Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

75 NRP
Big Data
National Research Programme

PhoneMD:
Learning to Diagnose Parkinson's Disease from Smartphone Data

Patrick Schwab and **Walter Karlen**

@schwabpa @mhsl_ethz

Institute of Robotics and Intelligent Systems
ETH Zurich


ALI
REZNICK Productions

Parkinson's Disease (PD)

- Slow **degeneration of motor skills**
- **Hard to diagnose**
    - Assessment of **symptoms**
        - **Similar** symptoms in other diseases
        - Symptom **fluctuations**
- Only ~80% of diagnoses are accurate[1]
- **~7m (0.3%)** affected, **120,000** deaths[2]

[1] Rizzo, G. et al. (2016) *Accuracy of clinical diagnosis of Parkinson disease: A systematic review and meta-analysis.* Neurology 86 (6).

[2] de Lau, LM and Breteler MM. (2006) *Epidemiology of Parkinson's disease.* Lancet Neurology 5 (6).

Wide Variety of Symptoms

Cognition

Speech

Dexterity

Movement

Motor Impairments

Tilted Posture

Rigidity

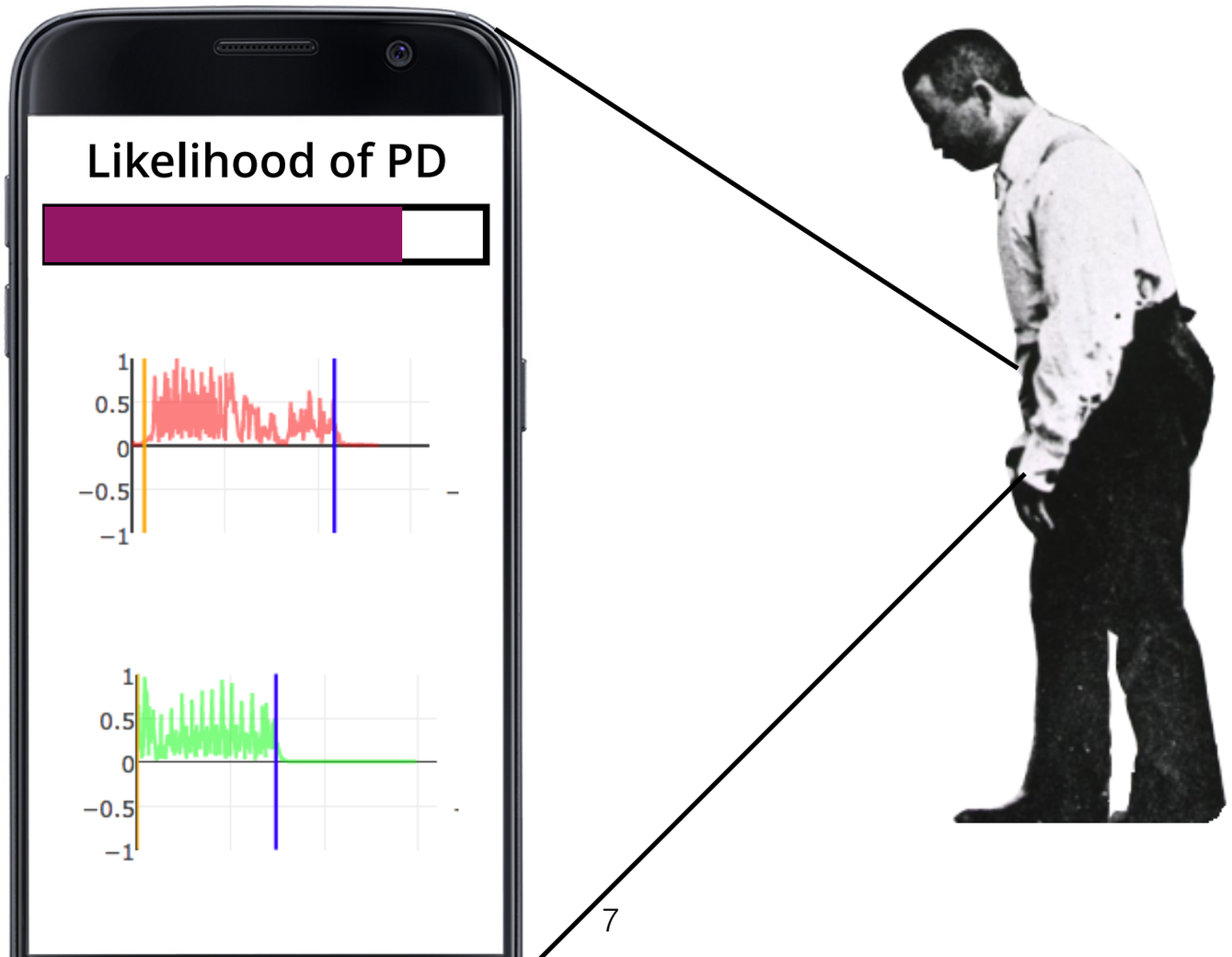
Reduced Arm Movement

Tremor of Extremities

Shuffling Gait & Short Steps

The Idea

Can we use machine learning on long-term smartphone data to diagnose Parkinson's?



The Dataset

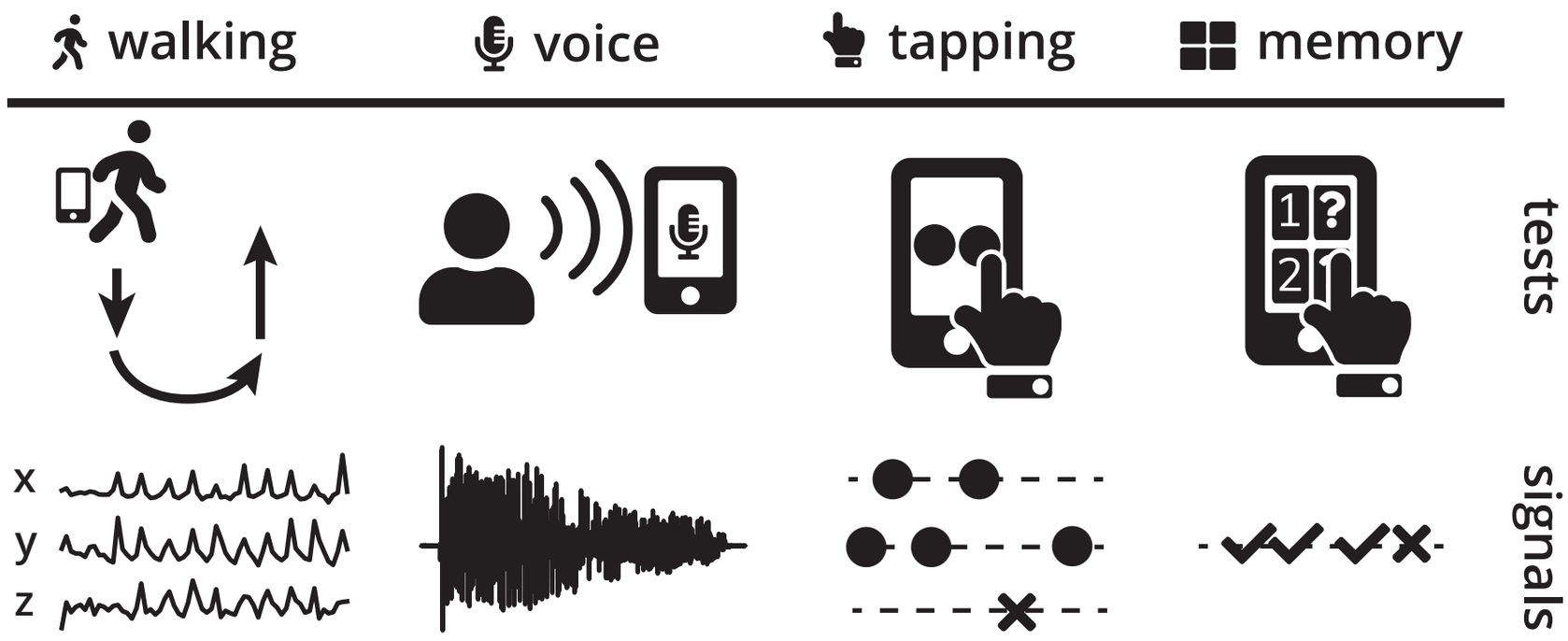
The mPower Study

- We use data collected in the **mPower study**[1]
    - Openly available on **Synapse**[2]
- App users (with and without Parkinson's, n=1853) were asked to perform several tests regularly
    - Outcome: **Prior clinical PD diagnosis**

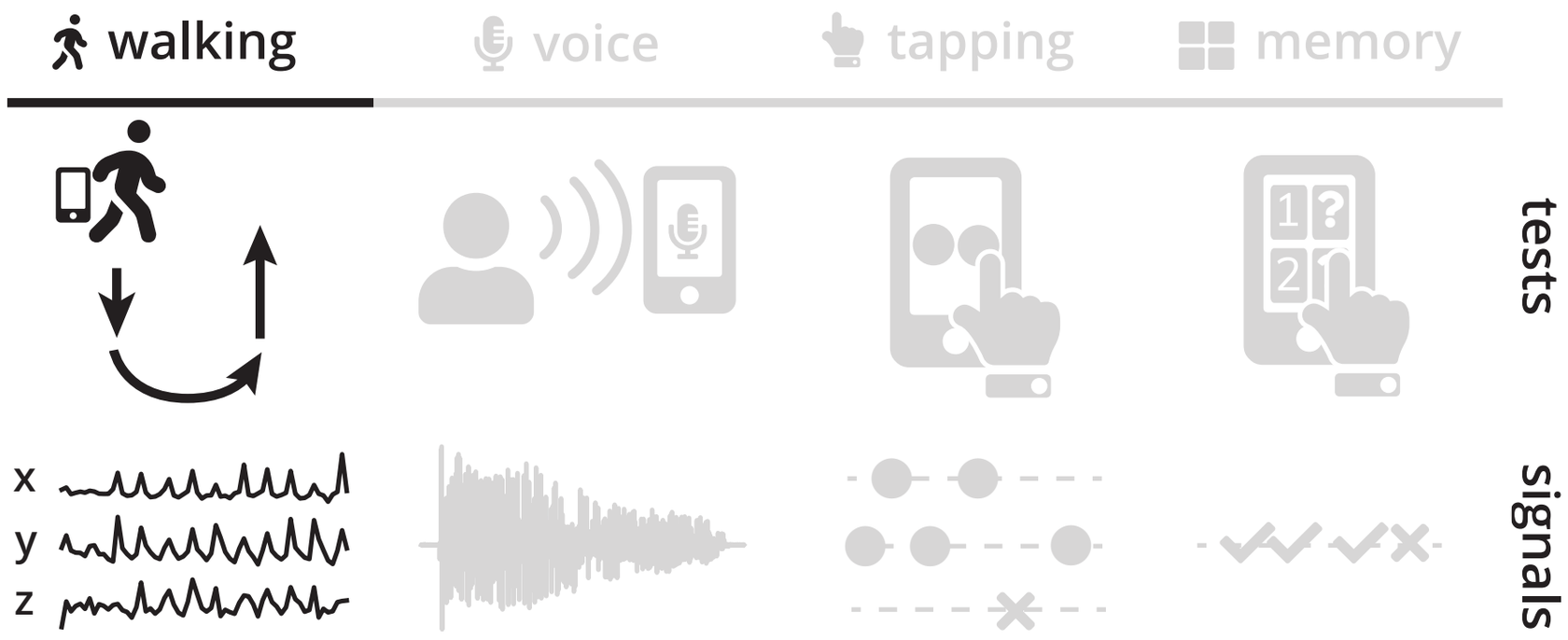
[1] Bot, B.M., et al. (2016) *The mPower study, Parkinson disease mobile data collected using ResearchKit.* Scientific data 3.

