Mind in Motion

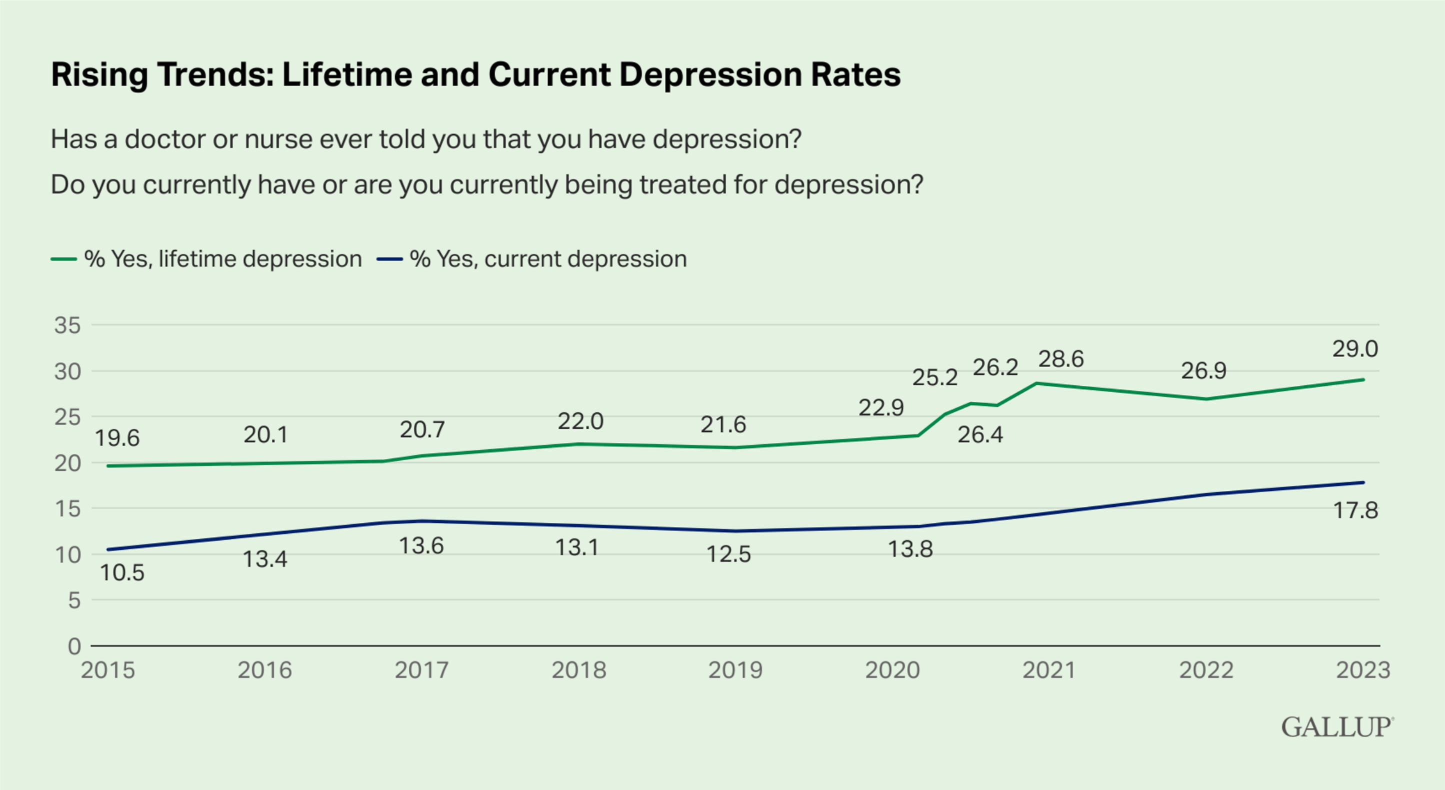
Machine Learning for Depression Classification using Motor Actigraphy Data

Baseline Presentation

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

Berkeley
UNIVERSITY OF CALIFORNIA

Our study aims to use machine learning to classify patients into depression status using actigraphy data



Depression

- 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
- 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?

