Accelerating the future of aerospace

Eye-tracking & AI: Classification of ATCOs Fatigue and Workload using Machine Learning

Laurie Marsman & Michelle Bont NLR | 08-05-25

BACKGROUND OF THE STUDY

In December 2022, EASA commissioned a research study on the **impact analysis, prevention, and management of ATCOs fatigue in the European Union**. The study, lead by NLR, was conducted in a scientific and objective manner, supported by data collection and various research methods.

The study included three tasks:

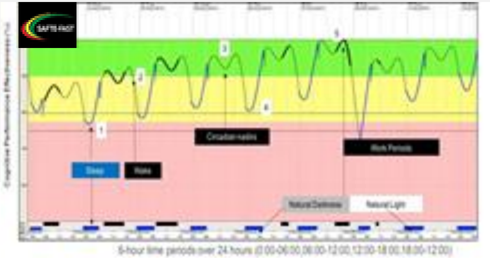
1. An evaluation of the implementation of EU regulations on this issue, notably Commission Implementing Regulation (EU) 2017/373, which imposed on air traffic service providers specific requirements linked to ATCOs stress, fatigue and rostering systems as part of their safety management systems
2. **Scientific research and data collection on ATCO fatigue causes and impacts through fatigue science methodologies**
3. An assessment of the possible impact of the introduction of new technologies on the ATCOs' workload and fatigue

Methodology

Validate subjective measurements

Roster Analysis

Involving 16 ATSPs and 24 actual rosters.



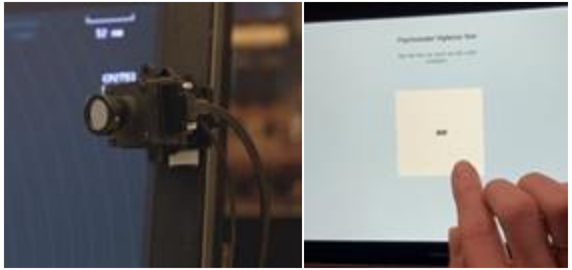
	0 points	1 point	2 points	4 points	8 points
Total hours over 7 days	≤ 36 h	36.1h - 43.9h	44h - 47.9h	48h - 54.9h	≥ 55h
Longest duty	≤ 8h	8.1h - 9.9h	10h - 11.9h	12h - 13.9h	≥ 14h
Shortest rest between duties	≥ 16h	15.9h - 13h	12.9h - 10h	9.9h - 7.9h	≤ 8h
Night work over 7 days	0h	0.1h - 8h	8.1h - 16h	16.1h - 23.9h	≥ 24h
Rest days	> 1 in 7 days	≤ 1 in 7 days	≤ 1 in 14 days	≤ 1 in 21 days	≤ 1 in 28 days

Data Collection (Subjective)

On fatigue and sleep for at least 10 days involving 6 ATSPs and 216 ATCOs.

Data Collection (Objective)

Using objectives measurements - Continuous **eye tracking** and a pre- and post-duty performance during shifts involving 5 ATSPs and 20 ATCOs.



Approach of objective measurements

- Objectives of the ATCO fatigue study
  - **Validate subjective fatigue measurements**
  - Determine the **feasibility** of objective measurement equipment to measure fatigue, in real-time, during the ATC operation

- 4 volunteering ATCOs within each participating ATSP
- Measurement of objective fatigue during main hotspots (as determined in roster analysis)
  - **Continuous eye tracking** during entire shift
  - **Subjective workload** ratings (RSME and ISA, hourly)
  - **Subjective fatigue** ratings (KSS and SP, hourly)

Objective measurements – Eye tracking

Data was collected during the shift(s) that were determined to be the main fatigue hotspots for each ATSP

Remote eye tracking

- SmartEye Pro & SmartEye Aurora
- Per ATCO during entire shift (6-8 hours) to validate feasibility and subjective fatigue measurement.
- Resources and practical/operational conditions limited sample to 4 ATCOs per ATSP.

This study resulted in a detailed dataset with both objective and subjective data on ATCO workload and fatigue

ATCO fatigue dataset

Null hypothesis significance testing*

Machine Learning classification

* Study on the Analysis, Prevention and Management of Air Traffic Controller Fatigue, via EASA Website
* Marsman, L.A. et al. (2024). Results and implications of objective fatigue and performance measurements in five European Air Traffic Service Providers. (*EAAP pre-print*)

Research question:

“To what extent can eye-tracking features accurately classify operator fatigue and workload in selected European ATCOs by applying Machine Learning classification?”

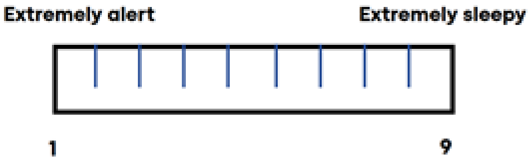
Sub-questions:

"Which Machine Learning models perform best in classifying fatigue vs. workload?"

