

BIG DATA - PROJECT

How our sleep is affected by activities during the day ? (physical activities, stress, ...)

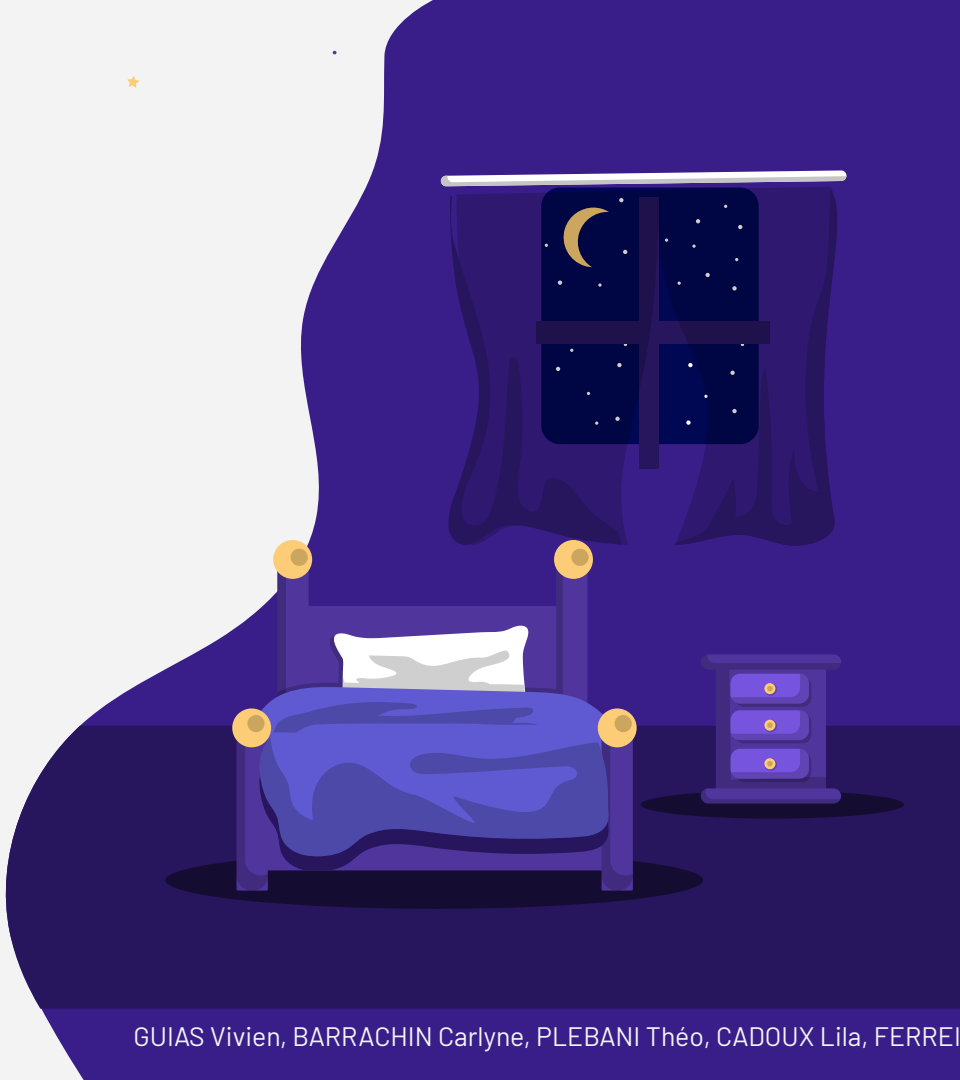




Table of contents

01

Introduction

Sleep deprivation induced illnesses
Sleep component

02

Data Ingestion

Get all the data needed (static and real time)

03

Data Storage and Processing

04

Machine Learning

XQBoost and Random Forest performance

05

Model usage

Using the Model for Predictions and Data Management

06

Visualization

Grafana



01

Introduction

IS Sleep Really Important?

