WIP : Parkinson's wearing-off Detection: Forecasting, App Design, and Behavior Feedback

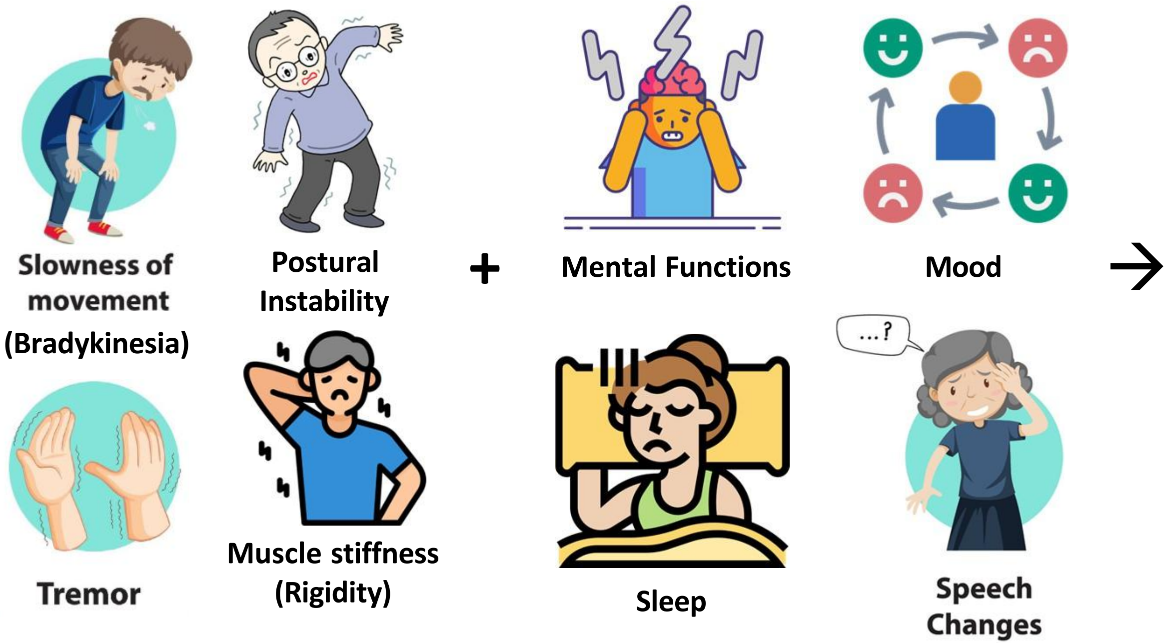
KANIA GUPTA

Kyushu Institute of Technology
Internship Student

Parkinson's Disease (PD)

PD is a slowly progressive disorder of the nervous system due to loss of dopamine-producing brain cells.

Motor Symptoms **Non-Motor Symptoms**



Symptoms and its severity vary for each patient

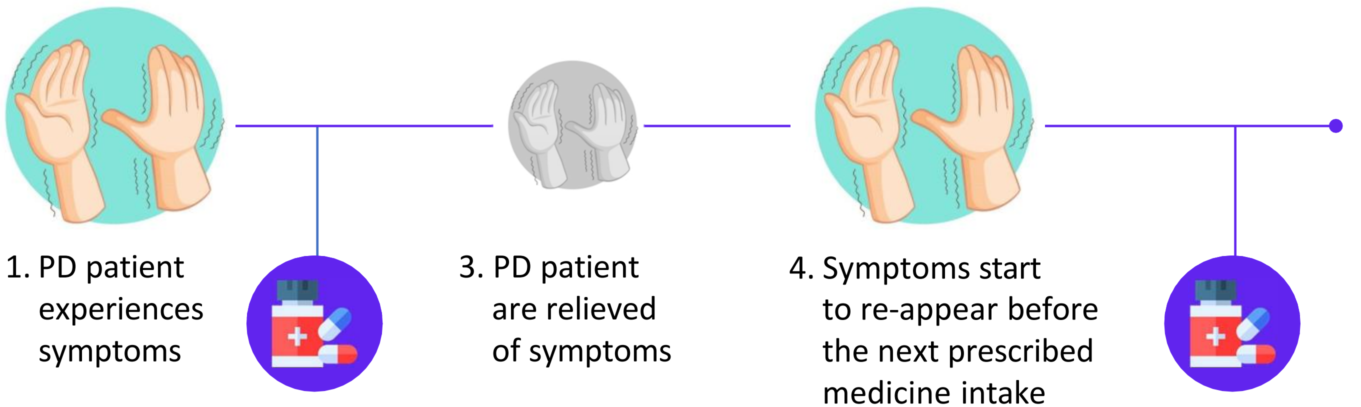
Affects daily life & Quality of Life (QoL)

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.

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/

Victorino, John Noel, et al. "Understanding wearing-off symptoms in Parkinson’s disease patients using wrist-worn fitness tracker and a smartphone." *Procedia Computer Science* 196 (2022): 684-691.

Victorino, John Noel, et al. "Predicting wearing-off of Parkinson’s disease patients using a wrist-worn fitness tracker and a smartphone: A case study." *Applied Sciences* 11.16 (2021): 7354.

Research Objectives

1. **Improve Existing Models**
   Refine and re-evaluate previously developed ML models for higher prediction accuracy.

2. Expand Algorithm Scope
Implement and compare additional machine learning and deep learning models.

3. Customized-Time Model Deployment
Deploy the best-performing model on a server to enable real-time, customized wearing-off prediction.

4. Application Enhancement
Improve the WoForecastProto app for intuitive visualization of prediction results.

5. Integrate Behavioral Feedback
Designed and embedded post-prediction questions in FonLog to collect user responses and enhance future model performance.

Related Search

1. Victorino et al. (2021)

Predicting Parkinson's Wearing-Off Using Wearables

•Used Garmin + FonLog data

•Achieved up to **76.9% balanced accuracy**

•Personalized ML models for real-life prediction

•[Applied Sciences, 2021]

2. Victorino et al. (2022)

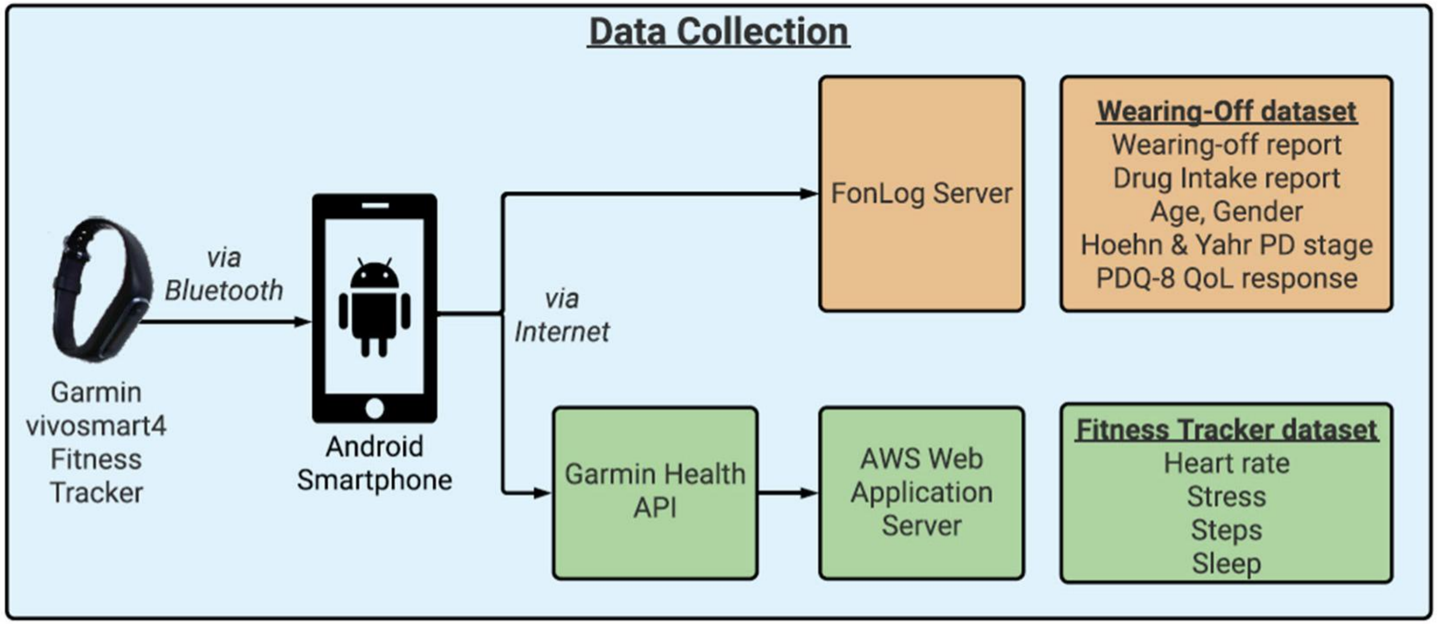
Understanding Wearing-Off Symptoms via Smart Devices

•Analyzed **sleep, steps, drug timing** as key indicators

•Showed feasibility of **commercial trackers**

•[Procedia Computer Science, 2022]

