synthétique a le potentiel de réduire le nombre d'accidents.

Deuxièmement, un pilote synthétique pourrait augmenter l'efficacité et réduire les coûts de vols. En intégrant ce pilote synthétique, les compagnies aériennes pourraient analyser le comportement de ses pilotes afin de les accompagner dans leurs tâches ; Celle-ci pourrait aider les pilotes à s'ajuster en temps réel. Le pilote synthétique serait aussi capable d'analyser plus de métriques telles que la météo ou le trafic aérien. Cette analyse des données en temps réel serait plus rapide que celle effectuée par un pilote humain.

Enfin, un tel pilote synthétique pourrait, dans un premier temps, être utilisé pour

l'entraînement des pilotes humains :

- Simulation : Le pilote synthétique peut être utilisé dans un environnement de simulation afin d'entraîner de nouveaux pilotes dans un environnement stable et contrôlé. Les humains peuvent donc s'entraîner sur différents scénarios puis comparer leurs performances à celles du pilote synthétique.
- Rétroaction : Le pilote synthétique peut aussi analyser les performances d'un pilote humain, l'analyse des données de vol et les actions du pilote afin de proposer des pistes d'amélioration. Ceci est très efficace pour les pilotes n'ayant aucune expérience.
- Personnalisation : Un pilote synthétique peut générer des scénarios personnalisés aux besoins de chaque pilote. Chaque scénario serait alors unique à chaque pilote en fonction du niveau actuel et du niveau désiré.
- Uniformisation : Un pilote synthétique assure que chaque pilote en entraînement ou formation ait accès aux mêmes ressources. La même intelligence artificielle entraîne, évalue et forme les pilotes selon des standards définis et fixes. Chaque pilote aurait donc accès aux mêmes ressources, ce qui permettrait d'uniformiser les entraînements et donc éliminer les potentiels biais humains.

Ces nombreux avantages ainsi que leur intégration dans le domaine de la simulation justifient notre recherche. Les différents partenaires du projet sont particulièrement intéressés par la perspective d'améliorer les processus d'entraînements de chaque pilote.

0.2 Questions de recherche

Le comportement humain étant par nature multimodal, ce mémoire essaie d'évaluer et répondre à divers hypothèses. Ces hypothèses évaluent nos capacités de

prédiction sur plusieurs modalités du comportement. Les interrelations entre ces modalités seront évaluées et présentées plus en détail dans le prochain chapitre.

- Est-il possible de prédire le rythme cardiaque, la dilatation de la pupille ainsi que la charge mentale d'un pilote lors du décollage ?
- Est-il possible de prédire certaines données sensorielles à partir d'autres métriques obtenues de façon non intrusive ? Certaines prises de mesure sensorielle nécessitent l'installation d'appareil intrusif sur le pilote impactant négativement la qualité du pilotage.

0.3 Méthodologie

Dans le cadre de ce mémoire, nous avons évalué les algorithmes reliés aux réseaux de neurones. Il existe plusieurs types de réseaux de neurones ayant chacun des avantages et des prédispositions à répondre plus efficacement à certains problèmes. Par exemple, les *transformers* présentés en 2017 (Vaswani *et al.*, 2017) sont devenus la référence dans le traitement de la langue naturelle en raison de leurs nombreux avantages, mais surtout leur parallélisation de l'entraînement des modèles.

Nous avons choisi d'évaluer et d'analyser les réseaux dits récurrents (RNN) ainsi que les réseaux mémoires à long terme et à court terme (LSTM). Nous présentons donc une analyse de ces réseaux ainsi que leur intégration dans le cadre d'une classification multimodale. Enfin, plusieurs modèles utilisant ces réseaux sont développés afin de pouvoir répondre aux questions de recherches présentées précédemment.

0.4 Objectifs de la recherche

L'objectif principal de ce mémoire est de développer un modèle intelligent de prédiction du comportement d'un pilote au temps $t + 1$. La prédiction de ce comportement doit se concentrer sur les différentes modalités disponibles afin d'anticiper les variations. Il est important de noter que nous avons décidé de nous concentrer sur $t + 1$ comme étant la prédiction dans 1 seconde à partir des données actuellement disponibles. Chacun de nos modèles utilise donc les données antérieures au moment actuel t afin de prédire certaines métriques dans 1 seconde.

0.5 Les acteurs du projet

Ce mémoire de recherche se concentre sur la prédiction du comportement des pilotes, mais il s'agit en réalité d'un sous-ensemble d'un plus gros projet nommé *Pilot-AI*. Ce projet est une collaboration entre l'Université du Québec à Montréal (UQAM), l'Université de Montréal (UDEM), CAE, Bombardier et BMU. Il regroupe donc cinq entités ayant pour objectif de concevoir le cockpit du futur intégrant l'intelligence artificielle dans les processus de formation et de pilotage.

La présence des entreprises du milieu (Bombardier et CAE) est cruciale pour le projet. Leur implication est détaillée lors de la présentation des données. Il est important de souligner leur participation pour la validation de nos modèles, mais également la disponibilité des experts du domaine et des pilotes pour l'élaboration des idées et les expérimentations, à travers des réunions régulières et des ateliers tout au long de ce travail.

0.6 Organisation de ce mémoire

Ce document comporte une introduction, quatre chapitres puis une conclusion.

Dans le premier chapitre — État de l’art — nous analysons la littérature afin d’étudier le comportement des pilotes. Nous parlons là de comportement des pilotes, nous allons donc définir ce que c’est, puis présenter les différentes métriques identifiées par la littérature. Nous évaluons les motivations de l’analyse et la prédiction du comportement des pilotes (littérature) tout en comparant avec les nôtres.

Dans le deuxième chapitre — Modèle multimodal pour la prédiction du comportement — nous présentons l’architecture multimodale de notre modèle tout en détaillant chaque modalité dans son ensemble. La définition exacte de ce que nous qualifions de comportement est aussi présentée.

Dans le troisième chapitre — Expérimentations — nous détaillons l’origine des données que nous avons collectées, les outils de mesures ayant été utilisés ainsi que le traitement effectué sur celles-ci.

Dans le quatrième chapitre — Analyse des résultats — nous évaluons les différents modèles développés au niveau de la qualité de l’entraînement, mais aussi par comparaison des prédictions obtenues par rapport aux données réelles. Le traitement préalable des données est également détaillé dans ce chapitre.

Ce mémoire se termine par une conclusion générale où nous évaluons nos contributions et leurs limitations. Nous évaluons la qualité de nos résultats afin de répondre aux différentes questions que nous avons soulevées dans ce document. Des améliorations et suggestions sont aussi proposées pour de possibles travaux futurs.

La lecture de ce mémoire ne devrait pas nécessiter de connaissances particulières dans le domaine de l’intelligence artificielle. Le chapitre quatre requiert une base de connaissances en traitement et analyse de données afin de comprendre les motivations de ce traitement. L’introduction ainsi que la conclusion ne demandent

aucune connaissance scientifique et permettent au lecteur de comprendre les motivations et débouchés de ce projet.

CHAPITRE I

REVUE DE LA LITTÉRATURE PERTINENTE

Ce chapitre a pour but d'étudier, observer et présenter le cadre théorique du travail. Les différentes études du comportement des pilotes d'avion nous permettent de valider notre approche au niveau de la sélection de l'architecture de nos modèles ainsi que les métriques collectées dans notre expérience. Les modèles que nous avons développés se concentrent sur l'analyse comportementale des pilotes. Ce comportement se base sur des signaux physiques tels que le rythme cardiaque, la dilatation de la pupille et la charge mentale, mais également les actions de pilotage. La nature multimodale du comportement humain dirige notre recherche vers le développement d'un modèle de prédiction lui aussi multimodal.

Les prochaines sections se concentrent donc sur l'analyse des différentes modalités considérées par la science ainsi que les technologies adaptées au développement de notre solution.

1.1 Analyse du comportement des pilotes dans l'aéronautique

Dans un premier temps, nous avons exploré des travaux ayant traité de l'analyse du comportement des pilotes. De nombreuses études ont été menées sur le comportement des pilotes (Borghini *et al.*, 2014), (Morris et Leung, 2006), (Wilson, 2002) et (Roscoe, 1992). Les travaux de Roscoe et al (Roscoe, 1992) portent sur

une analyse du comportement, plus précisément de la corrélation entre la charge mentale avec le rythme cardiaque et la respiration. Ces travaux ont été effectués suite à la montée en popularité de l’aviation civile et militaire dans les années 1990 suscitant différentes questions relative à la sécurité et aux performances des pilotes en vol. Au moment de leur étude, la seule méthode d’estimation de la charge mentale d’un pilote est un simple questionnaire effectué après le vol.

La prédiction de la charge mentale peut être effectuée par l’installation de différents outils ; un détecteur de monoxyde de carbone ou encore l’analyse du rythme cardiaque sont des approches possibles. Cependant, lors d’un vol, l’effort purement physique d’un pilote reste faible en raison de son immobilité. Roscoe et Alan (1992) proposent alors de se concentrer sur l’analyse de la charge mentale lors du décollage et atterrissage de l’avion.

Cette sélection de ces périodes n’est pas aléatoire. En observant le comportement d’un pilote lors de son vol, il se doit de collecter, filtrer, analyser et comprendre l’information afin d’effectuer des actions rapidement. Cette augmentation de l’activité neurologique est particulièrement importante lors du décollage ou l’atterrissage puisque le nombre de tâches à effectuer deviennent plus nombreuses et plus complexes.

1.1.1 Comportement des pilotes

Nous parlons de comportements des pilotes en référence au *Behavior modeling*. Cette technique consiste à modéliser le comportement d’un individu en utilisant des données et des algorithmes. Cela permet de comprendre et de pouvoir prédire comment l’humain interagit avec l’environnement. Dans le domaine de l’intelligence artificielle, il existe plusieurs structures permettant cette modélisation :

— Réseaux bayésiens : Ces modèles probabilistes sont utilisés pour la repré-

sentation des relations entre les différentes métriques à l'étude. Le comportement futur peut être prédit en fonction du comportement actuel et des métriques sélectionnées.

- Réseaux de neurones : Les modèles d'apprentissage profond cherchant à imiter le fonctionnement du cerveau humain peuvent modéliser le comportement. L'analyse de données non structurées peut être utilisée afin de prédire le comportement futur.
- Arbres de décisions : Les modèles de classification sont utilisés pour représenter l'ensemble des décisions prises à partir d'un ensemble de données. Le comportement (décisions) peut être prédit en fonction de l'historique de données collectées.

Dans le cadre de notre projet, nous nous concentrons sur le comportement sensoriel des pilotes. Le rythme cardiaque, la dilatation de la pupille et la charge mentale sont trois métriques se retrouvant dans la littérature pour l'analyse et la prédiction du comportement des pilotes.

1.1.2 Respiration et Rythme cardiaque

La respiration est un échange d'oxygène et dioxyde de carbone avec l'environnement. D'un point de vue du métabolisme, la quantité d'oxygène nécessaire pour l'organisme est déterminée par le niveau d'activité. Une augmentation de la demande en oxygène est accompagnée d'une augmentation du rythme de la respiration. Plus l'effort physique ou mental est élevé, plus le rythme de la respiration augmente. L'analyse du nombre de respirations par minute, normalement 12 pour une personne en bonne santé, peut permettre d'évaluer une augmentation ou diminution de la charge mentale. Un stress intense ou une augmentation de la charge mentale peut entraîner le phénomène d'hyperventilation ou sous-ventilation. Ce

phénomène humain permet, théoriquement, de corréler le rythme de respiration avec le niveau de charge mentale d'un pilote. Pour un pilote, une augmentation du rythme de respiration serait alors directement reliée à une augmentation de la charge mentale puisque son activité physique est faible. Une étude (Haward, 1967) propose alors de placer un capteur dans le masque à oxygène des pilotes ou encore dans le microphone afin de collecter le taux de dioxygène de carbone expiré. Cette donnée permettrait alors d'estimer la charge mentale des pilotes.

En 1945, une autre étude (RE, 1945) avait aussi été effectuée dans les avions de combat en observant directement le volume des bonbonnes à oxygène des avions.

Ces différentes études convergent généralement vers les mêmes conclusions :

- Il est difficile de mesurer la consommation exacte en oxygène.
- Les capteurs sur le microphone ou dans le masque respiratoire ne sont pas fiables lorsque le pilote communique avec la radio ou avec son copilote.

Le rythme de respiration n'est pas la seule métrique physiologique ayant une corrélation avec la charge mentale. Le rythme cardiaque est aussi une variable qui a été étudiée dans le passé.

Les valeurs du rythme cardiaque sont plus simples à obtenir que la consommation en oxygène. Un simple capteur placé sur le torse ou autour du poignet permet d'obtenir une mesure du rythme cardiaque.

Puisque la définition même du rythme cardiaque est une moyenne du nombre de battements de cœur par minute, les prises de valeurs sont fiables, et cela, indépendamment de la qualité du capteur. Contrairement au détecteur de consommation d'oxygène, une erreur de 5 % sur la prise de mesure est moins grave sur le rythme cardiaque. La consommation en oxygène d'un homme effectuant un effort physique est de 35 – 45 ml/kg/min d'après la définition du Larousse. Cette valeur de l'ordre

du millilitre est plus sensible à la qualité des outils de mesure en comparaison aux valeurs de rythme cardiaque aux alentours des 60 – 80 battements par minute.

1.2 Prédiction du rythme cardiaque

Comme discuté plus haut, le rythme cardiaque est une métrique intéressante. Elle est facilement mesurée et est corrélée au rythme respiratoire qui est lui corrélé à la charge mentale. Il est cependant important de noter que le rythme cardiaque peut évoluer et être affecté par plusieurs facteurs :

- Activité physique ;
- Charge mentale ;
- L'âge de la personne ;
- La santé physique.

Puisque notre étude se concentre sur la prédiction du comportement, nous devons évaluer la prédiction du rythme cardiaque.

L'étude de prédiction du rythme cardiaque à base de réseaux de neurones (Luo et Wu, 2020) est une preuve de concept de l'efficacité des réseaux de neurones récurrents. Elle met en avant les différents avantages de l'utilisation des LSTM et RNN en soulevant le problème de gradient de fuite (*vanishing gradient*). Les réseaux utilisés dans cette étude seront présentés plus en détail dans une section future.

L'information importante que nous souhaitons extraire de cette étude est l'utilisation d'une architecture simple. L'architecture globale de la solution exposée est une simple couche d'entrée, une couche cachée puis une couche de sortie. La couche de traitement est une LSTM.

Cette simple étude ((Luo et Wu, 2020)) présente des valeurs de *MAPE* (Mean

Absolute Percentage Error) faibles tout en offrant une prédiction du rythme cardiaque précise. Les résultats exposés sont bons, mais bruités. Cette étude utilise un jeu de données de 48 participants. Chaque participant générait entre 90 et 580 secondes de données. Dans le cadre de notre étude, nous devons donc prendre en considération le volume de données. Ce volume, que nous approximations entre 1 et 5 heures de données, permet une prédiction du rythme cardiaque. Un des objectifs de notre étude étant la prédiction multimodale du comportement, notre volume de données se doit d'être plus élevé que celui présenté ci-dessus. L'augmentation de la complexité d'un modèle d'intelligence artificielle entraîne une augmentation du volume de données nécessaires à son développement comme nous allons le voir dans les prochains chapitres. Nous avons donc décidé d'utiliser l'*apprentissage par transfert* (*transfer learning*) que nous allons expliquer dans une prochaine section comme outil contre le manque de données.

1.3 Prédiction de la dilatation de la pupille

Une modalité dont nous n'avons pas encore parlé est la dilatation de la pupille. La pupille est l'orifice circulaire situé au centre de l'iris. Celle-ci est responsable de réguler la quantité de lumière entrant dans notre œil.

La pupille se dilate lorsque la charge mentale augmente (Zekveld *et al.*, 2011). Cette dilatation de la pupille a été étudiée afin de pouvoir établir une corrélation entre ses mouvements et les niveaux de charge mentale. Selon l'étude d'analyse du clignement des yeux (Siegle *et al.*, 2008) il existe une corrélation entre le clignement et la dilatation de la pupille. La conjonction du clignement et de la dilatation permet de déduire / calculer / estimer la charge mentale.

Une étude sur la prédiction de la dilatation de la pupille (Fridman *et al.*, 2018) présente de nombreuses similitudes avec l'analyse de prédiction du rythme car-

diague présenté précédemment. Les réseaux de neurones utilisés sont encore les LSTM ce qui renforce une fois de plus notre choix d'utiliser ces réseaux.

Le volume de données de l'expérience de Fridman et Al. regroupe 58 personnes générant un total de 3249 blocs de 15 secondes. Ce volume de données est donc plus important que celui de l'étude du rythme cardiaque. Cependant, la précision cette fois-ci n'est que de 59%.

Cette étude confirme plusieurs réflexions. La prédiction de la dilatation de la pupille est plus complexe que celle du rythme cardiaque. Contrairement au rythme cardiaque, l'évolution de la dilatation de la pupille peut varier en moins de 1000 ms (Hyönä *et al.*, 1995). Cette rapide variation impose donc un volume de données conséquent pour laisser le temps au réseau de neurones d'apprendre.

1.4 Architectures neuronales pertinentes : RNN et LSTM

L'architecture des réseaux neuronaux utilisée dans notre projet est inspirée des différentes architectures proposées dans la littérature pour l'analyse du comportement. Le comportement humain, caractérisé de multimodalité, évolue dans le temps en fonction des interactions avec l'environnement. L'état comportemental précédent impacte l'état suivant. Cette nature séquentielle du comportement nous dirige vers l'utilisation de réseaux de neurones capables de prendre en considération l'état passé afin de prédire l'état futur. Les réseaux RNN et LSTM que nous présentons ici répondent à nos besoins.

1.4.1 RNN

Les réseaux de neurones récurrents (RNN) sont un type de réseau conçu pour le traitement de données séquentielles. La différence importante avec les réseaux

classiques est l'état interne de chaque nœud des RNN. Un statut caché conserve l'information des précédentes entrées. Cette conservation de l'information permet d'étudier le passé pour impacter l'état présent. Les prédictions issues de ces réseaux prennent donc en considération les événements ayant mené à ce moment présent.