Metabolic Disorders

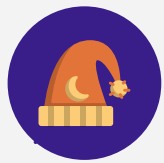


- Obesity
- Type II Diabetes
- Insulin Resistance

Cardiovascular Diseases



- Hypertension
- Angina,
- Myocardial infarction
- stroke

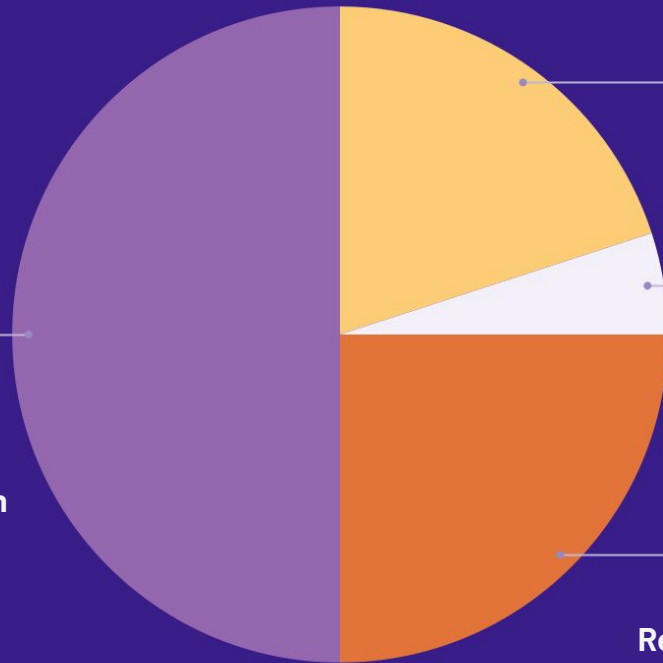


IDEAL SLEEP



LIGHT_SLEEP
50,0%

Transition
body regulation



REM_SLEEP
20,0%

Dreams, memory
processing

AWAKE
5,0%

normal
interruptions

DEEP_SLEEP
25,0%

Restoration, physical
repair



WHAT OTHER PARAMETERS COULD DETERMINE SLEEP QUALITY?

1

**Physical
Activity**

2

Stress

3

**Mid Day
Nap**

4

**Mental
Activity**

02

★ Data Ingestion

Get all the data needed (static and real time)



Static Data

Python API garminapi

Get static data and insert in the DataBase

Authentication

- Log into the API with a Garmin account (username/password)

Request

- Request the data :
 - Heart rate
 - Heart rate variability
 - Respiration
 - Stress
- Only if the data is not already in the database
- By checking the date of the data

Insertion

- Insert in MongoDB
- One table by type (steps, stress, etc.)
- Date has to be unique

```
_id: ObjectId('67641a4fbf8b0fd3b7cd025c')
date: "2024-05-16"
__v: 0
data: Object
  totalSteps: 7544
  totalDistance: 6292
  stepGoal: 6390
```

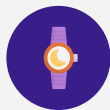
★ A record in MongoDB Steps Table



Real Time Data

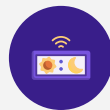
Garmin Watch Application

Get the Heart Rate in real time from a garmin Watch



Garmin App

- Garmin App Development
 - Send the heart rate to the server in real time (every 1s)
 - New language: MonkeyC
 - Limited documentation



NodeJS

- Server Development
- Host in *Render.com* (free)
- Handle data received from the app
- Insert data in MongoDB

03

Data Storage and Processing



Batch processing

Objective : process the raw data saved in mongodb

→ Making the visualization process and model prediction easier

- Connect to Mongo and access the data saved in multiple collections using Spark (pyspark) and mongo-spark connector
- Select the features to keep, needed for visualizations and predictions
- Handle missing values
- Join the data on date : summary for each day with selected features
- Insert data to MongoDB

