Multimodal Behavioral Sensing for Precision Mental Health Care

PhD Defense By

Prerna Chikersal

Thesis Committee:

Anind Dey (Co-chair, UW ISchool)

Mayank Goel (Co-chair, CMU ISR/HCII)

Geoff Kauffman (CMU HCII)

Andrew Campbell (Dartmouth)

Mary Czerwinski (Microsoft Research)

Thesis Outline

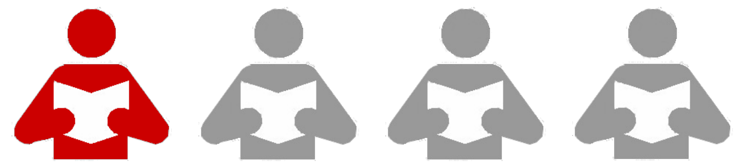
- Introduction
- The Curse of Dimensionality Challenge
- S1: Detecting Depression and Loneliness in College Students
- S2: Forecasting End of Semester Depression in College Students
- S3: Predicting the Mental Health of People with Multiple Sclerosis (MS) during the Covid-19 Stay-at-Home Period
- S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention
- S5: Predicting Periodically Assessed MS Outcomes Using Passively Sensed Behaviors and Ecological Momentary Assessments
- Thesis Contributions and Future Work

Thesis Outline

- Introduction
- The Curse of Dimensionality Challenge
- S1: Detecting Depression and Loneliness in College Students
- S2: Forecasting End of Semester Depression in College Students
- S3: Predicting the Mental Health of People with Multiple Sclerosis (MS) during the Covid-19 Stay-at-Home Period
- S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention
- S5: Predicting Periodically Assessed MS Outcomes Using Passively Sensed Behaviors and Ecological Momentary Assessments
- Thesis Contributions and Future Work

Mental Health Disorders are Very Prevalent!

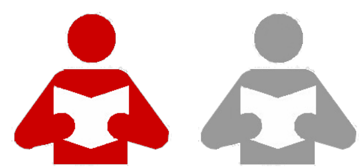
Lifetime occurrence within the US



Higher in certain groups.

- In any year, for college students

- Lifetime occurrence for patients with multiple sclerosis (pwMS)



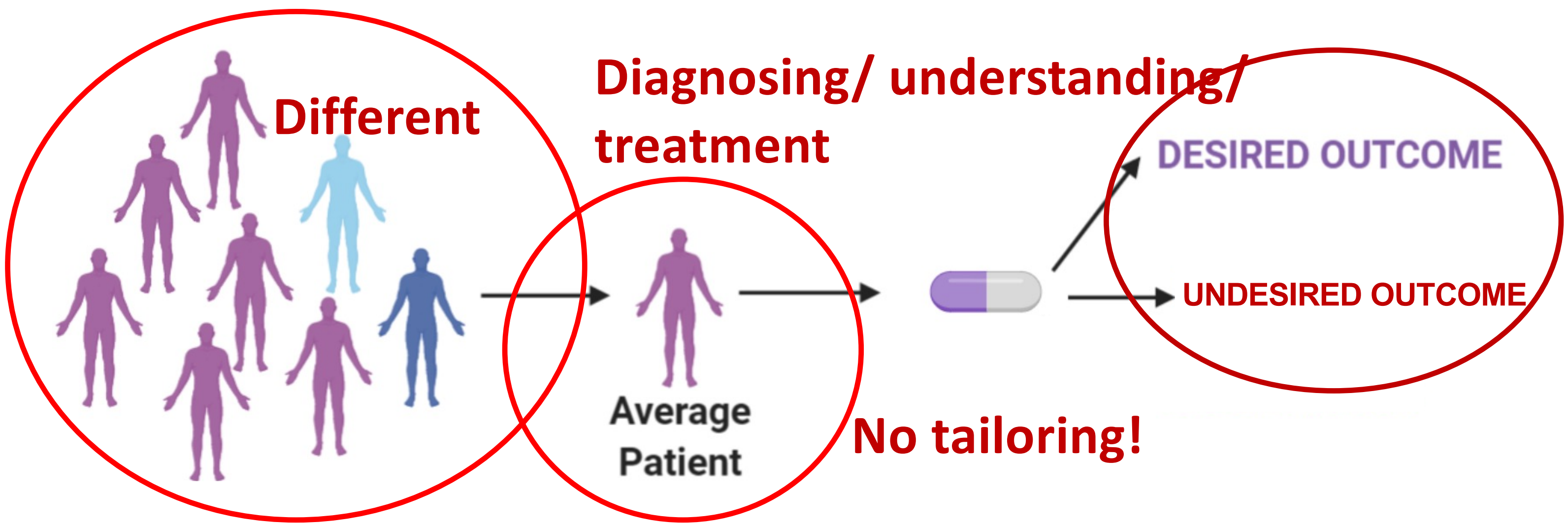
Covid-19 and social distancing → increased prevalence

The Burden of Mental Illnesses is Huge!

Diagnosis, understanding, and treatment → A public health priority

Barriers for Mental Health Care

- Many people do not seek help and thus, go undiagnosed.
  - Lack of awareness, stigma, limited access



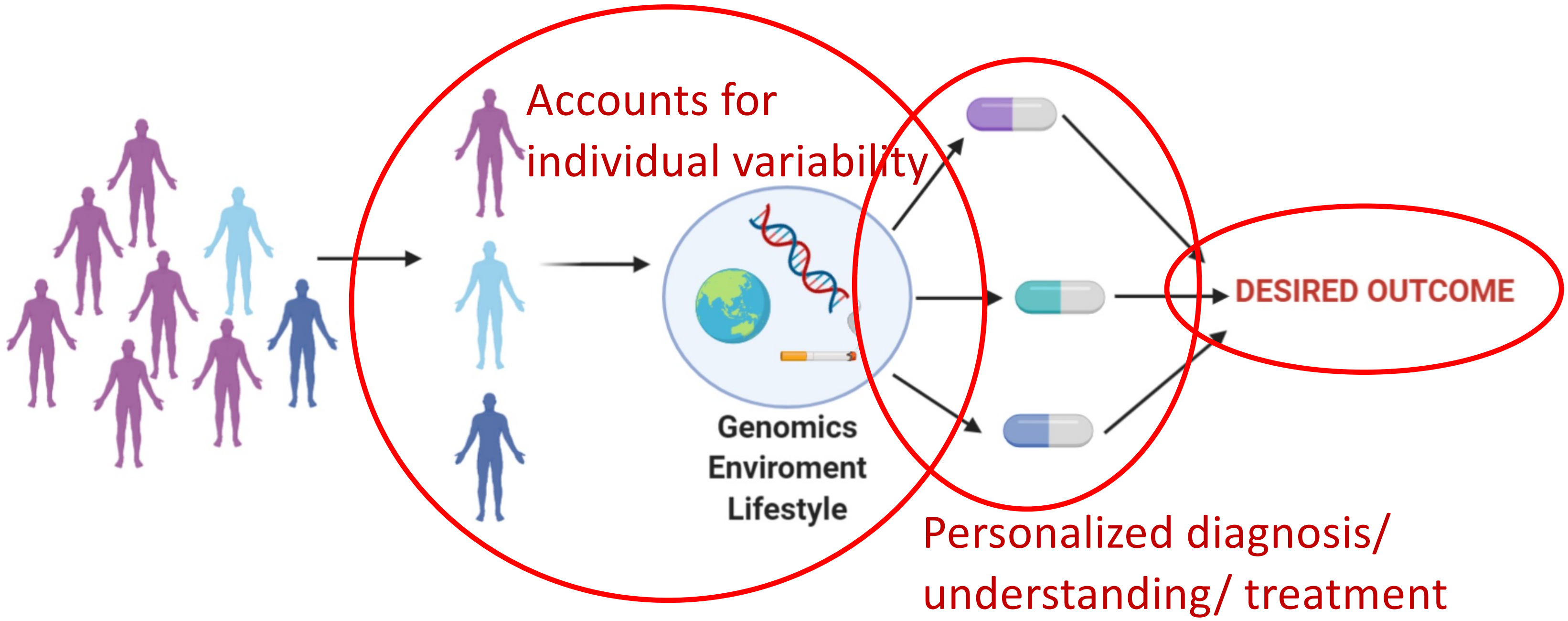
- For some people, diagnosis can take a long time → Delayed care.

The Need for Precision!

- Hence, there is a need to develop digital tools that
  - Increase access to mental health care,
  - while making diagnosis and treatment more precise by using data-driven insights and predictions.

Precision Health

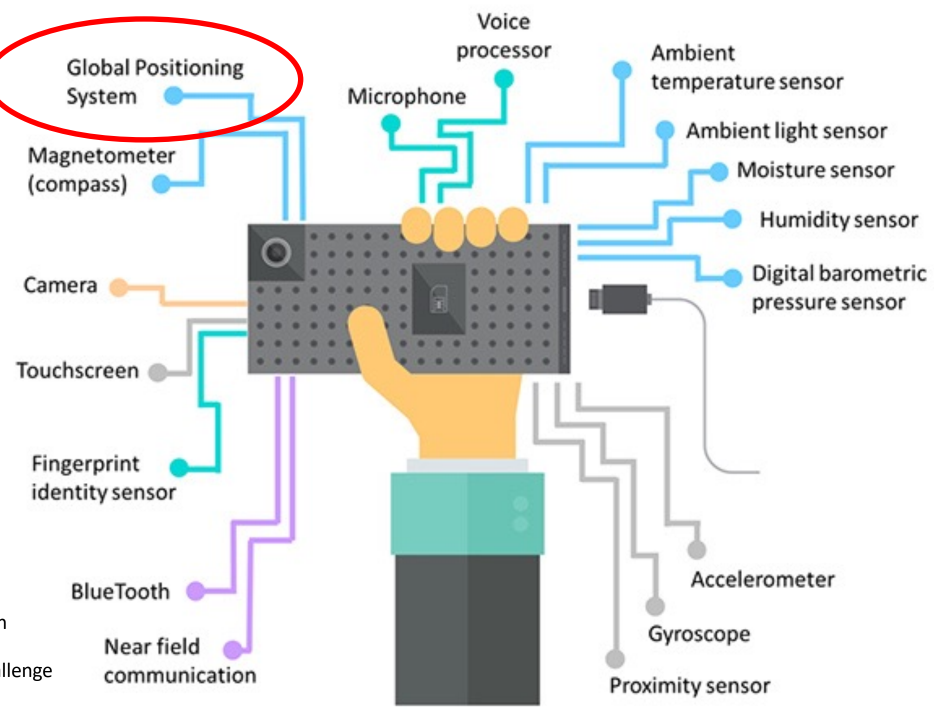
- An emerging data-driven approach to healthcare that:



Precision Mental Health (Precision MH)

- While genetics play a role in mental health, research has shown that behaviors, environment, and social context play a much larger role*. → Precision MH focus.
- Can be captured by smartphones and fitness trackers.

Time spent at home (rest/ engagement with the world)



*References:
1) Leonard Bickman, Aaron R Lyon, and Miranda Wolpert. Achieving precision mental health through effective assessment, monitoring, and feedback processes. 2016.
2) Michael Rutter. "The interplay of nature, nurture, and developmental influences: the challenge ahead for mental health". In: Archives of General Psychiatry 59.11 (2002), pp. 996–1000.

Precision Mental Health (Precision MH)

- While genetics plays a role in mental health, research has shown that behaviors, environment, and social context play a much larger role. → Precision MH focus.
- Can be captured by smartphones and fitness trackers.

The Biggest Problem in Precision MH

- Challenges
  - Variety of **data sources**
    - *e.g.*, data from many sensors in multiple devices.
  - From patients with different **multi-morbidities**
  - Patients in **different contexts**
    - e.g., demographics, medical history, past behaviors
- Such high-dimensional data
  - Creates new opportunities for precision
  - BUT makes it harder to derive robust insights and models for real-world scenarios.
  - Is the biggest problem in Precision MH!

    → **The Curse of Dimensionality Challenge**

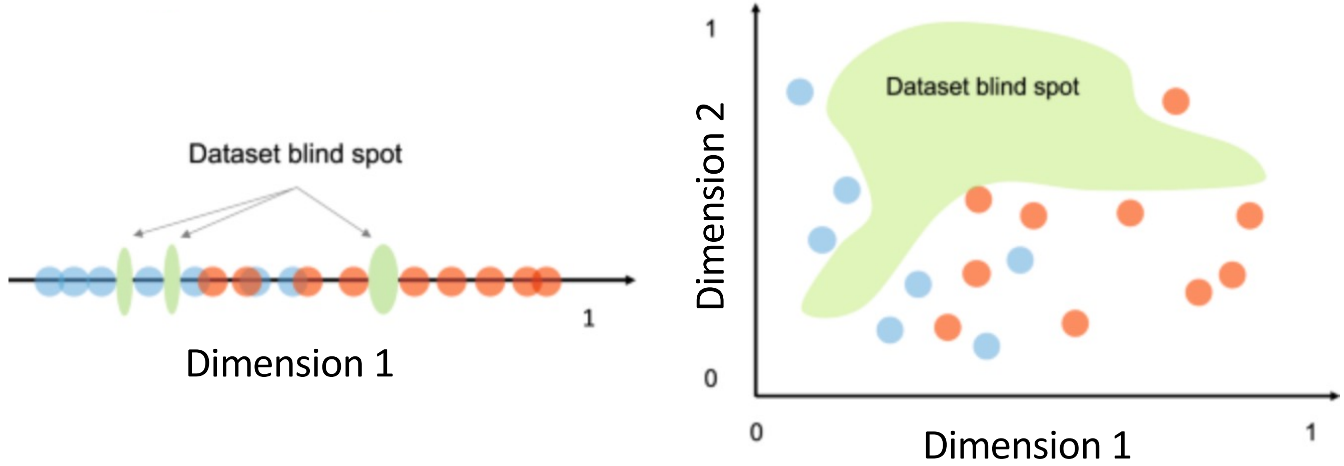
Thesis Outline

- Introduction
- The Curse of Dimensionality Challenge
- S1: Detecting Depression and Loneliness in College Students
- S2: Forecasting End of Semester Depression in College Students
- S3: Predicting the Mental Health of People with Multiple Sclerosis (MS) during the Covid-19 Stay-at-Home Period
- S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention
- S5: Predicting Periodically Assessed MS Outcomes Using Passively Sensed Behaviors and Ecological Momentary Assessments
- Thesis Contributions and Future Work

The Curse of Dimensionality Challenge

- Number of samples needed to estimate a function grows exponentially w.r.t. the number of input variables or features.

SAME NUMBER OF SAMPLES



- More blind spots → highly variable models (*e.g.*, vastly different selected features) and highly variable estimates of true model performance, across different subsamples of the same dataset.