Previously Collected Data Process

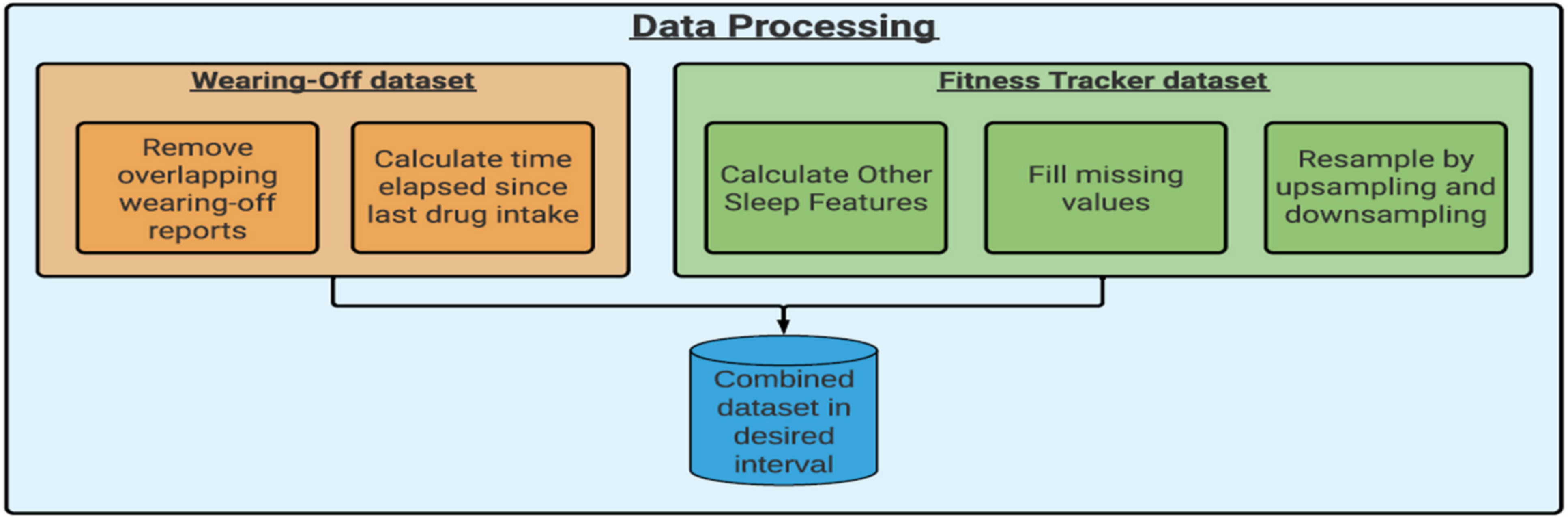


- PD patients who are aware of wearing-off
- PD participants were asked to contribute 30 days' worth of data from 23rd February'21 to 24th March'21
  - Wear fitness tracker
  - Report wearing-off period and drug intake period using a smartphone app

Victorino, John Noel, et al. "Understanding wearing-off symptoms in Parkinson's disease patients using wrist-worn fitness tracker and a smartphone." *Procedia Computer Science* 196 (2022): 684-691.

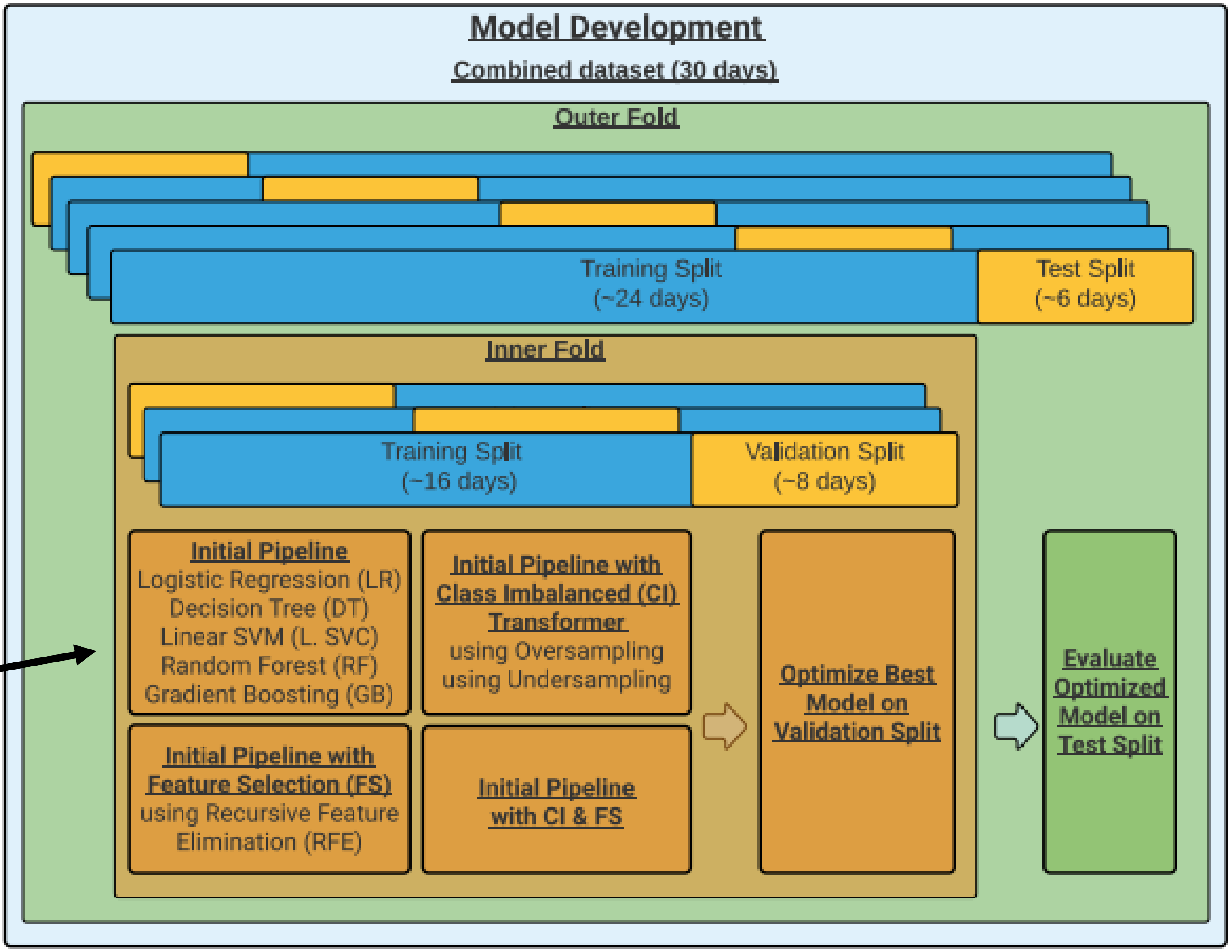
Victorino, John Noel, et al. "Predicting wearing-off of Parkinson's disease patients using a wrist-worn fitness tracker and a smartphone: A case study." *Applied Sciences* 11.16 (2021): 7354.

Data Processing for ML+DL




```
                 |
                 ▼
  ┌──────────────────────────────┐
  │ DL Prep & Modeling           │
  │ • Reshape to 3D              │
  │ • Train with Keras models    │
  └──────────────────────────────┘
```

Model Development for ML



+ LightGBM (LGBMC)

Model Development for DL

1. Data Preparation
   - Clean & select features
2. Feature Engineering
   - Label shift, scaling
3. Input Reshaping
   - MLP: 2D LSTM/CNN: 3D
4. Model Building
   - MLP / LSTM / 1D CNN
5. Train & Validate
   - Adam, BCE, val split
6. Evaluate Model
   - Metrics, confusion matrix

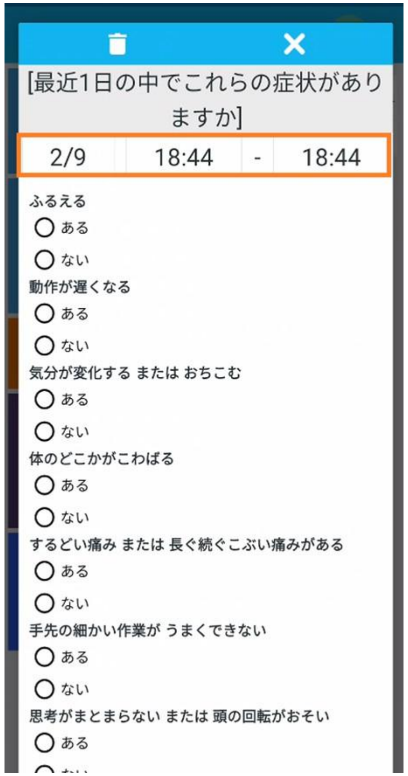
Garmin vivosmart4 Datasets

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 • Awake

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



- Age, Gender, Hoehn and Yahr PD Stage, PDQ-8 QoL response (one time)

Participant Demographics

	Participant 1	Participant 2
Age	43	38
Gender	Female	Female
H&Y	2: Bilateral or midline involvement without impairment of balance	3: Bilateral disease: mild to moderate disability with impaired postural reflexes; physically independent
JCLD	1: Little assistance is needed in daily life and outpatient visits	2: Partial assistance is required for daily life and outpatient visits
PDQ-8	37.5%	65.63%

Results and Discussion

Best hyperparameter configuration performance on the validation set among all ML models

Participant1_15mins

	Balanced Accuracy	F1 Score	Precision
Initial Pipeline	0.5680	0.4995	0.5908
Feature Selection	0.5713	0.4438	0.6791
Class Imbalance	0.5712	0.5325	0.5840
Feature Selection + Class Imbalance	0.5714	$\boxed{0.4688}$	0.6104

Participant1_15secs

	Balanced Accuracy	F1 Score	Precision
Initial Pipeline	0.5559	0.4852	0.5292
Feature Selection	0.5683	0.4510	0.6133
Class Imbalance	0.5700	$\boxed{0.5173}$	0.5395
Feature Selection + Class Imbalance	0.5629	0.4346	0.6080

FS+CI

Participant2_15mins

	Balanced Accuracy	F1 Score	Precision
Initial Pipeline	0.6726	0.1128	0.0610
Feature Selection	0.6794	0.1416	0.0832
Class Imbalance	0.7351	0.1843	0.1393
Feature Selection + Class Imbalance	0.7505	$\boxed{0.1583}$	0.0887

Participant2_15secs

	Balanced Accuracy	F1 Score	Precision
Initial Pipeline	0.6843	0.0855	0.0449
Feature Selection	0.7001	0.1086	0.0605
Class Imbalance	0.7321	0.1019	0.0541
Feature Selection + Class Imbalance	0.7529	$\boxed{0.1303}$	0.0725

Results and Discussion

Comparison of ML algorithms performance on the validation set using the FS+CI best pipeline

Participant1_15mins

	Balanced Accuracy	F1 Score	Precision
LR	0.5367	0.4074	0.6230
DT Classifier	0.5595	0.4525	0.5970
Linear SVC	0.5180	0.3821	0.5977
RF Classifier	0.5287	0.4445	0.5332
GB Classifier	0.5596	$\boxed{0.5130}$	0.5645
LGBMC	0.5549	0.5066	0.5587

Participant1_15secs

	Balanced Accuracy	F1 Score	Precision
LR	0.5294	0.3877	0.5642
DT Classifier	0.5325	0.4545	0.5005
Linear SVC	0.5253	0.4014	0.5547
RF Classifier	0.5227	0.4430	0.4846
GB Classifier	0.5365	0.4702	0.5045
LGBMC	0.5587	$\boxed{0.5054}$	0.5267