[2] Synapse Platform, https://www.synapse.org/#!Synapse:syn8717496 (Accessed: Nov 13, 2017)

Tests Overview



Tests Overview

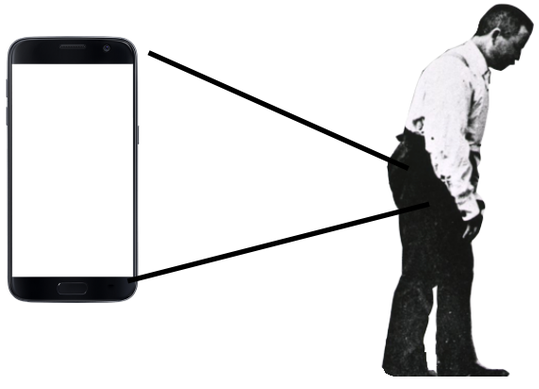


mPower Walking Test

- In the walking task, participants were asked to do the following three-segment task (each ~30s):

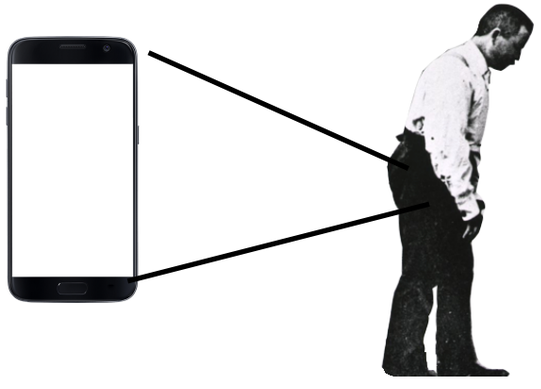
mPower Walking Test

- In the walking task, participants were asked to do the following three-segment task (each ~30s):



mPower Walking Test

- In the walking task, participants were asked to do the following three-segment task (each ~30s):

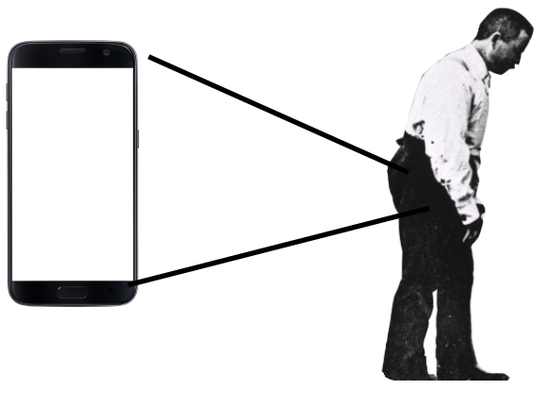


(1) Walk outbound

mPower Walking Test

- In the walking task, participants were asked to do the following three-segment task (each ~30s):

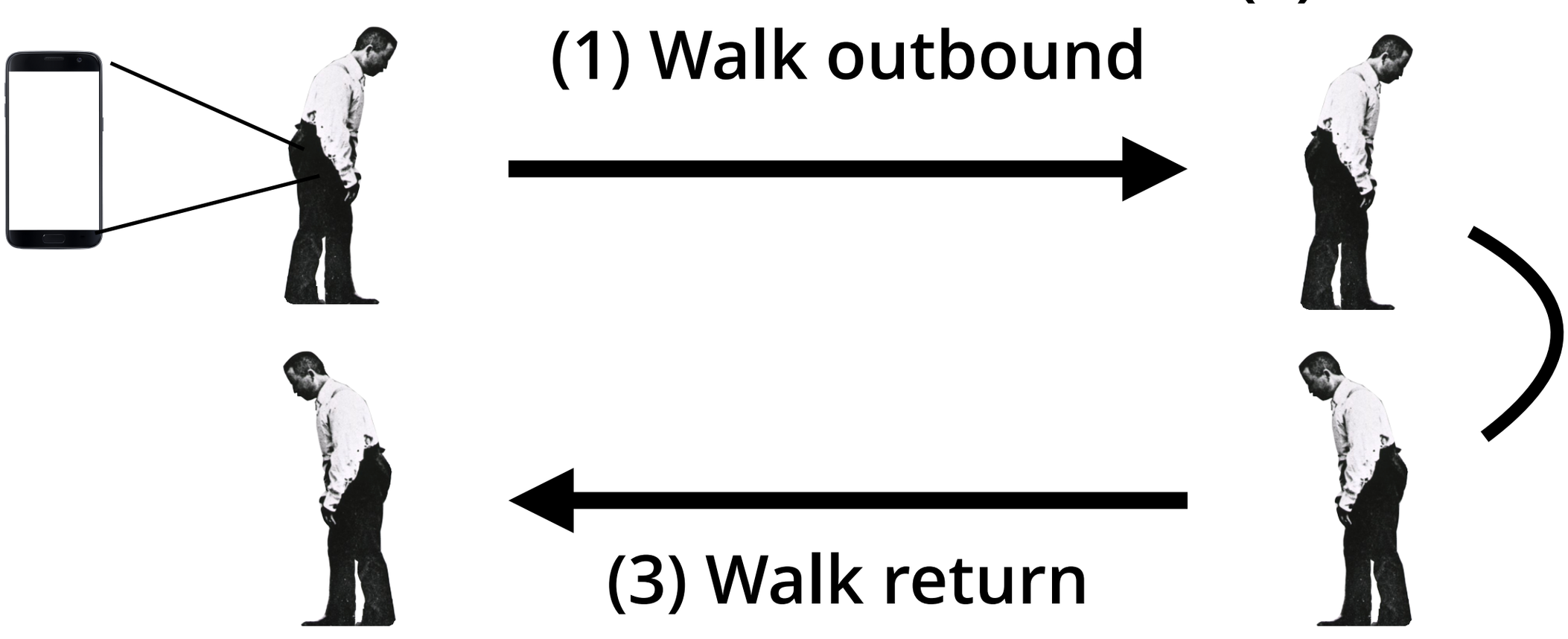
(2) Rest



(1) Walk outbound

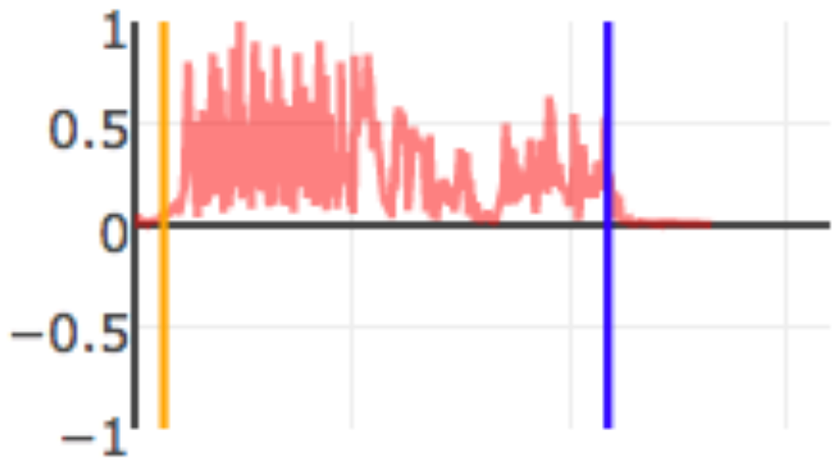
mPower Walking Test

- In the walking task, participants were asked to do the following three-segment task (each ~30s):