Les RNN ont été présentés par Elman en 1990 dans la publication *Finding Structure in Time* (Elman, 1990). Ces réseaux ont été utilisés dans la reconnaissance de phrases (Cho *et al.*, 2014) en combinaison avec les auto-encoder ou encore dans la prédiction de la consommation électrique (Tokgöz et Ünal, 2018) sur des données décrivant des séries chronologiques.

L'architecture interne d'un RNN est un graphe de neurones. La sortie d'un neurone est passée à l'entrée du prochain, mais aussi à lui-même. Cette liaison vers lui-même représente l'état caché permettant de garder l'information passée. Ainsi, lors de la prochaine entrée, le nœud courant aura aussi la sortie de son dernier état.

Il est important de comprendre la logique de transfert de connaissance interne précédemment évaluée comme étant une sorte de mémoire du réseau afin de comprendre les avantages des LSTM.

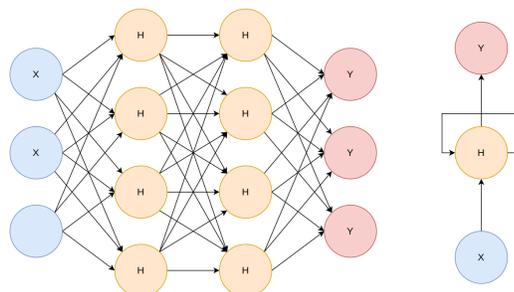


FIGURE 1.1 – Réseaux de neurones classiques et RNN.

La figure 1.1 illustre le concept de transfert de connaissance interne des RNN et LSTM. Pour les deux réseaux, X représentent les nœuds d'entrées, Y les nœuds de sortie et H les nœuds des couches cachées. Le réseau de gauche est un réseau classique où chaque nœud est entièrement connecté aux nœuds de la couche suivante. C'est un réseau de propagation en avant. Le réseau de droite est un RNN puisque le nœud de la couche caché présente une connexion vers lui-même, permettant ainsi le transfert de connaissance à l'interne.

1.4.2 LSTM

Les réseaux de neurones récurrents à mémoire à long terme (LSTM) sont une variante des réseaux de neurones récurrents (RNN). Ces LSTM sont adaptés au traitement de données en séries chronologiques avec des dépendances sur le long terme.

Ils ont été présentés pour la première fois par Hochreiter et Schmidhuber en 1997 dans l'article *Long Short-Term Memory* (Hochreiter et Schmidhuber, 1997).

Dans un LSTM, chaque nœud ou cellule est un ensemble de portes contrôlant le flux d'informations. L'entrée et la sortie de chaque cellule est donc contrôlé, permettant ainsi au LSTM de conserver l'information plus longtemps qu'un RNN classique. La mémoire d'une cellule LSTM est définie par l'historique des entrées passées qui sont sauvegardées pour pouvoir être réinjectées.

Un LSTM est composé de quatre parties :

- Une porte d'entrée contrôlant le flux d'informations entrant dans la cellule ;
- Une porte d'oubli contrôlant le flux d'informations de la cellule ;
- Une porte de sortie contrôlant le flux d'informations sortant de la cellule ;
- L'état de la cellule stockant les informations qui sont conservées dans la

cellule.

Chacune des différentes portes est contrôlée par des poids et des biais modifiés et optimisés lors de l'entraînement. L'ajustement de ces valeurs permet au LSTM de conserver certaines informations importantes tout en éliminant les informations non pertinentes.

Les RNN sont donc plus simples et faciles à comprendre, mais peuvent facilement oublier l'information à long terme. Les LSTM sont plus complexes. Dans le cas de la prédiction du rythme cardiaque et de la dilatation de la pupille, les LSTM sont privilégiés pour leurs meilleures gestions des dépendances passées. En raison de la structure interne plus complexe des LSTM, le volume de données d'entraînement consacré aux LSTM se doit d'être plus important.

1.5 L'apprentissage par transfert

L'apprentissage par transfert est une amélioration de l'apprentissage par transfert de connaissance. Cette connaissance transférée est issue d'un autre modèle ayant déjà appris sur une tâche. Cette méthode a pour but de recréer le processus d'apprentissage de l'être humain. Nos connaissances actuelles sont formées d'une accumulation de connaissances passées où nous transférons ce que nous avons appris avant, afin de l'appliquer à un nouveau problème. L'idée, en intelligence artificielle, est de faire apprendre un modèle sur une tâche générique puis de récupérer ces connaissances pour commencer un apprentissage sur une tâche spécifique.

D'après (Torrey et Shavlik, 2010), l'apprentissage par transfert permet un début d'apprentissage plus élevé au niveau des performances pour finalement atteindre une plus haute asymptote. Ce concept générique présente de nombreux avantages tels que :

- Accélération du temps d'apprentissage pour obtenir la même précision ;
- Augmentation théorique du volume de données ;
- Augmentation de la précision.

Ces avantages se retrouvent dans la littérature, une analyse des CNN (réseaux de neurones convolutifs) avec série chronologique pour la prédiction des inondations de (Kimura *et al.*, 2019) où leur modèle a été entraîné cinq fois plus rapidement avec l'apprentissage par transfert tout en proposant une MAE (Erreur absolue moyenne) 15 % inférieure en comparaison au modèle sans transfert. Le modèle utilisant l'apprentissage par transfert était capable de converger vers une solution cinq fois plus rapidement (moins d'itération nécessaire lors de l'entraînement). Du point de vue de l'architecture des réseaux de neurones, le transfert est applicable sur tout type de réseaux. D'un point de vue technique, il s'agit d'un transfert des poids de chaque neurone des couches cachées d'un modèle.

L'idée générale est d'utiliser des données facilement accessibles d'un domaine proche pour ensuite récupérer ces connaissances pour entraîner le modèle sur les données du problème à l'étude. Cependant, il n'est pas nécessaire que les deux domaines soient différents. Il est possible d'appliquer l'apprentissage par transfert dans le même domaine.

Dans notre projet, nous nous concentrons sur l'optimisation de modèles préentraînés. Cette technique de préentraînement des modèles est courante. Le jeu de données ImageNet par exemple, est souvent utilisé pour l'entraînement de modèle de classification d'images. ImageNet, jeu de donnée de plus de 14 millions d'images étiquetées, permet alors la conception de réseaux de neurones de classification d'objets spécifique présentant peu d'images, en corrélation avec l'apprentissage par transfert par exemple.

CHAPITRE II

MODÈLE MULTIMODAL POUR LA PRÉDICTION DU COMPORTEMENT

Le concept de modèle multimodal est important dans le domaine de l'intelligence artificielle pour sa capacité à comprendre et analyser différents types de données provenant de sources différentes (ou non) simultanément. Ce chapitre présente donc la théorie derrière ces modèles avant de présenter le modèle que nous proposons.

2.1 La théorie

Le concept de modèle multimodal est issu du domaine de sciences cognitives. Ces sciences ont pour objectif de comprendre comment les humains perçoivent, interprètent, interagissent et analysent le monde qui nous entoure. Les sciences cognitives de la fin du XXe siècle ont analysé et étudié ce comportement humain afin de mieux le comprendre. Les différentes études du cerveau ont pour objectif de comprendre sa capacité de traitement. Celui-ci est, par définition, un outil de traitement de données multimodales sensorielles. Cette multimodalité de l'humain se base sur la vue, l'ouïe, le toucher et l'odorat afin de pouvoir obtenir une perception unique et uniforme de notre entourage.

L'étude (Baltrusaitis *et al.*, 2019) analyse l'utilisation des modèles multimodaux. Différentes solutions au niveau de la conception des modèles sont présentées :

- Modèles de fusion précoce (*Early Fusion*) : Ces modèles combinent les caractéristiques de différentes modalités au niveau de l'entrée avant de les alimenter dans le modèle.
- Modèles basés sur les graphes (*Graph-based Models*) : Ces modèles représentent la relation entre les modalités sous forme de graphe et apprennent à prédire en fonction de la structure du graphe.
- Modèles basés sur l'apprentissage par transfert (*Transfer Learning-based Models*) : Ces modèles transfèrent les connaissances acquises d'une tâche à une autre tâche impliquant plusieurs modalités.

Ces différents modèles présentés sont utilisés dans la détection d'images, reconnaissance vocale, reconnaissance d'émotion ou encore le traitement du langage naturel.

L'objectif de l'intelligence artificielle a toujours été de vouloir imiter notre intelligence dans le but de la retranscrire dans une machine. L'analyse de données multimodales afin d'obtenir une observation capable de pondérer les valeurs de chaque modalité forme donc le début du concept des modèles multimodaux.

L'apprentissage profond (*Deep Learning*) abordé par Hilton et al. (Hinton *et al.*, 2006) est un sous-domaine de l'apprentissage machine impliquant les réseaux de neurones utilisant des données complexes, un grand volume de données et des architectures dites profondes. La théorie indique que l'apprentissage profond est capable de modéliser les relations entre les données. Cette capacité de traitement de grand volume de données complexes est ce que nous recherchons pour l'analyse et prédiction du comportement. Cette représentation est particulièrement efficace pour la détection d'images, la reconnaissance vocale et le traitement du langage naturel. Les modèles multimodaux rentrent donc dans cette catégorie où le volume de données est extrêmement important pour l'efficacité des modèles. Le volume de

données est important pour ces modèles puisqu'ils ont composé de plusieurs modèles. Là où un modèle a besoin de X données, un modèle multimodal regroupant plusieurs modalités a besoin de plusieurs fois X volumes de données.

Les modèles multimodaux consistent en plusieurs entrées de modalité. Chaque entrée est un modèle de traitement d'une modalité spécifique. Chaque modèle utilise généralement une forme de réseaux récurrents (Cui *et al.*, 2020) pour le traitement des données avant de fusionner chaque modalité afin d'obtenir une prédiction finale. L'utilisation de réseaux récurrents n'est pas garantie, mais ceux-ci sont souvent utilisés pour le traitement et l'analyse du comportement humain.

La littérature présente donc une architecture complexe de plusieurs modèles couplés à une couche de fusion permettant le traitement de plusieurs modalités partageant un contexte commun. Cette architecture est donc capable d'analyser les corrélations, les similitudes et autres facteurs communs de chaque modalité dans l'objectif de fournir une réponse considérant ces différentes modalités. Dans le cadre de notre projet, le résultat est une prédiction du comportement du pilote en fonction de tous les comportements observés.

On constate que les modèles multimodaux proposés dans la littérature ne se concentrent pas sur la modélisation du comportement humain, mais plutôt sur des tâches spécifiques comme la reconnaissance vocale ou encore la reconnaissance d'images. Cependant, le comportement humain étant de nature multimodale, il convient de s'intéresser à ces modèles pour la modélisation de ce dernier.

2.1.1 Analyse neuronale

Les modèles multimodaux sont spécialisés dans le traitement de différentes métriques en parallèle, afin de les intégrer. L'architecture générale de ces modèles dépend de la tâche à l'étude. Il existe cependant une architecture globale divisée

en trois parties :

- Un réseau neuronal par modalité ;
- Une couche de fusion pour combiner les sorties des réseaux précédents ;
- Une couche de sortie.

Les différents réseaux neuronaux composant le modèle multimodal s'occupent du traitement des données d'entrées. Ce sont ces réseaux qui définissent le format ainsi que les métriques que le modèle accepte. Puisque ces modèles sont uniques et différents, il est possible que chaque métrique d'entrées soit de type différent comme du texte, des images ou encore de l'audio. Il est seulement important que les réseaux utilisés soient spécialisés dans leur traitement d'entrées afin de pouvoir nourrir la couche de fusion.

La couche de fusion combine les sorties de chaque modèle. Cette couche a donc comme responsabilité de construire une représentation conjointe des données d'entrées. Cette couche devient alors la couche d'entrées du réseau effectuant le traitement final pour obtenir la prédiction voulue.

La figure 2.1 illustre le modèle multimodal générique présenté dans cette analyse neuronale. Le modèle présenté contient trois modalités, le nombre n'est pas limité. Chaque boîte orange est un réseau spécialisé par modalité. La couche de fusion est représentée au milieu par la couleur bleue. Le traitement final du réseau multimodal s'effectue dans les couches neuronales du réseau de la boîte verte avant de sortir la prédiction finale du modèle.

La fusion des modèles est une partie importante de notre modèle. Différents types de fusion existent ((Ramachandram et Taylor, 2017)) et se différencient par le moment et la manière dont les modalités sont combinées. Les trois principaux types de fusions sont les suivants :

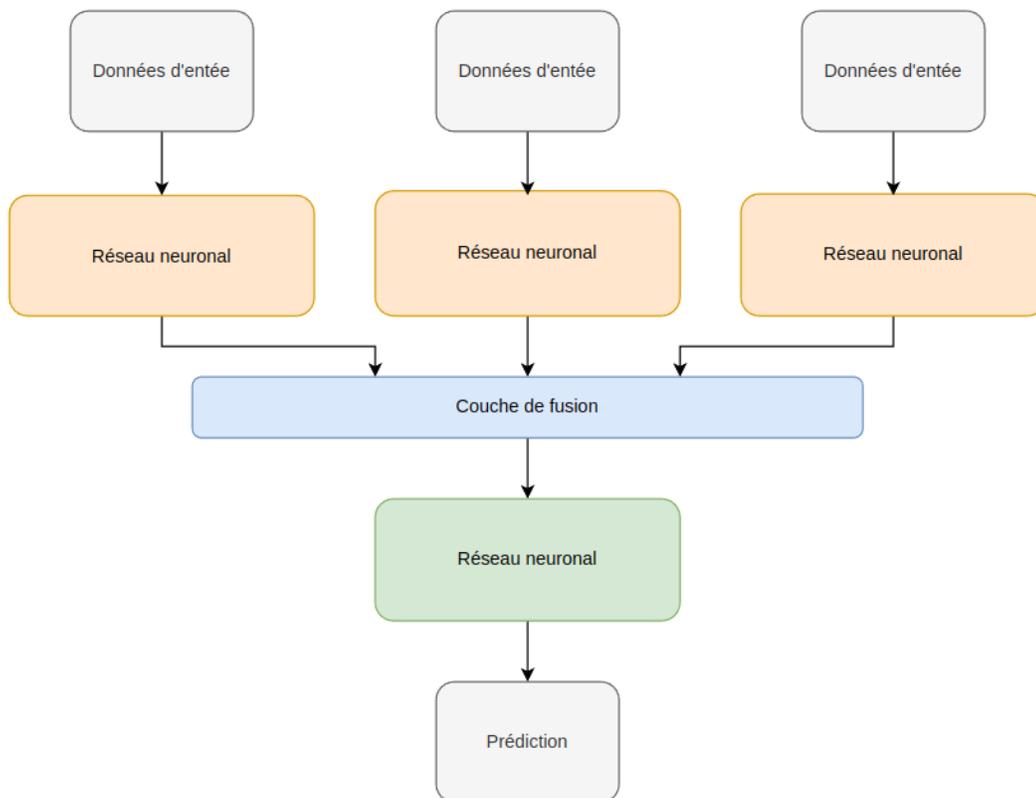


FIGURE 2.1 – Modèle multimodal théorique.

- Fusion précoce : Combinaison des caractéristiques de différentes modalités au niveau de l'entrée avant de les alimenter dans le modèle.
- Fusion tardive : Combinaison des sorties de plusieurs modèles, chacun traitant une seule modalité, au niveau de la sortie.
- Fusion hybride : Combinaison des techniques de fusion précoce et tardive, offrant une plus grande flexibilité dans la combinaison des modalités.

La fusion présentée dans la figure 2.1 est donc une fusion tardive des sorties des différents modèles.

2.2 Notre méthodologie

Ce chapitre se concentre sur la présentation et l'analyse des différents réseaux de neurones utilisés pour la prédiction du comportement des pilotes. La présentation des données ainsi que leur analyse sera présentée dans le prochain chapitre. Nous présentons cependant les différents réseaux et sous-réseaux en fonction de leurs données afin de pouvoir justifier nos différentes segmentations et séparations de modalités.

2.2.1 Modèle multimodal pour la prédiction du comportement

La solution proposée et implémentée a pour but de comprendre automatiquement les tâches de pilotage ainsi que les différents comportements des pilotes. Bien que dépendante des données collectées lors de l'expérience, la structure générale du modèle peut être présentée afin de pouvoir identifier les métriques et modalités importantes.

L'architecture complète de notre modèle de prédiction de comportements présenté à la figure 2.2 met en avant trois modalités différentes issues des données cognitives, le contexte et les manœuvres associées. La prédiction finale est un vecteur représentant la charge mentale, l'émotion, etc... dans lequel sera le pilote à un instant $t + 1$ sachant tout ce qui s'est passé avant. Ce vecteur contient également l'action prédite que fera le pilote à l'instant $t + 1$.

Les trois modalités orange, vertes et bleues sur la figure 2.2 représentent un ensemble de sous modèles de prédiction par catégorie. Ces trois sections ne sont là que pour aider la visualisation et la compréhension de l'architecture globale.

Le modèle multimodal final fusionne chacune de ces modalités afin d'obtenir un modèle complexe de prédiction. Cette division en sous-section nous permet d'éva-

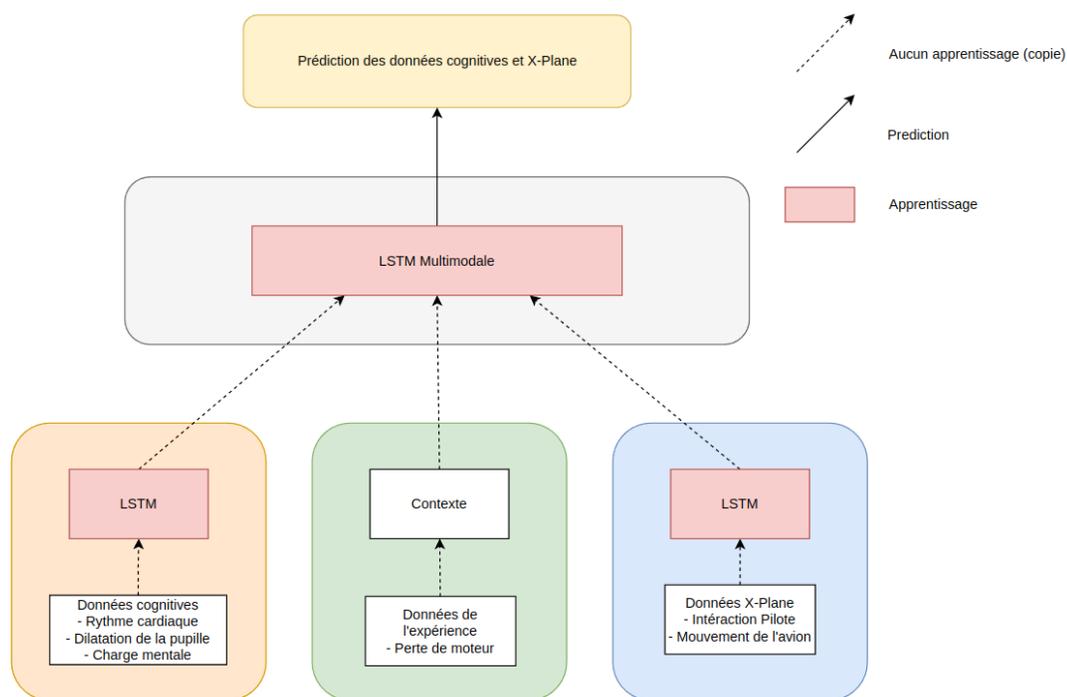


FIGURE 2.2 – Modèle multimodal hybride pour la prédiction du comportement.

luer chaque jeu de données indépendamment des autres afin de valider progressivement l'architecture du modèle complet.