GBC

Participant2_15mins

	Balanced Accuracy	F1 Score	Precision
LR	0.6291	0.0918	0.0486
DT Classifier	0.7146	0.1342	0.0742
Linear SVC	0.6328	0.1022	0.0552
RF Classifier	0.7348	$\boxed{0.1454}$	0.0809
GB Classifier	0.7348	$\boxed{0.1286}$	0.0693
LGBMC	0.7151	0.1255	0.0681

Participant2_15secs

	Balanced Accuracy	F1 Score	Precision
LR	0.6385	0.0760	0.0398
DT Classifier	0.7228	0.1074	0.0584
Linear SVC	0.6480	0.0772	0.0404
RF Classifier	0.7109	0.1094	0.0598
GB Classifier	0.6926	0.0911	0.0482
LGBMC	0.7326	$\boxed{0.1496}$	0.1406

Results and Discussion

Comparison of DL algorithms performance on the test set using keras except for MLP

Participant1_15mins

	F1 score	Recall	Precision
1D CNN	0.6955	0.6960	0.6965
LSTM	0.7413	0.7415	0.7440
MLP	0.6465	0.6465	0.6465

Participant1_15secs

	F1 score	Recall	Precision
1D CNN	0.8045	0.8020	0.8131
LSTM	0.9222	0.9224	0.9220
MLP	0.7800	0.7782	0.7852

LSTM

Participant2_15mins

	F1 score	Recall	Precision
1D CNN	0.6723	0.7004	0.6521
LSTM	0.6271	0.5821	0.8425
MLP	0.4911	0.5000	0.4826

Participant2_15secs

	F1 score	Recall	Precision
1D CNN	0.5936	0.5567	0.8792
LSTM	0.6946	0.6320	0.8935
MLP	0.6832	0.5624	0.8443

Results and Discussion

Average performance of best ML model on the test set using the FS+CI best pipeline + GB best ML model using final Pipeline

Participant1_15mins

	mean	std
Bal. Acc.	0.4852	0.0896
F1 Score	0.4878	0.1022
Acc.	0.4850	0.0893
Precision	0.4767	0.0841
Recall / Sn	0.5039	0.1309
Sp	0.4666	0.0896
AUC	0.4852	0.0896

Participant1_15secs

	mean	std
Bal. Acc.	0.4692	0.0803
F1 Score	0.4302	0.0998
Acc.	0.4712	0.0791
Precision	0.4206	0.0771
Recall / Sn	0.4472	0.1283
Sp	0.4913	0.1050
AUC	0.4692	0.0803

Participant2_15mins

	mean	std
Bal. Acc.	0.6650	0.1628
F1 Score	0.2527	0.3366
Acc.	0.7161	0.2971
Precision	0.2424	0.3802
Recall / Sn	0.6100	0.4247
Sp	0.7201	0.3175
AUC	0.6650	0.1628

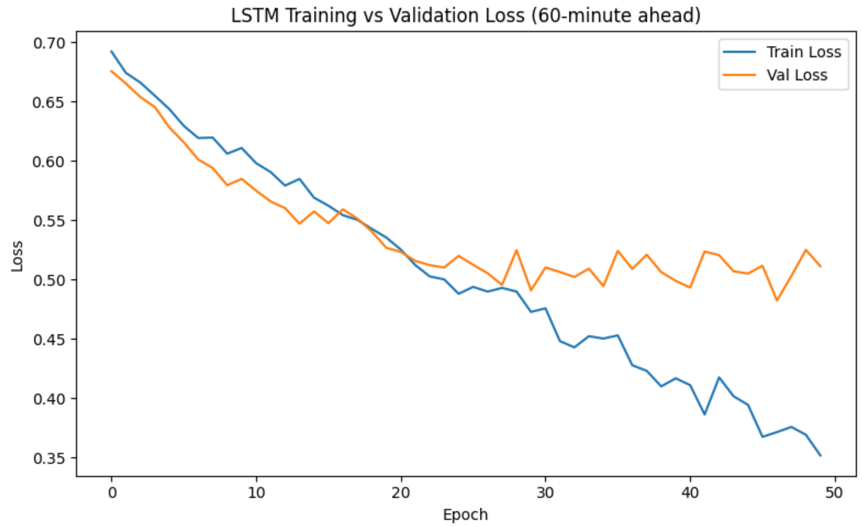
Participant2_15secs

	mean	std
Bal. Acc.	0.7002	0.1706
F1 Score	0.1339	0.0573
Acc.	0.6785	0.2449
Precision	0.0771	0.0327
Recall / Sn	0.7231	0.2742
Sp	0.6773	0.2525
AUC	0.7002	0.1706

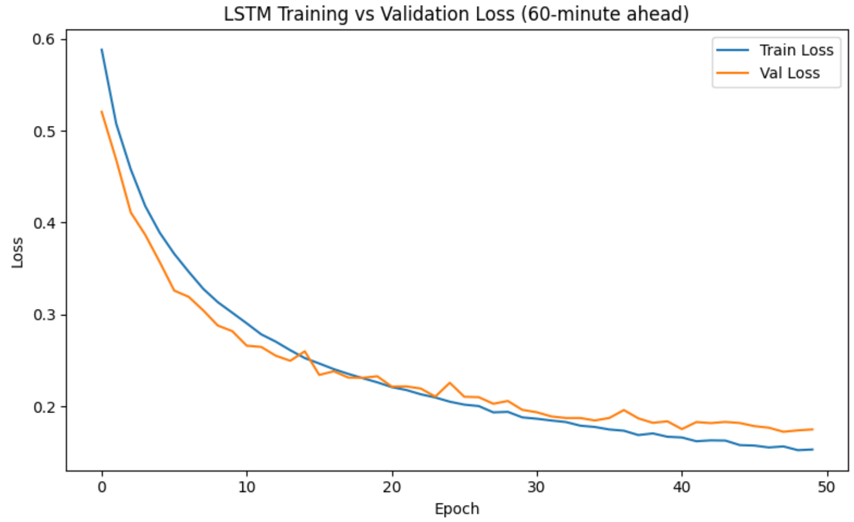
Results and Discussion

Training vs Validation curves for best DL model

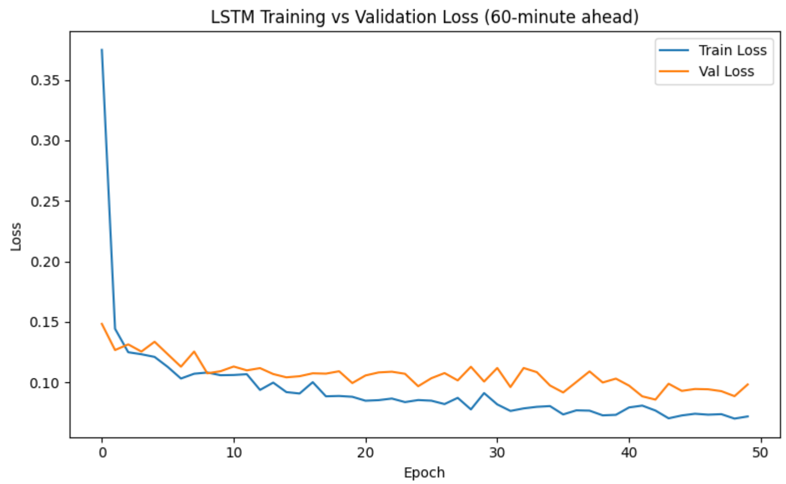
Participant1_15mins



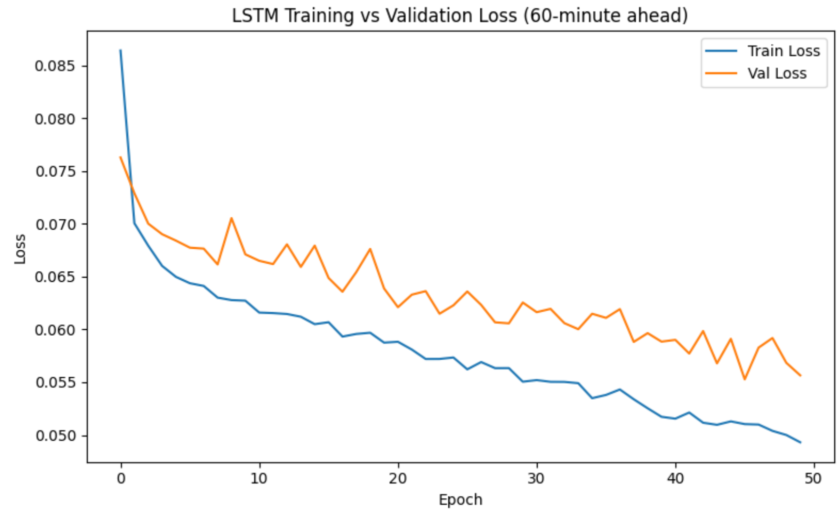
Participant1_15secs



Participant2_15mins



Participant2_15secs



Results and Discussion

Permutation feature importance for the FS+CI best pipeline + GB best model using final pipeline

Participant1_15mins

Features	Values
time from last drug taken	0.3242
heart rate	0.2377
stress score	0.1276
deep	0.1034
steps	0.0869
awake	0.0699
non rem percentage	0.0286
total	0.0159
light	0.0039
rem	0.0020
non rem total	0.0000
sleep efficiency	0.0000

Participant1_15secs

Features	Values
time from last drug taken	0.2828
heart rate	0.1460
steps	0.1294
stress score	0.0885
deep	0.0783
total	0.0540
non rem percentage	0.0530
light	0.0409
sleep efficiency	0.0377
awake	0.0346
rem	0.0333
non rem total	0.0215

