Mind in Motion

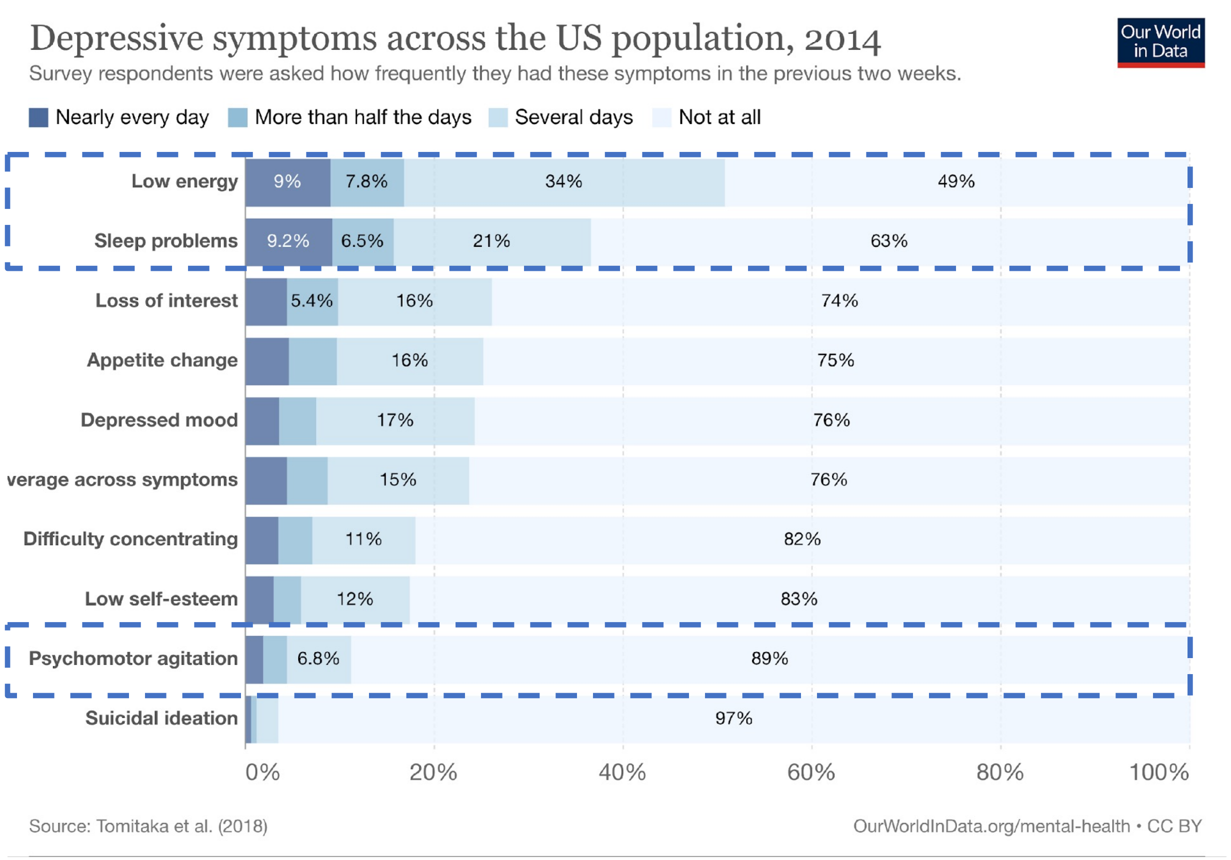
Machine Learning for Depression Classification using Motor Actigraphy Data

Final Presentation

Gary Kong, Julia Kauffman, Leo Le, Vishnu Paty, Melia Soque

Berkeley
UNIVERSITY OF CALIFORNIA

Motivation



RELATED TO MOTOR ACTIVITY

Depression

- World's leading cause of disability
- Associated with disrupted biological rhythms and changes in motor activity
- Subjective diagnosis (DSM-5 criteria)

Motor actigraphy data

- Non-invasive
- Available through wearables (e.g., Fitbit)
- Potentially objective method of diagnosing depression

