Forecasting Parkinson's Disease Patients' Wearing-Off using Wrist-Worn Fitness Tracker and Smartphone Dataset

John Noel VICTORINO, Yuko SHIBATA, Inoue SOZO, Tomohiro SHIBATA

Kyushu Institute of Technology

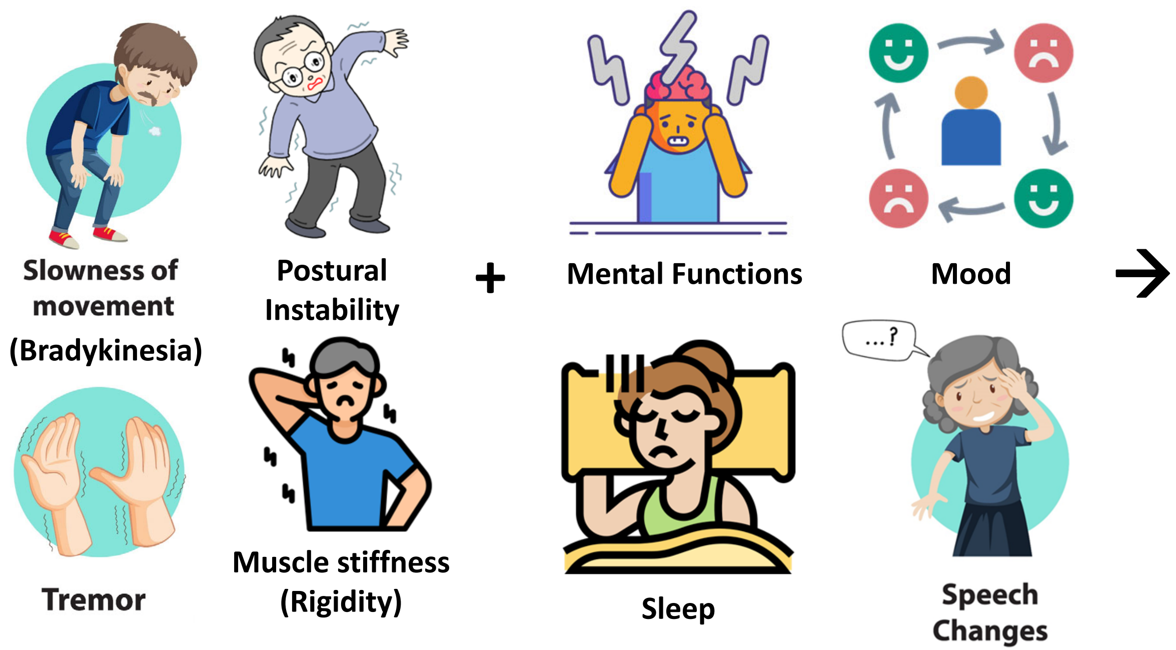
4th International Conference on Activity and Behavior Computing (ABC 2022)

University of East London, London

Parkinson's Disease (PD)

PD is progressive neurodegenerative movement disorder due to loss of dopamine-producing cells in the brain.

Motor Symptoms **Non-Motor Symptoms**



Symptoms and its severity vary for each patient

Affects daily life & Quality of Life (QoL)

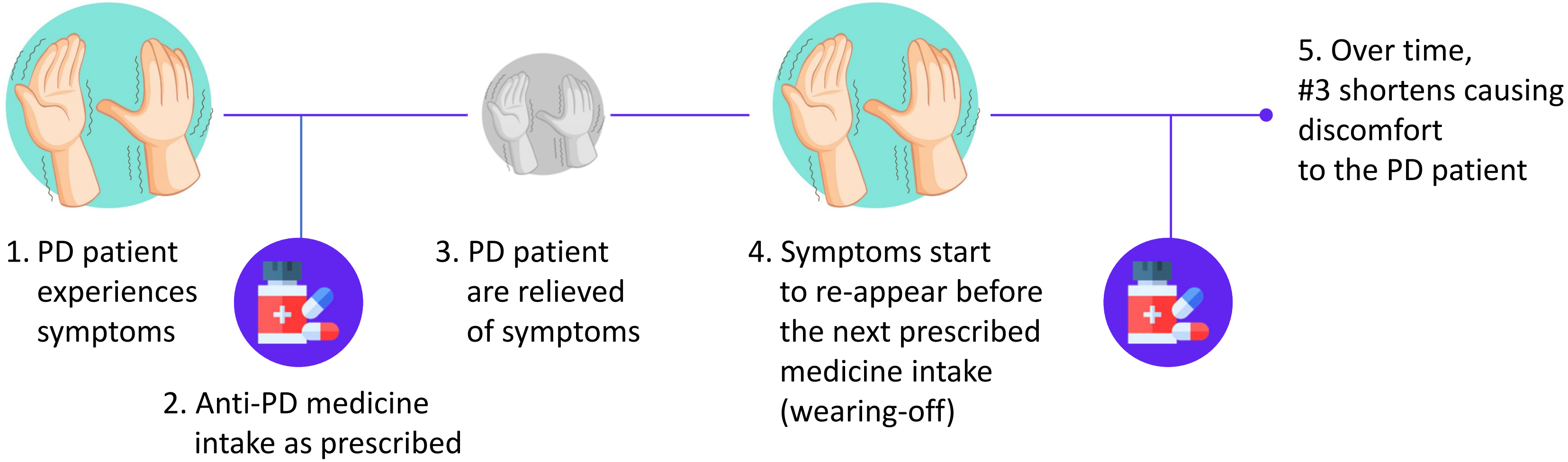
Collaborate with clinicians for customized treatment

J. Massano and K. P. Bhatia, 'Clinical Approach to Parkinson's Disease: Features, Diagnosis, and Principles of Management', *Cold Spring Harb Perspect Med*, vol. 2, no. 6, Jun. 2012, doi: 10.1101/cshperspect.a008870.

S. Sveinbjornsdottir, 'The clinical symptoms of Parkinson's disease', *Journal of Neurochemistry*, vol. 139, no. S1, pp. 318–324, 2016, doi: [10.1111/jnc.13691](10.1111/jnc.13691).

Wearing-Off Phenomenon (WO)

Patients undergo Levodopa treatment (L-dopa) to alleviate the symptoms.



Wearing-off needs to be monitored & reported by PD patients to re-adjust the treatment plan.

A. Antonini *et al.*, ‘Wearing-off scales in Parkinson’s disease: Critique and recommendations: Scales to Assess Wearing-Off in PD’, *Mov. Disord.*, vol. 26, no. 12, pp. 2169–2175, Oct. 2011, doi: 10.1002/mds.23875.