Berkeley
UNIVERSITY OF CALIFORNIA

Sources: Gallup (2023)

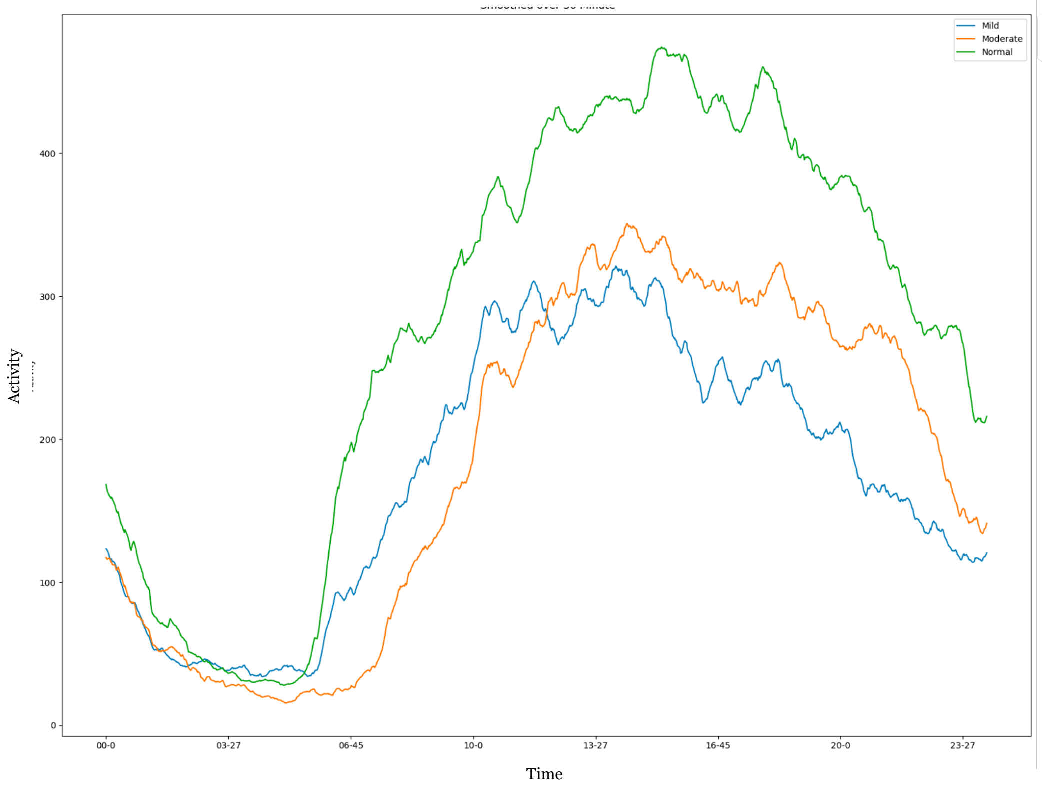
The Depresjon dataset contains actigraph data for controls and patients with mild/moderate depression

Summary Statistics of Depresjon Dataset

Variable	Values	Count	%
MADRS when measurement started (MADRS1)	Normal (<7)	32	58%
	Mild (7-20)	7	29%
	Moderate (20-35)	16	13%
MADRS when measurement stopped (MADRS2)	Normal (<7)	32	58%
	Mild (7-20)	11	22%
	Moderate (20-35)	12	20%
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	

Label

Average Daily Activity (30-Minute Rolling Average)



Berkeley
UNIVERSITY OF CALIFORNIA

We extracted features that capture differences in activity patterns across depression classes

