# Machine Learning for Fatigue Detection using Fitbit Fitness Trackers

icSports 2022

Erik Johannes Husom

October 28, 2022

# Machine Learning for Fatigue Detection using Fitbit Fitness Trackers

Authors:

- Erik Johannes Husom, *SINTEF Digital*, *Norway*
- Rustem Dautov, *SINTEF Digital*, *Norway*
- Adela-Aniela Videsjorden, *University of Oslo*, *Norway*
- Fotis Gonidis, *Gnomon Informatics SA*, *Greece*
- Spyridon Papatzelos, *Gnomon Informatics SA*, *Greece*
- Nikolaos Malamas, *Gnomon Informatics SA*, *Greece*

# Context and motivation

- Occupational health and safety
  - Manual labour
  - Field workers
  - Fatigue-caused accidents
- Long-term endurance sports
  - Training load
- Fatigue estimation and detection
  - Traditionally done using electroencephalogram (EEG)
  - Wearable devices (consumer-grade)
  - Continuous monitoring of biomarkers
  - Machine learning to produce predictive models
    - Extracting information from sequences
    - Suitable for learning from large amounts of data

SINTEF

# Fatigue

- Human tiredness
- Result of prolonged physical and mental activity
- Could be a symptom of medical conditions
- Safety hazard

# Fatigue assessment

- Electroencephalograms (EEG)[1]
  - Monitoring of brain waves
- Percentage eye openness tracking (PERCLOS)[2]
- General challenges:
  - Limited portability
  - Infrequent use

[1]Karuppusamy, N. S. and Kang, B.-Y. (2020). Multi-modal System to Detect Driver Fatigue Using EEG, Gyroscope, and Image Processing. *IEEE Access*, 8:129645–129667.

[2]Zhang, J., Chen, Z., Liu, W., Ding, P., and Wu, Q. (2021). A Field Study of Work Type Influence on Air Traffic Controllers' Fatigue Based on Data-Driven PERCLOS Detection. *International journal of environmental research and public health*, 18(22):11937.

# Fatigue assessment

- Wearable technology and fatigue
  - Activity levels
  - Number of steps
  - Heart rate
  - Sleep patterns
- Fitbit activity trackers (wristbands)

## Fatigue assessment

- Fatigue Assessment Scale (FAS)[3]
  - 10 statements rated on a scale from 1 (*never*) to 5 (*always*)
  - Fatigue score in the range 10-50
  - Estimated fatigue levels:
    - 10-21: No fatigue
    - 22-34: Fatigue
    - 35-50: Extreme fatigue

[3] Michielsen, H. J., De Vries, J., and Van Heck, G. L. (2003). Psychometric qualities of a brief self-rated fatigue measure: The fatigue assessment scale. *Journal of psychosomatic research*, 54(4):345–352.

# Fatigue Assessment Scale (FAS)

| | Never | Sometimes | Regularly | Often | Always |
|---|---|---|---|---|---|
| 1. I am bothered by fatigue (WHOQOL) | 1 | 2 | 3 | 4 | 5 |
| 2. I get tired very quickly (CIS) | 1 | 2 | 3 | 4 | 5 |
| 3. I don't do much during the day (CIS) | 1 | 2 | 3 | 4 | 5 |
| 4. I have enough energy for everyday life (WHOQOL) | 1 | 2 | 3 | 4 | 5 |
| 5. Physically, I feel exhausted (CIS) | 1 | 2 | 3 | 4 | 5 |
| 6. I have problems starting things (FS) | 1 | 2 | 3 | 4 | 5 |
| 7. I have problems thinking clearly (FS) | 1 | 2 | 3 | 4 | 5 |
| 8. I feel no desire to do anything (CIS) | 1 | 2 | 3 | 4 | 5 |
| 9. Mentally, I feel exhausted | 1 | 2 | 3 | 4 | 5 |
| 10. When I am doing something, I can concentrate quite well (CIS) | 1 | 2 | 3 | 4 | 5 |