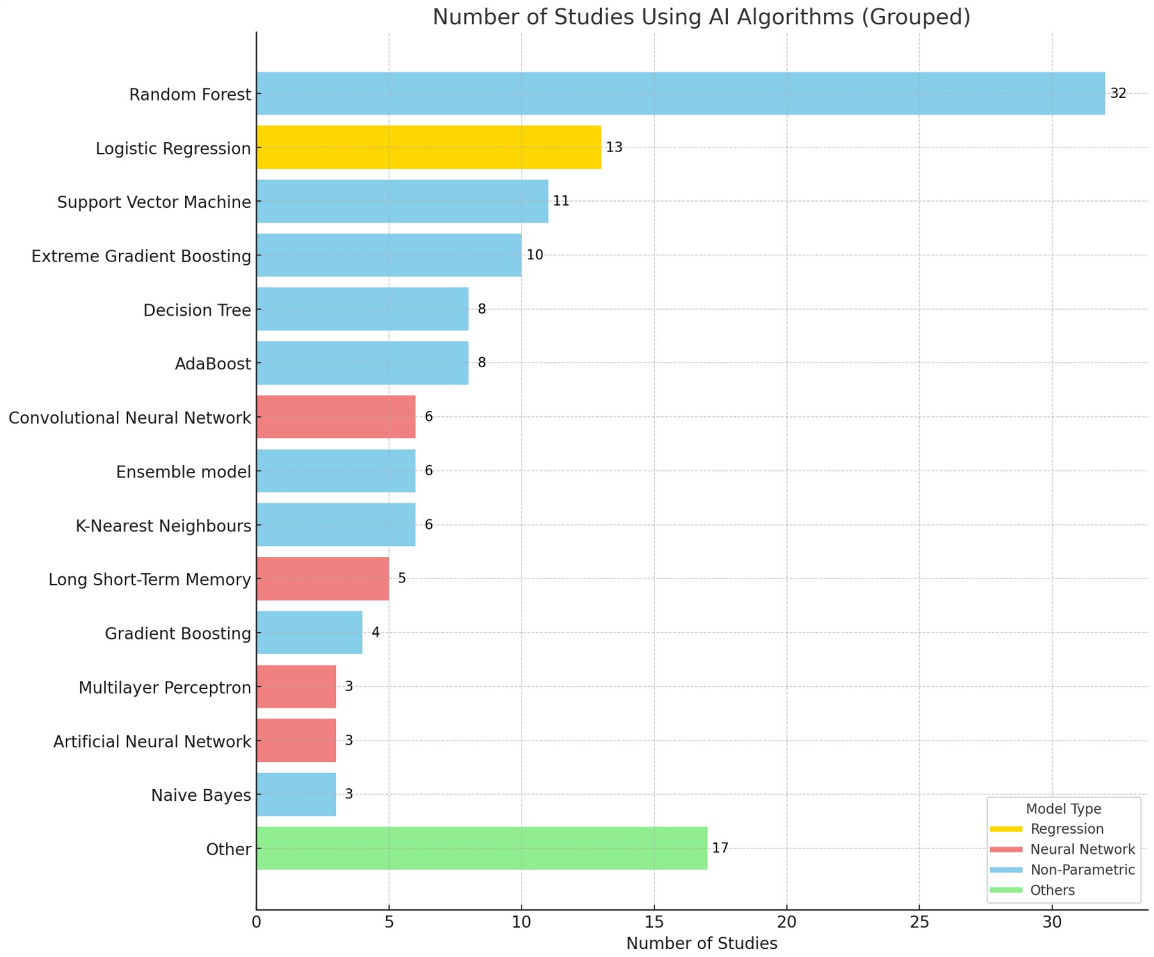
Research question

Can machine learning algorithms be used to accurately classify patients into depression status (normal, mild and moderate) using motor actigraphy data?

•Sources: UpToDate, WHO, Our World in Data

Past Studies

- **54 studies** used AI with wearables for depression
- **Non-parametric methods** were predominant
- **AI accuracy:** 70–89%
- **AdaBoost** excelled in performance, whereas logistic regression and decision trees were less effective



Berkeley
UNIVERSITY OF CALIFORNIA

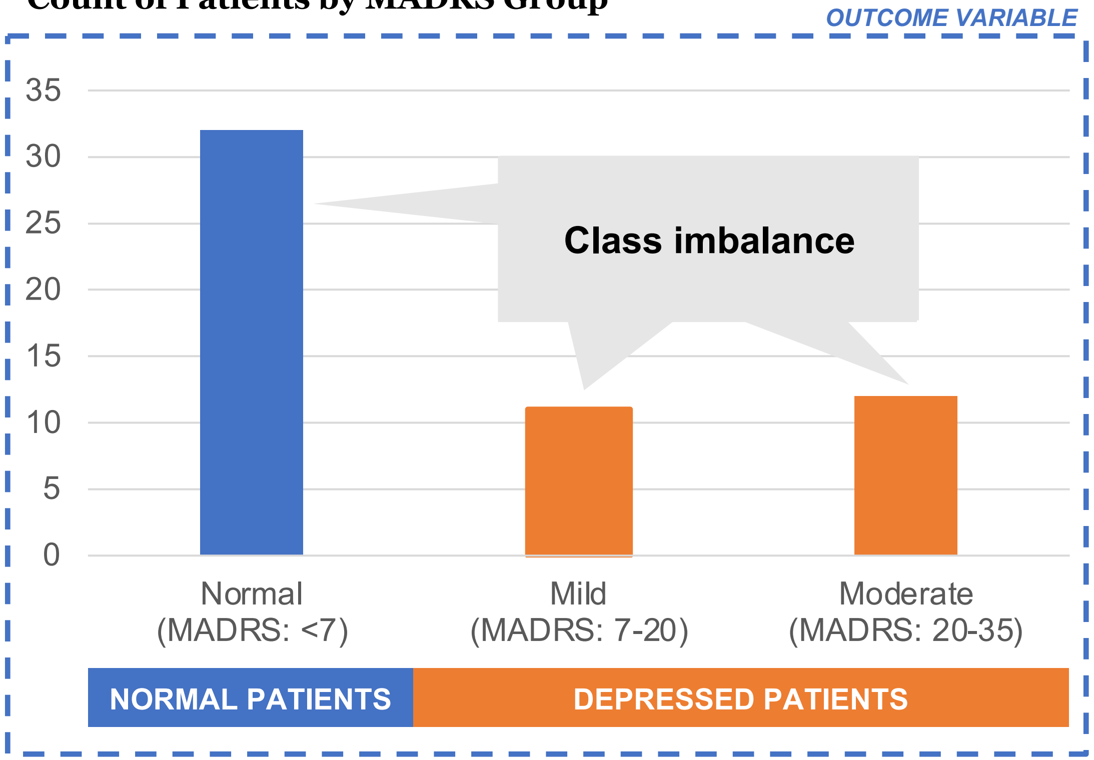
Source: Abd-Alrazaq, A., AlSaad, R., Shuweihdi, F. et al. (2023). Systematic review and meta-analysis of performance of wearable artificial intelligence in detecting and predicting depression. npj Digital Medicine, 6, 84. DOI: 10.1038/s41746-023-00828-5