Correlation of Features with Depression Classes

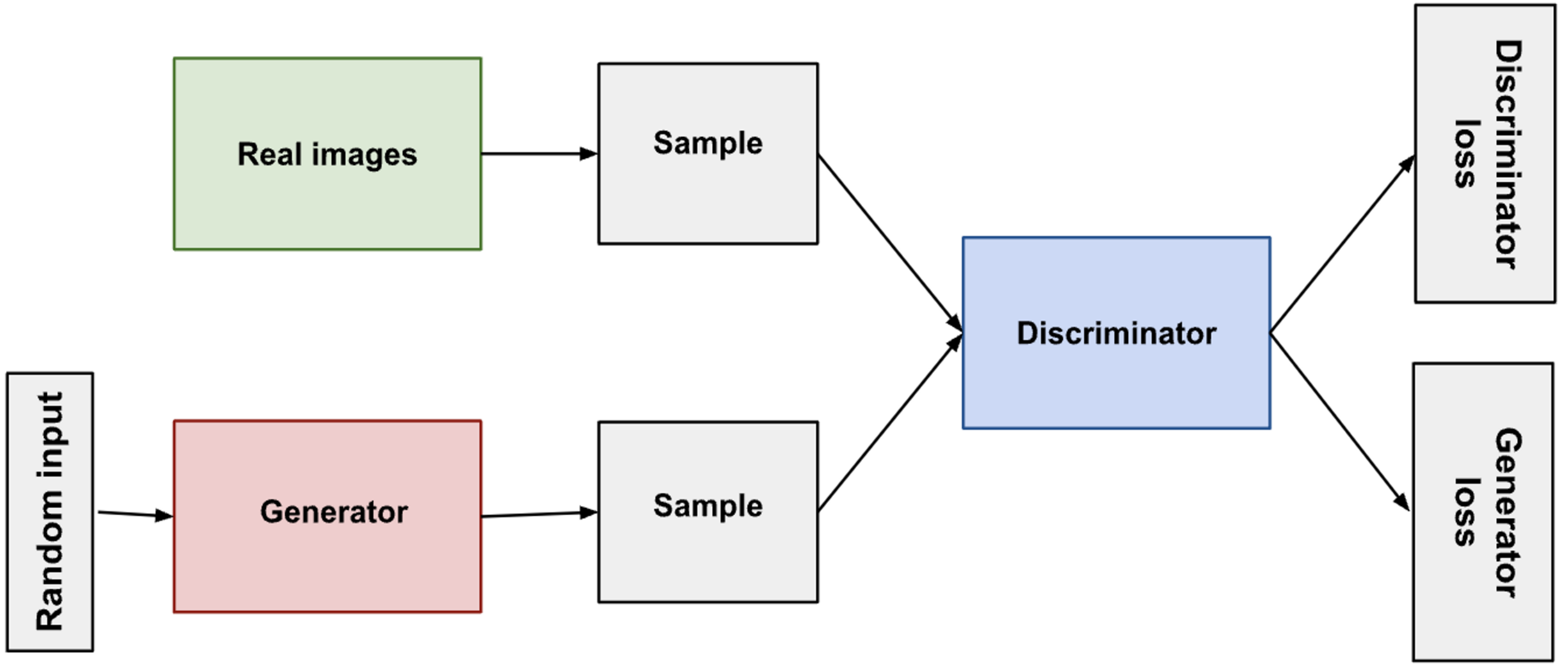
		Normal	Mild	Moderate
Statistical metrics capturing distribution of activity	Activity Intraday Variability	0.33	-0.25	-0.15
	Activity Mean	0.52	-0.35	-0.28
	Activity Skewness (Hourly)	-0.47	0.30	0.27
	Activity Entropy (Hourly)	0.39	-0.20	-0.27
Proportion of activity occurring in different periods of each day	% Activity in 00:00-06:00	-0.03	0.26	-0.21
	% Activity in 06:00-12:00	0.13	0.09	-0.24
	% Activity in 12:00-18:00	-0.09	-0.14	0.25
	% Activity in 18:00-00:00	0.01	-0.31	0.28
Statistical metrics capturing count and distribution of bouts of activity	Activity Bout Count	0.46	-0.25	-0.31
	Activity Bout Duration Mean	0.39	-0.32	-0.15
	Activity Bout Duration Standard Deviation	0.17	-0.07	-0.13
	Activity Bout Duration Coefficient of Variation	0.01	0.12	-0.12
	Activity Bout Duration Skewness	0.08	0.02	-0.11
	Activity Bout Duration Entropy	0.22	-0.25	-0.03
Statistical metrics capturing distribution of bouts of inactivity	Inactivity Bout Duration Mean	-0.42	0.19	0.31
	Inactivity Bout Duration Coefficient of Variation	0.42	-0.14	-0.37
	Inactivity Bout Duration Skewness	0.21	0.05	-0.30
	Inactivity Bout Duration Entropy	-0.20	0.09	0.15

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

To account for small Ns and address class imbalance, we will use GANs to augment our data