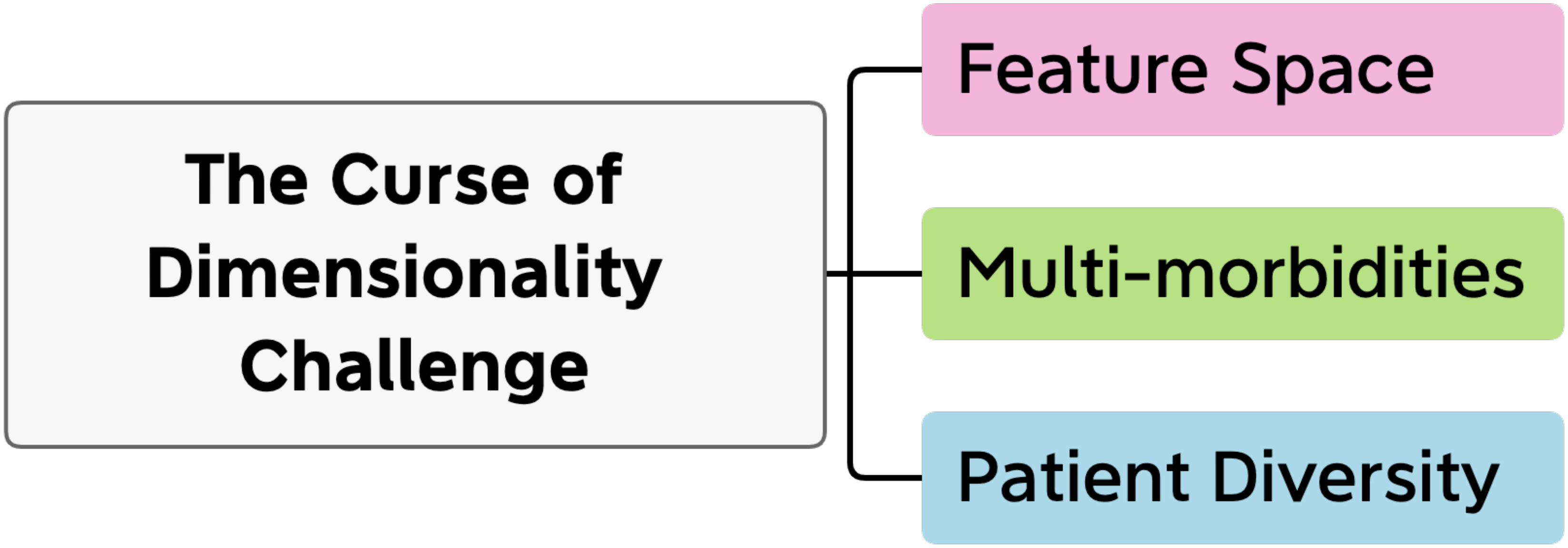
Thesis Problem Statement

This thesis focuses on developing and presenting novel methods that address **the curse of dimensionality challenge** with respect to:

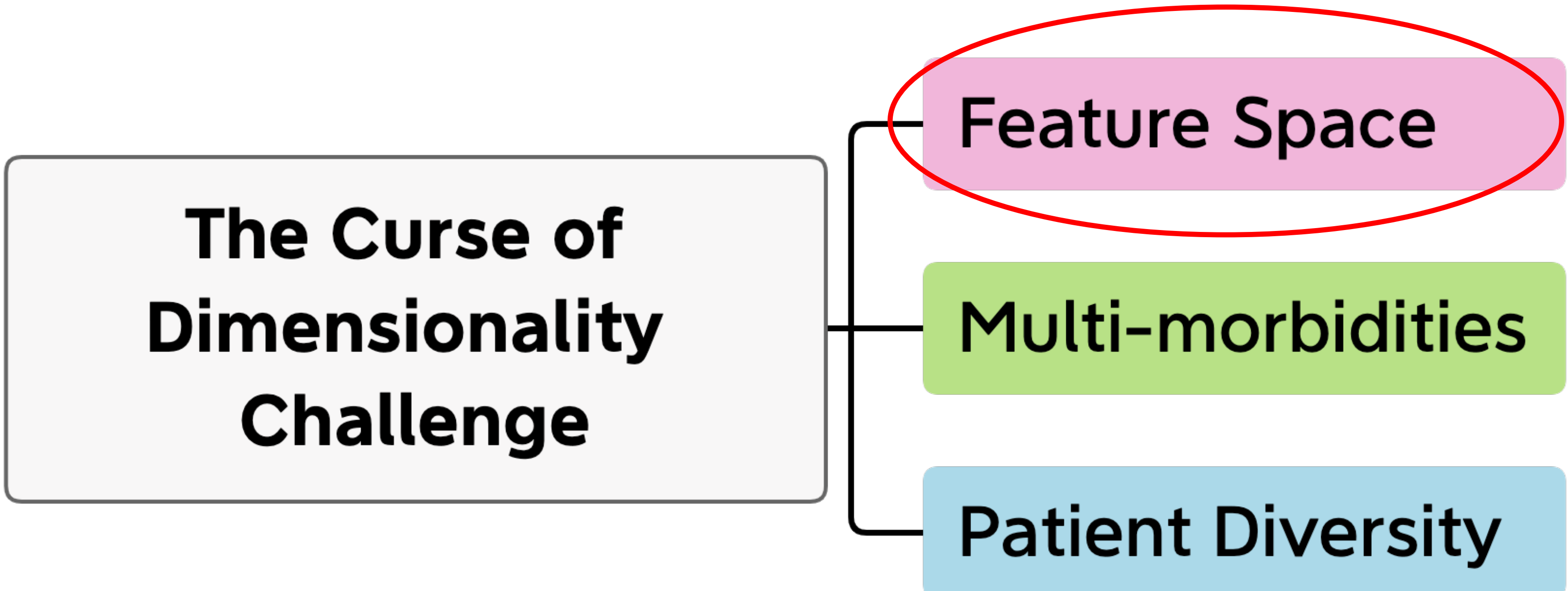
a) the feature space
b) multiple outcomes stemming from co-morbid medical conditions, and
c) diversity in patient contexts and characteristics.

Explanation to follow…

The Curse of Dimensionality Challenge Contd.

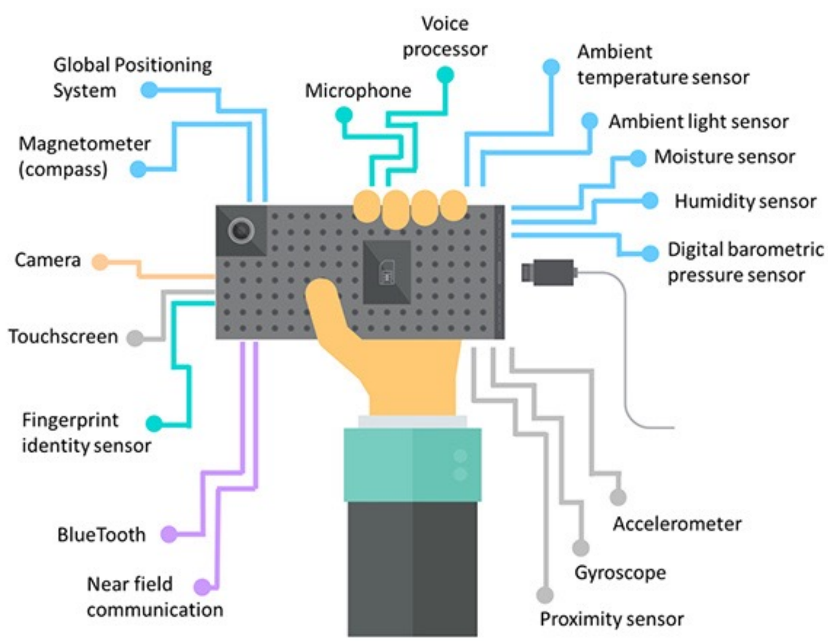


The Curse of Dimensionality Challenge Contd.



The Curse of Dimensionality w.r.t. the Feature Space

- When analyzing sensor data or interaction logs, the number of features >> number of samples:
  - Many sensors in multiple devices



The Curse of Dimensionality w.r.t. the Feature Space Contd.

- When analyzing sensor data or interaction logs, the number of features >> number of samples:
  - High cost of data collection → small sample size
  - Survey-based outcomes burden users

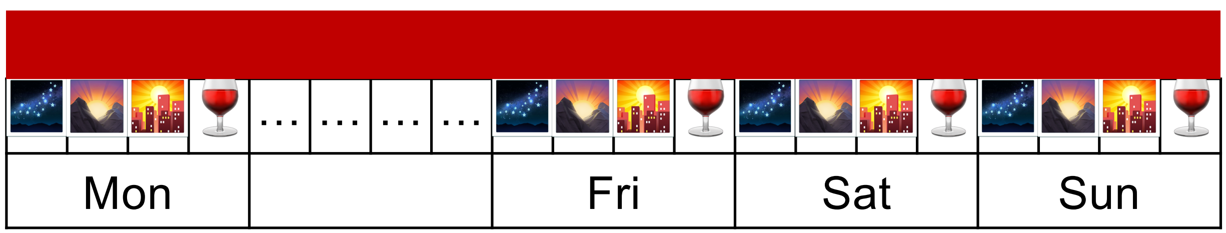
		Not at all	Several days	More than half the days	Nearly every day
1.	Little interest or pleasure in doing things	0	1	2	3
2.	Feeling down, depressed, or hopeless	0	1	2	3
3.	Trouble falling or staying asleep, or sleeping too much	0	1	2	3
4.	Feeling tired or having little energy	0	1	2	3
5.	Poor appetite or overeating	0	1	2	3
6.	Feeling bad about yourself — or that you are a failure or have let yourself or your family down	0	1	2	3
7.	Trouble concentrating on things, such as reading the newspaper or watching television	0	1	2	3
8.	Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	0	1	2	3
9.	Thoughts that you would be better off dead or of hurting yourself in some way	0	1	2	3

The Curse of Dimensionality w.r.t. the Feature Space Contd.

- When analyzing sensor data or interaction logs, the number of features >> number of samples:
  - Same features from different temporal slices are useful

Step_Count_All_All:

- A measure of overall physical activity

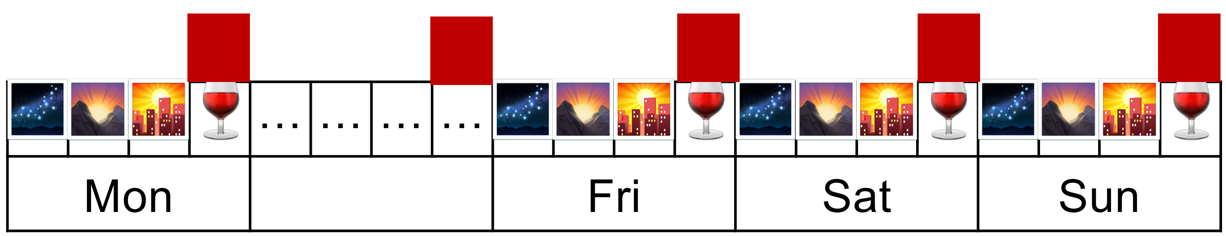


The Curse of Dimensionality w.r.t. the Feature Space Contd.

- When analyzing sensor data or interaction logs, the number of features >> number of samples:
  - Same features from different temporal slices are useful

Step_Count_All_Evenings:

- may indicate after work exercise/ activities

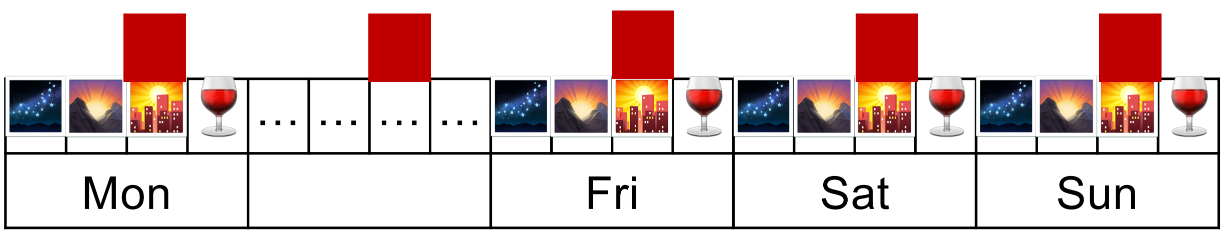


The Curse of Dimensionality w.r.t. the Feature Space Contd.

- When analyzing sensor data or interaction logs, the number of features >> number of samples:
  - Same features from different temporal slices are useful

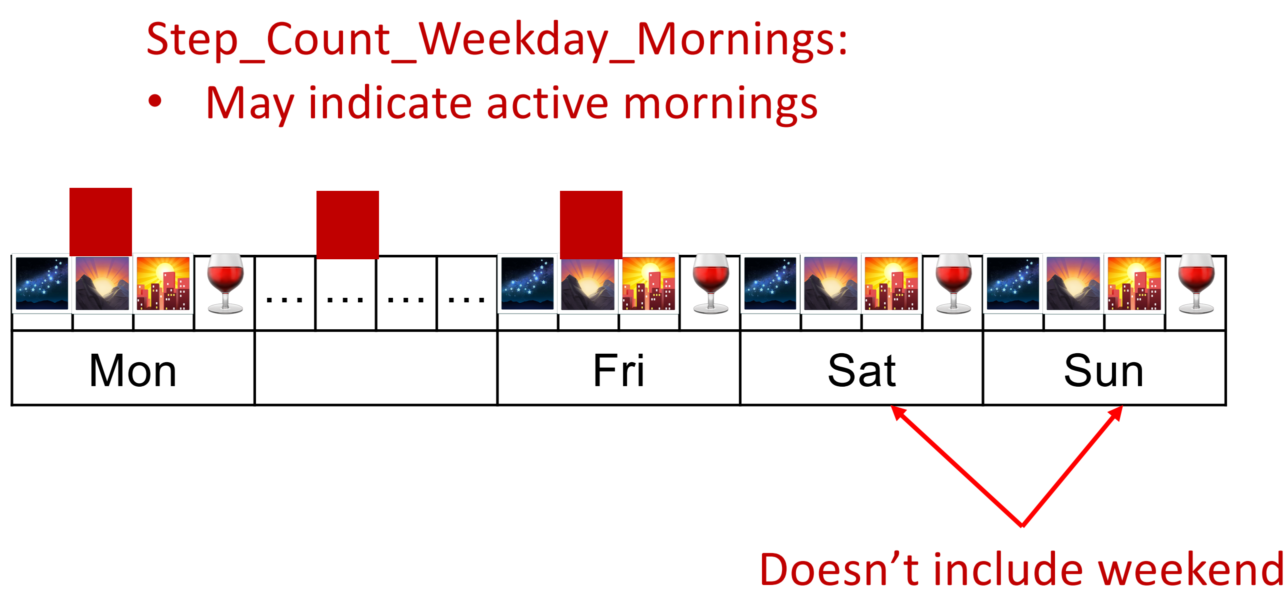
Whereas Step_Count_All_Afternoons:

- May be more related to occupation



The Curse of Dimensionality w.r.t. the Feature Space Contd.

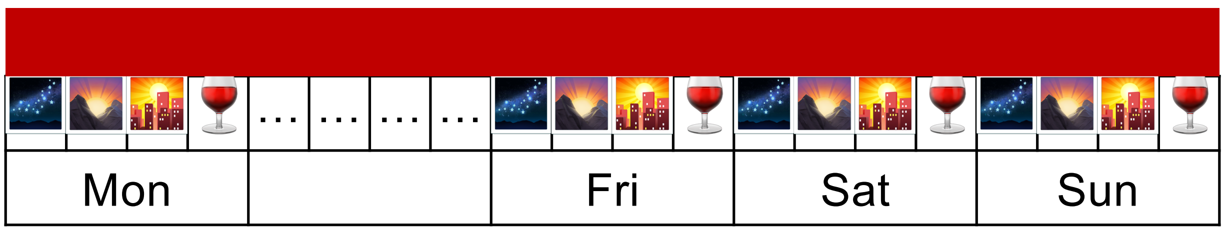
- When analyzing sensor data or interaction logs, the number of features >> number of samples:
  - Same features from different temporal slices are useful

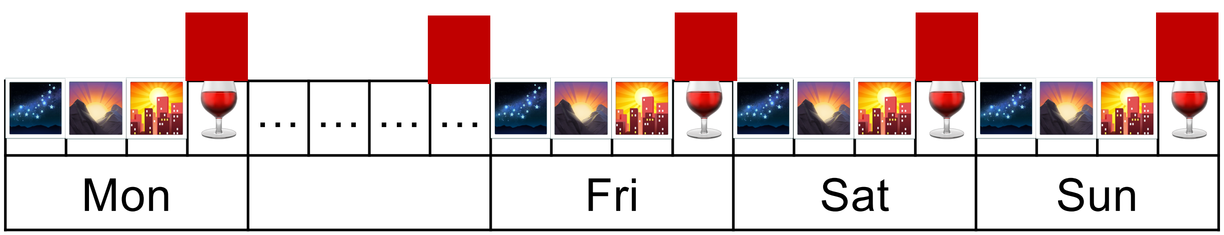


The Curse of Dimensionality w.r.t. the Feature Space Contd.

- When analyzing sensor data or interaction logs, the number of features >> number of samples:
  - Same features from different temporal slices are useful

Note: Temporal slices overlap

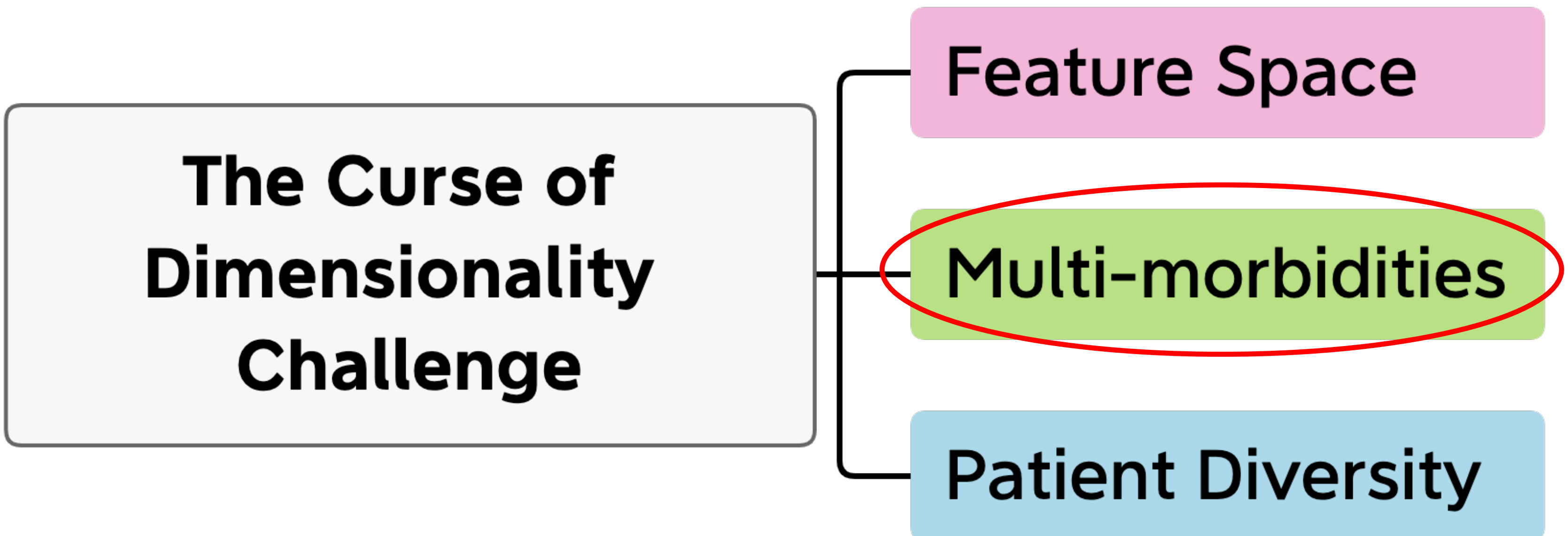




The Curse of Dimensionality w.r.t. the Feature Space Contd.

