

Child Mind Institute – Detect Sleep States

Detect sleep onset and wake from
wrist-worn accelerometer data

Feriel ABDI
Manel BENKORTEBI
Cassandra GBABOUA
Inès LEBIB



M2- Bio-Informatique
Université Paris Cité
27/10/2023

SOMMAIRE

- I. Introduction
- II. Matériels et méthodes
- III. Analyses préliminaires
- IV. Deep Learning
- V. Conclusion et perspectives

INTRODUCTION

Contexte du projet

CONTEXTE DU PROJET

Meilleure compréhension des mécanismes et des conséquences du sommeil sur la santé.

Détection avec précision les phases d'endormissement et de réveil chez les enfants et adolescents

Objectif : mettre en place un système innovant capable de déterminer l'état de sommeil chez les jeunes en se basant sur ces données.



-> Apprentissage automatique et analyse des données

Prédiction de deux types d'événements:

- 'wakeup' (réveil)
- 'onset' (endormissement).

MATERIELS & METHODES

DATA

Train_series :

Données d'entraînement

series_id	step	timestamp	anglez	enmo
31011ade7c0a	0	2017-08-15 16:45:00	-82.680603	0.0000
31011ade7c0a	1	2017-08-15 16:45:05	-82.680603	0.0000

Train_events :

Evenements 'onset' et 'wakeup'

series_id	night	event	step	timestamp
038441c925bb	1	onset	4992.0	2018-08-14T22:26:00-0400
038441c925bb	1	wakeup	10932.0	2018-08-15T06:41:00-0400

Test_series :

Données pour la prédiction

series_id	step	timestamp	anglez	enmo
038441c925bb	0	2018-08-14T15:30:00-0400	2.636700	0.0217
038441c925bb	1	2018-08-14T15:30:05-0400	2.636800	0.0215

Méthodes – Preprocessing

- Fusion du data 'Train' avec les événements associés 'Events'
- Nettoyage des données

series_id	step	timestamp	anglez	enmo	event
31011ade7c0a	0	2017-08-15 16:45:00	-82.680603	0.0000	wakeup
31011ade7c0a	1	2017-08-15 16:45:05	-82.680603	0.0000	wakeup

14 334 881 données

- Filtrage des données avec pas de 50
- Normalisation des données
- Transformation de la variable 'event' en binaire :

series_id	step	timestamp	anglez	enmo	event
31011ade7c0a	0	2017-08-15 16:45:00	0.040663	0.000000	1
31011ade7c0a	1	2017-08-15 16:45:05	0.040663	0.000000	1

286 298 données

- 0 : onset
- 1 : wakeup

Méthodes – Analyses

Préliminaires

ANALYSE EN COMPOSANTE PRINCIPALE

- Visualition des données

REGRESSION LOGISTIQUE

- Structure et tendances des données

Deep Learning

RESEAUX DE NEURONES RECURENTS (RNN)

- Apprend à partir de données passées

LONG SHORT-TERM MEMORY (LSTM)