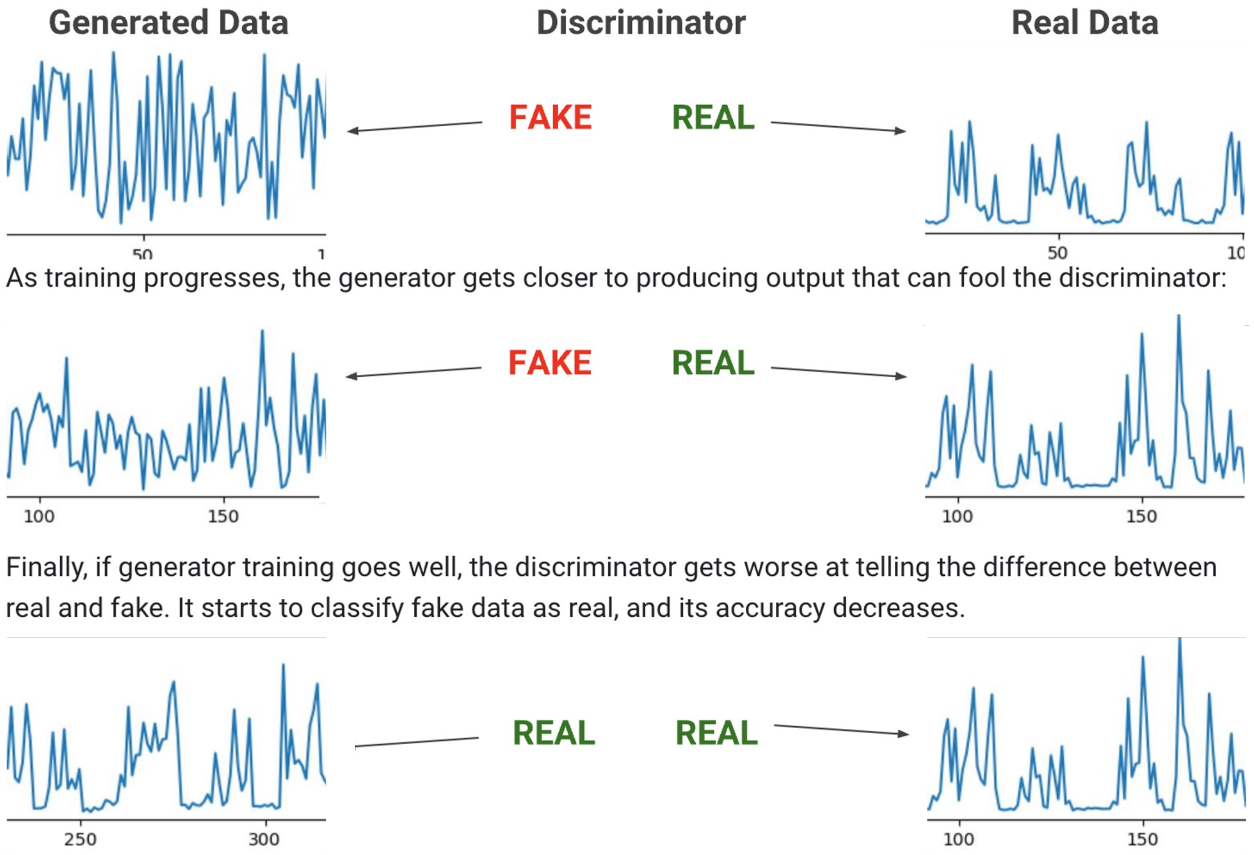
Explored Techniques:

- SMOTE (Synthetic Minority Over-sampling Technique)
- Cross-Subject Data Fusion
- Generative Adversarial Networks (GANs)