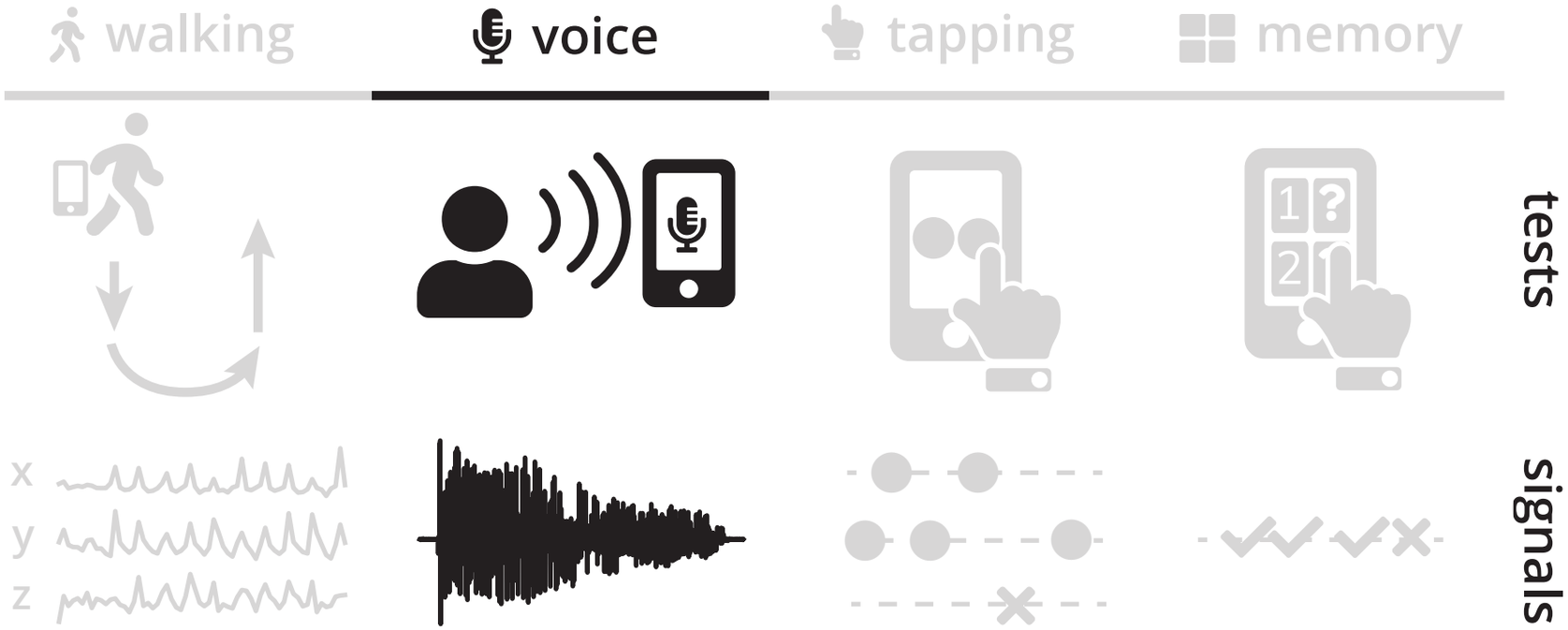


Data Streams

- Accelerometer time series:
  - Acceleration
  - Rotation Rate
  - Attitude



Tests Overview



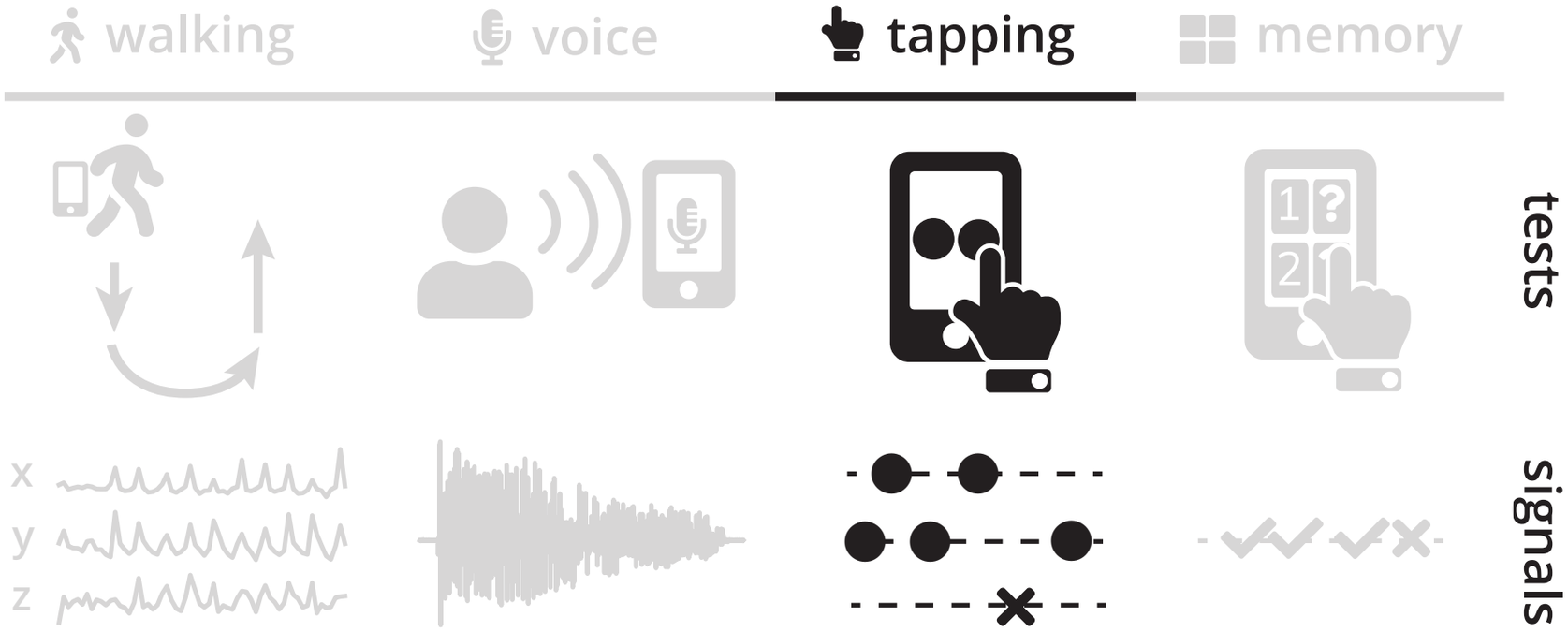
mPower Voice Test

“aaaaaah”

Data Streams

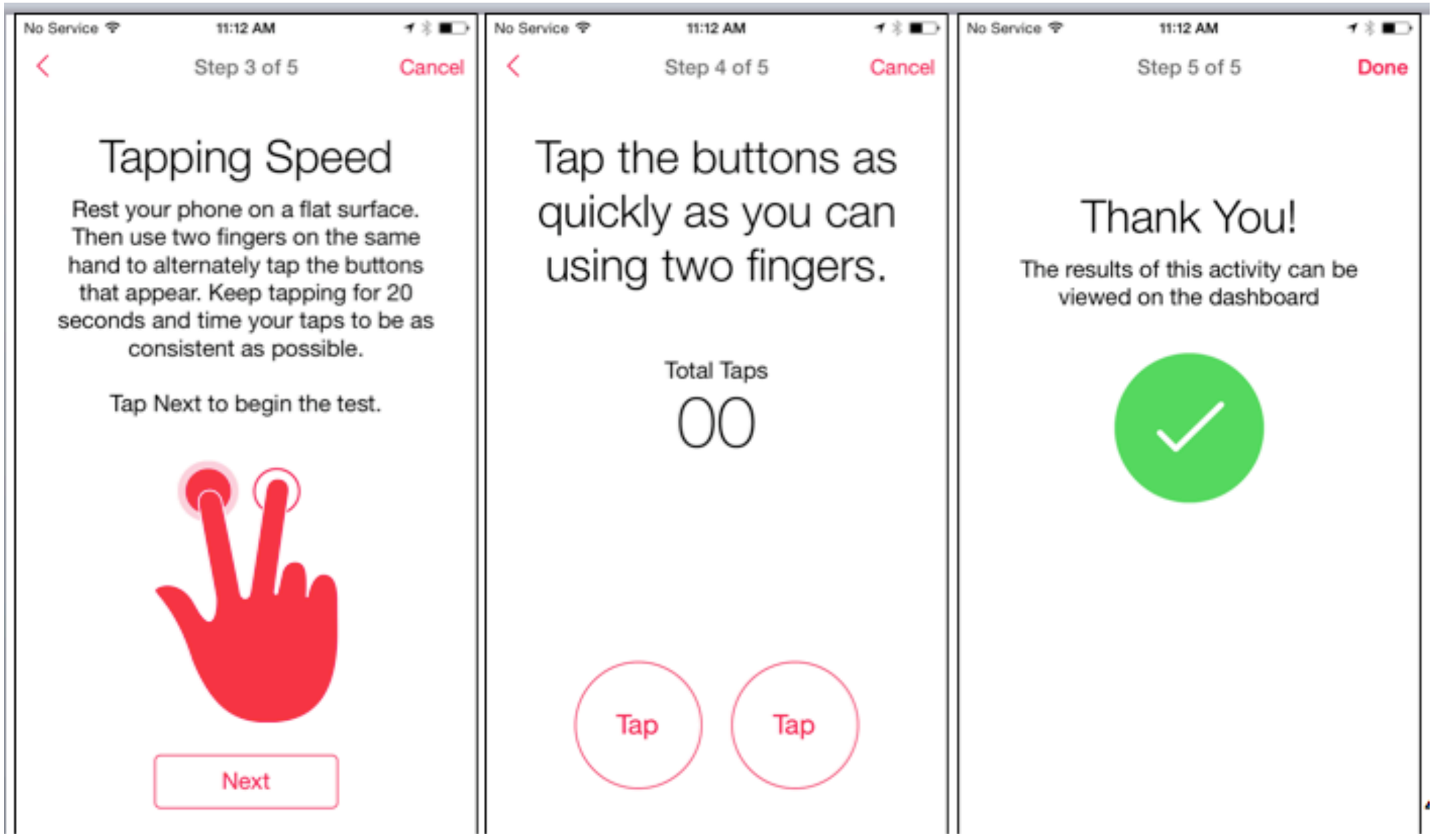
- Voice recording
  - 44100 Hz
  - ~30 seconds

Tests Overview

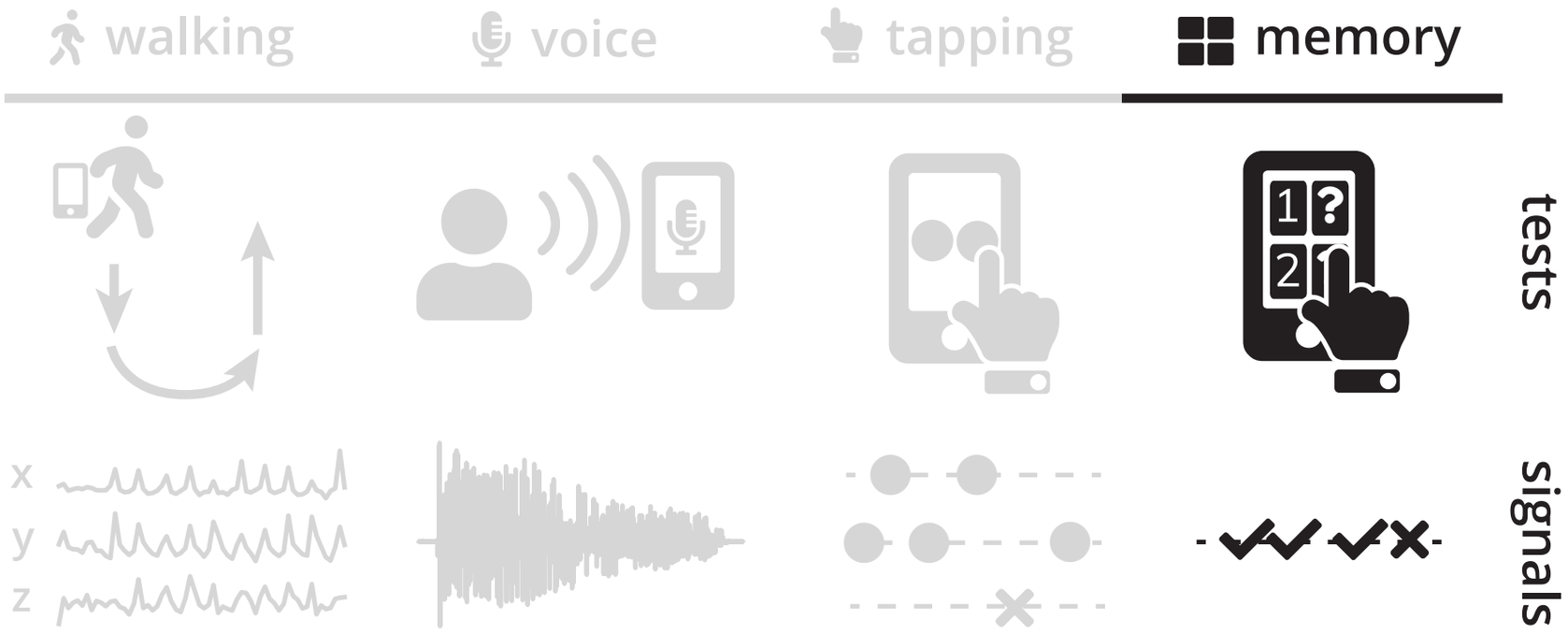


mPower Tapping Test

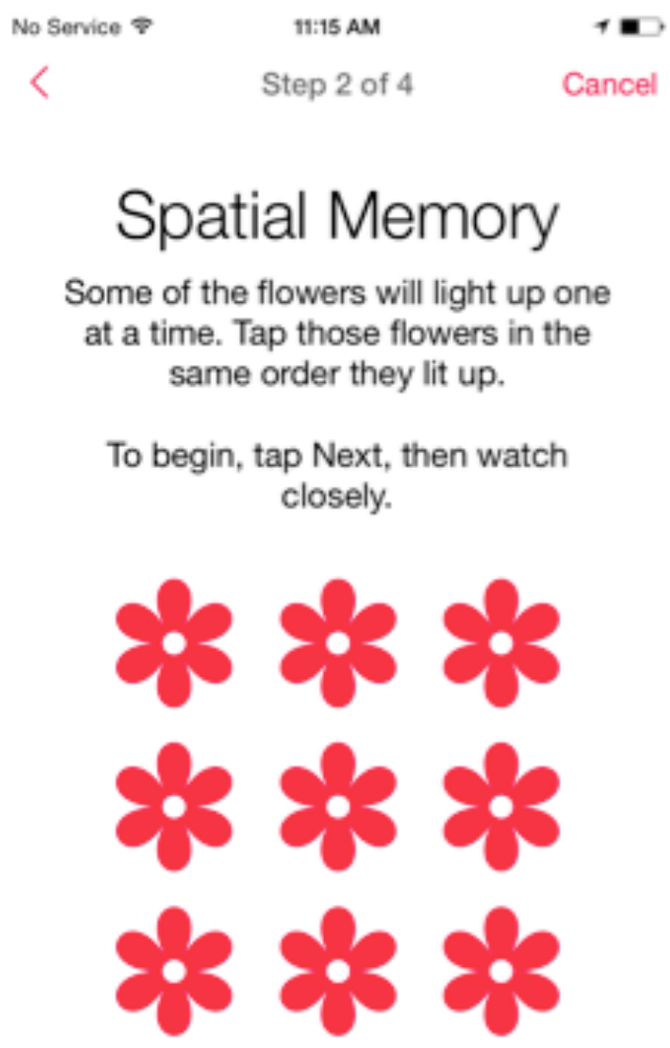
[1] Bot, B.M., et al. (2016) *The mPower study, Parkinson disease mobile data collected using ResearchKit.* Scientific data 3.