- When analyzing sensor data or interaction logs, the number of features >> number of samples:
  - Same features from different temporal slices are useful
  - As they may indicate different things
  - But this further increases the size of the feature space
  - And adds to the curse of dimensionality in the feature space

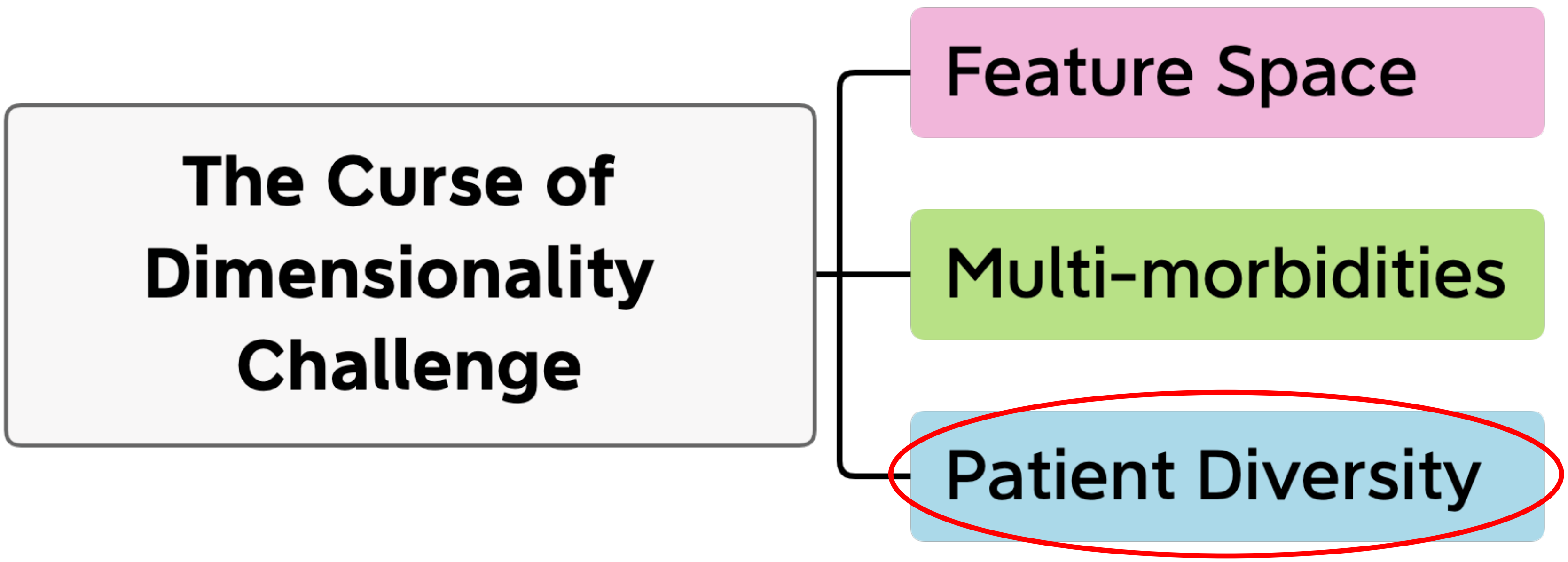
The Curse of Dimensionality Challenge Contd.



The Curse of Dimensionality w.r.t. Multi-morbidities

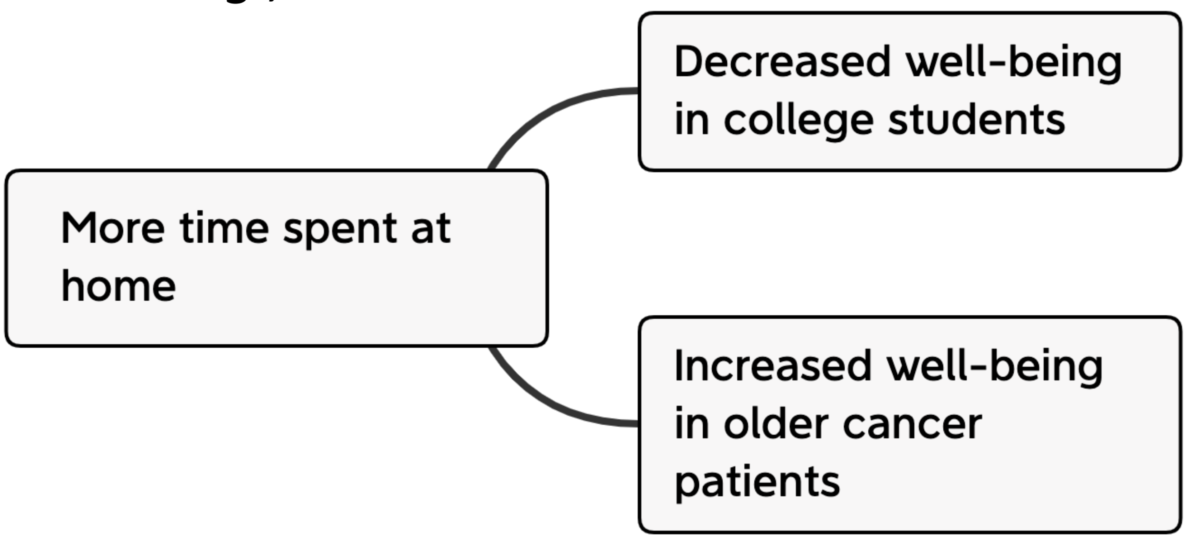
- Multimorbidity is the co-occurrence of >=2 chronic conditions.
- 70% of people with a MH disorder have co-morbidities.
- Co-morbid conditions could be the most important factor during diagnosis or treatment of the primary condition.
- Yet, they are often ignored in clinical trials and studies. In many studies, people with co-morbid conditions are explicitly excluded.
- Hence, it is important to consider conditions co-morbid with the primary condition, even though this will add to the curse of dimensionality by increases the number of outcomes.

The Curse of Dimensionality Challenge Contd.



The Curse of Dimensionality w.r.t. Patient Diversity

- Previous work shows that the relationship between behaviors and outcomes may be dependent on patient context and characteristics. *E.g.,*



- → Accounting for the patient's context is important.

Thesis Outline

- Introduction
- The Curse of Dimensionality Challenge
- S1: Detecting Depression and Loneliness in College Students
- S2: Forecasting End of Semester Depression in College Students
- S3: Predicting the Mental Health of People with Multiple Sclerosis (MS) during the Covid-19 Stay-at-Home Period
- S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention
- S5: Predicting Periodically Assessed MS Outcomes Using Passively Sensed Behaviors and Ecological Momentary Assessments
- Thesis Contributions and Future Work

S1: Detecting Depression and Loneliness In College Students

Capture behavior and symptoms

Detect Depression

Detect Loneliness

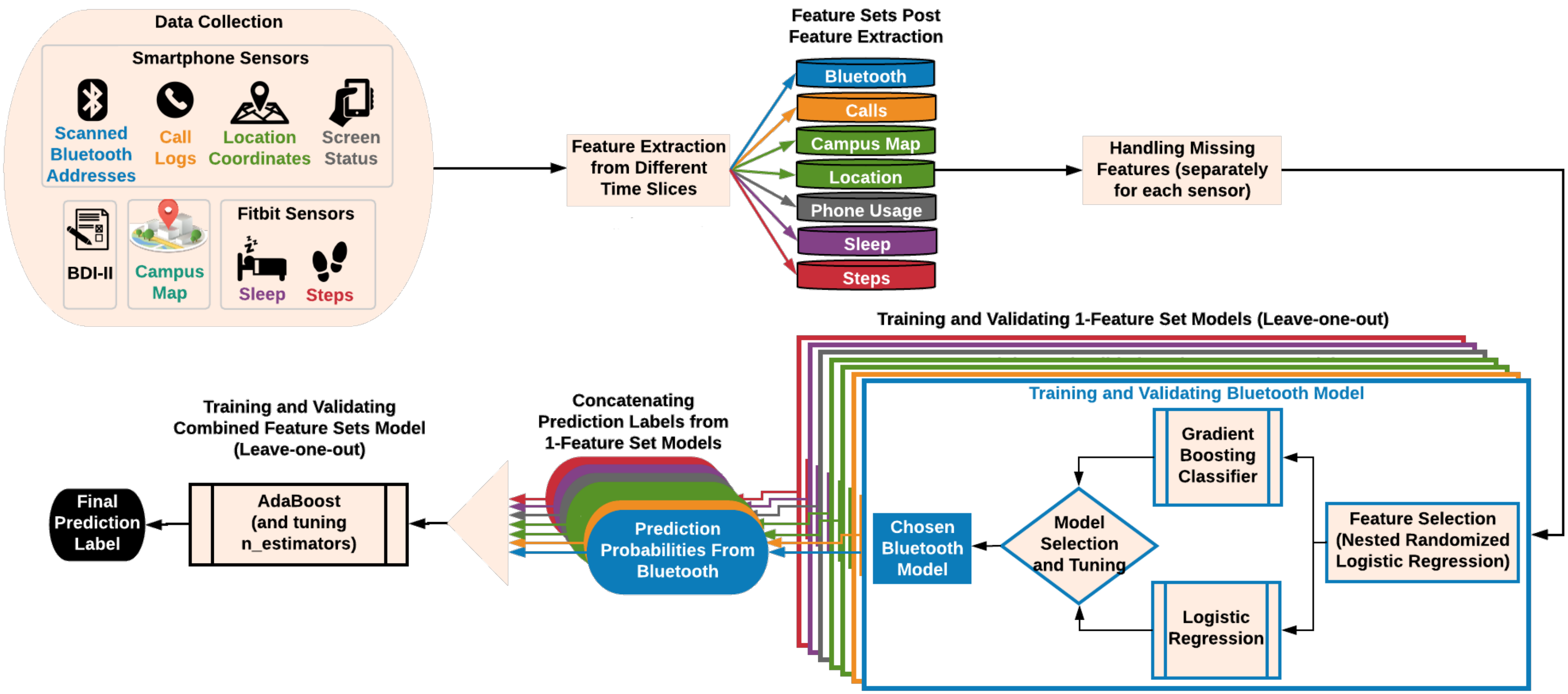
S1: Gaps in Previous Work

- Does not address situations where we have limited ground truth as compared to the feature space:
  - Relies on frequent measurement (*e.g.* every week) → burden
  - Limits the number of features (*e.g.* no temporal slicing)
  - → doesn't face the curse of dim. w.r.t. the feature space
- Rarely evaluates if the same approach can be used to detect multiple co-morbid outcomes.
  - → doesn't face the curse of dim. w.r.t. multi-morbidities

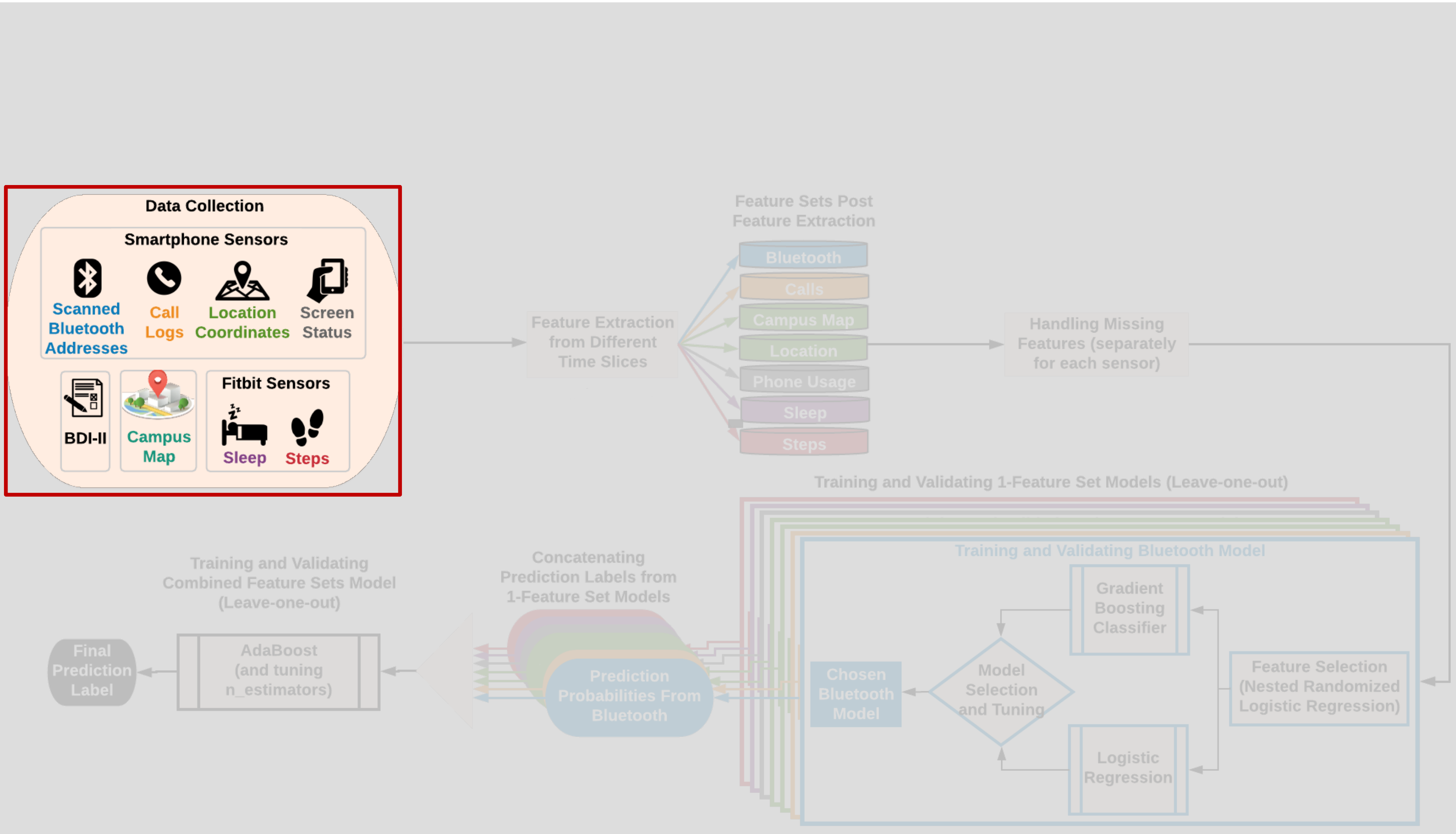
References

- Luca Canzian and Mirco Musolesi. 2015. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing. ACM, 1293–1304.
- Asma Ahmad Farhan, Chaoqun Yue, Reynaldo Morillo, Shweta Ware, Jin Lu, Jinbo Bi, Jayesh Kamath, Alexander Russell, Athanasios Bamis, and Bing Wang. 2016. Behavior vs. introspection: refning prediction of clinical depression via smartphone sensing data.. In Wireless Health. 30–37.
- Sohrab Saeb, Mi Zhang, Christopher J Karr, Stephen M Schueller, Marya E Corden, Konrad P Kording, and David C Mohr. 2015. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. Journal of medical Internet research 17, 7 (2015).
- Fabian Wahle, Tobias Kowatsch, Elgar Fleisch, Michael Rufer, and Stef Weidt. 2016. Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR mHealth and uHealth 4, 3 (2016).
- Rui Wang, Weichen Wang, Alex daSilva, Jeremy F Huckins, William M Kelley, Todd F Heatherton, and Andrew T Campbell. 2018. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2, 1 (2018), 43.