- Mémorise informations sur une longue période

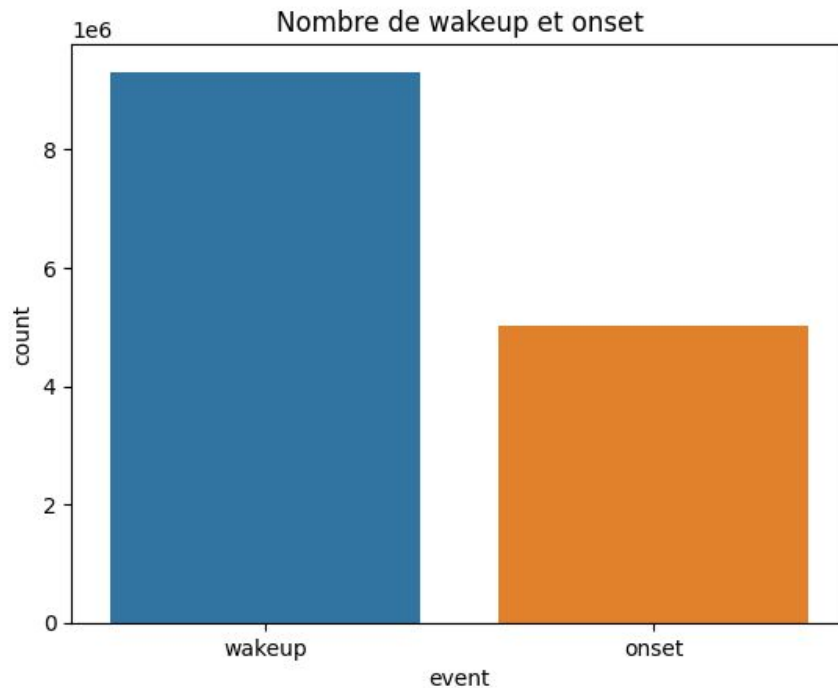
ANALYSES PRELIMINAIRES

Données, ACP et Régression Logistique

Visualisation des données

9 312 336 de wakeup

5 022 545 de onset



Analyse en Composante Principale (ACP)

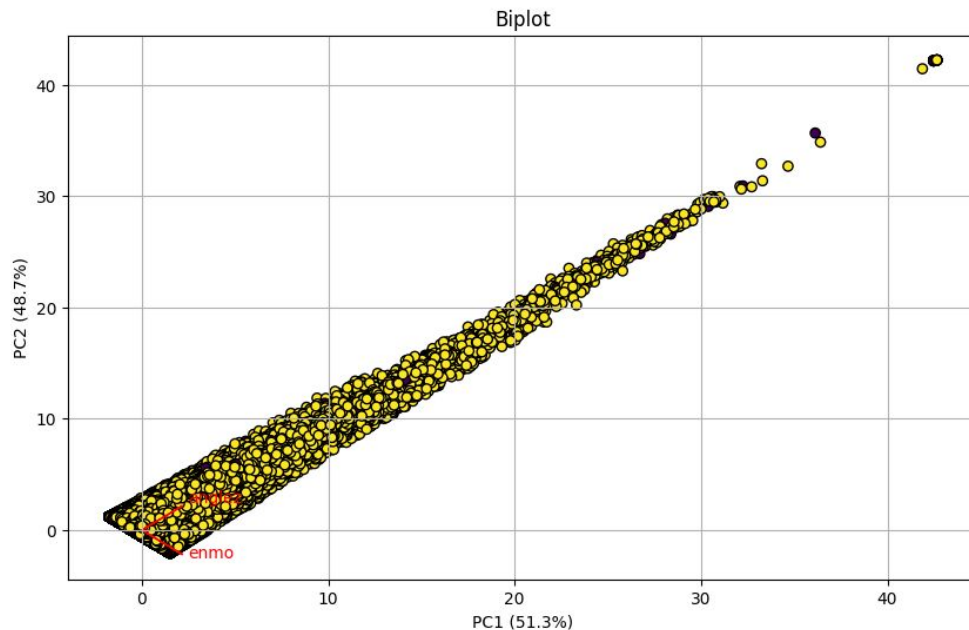
Wakeup en jaune & Onset en violet

PC1 explique 51,3% de la variabilité

PC2 explique 48,7% de la variabilité

Trajectoire diagonale marquée dominée par wakeup, particulièrement dans la zone où la variable enmo est mise en évidence

Événements onset moins présents et semblent plus dispersés (zone périphérique du graphique)