Tests Overview



Tests Overview

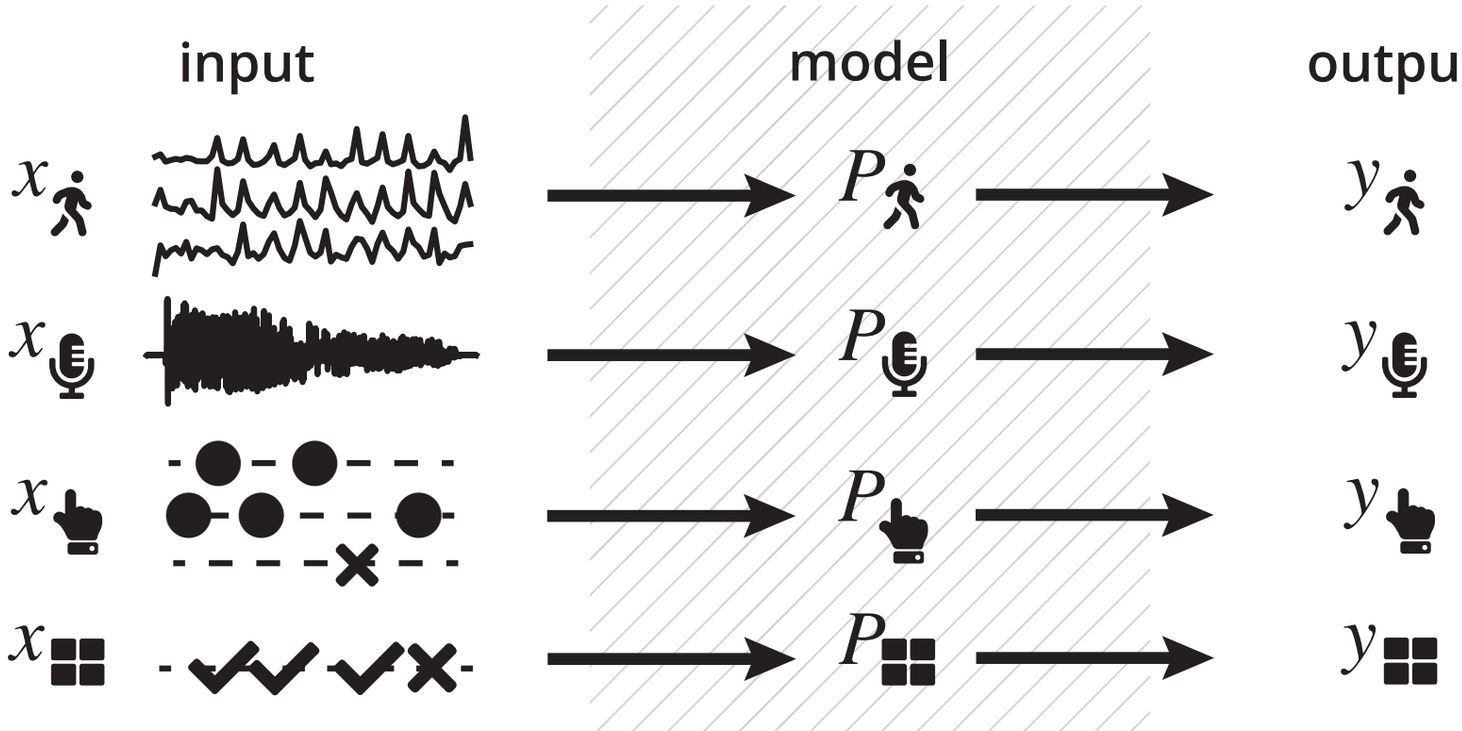


[1] Bot, B.M., et al. (2016) *The mPower study, Parkinson disease mobile data collected using ResearchKit.* Scientific data 3.

Approach

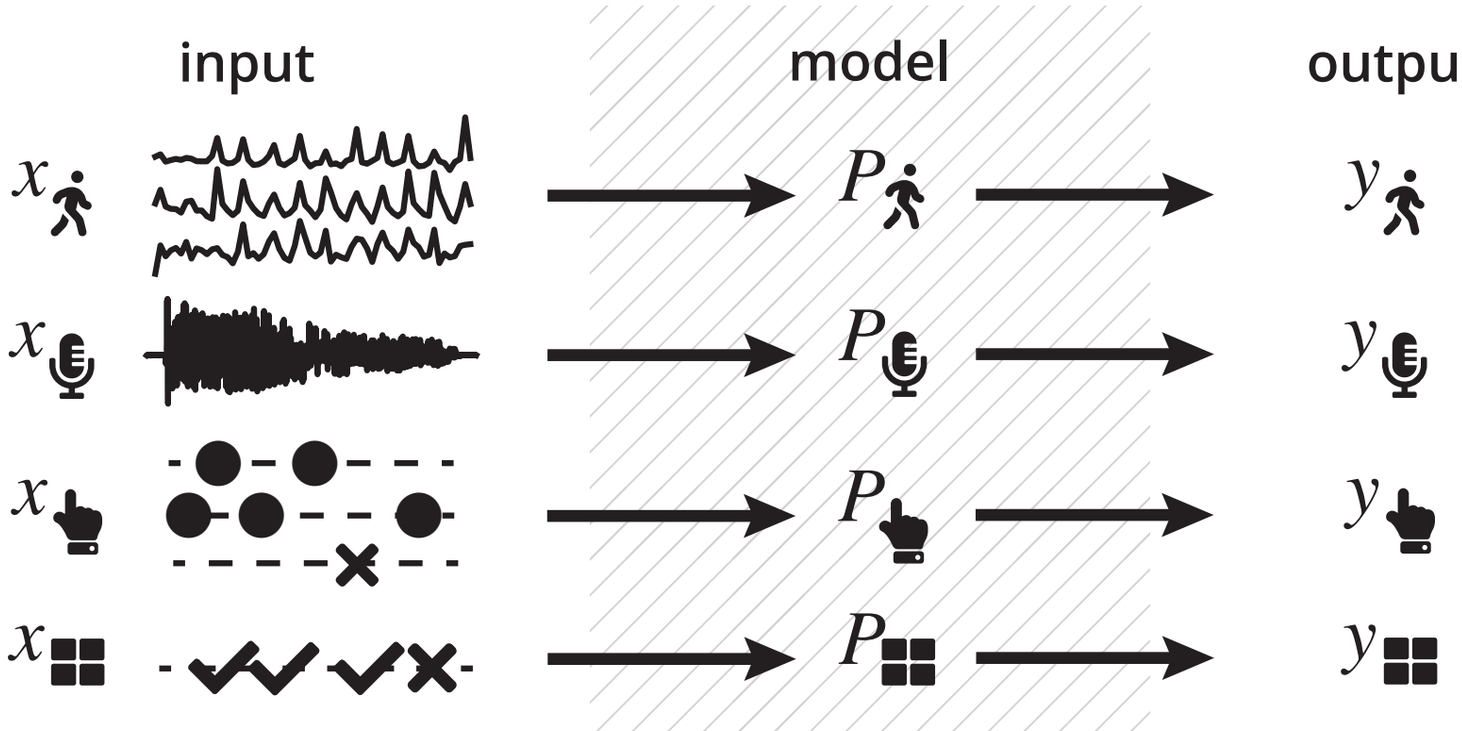
1- Hierarchical Approach

Per-test Models: Specialised in each test type.



1- Hierarchical Approach

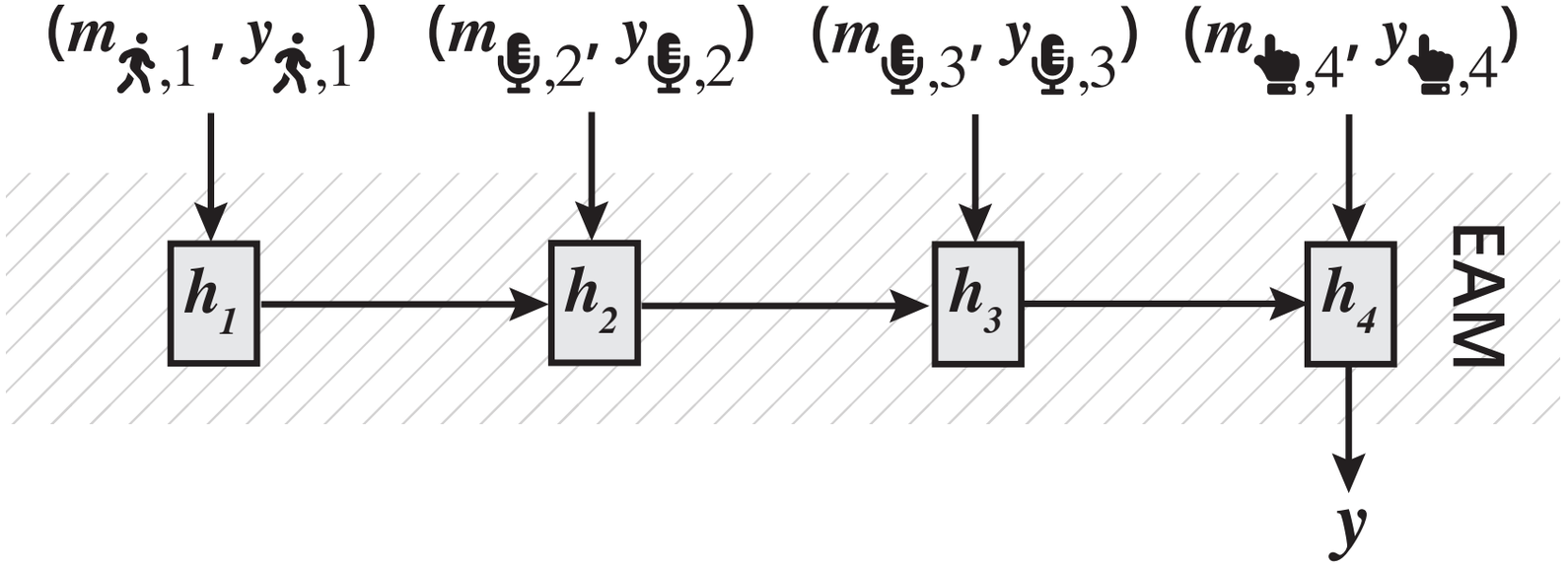
Per-test Models: Specialised in each test type.



Independent models

2- Hierarchical Approach

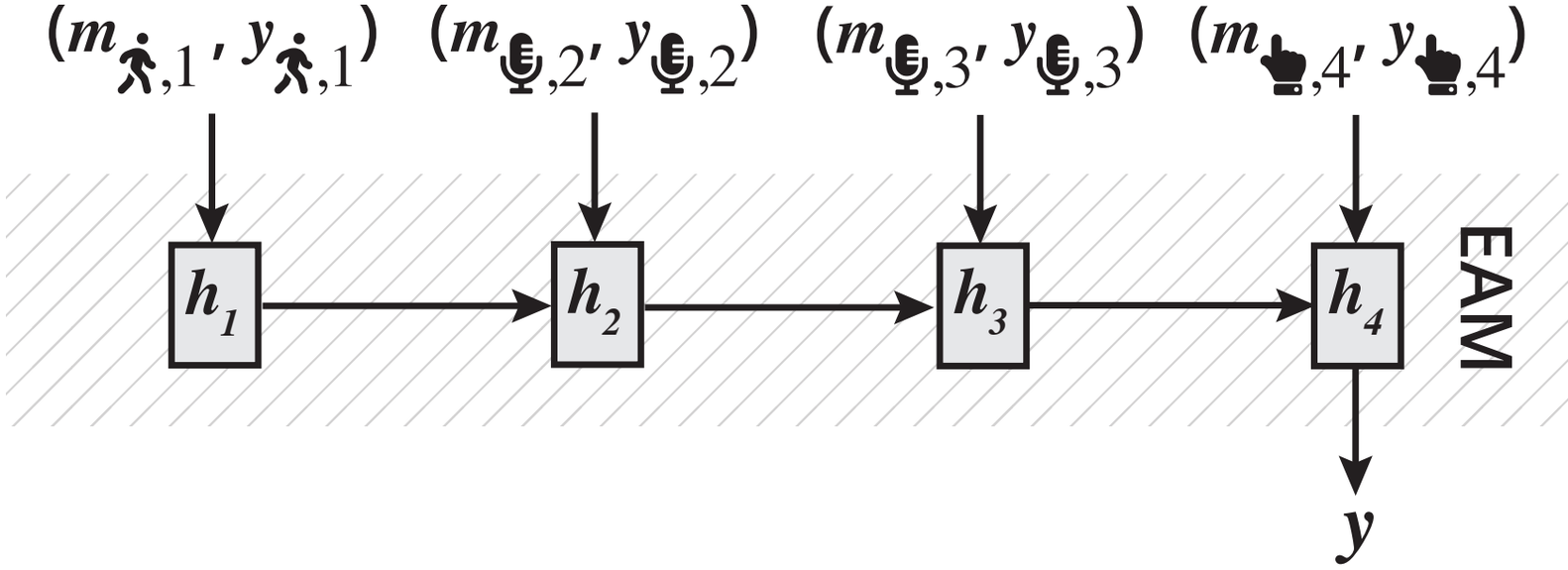
Evidence Aggregation Model (EAM):
Integrate available test data over time.