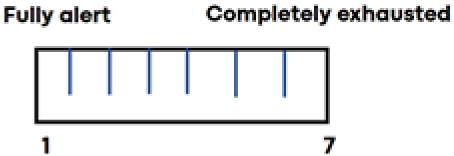
“Do different features have different importance in fatigue vs. workload?”

Subjective measures for ‘ground truth’

Fatigue scales

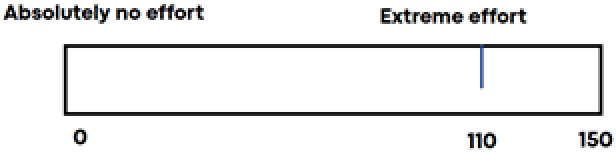


Karolinska Sleepiness Scale (KSS)

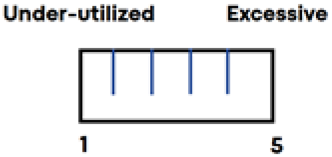


Samn-Perelli Scale (SP)

Workload scales



Rating Scale Mental Effort (RSME)



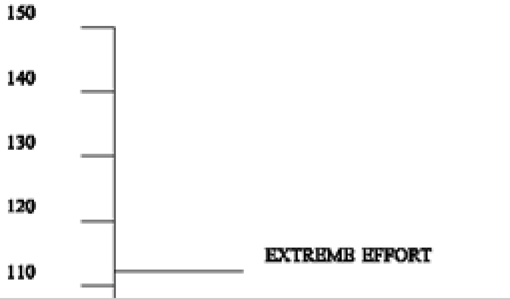
Instantaneous Self-Assessment (ISA)

Steps:

- Ob- and subjective measures
- ± Every hour
- Answers truth for binary classification
- 0 = low, 1 = high
- Determine cut-off

Methodology: two lal

Median-based



Rating	Workload	Description
1	Under-utilised	Nothing to do. Rather boring
2	Relaxed	More than enough time for all tasks. Active on the task less than 50% of the time
3	Comfortably busy pace	All tasks well in hand. Busy but stimulating pace. Could keep going continuously at this level
4	High	Non-essential task suffering. Could not work at this level very long
5	Excessive	Behind on tasks; losing track of the full picture

Source: Adapted from Kirwan et al. (1997)

Visualisation

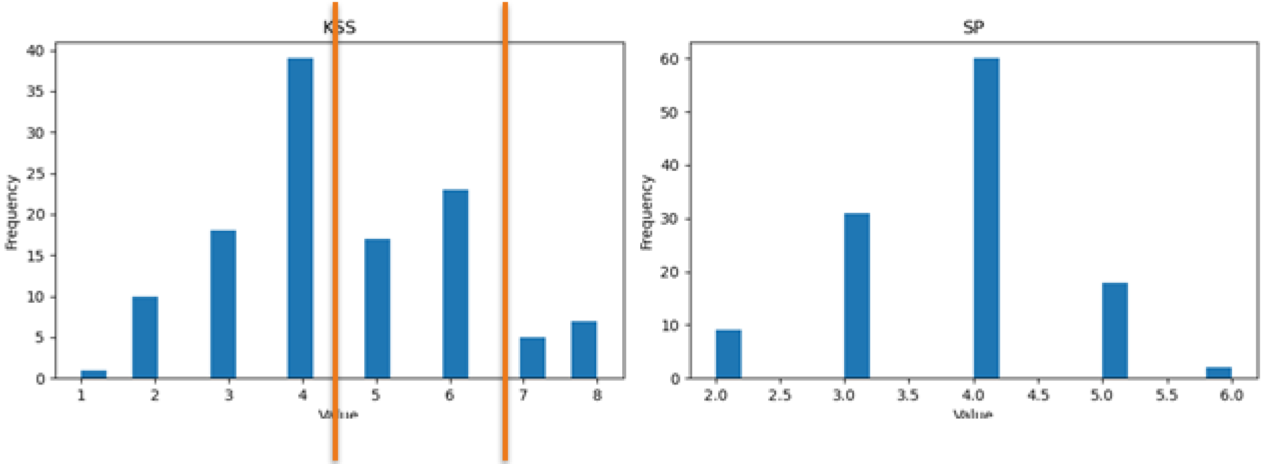


Fig. 1: Data distribution of the KSS and the SP

Eye-tracking features for classification

- Blink Duration in seconds
- Blink Frequency in blinks per minute
- PERCLOS 60, 70 & 80 (PERcentage of eye CLOSure)

- Cat or dog?