S1: Methodology



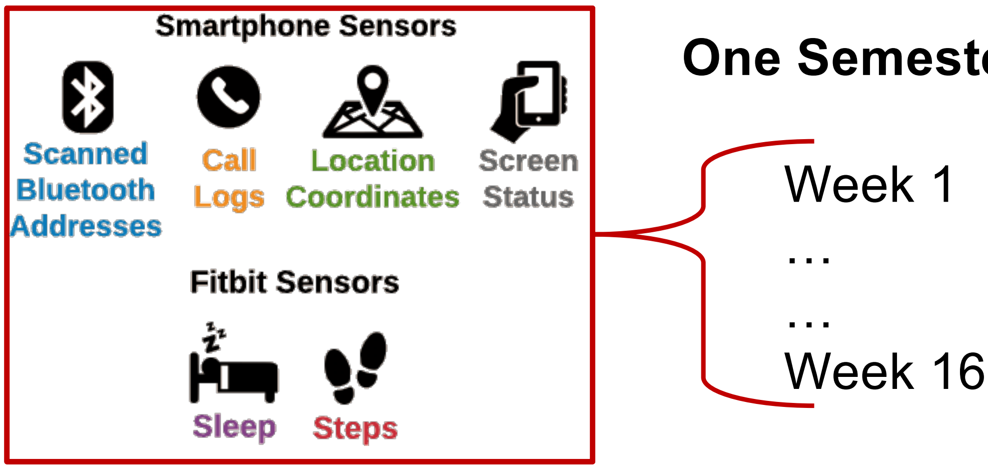
S1: Methodology



S1: Methodology – Data Collection

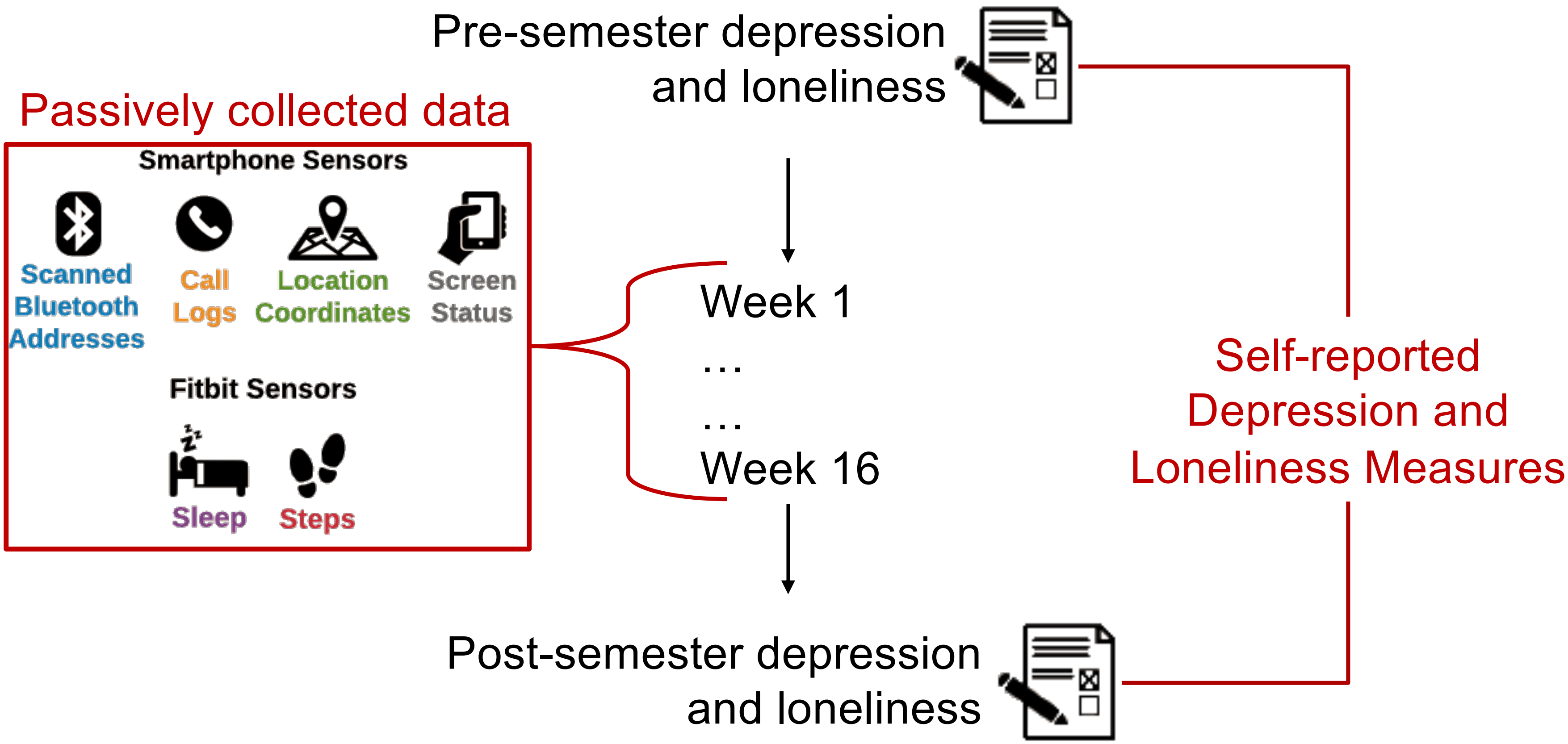
Passively collected data

One Semester



- 138 First Year College Students at the same University

S1: Methodology – Data Collection

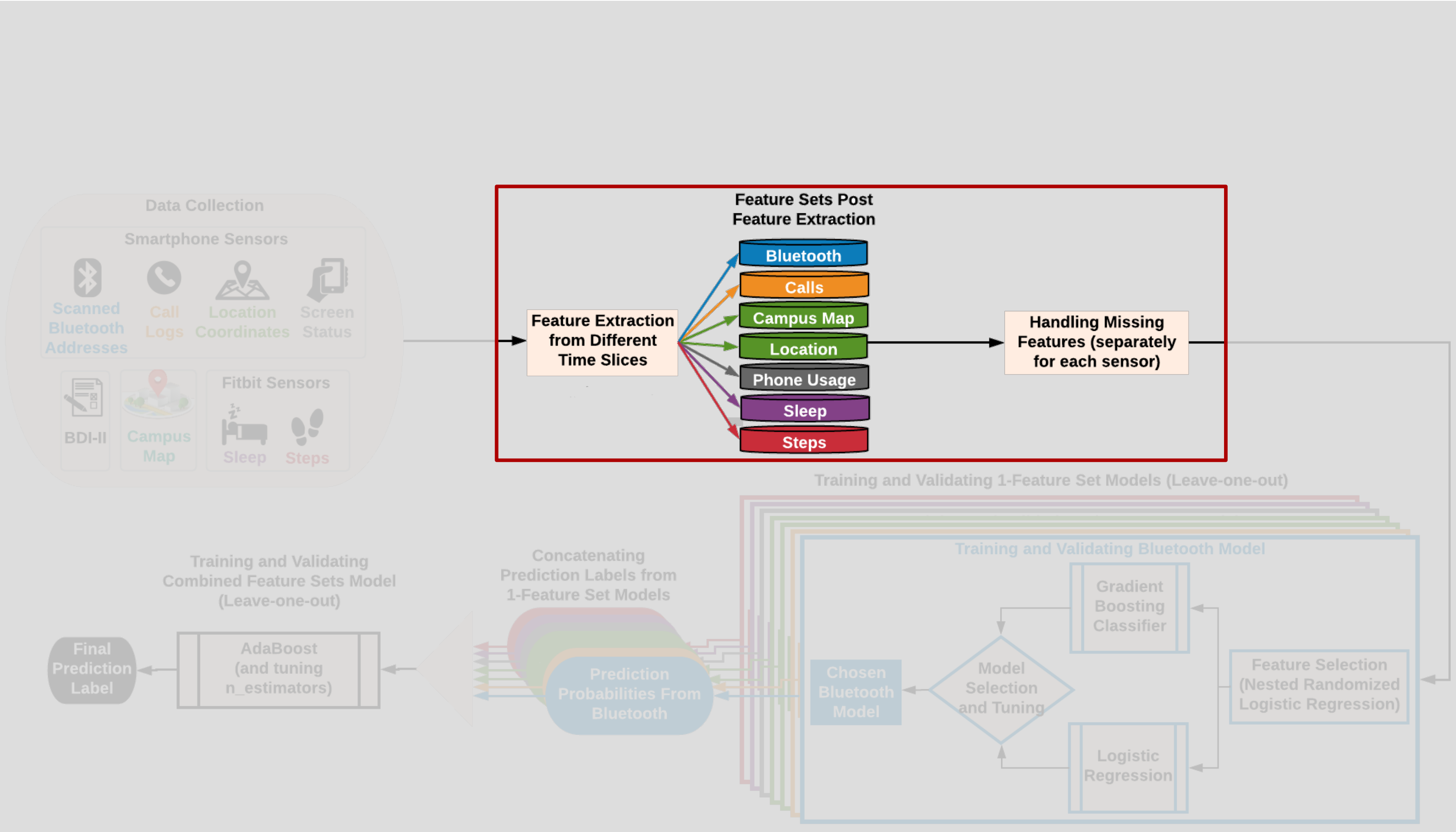


- 138 First Year College Students
- Aware API for iOS and Android, and Fitbit

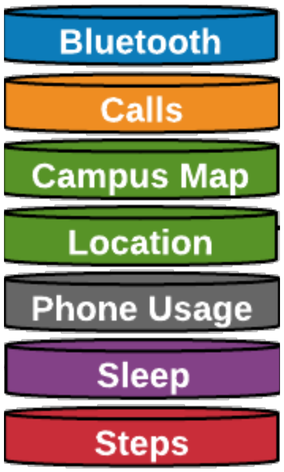
S1: Methodology – Data Collection Outcomes

- Post-semester Depression
  - Binary: “no depression” vs. “has depression”
- Change in Depression
  - Binary: “severity level remains the same” vs. “severity level worsens” (No one improved)
- Post-semester Loneliness
  - Binary: “high loneliness” vs. “low loneliness”
- Change in Loneliness
  - 3-class: “increased” vs. “decreased” vs. “remained the same”

S1: Methodology

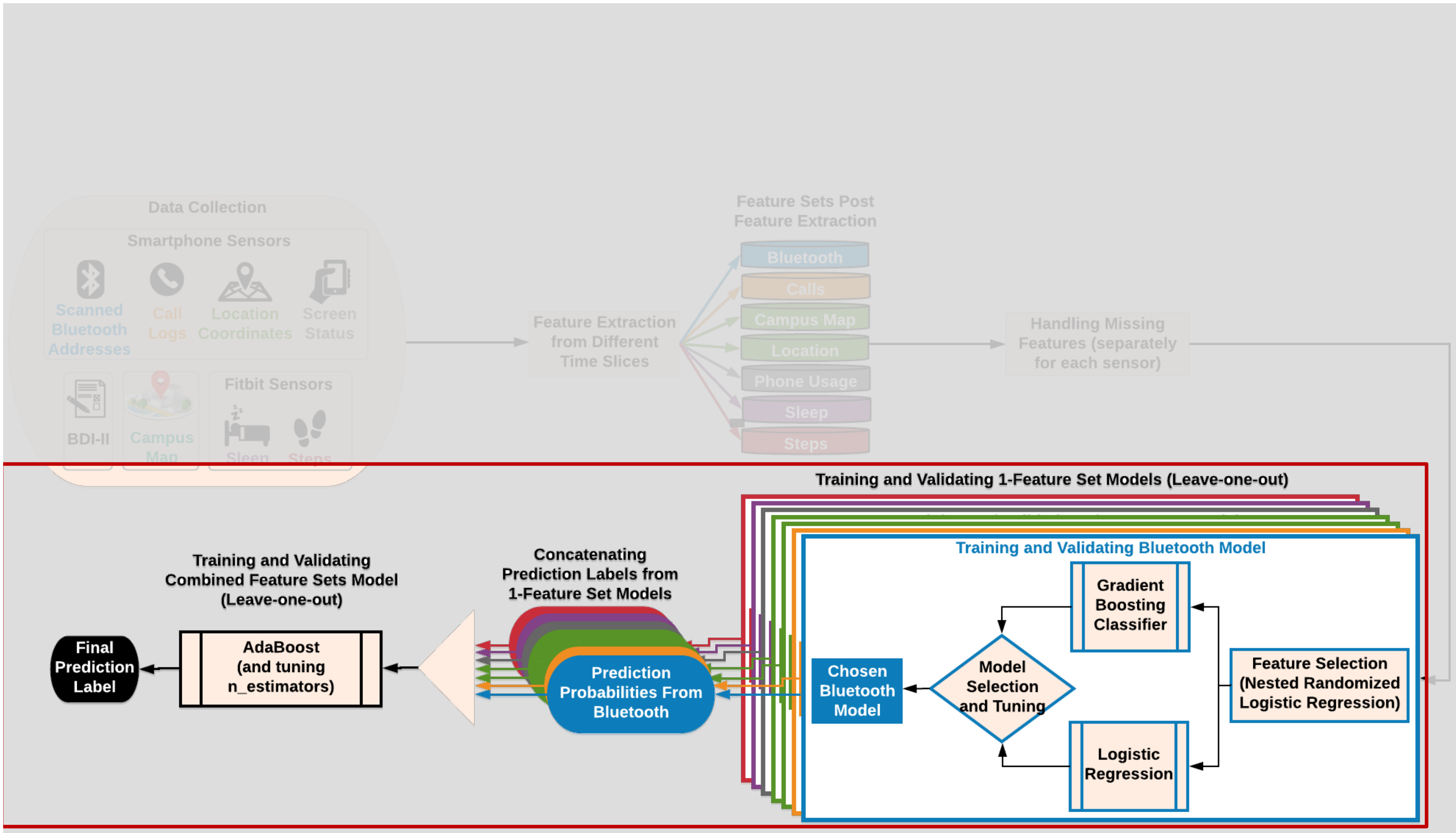


S1: Methodology – Feature Extraction



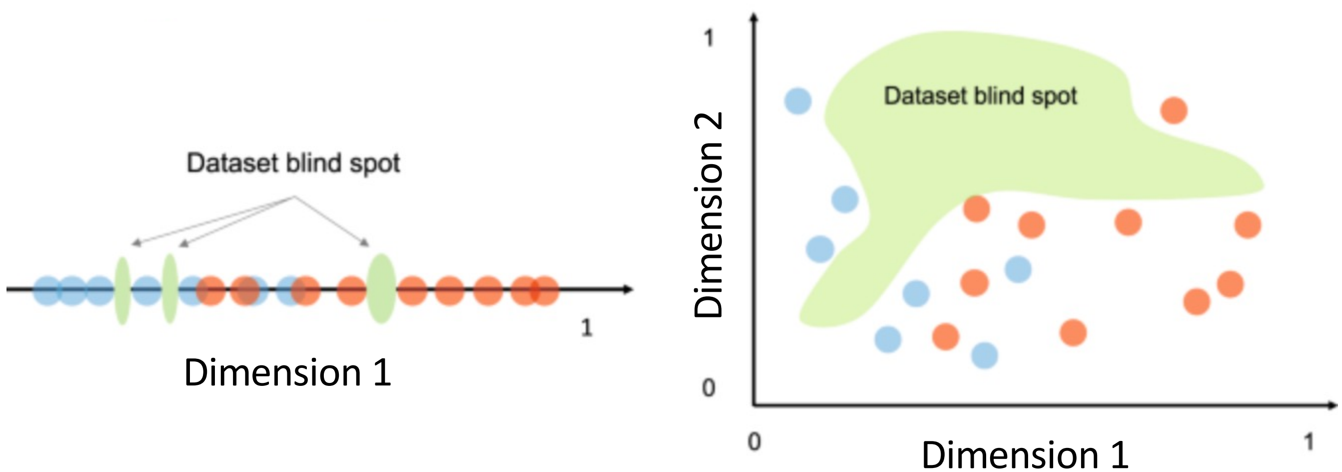
- We collect data from 7 feature sets/ sensors.
- From each, we extract features from 45 temporal slices.
- 50K features and only 79 people from all feature sets!
  → The curse of dim. w.r.t. the feature space applies.

S1: Methodology



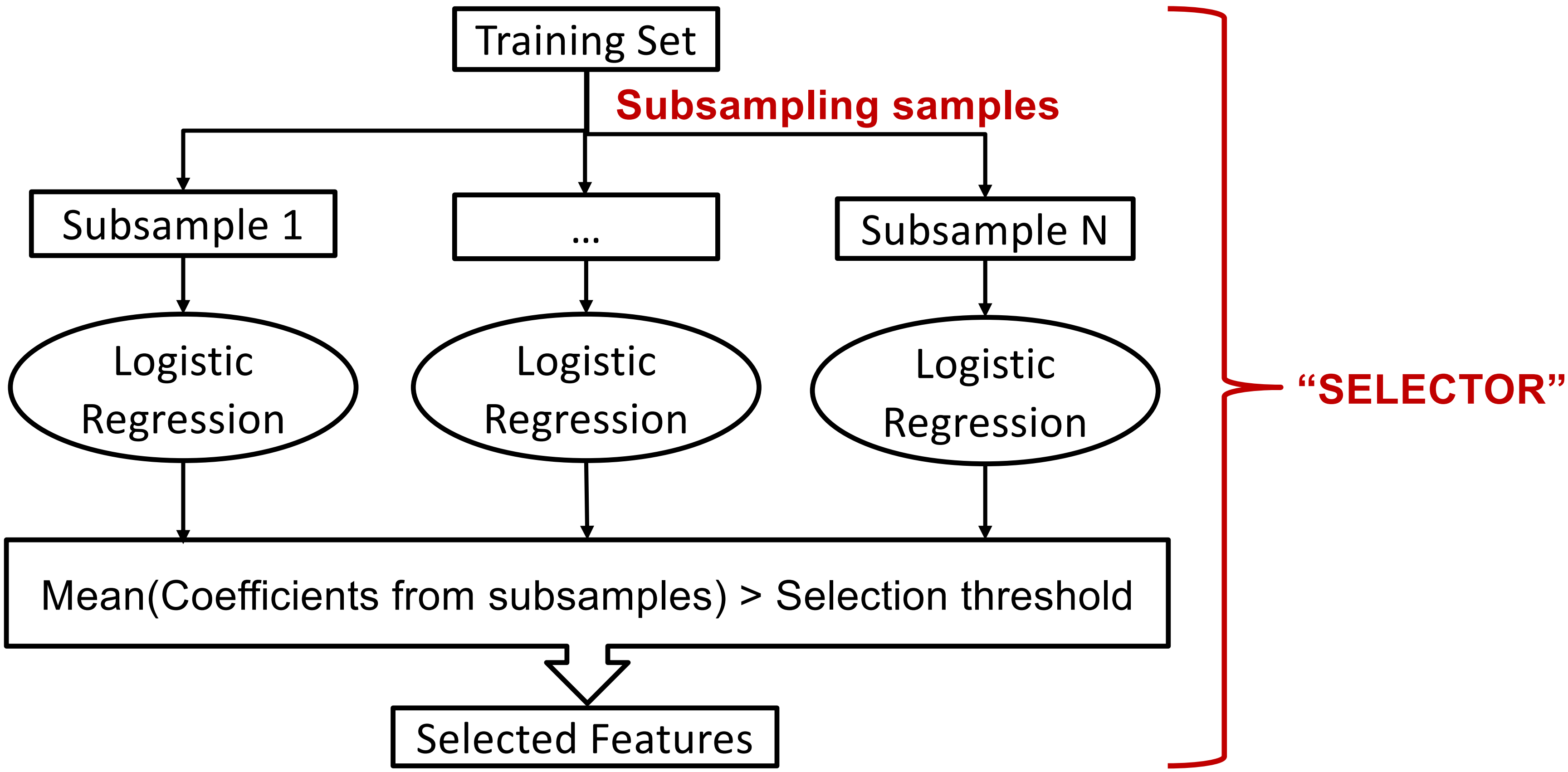
S1: Methodology – Feature Selection & Modeling Contd.