Data Overview

Data Source

Depresjon: A Motor Activity Database of Depression Episodes in Unipolar and Bipolar Patients

Count of Patients by MADRS Group



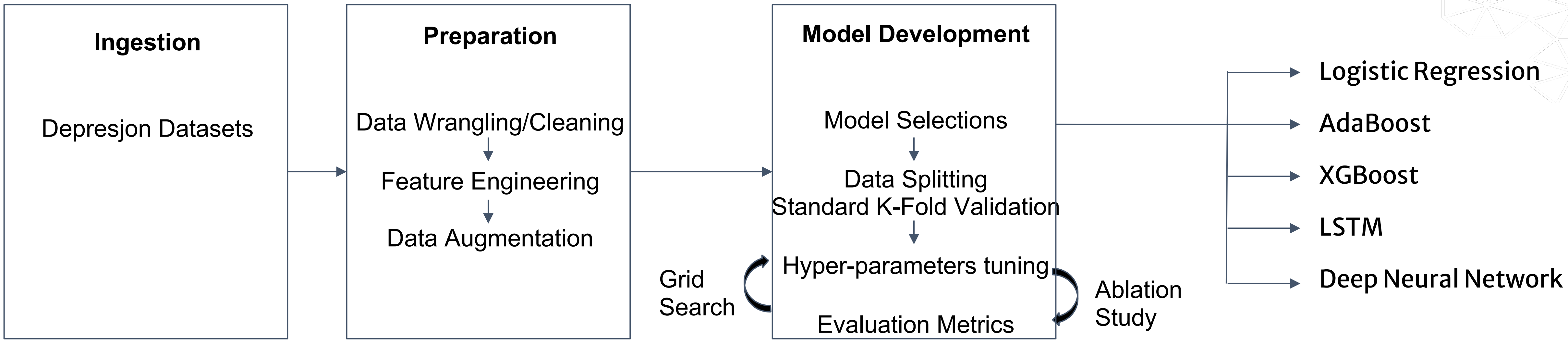
Summary Statistics

Variable	Values	Count	%
Gender	Female	30	55%
	Male	25	45%
Affective Type	Normal	32	58%
	Unipolar	15	27%
	Bipolar	8	15%
Days of Observation	Mean	13 days	
Age	Mean	40 years	
Education	Mean	11 years	

Berkeley
UNIVERSITY OF CALIFORNIA

Source: Enrique Garcia-Ceja, Michael Riegler, Petter Jakobsen, Jim Tørresen, Tine Nordgreen, Ketil J. Oedegaard, Ole Bernt Fasmer. "Depresjon: A Motor Activity Database of Depression Episodes in Unipolar and Bipolar Patients." Proceedings of the 9th ACM on Multimedia Systems Conference (MMSys'18), 2018. DOI: 10.1145/3204949.3208125

Approach



Baseline Accuracy : 33.33%

Feature Engineering

Feature Extraction

57 Features

- Activity distribution metrics
- Activity proportions by time of day / week
- Active / inactive bout metrics
- Cosine parameters