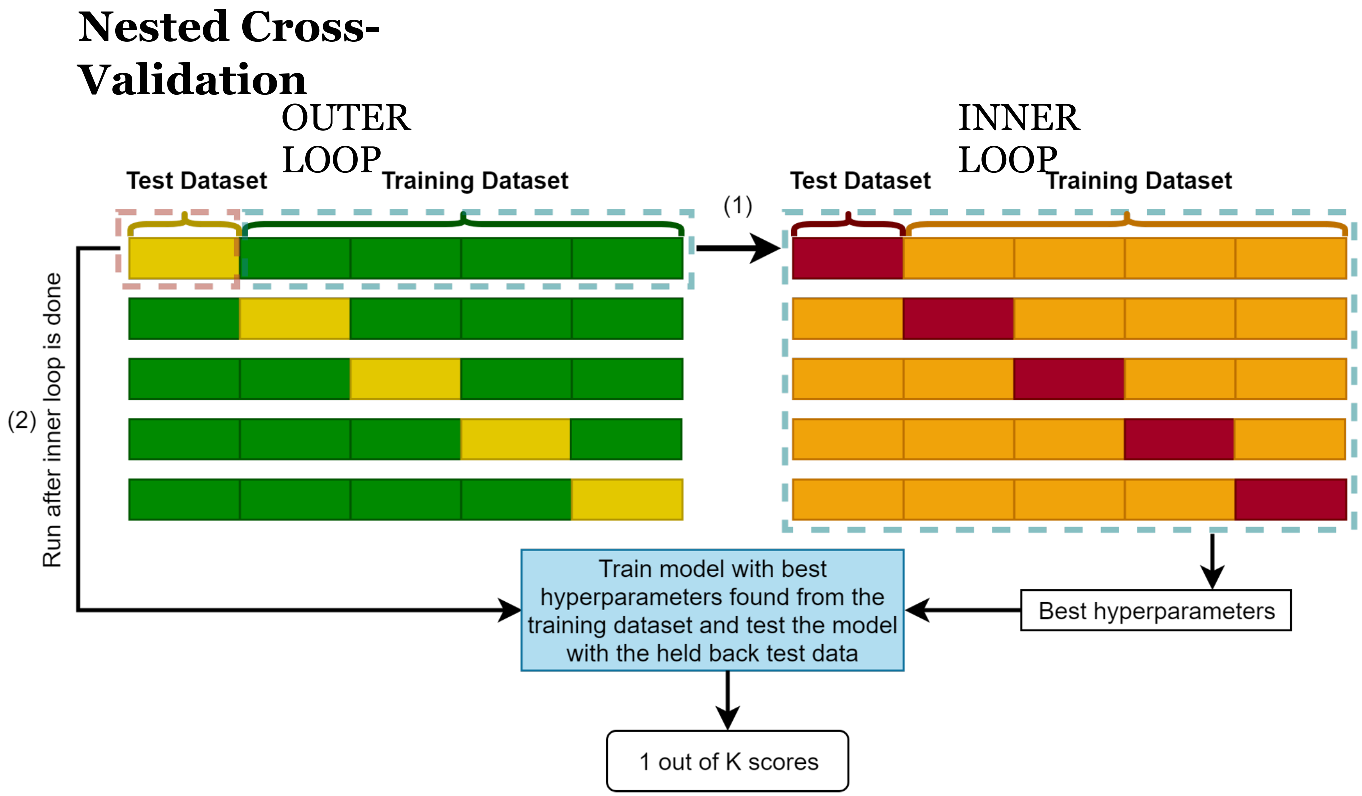


Berkeley
UNIVERSITY OF CALIFORNIA

Adversarial training involves a “game of chess” - the generator creates data and discriminator ‘critiques’ it

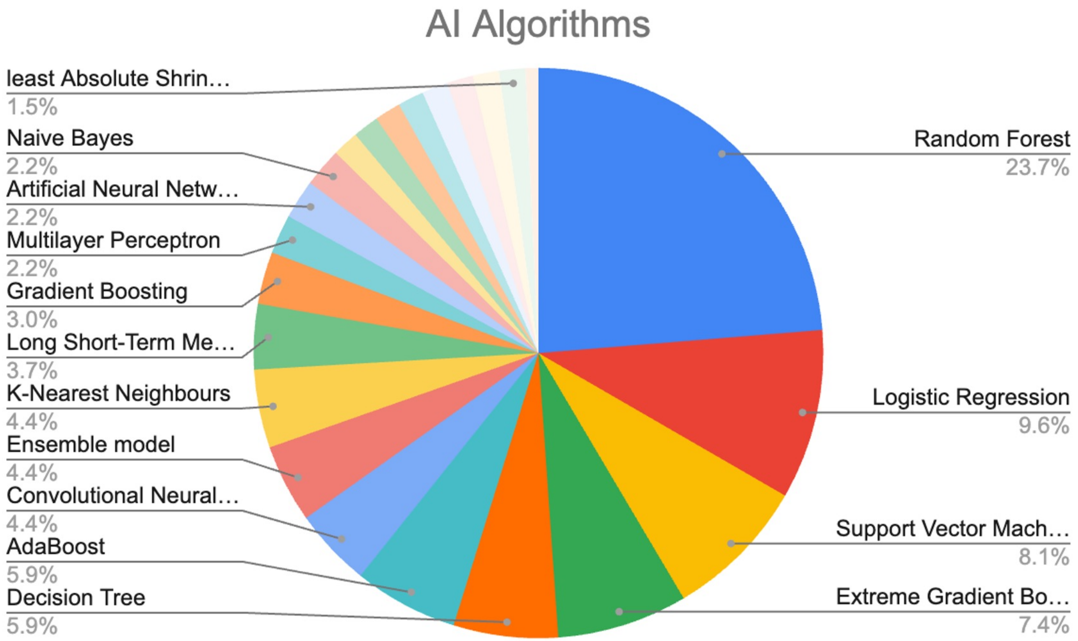


We will use nested cross-validation to validate our model



- All outer loops provides a robust estimate of a model's performance
- Inner loop prevents information leakage from the test set
- K-folds 3, 5, 10
- K iterations provides a more reliable estimate of a model's generalization ability