Participant2_15mins

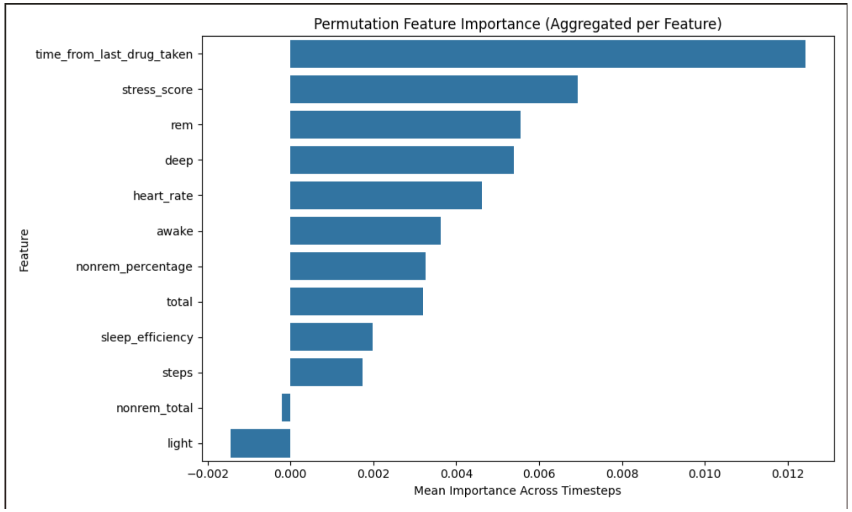
Features	Values
time from last drug taken	0.8102
light	0.0914
heart rate	0.0579
awake	0.0204
non rem total	0.0055
total	0.0042
deep	0.0041
sleep efficiency	0.0032
non rem percentage	0.0027
stress score	0.0003
steps	0.0000
rem	0.0000

Participant2_15secs

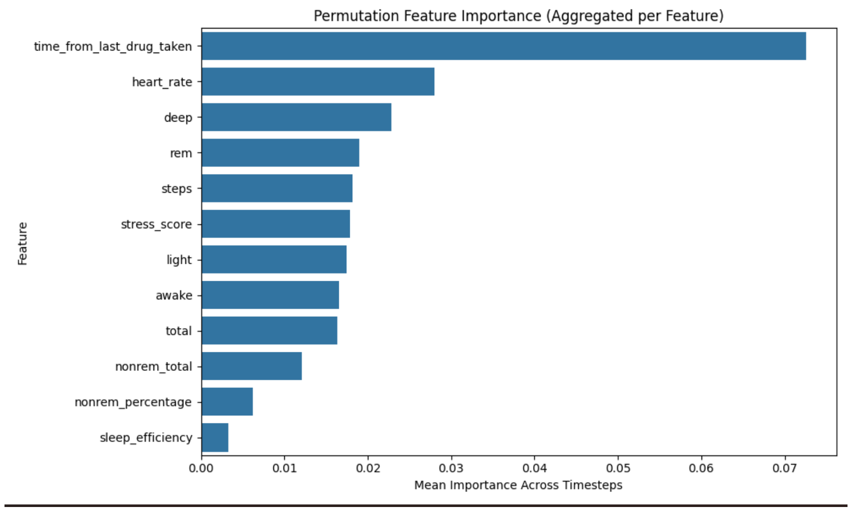
Features	Values
time from last drug taken	0.9115
light	0.0788
heart rate	0.0095
steps	0.0000
awake	0.0000
stress score	0.0000
rem	0.0000
deep	0.0000
non rem total	0.0000
total	0.0000
non rem percentage	0.0000
sleep efficiency	0.0000