- Even 1-sensor models: 6-10K features and only ~110 people
- Off-the-shelf approaches for 1-sensor predictions:
  - Poor accuracy
  - High variability in selected features across LOO folds
  - RECALL: Blind spots → high variability, low robustness

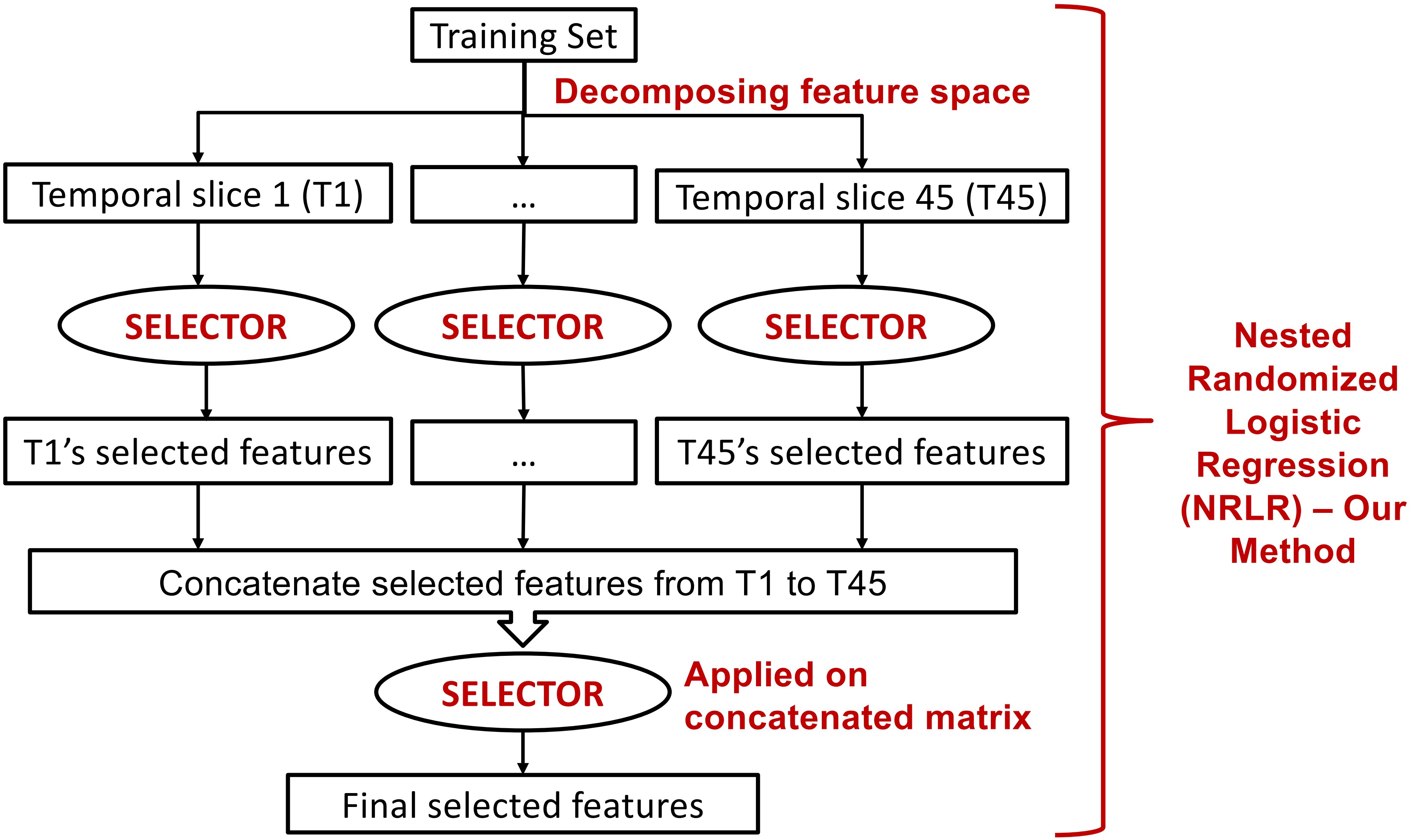


- Need a new method for stable feature selection
  - Feature space decomposition → reduce blind spots

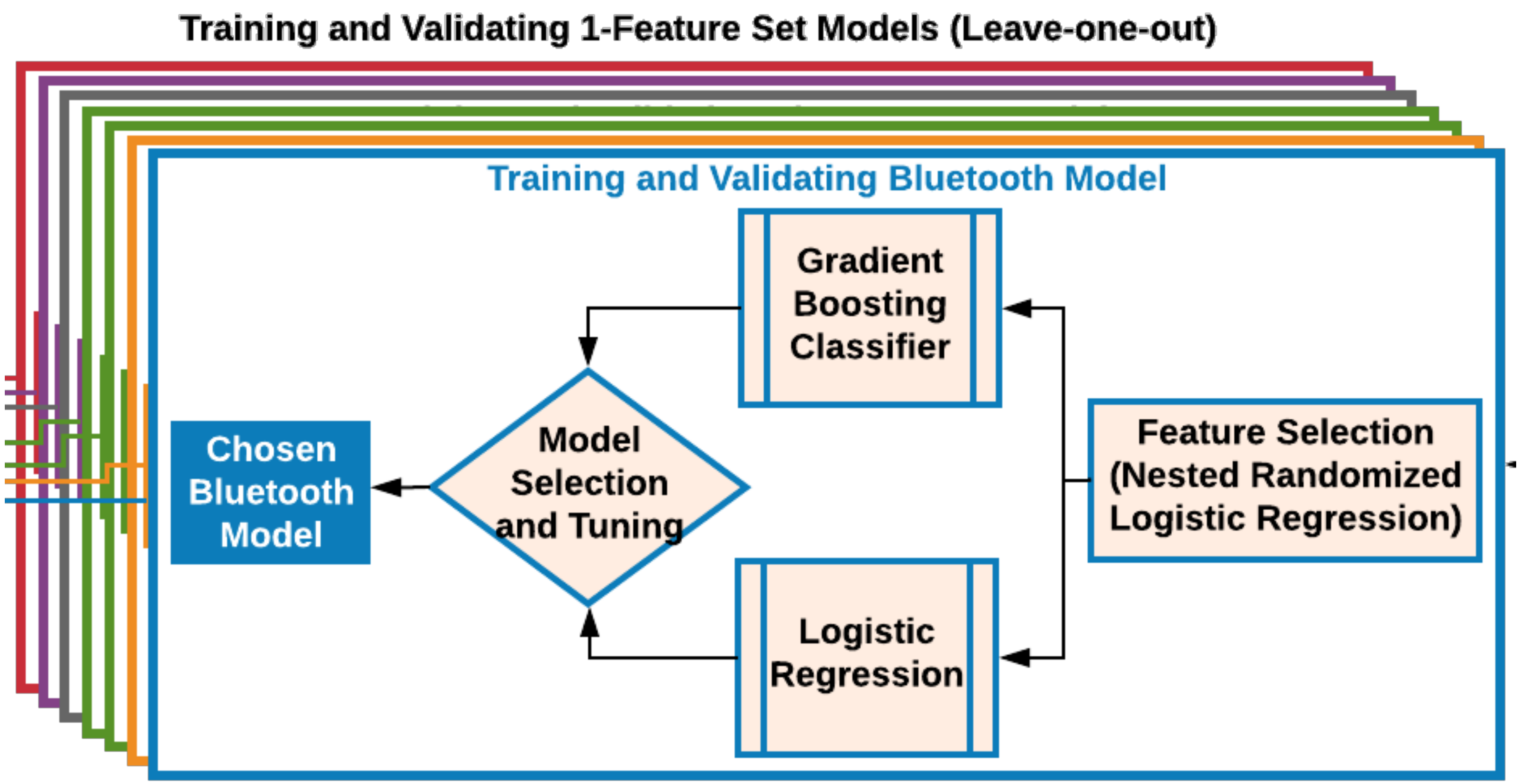
S1: Methodology – Feature Selection & Modeling Contd.



S1: Methodology – Feature Selection & Modeling Contd.

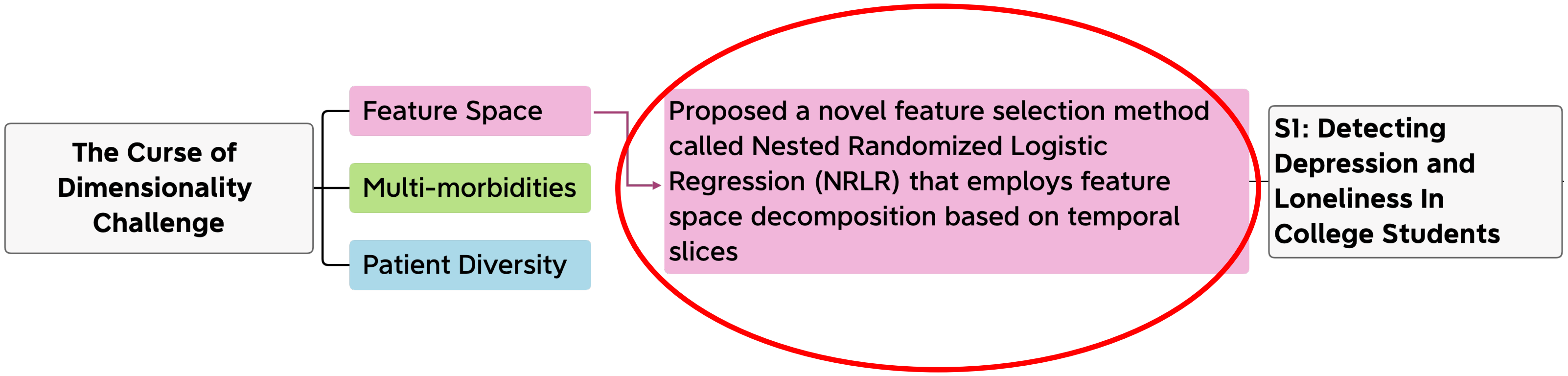


S1: Methodology – Feature Selection & Modeling Contd.

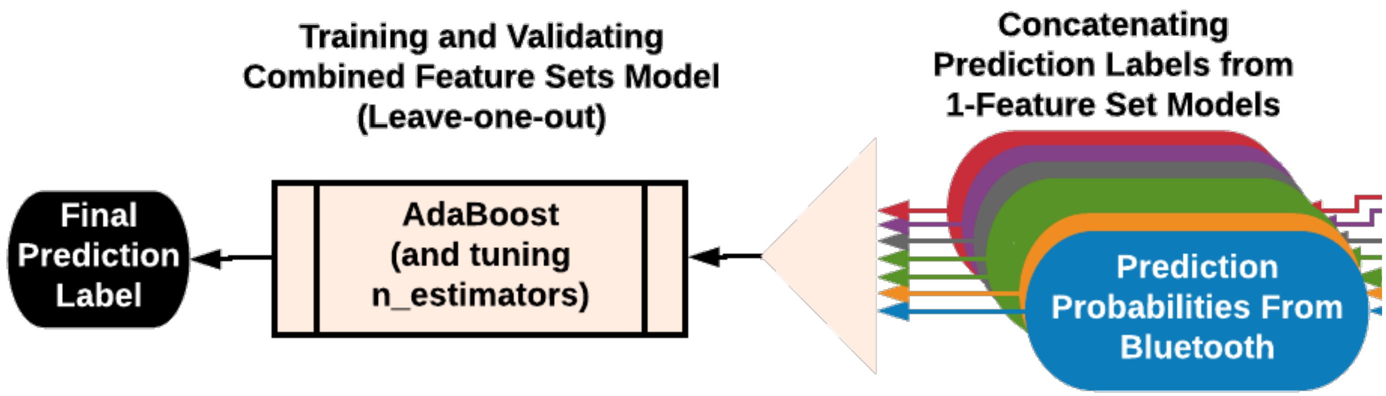


S1: Methodology – Feature Selection & Modeling Contd.

How does this address the curse of dimensionality?



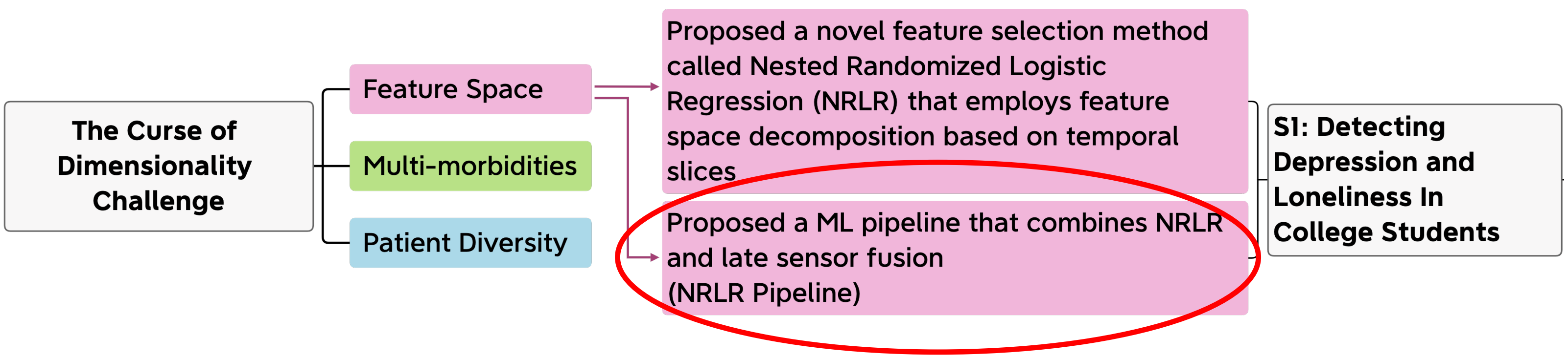
S1: Methodology – Feature Selection & Modeling Contd.



- Two methods for combining sensors:
  - Early sensor fusion: Combine sensor data or features, and then do ML to get the final prediction.
  - Late sensor fusion: We use ML to get a prediction for each sensor, and later, combine those predictions.
- We do late sensor fusion → aids in feature space decomposition.

S1: Methodology – Feature Selection & Modeling Contd.

How does this address the curse of dimensionality?