2- Hierarchical Approach

Evidence Aggregation Model (EAM):
Integrate available test data over time.

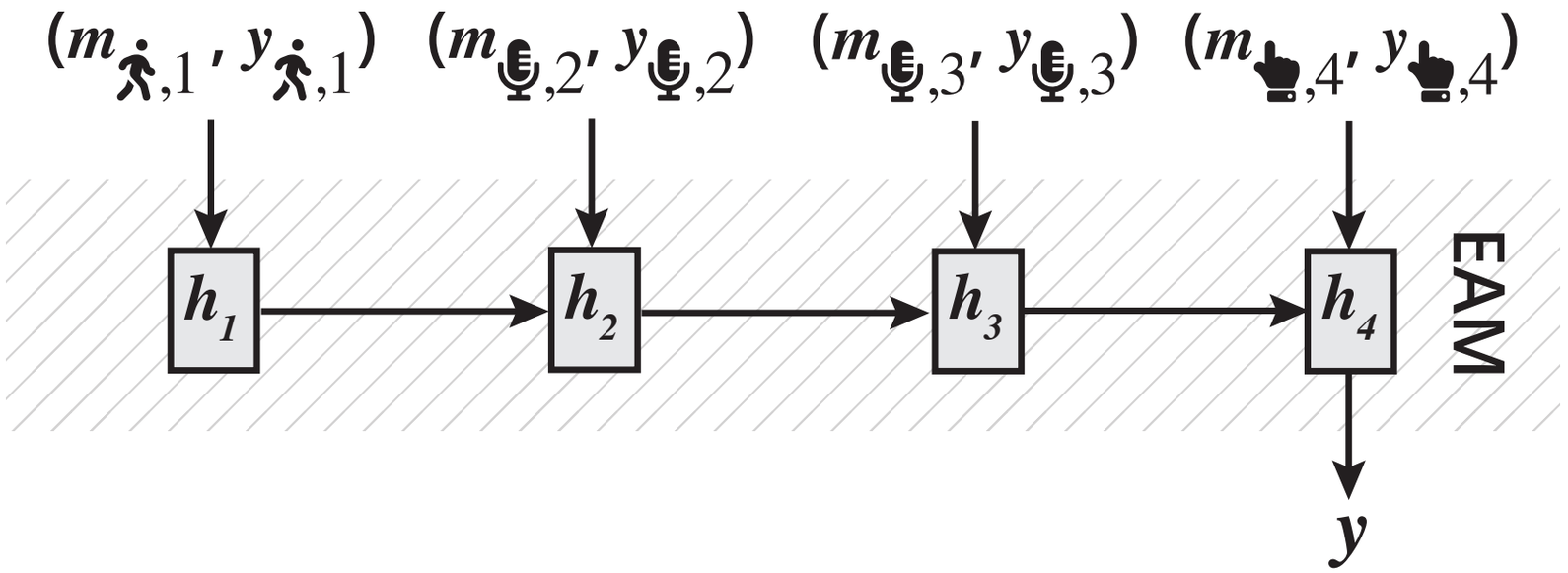
Any number of tests



2- Hierarchical Approach

Evidence Aggregation Model (EAM):
Integrate available test data over time.

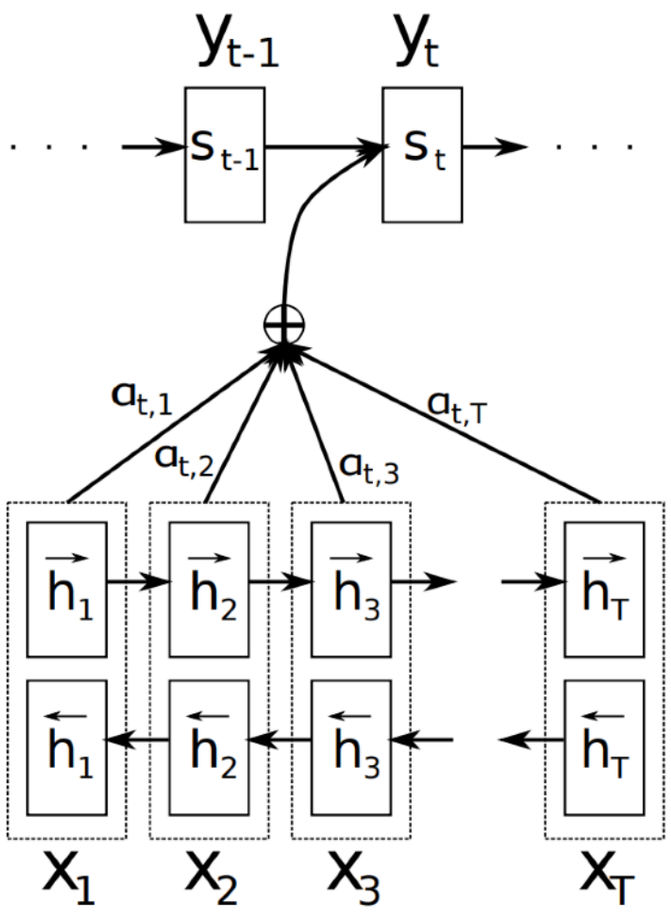
Any number of tests

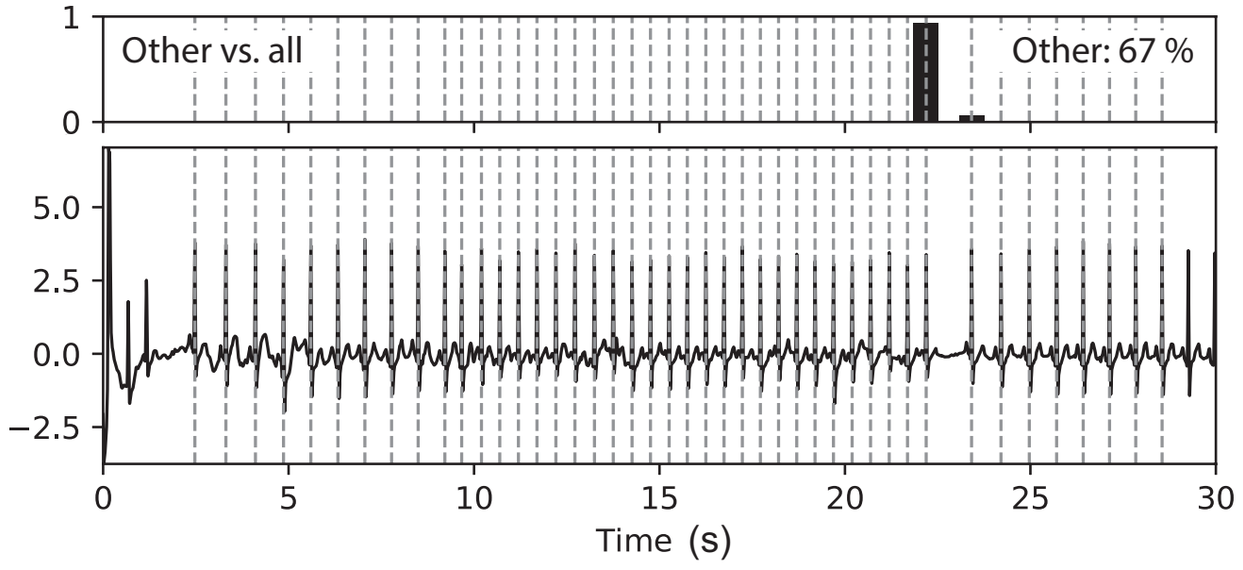


Final diagnostic score

Neural Soft Attention

- A **soft attention mechanism**[1] allows us to relate the **decisions** to the most relevant **(1) input segments**[2] and **(2) tasks.**



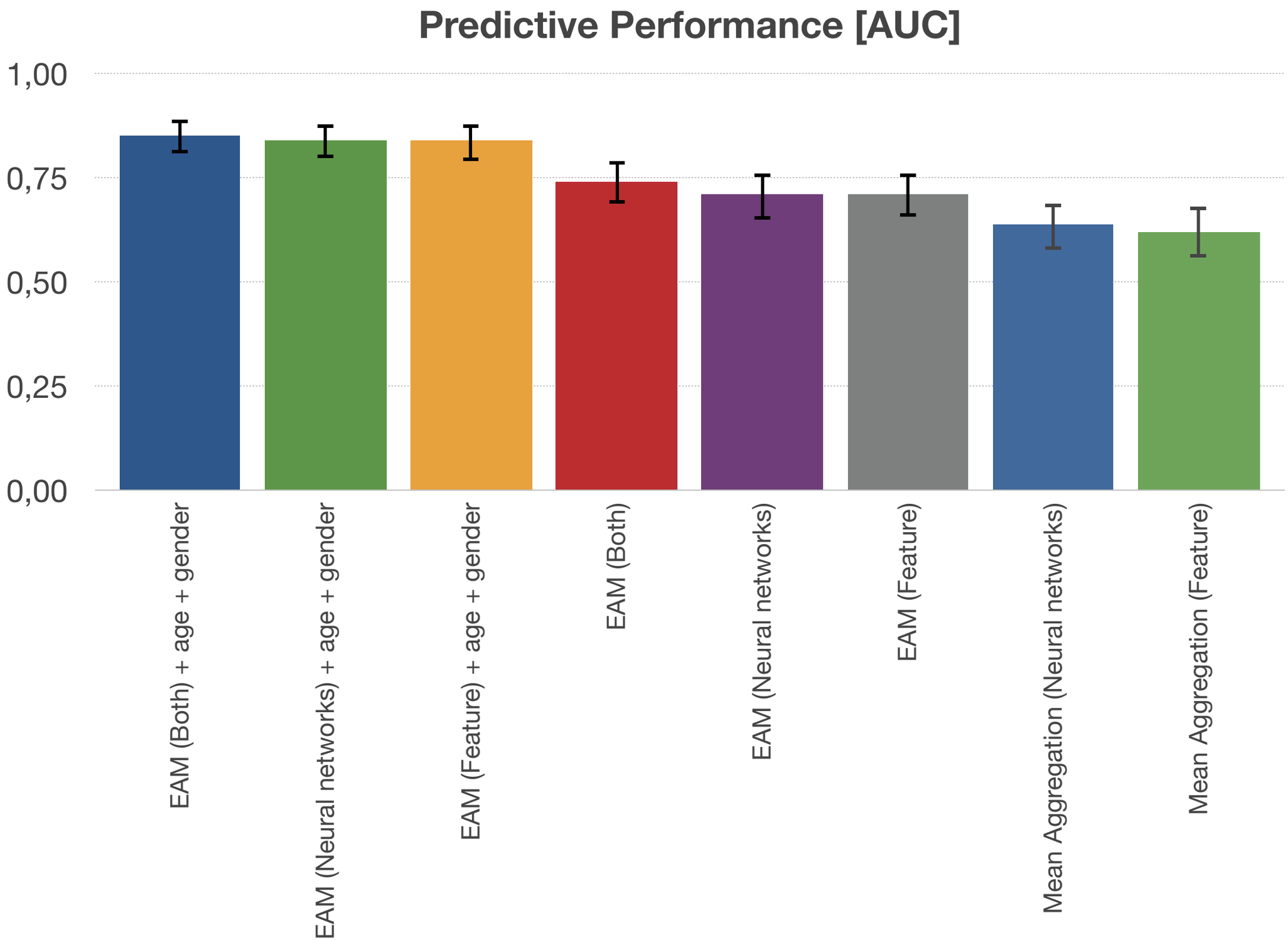


[1] Bahdanau, D. et al. (2014). *Neural Machine Translation by Jointly Learning to Align and Translate.* ICLR.