Nous parlons depuis le début de la prédiction du comportement du pilote, cependant nous avons aussi ajouté la prédiction des données *X-Plane* dans notre figure 2.2. Les données de simulation propres aux interactions des pilotes avec l'avion sont prédites par notre modèle. Nous parlons là de prédire le mouvement des pédales de l'avion, du manche tenu par le pilote ou encore de la manette d'accélération.

La modalité verte se différencie des autres puisqu'il ne s'agit pas d'un modèle appris. Les données du contexte sont là afin de fournir de l'information annexe aux autres modèles. Cette modalité a été prise en considération, car elle représente une

partie importante de notre étude. Puisque nous souhaitons corréler les événements ayant eu lieu avec les augmentations de rythme cardiaque et la dilatation de la pupille afin de valider leur impact sur la charge mentale du pilote.

Notre modèle multimodal 2.2 s'est inspiré du modèle théorique présenté à la figure 2.1. Les différents réseaux neuronaux de la théorie se retrouvent dans nos modalités orange, vertes et bleues. Celle-ci présente de l'apprentissage ou non dépendamment de la modalité à l'étude. Certaines de ces modalités peuvent définir un contexte général sans nécessiter l'apprentissage d'un réseau. Ensuite, la couche LSTM multimodale de notre modèle regroupe la couche de fusion et le réseau neuronal du modèle théorique.

Notre approche générale pour ce qui est de la prédiction est basée sur les données. Notre objectif est d'observer les données pour en ressortir un modèle d'apprentissage automatique pour pouvoir prédire l'évolution de celle-ci. Les réseaux bayésiens auraient pu être utilisés dans le cadre de notre projet, mais la nature séquentielle des données nous a poussés à choisir les réseaux LSTM.

Afin de pouvoir comprendre la conception des différents modèles suivants, il faut comprendre que la collecte de données ne fait pas partie de mon projet. Cette collecte de donnée a été prise en charge par un autre étudiant du projet. Cette séparation des tâches nous a donc forcés à analyser le format des données attendues afin de pouvoir concevoir l'architecture des modèles. Ainsi le format et les *features* disponibles pour notre projet sont limités par l'expérience décrite dans le prochain chapitre. Cependant, il est important de noter que toutes les données sont sur des séries chronologiques.

Chaque sous-modèle et modalité est présenté dans sa propre sous-section.

2.2.2 Sous-modèles de prédiction cognitive

Comme il a été défini dans le premier chapitre, nous utilisons les réseaux de neurones LSTM pour l'ensemble de notre projet. Les prédictions des données cognitives sont sensibles au volume de données ainsi que leur qualité. En raison des conclusions portées par la littérature et ne pouvant pas contrôler la qualité des données, nous avons décidé d'opter pour l'apprentissage par transfert afin de pallier notre manque de données.

Ce transfert de connaissances est utilisé pour la prédiction du rythme cardiaque ainsi que la dilatation de la pupille, en raison de la disponibilité de la donnée publique que nous allons détailler dans le prochain chapitre. La figure 2.3 présente les trois modèles que nous avons développés dans le cadre de la prédiction cognitive. Chaque modèle respecte l'architecture globale de tout modèle d'intelligence artificielle présenté. Une entrée de données, un traitement neuronal puis enfin une prédiction.

Le modèle de prédiction du rythme cardiaque ainsi que le modèle de prédiction de la dilatation de la pupille présentent deux entrées de données dans leur réseau neuronal. Cette entrée de données de la boîte jaune représente l'apprentissage par transfert. Afin d'augmenter le volume de données disponibles à la prédiction, un jeu de données publiques est utilisé afin de préentraîner les deux modèles. Puisque le contexte de notre apprentissage par transfert est identique, la même architecture est utilisée. La méthodologie suivante est appliquée :

- 1. Envoi des données publiques dans le modèle ;
- 2. Entraînement du modèle ;
- 3. Envoi des données de l'expérience ;
- 4. Entraînement du modèle.

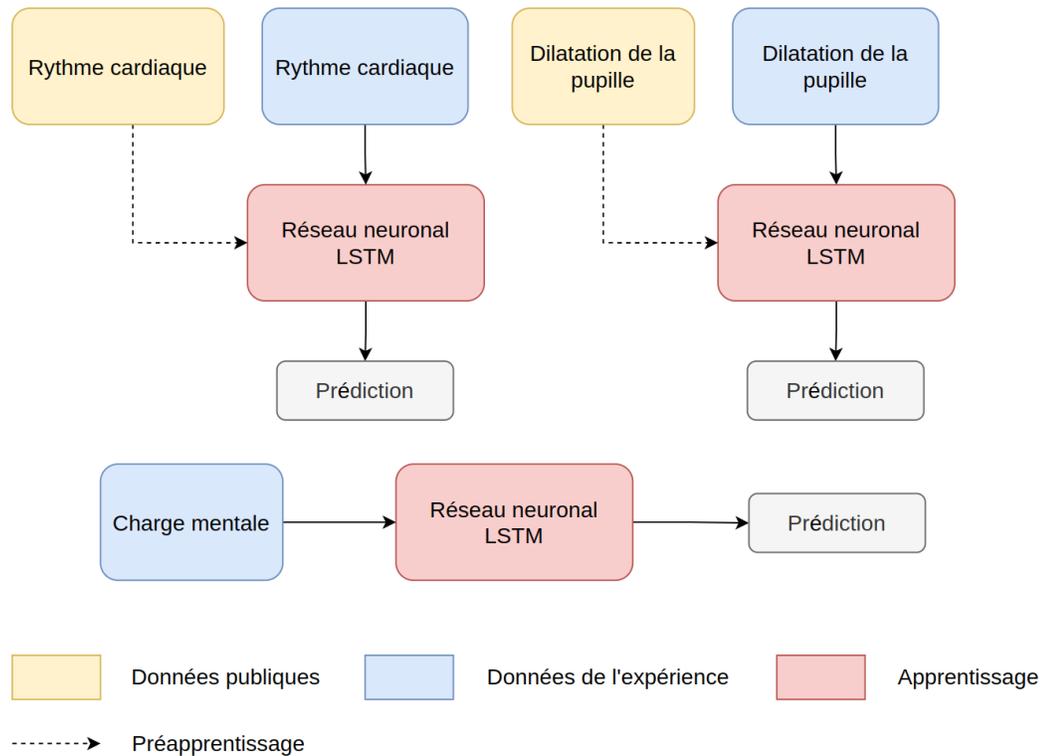


FIGURE 2.3 – Architecture des sous-modèles de prédiction cognitive.

Nous n’effectuons pas d’apprentissage par transfert pour ce qui est de la prédiction de la charge mentale, ceci sera détaillé dans le prochain chapitre.

2.2.3 Prédiction du rythme cardiaque

La prédiction du rythme cardiaque est la première modalité mise à l’étude. Celle-ci fut sélectionnée pour sa simplicité au niveau du format de données, mais aussi sa compréhension commune de tous. Il est plus simple d’expliquer l’évolution du rythme cardiaque que l’évolution de la dilatation de la pupille. Le rythme cardiaque est un phénomène physique que chaque personne connaît et comprend facilement.

L'objectif final étant la fusion des sous modèles afin de concevoir le modèle multimodal, ce modèle de prédiction du rythme cardiaque décide du format et de certains paramètres que nous allons imposer au futur modèle. Un des paramètres dont nous n'avons pas parlé encore est la taille de la fenêtre d'historique que nous donnons aux couches LSTM.

La nature chronologique des données implique de sélectionner la taille de l'historique de données envoyées à la couche LSTM. Celle-ci est modifiable lors de la création des réseaux, mais cela impacte la mémoire interne de cette couche LSTM. Cet historique est une fenêtre représentant un ensemble de valeurs successives ayant eu lieu avant la valeur au temps t .

Prenons comme exemple la figure 2.4. Les données sont représentées sous forme de série chronologique. Le jeu de données complet est une suite de 10 chiffres croissants. Une fenêtre chronologique sélectionnant cinq valeurs est ajoutée en rouge sur la première ligne de données. Cette première ligne de données représente le traitement de la valeur 6 en considérant les cinq valeurs précédentes comme historiques. Pour le traitement de la valeur 7, la fenêtre se déplace d'une case vers la droite pour toujours sélectionner les cinq dernières valeurs.

Ce concept de fenêtre glissante (*sliding window*) est fortement utilisé avec les couches LSTM afin de fournir un historique fixe et constant à chaque itération. La taille de cette fenêtre glissante est donc la première métrique que nous avons analysée puisque nous souhaitons uniformiser tous nos modèles. Cette uniformisation est nécessaire pour la fusion des modèles.

La prédiction du rythme cardiaque consiste à utiliser les valeurs précédentes du rythme cardiaque pour en définir les suivantes. Pour le projet, nous nous concentrons sur les mesures directes du rythme cardiaque. Le rythme cardiaque mesuré est simplement le nombre de battements par minute, 60 battements par minute

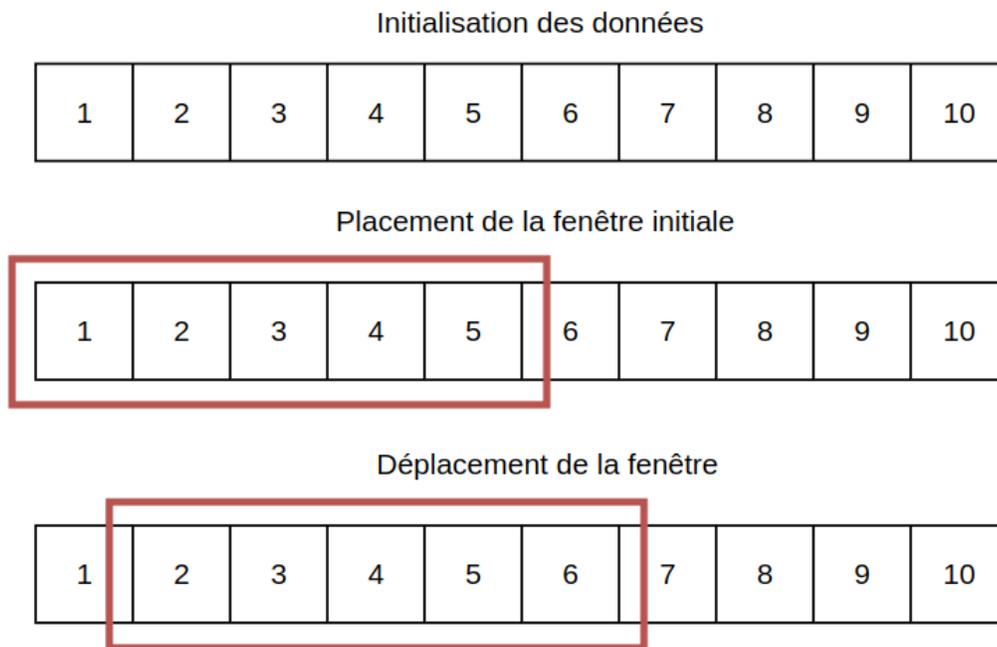


FIGURE 2.4 – Fenêtre coulissante sur un jeu de données.

par exemple. Le format de la donnée à l'étude est donc une valeur de rythme cardiaque collectée toutes les secondes. L'architecture du modèle est par conséquent, elle aussi, simple : une couche d'entrée, une couche de sortie et une seule couche cachée LSTM.

La figure 2.5 représente les différentes couches neuronales de notre modèle de prédiction du rythme cardiaque, les différentes métriques (nombre de paramètres d'entrés, nombre de paramètres des couches LSTM) seront présentées et expliquées dans le prochain chapitre. Puisque le rythme cardiaque est une mesure présentant une faible variation à l'échelle de la seconde, un réseau simple ne contenant qu'une seule couche LSTM est suffisant pour prédire le rythme cardiaque au temps $T + 1$ seconde.

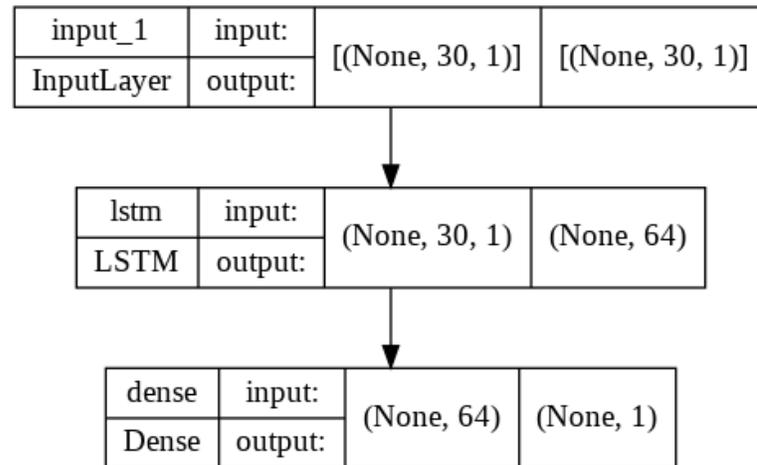


FIGURE 2.5 – Couche neuronale du modèle de prédiction du rythme cardiaque.

2.2.4 Prédiction de la dilatation de la pupille

La dilatation de la pupille étant capable de varier en moins de 1000 ms (Hyönä *et al.*, 1995), le réseau neuronal est plus complexe que celui du rythme cardiaque. Les couches LSTM sont encore utilisées pour leur capacité de traitement des données sur série chronologique. Nous avons privilégié l’approche de superposer les couches LSTM afin de permettre au réseau de traiter les données. Le nombre de couches superposées sera expliqué dans le prochain chapitre.

L’analyse et le suivi des mouvements oculaires sont composés de plusieurs mesures, telles que la durée de fixation, l’amplitude de la saccade, la dilatation de la pupille, chacune fournissant des informations uniques sur la perception visuelle. Les différents outils d’analyse de mouvements des yeux collectent toutes ces données en raison de leur forte corrélation. Bien que nous parlons seulement de la dilatation de la pupille, car la littérature confirme sa corrélation avec nos autres modalités cognitives, ces différentes données ont un impact sur la dilatation. L’analyse de ces données est donc complexe (Holmqvist *et al.*, 2011), notre modèle de prédiction

de la dilatation de la pupille le sera aussi.

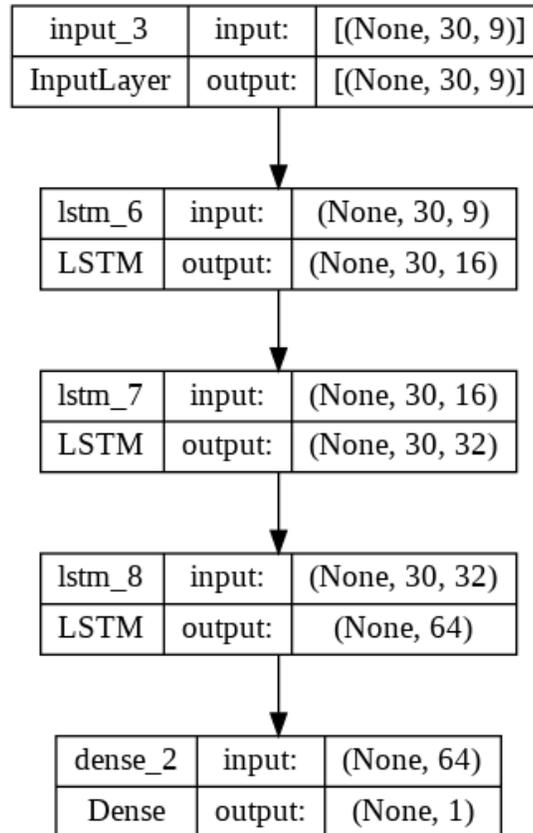


FIGURE 2.6 – Couche neuronale du modèle de prédiction de la dilatation de la pupille.

La figure 2.6 représente les différentes couches neuronales de notre modèle de prédiction de la dilation de la pupille. Il s’agit d’une architecture neuronale de superposition de trois couches LSTM entourées d’une simple couche d’entrée et d’une de sortie.

2.2.5 Prédiction de la charge mentale

La charge mentale est une métrique assez différente des deux précédentes. Elle est généralement évaluée à l’aide de mesures d’auto-évaluation, de mesures de perfor-

mance comportementale ou physiologique. Dans le cadre du projet, nous utilisons les données issues d'un électroencéphalogramme (EEG). La structure neuronale de ce modèle est fortement dépendante de l'outil utilisé pour la collecte de données. Certains outils retournent des données brutes, comme la plupart des outils d'analyse oculaire, mais d'autres retournent simplement une mesure interprétée. Le modèle peut donc ressembler au modèle de prédiction du rythme cardiaque ou au modèle de prédiction de la dilatation de la pupille. Comme nous allons le voir dans le prochain chapitre, nos données sont interprétées par le logiciel de l'outil de prélèvement. Notre modèle est donc simple.

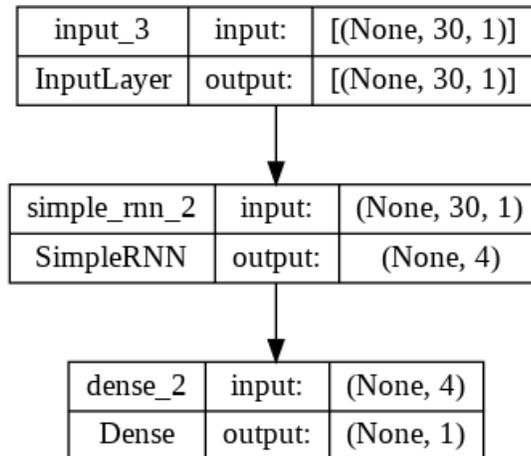


FIGURE 2.7 – Couche neuronale du modèle de prédiction de la charge mentale.