Figure 2: FAS questionnaire

# Approach

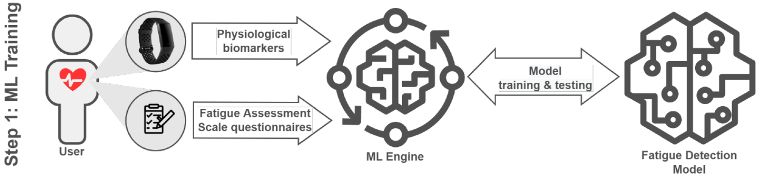



Figure 3: Conceptual archtiecture of our approach

# Approach

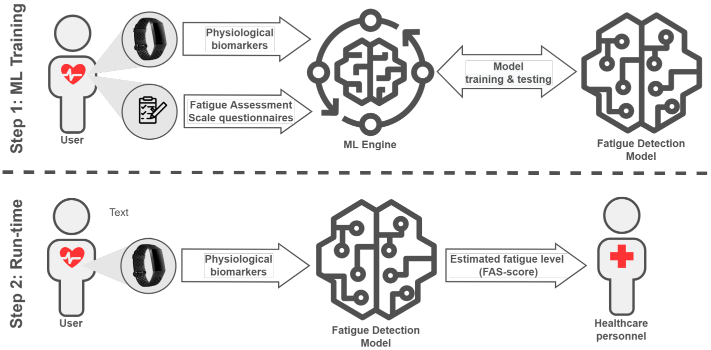



Figure 4: Conceptual architecture of our approach

# Approach

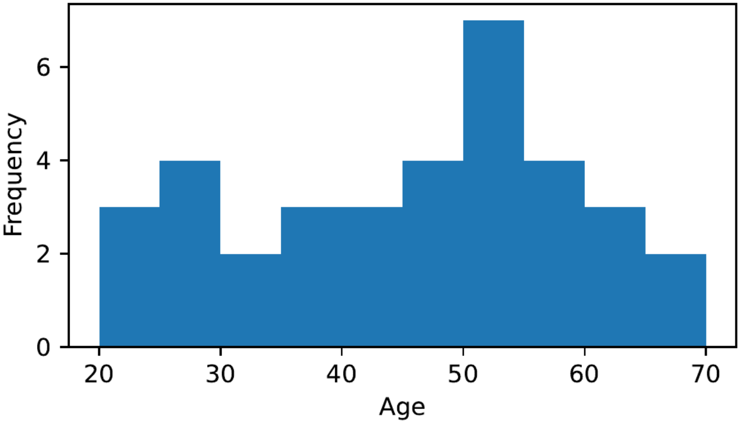
- Data collection
- Data preprocessing / feature selection
- Creating ML models

SINTEF

# Data collection participants

- 35 subjects
  - 31 female
  - 4 male
- Age: $45 \pm 13$ years

# Data collection participants



# Data collection participants

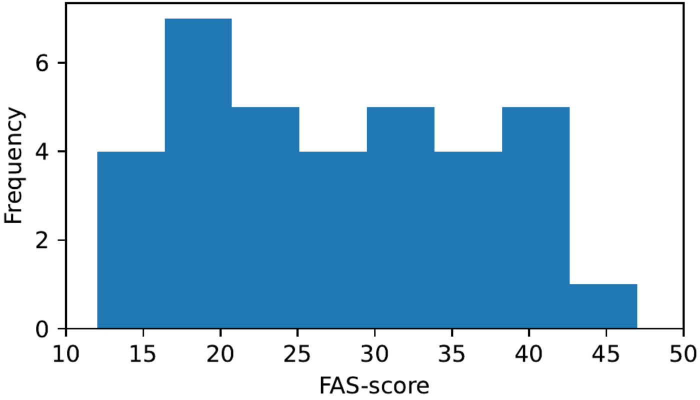



Figure 6: FAS-score distribution among participants

# Data collection

- Fitbit wristband worn for 7 days
- FAS-questionnaire filled out once

# Data preprocessing

| Type | Variable | Granularity |
|---|---|---|
| Activity | Calories burned | Daily |
| | Number of floors | Daily |
| | Sedentary minutes | Daily |
| | Lightly active minutes | Daily |
| | Fairly minutes | Daily |
| | Very minutes | Daily |
| | Number of steps | Daily |
| | Distance walked | Daily |
| Heart rate | Heart rate time series | 1 second |
| | Resting heart rate | Daily |
| Sleep | Duration | Daily |
| | Efficiency | Daily |
| | Start time | Daily |
| | End time | Daily |
| | Main sleep or nap | Daily |
| | Sleep stage duration | 1 second |
| | Number of occurences of sleep stage | Daily |