D. Colombo *et al.*, ‘The “Gender Factor” in Wearing-Off among Patients with Parkinson’s Disease: A Post Hoc Analysis of DEEP Study’, *The Scientific World Journal*, Jan. 20, 2015. https://www.hindawi.com/journals/tswj/2015/787451/

Cleveland Clinic, ‘Parkinson’s disease: Causes, Symptoms, Stages, Treatment, Support’, *Cleveland Clinic*, May 01, 2020. https://my.clevelandclinic.org/health/diseases/8525-parkinsons-disease-an-overview

Executive Summary

Can wrist-worn fitness tracker datasets be used to forecast wearing-off in the next hour?

YES, forecasting wearing-off using commercial wrist-worn fitness tracker datasets was **feasible.**

- With current time data
- With 1-day's worth of data

Which among the DL architectures performed well in forecasting wearing-off in the next hour?

Multi time-step Dense	CNN	Single time-step Dense
Highest balanced accuracy: 80.64% ± 10.36%		Highest AUC: 69.14% ± 10.60%

Forecasting wearing-off was feasible within a certain limit, even with only commercial fitness tracker features.

Forecasting models can be used in PD management

Current Landscape in Detecting PD symptoms or wearing-off

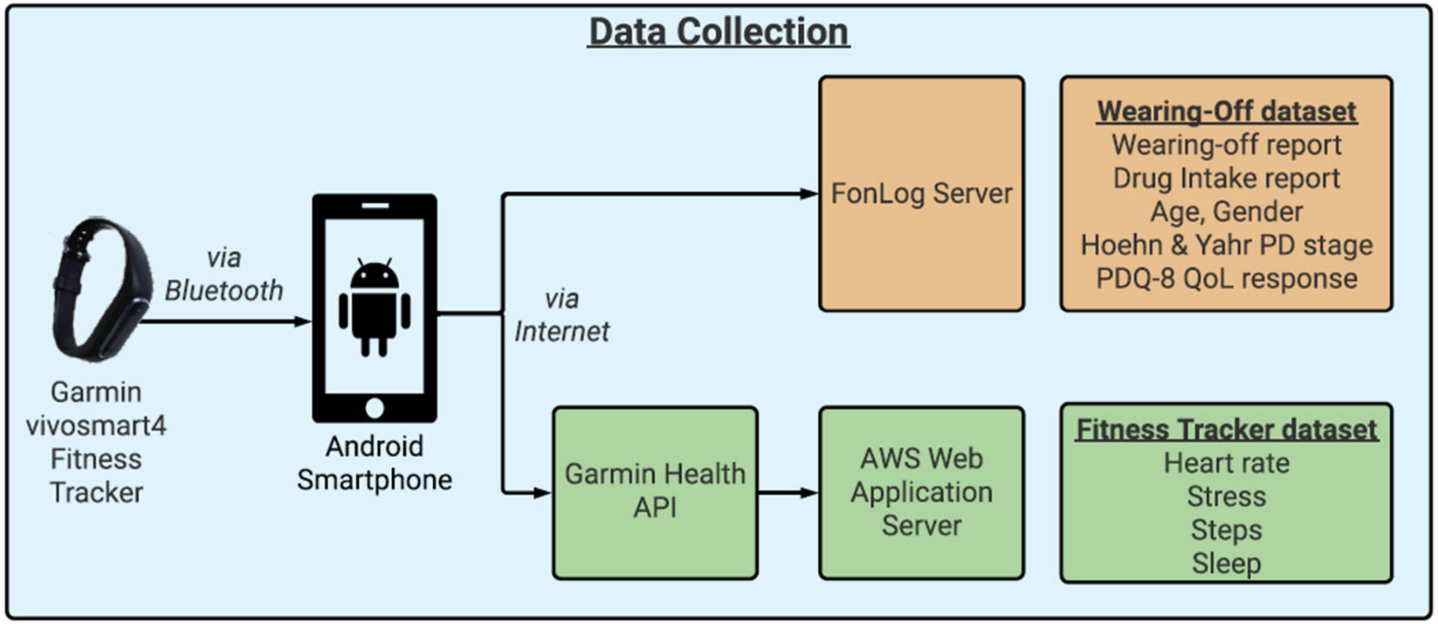
In-Clinic Data Collection

	Subjective Assessment	Motor Aspect	Non-Motor Aspect
In-Clinic Data Collection	UPDRS MDS-UPDRS	Keijsers, 2006 Jeon, 2017 Sama, 2017 Aich, 2018 & 2020 Steinmetzer, 2019 Hssayeni, 2019	Pursiainen, 2007 Salsone, 2016 Lee, 2018 van der Heide, 2021 & other clinical research
Continuous Home Data Collection	Parkinson's Disease Diary WoQ	Griffiths, 2012, Farzanehfar, 2018 & other clinical studies that use Parkinson's Kinetigraph (PKG)	Victorino, 2021: Detection of Wearing-off ★ **Current Study Forecasting Wearing-off**

Subjective Assessment ⟷ **Objective Assessment**

Continuous Home Data Collection

Data Collection Process



- PD patients who are aware of wearing-off
- PD participants were asked to contribute 7 days' worth of data
  - Wear fitness tracker
  - Report wearing-off using smartphone app

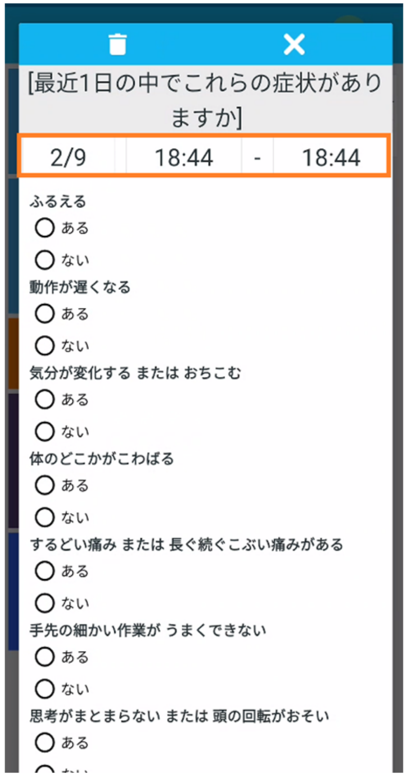
Garmin vivosmart4 Datasets

Data Type	Granularity	Description
Heart rate	15-second interval	Beats per minute (bpm)
Steps	15-minute interval	Cumulative count per interval (minimum: 0)
Stress score	3-minute interval	Estimated stress score (0 – 100) • 0 – 25: Resting state • 26 – 50: Low stress • 51 – 75: Medium stress • 76 – 100: High stress • -1: not enough data to detect stress • -2: too much motion
Sleep classification & Sleep period	Per calendar date	Start and end time per sleep classification • Light sleep • Rapid eye movement (REM) sleep • Deep sleep