La figure 2.7 représente les différentes couches neuronales de notre modèle de prédiction de la dilatation de la pupille. Il s'agit d'une architecture neuronale RNN simple d'une seule couche entourée d'une simple couche d'entrée et d'une de sortie.

2.2.6 Sous modèle du contexte

Le modèle du contexte existe simplement comme support aux autres modèles au niveau multimodal. Si l'on reprend notre figure 2.2, le modèle de contexte

est représenté par la modalité verte (la modalité qui se trouve au milieu). Il est important de la considérer comme une modalité à part entière afin de respecter notre architecture générique. Ce contexte se différencie des autres modalités par son absence d'apprentissage.

Dans le cadre de notre projet, le contexte se définit par plusieurs métriques externes. Nous définissons le contexte comme étant les événements ou actions hors du contrôle du pilote et ne pouvant pas directement être anticipés par le pilote. Celui-ci a donc un fort impact sur les actions que le pilote doit prendre lors de son décollage. C'est pour cela que nous avons ajouté cette modalité dans notre modèle. Dans notre étude, nous n'avons qu'une seule métrique de contexte qui sera présentée dans le prochain chapitre.

Au niveau neuronal, ce modèle est simplement une couche faisant office d'entrée et de sortie qui est envoyée directement dans la couche de fusion. Elle permet d'informer le réseau lors de l'apprentissage.

2.2.7 Sous modèle de l'environnement de simulation

L'environnement de simulation est évidemment une métrique importante si l'on évalue les actions d'un pilote. Les métriques sensorielles sont importantes, mais il faut pouvoir les corrélérer à des actions physiques pour pouvoir comprendre l'origine de l'augmentation du rythme cardiaque d'un pilote par exemple.

Le premier groupe des métriques que nous avons sélectionnées sont les conditions météorologiques qui font partie intégrante de la simulation. Les pilotes et les contrôleurs aériens surveillent de près les conditions météorologiques telles que la vitesse et la direction du vent, la visibilité, la couverture nuageuse, les précipitations et les orages. Des conditions météorologiques défavorables peuvent entraîner des retards, des annulations de vols et impactent donc les manœuvres que le pi-

lote effectuée. En plus des données météorologiques, les interactions du pilote avec l'avion font partie de cette modalité. Les mouvements du manche, des pédales et des manettes des gaz sont des composantes importantes qui ont été identifiées par les partenaires du projet.

L'objectif du sous-modèle de l'environnement est, lui aussi, la prédiction. Puisque cette modalité varie en fonction de la donnée, nous avons décidé d'opter pour une architecture simple semblable à celle de la prédiction du rythme cardiaque.

Il s'agit de la modalité bleue (la modalité à droite) dans la figure 2.2.

2.3 Conclusion

Le modèle multimodal est donc un regroupement de différents modèles spécialisés dans leur modalité. Chaque sous-modèle défini peut être interprété comme un expert du domaine évaluant ses données pour partager ses résultats avec les autres. Dépendamment de la nature de la donnée et de la complexité de celle-ci au niveau de la science, les modèles sont plus ou moins complexes au niveau neuronal.

Chaque modèle, à l'exception du contexte, est entraîné afin de prédire une ou plusieurs métriques. Le nombre et la nature de ces métriques ne sont pas importants pour comprendre le concept et le fonctionnement d'un modèle multimodal. Ces métriques sont fortement impactées par la donnée disponible lors de l'entraînement des modèles.

Une fois que les modèles de chaque modalité ont prouvé leur capacité de prédiction indépendamment des autres, le modèle multimodal peut être conçu en utilisant l'architecture de chaque modèle combinée à une couche de fusion. Cette couche de fusion est généralement une simple mise en commun des sorties des modèles.

Le modèle multimodal final que nous présentons et analysons dans les prochains

chapitres est présenté dans la figure 2.8.

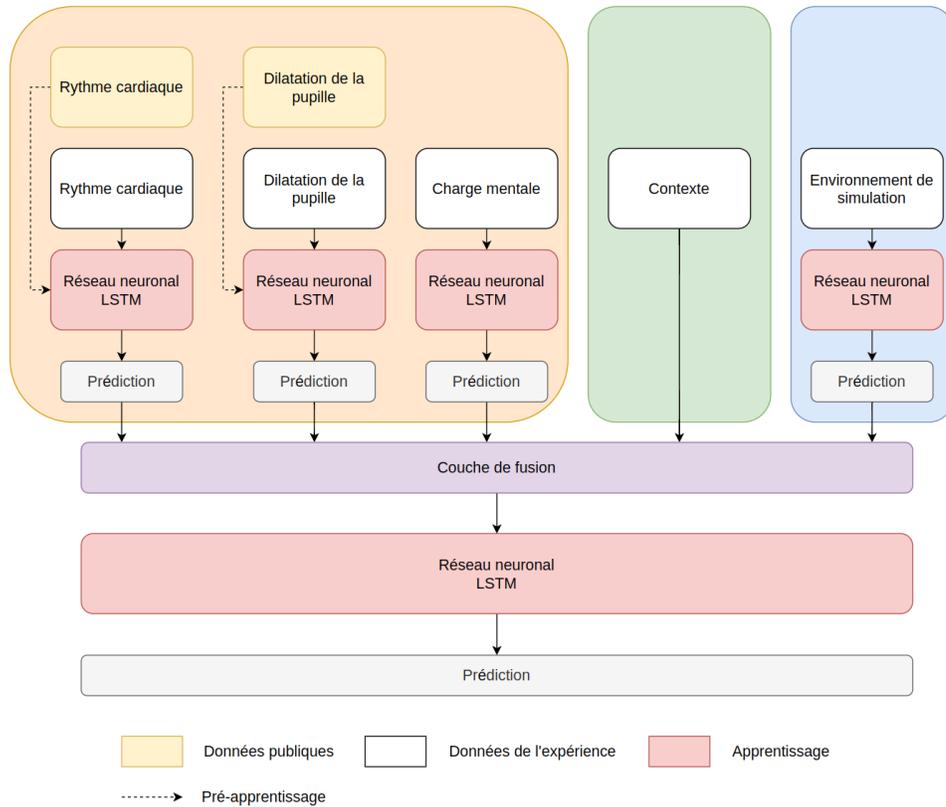


FIGURE 2.8 – Modèle multimodal détaillé.

Ce modèle est un regroupement de quatre modèles de prédiction, deux modèles de préentraînement grâce à l'apprentissage par transfert ainsi qu'un modèle simple de contexte.

Le prochain chapitre se concentre sur l'analyse et la collecte des données utilisées dans notre projet.

CHAPITRE III

EXPÉRIMENTATION

3.1 Mise en contexte

Notre analyse du comportement des pilotes est une partie d'un plus grand projet. Ce projet nommé *PilotAI* a pour objectif de concevoir une intelligence artificielle capable de piloter un avion.

L'architecture globale du projet se divise en trois WP. Ces WP (voir figure 4.1) représentent chacun une tâche spécifique. Le WP1, représentant le modèle de pilotage, est chargé d'établir l'ontologie des tâches nécessaires au pilotage de l'avion ainsi qu'analyser les différents profils de pilote. Le WP2 se concentre sur l'analyse du comportement du pilote ainsi que la collecte de données nécessaires à l'entraînement des différents modèles développés. Le dernier WP se concentre sur l'élaboration d'un modèle de déviation permettant d'évaluer les décisions prises et y attribuer un score de validité.

Notre modèle multimodal se place dans le WP2. La partie génération de données ainsi que la mise en place de sa collecte ne font pas partie de mon mandat sur le projet. Cette partie a été en charge par un autre étudiant de l'UDEM (Université de Montréal).

3.1.1 Partenaires du projet PilotAI

Le projet PilotAI est difficilement réalisable seul. De nombreux partenaires ont contribué et continuent de contribuer activement au projet. Puisque ce projet nécessite une prise de mesures sur des humains, la recherche complète est validée par un certificat d'éthique approuvé par le comité d'éthique de la recherche en sciences et en santé (CERSES) de l'Université de Montréal.

Différentes entreprises participent à l'évolution du projet ainsi que l'UDEM. Ce projet regroupe un peu plus de 25 personnes travaillant activement sur le projet.

Les entreprises offrant leur point de vue d'expert sur le domaine de l'aviation ainsi que de nombreux contacts avec des pilotes afin de pouvoir collecter des données sont les suivantes :

- CAE : "Leader mondial dans le domaine des technologies de simulation et de modélisation et des solutions intégrées de formation destinées à l'aviation civile, aux forces de défense et de sécurité, et du domaine médical, dont le siège social est situé à Montréal." (Wikipedia, 2023b)
- Bombardier Inc. : "Bombardier Inc. est une entreprise multinationale canadienne dont le siège social est situé à Montréal, au Québec. Depuis 2021, l'entreprise est recentrée sur les avions d'affaires." (Wikipedia, 2023a)

L'expérience de collecte de données a été gérée par un étudiant de l'UDEM avec l'aide des entreprises. Les entreprises ont participé à la vérification et validation de l'expérience.

3.1.2 Objectif de l'expérience

L'objectif de l'expérience est de générer des données utilisables pour l'ensemble des modèles du projet Pilot AI. Puisque l'étude globale de ce WP est l'analyse sensorielle des pilotes, il était nécessaire de contrôler l'environnement afin de pouvoir forcer certaines réactions sensorielles des pilotes.

Premièrement, chaque pilote ayant participé se voyait présenter un objectif simple. Dans le simulateur, à partir de la piste de décollage 06L de l'aéroport international Pierre-Elliott-Trudeau de Montréal (YUL) celui-ci effectue les actions suivantes :

- Relâcher les freins de l'avion ;
- Faire décoller l'avion ;
- Atteindre une altitude de 3000 pieds sans utiliser l'autopilote.

Ces critères relativement simples et fixes permettent d'uniformiser le contexte de chaque pilote. La recherche effectuée par Antoine présente des scénarios inconnus non communiqués avec le pilote afin de maximiser la charge mentale du pilote. Un total de six scénarios (3.1) ont été conçus où l'heure et les conditions météorologiques varient. De plus, une panne de moteur est planifiée dans certains scénarios afin d'évaluer les mesures prises par le pilote.

Ces scénarios sont donc identiques pour chaque pilote. Une session de 20 minutes de décollages normaux suivies de 20 minutes de décollages difficiles où une panne de moteur a lieu. Chaque participant de l'expérience va alors générer le même volume de données.

TABLEAU 3.1 – Scénarios des expériences

Scénario	Heure	Condition Météorologique	Panne de moteur
1	13 h 45	Pas de vents, pas de nuages	Non
2	06 h 00	Nuage à 2700 pieds, pluie	Non
3	21 h 00	Pas de vents, pas de nuages	Non
4	05 h 30	Pas de vents, pas de nuages	Oui
5	06 h 00	15 nœuds de <i>crosswind</i>	Oui
6	06 h 00	Faible visibilité, pluie	Oui

3.1.3 Participants de l'expérience

Il est important de présenter les profils des différents participants de l'expérience. Chaque étude demandant la participation de personnes issues d'un milieu spécifique, risque de tomber dans un biais volontaire ou non (Hammer *et al.*, 2009). Puisque notre modèle de prédiction du comportement se base sur les données collectées lors de l'expérience, il est nécessaire de valider l'expérience des pilotes par leur nombre d'heures de vol ou encore leur ancienneté dans la profession par exemple.

Au total, l'expérience a regroupé 13 pilotes ayant complété un total de 136 décollages. Tous les participants sont des hommes avec une moyenne d'âge de 36 ans (± 8 ans) et 7 des participants étant des pilotes de profession, l'expérience présente un certain biais au niveau du genre et de l'expérience de chacun. Les participants qui ne sont pas des pilotes ont de l'expérience en simulateur de vol (98h en moyenne) (Antoine, 2022). Nous sommes conscients du biais potentiel de notre jeu de données.

3.2 Outils de prises de mesures

Nos modèles étant définis par la donnée, les outils de prises de mesures utilisés ont un rôle important dans ce que nous pouvons faire par la suite. Les différents outils ont été validés avec les partenaires afin d'assurer que leur installation ne gêne pas les pilotes lors de l'expérience. La précision des mesures est aussi un facteur important de la sélection des outils. Les données récupérées doivent être précises.

3.2.1 Polar H10

Le rythme cardiaque est une mesure relativement simple à obtenir. De nos jours, il est possible de trouver des capteurs cardiaques dans les montres connectées. Il a été décidé d'utiliser le Polar H10 puisqu'il propose une meilleure précision.

Le Polar H10 3.1 est un capteur cardiaque s'accrochant autour du torse de la personne. Cette méthode peut être considérée comme intrusive en comparaison à une montre connectée. L'ensemble des acteurs du projet ont jugé l'intrusivité de cette technique négligeable. Cet appareil offre des données plus précises.



FIGURE 3.1 – Capteur fréquence cardiaque polar H10.

Lors de l'expérience, cet appareil mesure le rythme cardiaque et communique par

Bluetooth avec l'ordinateur. Les données récupérées sont les suivantes :

- *Time* : Horodatage de la mesure
- *HR* : Valeur du rythme cardiaque (battements par minute)

3.2.2 Gazepoint GP3

La dilatation de la pupille est une mesure complexe à obtenir. De nombreux outils et caméra se concentrent sur la position du regard sur un écran. La valeur brute de dilatation de la pupille est rarement offerte directement par les outils. Finalement, les responsables de l'expérience ont choisi le Gazepoint GP3. Cet appareil offre les données brutes de mesure de la dilatation de la pupille ainsi que de nombreuses autres mesures. Cet outil compact a été validé par les partenaires pour son utilisation dans un environnement de simulation.

Le Gazepoint GP3 3.2 est une caméra placée devant l'écran. Le logiciel de cet outil détecte le sujet devant lui et se calibre automatiquement.



FIGURE 3.2 – Un système de suivi oculaire.

Cet outil de mesure est un outil parfait pour la collecte de données. Une étude d'analyse de la qualité des mesures (Brand *et al.*, 2020) faite avec le Gazepoint

GP3 par des chercheurs est similaire aux résultats obtenus par des experts en suivi oculaire.

Lors de l'expérience, cet appareil mesure la dilatation de la pupille en temps réel et communique par USB avec l'ordinateur. Les données récupérées sont les suivantes :

- *Time* : Horodatage de la mesure ;
- *LPD* : Dilatation de la pupille gauche ;
- *LPS* : Facteur de calibration de la pupille gauche ;
- *LPV* : Validateur de mesure de la pupille gauche ;
- *RPD* : Dilatation de la pupille droite ;
- *RPS* : Facteur de calibration de la pupille droite ;
- *RPV* : Validateur de mesure de la pupille droite ;
- *BPOGX* : Position du point de regard en X ;
- *BPOGY* : Position du point de regard en Y ;
- *BPOGV* : Validateur de la mesure du point de regard.

3.2.3 NCO EEG Headset

Le dernier outil de mesure sensorielle de l'expérience est le *NCO EEG Headset*. Ce casque est un électroencéphalogramme analysant l'activité du cerveau en temps réel. L'appareil calcule directement la charge mentale et nous avons accès aux valeurs transformées et non aux valeurs brutes captées par l'appareil. Pour ce faire, le module doit utiliser un logiciel tiers appelé Mentor (Chaouachi *et al.*, 2015). Ce logiciel est un module du logiciel NCO, logiciel propriétaire du laboratoire BMU issu de plusieurs années de recherche dans un vaste programme appelé NCO (Benlamine *et al.*, 2016).

Ce casque est évidemment intrusif. Les différentes études effectuées dans le projet

Pilot AI ont comme objectif d’obtenir la charge mentale du pilote à partir des autres capteurs. Dans le cadre de l’expérience, l’intrusivité du casque EEG est acceptée.

Lors de l’expérience, cet appareil communiqué par USB avec l’ordinateur. Les données récupérées sont les suivantes :

- *Time* : Horodatage de la mesure
- *Workload* : Valeur de charge mentale

3.2.4 *X-Plane*

Finalement, le simulateur *X-Plane* 11 a été choisi pour son réalisme et son intégration de plugin externe. L’entièreté des vols effectués par les pilotes a eu lieu sur l’Airbus A320. Le choix de cet avion a été validé par les partenaires en raison de sa popularité au sein des pilotes participants à l’expérience.

L’objectif initial de l’utilisation d’un simulateur est de collecter les données relatives aux interactions du pilote avec l’avion. Comme nous l’expliquons dans la section suivante (sous-section 3.3.3), ce simulateur n’a pas fonctionné comme prévu. L’expérience n’a pas pu collecter les données d’interactions du pilote.

Les données récupérées sont les suivantes :

- *Speed brake* : État des freins de vitesse ;
- *Landing gears handle* : État du train d’atterrissage ;
- *Wind speed* : Vitesse du vent réelle, en nœuds ;
- *Wind direction* : Direction du vent, en degrés ;
- *Wind speed indicated* : Vitesse du vent indiqué dans le cockpit, en nœuds ;
- *True heading* : Cap de l’avion, mesuré en angles d’Euler dans l’axe du corps ;

- *Magnetic heading* : Le cap magnétique de l’avion, en degrés ;
- `textitDynamic pressure` : Pression dynamique, en livres par pied carré.

3.3 Analyse des données

L’expérience du projet a eu lieu au cours du mois de mai 2022. Cependant, les différents outils utilisés ont été validés six mois plus tôt. Ce délai d’attente entre la validation des outils et l’expérience nous a permis d’étudier l’architecture de nos modèles, mais aussi de collecter différents jeux de données publiques. Ces données supplémentaires sont utilisées pour le transfert d’apprentissage.

L’analyse et le traitement des données généralement utilisés dans l’entraînement de notre modèle multimodal sont les mêmes pour l’ensemble des données. Puisque les outils de mesure ont été choisis dès le début du projet, les données utilisées dans l’apprentissage par transfert sont des données issues des mêmes appareils de mesure. Comme expliqué dans l’état de l’art, l’apprentissage par transfert permet une augmentation du volume de données. Dans notre cas, le contexte des données publiques récupérées est la seule métrique hors de notre contrôle.