# Feature selection

| Feature name | Unit | PCC |
|---|---|---|
| Sleep total duration | minutes | 0.072 |
| Sleep efficiency | score, 0-100 | -0.066 |
| Deep sleep duration | minutes | -0.054 |
| **Light sleep duration** | minutes | **0.173** |
| REM sleep duration | minutes | -0.059 |
| Awake in bed duration | minutes | -0.029 |
| Deep sleep count | count | -0.047 |
| Light sleep count | count | 0.085 |
| REM sleep count | count | -0.109 |
| Awake in bed count | count | -0.012 |
| **Sedentary minutes** | minutes | **0.138** |
| Lightly active minutes | minutes | -0.082 |
| Fairly active minutes | minutes | 0.027 |
| **Very active minutes** | minutes | **-0.149** |
| **Average heart rate** | beats per min | **0.211** |
| **Minimum heart rate** | beats per min | **0.333** |
| Maximum heart rate | beats per min | 0.027 |
| **Resting heart rate** | beats per min | **0.265** |
| **Calories burned** | kcal | **-0.092** |
| Steps | count | -0.055 |
| Distance | meters | -0.068 |
| **Age** | years | **0.338** |
| **Gender** | female/male | **-0.394** |
| **Weight** | kg | **0.143** |
| **Height** | cm | **-0.193** |
| **Body Mass Index** | $kg/m^2$ | **0.275** |

# Feature selection

The input features were based on:

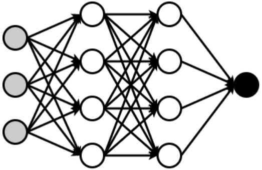
- Sleep
- Amount of activity/inactivity
- Heart rate
- Age
- Gender
- Weight
- Height

# Sequence as input

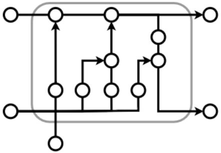
- Using multiple time steps as input to the model
- Capture temporal information
- Sequence length: $d$

# Machine learning algorithms

- Decision Tree (DT)
- Random Forest (RF)
- XGBoost, Extreme Gradient Boosting (XGB)
- k-Nearest-Neighbors (kNN)
- Fully-Connected Neural Network (FCNN)
- Long Short-Term Memory network (LSTM)



FCNN



LSTM

SINTEF

# Performance metrics

- Mean Squared Error (MSE):

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

- $\mathrm{R}^2$-score, coefficient of determination:

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

# Fine-tuning hyper-parameters of ML algorithms

| ML algorithm | Hyper-parameter | Values |
|---|---|---|
| DT | Max depth | $2, 5, 10, 15, 20, 50, 100$ |
| | Minimum samples split | $2, 5, 10$ |
| | Minimum samples leaf | $1, 3, 5$ |
| RF | Max depth | $2, 5, 10, 15, 20, 50, 100$ |
| | Number of estimators | $50, 100, 200, 400, 600, 800, 1000, 1200$ |
| | Minimum samples split | $2, 5, 10$ |
| | Minimum samples leaf | $1, 3, 5$ |
| XGB | Max depth | $2, 5, 10, 15, 20, 50, 100$ |
| | Number of estimators | $50, 100, 200, 400, 600, 800, 1000, 1200$ |
| | Learning rate | $0.3, 0.1, 0.001, 0.0001$ |
| kNN | Number of neighbors | $2, 4, 5, 6, 10, 15, 20, 30$ |
| | Weights | `uniform` or `distance` |
| | Leaf size | $10, 30, 50, 80, 100$ |
| | Algorithm | `ball_tree`, `kd_tree` or `brute` |

Figure 10: Search space of hyper-parameters

# Fine-tuning hyper-parameters of ML algorithms

- FCNN: 1 hidden layer with 8 nodes and Rectified Linear Unit (ReLU) activation in each node
- LSTM: 1 LSTM layer with 8 hidden units and sigmoid activation in each unit.