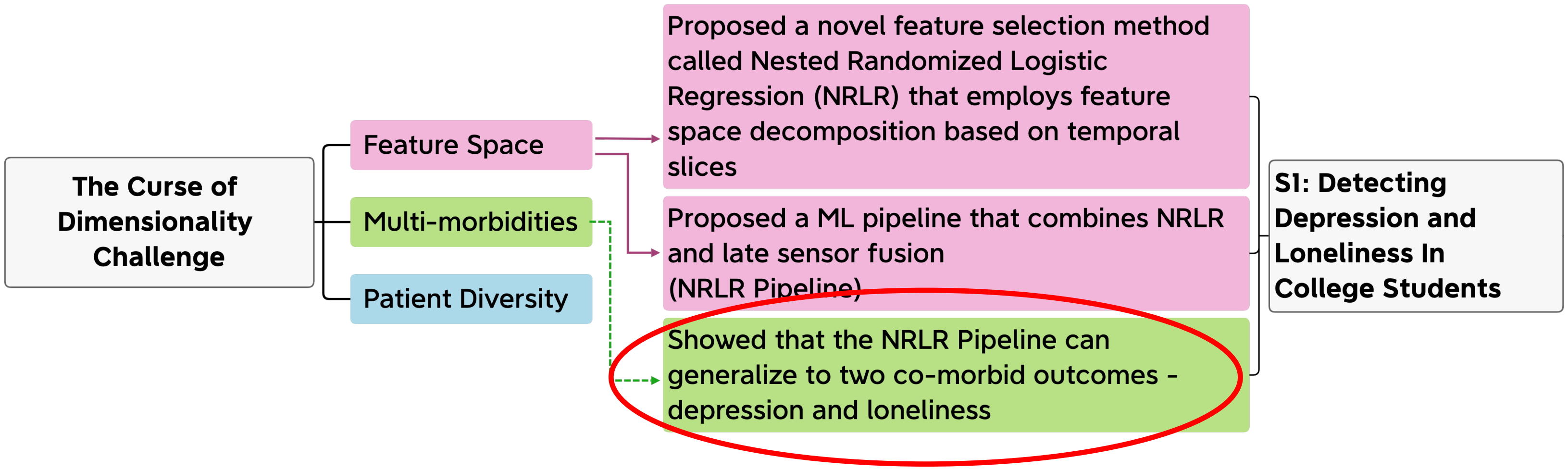
S1: Results – All

- Post-semester depression (binary): 85.7% accuracy
- Change in depression (binary): 85.4% accuracy
- Post-semester loneliness (binary): 80.2% accuracy
- Change in loneliness: 88.4% accuracy

- Results in comparison with existing methods:
  - Our method outperformed Lasso and KNN for all except one sensor.
  - It also selected fewer features than Lasso and KNN.
  - These findings were consistent for depression and loneliness.

S1: Addressing the Curse of Dimensionality

How does this address the curse of dimensionality?

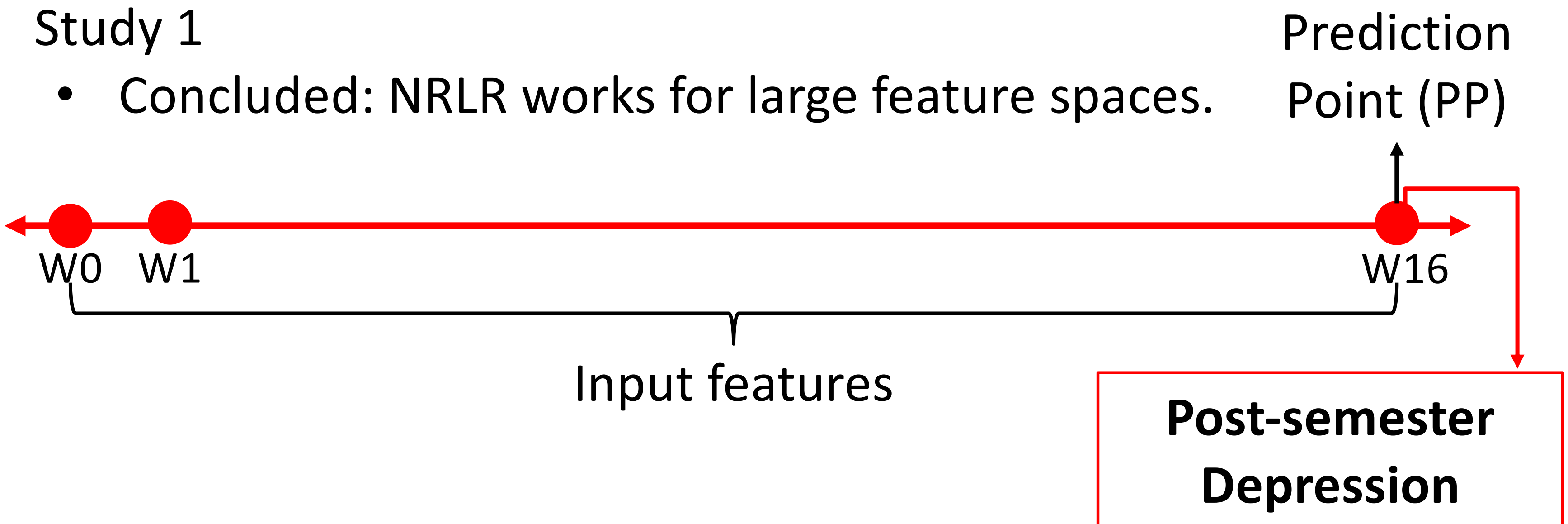


Thesis Outline

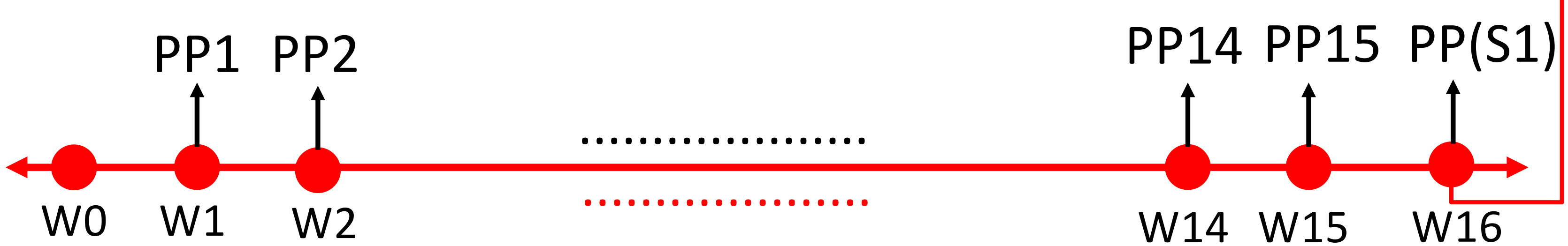
- Introduction
- The Curse of Dimensionality Challenge
- S1: Detecting Depression and Loneliness in College Students
- S2: Forecasting End of Semester Depression in College Students
- S3: Predicting the Mental Health of People with Multiple Sclerosis (MS) during the Covid-19 Stay-at-Home Period
- S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention
- S5: Predicting Periodically Assessed MS Outcomes Using Passively Sensed Behaviors and Ecological Momentary Assessments
- Thesis Contributions and Future Work

S2: Forecasting End of Semester Depression In College Students

- Study 1
  - Concluded: NRLR works for large feature spaces.



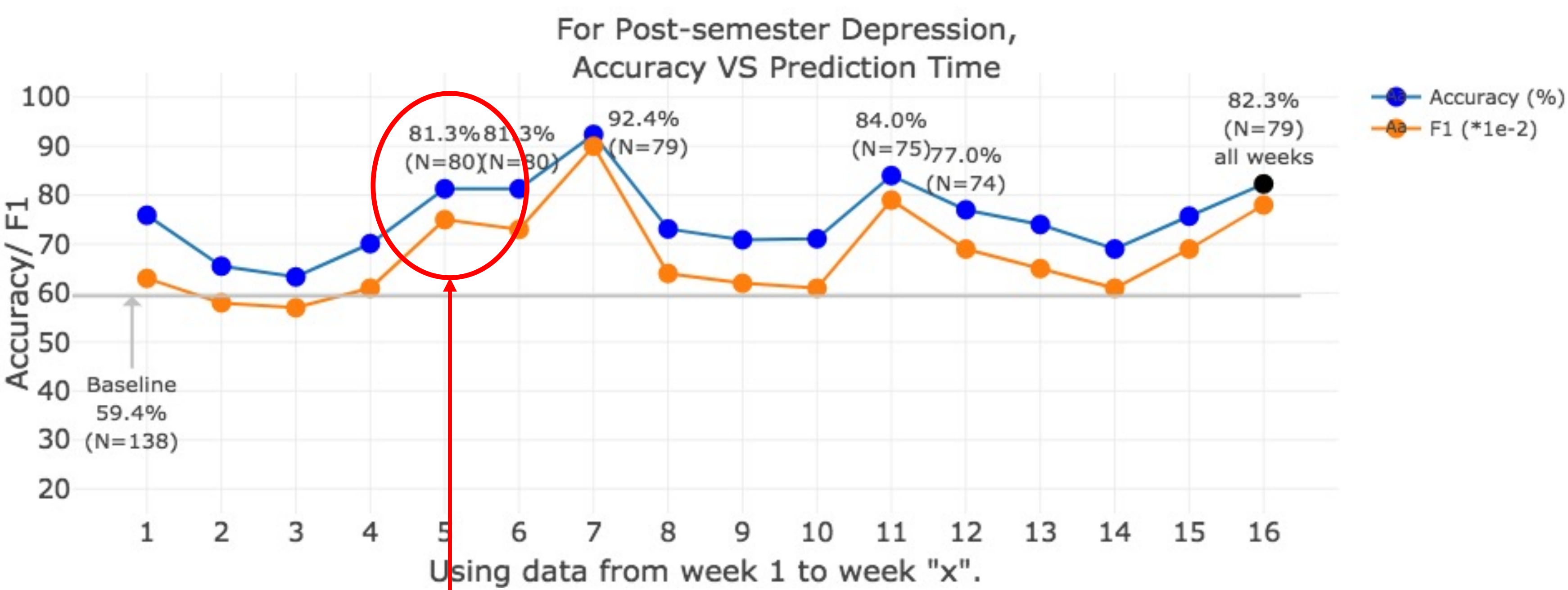
- Q) If our feature space is smaller, would NRLR work?
  - Inspired by the prediction/ forecasting problem



S2: Forecasting End of Semester Depression In College Students

- Study 2:
  - Forecasting NOT weekly prediction
  - Same data as study 1

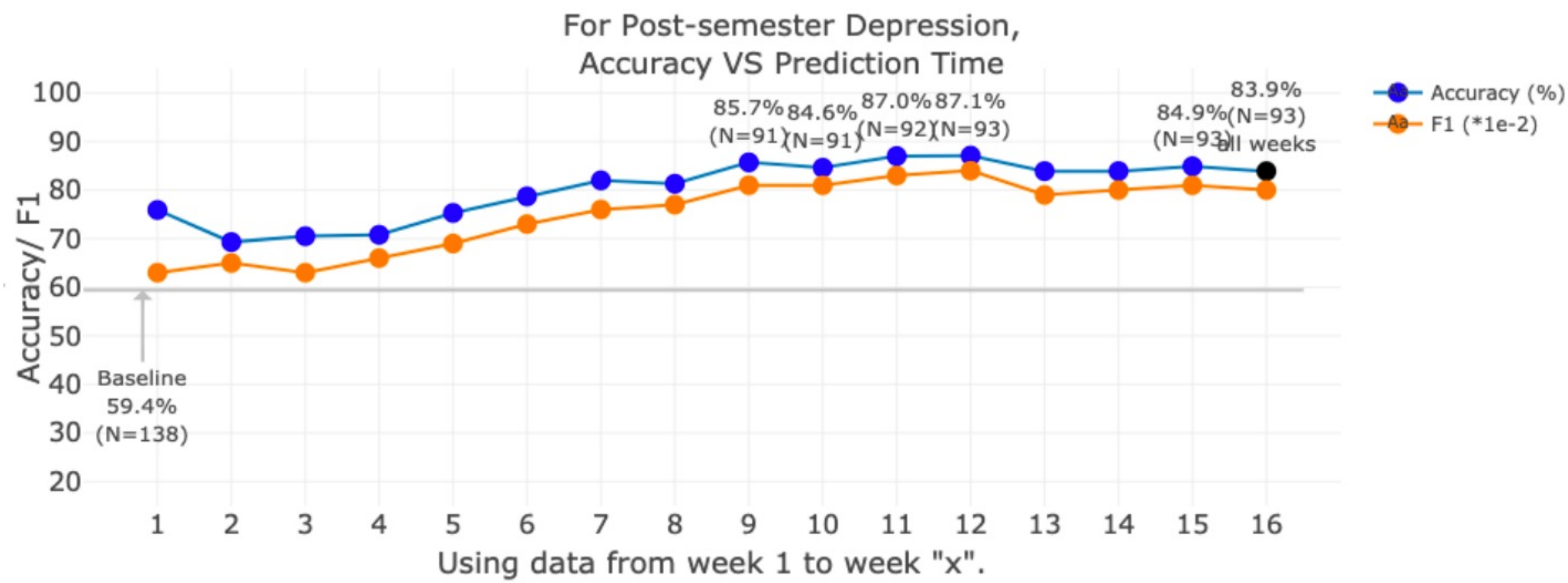
S2: Results



- We achieve an accuracy of 81.3% as early as the end of week 5.

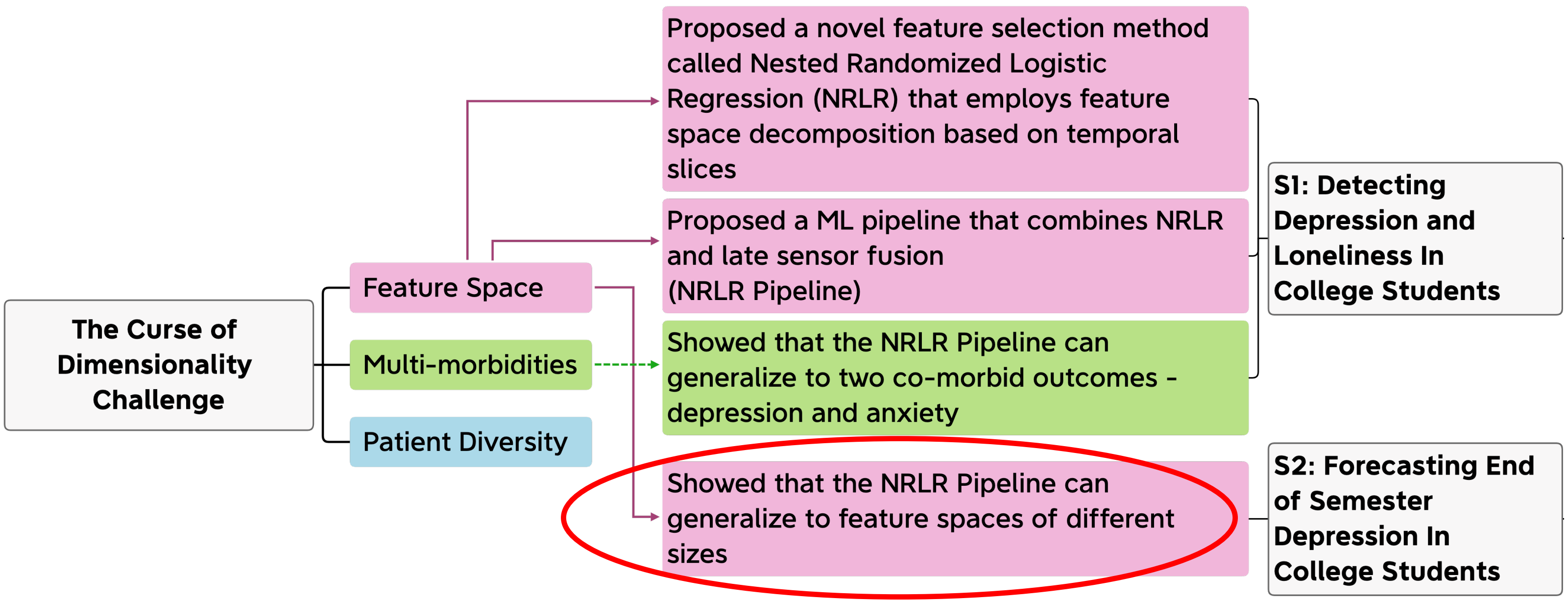
S2: Results

- After majority class voting:



S2: Addressing the Curse of Dimensionality

How does this address the curse of dimensionality?