Daily summary example

```
_id: ObjectId('6784de7937a14208148ae5fa')
date: "2024-05-16"
sleepScore: 70
overall_sleep_quality: "FAIR"
lightSleepSeconds: 16680
remSleepSeconds: 2460
deepSleepSeconds: 3780
sleepScoreFeedback: "NEGATIVE_RESTLESS"
sleepTimeSeconds: 22920
lowestRespirationValue: 0
avgWakingRespirationValue: 0
highestRespirationValue: 0
avgStressLevel: 0
maxStressLevel: 0
totalDistance: 6292
```

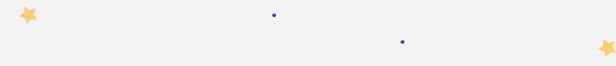


Processing on real-time data

Real-time processing

- Implementing Spark in the data retrieving process
- Unable to with the preexisting structure
- Near real-time, using only the structure presented before

Average heart-rate by minute

- Connect to MongoDB using Mongoose
 - Function to calculate average heart-rate on a defined time interval
 - Use cron to schedule the function to run every minute : **periodic batch-processing**
 - Insert result in MongoDB cleaned_data summaries
- 



04

Machine Learning

Predict sleep quality based on specific data



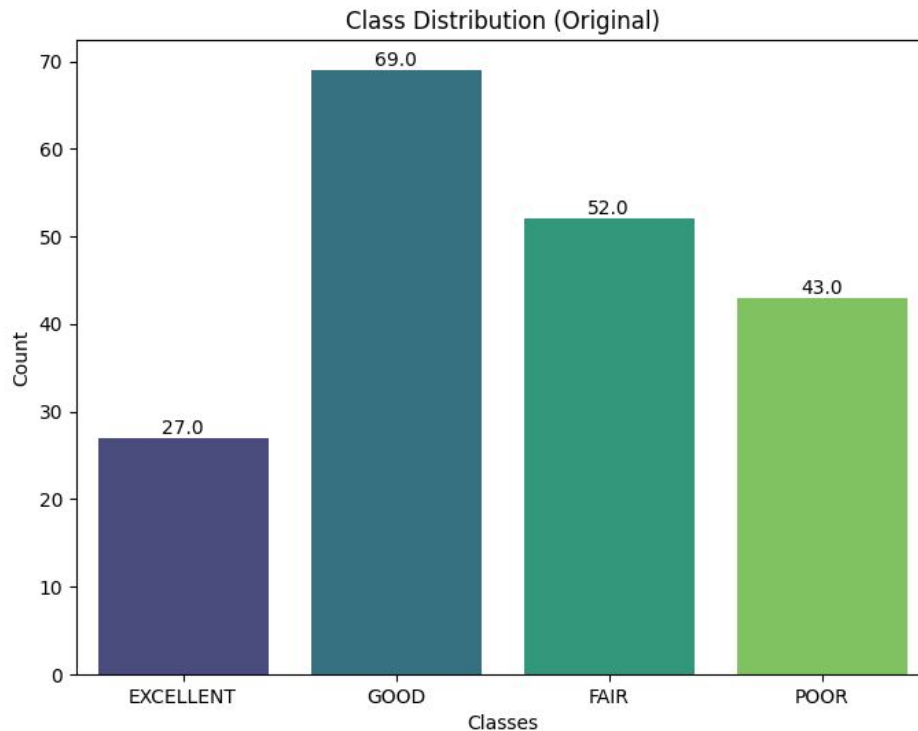
Initial Data

Total: 191

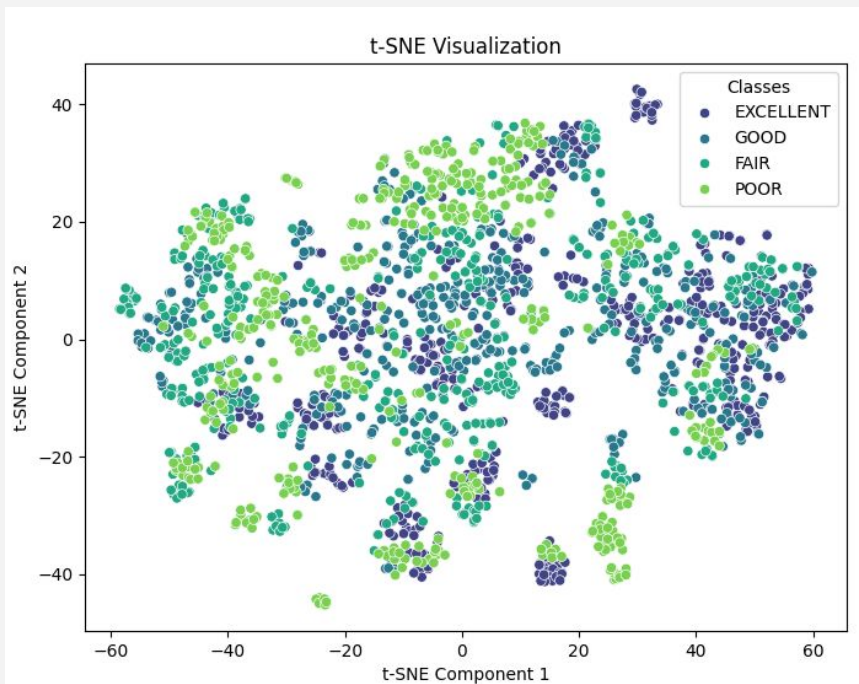
**Underrepresented
"EXCELLENT"**



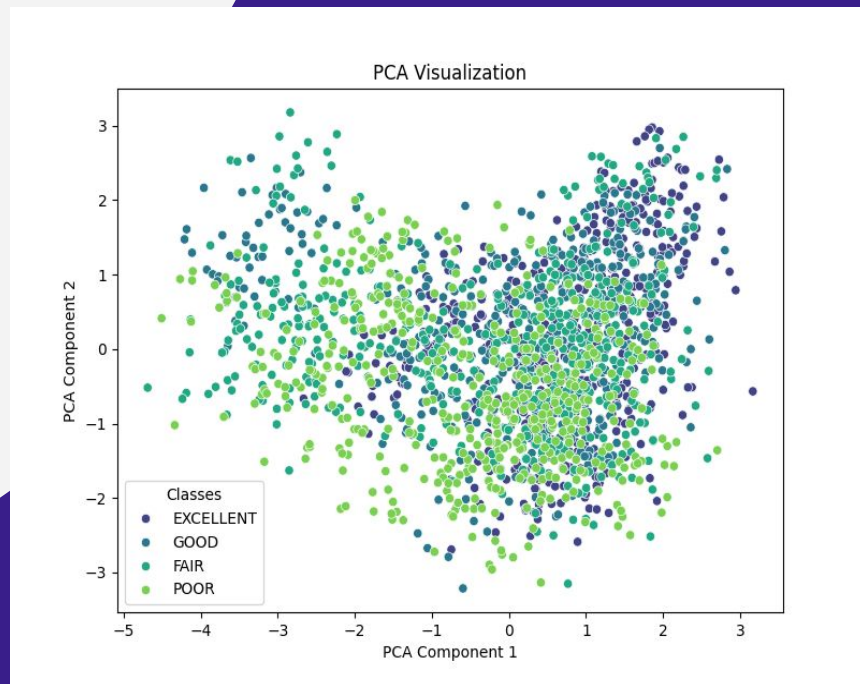
**Data_ augmentation
AUTOENCODER**



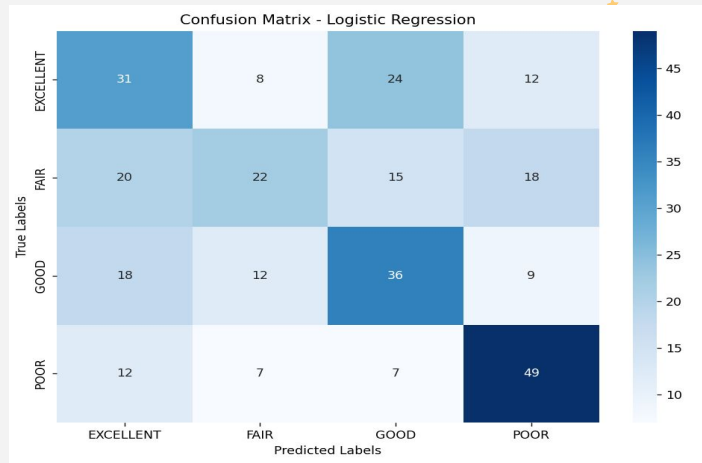
T-SNE



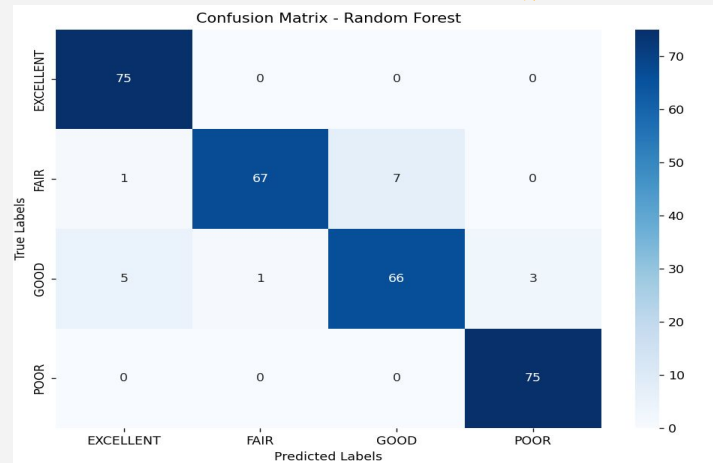
PCA



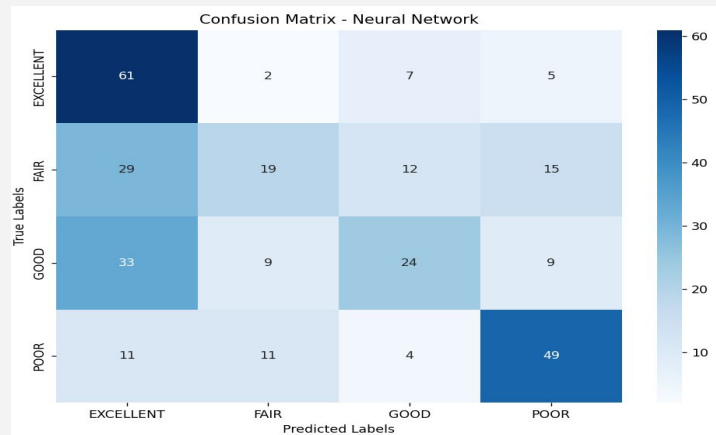
LR(46%)



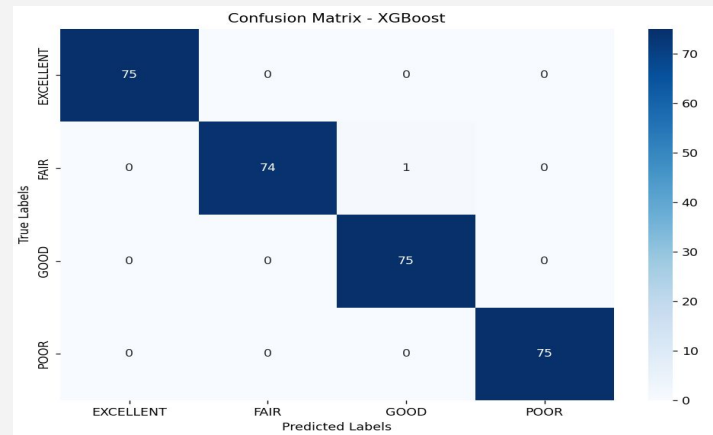
RF(93.4%)