SINTEF

# Results: Hyper-parameters

| ML algorithm | Hyper-parameter | $d=1$ | $d=2$ | $d=3$ | $d=4$ | $d=5$ | $d=6$ | $d=7$ |
|---|---|---|---|---|---|---|---|---|
| DT | Max depth | 50 | 20 | 15 | 100 | 20 | 50 | 100 |
| | Minimum samples split | 10 | 10 | 10 | 2 | 5 | 10 | 2 |
| | Minimum samples leaf | 1 | 3 | 3 | 5 | 3 | 1 | 5 |
| RF | Max depth | 2 | 2 | 10 | 10 | 100 | 100 | 10 |
| | Number of estimators | 600 | 600 | 50 | 200 | 200 | 200 | 100 |
| | Minimum samples split | 5 | 5 | 10 | 10 | 2 | 2 | 5 |
| | Minimum samples leaf | 3 | 3 | 5 | 5 | 5 | 5 | 5 |
| XGB | Max depth | 50 | 15 | 15 | 15 | 15 | 5 | 20 |
| | Number of estimators | 400 | 50 | 50 | 50 | 50 | 800 | 400 |
| | Learning rate | 0.3 | 0.1 | 0.1 | 0.1 | 0.1 | 0.3 | 0.1 |
| kNN | Number of neighbors | 30 | | | | | | |
| | Weights | `distance` | | | | | | |
| | Leaf size | 10 | | | | | | |
| | Algorithm | `kd_tree` | | | | | | |
| FCNN | Number of layers | 1 | | | | | | |
| | Number of nodes in each layer | 8 | | | | | | |
| LSTM | Number of units | 8 | | | | | | |
| | Dropout rate | 0.2 | | | | | | |

Figure 11: Optimal configuration of hyper-parameters

# Results: Model comparison

| | DT | | RF | | XGB | | kNN | | FCNN | | LSTM | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *d* | MSE | $R^2$ | MSE | $R^2$ | MSE | $R^2$ | MSE | $R^2$ | MSE | $R^2$ | MSE | $R^2$ |
| 1 | 141.7 | -0.410 | 159.6 | -0.588 | 127.2 | -0.266 | 84.5 | 0.159 | 29.4 | 0.708 | 104.3 | -0.038 |
| 2 | 136.9 | -0.334 | 156.9 | -0.529 | 123.2 | -0.201 | 91.9 | 0.104 | 24.1 | 0.765 | 117.8 | -0.149 |
| 3 | 150.9 | -0.430 | 154.7 | -0.467 | 125.3 | -0.188 | 102.3 | 0.030 | 22.4 | 0.787 | 110.3 | -0.046 |
| 4 | 194.3 | -0.795 | 147.8 | -0.364 | 121.2 | -0.119 | 108.9 | -0.005 | 21.4 | 0.803 | 126.5 | -0.169 |
| 5 | 146.7 | -0.304 | 138.6 | -0.232 | 102.8 | 0.086 | 115.6 | -0.027 | **20.8** | **0.815** | 130.4 | -0.159 |
| 6 | 144.2 | -0.207 | 137.7 | -0.153 | 119.5 | -0.001 | 116.7 | 0.023 | 25.3 | 0.788 | 205.9 | -0.724 |
| 7 | 180.8 | -0.362 | 151.6 | -0.142 | 137.2 | -0.033 | 146.0 | -0.022 | 27.4 | 0.794 | 135.8 | -0.023 |

Figure 12: Model performance for the various machine learning algorithms

# Results: Best performing model

Best performing model:

- Fully-Connected Neural Network (FCNN)
  - 1 hidden layer
  - 8 nodes
- 5 time steps (days) as input
- $R^2$-score: 0.815
- Average error of 18% on the test set.

# Results

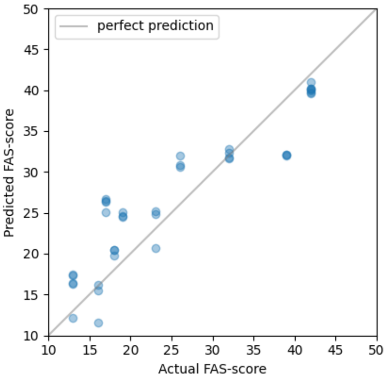



Figure 13: True vs predicted values of the best model

# Results

- Similar research showed using deep learning outperformed traditional ML method[4]

[4] Bai, Y., Guan, Y., and Ng, W.-F. (2020). Fatigue assessment using ECH and Actigraphy sensors. In *Proceedings of the 2020 International Symposium on Wearable Computers*, pages 12–16.

# Limitations

- Limited number of participants
- Gender imbalance
- High-level “black box” features from Fitbit

# Conclusion

- Fatigue estimation using ML
- Predictions based on biomarkers from wearable technologies
- Occupational fatigue-related hazards
- Future works
  - Validation with more users over longer timeframes
  - Continual learning
  - Federated learning

Technology for a better society