Permutation feature importance by the best DL model

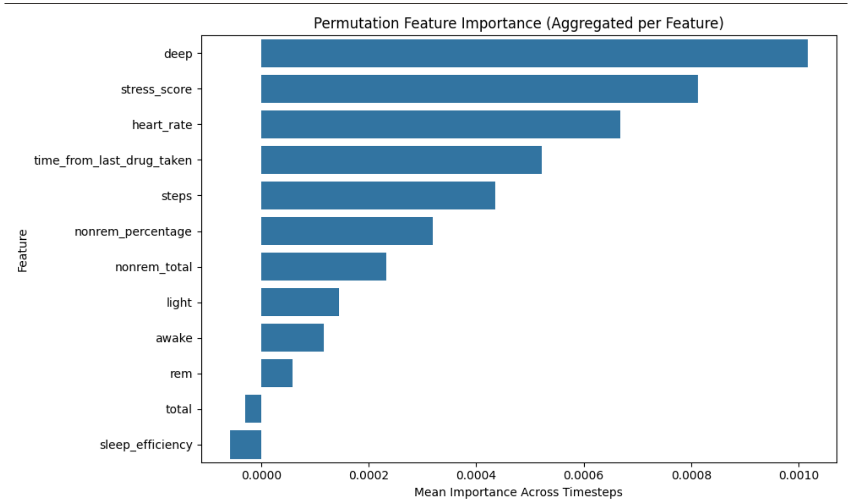
Participant1_15mins



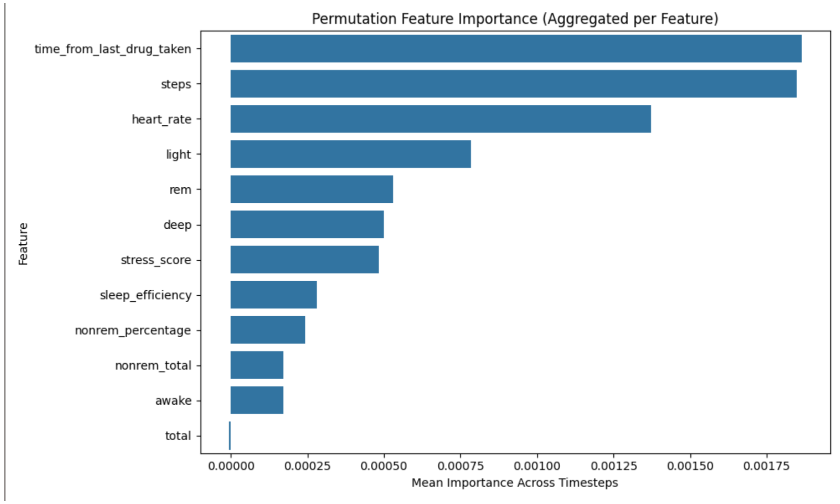
Participant1_15secs



Participant2_15mins

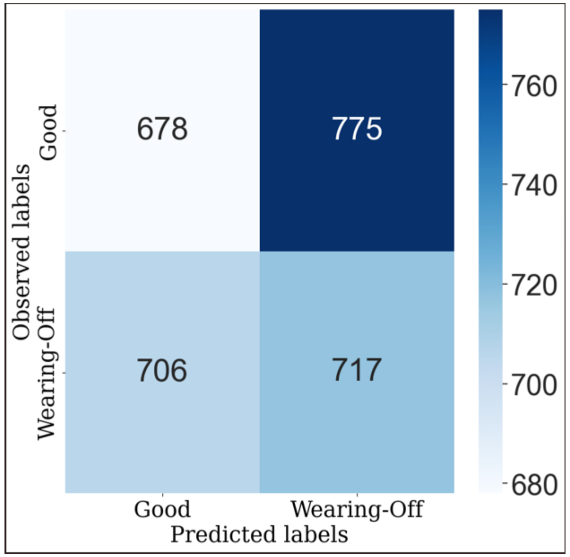


Participant2_15secs

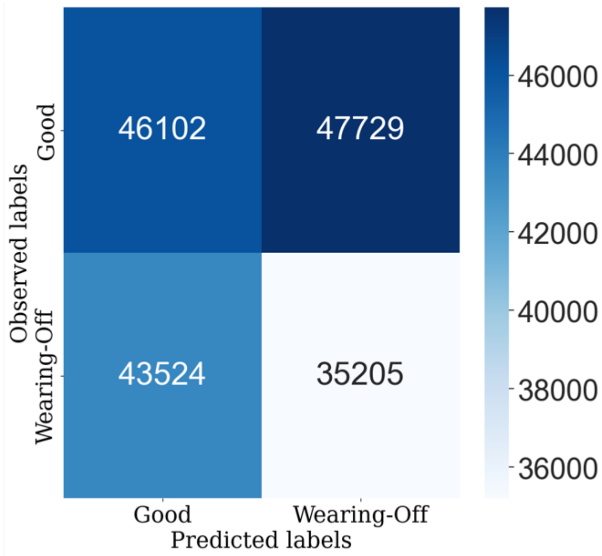


CONFUSION MATRIXs by best ML case

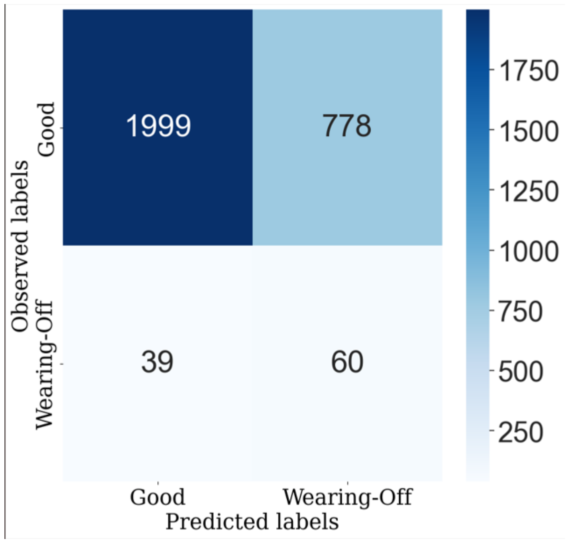
Participant1_15mins



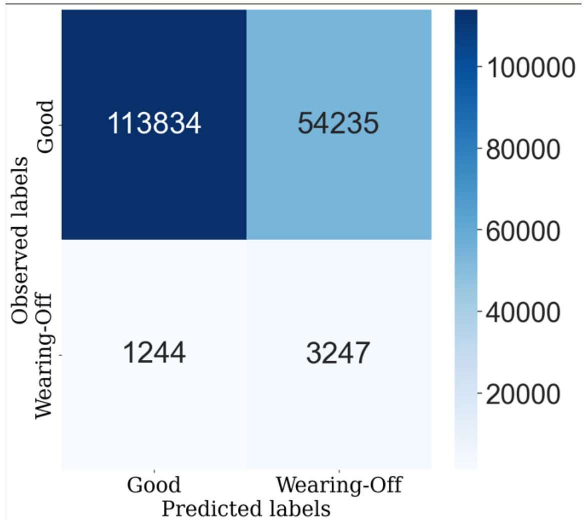
Participant1_15secs



Participant2_15mins

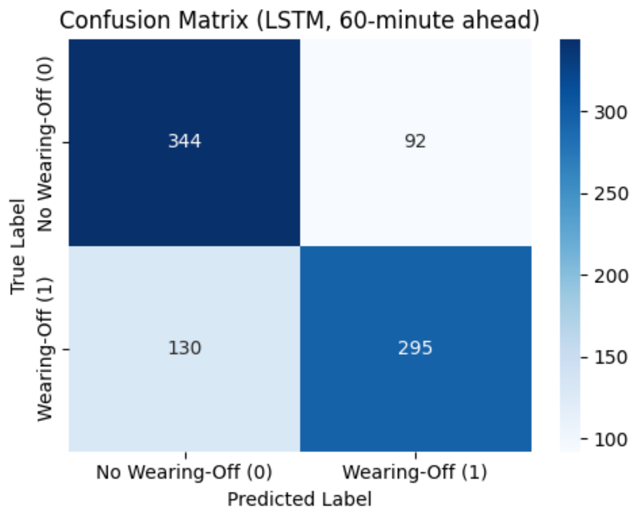


Participant2_15secs

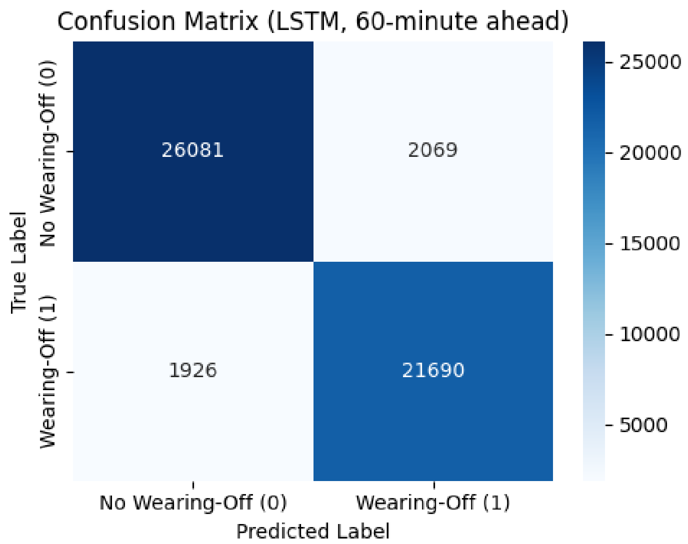


CONFUSION MATRIXs by best DL case

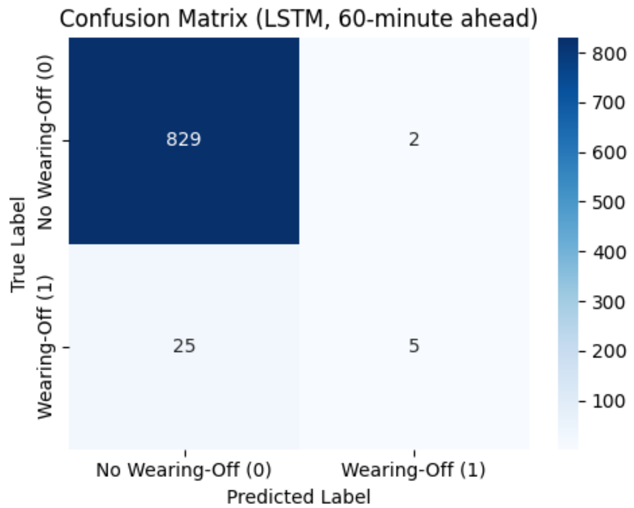
Participant1_15mins



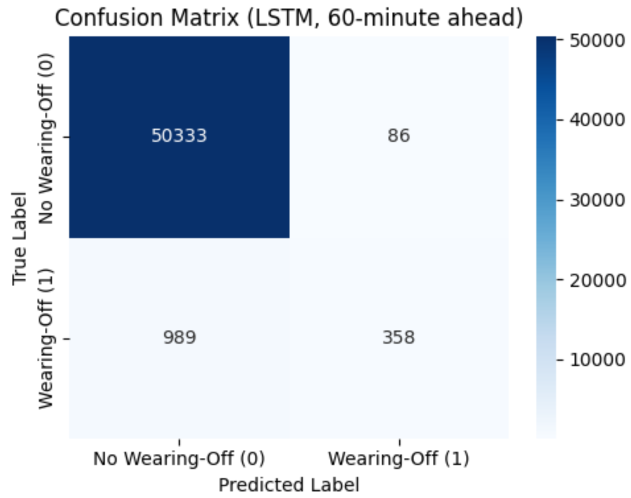
Participant1_15secs



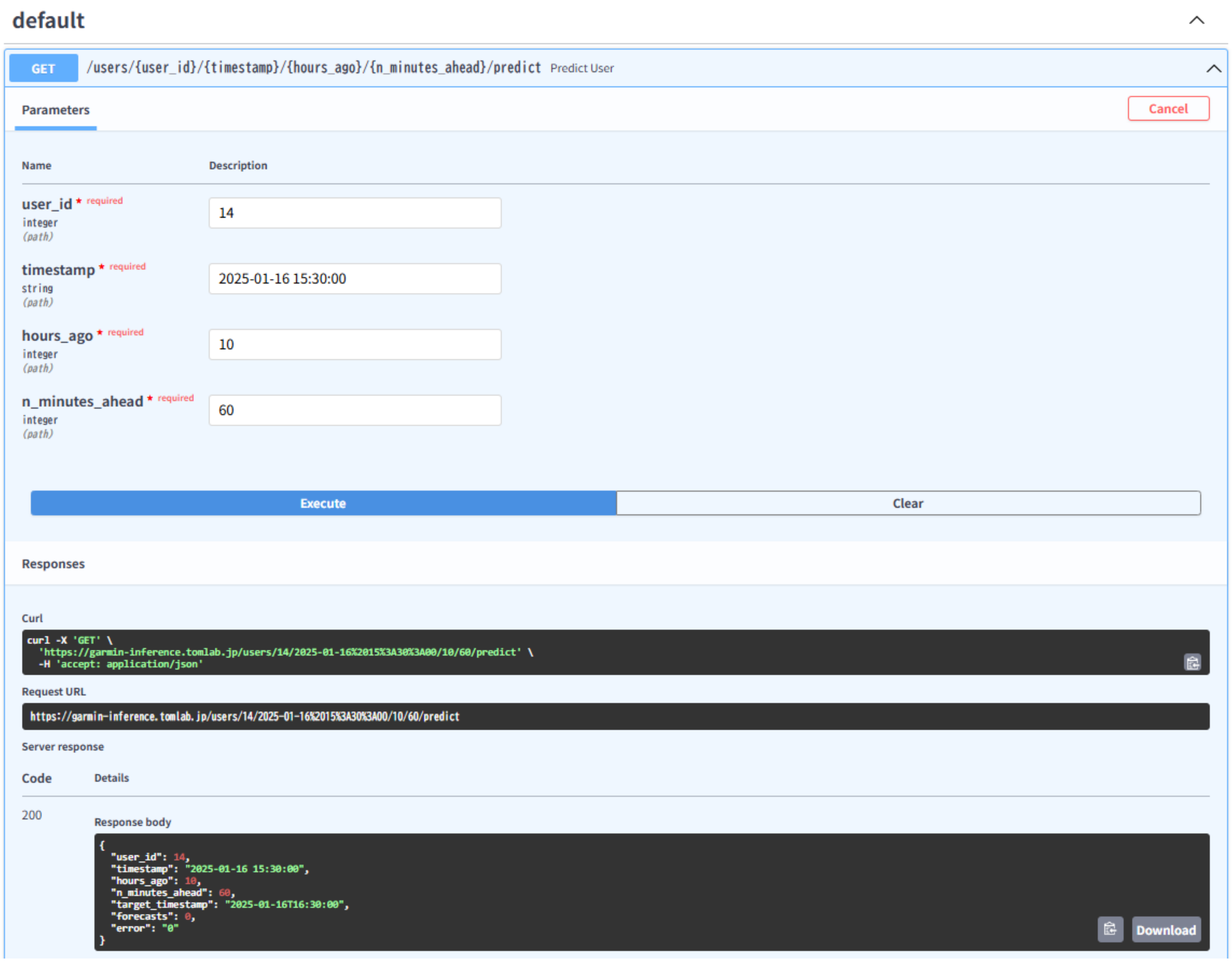
Participant2_15mins



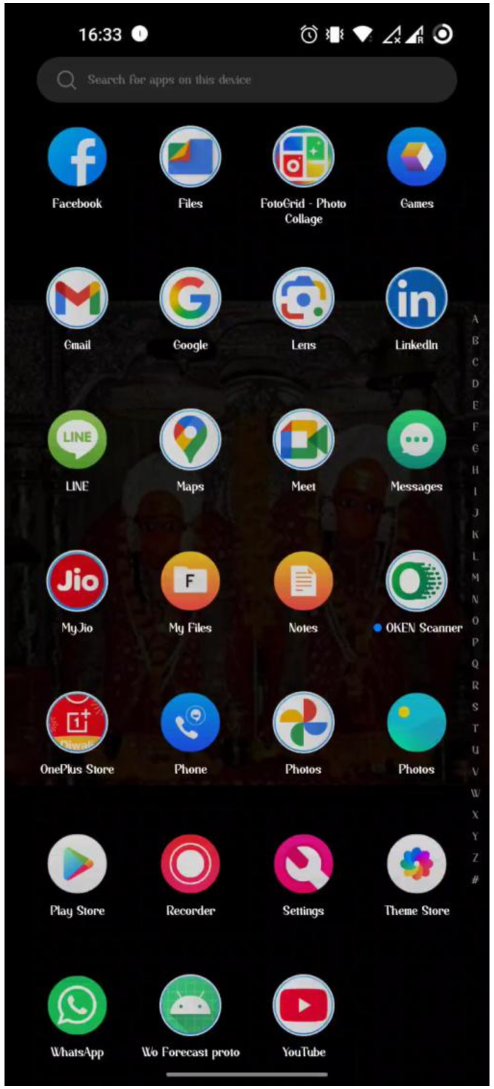
Participant2_15secs



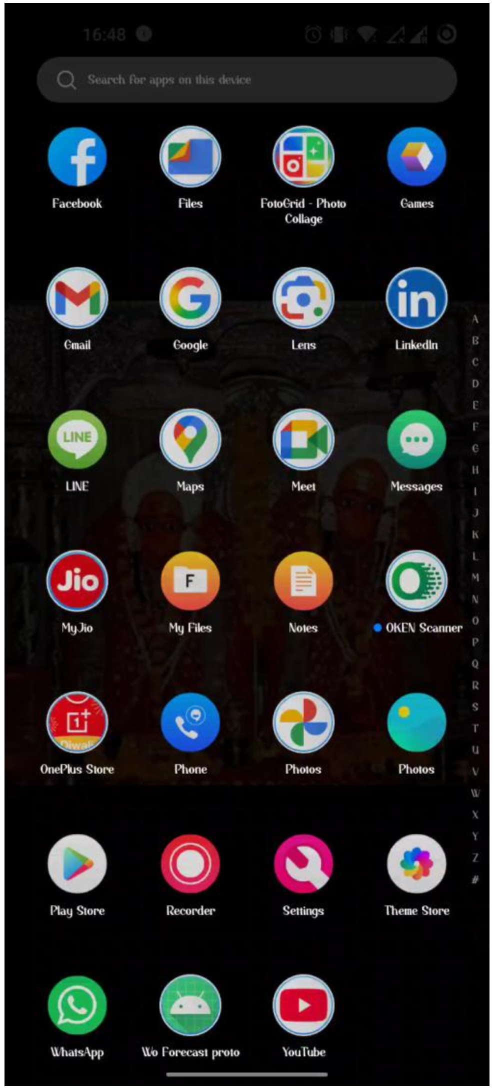
Cutomized-Time Deployment on Garmin server for Wearing-Off Predic



Application Enhancement-WoForecastProto (Intervention App) Demo



On-Off Case



Error Case

Results and Discussion

Post-Prediction Behavioral Questionnaire

Record Types

Name	Form Type	order	Show in summary	Show in heat	Recording Rules	Handover	Control
1. Did you receive a wearing off prediction from the app?	radio	1	Yes	Yes		No	Edit Disable
2. What actions did you take in response to the prediction? (Multiple choices possible)	Checkbox	2	Yes	Yes		No	Edit Disable
3. (Only for those who selected "Other" in question 2) Please describe the actions you took.	text	3	Yes	Yes		No	Edit Disable
4. (Only for those who selected "did nothing" in question 2) What was the reason you did not take any action? (Multiple answers possible)	Checkbox	4	Yes	Yes		No	Edit Disable
5. (Only for those who selected "Other" in question 4) Please provide specific reasons why you did not take action.	text	5	Yes	Yes		No	Edit Disable