NN(51%)



XGB(99.6%)





TOP FEATURES

RANDOM FOREST CLASSIFIER

- **Activity**
- **Stress**
- **HeartRate**
- **Napping**

Top Features:

	Feature	Importance
0	totalSteps	0.193610
1	totalDistance	0.190885
2	napTimeSeconds	0.107042
11	stress_avgStressLevel	0.093859
3	restingHeartRate	0.088402



05

★ Model usage

Predict sleep quality based on specific data

Model and Data Retrieval

Loading the Model and Scaler

- Use **joblib** to load the saved model and scaler
- Function **load_model_and_scaler** to load best_model.pkl and scaler.pkl

heartRate	spo2
heartRateVariability	steps
heartrates	stress
prediction	summaries
respiration	trainingStatus
sleep	users

Fetching Data from MongoDB

- Connect to MongoDB using **MongoClient**
- Function **get_data_from_mongo** to fetch data for a specific date
- Retrieve data from various collections (steps, sleep, heartRate, etc.)
- Handle missing values and errors

Prediction and Storage

Preparing Data for Prediction

- Convert fetched data to a numpy array
- Standardize feature values using the loaded scaler

Model Prediction

- Use the loaded model to make a prediction
- Map the prediction to a corresponding label
(**POOR, FAIR, GOOD, EXCELLENT**)

Storing Prediction in MongoDB

- Function **insert_prediction_to_mongo** to insert the prediction into the prediction collection
- Insert features, prediction, date, and timestamp into MongoDB

```
_id: ObjectId('6783f8fc25aa78209ce8722c')  
features: Object  
prediction: "FAIR"  
date: "2025-01-11"  
timestamp: 2025-01-12T19:16:44.221+00:00
```

06

Visualization



Grafana

Open-source data **visualization** and **monitoring** tool

Dynamic dashboards

Many **different** data sources: **MongoDB**

Real-time updates every **5s**: **synchronization** with the database

... Panels



Gauge

Real Time: Heart Rate

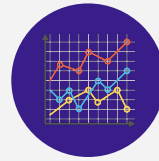
→ Identification of critical thresholds



Pie Chart

Overall Sleep Quality

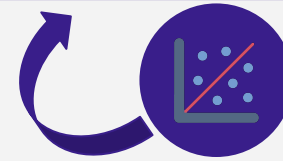
→ Distribution



Time Series Lines

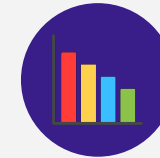
Features across **Time**:

- Distance
- Stress
- Sleep Duration
- Sleep Quality
- Napping
- Resting HR
- Real Time: Heart Rate



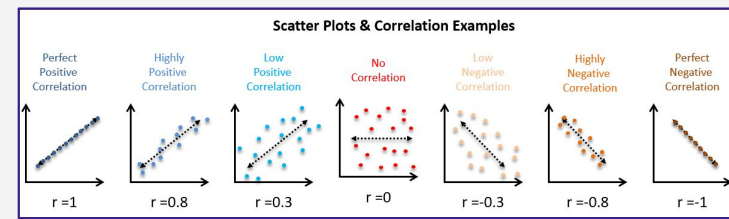
Scatter Plot (XY Chart)