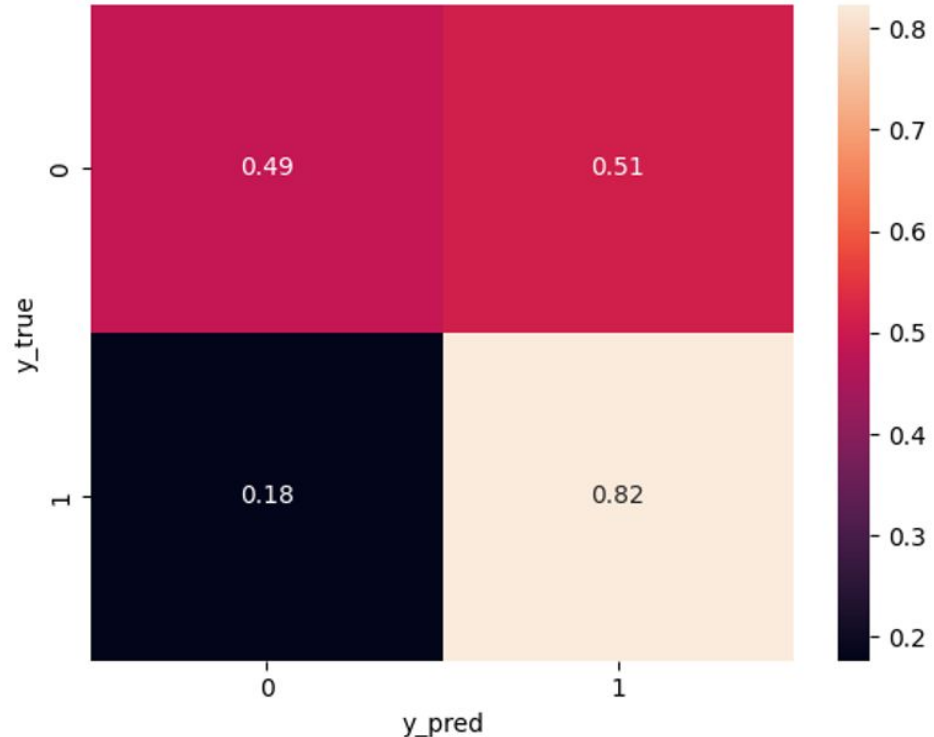
Régression Logistique

0,49 : Vrais Négatifs

0,51 : Faux Positifs

0,18 : Faux Négatifs

0,82 : Vrais Positifs

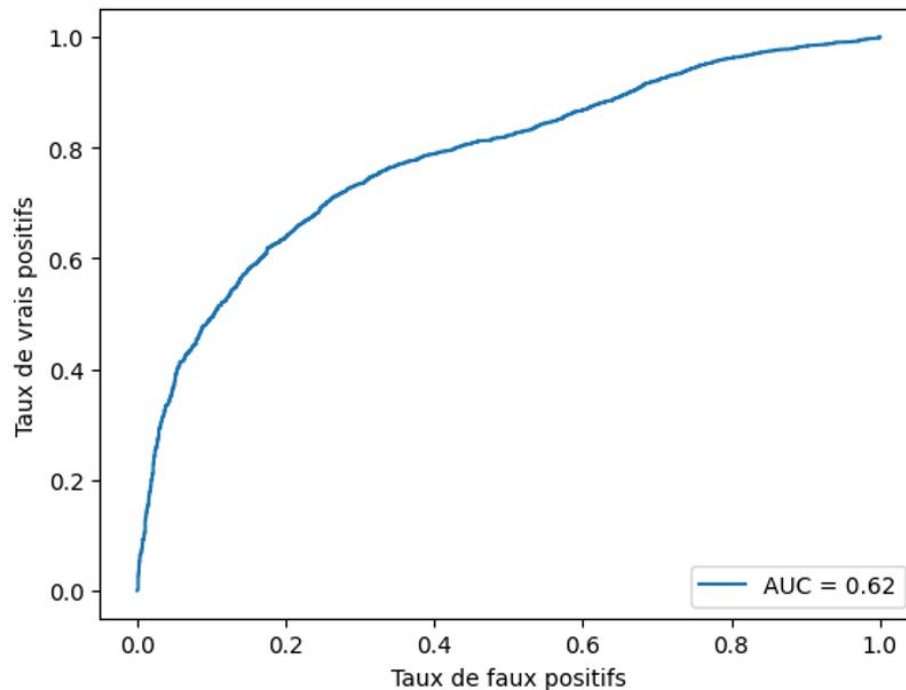


Régression Logistique

AUC de 0.62

La courbe s'écarte progressivement de la diagonale, ce qui indique que le modèle a une certaine capacité à discriminer nos classes

Cependant le modèle montre une capacité modérée à classer correctement les échantillons 'onset' et 'wakeup'



DEEP LEARNING

RNN ET LSTM

Long Short-Term Memory (LSTM)

Model: "sequential_1"

Layer (type)	Output Shape	Param #
=====		
lstm_2 (LSTM)	(None, 1, 74)	22792
dropout_2 (Dropout)	(None, 1, 74)	0
lstm_3 (LSTM)	(None, 1, 60)	32400
dropout_3 (Dropout)	(None, 1, 60)	0
time_distributed_1 (TimeDistributed)	(None, 1, 1)	61