Garmin, 'vívosmart 4 - Heart Rate Variability and Stress Level', *Heart Rate Variability and Stress Level*. Oct. 2020. Accessed: Apr. 06, 2021. [Online]. Available: https://www8.garmin.com/manuals/webhelp/vivosmart4/EN-US/GUID-9282196F-D969-404D-B678-F48A13D8D0CB.html

Smartphone Application Dataset

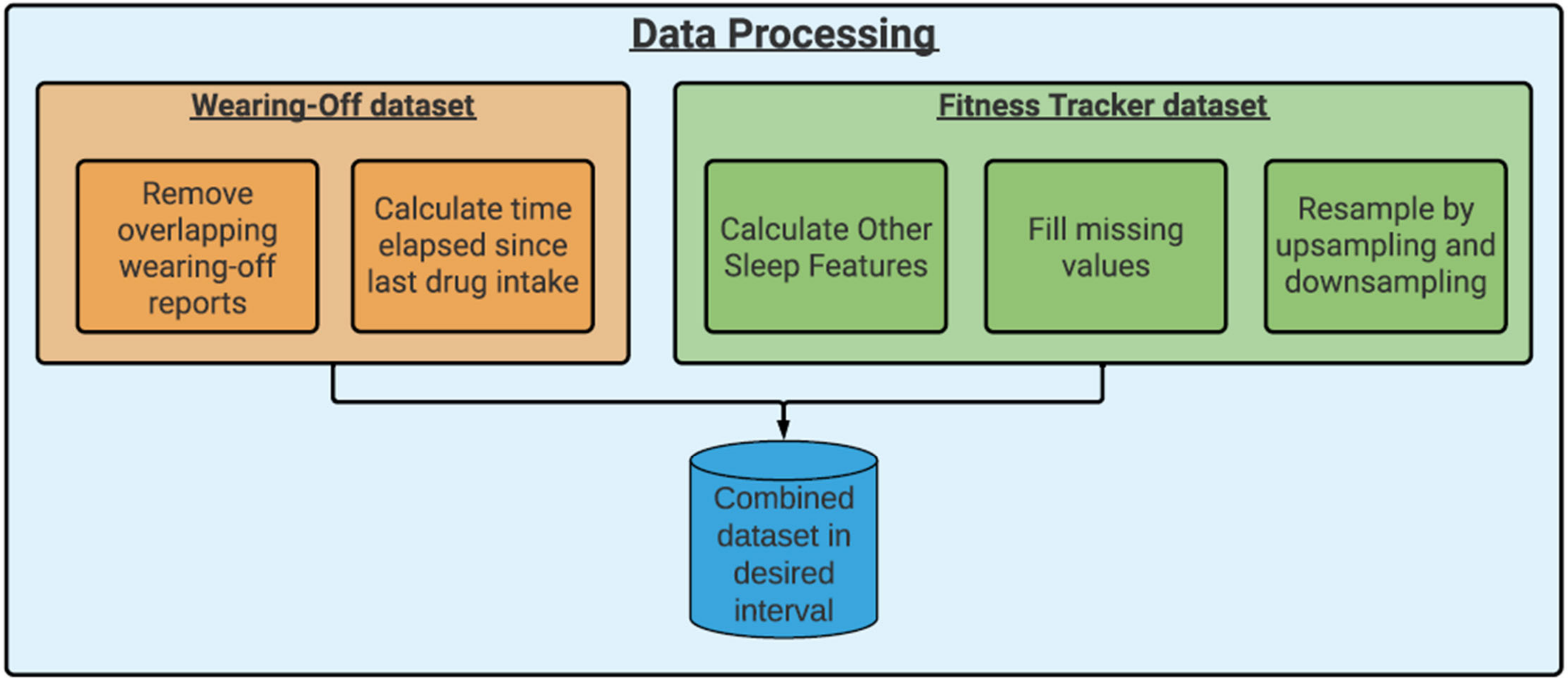
- Data collection tool to record experienced symptoms (using Japanese Wearing-Off Questionnaire or WoQ-9)
  - Tremors
  - Slowing down of movement
  - Change in mood / Depression
  - Rigidity of muscles
  - Sharp pain / Prolonged dull pain
  - Impairment of complex movement of hands & fingers
  - Difficulty integrating thoughts / slowing down of thought
  - Anxiety / Panic attacks
  - Muscle spasm
- Drug intake and its effects were also part of WoQ-9



FonLog Smartphone Application Dataset

Data	Description
WoQ-9	Symptoms onset and drug intake time
Basic Information	Age, Gender
Hoehn and Yahr Scale (H&Y), Japan Ministry of Health, Labor, and Welfare's classification of living dysfunction (JCLD)	Participant's PD stage
Parkinson's Disease Questionnaire (PDQ-8)	Participant's QoL measurement specific to PD 0 – 100%, with 100% showing worst QoL

Data Processing



Model Development

- Developed personalized models
- Data split
  - Training Set: 60%
  - Validation & Test Sets: 20%

Features Used

x_1: Heartrate (HR)
x_2: Number of steps (Step)
x_3: Stress score (Stress)
x_4: Awake duration (Awake)
x_5: Deep sleep duration (Deep)
x_6: Light sleep duration (Light)
x_7: REM sleep duration (REM)
x_8: Total non-REM sleep duration (NonREMTotal)
x_9: Total sleep duration (Total)
x_{10}: Time non-REM sleep percentage (NonREMPercentage)
x_{11}: Sleep efficiency (SleepEfficiency)

x_{12}: Day of the week (TimestampDayofWeek)
x_{13}: Sine value of Hour of the day (TimestampHourSin)
x_{14}: Cosine value of hour of the day (TimestampHourCos)