Feature Selection

Greedy feature selection (correlation-based) score

+

Decision-tree feature importance score

=

Composite feature importance score

↓

Rank of feature importance

Top 30 Features

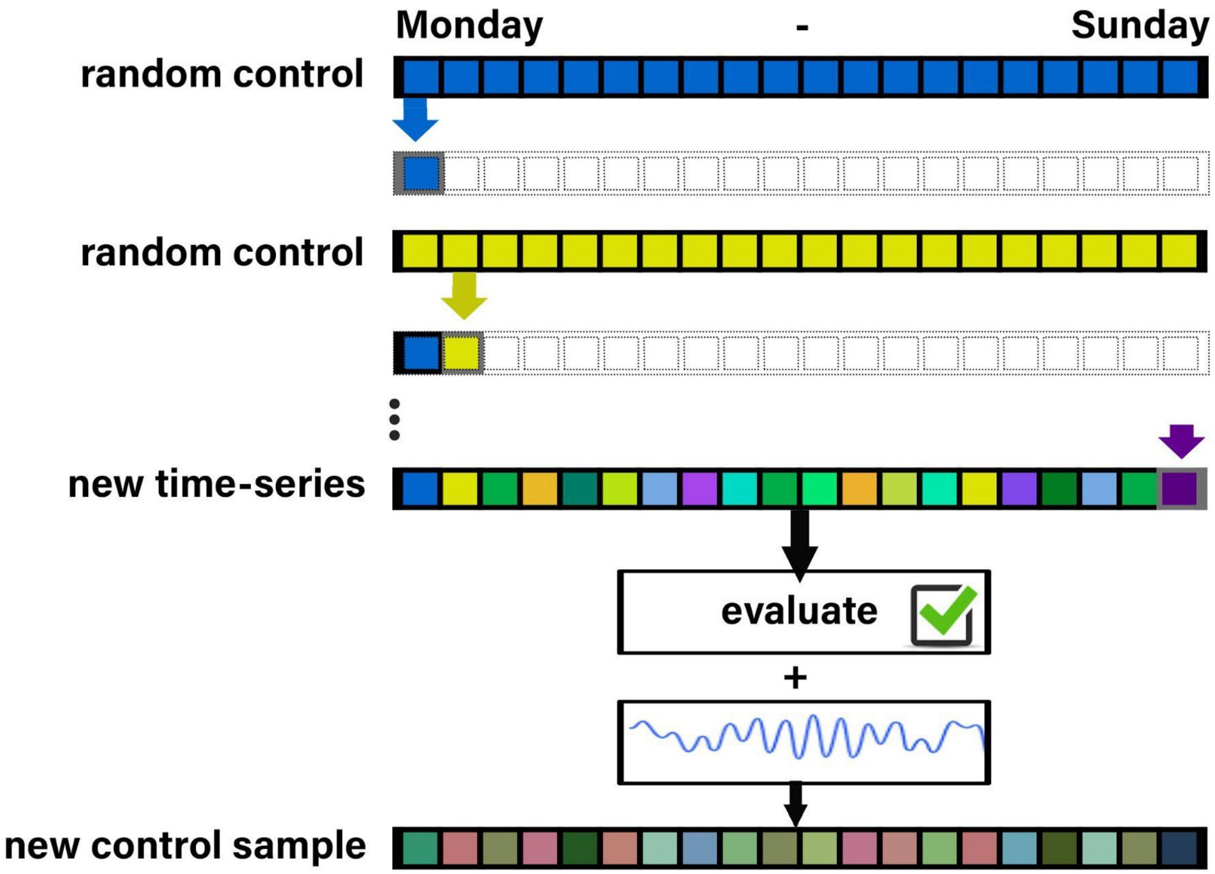
Feature Correlation

	Normal	Mild	Moderate
Activity Entropy (by Hour)	0.42	-0.25	-0.26
Activity Entropy (by Minute)	0.54	-0.33	-0.33
Activity Intraday Variability	0.28	-0.26	-0.09
Activity Kurtosis (by Day)	-0.31	0.12	0.26
Activity Kurtosis (by Hour)	-0.48	0.33	0.26
Activity Kurtosis (by Minute)	-0.24	0.05	0.23
Activity Mean (Overall)	0.46	-0.32	-0.25
Activity Mean (Weekdays, 00:00 - 06:00)	0.33	-0.07	-0.32
Activity Mean (Weekdays, 06:00 - 12:00)	0.44	-0.16	-0.36
Activity Mean (Weekends, 12:00 – 18:00)	0.18	-0.12	-0.10
Activity Mean (Weekends, 12:00 – 18:00)	0.38	-0.39	-0.09
Activity Mean (Weekends, 06:00 – 12:00)	0.45	-0.20	-0.34
% Activity (00:00 – 06:00)	-0.07	0.30	-0.20
% Activity (12:00 – 18:00)	-0.14	-0.18	0.34
% Activity (18:00 – 00:00)	0.03	-0.39	0.33
% Activity (06:00 – 12:00)	0.15	0.17	-0.33
% Activity (Weekdays)	0.17	0.00	-0.21
Activity Skewness (by Day)	-0.45	0.28	0.26
Activity Skewness (by Hour)	-0.48	0.33	0.27
Activity Skewness (by Minute)	-0.33	0.21	0.19
Activity Standard Deviation (by Hour)	0.28	-0.17	-0.18
Active Bout Duration Standard Kurtosis	0.30	-0.14	-0.22
Active Bout Duration Standard Deviation	0.32	-0.20	-0.19
Number of Active Bouts	0.45	-0.34	-0.22
Inactive Bout Duration Coef. of Variation	0.33	-0.14	-0.26
Inactive Bout Duration Entropy	-0.21	0.10	0.16
Inactive Bout Duration Mean	-0.41	0.30	0.21
Inactive Bout Duration Standard Deviation	-0.05	0.09	-0.03
Cosine Amplitude (by Day of Week)	-0.19	0.25	-0.02
Cosine Phase (by Minute of Day)	-0.14	0.23	-0.05