In other studies RF and Logistic Regression are frequently used; accuracy benchmarks are c.70-90%



Evaluation metric	Average top accuracy studies	Average bottom accuracy studies
Accuracy	89%	70%
Sensitivity	87%	61%
Specificity	93%	73%
RMSE	4.55	3.76

Berkeley
UNIVERSITY OF CALIFORNIA

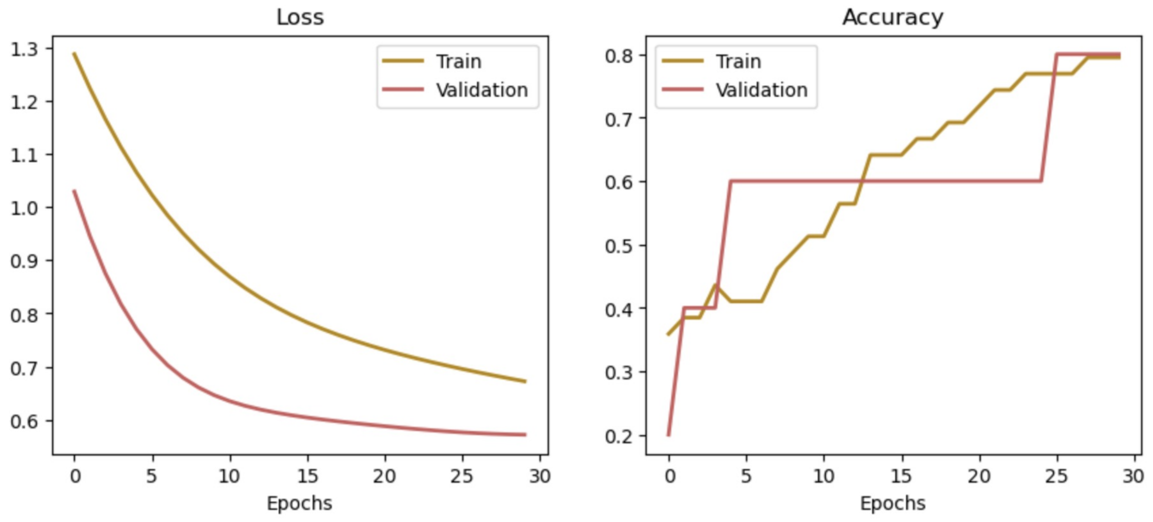
Source: Abd-Alrazaq, A., AlSaad, R., Shuweihdi, F., Ahmed, A., Aziz, S., & Sheikh, J. (2023). Systematic Review and meta-analysis of performance of wearable artificial intelligence in detecting and predicting depression. Npj Digital Medicine, 6(1). https://doi.org/10.1038/s41746-023-00828-5

Our baseline model will use logistic regression; we also plan to use random forest and boosted models

Features: Motor Actigraphy Data with 19 features

Implementation: Logistics Regression (3 classes classification)
->Complex model: Random Forest and Boosted Tree Models (Adaboost and XGBoost)

Learning rate : 0.01
Epochs : 30
Batch Size : 64
Validation Split : 0.1



We will use accuracy, F1 score, and ROC-AUC alongside statistical tests to evaluate performance

Classifier Metrics			**Statistical Tests**	
Accuracy	**F1 Score**	**ROC Curve + Area Under the Curve (AUC)**	**Omnibus Tests**	**Post-hoc Tests with Bonferonni correction**
Measures the fraction of correct predictions (classifications) made by the model Sensitive to class imbalances	Harmonic mean of: • Precision (proportion of positive indications actually correct) • Recall (proportion of actual positives identified correctly) FPs and FNs considered equally important	Shows tradeoff between True Positive Rate (TPR) and False Positive Rate (FPR) (i.e., trade-off between sensitivity and specificity) AUC measures the entire two-dimensional area underneath the ROC curve	• **ANOVA** • Cochran's Q test	• **Alpaydin's combined 5x2cv F-test** • 5x2cv paired t-test • k-foldcv paired t-test • McNemar's test

$\underbrace{\hspace{20em}}_{}$

Used to benchmark the performance of each algorithm. Samples of these metrics will be collected to reduce variance.

Used to compare the performance between algorithms.

Berkeley
UNIVERSITY OF CALIFORNIA

Source: Raschka, Sebastian. "Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning." arXiv.Org, 11 Nov. 2020, arxiv.org/abs/1811.12808.