Metrics

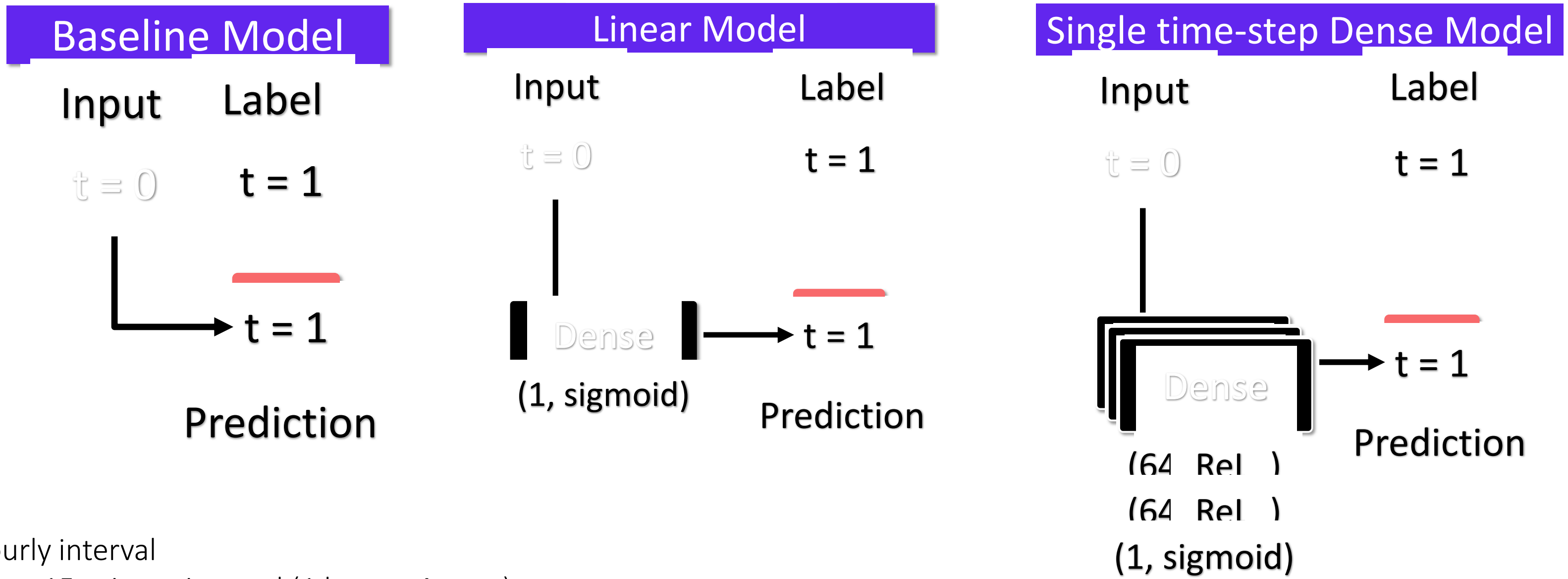
- Balanced Accuracy
- AUC
- Other metrics:
  - Accuracy, F1 Score, Precision, Recall

Models Considered

Current Time Models:

- Input: current time step (*t = 0*)
- Output: next hour (*t = 1)*

$$y_{t+1} = f\ (X_t, y_t), X_t = \{x_1, x_2, \ \dots, x_{14}\}$$



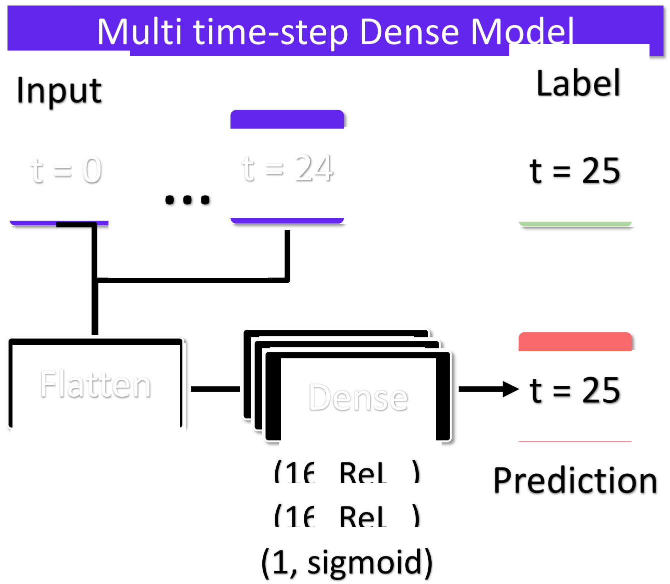
Note:
- Shown as hourly interval
- Data is in every 15-minute interval (1 hour = 4 rows)

Models Considered

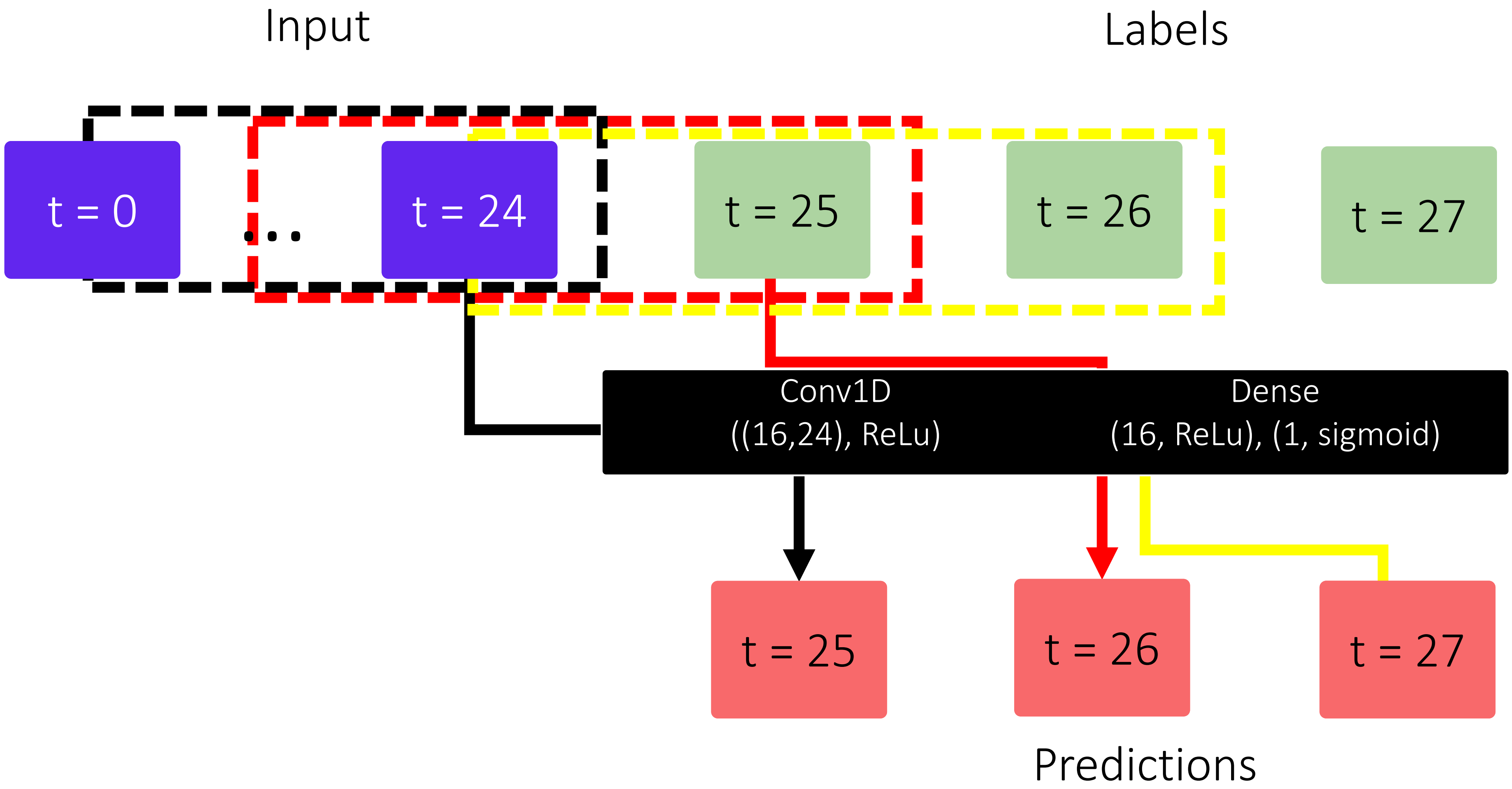
1-Day Models:

- Input: from the last day (*t = 0…24*)
- Output: next hour (*t = 1)*

$$y_{t+1} = f(M(X, y, w)), M = \begin{bmatrix} X_t & y_t \\ X_{t-1} & y_{t-1} \\ \vdots & \vdots \\ X_{t-w} & y_{t-w} \end{bmatrix}$$

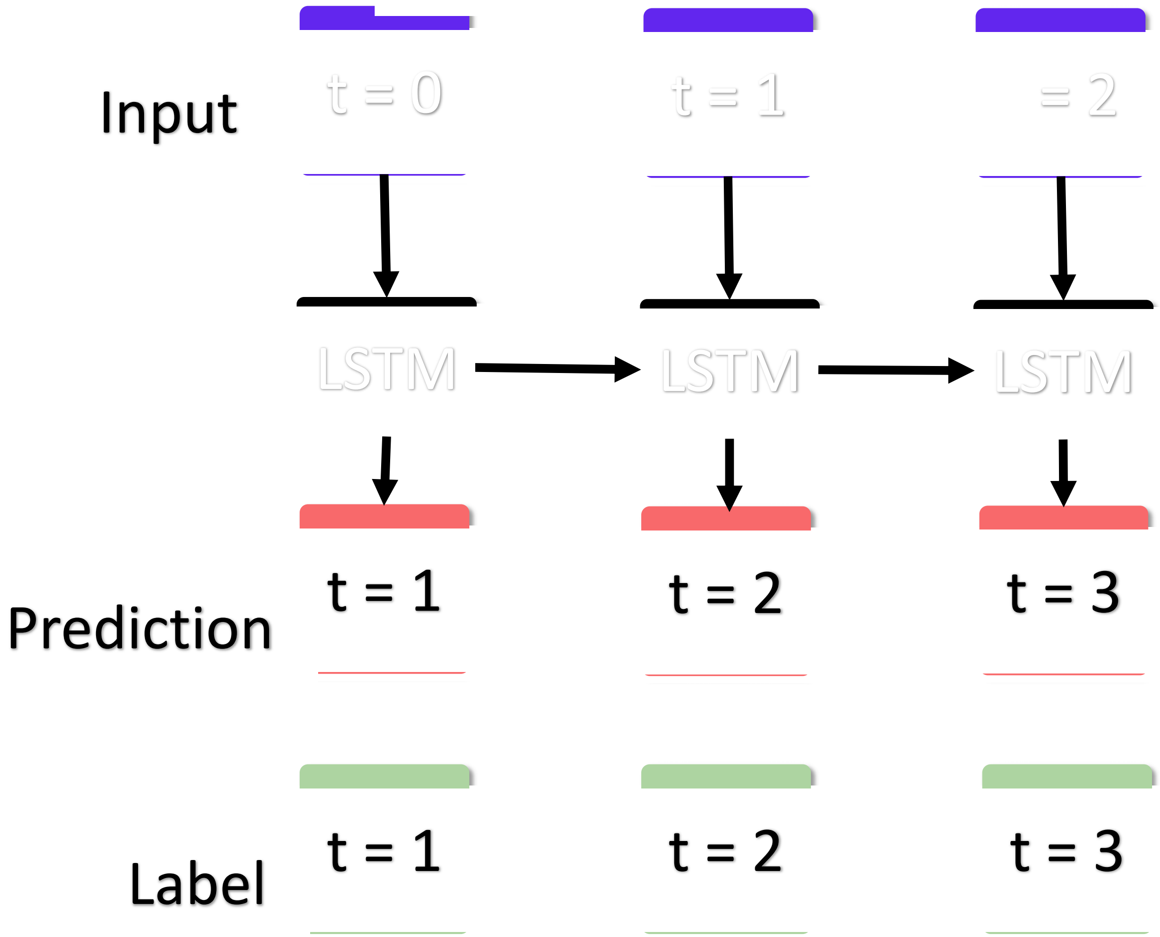


CNN Model



Input
Labels
t = 0
...
t = 24
t = 25
t = 26
t = 27
Conv1D
((16,24), ReLu)
Dense
(16, ReLu), (1, sigmoid)
t = 25
t = 26
t = 27
Predictions

LSTM Model



Participants' Demographics

Participants

4 6

Mean Age
57.1 yrs. old
$(\sigma = 15.059)$

H&Y Score

2.8
$(\sigma = 0.632)$

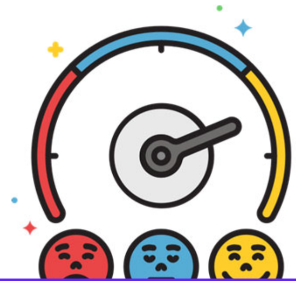
2 = bilateral involvement without impairment of balance,