makido-exp.fonlog.com/activity_types/4863

4. (Only for those who selected "did nothing" in question 2) What was the reason you did not take any action? (Multiple answers possible)	Checkbox	4	Yes	Yes		No	Edit Disable
5. (Only for those who selected "Other" in question 4) Please provide specific reasons why you did not take action.	text	5	Yes	Yes		No	Edit Disable
6. Did the actions you took help you manage your symptoms?	radio	6	Yes	Yes		No	Edit Disable
7. Overall, how useful was the forecast?	radio	7	Yes	Yes		No	Edit Disable
8. Did the predictions help you feel prepared or in control of your situation?	radio	8	Yes	Yes		No	Edit Disable

Add Record Type

Edit Back to list

Summary

•**Best Models Identified**
GBC was the top ML model; **LSTM** outperformed all and was selected as the overall best.

•**Algorithm Expansion**
Evaluated and compared various ML and DL models to enhance prediction accuracy.

•**Customized Real-Time Deployment**
Deployed the LSTM model on a server for **instant predictions** with **user-defined forecast intervals** (e.g., 15–60 mins).

•**App Enhancement**
Improved the **WoForecastProto** app with intuitive, real-time prediction visualization.

•**Behavioral Feedback Integration**
Embedded follow-up questions in **FonLog** to collect patient input and **continuously improve model performance**.

Research Future Work

Analyzing the results with more experiments with the support of PD patients

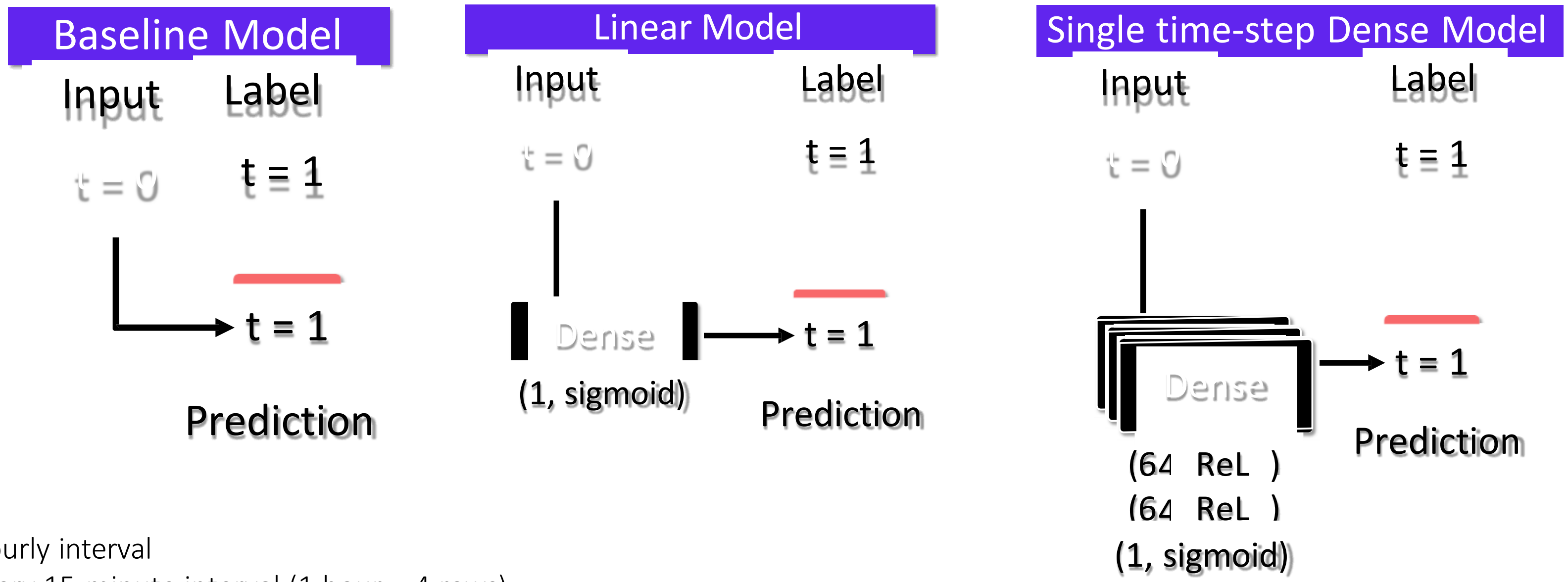
THANK YOU

Models Considered

Current Time Models:

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

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



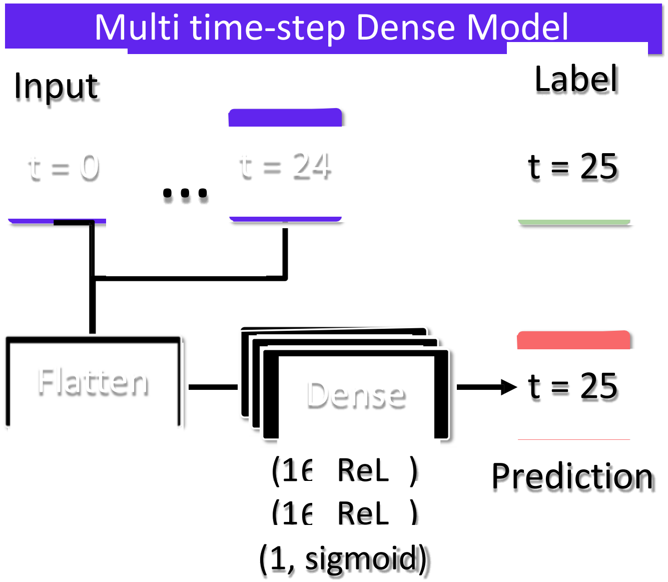
Note:
- Shown as hourly interval
- Data is in every 15-minute interval (1 hour = 4 rows)