Color scale: 0.4, 0.2, 0.0, −0.2, −0.4

Berkeley
UNIVERSITY OF CALIFORNIA

Data Augmentation via Cross-Subject Data Fusion



Process Overview

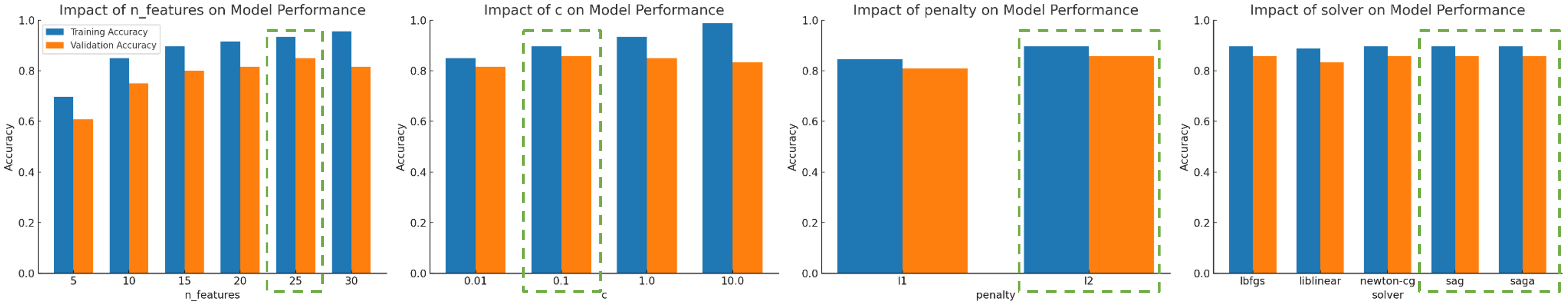
- Randomly select sample
- Time-segment slicing
- Repeat – incorporating slices from different random subjects
- Assess the quality
- Add noise

	Control	Mild	Moderate
Before	23	11	12
After	40	40	40

Logistic Regression

Best hyperparameter(s)

Ablation analysis suggests optimal hyperparameters are: n_features=25, c=0.1, penalty=L2, solver=sag/saga