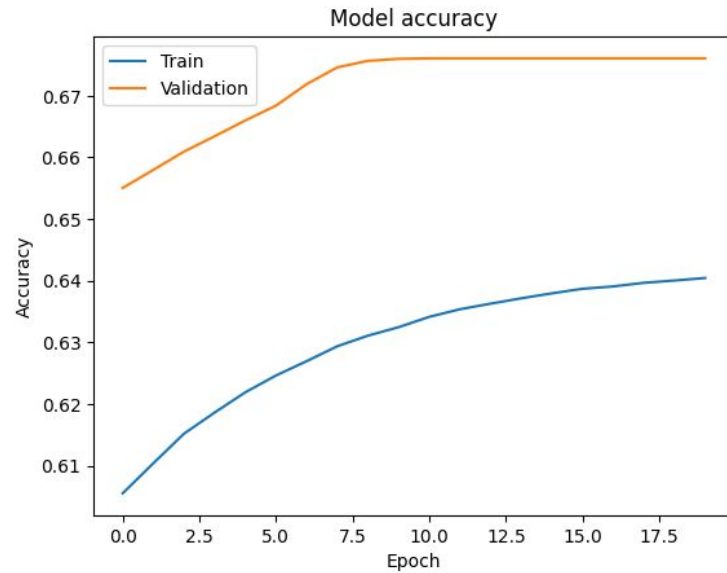
=====

Total params: 55253 (215.83 KB)

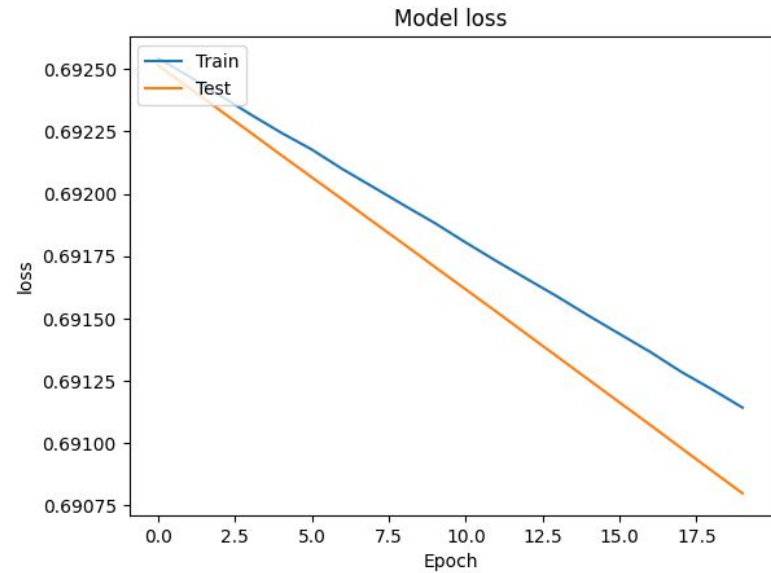
Trainable params: 55253 (215.83 KB)

Non-trainable params: 0 (0.00 Byte)

Long Short-Term Memory (LSTM)



Accuracy : 0.65



Loss : 0.69

Réseaux de Neurones Récurrents (RNN)

Model: "sequential_3"