3.3.1 Données du transfert d’apprentissage

Les données publiques collectées sont utilisées pour le transfert d’apprentissage du modèle de prédiction du rythme cardiaque ainsi que la dilatation de la pupille. Le Polar H10 ainsi que le Gazepoint GP3 sont deux outils utilisés dans la littérature. Pour les valeurs de rythme cardiaque, puisque les mesures fournies par le Polar H10 sont les mêmes que celles offertes par une simple montre connectée, nous avons simplement cherché des données de rythme cardiaque (Une mesure de battement par minute sera la même sur différents outils). Les données obtenues par le Gazepoint GP3 sont spécifiques à ce modèle, il était donc nécessaire d’obtenir

un jeu de données d’une étude ayant utilisé exactement le même modèle.

Le projet suivant présente un modèle de prédiction du rythme cardiaque en fonction de l’effort physique du participant. Notre expérience présente un contexte différent puisque nous évaluons l’effort mental et non physique. Ce jeu de données comprend 55 fichiers d’entraînements ainsi que 5 fichiers de tests offrant un peu plus de 152000 lignes de données. Chaque ligne comprend les informations suivantes :

- *timestamp* : Horodatage de la mesure ;
- *heart rate* : Valeur du rythme cardiaque (battements par minute) ;
- *cadence* : Cadence de course ;
- *distance* : Distance parcourue ;
- *enhanced altitude* : Altitude ;
- *enhanced speed* : Vitesse ;
- *calories* : Calories consommées ;
- *temperature* : Température.

En comparant notre expérience avec celle ci-dessus, seule la valeur du rythme cardiaque ainsi que l’horodatage nous intéresse puisque les autres métriques définissent le contexte. Il est aussi important de noter la fréquence de prise des mesures des différents outils. Dans l’expérience publique et dans notre expérience, la fréquence de mesure du rythme cardiaque est de 1 hertz (une prise de mesure par seconde).

Le jeu de données (Pillai *et al.*, 2020) est issu d’une analyse du temps de réaction sur une expérience DRT (*Detection Reponse Task*). Cette expérience étudie la dilatation de la pupille avec un Gazepoint GP3. En comparant le format de données de leur expérience avec la nôtre, le format est presque identique. Leur expérience collecte certaines données supplémentaires. Les 28 participants de l’ex-

périence (Pillai *et al.*, 2020) ont chacun 8 fichiers différents regroupés sous 2 catégories représentant les niveaux de difficulté. La première *Single* (sans DRT) et *Dual* (avec DRT). Chaque difficulté est un ensemble de 3 ou 4 expériences, une de contrôle (seulement pour le *DUAL*) et ensuite 3 niveaux de difficultés différents. Cet ensemble de 196 fichiers de données représente un peu plus de 2.3 millions de lignes.

La fréquence de mesure du Gazepoint GP3 est de 60 hertz afin de pouvoir détecter les rapides variations de dilatation de la pupille. Puisque nous utilisons le même outil de mesure, la validation de la fréquence est automatique.

3.3.2 Traitement des données

Le traitement des données est une étape importante de tout projet d'intelligence artificielle. La première vérification effectuée sur l'ensemble de nos données est la présence de *NaN* (Not A Number). Une valeur *NaN* apparaît lorsque la mesure est corrompue. Par exemple, un rythme cardiaque de "A" ne représente pas une valeur cohérente puisque le rythme cardiaque attendu est un nombre. Dans le cadre de notre projet, toutes les données attendues se doivent d'être des nombres. Ainsi, seulement quatre *NaN* sont présents dans notre jeu de données.

Une deuxième analyse effectuée est la détection de valeurs aberrantes. Ces valeurs sont des anomalies ou des incohérences avec les valeurs théoriques attendues. Pour ce qui est du rythme cardiaque, une valeur négative sera considérée comme une valeur aberrante devant être supprimée. Dépendamment de la position de ces valeurs et de leur fréquence, il est possible de simplement les remplacer par leur valeur voisine. Puisque l'ensemble de nos données sont sur des séries chronologiques, la valeur de rythme cardiaque à X seconde est proche de la valeur à $X - 1$ et $X + 1$ seconde.

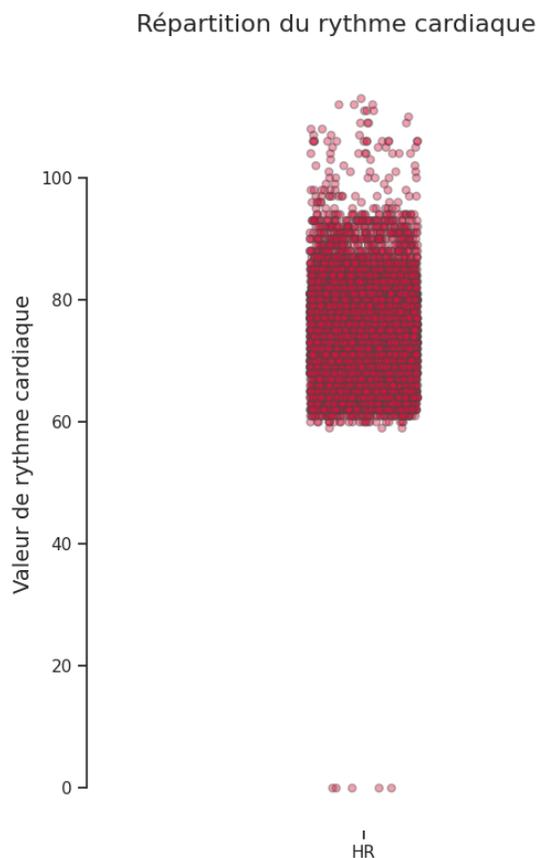


FIGURE 3.3 – Répartition du rythme cardiaque des pilotes de l’expérience.

Une analyse de la répartition des données du rythme cardiaque nous permet d’identifier certaines données erronées. La figure 3.3 nous permet de détecter les erreurs de prise de mesure lors de l’expérience. Certaines valeurs issues du capteur donnent un rythme cardiaque de zéro (les points visibles en bas de la figure). Ces valeurs aberrantes se trouvent dans un fichier spécifique, celui du pilote 3 lors de l’expérience 2-5. Le tableau 3.2 montre l’évolution des données entre l’index 28 et 36. Les cinq valeurs de rythme cardiaque à 0 sont évidemment une erreur et peuvent être modifiées en les remplaçant par une valeur au choix entre 70 et 72. Nous avons remplacé ces valeurs par 71 puisque le rythme cardiaque est une valeur continue qui évolue graduellement dans le temps.

TABLEAU 3.2 – Rythme cardiaque du pilote 2 lors de l’expérience 2-5

Index	Timestamp	HR
28	1651681668	72.0
29	1651681669	72.0
30	1651681670	0.0
31	1651681671	0.0
32	1651681672	0.0
33	1651681673	0.0
34	1651681674	0.0
35	1651681675	70.0
36	1651681676	70.0

Une fois les valeurs nettoyées et vérifiées pour leur continuité, 50% des pilotes de l’expérience présentent un rythme cardiaque entre 70 et 80 battements par minute 3.4.

Pour les données issues de l’*eye tracker* gazeport GP3, le même nettoyage a été effectué. Aucune valeur aberrante ni erreur n’a été corrigée dans le jeu de données. La figure 3.5 montre une concentration des valeurs de dilatation entre 13 et 18 pour les deux pupilles (RPD, LPD).

Les valeurs du rythme cardiaque ainsi que les dilatations des pupilles des pilotes présentent chacun, par le biais des *boxplot*, une médiane centrée entre le premier et troisième quartile. En comparant la répartition du rythme cardiaque des pilotes 3.4 avec celles des athlètes du jeu de données publiques, on remarque que les valeurs des pilotes sont plus faibles. Les athlètes ont une médiane de 144. Ces différences au niveau de la répartition des valeurs constituent les variations de contexte entre nos deux jeux de données. Afin d’appliquer un transfert d’apprentissage fonctionnel,

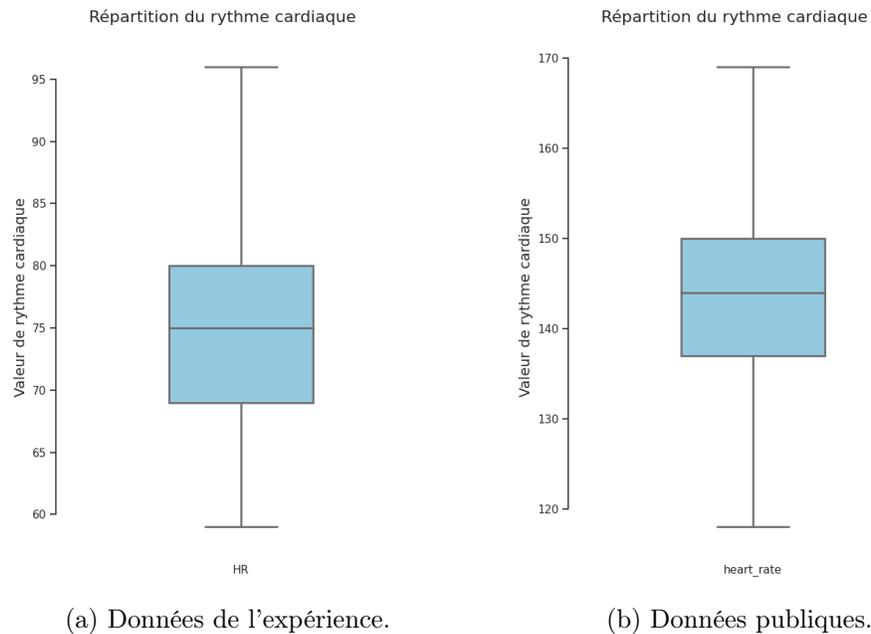


FIGURE 3.4 – Répartition du rythme cardiaque des pilotes de l'expérience et du jeu de donnée publique.

il faut que les deux jeux de données présentent ou non un contexte identique. Cette différence au niveau des valeurs de rythme cardiaque n'est pas un problème puisque l'objectif du transfert d'apprentissage est de généraliser un apprentissage afin de pouvoir se spécialiser par la suite.

Le nettoyage des données issues du casque NCO EEG est identique au nettoyage effectué ci-dessus. Aucune valeur aberrante n'a été détectée et aucune modification majeure n'a été effectuée.

L'analyse des données de *X-Plane* a apporté un grand nombre de problèmes par rapport au système de prise de mesures de l'expérience. Nous allons présenter ces problèmes dans la prochaine section.

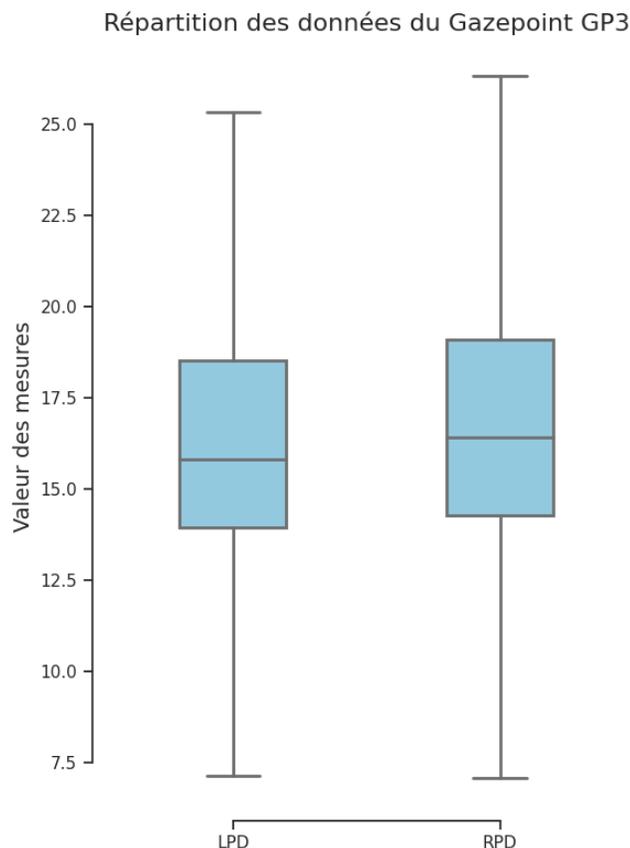


FIGURE 3.5 – Répartition de la dilatation des pupilles des pilotes de l’expérience.

3.3.3 Uniformisation du format de données

Les données de *X-Plane* sont obtenues à l’aide du système de log officiel de l’application. Cependant, l’avion utilisé lors des expériences est l’A320 qui n’est pas un avion officiel du programme. Cet avion est disponible par le biais d’un plugin développé par la communauté. Puisque l’avion n’est pas une partie officielle du programme, celui-ci ne semble pas utiliser le même système de log. Ainsi les données que nous pensions avoir après l’expérience ne sont pas là.

Ce problème nous empêche donc d’obtenir les interactions directes du pilote. Par

exemple, toutes les actions effectuées sur le manche, sur les pédales ou encore sur les manettes des gaz ne sont pas dans notre jeu de données.

Ces valeurs faisaient partie de notre analyse préliminaire. Nous avons comme pour objectif de les incorporer dans notre modèle afin d'augmenter le volume de données disponibles et donc augmenter, théoriquement, le nombre de corrélations entre nos données. Mis à part ces données, nous avons malgré tout récupéré les données de l'environnement de l'avion. Nous parlons de la direction de l'avion, mais aussi de la direction et vitesse du vent.

Un deuxième problème majeur est la fréquence de prise de mesures des données. Voici les différentes fréquences de chaque appareil de mesures :

- Polar H10 - 1 hertz ;
- Gazepoint GP3 - 60 hertz ;
- NCO EEG - 1 hertz ;
- *X-Plane* - Non constante.

Le système de log d'*X-Plane* n'est pas constant. Durant certaines prises de mesures, le logiciel s'arrête d'enregistrer pendant plus de 2 secondes. En observant le temps d'attente entre chaque mesure 3.6, on remarque cette non-constance des prises de mesures couplées aux données manquantes de cet outil qui rend l'ensemble des données obtenues par *X-Plane* moins importantes que prévues pour notre modèle. La prédiction des interactions des pilotes avec l'avion faisait partie des enjeux du projet PilotAI au départ. Malheureusement, la qualité des données obtenues nous oblige à abandonner complètement les données d'interaction du pilote avec l'avion.

Le modèle de prédiction de l'environnement de simulation va donc être développé afin de prédire les quelques données que nous avons pu récupérer (sous-

section 3.2.4).

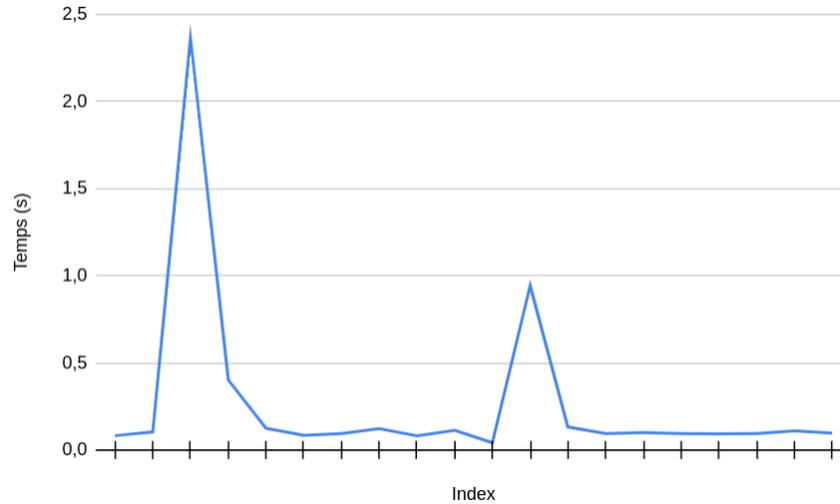


FIGURE 3.6 – Temps d’attente entre chaque prise de mesure.

Afin de préparer l’ensemble de données à l’entraînement d’un modèle multimodal, il a fallu les uniformiser. Puisque les données de dilatation de la pupille sont sur 60 Hz, nous avons ramené chaque donnée sur cette même fréquence. Nous ne ramenons pas les données de dilatation de la pupille sur 1 Hz puisque celle-ci varie rapidement sur l’échelle de la seconde.

L’étalage des données de rythme cardiaque et de charge mentale consiste simplement à dupliquer soixante fois les valeurs de chaque seconde. Ces deux jeux de données passent donc de 1 Hz à 60 Hz.

Les données de X-Plane ont dû être ramenées sur 1 hertz avant d’être dupliquées soixante fois. Il y a alors deux scénarios possibles pour la reconstruction des données *X-Plane* :

- 1. S’il existe des données entre X et $X + 1$ seconde. La moyenne de chaque métrique de cet intervalle est utilisée afin de reconstruire une seule valeur

par seconde. Le nombre de lignes de données entre X et $X + 1$ n'est jamais constant.

- 2. S'il n'existe pas des données entre X et $X + 1$ seconde. La moyenne des valeurs de $X - 1$ et $X + 2$ sont alors utilisée.

L'objectif du traitement des données d'*X-Plane* est de récupérer le maximum d'informations utiles malgré les différents problèmes. Puisque les données météorologiques sont définies par le scénario, il est assez simple de reconstruire les secondes manquantes par le système de log d'*X-Plane*. Les données météorologiques sont constantes. Pour ce qui est de la direction de l'avion, nous avons appliqué les deux scénarios présentés ci-dessus.

La dernière modification effectuée sur les jeux de données est le découpage. Puisque l'expérience regroupe plusieurs outils de mesures, chaque outil démarre et termine son enregistrement de mesures à un moment différent. Il est donc nécessaire de supprimer les mesures ayant eu lieu avant et après le scénario. Les données du rythme cardiaque pourraient être gardées pour le développement d'un modèle ne prenant pas en compte les autres modalités.

3.3.4 Analyse du volume de données

Une fois l'ensemble du nettoyage de données effectué sur l'ensemble de nos jeux de données, nous avons le volume suivant 3.3.

Une grande partie du volume de données a été supprimée à cause de la phase de découpage. Le jeu de données brutes présente souvent presque la moitié des données collectées entre les scénarios 3.7 3.8. Ces données rattachées à aucun contexte sont non utilisables et donc supprimées. Une fois le nettoyage complété, notre jeu de données contient un total de 9394 secondes de données, soit 2.6 heures.