→ Show the relationship between two variables

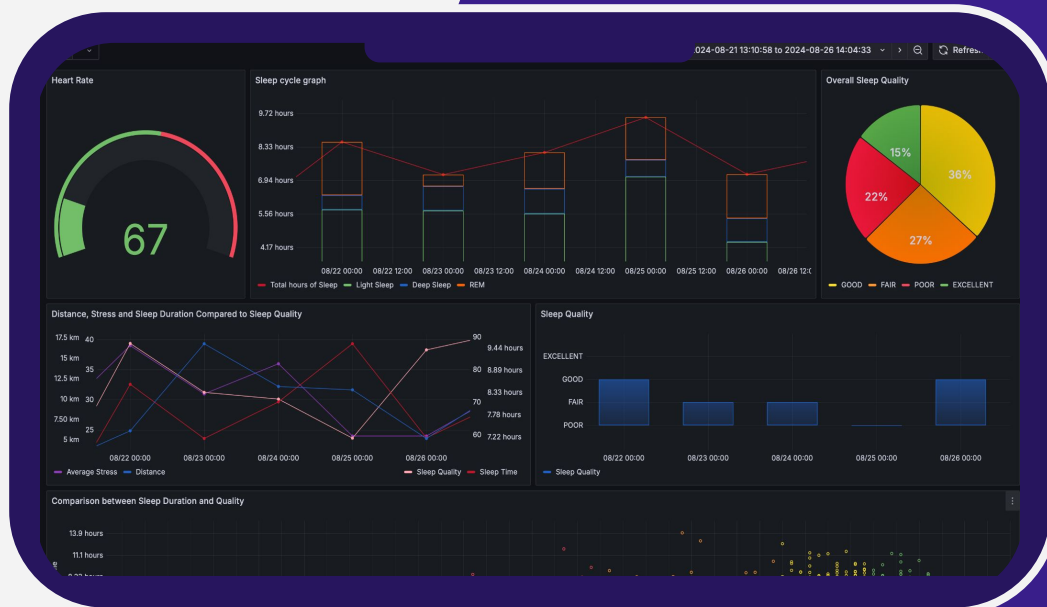


Time Series Bars

Sleep **Stages**
Sleep **Quality**



Demonstration





Resources

- (2023). Induced insufficient sleep syndrome. 402-410. doi: 10.1016/b978-0-12-822963-7.00063-3
- Bengkel, Mecca, Medina. (2023). Sleep deprivation. 137-141. doi: 10.1016/b978-0-323-91890-9.00023-4
- Joseph, Malachy, Burke., Joseph, Malachy, Burke., Mark, C., Haigney., Ryan, T., Borne., Mori, J., Krantz., Mori, J., Krantz. (2020). Smartwatch detection of ventricular tachycardia: Case series. *Heartrhythm Case Reports*, 6(10):800-804. doi: 10.1016/J.HRCR.2020.08.003
- Daisy, Duan., Lenise, Jihe, Kim., Jonathan, C., Jun., Vsevolod, Y., Polotsky. (2022). Connecting insufficient sleep and insomnia with metabolic dysfunction. *Annals of the New York Academy of Sciences*, 1519(1):117-94. doi: 10.1111/nyas.14926
- Veronika, Doroshenko., Kseniia, P., Meleha., Vladimir, Favoritov. (2024). The role of physical activity in improving the quality of sleep of patients with various sleep disorders. *Art of medicine*, 226-230. doi: 10.21802/artm.2024.3.31.226
- Jun, Zhang., Shungui, Xiang., Xiaowen, Li., Yin, Tang., Qun, Hu. (2024). The impact of stress on sleep quality: a mediation analysis based on longitudinal data. *Frontiers in Psychology*, 15 doi: 10.3389/fpsyg.2024.1431234
- Jianghong, Liu., Rui, Feng., Xiaopeng, Ji., Xiaopeng, Ji., Naixue, Cui., Naixue, Cui., Adrian, Raine., Sara, C., Mednick. (2019). Midday napping in children: associations between nap frequency and duration across cognitive, positive psychological well-being, behavioral, and metabolic health outcomes. *Sleep*, 42(9) doi: 10.1093/SLEEP/ZSZ126



THANK YOU

GUIAS Vivien, BARRACHIN Carlyne, PLEBANI Théo, CADOUX Lila, FERREIRA Mathieu