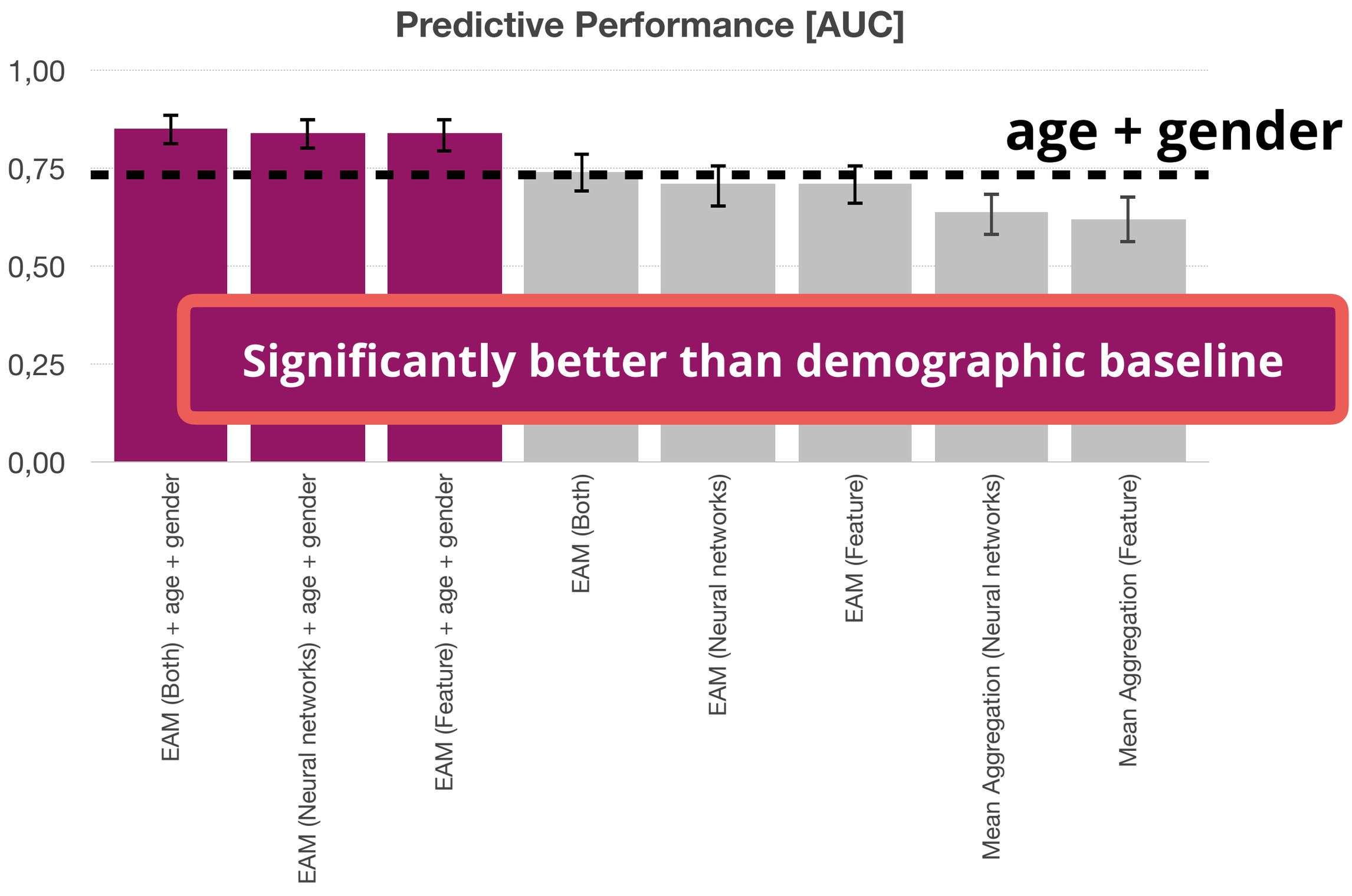
[2] Schwab, P., et al. (2017). *Beat by Beat: Classifying Cardiac Arrhythmias with Recurrent Neural Networks.* Computing in Cardiology.