TABLEAU 3.3 – Comparaison du nombre de lignes de données

	Donnée publique	Donnée de l'expérience avant nettoyage	Donnée de l'expérience après nettoyage
HR	152 115	21 929	9 394
ET	2 334 805	1 335 850	564 640
WL	0	17 999	9 394
X-Plane	0	434 641	9 394

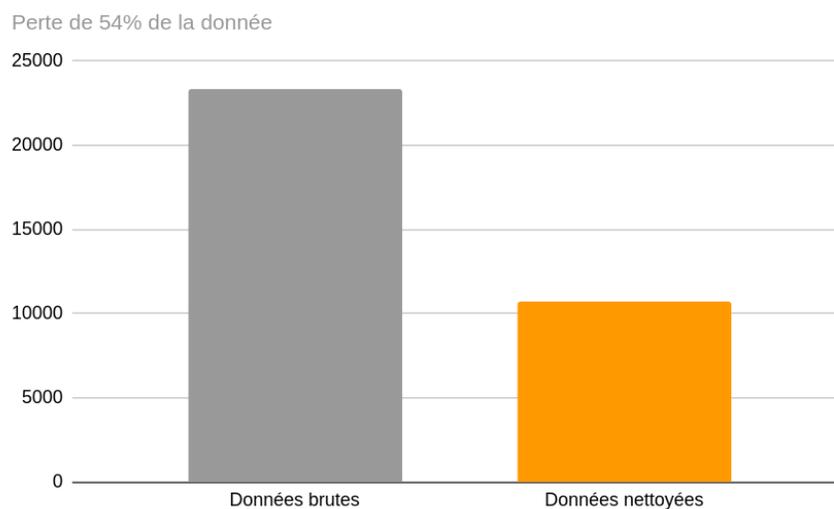


FIGURE 3.7 – Volume de données avant et après nettoyage (Rythme cardiaque).

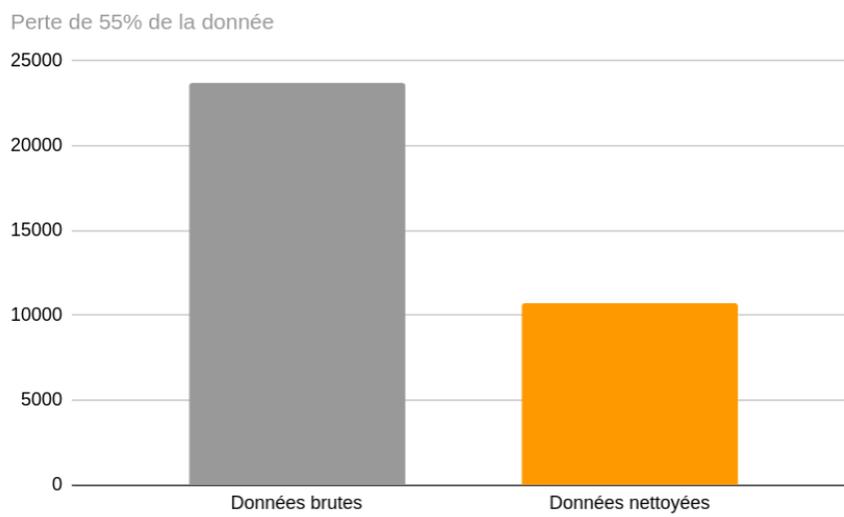


FIGURE 3.8 – Volume de données avant et après nettoyage (Dilatation de la pupille).

CHAPITRE IV

ENTRAÎNEMENT DES MODÈLES

Maintenant que l'architecture de notre modèle ainsi que les données disponibles sont définies, nous pouvons passer à l'entraînement. Il est important de comprendre que les différents modèles présentés sont fortement corrélés à notre ensemble de données. Ce sont elles qui ont défini l'architecture neuronale de chaque réseau. Comme expliqué dans le chapitre 2, la complexité de la donnée et sa forte variabilité nécessite une architecture neuronale complexe. Cette architecture neuronale n'est pas la seule composante impactant la qualité d'un modèle.

Nous avons établi une liste de cinq paramètres importants ayant un impact sur la qualité d'un modèle :

- Qualité de la donnée : La qualité et le volume de la donnée peuvent introduire du bruit faussant les prédictions.
- Traitement de la donnée : Le nettoyage de la donnée, sa préparation et son uniformisation sont nécessaires à l'entraînement des modèles.
- Architecture du modèle : Une architecture trop complexe ou non adaptée à un type de données peut limiter l'apprentissage.
- Hyperparamètres : Les réglages du logiciel d'apprentissage permettent de spécialiser l'outil à nos besoins. Ils permettent de réduire le temps d'entraînement tout en maximisant la qualité de la prédiction.

- Algorithme d’entraînement : Le choix de l’algorithme interne d’optimisation impacte la performance du modèle, pouvant entraîner un modèle à ignorer certaines données d’entrée.

La qualité de la donnée, le traitement de celle-ci ainsi que l’architecture du modèle sont des parties ayant déjà été présentées et analysées dans les chapitres précédents. Analysons maintenant les hyperparamètres ainsi que l’algorithme d’entraînement sélectionné avant de présenter nos résultats.

4.1 Hyperparamètres

La sélection des hyperparamètres est souvent longue et complexe. Il est important de pouvoir évaluer l’impact sur la qualité de la prédiction entre chaque changement de valeurs. Pour le projet, nous avons pris la décision de prendre le modèle de prédiction du rythme cardiaque comme référence. Ce modèle a été sélectionné par sa simplicité en comparaison aux autres modèles. La donnée est unidimensionnelle (seulement une métrique à l’étude) et son architecture est simple (une seule couche neuronale), ce qui en fait un modèle rapide à entraîner. De plus, l’utilisation de l’apprentissage par transfert nous a permis d’entraîner ce modèle avant que l’expérience ait eu lieu.

Afin d’uniformiser au maximum l’impact de chaque modalité sur la prédiction finale, nous avons décidé d’appliquer les mêmes hyperparamètres pour tous les modèles. Notre approche a donc été de les optimiser sur le modèle de prédiction du rythme cardiaque. D’un point de vue technique, nous avons utilisé l’outil KerasTuner (O’Malley *et al.*, 2019). Cet outil est un optimiseur d’hyperparamètres par la méthode d’essai-erreur. Celui-ci a permis d’obtenir les hyperparamètres suivants :

- $epochs = 100$: Nombre d’itérations maximum d’entraînement ;

- *prediction_timestep* = 1 : Nombre de *timestep* futur ;
- *training_timestep* = 30 : Nombre de *timestep* antérieur utilisé ;
- *batch_size* = 500 : Nombre de batchs envoyés en simultanément pour accélérer le temps d'entraînement ;
- *learning_rate* = 0.001 : Taille du pas à chaque itération ;
- *decay_rate* = 0.001 : Facteur de ralentissement du *learning_rate*.

Les trois premières métriques sont des choix que nous avons imposés au logiciel d'optimisation. Le nombre d'itérations est limité à 100 afin de laisser assez de temps au modèle pour effectuer leur apprentissage. Cette valeur est un maximum après lequel l'entraînement se termine, cela ne veut pas dire que le modèle issu de la centième itération est le meilleur. Nous sauvegardons toujours le meilleur modèle en fonction des métriques de validation que nous présentons par la suite.

Le nombre de *timestep* antérieur et futur représentent le temps de la prédiction ($t + 1$ seconde dans le cadre du projet) et la taille de la fenêtre glissante dans le cadre de notre projet. Ces valeurs sont fixes pour l'ensemble de nos modalités afin d'obtenir une uniformité des données lors de la fusion des sous-modèles.

La taille de la *batch_size* est une valeur impactant le temps d'entraînement du modèle ; il s'agit du nombre de sous jeux de données envoyés simultanément au réseau lors de l'entraînement. Cette variable est généralement proportionnelle à la taille complète du jeu de données. L'ensemble de nos modèles ont été développés sur python à l'aide des bibliothèques *Keras* et *TensorFlow*. Celle-ci requiert l'utilisation d'un GPU afin d'accélérer la vitesse de calcul. N'ayant pas de GPU à disposition, nous avons entraîné nos modèles sur *Google Colab* où l'utilisation de GPU est possible. En raison des limitations matérielles, nous avons opté pour une taille de lot relativement élevée afin de réduire le temps d'entraînement (celui-ci pouvait atteindre plus de 8 heures en fonction des modalités).

Les valeurs de *learning_rate* (taux d'apprentissage) et de *decay_rate* (décroissance du taux d'apprentissage) sont les deux valeurs optimisées par *KerasTuner*. Ces deux valeurs sont responsables de la capacité de convergence de notre modèle vers une bonne solution. Une valeur trop grande entraîne une convergence rapide, mais peut entraîner le modèle vers une solution non optimale, alors qu'une valeur trop faible peut ralentir la convergence et donc présenter une meilleure précision.

Puisque notre étude se concentre sur la prédiction du comportement, plusieurs métriques de validation peuvent être utilisées. La prochaine section détaille l'analyse et la sélection de ces métriques.

4.2 Métriques de validation

Toujours dans notre volonté d'uniformisation des modalités et de leur cohérence au niveau de l'entraînement, nous avons utilisé deux métriques. Ces métriques permettent d'évaluer l'efficacité du modèle après chaque itération d'entraînement. Si la nouvelle valeur calculée est meilleure que l'ancienne, le modèle est sauvegardé et continue l'entraînement avec les poids calculés par cette itération. À l'inverse, si la valeur n'est pas meilleure, une nouvelle itération est effectuée afin de trouver un nouvel agencement des poids. Les sections suivantes présentent les deux métriques retenues.

4.2.1 *Mean Squared Error* (MSE)

La première métrique est l'erreur quadratique moyenne (MSE) qui calcule la moyenne des écarts au carré entre les valeurs prédites et les valeurs réelles. Cette méthode utilise la fonction mathématique suivante :

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

- n est le nombre d'échantillons dans l'ensemble du jeu de données ;
- y_i est la valeur cible réelle attendue en utilisant le i ème échantillon ;
- \hat{y}_i est la valeur prédite par le modèle.

Cette métrique attribue un poids plus élevé aux grandes erreurs la rendant sensible aux valeurs aberrantes.

4.2.2 Mean Absolute Error (MAE)

La deuxième métrique est l'erreur absolue moyenne (MAE) qui calcule la moyenne des différences absolues entre les valeurs prédites et les valeurs réelles. Cette méthode utilise la fonction mathématique suivante :

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- n est le nombre d'échantillons dans l'ensemble du jeu de donnée ;
- y_i est la valeur cible réelle attendue en utilisant le i ème échantillon ;
- \hat{y}_i est la valeur prédite par le modèle.

Cette métrique donne un poids égal à toutes les erreurs, ce qui la rend moins sensible aux valeurs aberrantes. De plus, il est de consensus que la valeur MAE est plus rapide à calculer que MSE. L'impact réel de cette rapidité se remarque principalement pour les grands jeux de données. Puisque notre approche consiste à utiliser plusieurs jeux de données pour l'apprentissage par transfert, nous avons décidé d'utiliser la valeur de MAE pour l'ensemble de nos modèles.

4.3 Sous modèles

La section suivante a pour objectif de présenter la capacité de prédiction de nos modèles. Nous présentons donc les analyses suivantes pour l'ensemble des modèles :

- Le temps d'entraînement : Ce temps est important puisqu'il s'agit d'un des facteurs ayant limité notre progression. Certains modèles présentant un temps d'entraînement trop élevé ne peuvent pas être complexifiés au niveau neuronal pour améliorer la précision par exemple.
- La valeur de MAE : Cette valeur nous permet d'évaluer la précision de notre modèle comme expliqué dans la section précédente.
- La courbe de perte : Ce graphique dit *Loss curve* est une évolution de la valeur de MAE tout le long de l'entraînement du modèle. Ce graphique est utilisé afin d'évaluer la qualité de l'apprentissage.
- Un exemple de prédiction : Nous présentons une prédiction pour pouvoir évaluer visuellement la qualité de nos modèles.

4.3.1 Prédiction du rythme cardiaque

Le premier modèle développé dans le cadre du projet est celui de la prédiction du rythme cardiaque. Comme expliqué dans les sections précédentes, nous avons deux modèles partageant la même architecture puisque nous appliquons le transfert d'apprentissage.

Chaque modèle de prédiction du rythme cardiaque a été entraîné pendant 1 heure, que ce soit avec les données publiques ou celle de notre expérience. Les données d'entrées sont le rythme cardiaque des trente dernières secondes afin de prédire le rythme cardiaque futur dans une seconde.

L'évolution de la courbe de perte 4.2 pour les données publiques est une courbe idéale. Les valeurs de MAE d'entraînement et de validation diminuent graduellement lors de l'entraînement sans pour autant être parfaitement superposées, ce qui serait alors un marqueur de surajustement.

L'évolution de la courbe de perte 4.3 pour les données de l'expérience présente les caractéristiques de comportements répétitifs. Le transfert d'apprentissage est visible dans cette courbe. Les valeurs de MAE commencent très faibles et nous remarquons que le modèle essaie de se spécialiser sur notre jeu de données. Cette courbe nous indique un manque de données pour être lissée (goo,). Il est important de noter que les valeurs de MAE sont très faibles, aux alentours de 0.1 pour la validation et l'entraînement. Cette faible valeur permet à notre modèle une excellente prédiction du rythme cardiaque comme nous pouvons le voir sur la figure 4.8. Cette figure montre l'évolution réelle du rythme cardiaque ainsi que la prédiction de notre modèle.

Le modèle développé est donc capable de prédire le rythme cardiaque une seconde dans le futur à partir des trente dernières secondes. Le transfert d'apprentissage nous permet de détecter un manque de données sur nos données de l'expérience puisque la courbe de perte oscille fortement lors de l'entraînement. L'entraînement du modèle manque de données puisqu'il n'est pas capable de se spécialiser sur le nouveau jeu de données (oscillation sur la courbe de perte).

4.3.2 Prédiction de la dilatation de la pupille

Chaque modèle de prédiction de la dilatation de la pupille a été entraîné pendant 8 heures, que ce soit avec les données publiques ou celle de notre expérience. Ce temps d'entraînement est très élevé en comparaison avec nos autres modèles puisque le modèle utilise une superposition de couches LSTM. Les données d'en-

trées sont les 9 métriques identifiées (LPD, RPD) dans le chapitre précédent des trente dernières secondes afin de prédire la dilatation de la pupille gauche future dans une seconde.

Les deux figures de l'évolution de la courbe de perte pour l'entraînement de ces modèles 4.4 et 4.5 montre un excellent apprentissage de nos modèles sur les données publiques et de l'expérience. Dans les deux cas, la courbe diminue graduellement tout le long afin d'atteindre des valeurs de MAE proche de 0.1. La courbe 4.5 présente, elle aussi, une oscillation semblable à celle du rythme cardiaque avant de se stabiliser en fin d'entraînement. Cette oscillation est toujours un marqueur du manque de données.

La figure 4.9 est un exemple de prédiction de notre modèle sur les données d'un pilote. La courbe bleue est la valeur réelle alors que la courbe orange est la prédiction progressive de notre modèle en utilisant l'historique de données. La dilatation de la pupille est précise. L'évolution progressive croissante de cette dilatation ainsi que les parties d'évolution décroissante se retrouvent dans la prédiction. Nous remarquons que le modèle n'est pas parfait lors de fortes variations sur un court espace de temps. Cela est sûrement dû à la taille du jeu de données utilisées pour l'entraînement.

Le modèle développé est donc capable de prédire la dilatation de la pupille une seconde dans le futur à partir des trente dernières secondes.

4.3.3 Prédiction de la charge mentale

La charge mentale est la dernière métrique de notre modalité sensorielle. Celui-ci ne présente pas de transfert d'apprentissage afin de pallier le manque de données et cela se remarque lorsque nous observons l'évolution de la courbe de perte 4.6. Ce graphique montre un clair manque d'apprentissage de la part de notre modèle.

La donnée se trouve donc être trop complexe et nous n'en avons pas assez pour que le modèle puisse apprendre.

Afin de confirmer que le problème vienne bel et bien du manque de données, nous avons développé plusieurs variantes de ce modèle :

- Substitution des couches LSTM par des couches RNN : Puisque les LSTM sont efficaces pour traiter les données plus complexes, elles peuvent aussi ne pas être efficaces lorsque le volume de données n'est pas élevé. L'utilisation de couche RNN n'a pas amélioré l'entraînement du modèle.
- Superposition de couches LSTM / RNN : L'étude d'un autre étudiant du projet PilotAI, qui développe lui aussi un modèle de prédiction de la charge mentale, indique que l'utilisation de couches superposées LSTM lui permet de prédire la charge mentale cinq secondes dans le futur (Antoine, 2022). Nous avons donc évalué cette approche pour la prédiction une seconde dans le futur, malheureusement sans succès pour nous deux.
- Modification des hyperparamètres : Les hyperparamètres que nous utilisons sont toujours ceux définis par le modèle du rythme cardiaque. Le programme KerasTuner que nous avons utilisé précédemment n'a pas réussi à obtenir des paramètres permettant un meilleur apprentissage pour ce modèle.

L'apprentissage par transfert n'est pas possible pour ce modèle puisque les données collectées par l'EEG sont transformées par le logiciel propriétaire de BMU. L'utilisation de ce logiciel nous empêche de trouver des jeux de données publiques utilisant des données semblables. Nous allons discuter de l'impact de ce manque d'apprentissage dans la section section 4.5.

Le modèle développé n'est donc pas capable de prédire de façon précise la charge mentale à une seconde près dans le futur à partir des trente dernières secondes.

Une observation de la qualité de la prédiction (4.7) nous permet de mieux comprendre ce qu'il se passe avec ce modèle. La courbe bleue de cette figure représente les valeurs réelles et la courbe orange la prédiction du modèle en se basant sur les trente dernières secondes. Les valeurs réelles de charge mentale évoluent très rapidement au fil du temps et la prédiction semble suivre la tendance générale de l'évolution de la charge mentale, mais ne respecte pas l'intensité des pics. Notre modèle est donc sur la bonne voie, mais ce manque de spécialisation de notre modèle nous confirme le manque de données pour ce modèle.