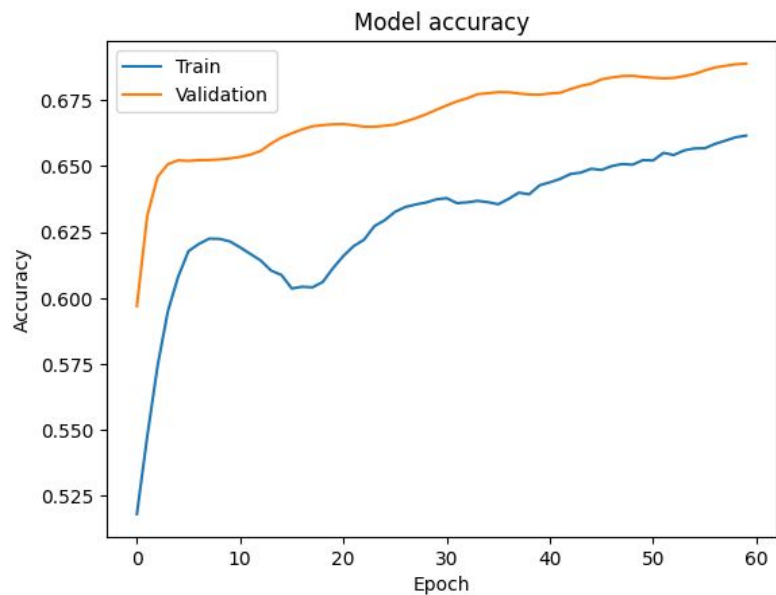
Layer (type)	Output Shape	Param #
masking_2 (Masking)	(None, None, 2)	0
simple_rnn_6 (SimpleRNN)	(None, None, 64)	4288
dropout_6 (Dropout)	(None, None, 64)	0
simple_rnn_7 (SimpleRNN)	(None, None, 128)	24704
dropout_7 (Dropout)	(None, None, 128)	0
simple_rnn_8 (SimpleRNN)	(None, None, 64)	12352
dropout_8 (Dropout)	(None, None, 64)	0
dense_2 (Dense)	(None, None, 1)	65

Total params: 41409 (161.75 KB)

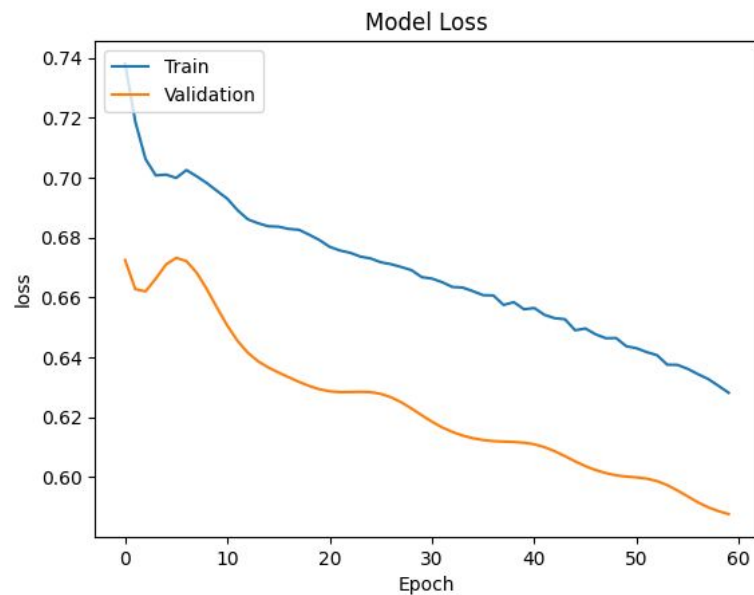
Trainable params: 41409 (161.75 KB)

Non-trainable params: 0 (0.00 Byte)

Réseaux de Neurones Récurrents (RNN)