Results & Discussion

Results on Test Set

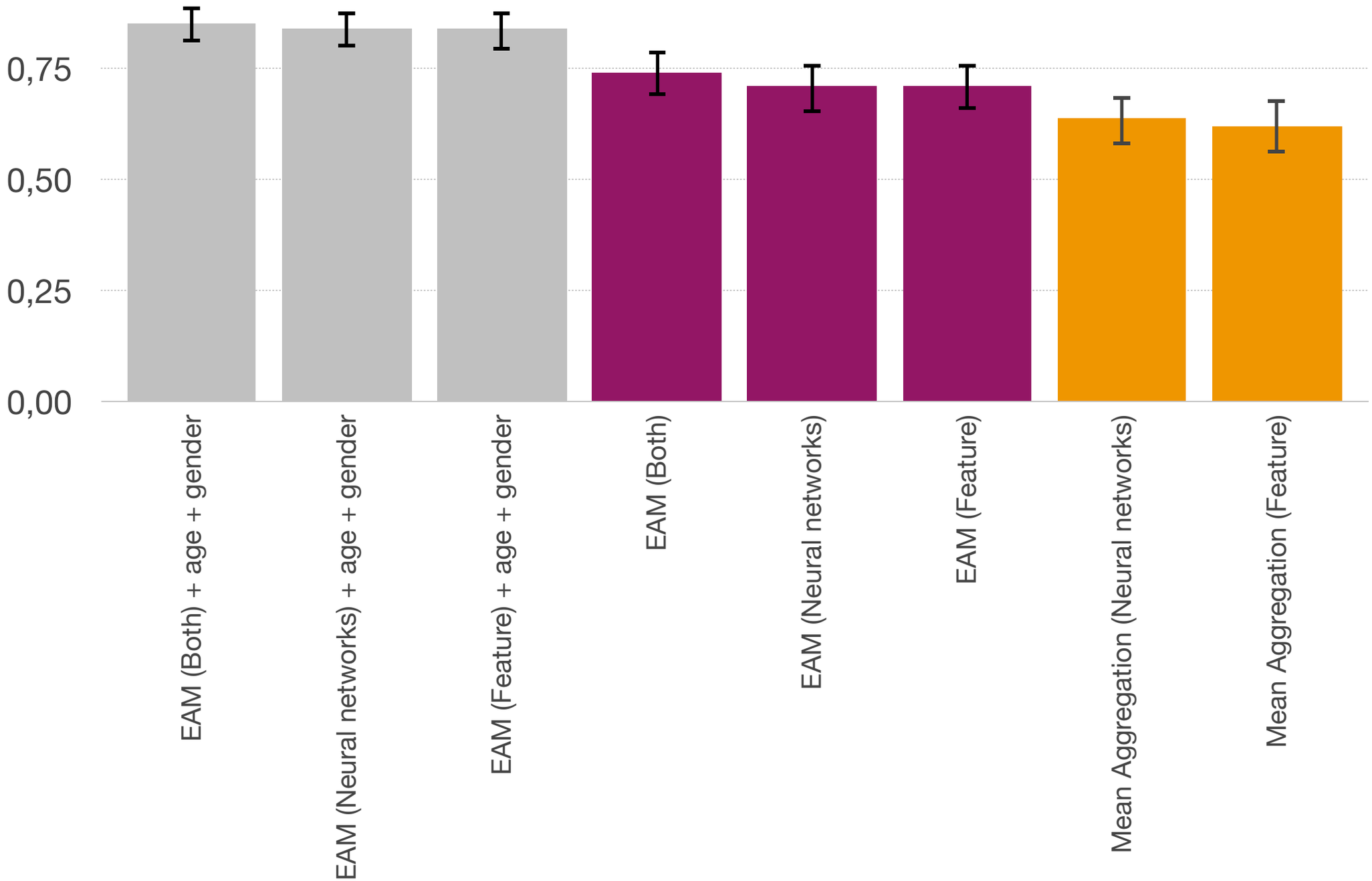


Results on Test Set

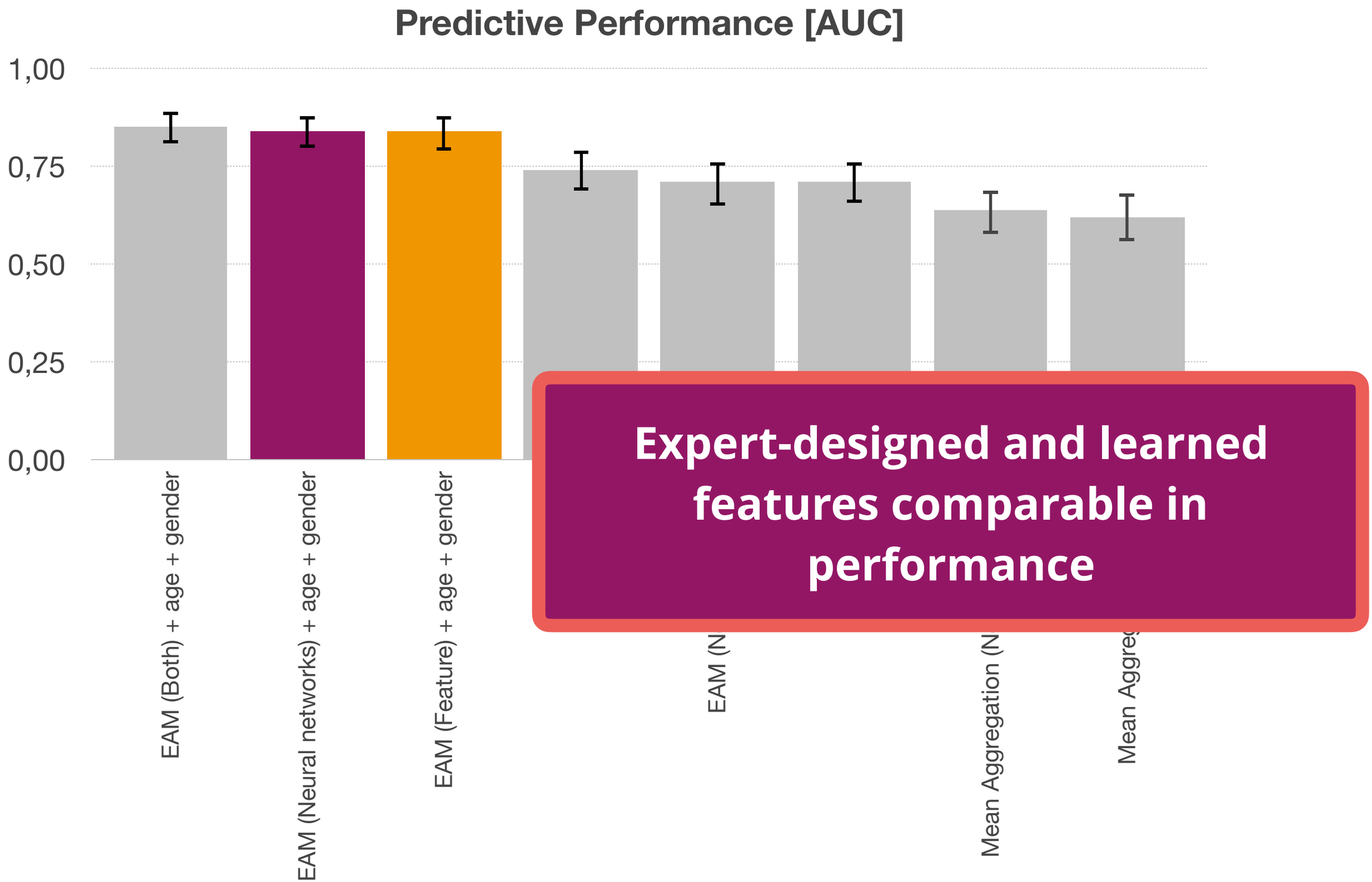


Results on Test Set

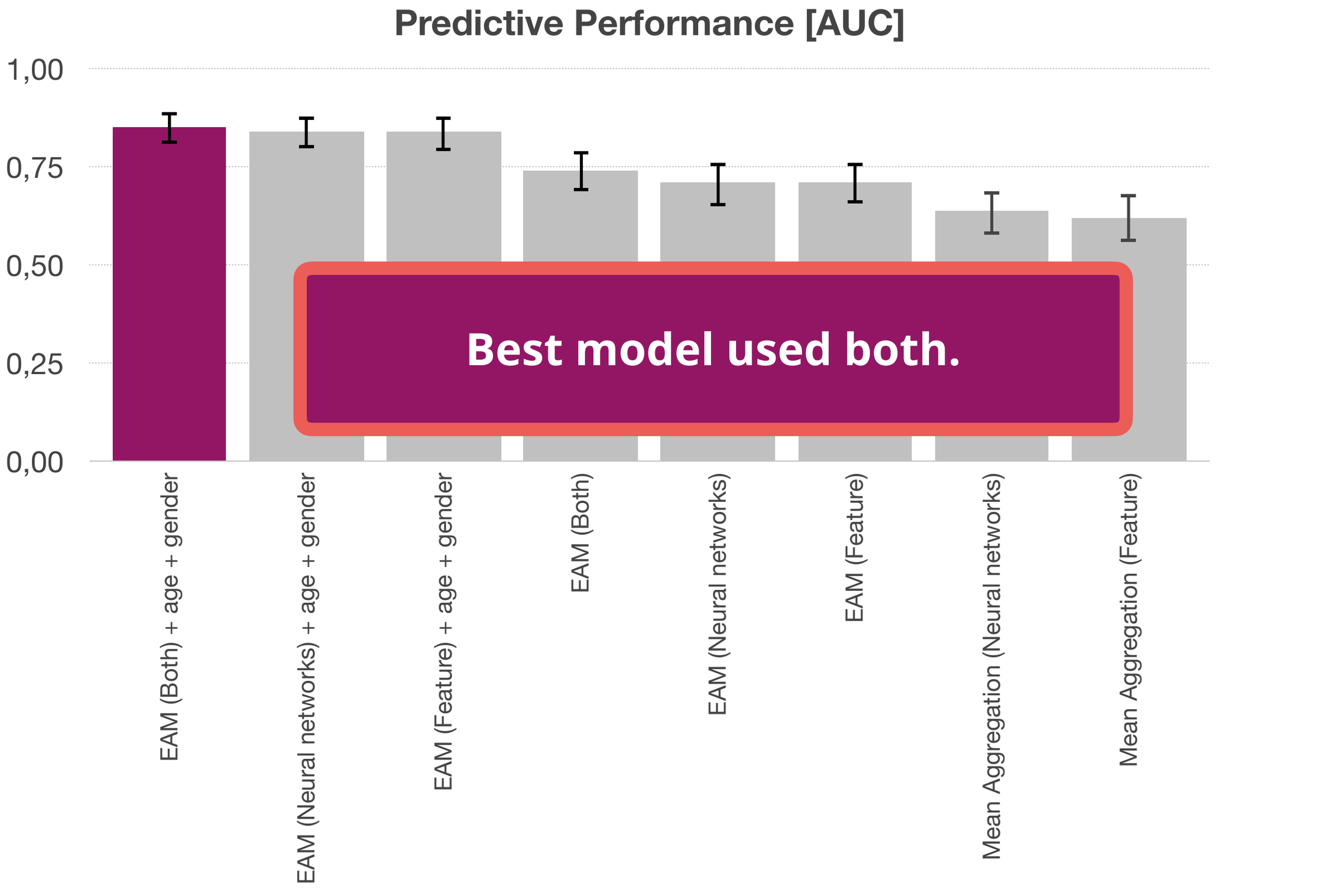
EAM better integrates the available test data



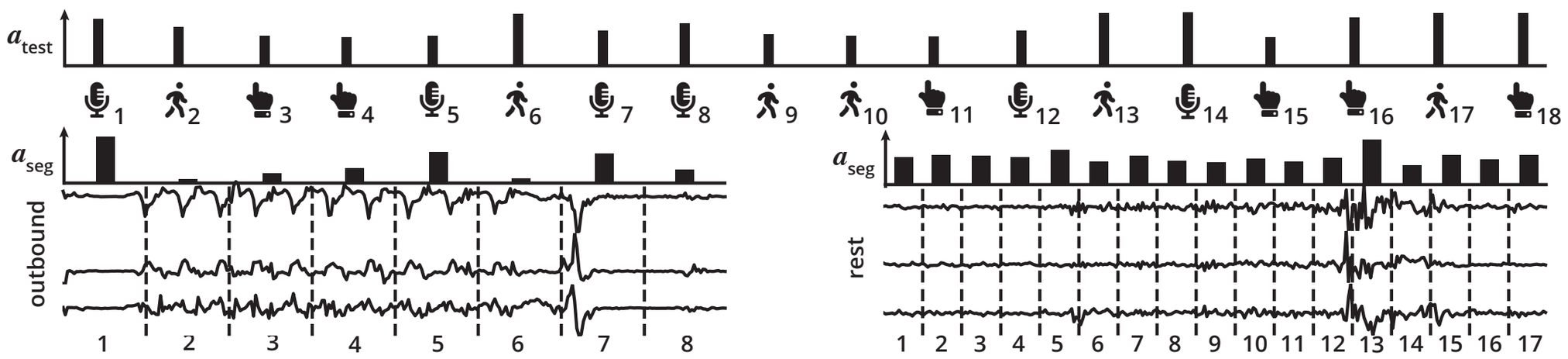
Results on Test Set



Results on Test Set

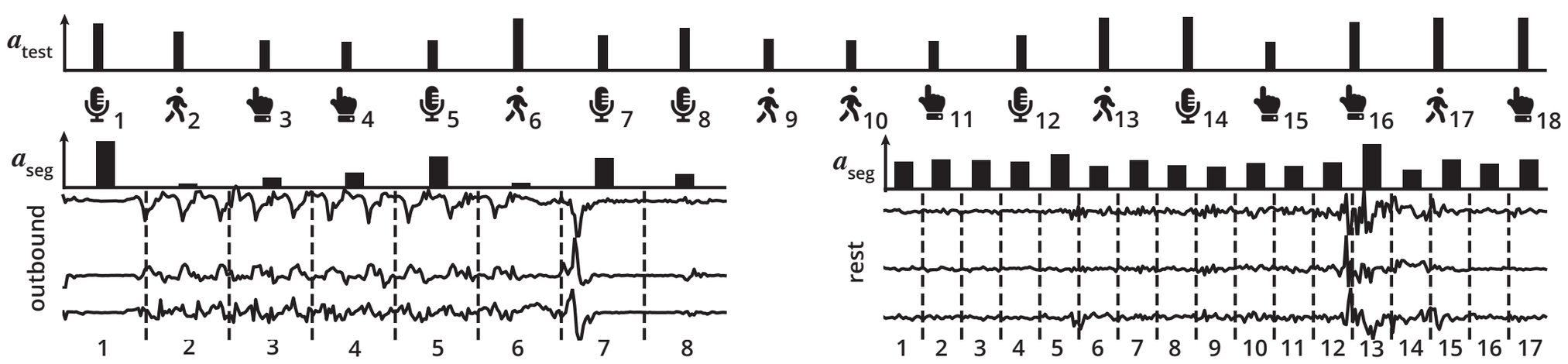


Neural Attention



Neural Attention

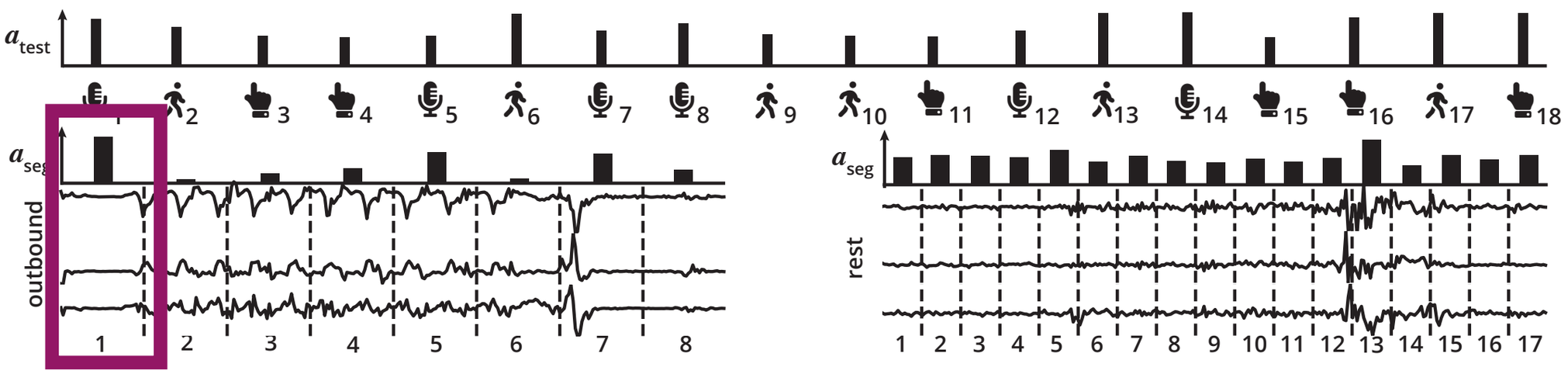
Importance over tests



Importance within test

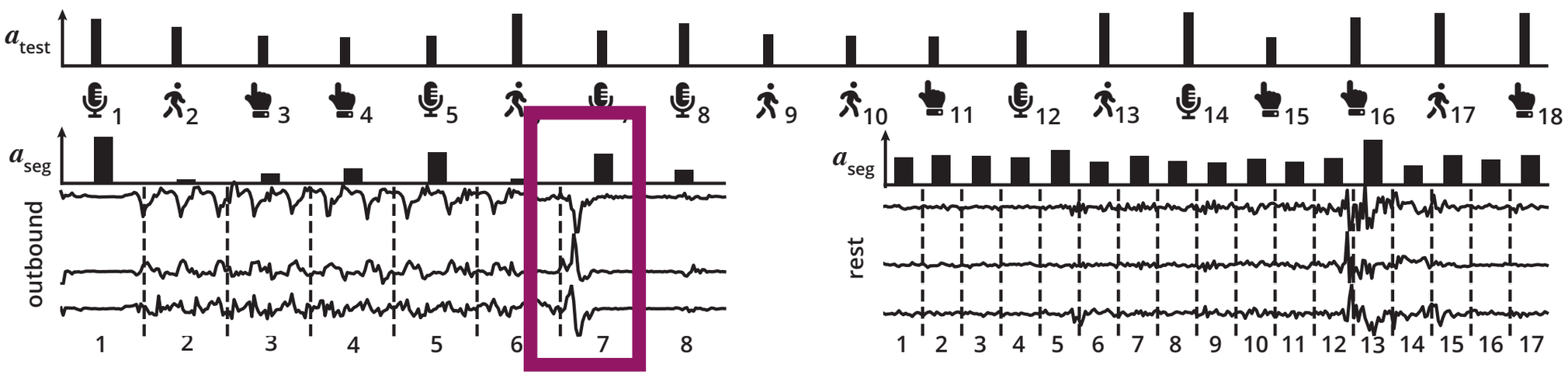
Importance within test

Neural Attention (Subject with PD)