Models Considered

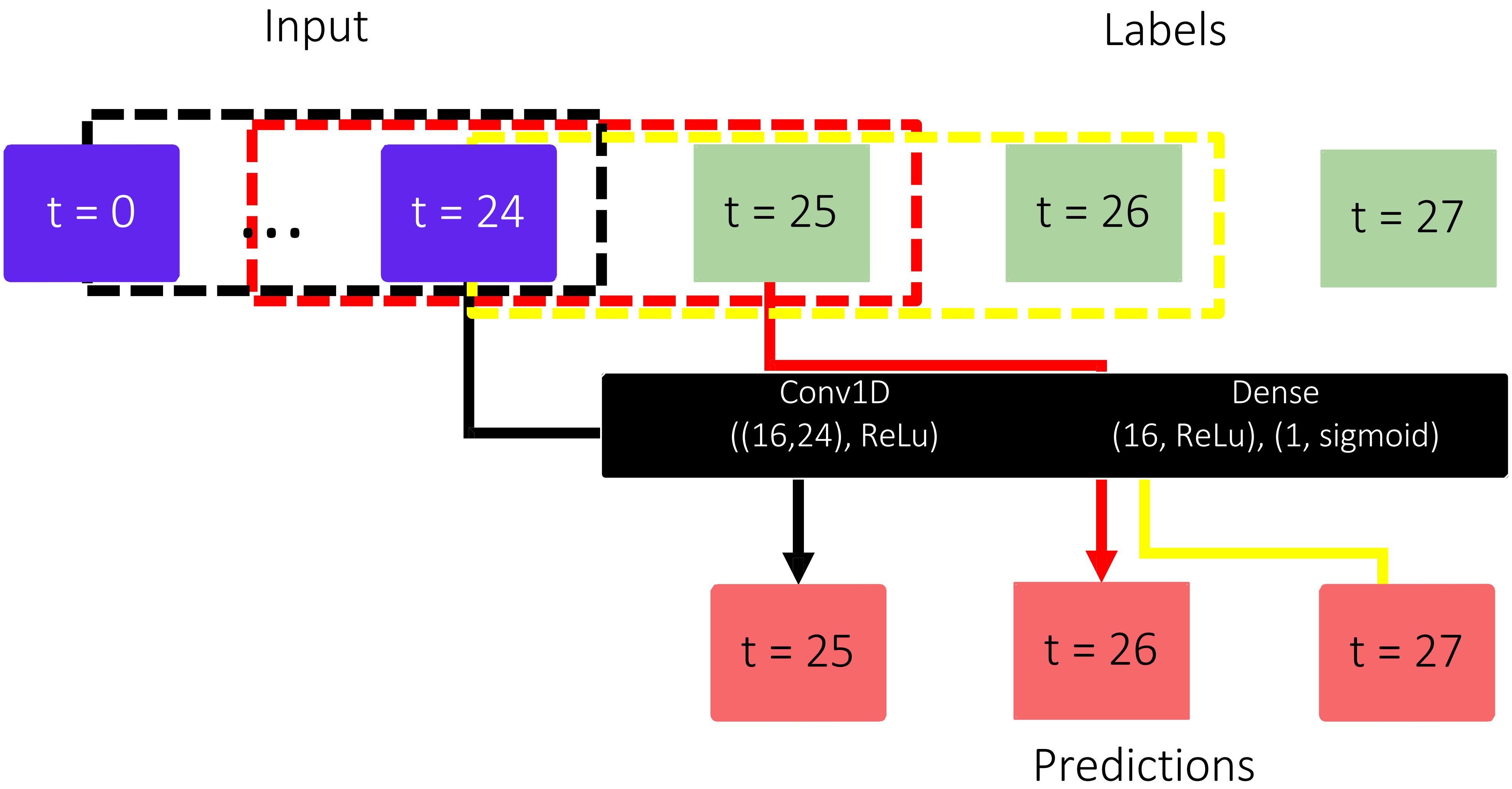
1 Day Models:

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

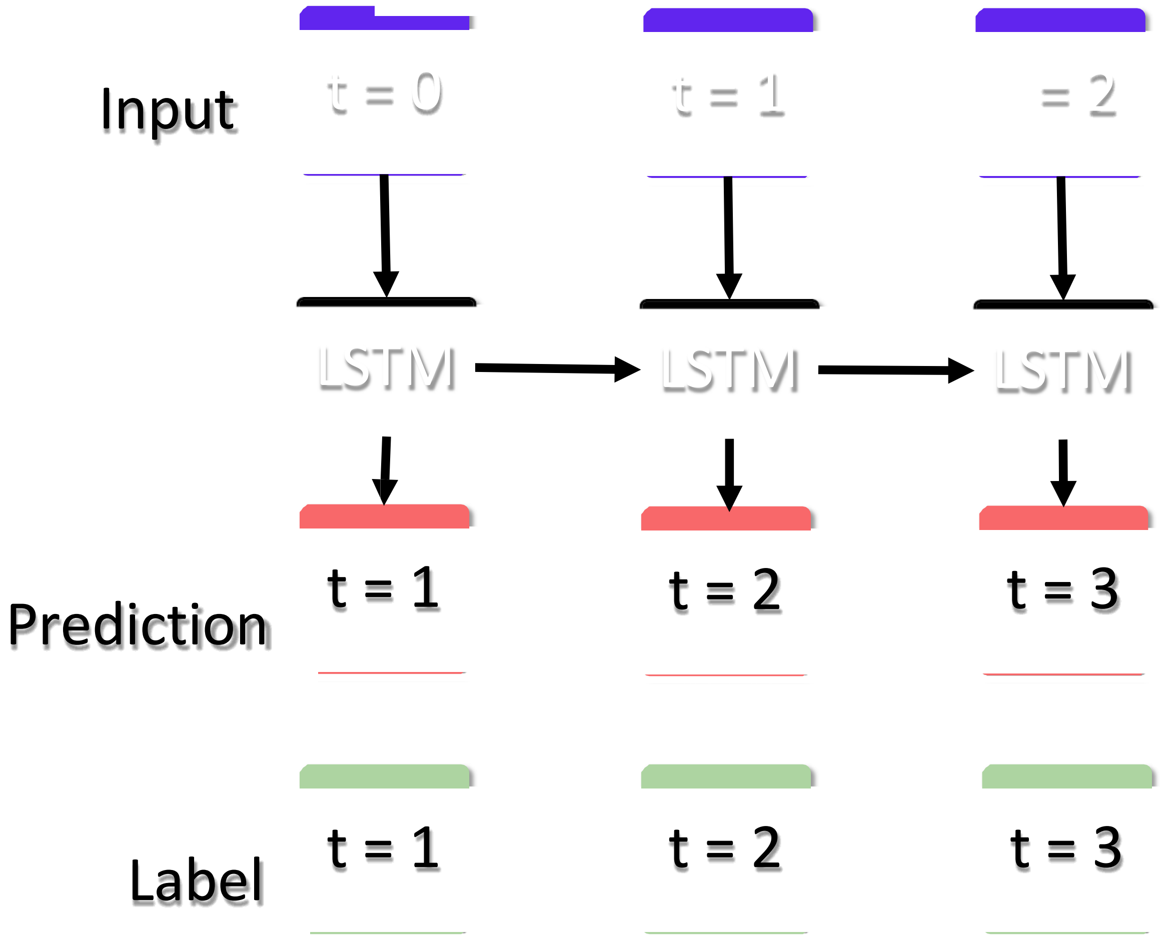
$$y_{t+1} = f(M(X, y, w)), M = \begin{bmatrix} X_t & y_t \\ X_{t-1} & y_{t-1} \\ \vdots & \vdots \\ X_{t-w} & y_{t-w} \end{bmatrix}$$