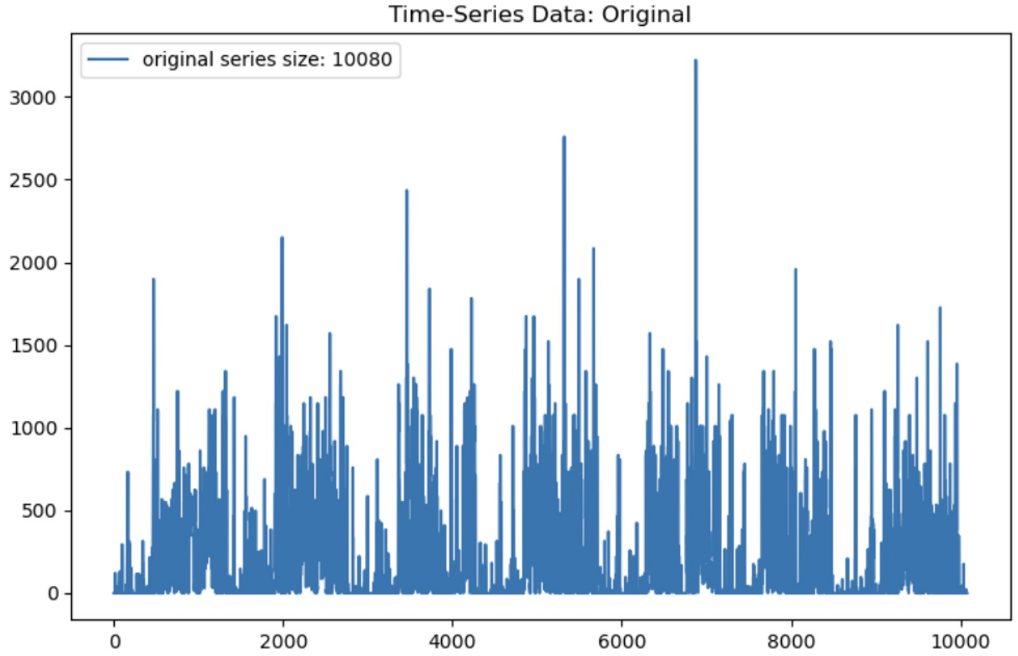
Grid search confirms this

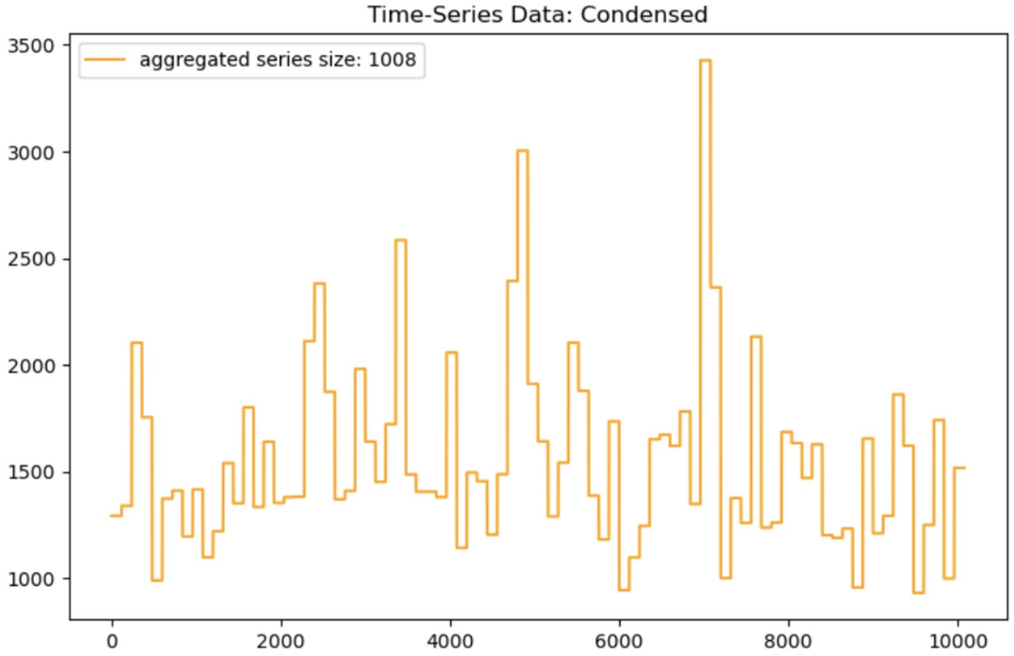
Grid Search Results (Sorted / Ranked by Mean Test Score)

rank	n_features	param_C	param_penalty	param_solver	mean_train_score	mean_test_score	mean_variance
1	25	0.1	l2	sag	0.893750	0.858333	0.035417
2	25	0.1	l2	saga	0.893750	0.858333	0.035417
3	25	0.1	l2	newton-cg	0.895833	0.858333	0.037500
4	25	0.1	l2	lbfgs	0.895833	0.858333	0.037500
5	25	1	l2	newton-cg	0.933333	0.850000	0.083333

LSTM Recurrent Neural Network

Using only time-series readings





Berkeley
UNIVERSITY OF CALIFORNIA

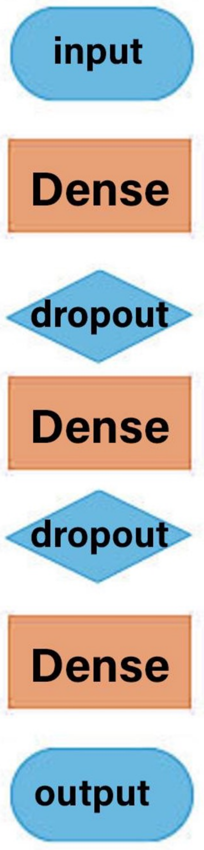
Note: Refer to the notebook at https://github.com/UC-Berkeley-I-School/mids-207-final-project-summer23-Kong-Kauffmann-Soque-Le-Paty/blob/main/notebooks/w207_2_models_and_evaluation.ipynb for hyperparameter tuning and model results

Dense Neural Network

Using selected extracted features

Model performs on par with our best models:

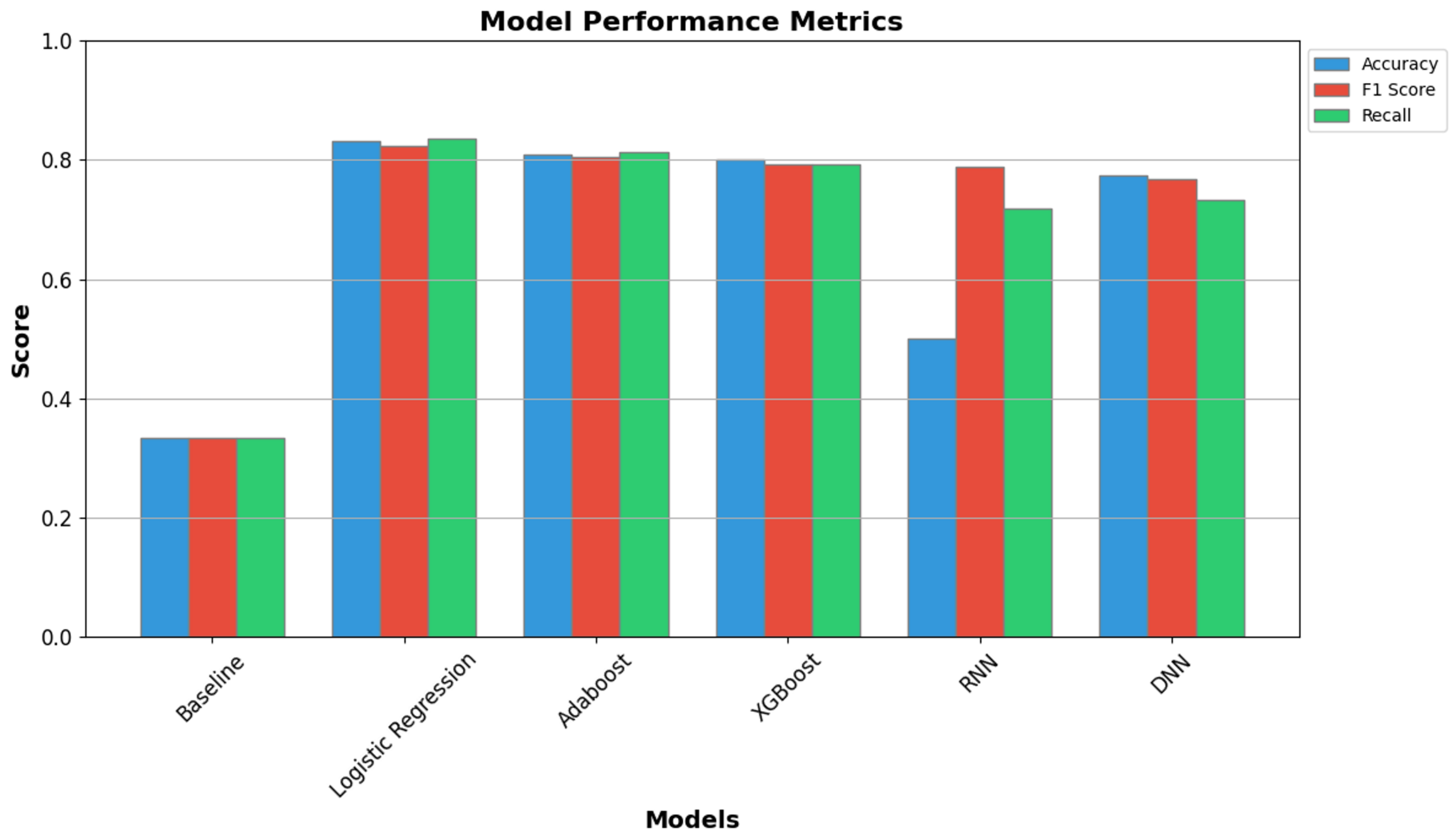
- Accuracy 82.5%
- Precision 85.1%
- Recall 80.0%
- F1-Score 82.4%



Berkeley
UNIVERSITY OF CALIFORNIA

Note: Refer to the notebook at https://github.com/UC-Berkeley-I-School/mids-207-final-project-summer23-Kong-Kauffmann-Soque-Le-Paty/blob/main/notebooks/w207_2_models_and_evaluation.ipynb for hyperparameter tuning and model results