JCLD

1.7
$(\sigma = 0.483)$

1 = Little assistance
2 = Partial assistance

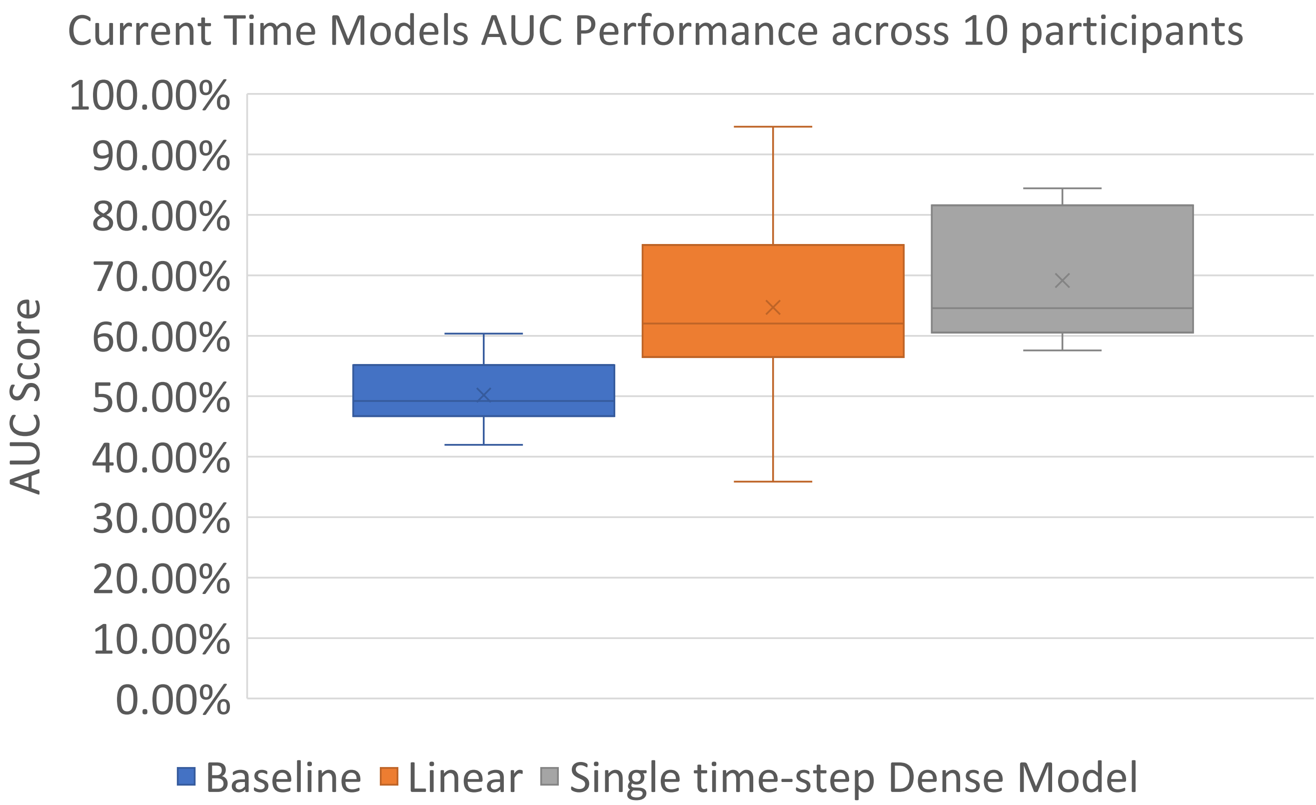
PDQ-8



42.50%
$(\sigma = 0.192)$

Higher score means worse QoL

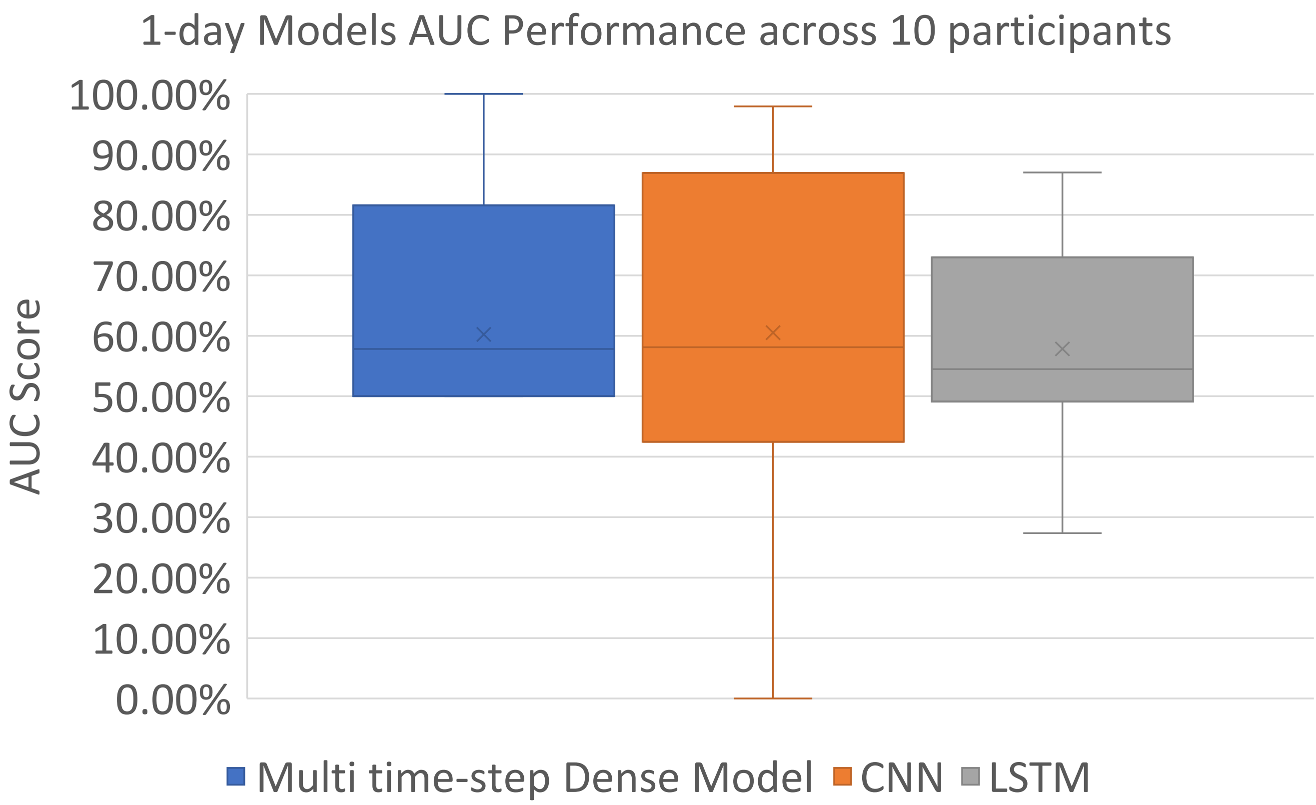
Can a wrist-worn fitness tracker datasets be used to forecast wearing-off in the next hour?



Current Time Models

- Single time-step Dense model
- Balanced Accuracy: 79.05% ± 7.09%
- AUC: 69.14% ± 10.60%

Can a wrist-worn fitness tracker datasets be used to forecast wearing-off in the next hour?