Results classification

Table 1. Fatigue models comparing metrics Median vs Literature Split:

Models	Median Split				Literature Split			
	Accuracy	AUC	F1 Score	MCC	Accuracy	AUC	F1 Score	MCC
CatBoost Classifier	0.88	0.93	0.83	0.74	0.92	0.79	0.00	0.00
Random Forest Classifier	0.92	0.99	0.88	0.84	0.92	0.50	0.00	0.00
Gradient Boosting Classifier	0.92	0.98	0.90	0.84	0.92	0.82	0.00	0.00

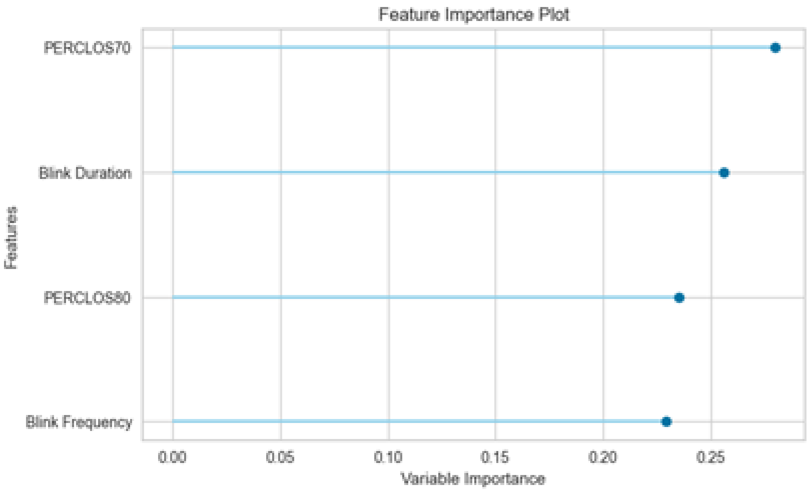
Table 2. Workload models comparing metrics Median vs Literature Split:

Models	Median Split				Literature Split			
	Accuracy	AUC	F1 Score	MCC	Accuracy	AUC	F1 Score	MCC
SVM	0.48	0.50	0.07	-0.01	0.76	0.55	0.25	0.11
K Neighbors Classifier	0.52	0.55	0.45	0.05	0.92	0.92	0.75	0.70
Ridge Classifier	0.46	0.47	0.37	-0.07	0.80	0.53	0.17	0.07

Class imbalance:
- F1 & MCC

Results feature importance

Fatigue



Workload

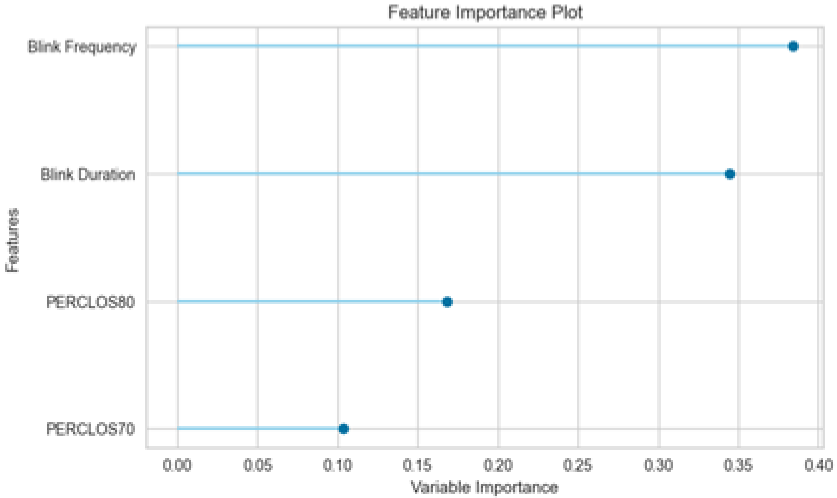


Fig. 2: Feature importance plots for fatigue vs. workload

Discussion & Conclusion

- 92% accuracy
- Different models for fatigue and workload
- Different features important

1. Class imbalance
2. Multi-class classification: low, med, high
3. PERCLOS
4. Report fatigue & workload

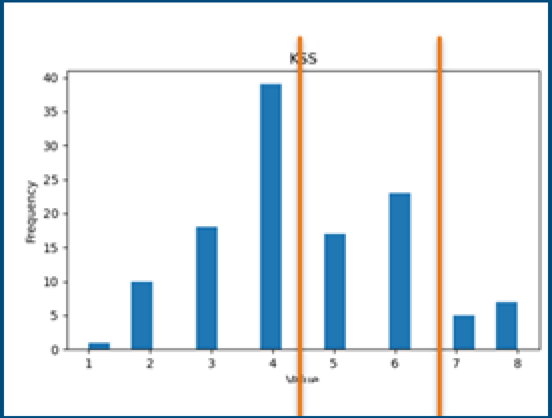


Fig. 3: Visualisation class imbalance

Next steps…

- Validation experiment
- Training shifts of 45 min, continuous eye-tracking
- Labels asked before, during and after

Accelerating
the future
of aerospace

Thank you for your attention!

+31 6 83999143

Michelle.bont@nlr.nl

LinkedIn

Michelle Bont

Student MSc Artificial Intelligence