4.3.4 Prédiction des données extraites du simulateur

La prédiction des données extraites du simulateur est une des modalités que nous avons partiellement modifiées par rapport à nos attentes initiales (sous-section 3.3.3). Le modèle de prédiction utilise donc seulement les données définies précédemment (sous-section 3.2.4). Ce modèle utilise donc les données d'entrée afin de prédire ces mêmes métriques une seconde dans le futur. Contrairement aux données sensorielles, ces données évoluent très peu sur une échelle d'une seconde. Ainsi, en regardant l'évolution de la courbe de perte 4.10, nous remarquons une très bonne évolution de la valeur de MAE. Les courbes d'entraînement et de validation diminuent graduellement au fil de l'entraînement afin d'atteindre une faible valeur de MAE comprise entre 0.05 et 0.15.

Cette courbe n'est pas aussi bonne que celle du modèle de prédiction de la dilatation de la pupille. Ce modèle calcule la prédiction de huit métriques et non pas une seule. Il est donc normal d'obtenir une valeur de MAE plus élevée.

Le modèle développé est donc capable de prédire l'état de l'environnement de simulation une seconde dans le futur à partir des trente dernières secondes.

4.4 Modèle multimodal pour la prédiction du comportement

Maintenant que nous avons validé que l'architecture de nos modèles composant le modèle multimodal est capable de prédire chaque métrique indépendamment, nous pouvons entraîner notre modèle multimodal global. La prédiction de la charge mentale n'est pas précise. Le manque de donnée qui a été établie ci-dessus peut prendre avantage de la multimodalité pour améliorer la précision de la charge mentale.

Le modèle multimodal est donc entraîné sur l'ensemble des données, le transfert d'apprentissage est appliqué avec les données publiques de rythme cardiaque et de dilatation de la pupille. L'ensemble des données de l'expérience sont utilisées sur le modèle multimodal une fois le transfert effectué. En plus, pour la première fois jusqu'ici, les données du contexte contenant l'état des moteurs sont utilisées dans l'entraînement des modèles. Le temps d'entraînement de ce modèle est de 8 heures. Le modèle que nous présentons a pour objectif de prédire l'ensemble des métriques définies dans les modèles ci-dessus (Rythme cardiaque, Dilatation de la pupille, Charge mentale, Donnée de l'environnement de simulation).

La tendance d'évolution de la courbe de perte 4.11 est une courbe d'absence d'apprentissage. Les données d'entraînement sont maîtrisées, cependant le modèle est incapable d'améliorer la qualité de sa prédiction sur les données de validation. Le modèle multimodal, tel quel, n'est pas capable de prédire les données demandées.

La prochaine section rentre en détail dans ce problème puis présente la solution que nous avons utilisée afin de valider notre modèle multimodal.

4.5 Absence d'apprentissage

L'absence d'apprentissage est un des problèmes majeurs de notre étude. Nous remarquons un faible apprentissage pour le modèle de prédiction de la charge mentale et un apprentissage nul sur les données de validation du modèle multimodal. Les différentes approches appliquées au modèle de la charge mentale ont déjà été détaillées plus haut. Puisque le modèle multimodal constitue une grande partie de notre étude, nous avons appliqué et testé les différentes méthodes suivantes. L'entraînement de chaque variation du modèle multimodal était arrêté si aucun apprentissage n'avait lieu après 10 itérations d'entraînement afin de ne pas perdre 8 heures pour chaque variation.

- Substitution des couches LSTM par des couches RNN : La même approche que celle utilisée sur le modèle de la charge mentale fut testée. En gardant les couches des sous modèles intactes, nous avons modifié la couche LSTM d'apprentissage après la couche de fusion. Ce changement n'a pas impacté la qualité de l'entraînement.
- Modification des hyperparamètres : Une modification graduelle des hyperparamètres n'a montré aucune amélioration de l'entraînement.
- Diminution du nombre de couches d'apprentissage : Le modèle tel que présenté contient un total de 8 couches LSTM. Certaines de ces couches sont superposées, dans le cas du modèle de dilatation de la pupille par exemple, ce qui peut entraîner un biais de surspécialisation des sous-modèles. Nous avons donc essayé de modifier graduellement le nombre de couches.
- Réduire le nombre de métriques prédites : Le modèle présenté essaie de prédire l'ensemble des métriques précédemment présenté. Nous avons alors décidé de simplifier cela en ne prédisant que le rythme cardiaque à partir de l'ensemble des données par exemple. Après avoir essayé cela sur l'ensemble

des métriques, ce changement n'a pas amélioré le modèle.

- Simplification de l'architecture générale : Puisque le modèle de la charge mentale n'est pas aussi précis que celui du rythme cardiaque et de dilatation de la pupille, nous avons simplifié l'architecture du modèle. Un nouveau modèle multimodal simple, simplement composé de deux métriques, a alors été testé. Ce petit modèle utilise le rythme cardiaque et la dilatation de la pupille, en faisant alors le modèle multimodal le plus simple que nous pouvons faire avec deux de nos meilleurs modèles de prédiction. Cette simplification ne montre aucun apprentissage.

L'ensemble des variations du modèle multimodal que nous avons essayé ci-dessus nous entraîne encore une fois vers la même conclusion, nous manquons de données afin de pouvoir concevoir un tel modèle multimodal. Afin de confirmer ce problème, nous avons décidé de valider l'architecture multimodale sur un jeu de données multimodales publiques. La prochaine section présente cette solution.

4.5.1 Validation du modèle multimodal

Afin de valider le travail effectué sur le modèle multimodal, nous avons décidé de tester notre architecture simplifiée sur un jeu de donnée publique. L'architecture simplifiée est celle présentée ci-dessus, c'est-à-dire, seulement deux sous-modèles, afin d'avoir une multimodalité simple respectant l'ensemble des critères que nous avons posés jusqu'ici.

Le jeu de données publiques que nous avons sélectionné est un jeu de données multimodales issues de plusieurs capteurs (Gashi *et al.*, 2022). Leur expérience collecte les données du rythme cardiaque, d'un accéléromètre, de la consommation en oxygène, d'un gyroscope et du rythme respiratoire. Comme nous l'avons vu dans le chapitre 1, la consommation d'oxygène est une des métriques utilisées

dans différentes études du comportement des pilotes. Nous avons donc décidé d'utiliser notre architecture simplifiée afin de prédire le rythme cardiaque à partir du rythme cardiaque et de la consommation en oxygène.

Pour ce qui est de l'architecture neuronale du nouveau modèle, nous avons décidé de modifier le moins possible ce que nous avons déjà développé. En ne modifiant que les données d'entrée et les valeurs prédites, nous pouvons voir le réel impact du volume de données sur notre architecture multimodale. Le sous-modèle de prédiction du rythme cardiaque ne change pas. Nous utilisons l'architecture LSTM superposée de notre modèle de prédiction de la dilatation de la pupille afin de traiter les données de consommation en oxygène.

Les données exactes d'entrée du modèle sont donc les suivantes :

- *HR* : Rythme cardiaque ;
- *VO2* : Consommation d'oxygène relative au poids ;
- *Rf* : Fréquence respiratoire ;
- *Tv* : Volume d'air respiré ;
- *Ve* : Ventilation, volume d'air déplacé par minute ;
- *Ve/VO2* : Proportion de volume d'air déplacé sur la consommation d'oxygène ;
- *FeO2* : Pourcentage d'oxygène expiré ;
- *Pressure* : Pression ;
- *Temp* : Température ;
- *HUM* : Taux d'humidité.

L'évolution de la courbe de perte du modèle multimodal simplifié 4.12 présente de l'apprentissage. La différence est flagrante lorsque nous comparons les courbes de validations avec celle issue de nos données. Évidemment, la précision n'est pas très bonne pour ce modèle. Une bonne précision n'est pas ce que nous voulions obtenir

dans notre approche. Il est normal que l'architecture du modèle de prédiction de la dilatation de la pupille ne soit pas spécialisée dans le traitement des données de consommation d'oxygène. La présence d'apprentissage lors de l'entraînement du modèle renforce encore une fois notre hypothèse du manque de données pour notre projet.

Le jeu de données publiques utilisé pour ce modèle regroupe un peu moins de 30 000 secondes de données alors que notre jeu de données n'en contient que 9 394 secondes. Un jeu de données trois fois plus gros nous permettrait donc de concevoir le modèle multimodal simplifié.

CONCLUSION

Notre recherche étudie le comportement multimodal d'un pilote afin d'en prédire son futur comportement. Nous avons collecté, analysé et préparé plusieurs jeux de données différents afin de concevoir différents modèles d'intelligence artificielle. Cette recherche s'est concentrée sur les deux questions de recherches posées en introduction.

Premièrement, nous avons comme objectif d'étudier la prédiction du rythme cardiaque, de la dilatation de la pupille ainsi que la charge mentale d'un pilote au décollage. Nous avons donc développé cinq modèles de prédiction utilisant des réseaux de neurones spécialisés dans le traitement de données chronologiques (LSTM). Le concept d'apprentissage par transfert a été utilisé afin de pallier le manque de données. Nos modèles de prédiction du rythme cardiaque et de dilatation de la pupille sont précis et fiables avec une MAE de 0.1 pour les deux métriques. Notre modèle de prédiction de la charge mentale propose des résultats prometteurs souffrant d'un manque de données.

Deuxièmement, nous avons comme objectif d'étudier la prédiction du comportement en utilisant une architecture multimodale. Nous souhaitons utiliser les données sensorielles prélevées de façon non intrusive afin de prédire celles prélevées de façon intrusive. Nous avons développé une architecture multimodale facilement modifiable par sa segmentation en modalité. Un total de sept réseaux neuronaux compose le modèle multimodal proposé, ce qui en fait un modèle d'apprentissage profond. Le volume de données que nous avons récolté lors de l'expérience ne nous permet pas de valider complètement notre deuxième hypothèse. Cependant,

nous avons validé le fonctionnement de l'architecture proposée en utilisant un jeu de données multimodales publiques en constatant un apprentissage significatif de notre modèle.

La prédiction du comportement d'un pilote d'avion au décollage est donc possible à l'aide des différents modèles développés à condition que le volume de données disponibles pour l'entraînement du modèle soit conséquent.

4.6 Limitations

Notre recherche présente plusieurs limitations. Premièrement, la capacité d'apprentissage de nos modèles est limitée par la quantité de données issues de notre expérience. Le volume de données est jugé insuffisant par rapport à la complexité du modèle que nous présentons.

Deuxièmement, l'ensemble de nos modèles calculent une prédiction au temps $t + 1$ seconde. Une prédiction autre, disons $t + 5$ secondes, n'est pas couverte par nos modèles actuels et nécessitera un entraînement complet et plus conséquent des modèles pour y arriver.

4.7 Perspectives futures

Puisque cette recherche fait partie du projet PilotAI, les modèles développés vont être utilisés afin de concevoir un modèle d'analyse de déviation du comportement du pilote. De plus, une deuxième expérience peut être effectuée afin de collecter un plus grand volume de données afin d'entraîner et prédire le comportement du pilote de façon multimodale et plus efficacement. Les métriques utilisées par nos modèles ont été imposées par les outils de mesure à notre disposition. La variabilité du rythme cardiaque est une métrique identifiée par les partenaires qui n'a pas été utilisée. De nouvelles métriques et modalités pourraient être ajoutées

à condition d'augmenter le volume de données disponibles en conséquence.

L'architecture générale du modèle multimodal que nous avons présentée est générique. Elle est facilement réutilisable et adaptable à de nouvelles modalités, comme nous l'avons démontré par l'utilisation d'un nouveau jeu de données utilisant la consommation d'oxygène. Cette architecture peut donc être utilisée à d'autres études, qu'elles soient comportementales ou non.

Nous avons concentré nos recherches autour des réseaux RNN et LSTM. Depuis la montée en popularité et l'efficacité prouvée des *transformers* dans le domaine de l'analyse de la langue naturelle (*ChatGPT*), ces réseaux pourraient être utilisés afin de développer un nouveau modèle multimodal suivant la même architecture proposée. Des méthodes d'apprentissage automatique traditionnelles (non profondes), telles que les arbres de décision, le Random Forest ou le SVM pourraient, eux aussi, être développés sur le jeu de données de l'expérience. Une comparaison de l'efficacité, la précision et le temps d'entraînement des différents modèles et architectures pourrait être envisagée.

ACRONYMES

- CNN** Réseau de neurones convolutifs (Convolutional Neural Network). 17
- EEG** Electroencéphalogramme. iii, 32, 42, 43, 49, 51, 64
- ET** Oculomètre (Eye Tracker). 54
- GPU** Carte graphique (Graphics Processing Unit). 58
- HR** Rythme cardiaque (Heart Rate). 41, 48, 54, 69
- LPD** Dilatation de la pupille gauche (Left Pupil Dilation). vii, 48, 63, 80
- LSTM** Réseaux de neurones récurrents à mémoire à long terme (Long Short-Term Memory networks). ii, 3, 11, 13–16, 25, 26, 28–31, 62, 64, 67, 69, 71, 73
- MAE** Erreur absolue moyenne (Mean Absolute Error). iv, vii, viii, 17, 60–63, 65, 71, 75–79, 81, 82
- MSE** Erreur quadratique moyenne (Mean Squared Error). iv, 59, 60
- RNN** Réseau de neurones récurrents (Recurrent Neural Network). ii, vi, 3, 11, 13–16, 32, 64, 67, 73
- RPD** Dilatation de la pupille droite (Right Pupil Dilation). 48, 63
- WL** Charge mentale (Workload). 54
- WP** Module du projet (Work Packages). 36, 38

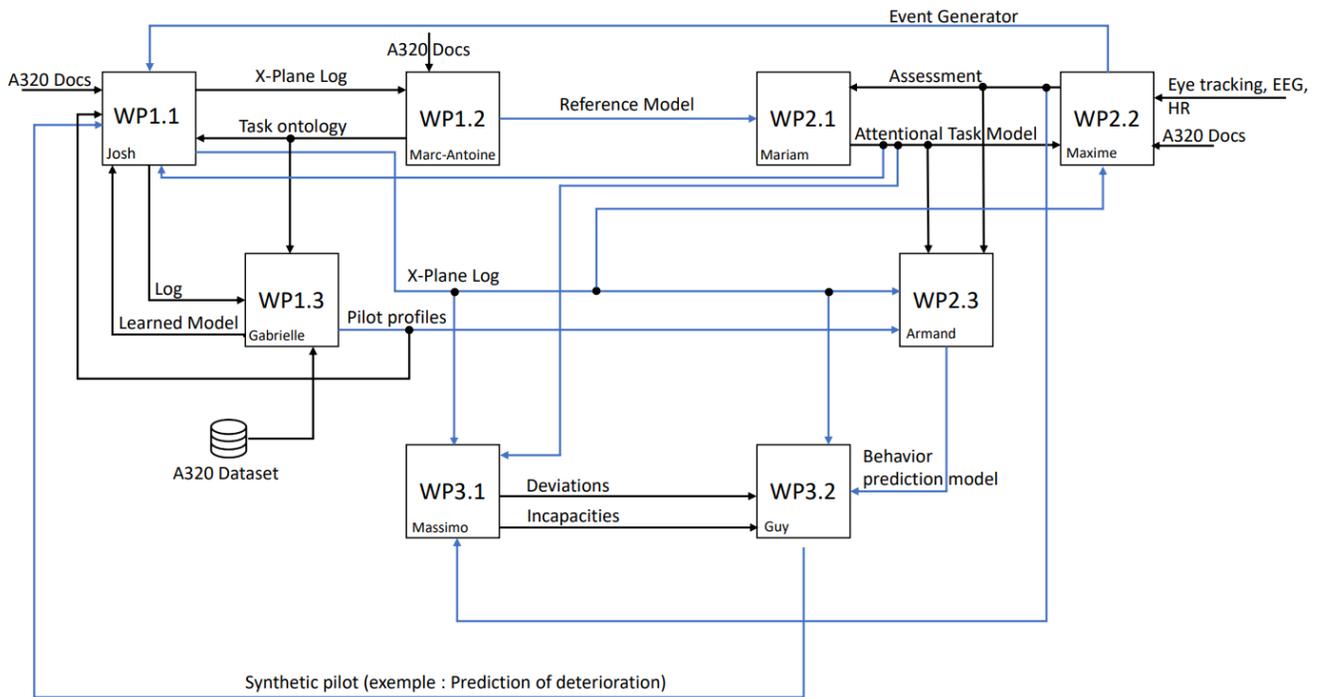


FIGURE 4.1 – Connexion entre les WP du projet PilotAI.