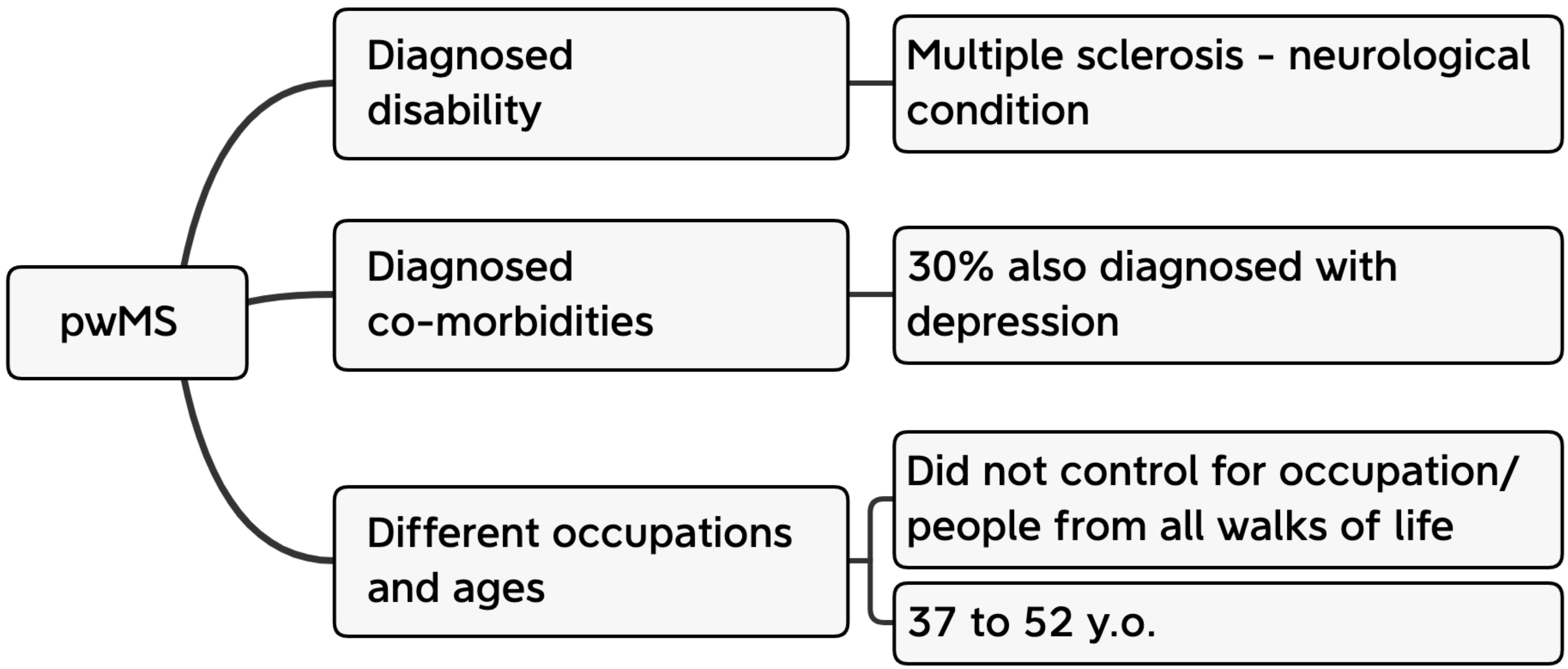


Thesis Outline

- Introduction
- The Curse of Dimensionality Challenge
- S1: Detecting Depression and Loneliness in College Students
- S2: Forecasting End of Semester Depression in College Students
- S3: Predicting the Mental Health of People with Multiple Sclerosis (MS) during the Covid-19 Stay-at-Home Period
- S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention
- S5: Predicting Periodically Assessed MS Outcomes Using Passively Sensed Behaviors and Ecological Momentary Assessments
- Thesis Contributions and Future Work

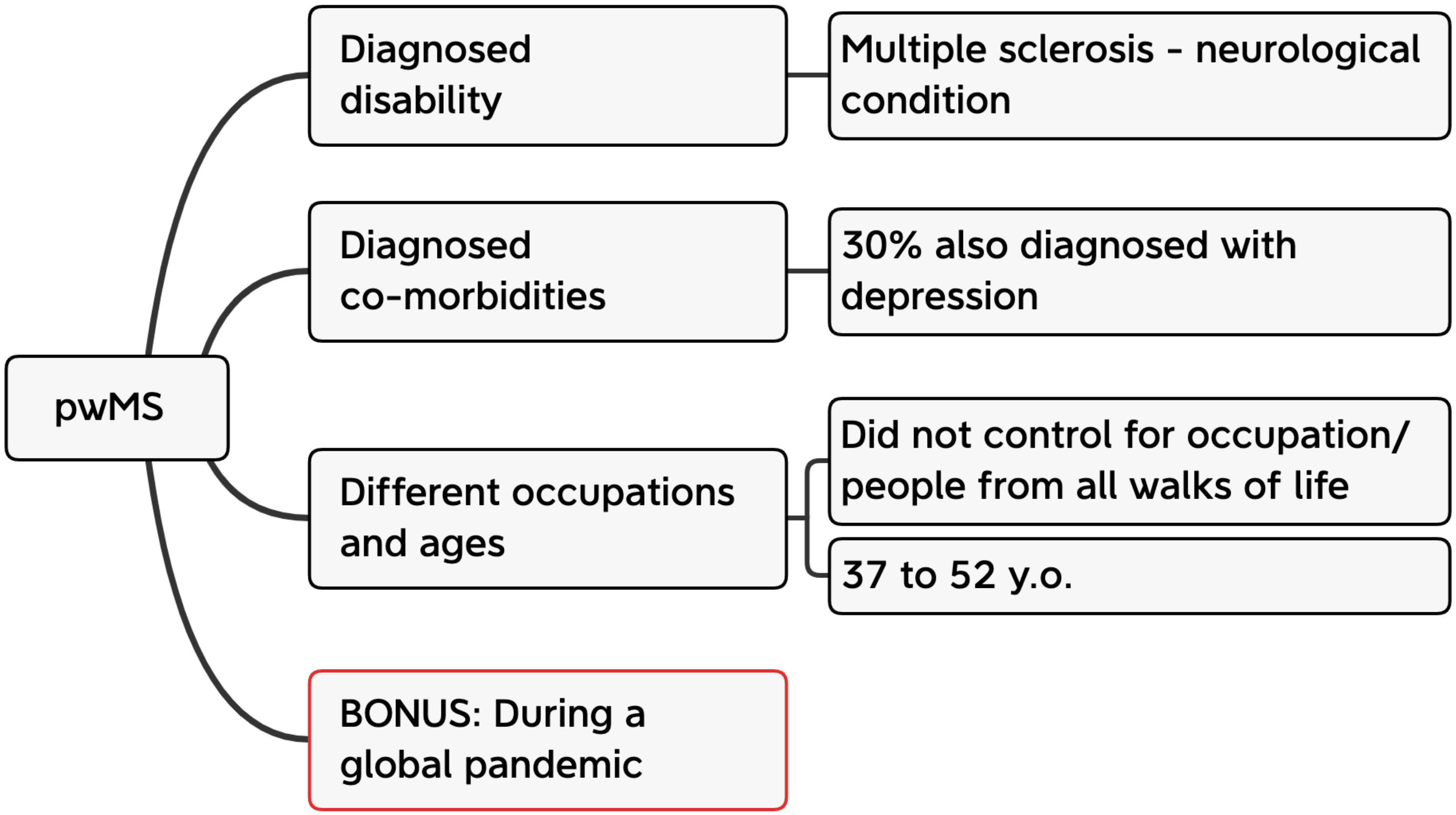
S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period

- Would NRLR generalize to a more complex population?
- Patients with Multiple Sclerosis (pwMS)



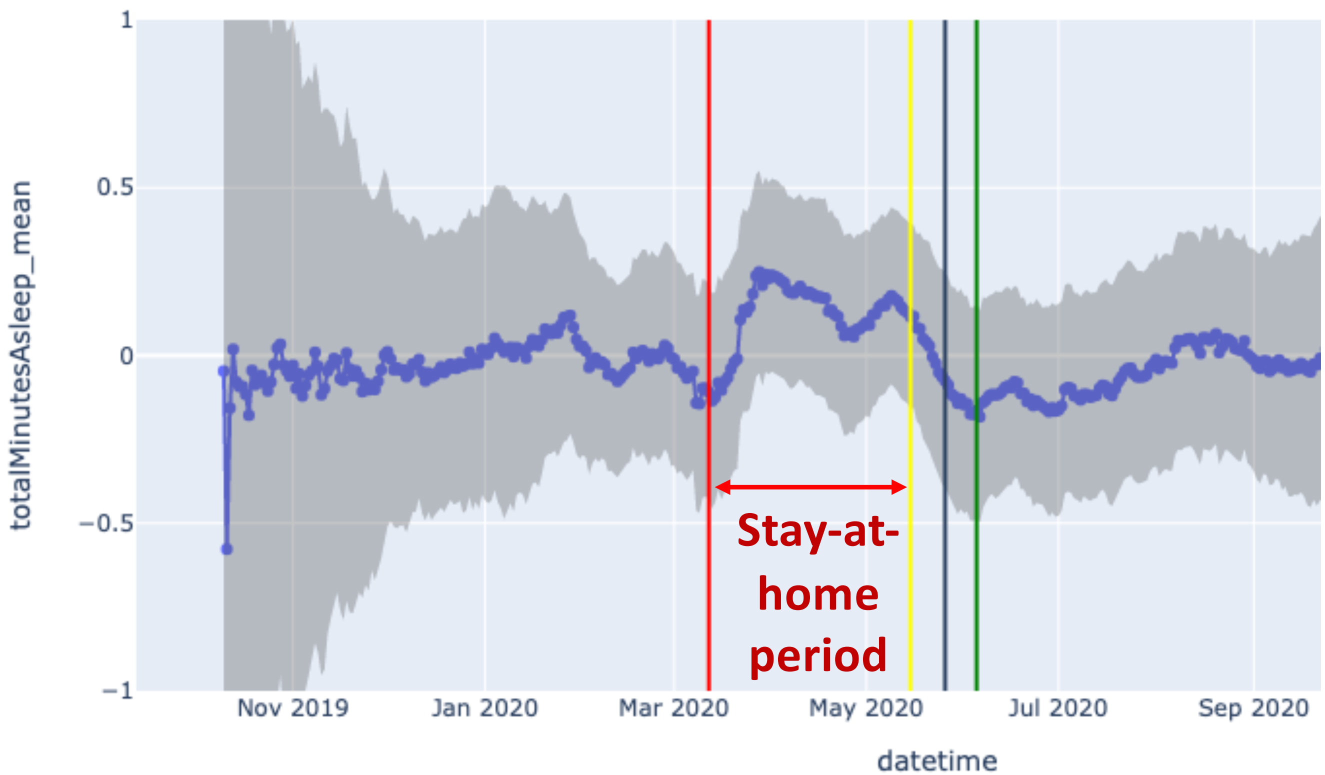
S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period

- Patients with Multiple Sclerosis (pwMS)



S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period

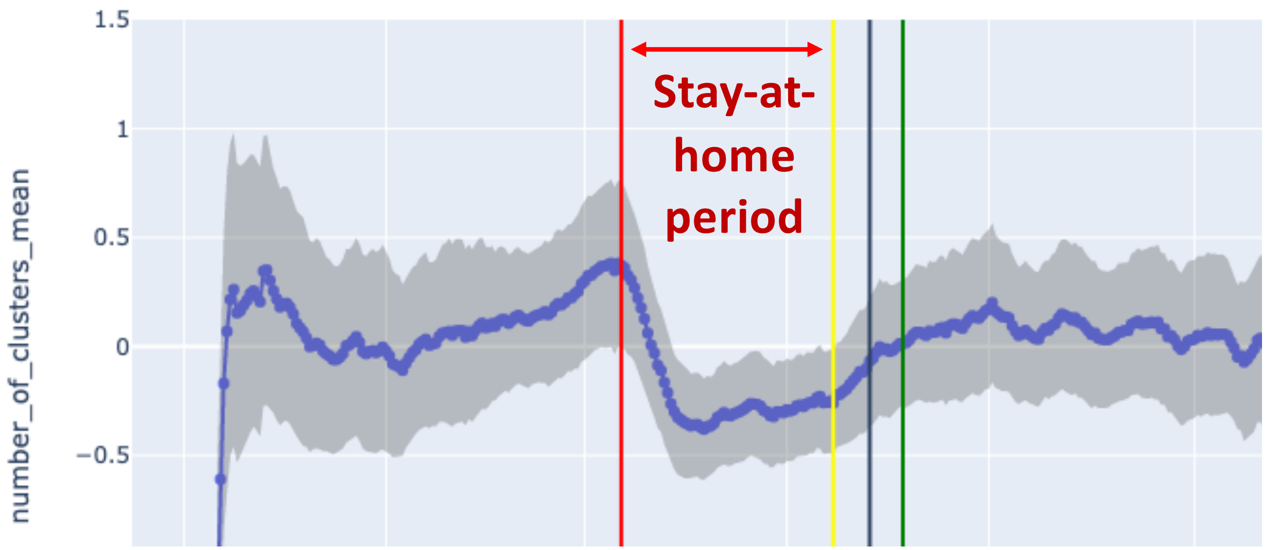
- We found that the stay-at-home period imposed during the pandemic had a major effect on people's behaviors.



Time spent asleep per day (normalized per person)

S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period

- We found that the stay-at-home period imposed during the pandemic had a major effect on people's behaviors.



Number of significant locations (normalized per person)

What does this mean for
Multimodal Behavioral Sensing research?

S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period Contd.

- Q) Can we use changes in behavior that occurred after the stay-at-home period was imposed to predict health outcomes during the stay-at-home period?

Change in behaviors before and after the "stay-at-home" order is instated

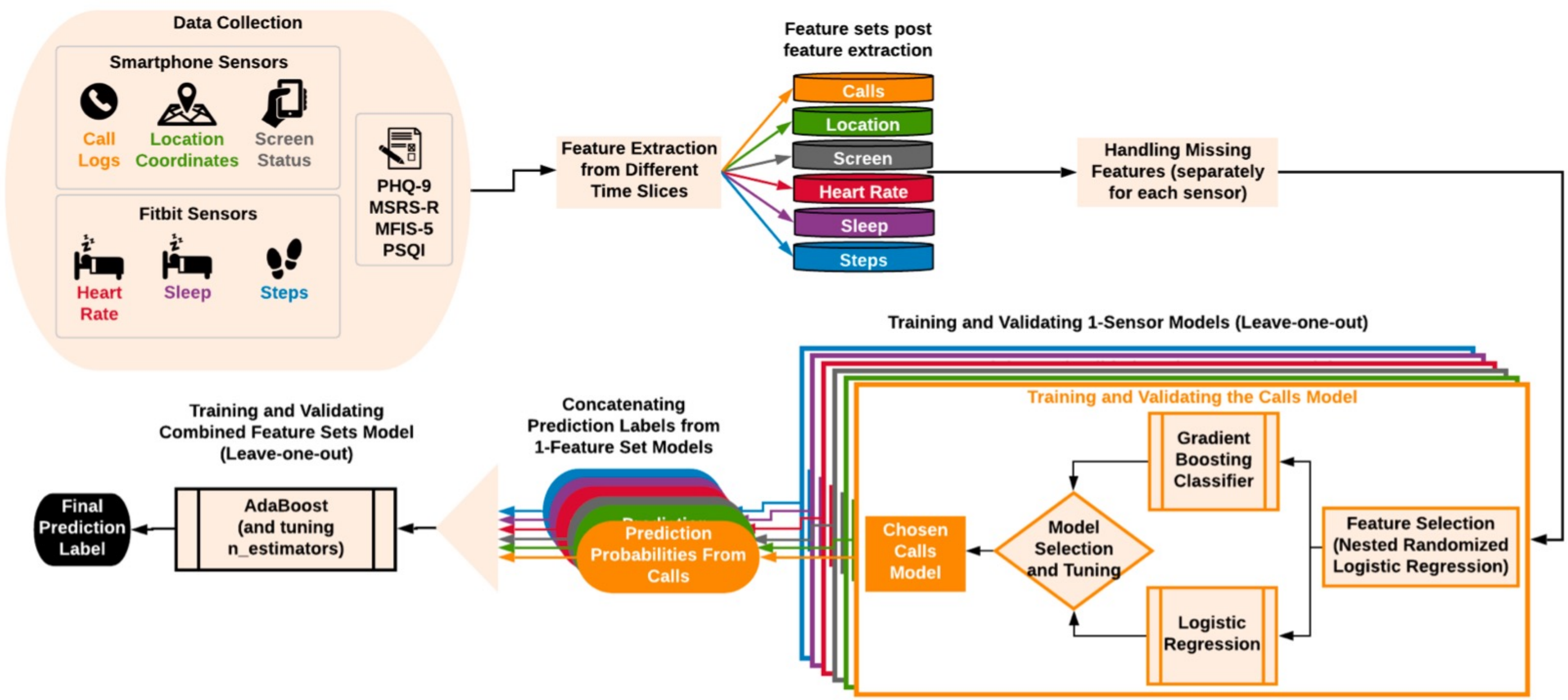
- Detect Depression
- Detect Fatigue
- Detect Global MS Symptom Burden
- Detect Sleep Quality

Multiple ground truth values are averaged over the Covid-19 stay at-home period to get outcomes

- 56 Patients with Multiple Sclerosis (pwMS)

S3: Methodology - Modeling

- Same modeling approach as before



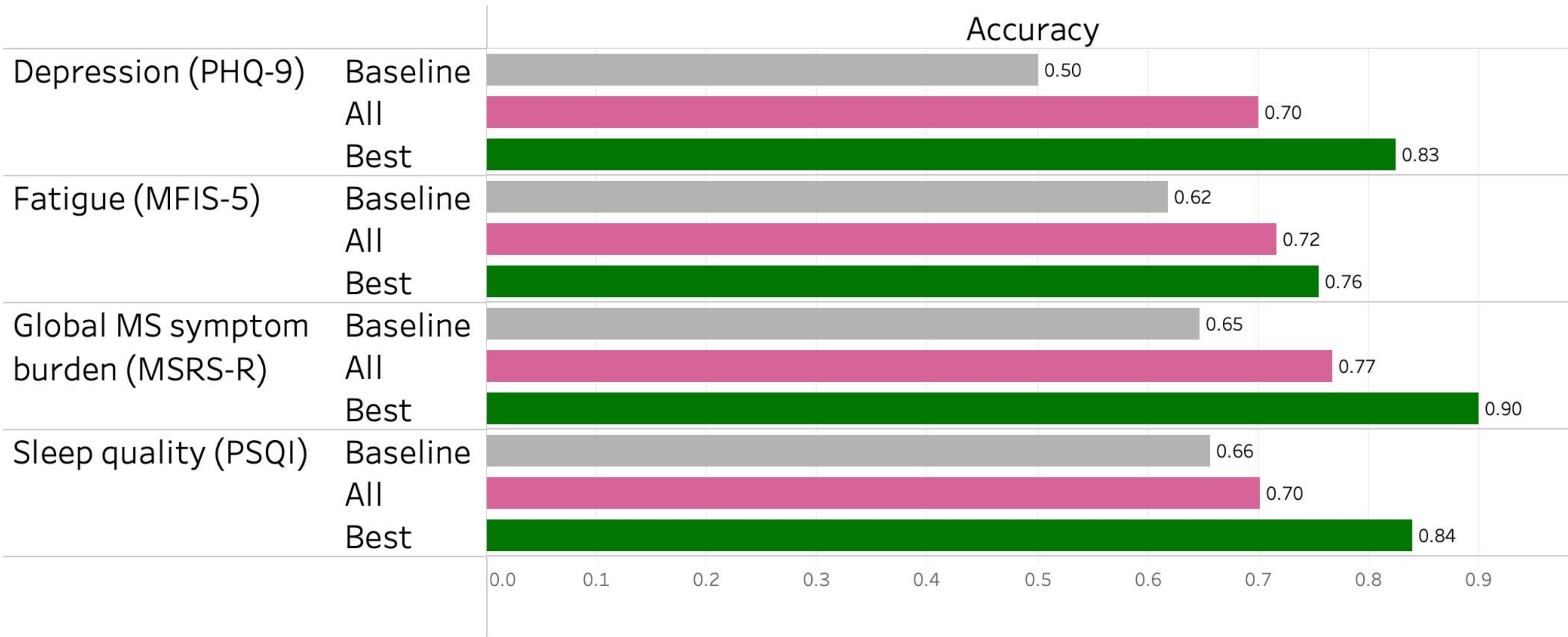
S3: Methodology – Feature Extraction

- Collected data from 6 sensors.
- For every sensor, extracted features from 15 time slices from the pre-covid-19 and stay-at-home periods.