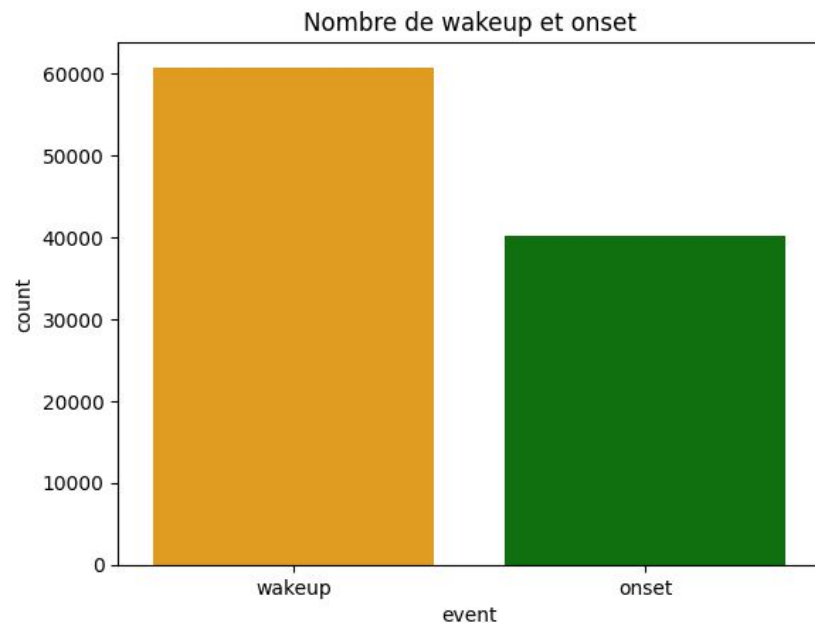
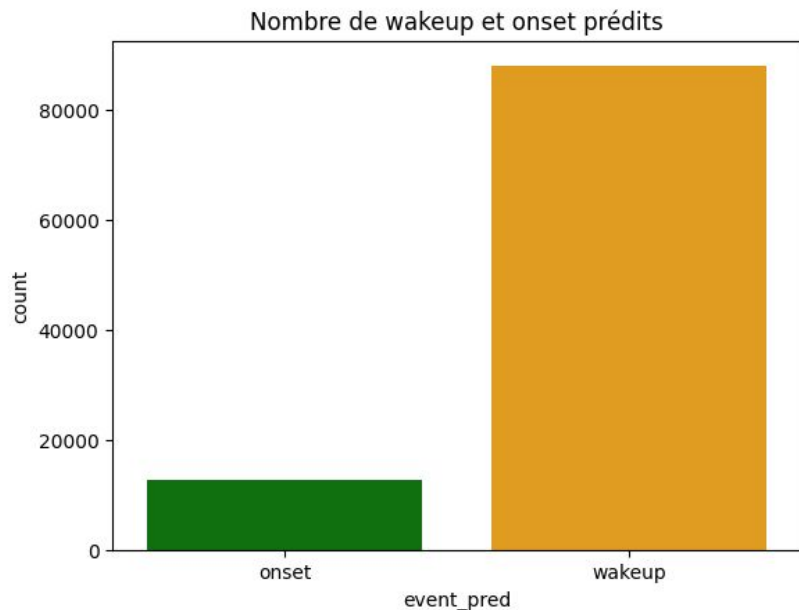