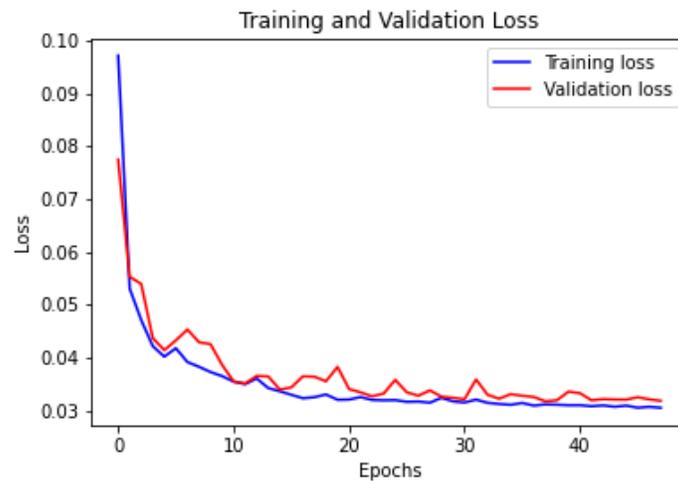


FIGURE 4.2 – Fonction de perte durant l'entraînement du modèle de prédiction du rythme cardiaque avec les données publiques. (La *loss* est basé sur la MAE)

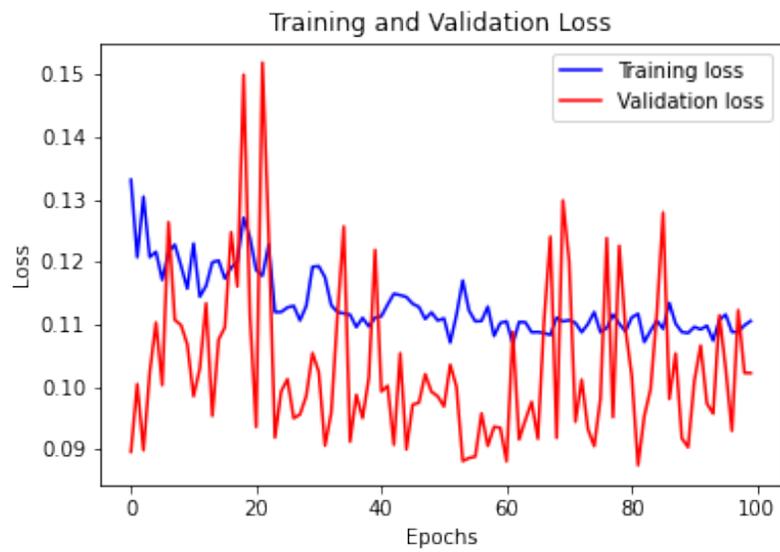


FIGURE 4.3 – Fonction de perte durant l’entraînement du modèle de prédiction du rythme cardiaque avec les données de l’expérience. (La *loss* est basé sur la MAE)

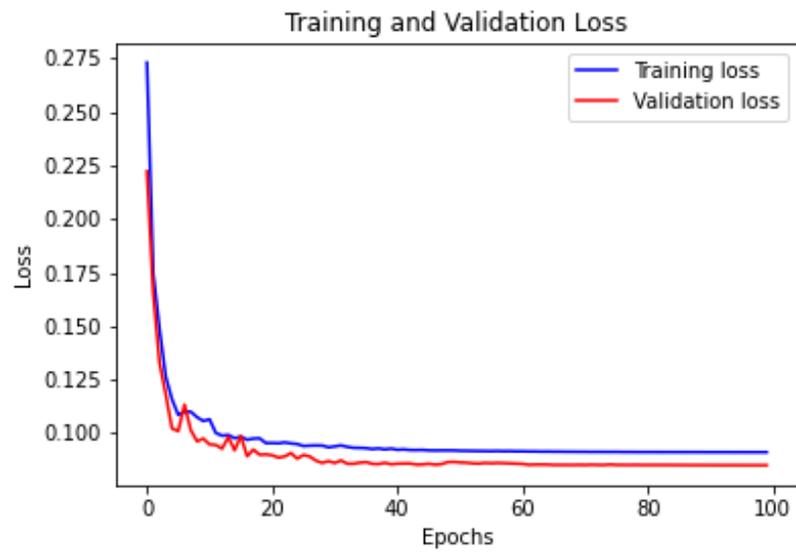


FIGURE 4.4 – Fonction de perte durant l’entraînement du modèle de prédiction de la dilatation de la pupille avec les données publiques. (La *loss* est basé sur la MAE)

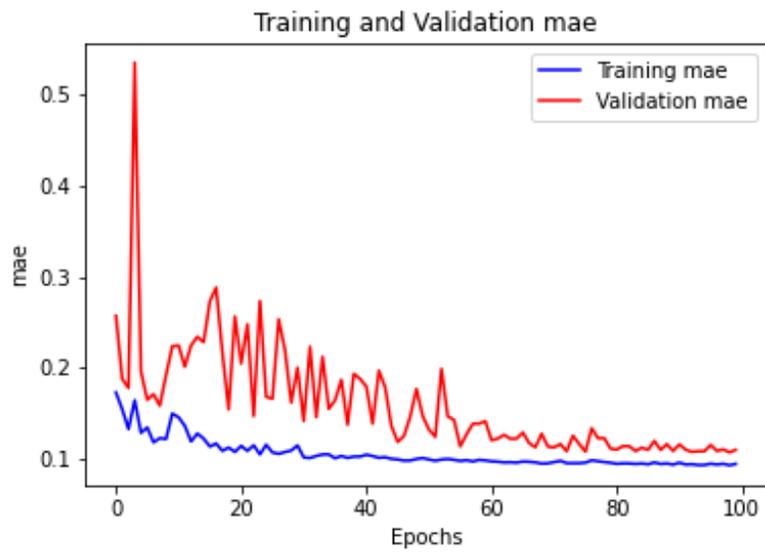


FIGURE 4.5 – Fonction de perte durant l’entraînement du modèle de prédiction de la dilatation de la pupille avec les données de l’expérience. (La *loss* est basé sur la MAE)

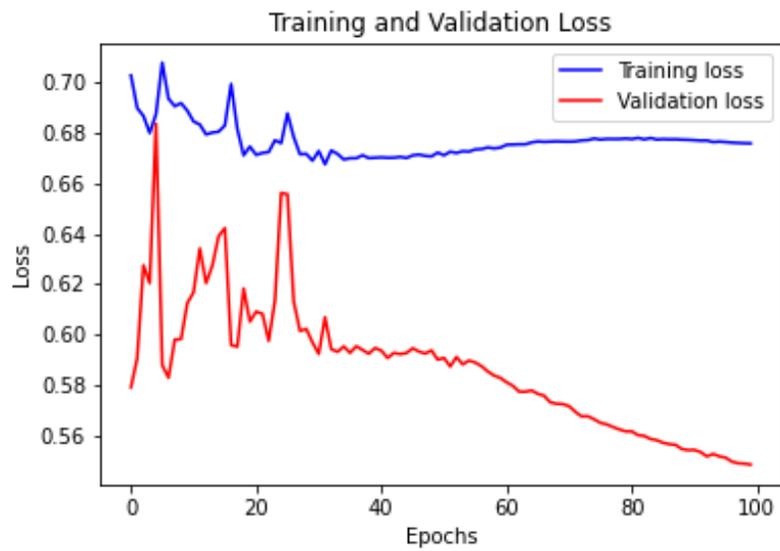


FIGURE 4.6 – Fonction de perte durant l’entraînement du modèle de prédiction de la charge mentale avec les données de l’expérience. (La *loss* est basé sur la MAE)

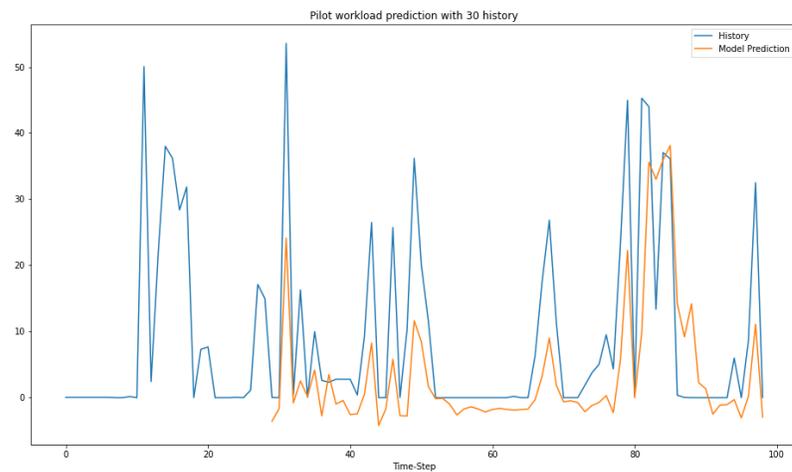


FIGURE 4.7 – Prédiction de la charge mentale avec les données de l’expérience. (charge mentale sur l’axe vertical et le temps sur l’axe horizontal)

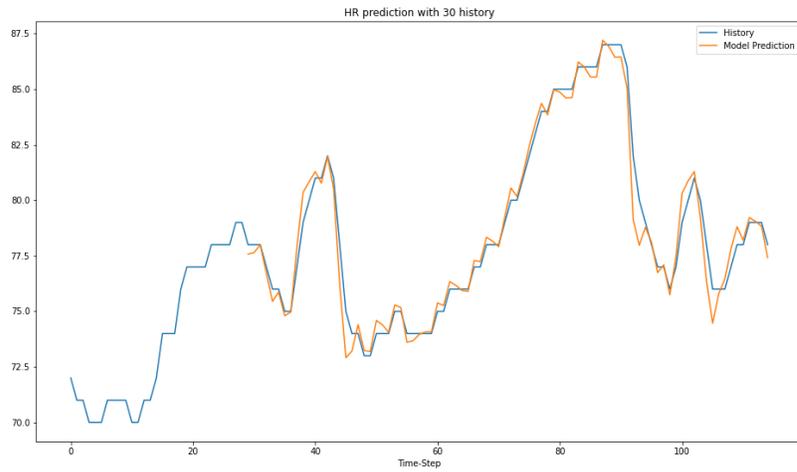


FIGURE 4.8 – Prédiction du rythme cardiaque avec les données de l'expérience. (nombre de battements cardiaque par minute sur l'axe vertical et le temps sur l'axe horizontal)

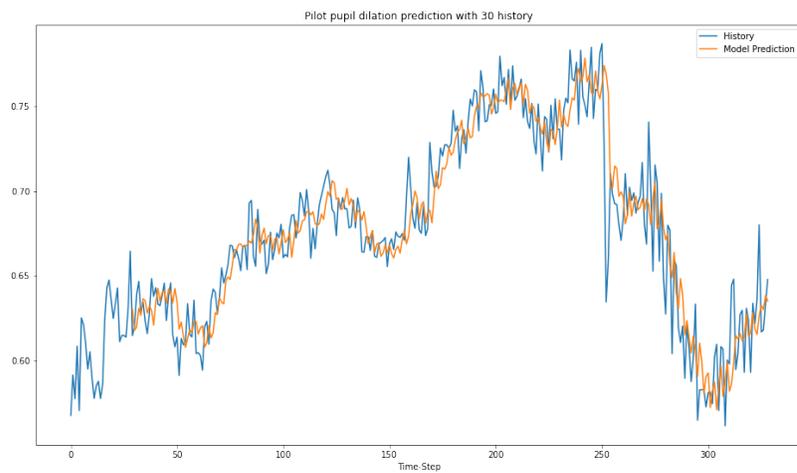


FIGURE 4.9 – Prédiction de la dilatation de la pupille avec les données de l'expérience. (dilatation de la pupille gauche (LPD) sur l'axe vertical et le temps sur l'axe horizontal)

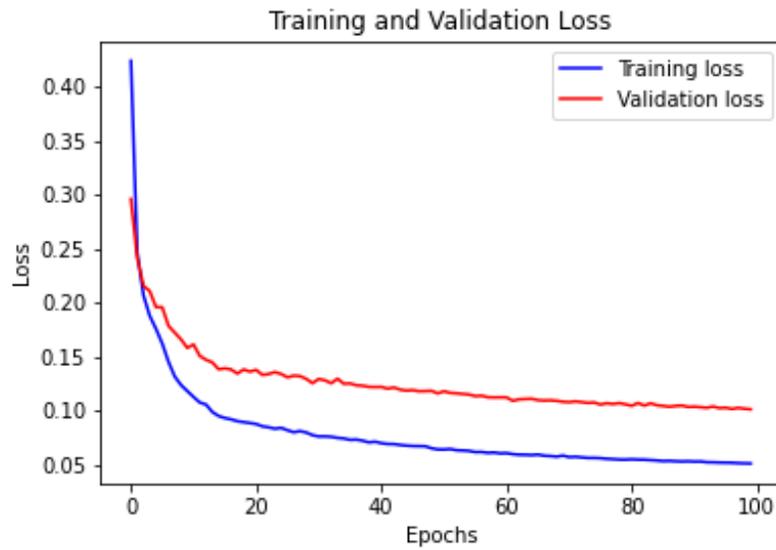


FIGURE 4.10 – Fonction de perte durant l’entraînement du modèle de prédiction de l’environnement de simulation avec les données de l’expérience. (La *loss* est basé sur la MAE)

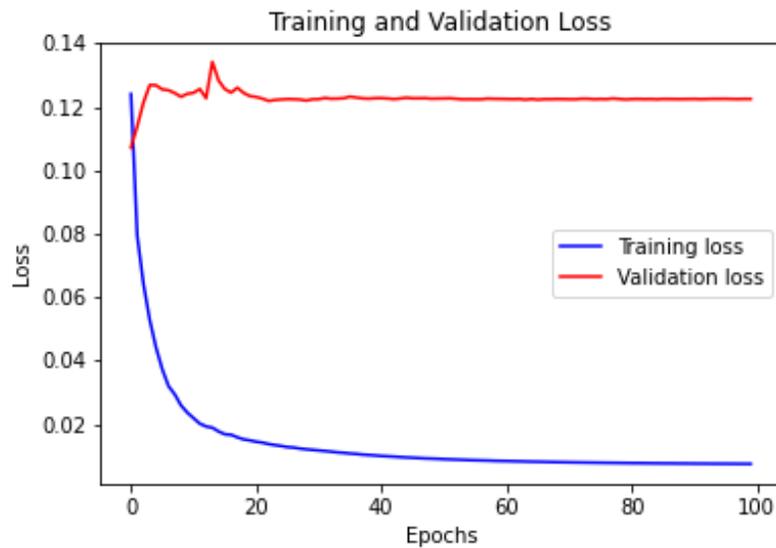


FIGURE 4.11 – Fonction de perte durant l’entraînement du modèle multimodal avec les données de l’expérience. (La *loss* est basé sur la MAE)



FIGURE 4.12 – Fonction de perte durant l’entraînement du modèle multimodal avec les données publiques. (La *loss* est basé sur la MAE)

RÉFÉRENCES

Interpreting loss curves | machine learning | google developers. Récupéré de <https://developers.google.com/machine-learning/testing-debugging/metrics/interpretic>

Antoine, M. (2022). Évaluation de la charge mentale des pilotes en situation de fausses manœuvres.

Baltrusaitis, T., Ahuja, C. et Morency, L.-P. (2019). Multimodal machine learning : A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423–443.
<http://dx.doi.org/10.1109/tpami.2018.2798607>

Benlamine, M. S., Chaouachi, M., Frasson, C. et Dufresne, A. (2016). Physiology-based recognition of facial micro-expressions using eeg and identification of the relevant sensors by emotion. Dans *PhyCS*, 130–137.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D. et Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, 44, 58–75.

Brand, J., Diamond, S. G., Thomas, N. et Gilbert-Diamond, D. (2020). Evaluating the data quality of the gazeport GP3 low-cost eye tracker when used independently by study participants. *Behavior Research Methods*, 53(4), 1502–1514.

Chaouachi, M., Jraidi, I. et Frasson, C. (2015). Mentor : a physiologically controlled tutoring system. Dans *User Modeling, Adaptation and Personalization : 23rd International Conference, UMAP 2015, Dublin, Ireland, June 29–July 3, 2015. Proceedings 23*, 56–67. Springer.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. et Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv :1406.1078*.

- Cui, C., Luo, J. et Chen, H. (2020). Deep multimodal learning : A survey on recent advances and trends. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8), 1857–1879.
- Elman, J. L. (1990). Finding structure in time. *Cognitive science*, 14(2), 179–211.
- Fridman, L., Reimer, B., Mehler, B. et Freeman, W. T. (2018). Cognitive load estimation in the wild. Dans *Proceedings of the 2018 chi conference on human factors in computing systems*, 1–9.
- Gashi, S., Min, C., Montanari, A., Santini, S. et Kawsar, F. (2022). A multidevice and multimodal dataset for human energy expenditure estimation using wearable devices. *Scientific Data*, 9(1), 537.
- Hammer, G. P., du Prel, J.-B. et Blettner, M. (2009). Avoiding bias in observational studies : part 8 in a series of articles on evaluation of scientific publications. *Deutsches Ärzteblatt International*, 106(41), 664.
- Haward, L. (1967). Assessment of stress-tolerance in commercial pilots. *Flight safety*.
- Hinton, G. E., Osindero, S. et Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), 1527–1554.
- Hochreiter, S. et Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–1780.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H. et Van de Weijer, J. (2011). *Eye tracking : A comprehensive guide to methods and measures*. OUP Oxford.
- Hyönä, J., Tommola, J. et Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology*, 48(3), 598–612.
- Kimura, N., Yoshinaga, I., Sekijima, K., Azechi, I. et Baba, D. (2019). Convolutional neural network coupled with a transfer-learning approach for time-series flood predictions. *Water*, 12(1), 96.
- Luo, M. et Wu, K. (2020). Heart rate prediction model based on neural network. Dans *IOP Conference Series : Materials Science and Engineering*, volume 715, p. 012060. IOP Publishing.
- Morris, C. H. et Leung, Y. K. (2006). Pilot mental workload : how well do pilots really perform ? *Ergonomics*, 49(15), 1581–1596.

- O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L. *et al.* (2019). Keras Tuner. <https://github.com/keras-team/keras-tuner>.
- Pillai, P., Ayare, P., Balasingam, B., Milne, K. et Biondi, F. (2020). Response time and eye tracking datasets for activities demanding varying cognitive load. *Data in Brief*, 33, 106389.
- Ramachandram, D. et Taylor, G. W. (2017). Deep multimodal learning : A survey on recent advances and trends. *IEEE Signal Processing Magazine*, 34(6), 96–108. <http://dx.doi.org/10.1109/msp.2017.2738401>. Récupéré de <https://doi.org/10.1109/msp.2017.2738401>
- Rankin, W. (2007). Meda investigation process. *Boeing Commercial Aero*.
- RE, K. (1945). A physiological study of aviators during combat flying. *The Journal of aviation medicine*, 16, 376–384.
- Roscoe, A. H. (1992). Assessing pilot workload. why measure heart rate, hrv and respiration ? *Biological psychology*, 34(2-3), 259–287.
- Siegle, G. J., Ichikawa, N. et Steinhauer, S. (2008). Blink before and after you think : Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*, 45(5), 679–687.
- Tokgöz, A. et Ünal, G. (2018). A rnn based time series approach for forecasting turkish electricity load. Dans *2018 26th Signal processing and communications applications conference (SIU)*, 1–4. IEEE.
- Torrey, L. et Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends : algorithms, methods, and techniques* 242–264. IGI global.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. et Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Wikipedia (2023a). Bombardier (entreprise) — Wikipedia, the free encyclopedia. [http://fr.wikipedia.org/w/index.php?title=Bombardier%20\(entreprise\)&oldid=194939329](http://fr.wikipedia.org/w/index.php?title=Bombardier%20(entreprise)&oldid=194939329).
- Wikipedia (2023b). CAE (entreprise) — Wikipedia, the free encyclopedia. [http://fr.wikipedia.org/w/index.php?title=CAE%20\(entreprise\)&oldid=200547209](http://fr.wikipedia.org/w/index.php?title=CAE%20(entreprise)&oldid=200547209).
- Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of*

Aviation Psychology, 12(1), 3–18.

Zekveld, A. A., Kramer, S. E. et Festen, J. M. (2011). Cognitive load during speech perception in noise : The influence of age, hearing loss, and cognition on the pupil response. *Ear and hearing*, 32(4), 498–510.