Evaluation

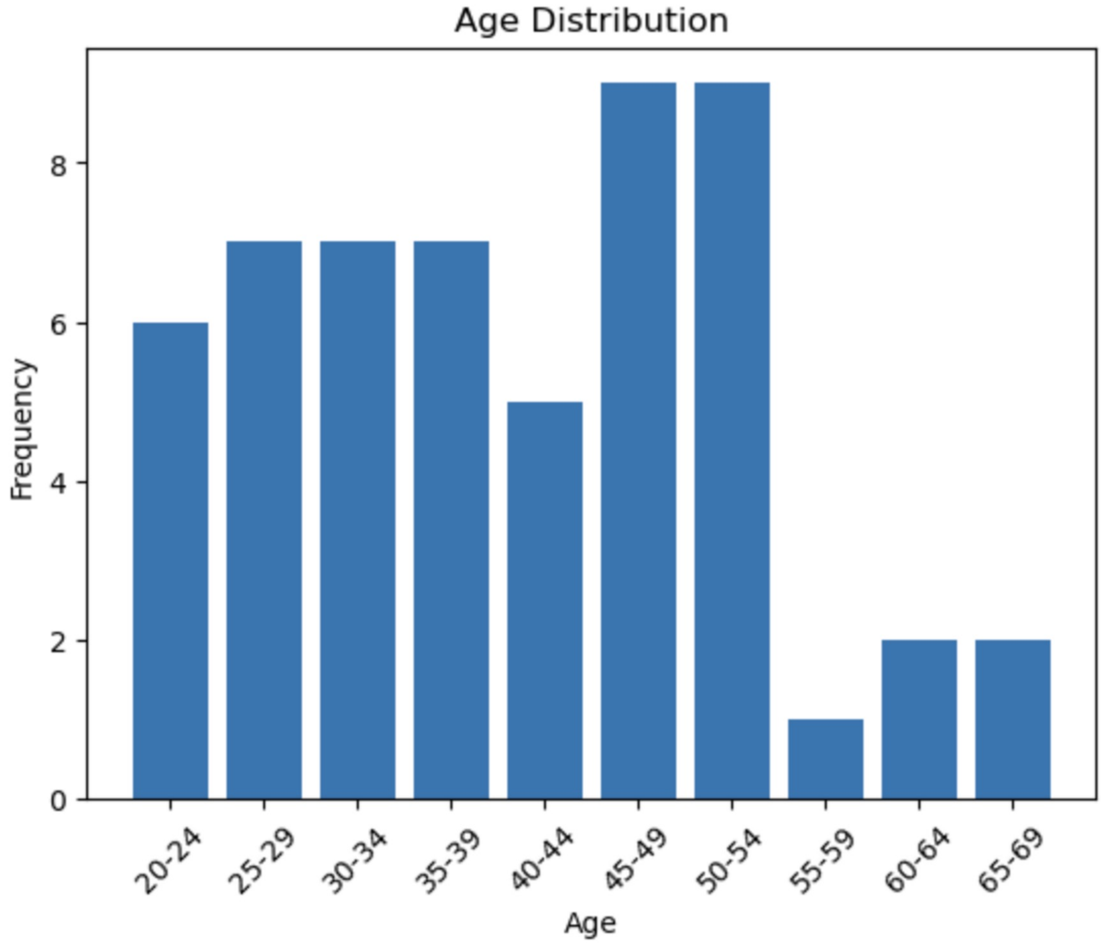


- Statistical tests:
  - ANOVA
  - Post-hoc pairwise comparisons with Bonferroni correction
- Logistic regression significantly outperforms other models
  - 85.83% accuracy
  - 85.74% F1 score
  - 87.8% recall
  - Lowest variance
- Sometimes the simplest model is best!

Berkeley
UNIVERSITY OF CALIFORNIA

Limitations

- Small Sample Size
- Unrepresentative Control Group
- Unbalanced Age Group
- Time Period Sampling
- MADRS Score Variability

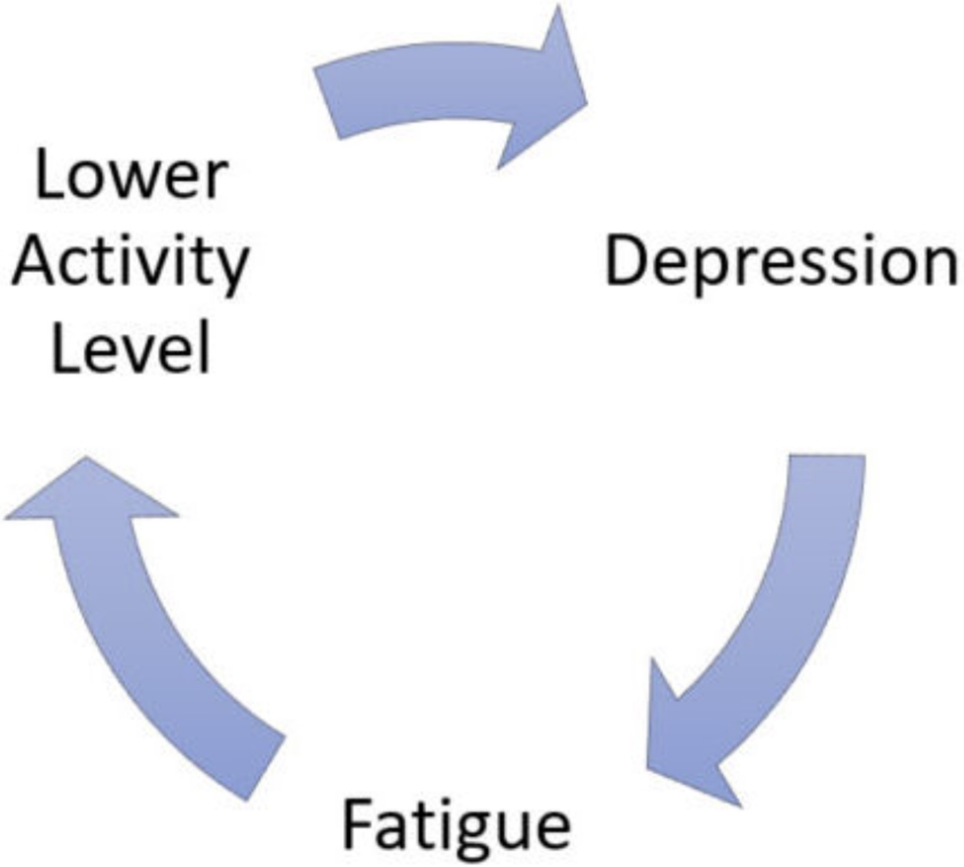


Biases

- Demographic Balance
- Urban vs. Rural Disparity
- Accessibility Generalization
- Time Period Sampling
- Observer Effect and Compliance

Berkeley
UNIVERSITY OF CALIFORNIA

Conclusions



Next Steps:

- Expand the Dataset
- Decide on model: DNN?
- Further fine-tune parameters
- Collaborate with mental health professionals
- Address ethical considerations
- Develop user-friendly tools

Contributions

	Gary Kong	Julia Kauffman	Leo Le	Vishnu Paty	Melia Soque
Literature Search	X	X	X	X	X
Data Wrangling and Cleaning	X				
Feature Extraction and Feature Selection	X				
Data Augmentation		X			
Data Splitting and Cross-validation		X	X	X	X
Model Development and Hyperparameter Tuning	Logistic regression	LTSM Deep neural network	AdaBoost (Lead) XGBoost (Lead)	AdaBoost (Support) XGBoost (Support)	Logistic regression
Model Evaluation				X	

Berkeley
UNIVERSITY OF CALIFORNIA