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

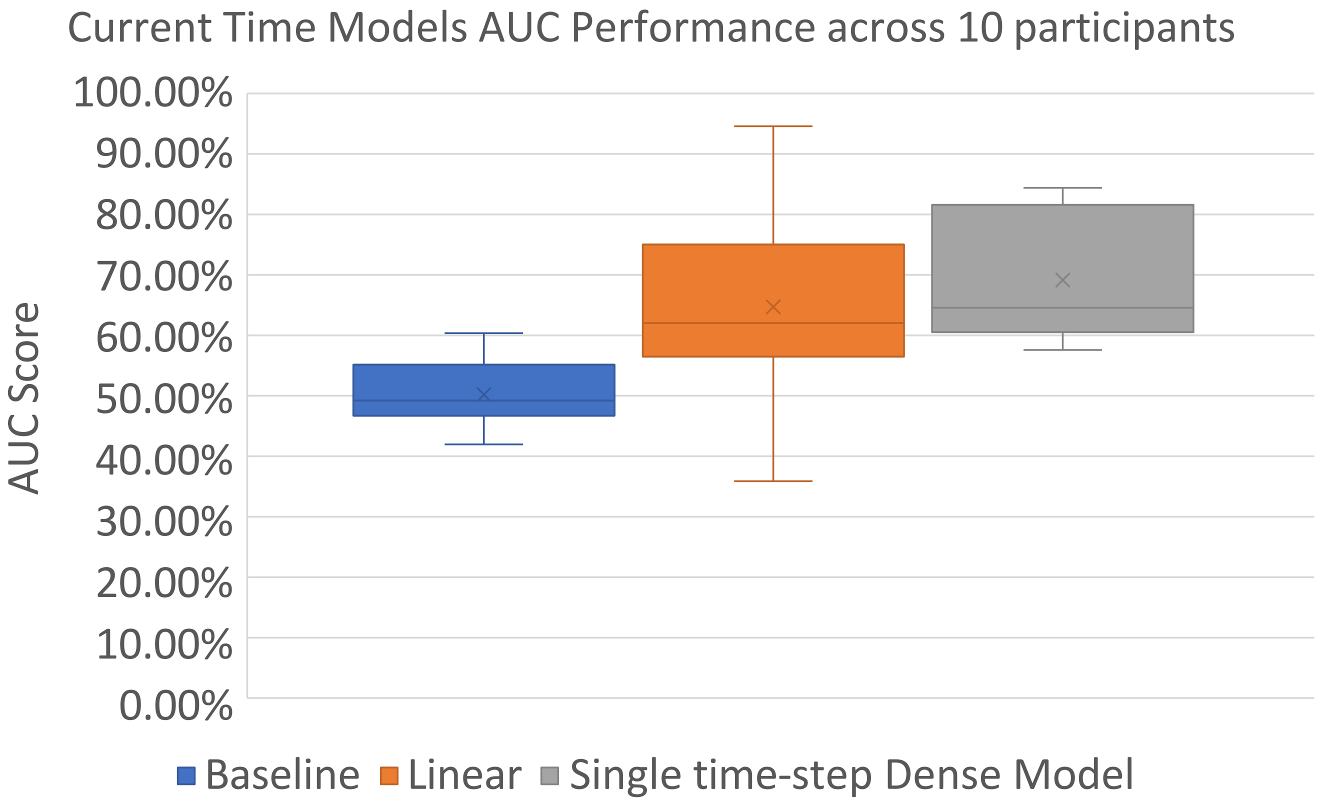
2 = Partial assistance

PDQ-8

42.50%
$(\sigma = 0.192)$

Higher score means worse QoL

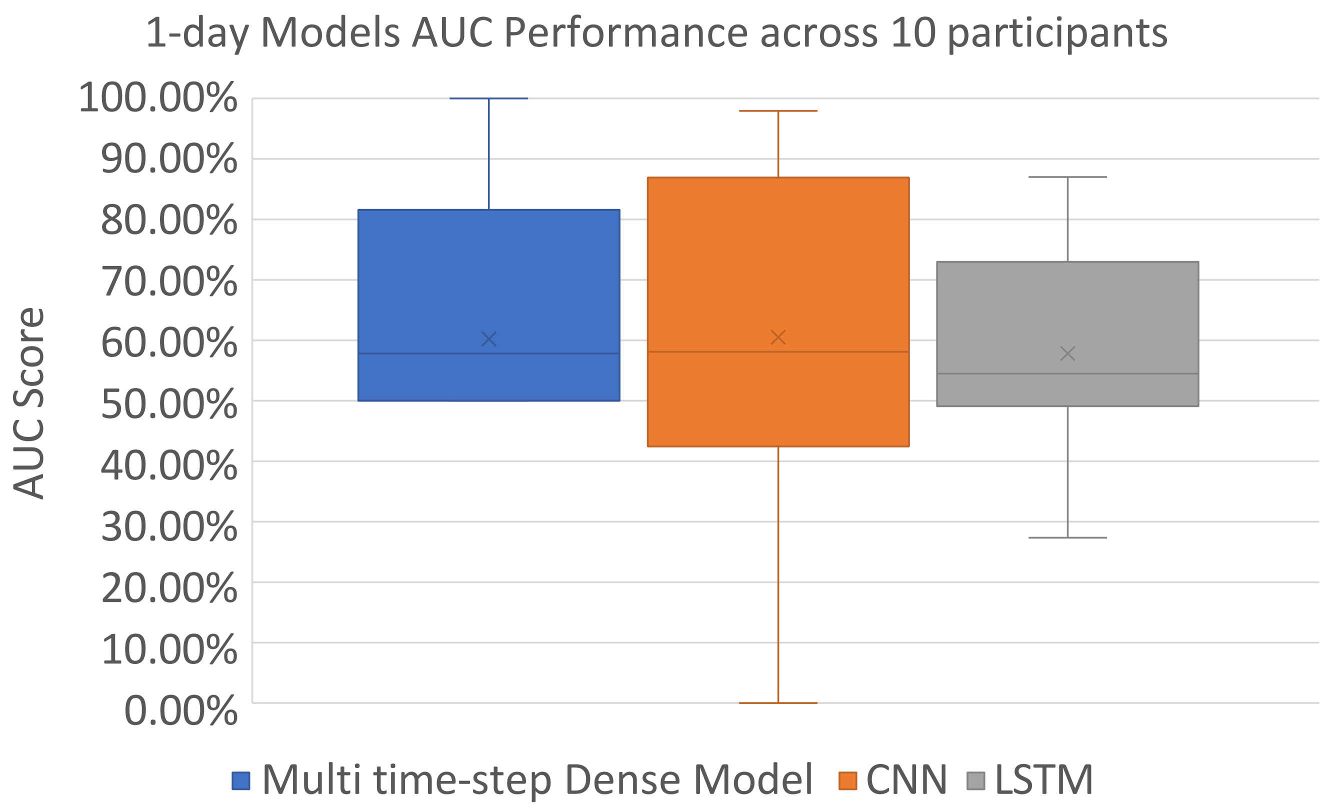
Can a wrist-worn fitness tracker datasets be used to forecast wearing-off in the next hour?



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%

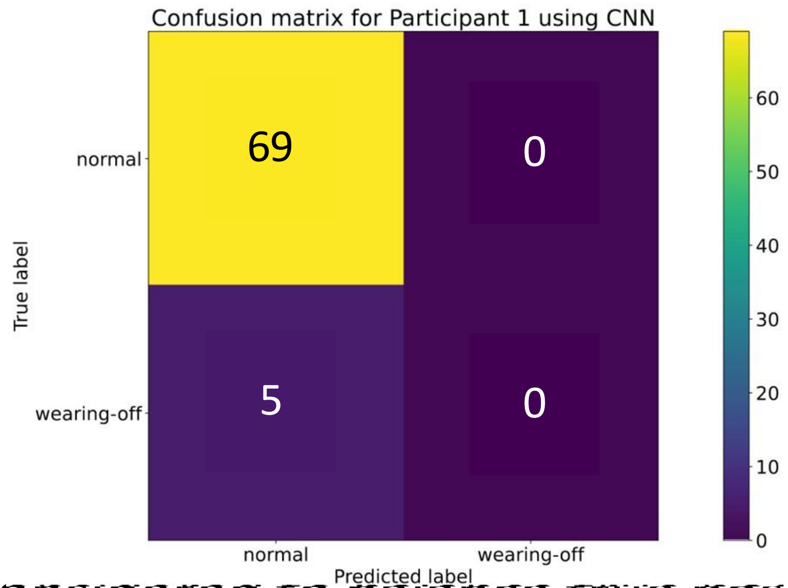
Which among the six deep learning architectures performed well in forecasting wearing-off in the next hour?

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

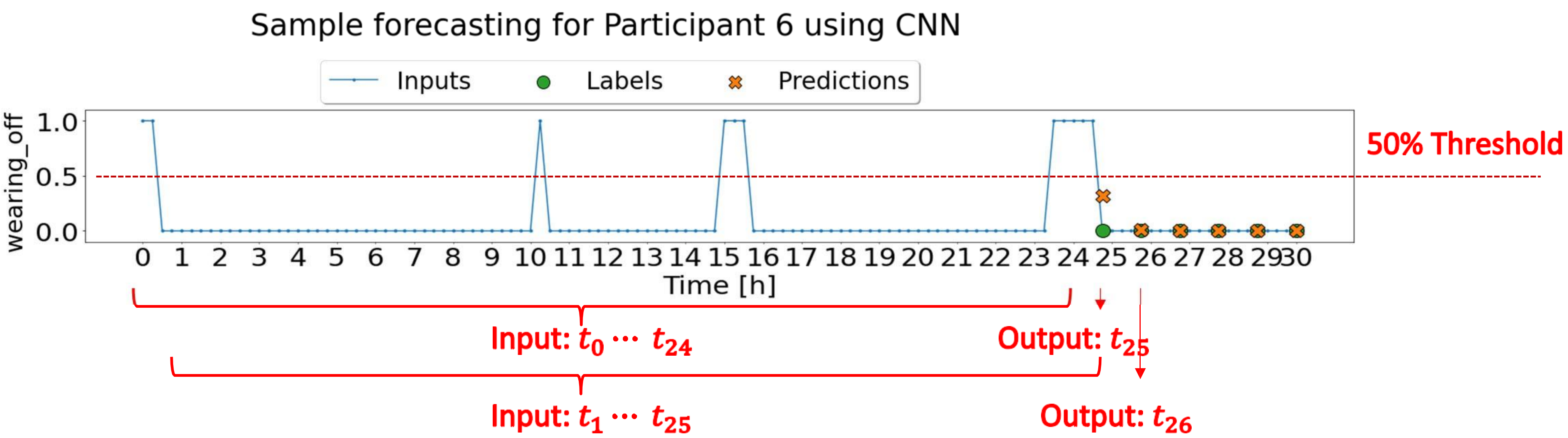
- Multi time-step Dense & CNN models still had a recall score lower than 50%
  - Low recall score = Missed wearing-off forecast



- 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



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.

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.

Sleep Features

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

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

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

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

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

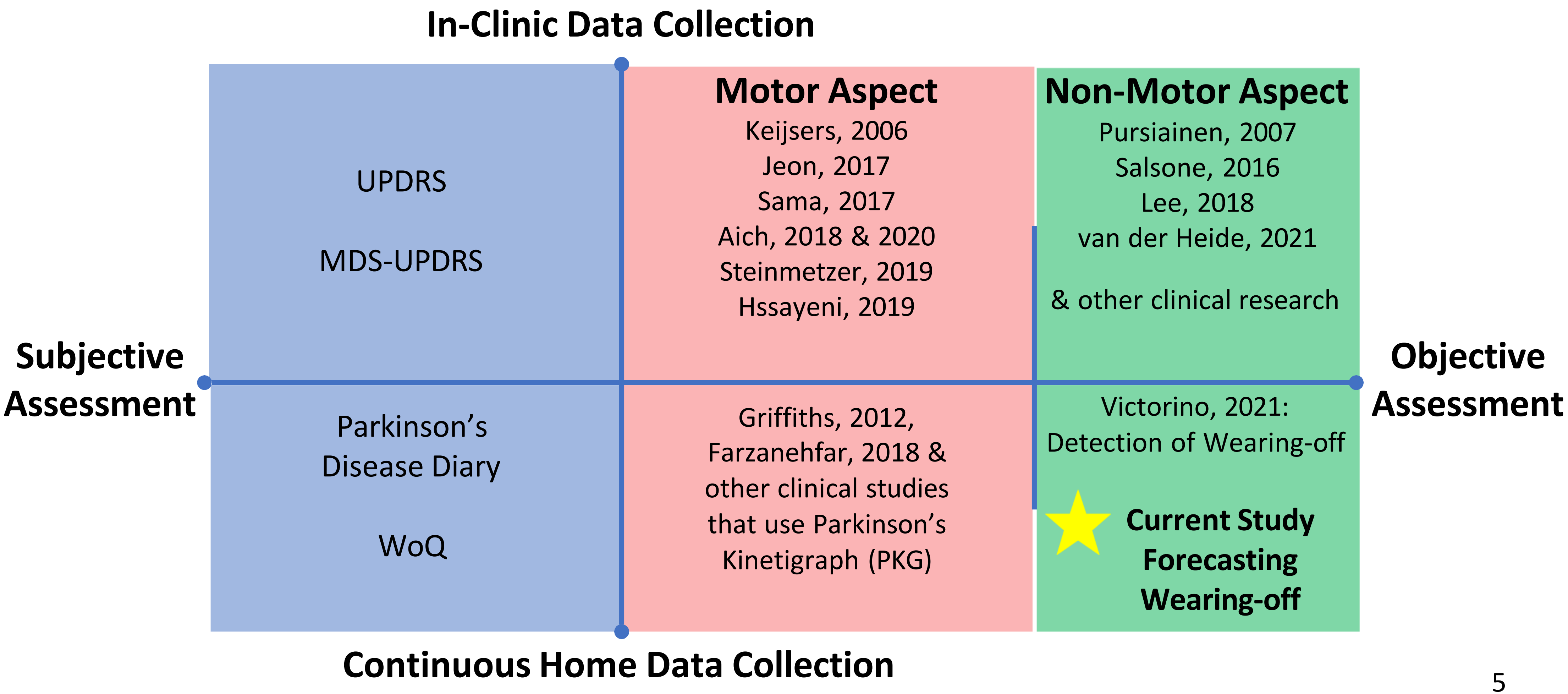
Which among the DL architectures performed well in forecasting wearing-off in the next hour?

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

Current Landscape in Detecting PD symptoms or wearing-off



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

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

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

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

Results and Discussion