Accuracy : 0.64



Loss : 0.64

Focus sur le RNN



Prédiction de 12 741 onset et 88 142 wakeup

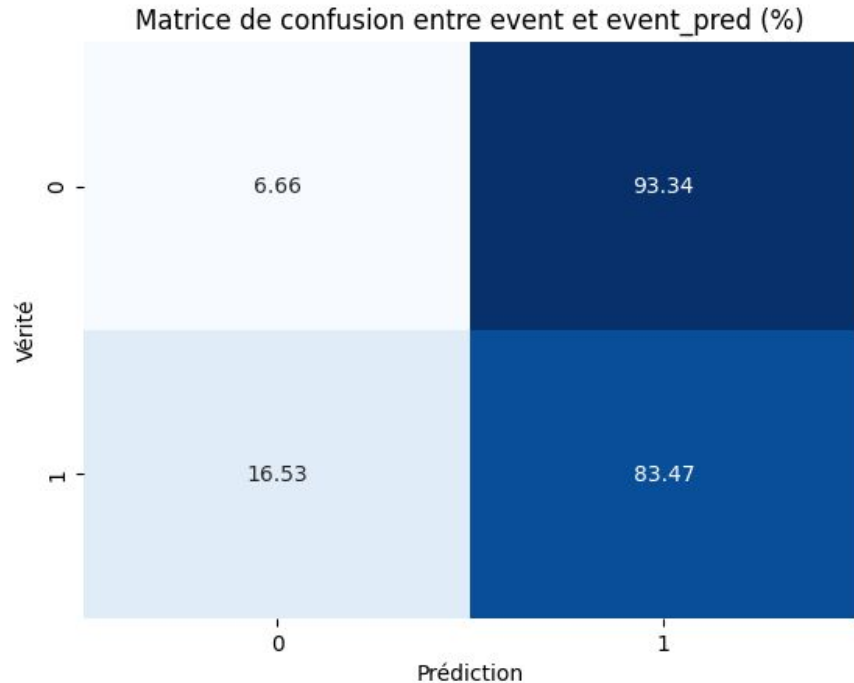
Focus sur le RNN

06.66 : Vrais Négatifs

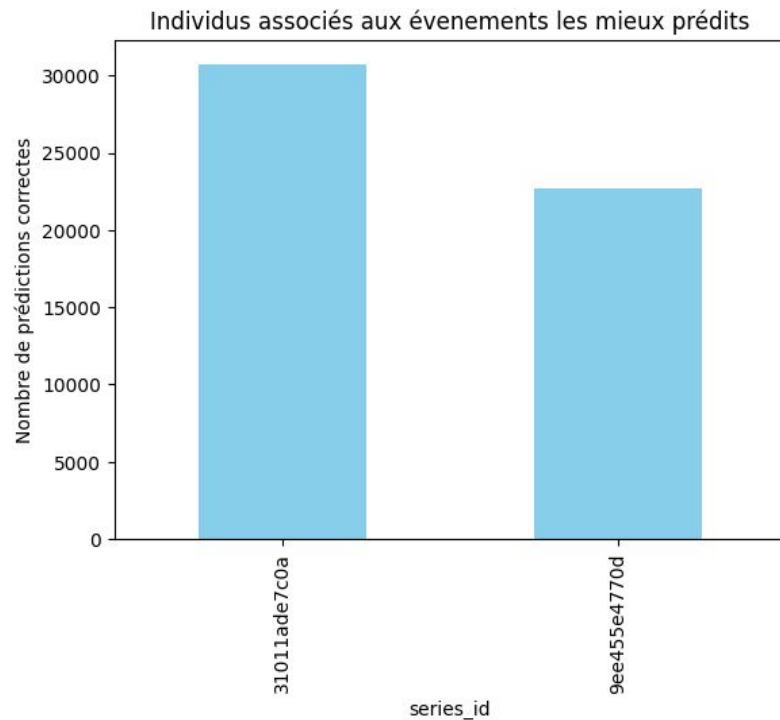
93.34 : Faux Positifs

16.53 : Faux Négatifs

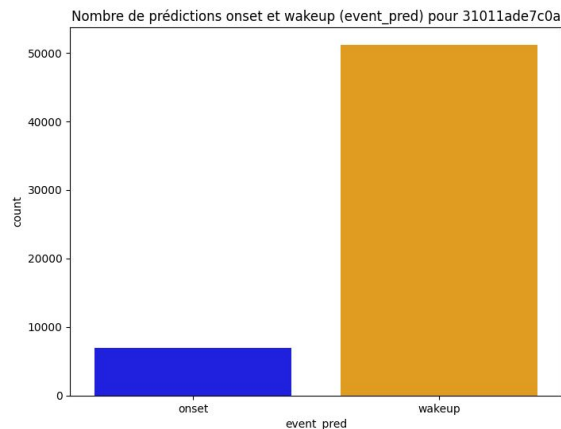
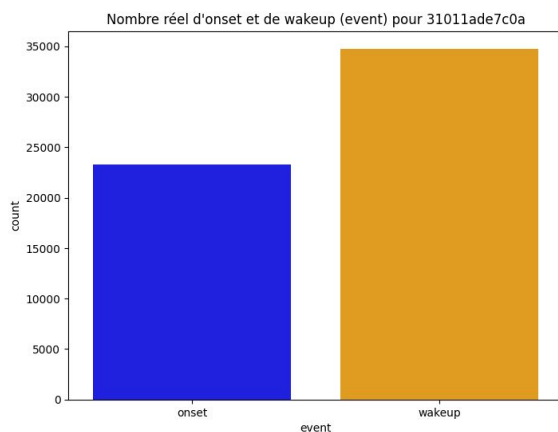
83.47 : Vrais Positifs



Focus sur le RNN



Focus sur l'individu 31011ade7c0a



Dispose du nombre de lignes le plus élevé dans notre ensemble de validation → plus de données pour une prédiction précise.

Tendance observée dans l'ensemble des données est observée ici: bonne prédiction des 'wakeup' mais prédiction plus complexes des 'onset'

CONCLUSION

Nos modèles :

Accuracy des modèles entre 64 et 65% et les loss ne convergent pas malgré une diminution progressive tout au long de l'entraînement du modèle.

Pas de discrimination possible entre nos deux classes avec la méthode ACP.

Bonnes prédictions de wakeup (83%) mais prédictions plus complexes pour onset (6%)

Perspectives :

Revoir pre-processing

Faire k-means ou UMAP

Faire CNN

REFERENCES

<https://www.kaggle.com/competitions/child-mind-institute-detect-sleep-states>