1-day Models

- CNN model
- Balanced Accuracy: 80.64% ± 10.36%
- AUC: 60.52% ± 30.26%

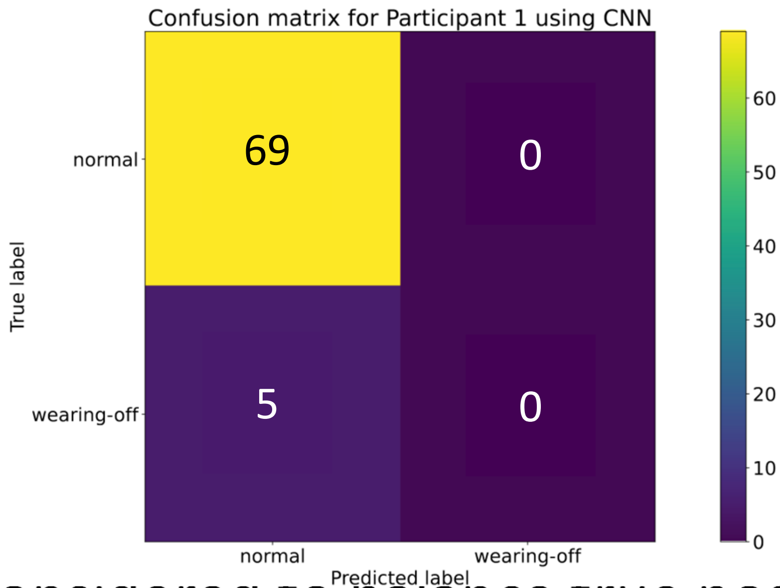
Which among the six deep learning architectures performed well in forecasting wearing-off in the next hour?

Architectures	Bal. Acc.	AUC	Precision	Recall
Baseline	79.05% ± 07.09%	50.20% ± 05.68%	07.81% ± 10.16%	07.59% ± 09.66%
Linear	79.05% ± 07.09%	64.71% ± 16.25%	05.08% ± 06.34%	28.73% ± 33.58%
Single time-step Dense	79.05% ± 07.09%	69.14% ± 10.60%	06.25% ± 13.50%	06.15% ± 15.83%
Multi time-step Dense	80.64% ± 10.36%	60.23% ± 28.33%	18.55% ± 32.29%	36.17% ± 42.24%
CNN	80.64% ± 10.36%	60.52% ± 30.26%	18.61% ± 31.98%	25.06% ± 39.10%
LSTM	50.00% ± 00.00%	57.83% ± 17.13%	03.20% ± 05.17%	08.10% ± 14.78%

- Multi time-step Dense & CNN had high scores in Bal. Acc. Precision, & Recall
- Single time-step Dense had the highest AUC

Discussion

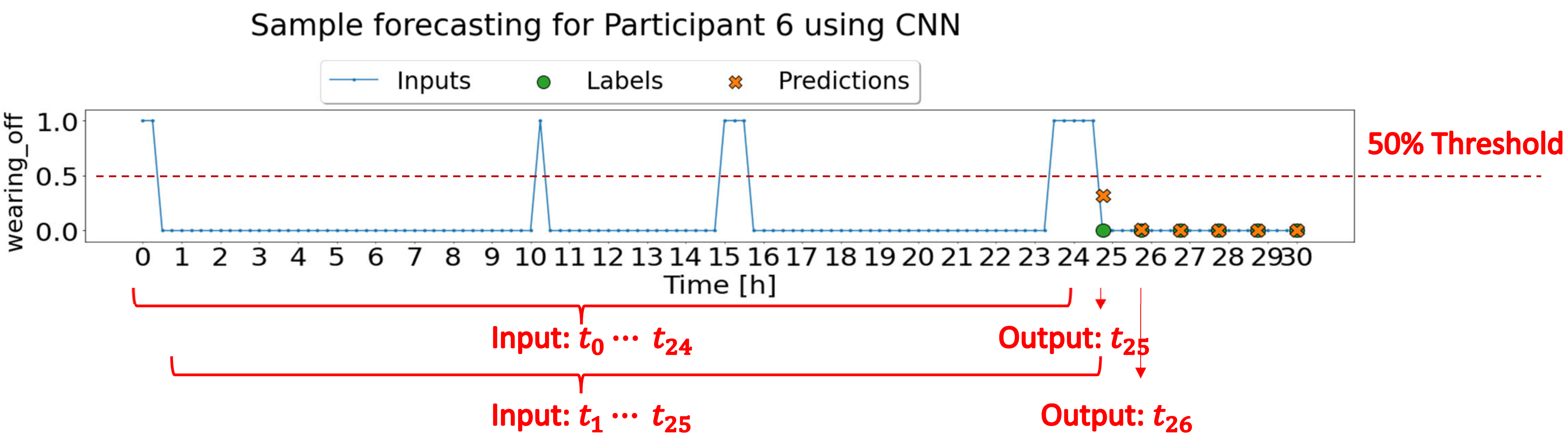
- Multi time-step Dense & CNN models still had a recall score lower than 50%
  - Low recall score = Missed wearing-off forecast



- AUC metric had been considered to balance true positive rate (recall) & false positive rate

Discussion

- Need to adjust the threshold value for the forecast probability



Conclusion

Can wrist-worn fitness tracker datasets be used to forecast wearing-off in the next hour?

<u>YES, forecasting</u> wearing-off using commercial wrist-worn fitness tracker datasets was **<u>feasible</u>.**

- With current time data
- With 1-day's worth of data

Which among the DL architectures performed well in forecasting wearing-off in the next hour?

Single time-step Dense	Multi time-step Dense	CNN
Highest AUC: 69.14% ± 10.60%		Highest balanced accuracy: 80.64% ± 10.36%

Detecting & forecasting earing-off was feasible within a certain limit, even with only commercial fitness tracker features.

Detection & Forecasting models can be used in PD management

Future Work

- Will an optimized threshold value provide better forecasts in real-life application with PD patients?
- Challenge in collecting wearing-off labels or lack of ground truth data
  - How does the model handle missing labels when making wearing-off forecasts in real-life application with PD patients?
  - Can the model approximate the missing labels using prior information, then update the model upon new information?

Related Works: Non-Motor Aspect of PD & WO

Heart Rate

- PD patients’ blood pressure & heart rate were studied during WO.
  - There’s statistical difference in blood pressure change

Stress

- Highly sensitive to the effects of stress with clinical evidence
  - Worsen tremors, freezing of gait (FoG) and dyskinesia

Sleep & Sleep Stages

- Rapid eye movement sleep behavior disorder, sleep disturbances affect PD patients.

V. Pursiainen, J. T. Korpelainen, T. H. Haapaniemi, K. A. Sotaniemi, and V. V. Myllylä, ‘Blood pressure and heart rate in parkinsonian patients with and without wearing-off’, *European Journal of Neurology*, vol. 14, no. 4, pp. 373–378, 2007, doi: 10.1111/j.1468-1331.2007.01672.x.

M. Salsone *et al.*, ‘Cardiac sympathetic index identifies patients with Parkinson’s disease and REM behavior disorder.’, *Parkinsonism & related disorders*, 2016, doi: 10.1016/j.parkreldis.2016.03.004.

J.-E. Lee, J.-S. Kim, D.-W. Ryu, Y.-S. Oh, I. R. Yoo, and K.-S. Lee, ‘Cardiac Sympathetic Denervation Can Predict the Wearing-off Phenomenon in Patients with Parkinson Disease’, *J Nucl Med*, vol. 59, no. 11, pp. 1728–1733, Nov. 2018, doi: 10.2967/jnumed.118.208686.