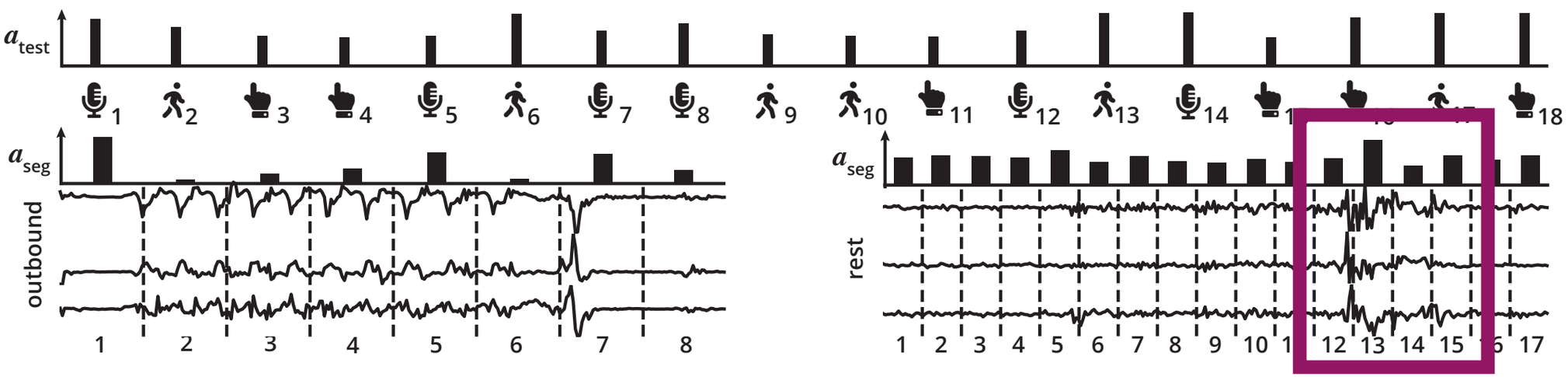
Difficulty starting to move

Neural Attention (Subject with PD)



Abrupt stop

Neural Attention (Subject with PD)



Potential resting tremor

Conclusion

Conclusion

- We present an approach to diagnosing PD that …

  ✔ **works** based on multiple smartphone-based tests that cover a **wide range of symptoms** across **long time frame**

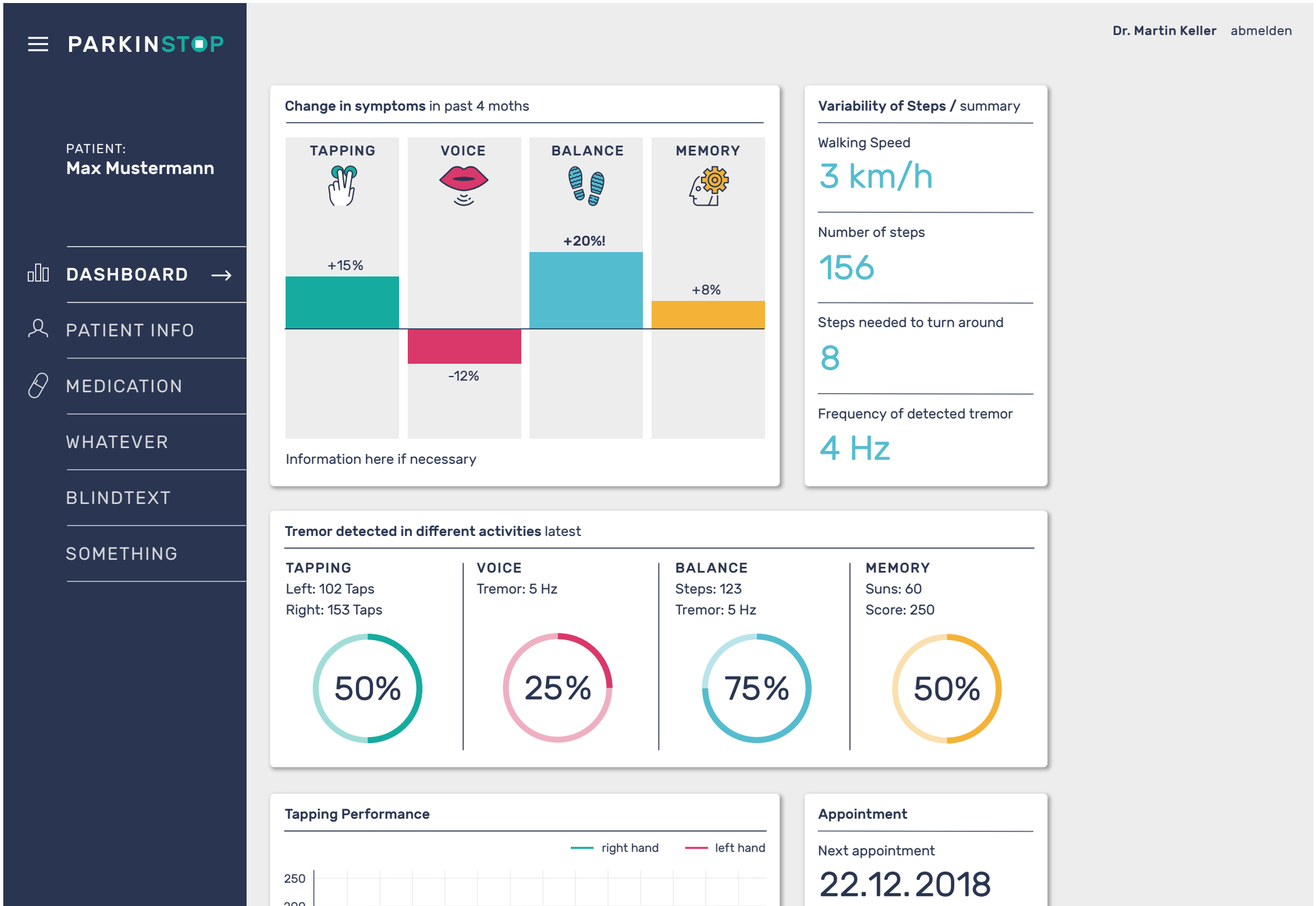
  ✔ informs the clinician about the **importance of tests and segments within those tests** using neural attention

  ✔ achieves **strong performance** in a **representative cohort (n=1853)** with an **AUC of 0.85 (95% CI: 0.81, 0.89)**

- We highlight **potential of smartphones** as accessible tools for gathering **clinically relevant data in the wild**

Future Work

[1] Matas Pocevicius (2018), Intelligent Decision-Support for Diagnosis and Monitoring of Parkinson's Disease. MSc Thesis, ETH Zurich



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

75 NRP Big Data
National Research Programme

Questions?

Patrick Schwab

@schwabpa

patrick.schwab@hest.ethz.ch
Institute for Robotics and Intelligent Systems
ETH Zurich

Schwab, Patrick and Karlen, Walter.
PhoneMD: Learning to Diagnose Parkinson's Disease with Smartphone Data.
AAAI 2019

Dataset Overview

Property	Training	Validation	Test
Subjects (#)	1314 (70%)	192 (10%)	347 (20%)
PD (%)	52.36	50.00	56.20
Female (%)	28.00	36.98	25.94
Age (years)	59.29 ± 9.40	59.53 ± 9.03	58.90 ± 9.24
Walking (#)	13.89 ± 35.07	15.58 ± 33.90	14.03 ± 45.20
Voice (#)	16.11 ± 40.21	19.47 ± 44.55	14.88 ± 45.12
Tapping (#)	15.20 ± 38.04	18.50 ± 43.12	14.78 ± 42.67
Memory (#)	14.01 ± 33.30	20.78 ± 35.92	17.58 ± 38.11
Usage (days)	24.27 ± 41.01	29.66 ± 45.73	25.43 ± 43.24

Table 3: Population statistics of the training, validation, and test set. Numbers (#) shown are mean ± standard deviation.

Dataset Overview

Largest cohort to date

Property	Training	Validation	Test
Subjects (#)	1314 (70%)	192 (10%)	347 (20%)
PD (%)	52.36	50.00	56.20
Female (%)	28.00	36.98	25.94
Age (years)	59.29 ± 9.40	59.53 ± 9.03	58.90 ± 9.24
Walking (#)	13.89 ± 35.07	15.58 ± 33.90	14.03 ± 45.20
Voice (#)	16.11 ± 40.21	19.47 ± 44.55	14.88 ± 45.12
Tapping (#)	15.20 ± 38.04	18.50 ± 43.12	14.78 ± 42.67
Memory (#)	14.01 ± 33.30	20.78 ± 35.92	17.58 ± 38.11
Usage (days)	24.27 ± 41.01	29.66 ± 45.73	25.43 ± 43.24

Table 3: Population statistics of the training, validation, and test set. Numbers (#) shown are mean ± standard deviation.

Dataset Overview

Nearly balanced

Property	Training	Validation	Test
Subjects (#)	1314 (70%)	192 (10%)	347 (20%)
PD (%)	52.36	50.00	56.20
Female (%)	28.00	36.98	25.94
Age (years)	59.29 ± 9.40	59.53 ± 9.03	58.90 ± 9.24
Walking (#)	13.89 ± 35.07	15.58 ± 33.90	14.03 ± 45.20
Voice (#)	16.11 ± 40.21	19.47 ± 44.55	14.88 ± 45.12
Tapping (#)	15.20 ± 38.04	18.50 ± 43.12	14.78 ± 42.67
Memory (#)	14.01 ± 33.30	20.78 ± 35.92	17.58 ± 38.11
Usage (days)	24.27 ± 41.01	29.66 ± 45.73	25.43 ± 43.24

Table 3: Population statistics of the training, validation, and test set. Numbers (#) shown are mean ± standard deviation.

Dataset Overview

Property	Training	Validation	Test
Subjects (#)	1314 (70%)	192 (10%)	347 (20%)
PD (%)	52.36	50.00	56.20
Female (%)	28.00	36.98	25.94
Age (years)	59.29 ± 9.40	59.53 ± 9.03	58.90 ± 9.24
Walking (#)	13.89 ± 35.07	15.58 ± 33.90	14.03 ± 45.20
Voice (#)	16.11 ± 40.21	19.47 ± 44.55	14.88 ± 45.12
Tapping (#)	15.20 ± 38.04	18.50 ± 43.12	14.78 ± 42.67
Memory (#)	14.01 ± 33.30	20.78 ± 35.92	17.58 ± 38.11
Usage (days)	24.27 ± 41.01	29.66 ± 45.73	25.43 ± 43.24

Table 3: Population stat
test set. Numbers (#) sh

Wide range of usage patterns