A. van der Heide, M. J. Meinders, A. E. M. Speckens, T. F. Peerbolte, B. R. Bloem, and R. C. Helmich, ‘Stress and Mindfulness in Parkinson’s Disease: Clinical Effects and Potential Underlying Mechanisms’, *Movement Disorders*, vol. 36, no. 1, pp. 64–70, 2021, doi: 10.1002/mds.28345.

Related Works: Detecting Wearing-Off

Study	Goal	Data Used	Result
Hoff et al, 2004	Detection: wearing-off, dyskinesia	Accelerometer	Sensitivity: 60% - 71% Specificity: 66% - 76%
Keijsers, 2006	Detection: wearing-off, bradykinesia, hypokinesia, tremor	Accelerometer	Sensitivity: 97% Specificity: 97%
Hssayeni, 2019	Detection: wearing-off	Accelerometer, Gyroscope	Accuracy: 90.5% Sensitivity: 94.2% Specificity: 85.4%
Aich, 2020	Detection: Wearing-off, Gait features	Accelerometer	Random Forest: 96.72% accuracy

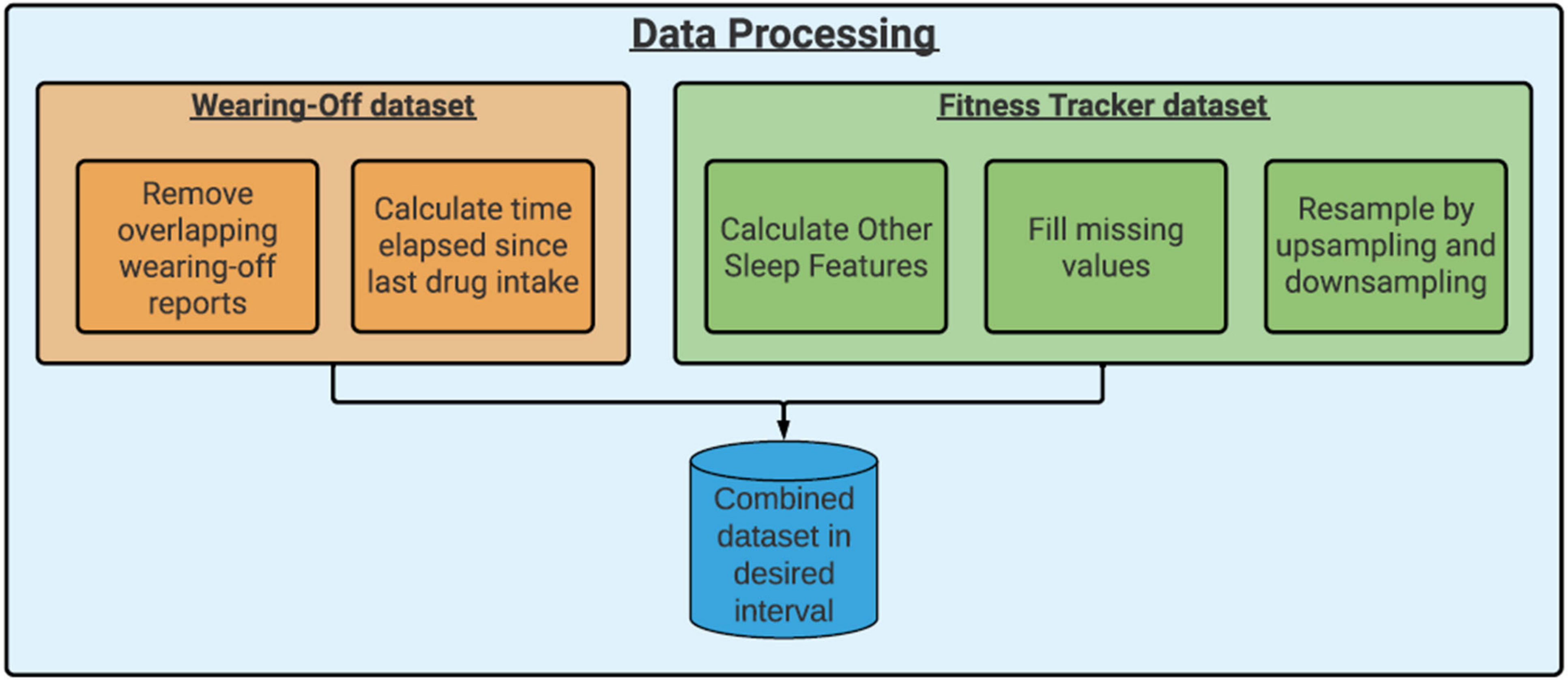
- Focused on using motion data
- Generalized models for the participants (except for Hssayeni, 2019)
- Detection of wearing-off rather than prediction

J. I. Hoff, V. van der Meer, and J. J. van Hilten, 'Accuracy of Objective Ambulatory Accelerometry in Detecting Motor Complications in Patients With Parkinson Disease', *Clinical Neuropharmacology*, vol. 27, no. 2, pp. 53–57, Apr. 2004.
N. L. W. Keijsers, M. W. I. M. Horstink, and S. C. A. M. Gielen, 'Ambulatory motor assessment in Parkinson's disease', *Movement Disorders*, vol. 21, no. 1, pp. 34–44, 2006, doi: https://doi.org/10.1002/mds.20633.
M. D. Hssayeni, M. A. Burack, J. Jimenez-Shahed, and B. Ghoraani, 'Assessment of response to medication in individuals with Parkinson's disease', *Med Eng Phys*, vol. 67, pp. 33–43, May 2019, doi: 10.1016/j.medengphy.2019.03.002.
S. Aich *et al.*, 'A Supervised Machine Learning Approach to Detect the On/Off State in Parkinson's Disease Using Wearable Based Gait Signals', *Diagnostics*, vol. 10, no. 6, p. 421, Jun. 2020, doi: 10.3390/diagnostics10060421.

Garmin vivosmart4

- Chosen due to sleek, lightweight; waterproof
- Weight: 16.5g – 17.1g
- Dimensions: 15 x 10.5 x 197 mm
- Communicates with Garmin Health API via Garmin Connect smartphone app

Data Processing



Sleep Features

$$\text{Total non-REM duration} = \text{Deep sleep duration} + \text{Light sleep duration}$$

$$\text{Total sleep duration} = \text{Total non-REM duration} + \text{REM sleep duration}$$

$$\text{Total non-REM percentage} = \frac{\text{Total non-REM duration}}{\text{Total sleep duration}}$$

$$\text{Sleep efficiency} = \frac{\text{Total sleep duration}}{\text{Total sleep duration} + \text{Total awake duration}}$$