Final Feature Matrix = Stay-at-home Feature Matrix – Pre-Covid-19 Feature Matrix

- E.g., for person A:
  - Average Steps Pre-Covid-19 = 7000
  - Average Steps during the Stay-at-Home Period = 4000
  - Final Feature = -3000

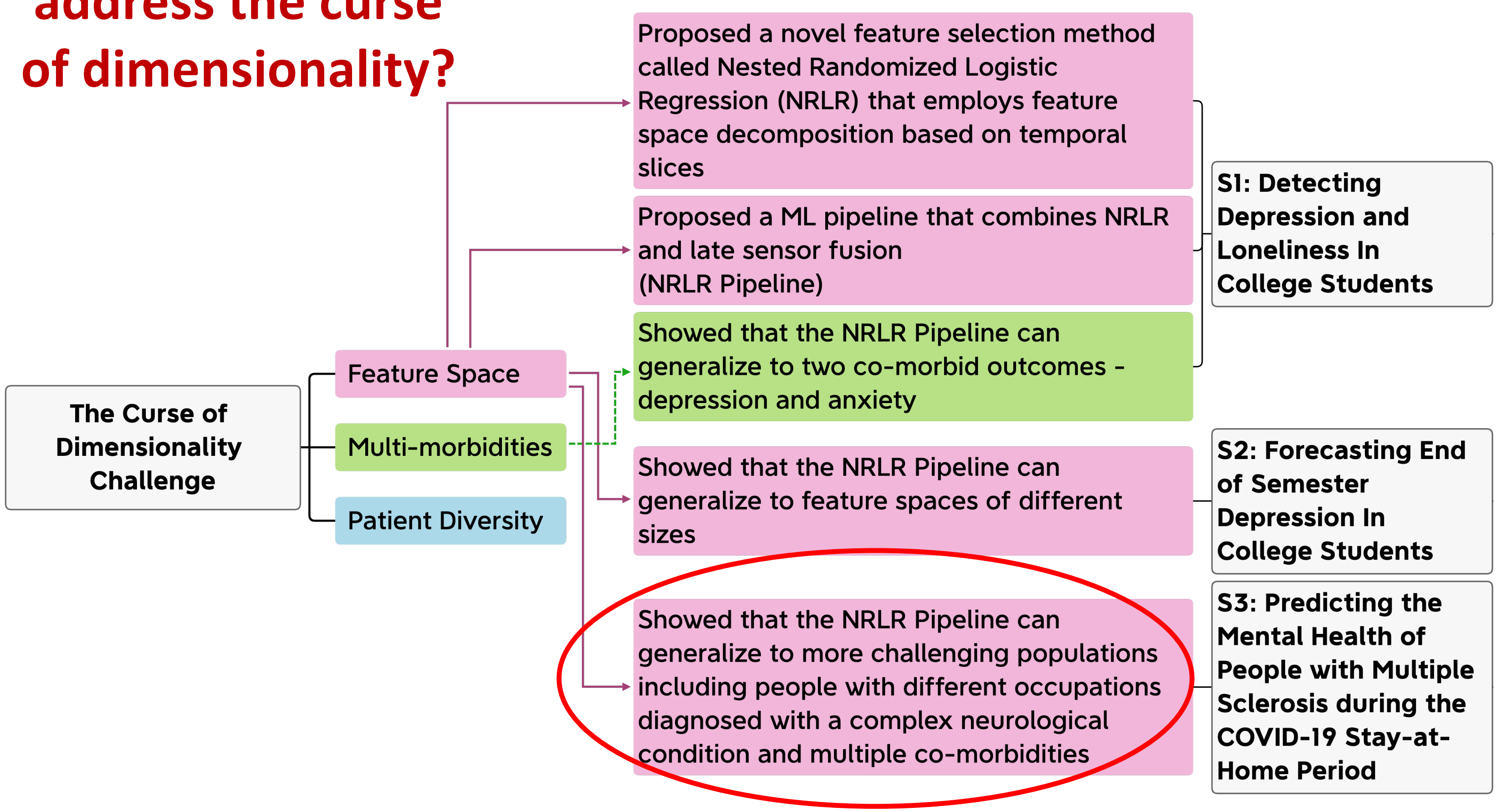
S3: Results



Baseline: Majority class
All: All 6 sensors
Best: Best combination of sensors

S3: Addressing the Curse of Dimensionality

How does this address the curse of dimensionality?



S3: Addressing the Curse of Dimensionality

How does this address the curse of dimensionality?

The Curse of Dimensionality Challenge

- Feature Space
- Multi-morbidities
- Patient Diversity

Proposed a novel feature selection method called Nested Randomized Logistic Regression (NRLR) that employs feature space decomposition based on temporal slices

Proposed a ML pipeline that combines NRLR and late sensor fusion (NRLR Pipeline)

Showed that the NRLR Pipeline can generalize to two co-morbid outcomes - depression and anxiety

S1: Detecting Depression and Loneliness In College Students

Showed that the NRLR Pipeline can generalize to feature spaces of different sizes

S2: Forecasting End of Semester Depression In College Students

Showed that the NRLR Pipeline can generalize to more challenging populations including people with different occupations diagnosed with a complex neurological condition and multiple co-morbidities

Showed that the NRLR Pipeline can generalize to 4 health outcomes that are frequently co-morbid in patients with MS

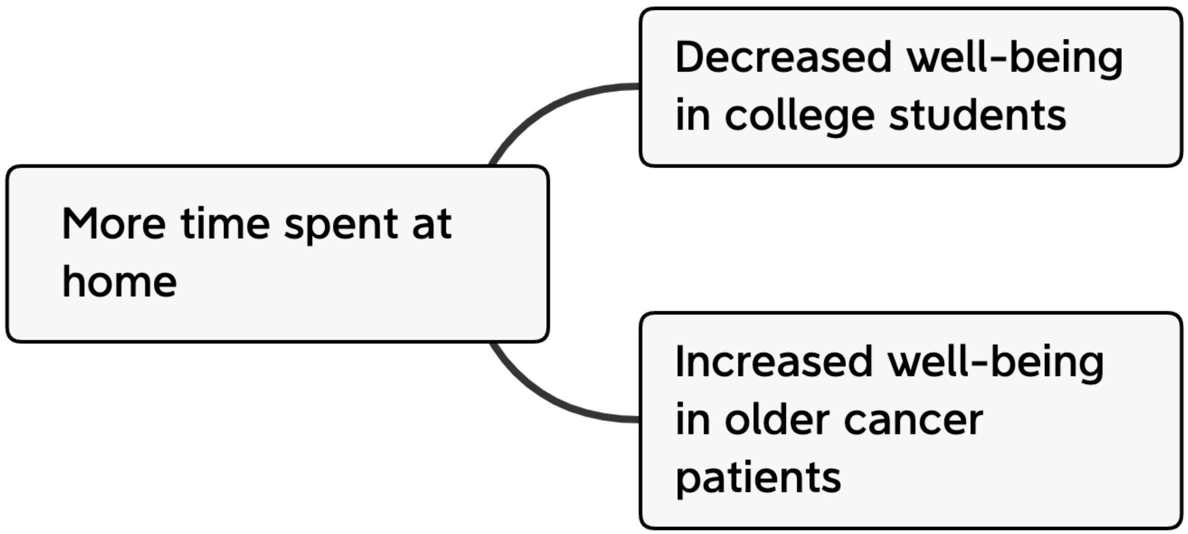
S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period

Thesis Outline

- Introduction
- The Curse of Dimensionality Challenge
- S1: Detecting Depression and Loneliness in College Students
- S2: Forecasting End of Semester Depression in College Students
- S3: Predicting the Mental Health of People with Multiple Sclerosis (MS) during the Covid-19 Stay-at-Home Period
- S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention
- S5: Predicting Periodically Assessed MS Outcomes Using Passively Sensed Behaviors and Ecological Momentary Assessments
- Thesis Contributions and Future Work

S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention

- **RECAP:** The relationship between behaviors and outcome may be dependent on patient context and characteristics.



- We have not yet addressed the curse of dimensionality in the diversity in patient context and characteristics.

S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention

- Further, interventions are an important part of this kind of work.
- Hence, study 4 focuses on interventions

- Analyzing behaviors in the context in which they occur, can lead to more interpretable insights,
  - Interventions require more trust and transparency.
  - So, interpretability is even more important for interventions.

- **S4's GOAL:** Analyze the users' interaction with a MH intervention app and the human supporters on the app to understand how supporter behaviors correlate with patient outcomes for patients in different contexts or situations.

S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention

- Refer to thesis for more information on this study.

Thesis Outline

- Introduction
- The Curse of Dimensionality Challenge
- S1: Detecting Depression and Loneliness in College Students
- S2: Forecasting End of Semester Depression in College Students
- S3: Predicting the Mental Health of People with Multiple Sclerosis (MS) during the Covid-19 Stay-at-Home Period
- S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention
- S5: Predicting Periodically Assessed MS Outcomes Using Passively Sensed Behaviors and Ecological Momentary Assessments
- Thesis Contributions and Future Work

Gaps so far

The Curse of Dimensionality Challenge

- Feature Space
- Multi-morbidities
- Patient Diversity

Proposed a novel feature selection method called Nested Randomized Logistic Regression (NRLR) that employs feature space decomposition based on temporal slices

Proposed a ML pipeline that combines NRLR and late sensor fusion (NRLR Pipeline)

Showed that the NRLR Pipeline can generalize to two co-morbid outcomes - depression and anxiety

S1: Detecting Depression and Loneliness In College Students

Showed that the NRLR Pipeline can generalize to feature spaces of different sizes

S2: Forecasting End of Semester Depression In College Students

Showed that the NRLR Pipeline can generalize to more challenging populations including people with different occupations diagnosed with a complex neurological condition and multiple co-morbidities

Showed that the NRLR Pipeline can generalize to 4 health outcomes that are frequently co-morbid in patients with MS

S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period

Defined a "combined" target outcome based on both depression and anxiety, in order to identify best intervention strategies that improve both depression and anxiety

Used a statistical approach to identify the best "general" intervention strategies that work for clients across all contexts.

Proposed a data mining approach to identify the best "specific" intervention strategies that work in narrow client contexts

S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention

Have not yet leveraged context for detecting MH outcomes → can increase performance

Gaps so far

Conditions like MS → Need for enabling periodic monitoring using passive sensors.

None of these studies enable periodic monitoring of mental health outcomes over time.

Periodic monitoring might also need us to account for the patient's historical context.

S1: Detecting Depression and Loneliness In College Students

S2: Forecasting End of Semester Depression In College Students

S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period

S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention

S5: Predicting Periodically Assessed MS Outcomes Using Passively Sensed Behaviors and EMAs

- Q) Can we use recent behaviors and past behaviors to predict health outcomes periodically to enable health monitoring?

Recent behaviors (action features) + Past behaviors (context features)

Detect Depression — Every 2 weeks

Detect Fatigue

Detect Global MS Symptom Burden

Detect Sleep Quality

Every 4 weeks

- 104 Patients with Multiple Sclerosis (pwMS) for 12 weeks
  - 44 out of 104 extended to 24 weeks.

S5: Predicting Periodically Assessed MS Outcomes Using Passively Sensed Behaviors and EMAs

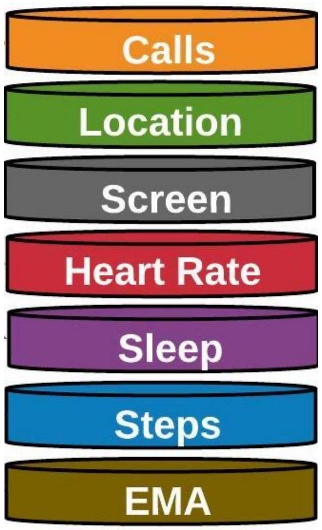
- In addition to passively sensed features, we include participants' answers to two short questions asked three times a day:
  - How depressed do you feel?
  - How tired do you feel?

  Likert scale 0 to 4 (least to most depressed/tired)

- A.K.A. Ecological Momentary Assessments (EMAs).
- Commonly used for "Repeated sampling of current behaviors and experiences in real-time in their natural environment."
- Sent to participants via mobile app and take 15s to complete.

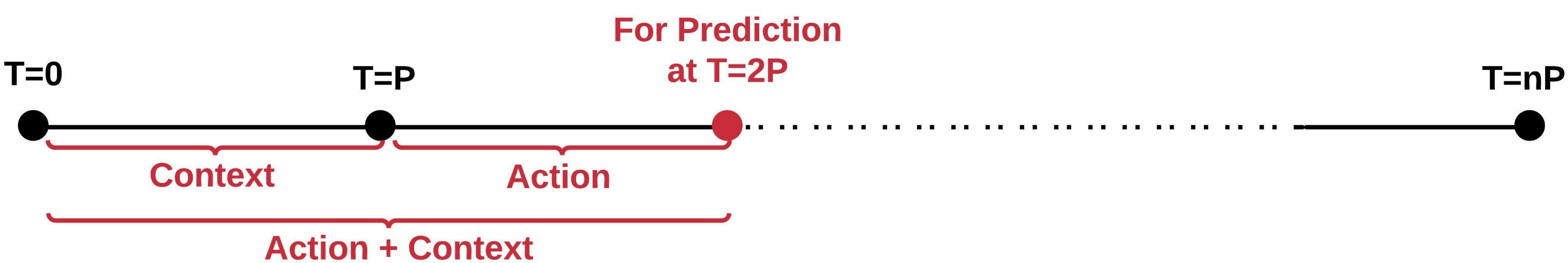
- We wanted to evaluate if adding EMA improves model performance.

S5: Methodology – Feature Extraction



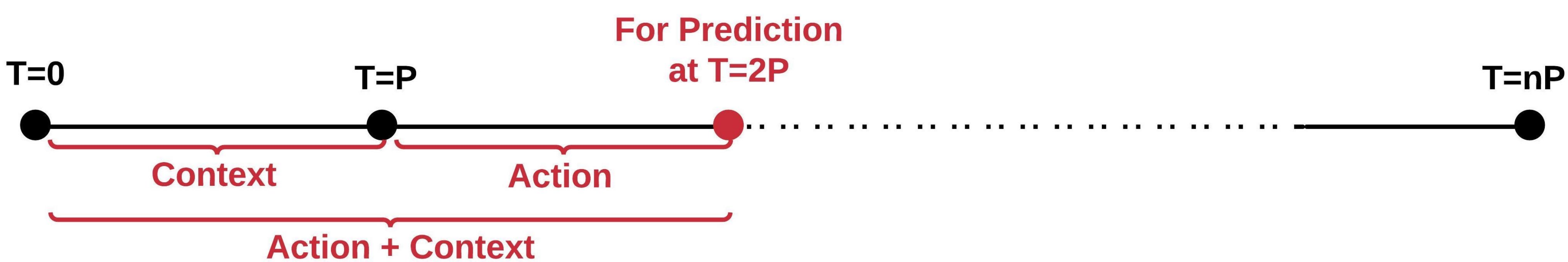
- Collected data from 6 sensors + EMA
- For every sensor, extracted features from 15 time slices from the action and context periods of each prediction time point.

P = { 2 weeks for Depression
4 weeks for Global MS Symptom Burden, Fatigue, and Sleep Quality

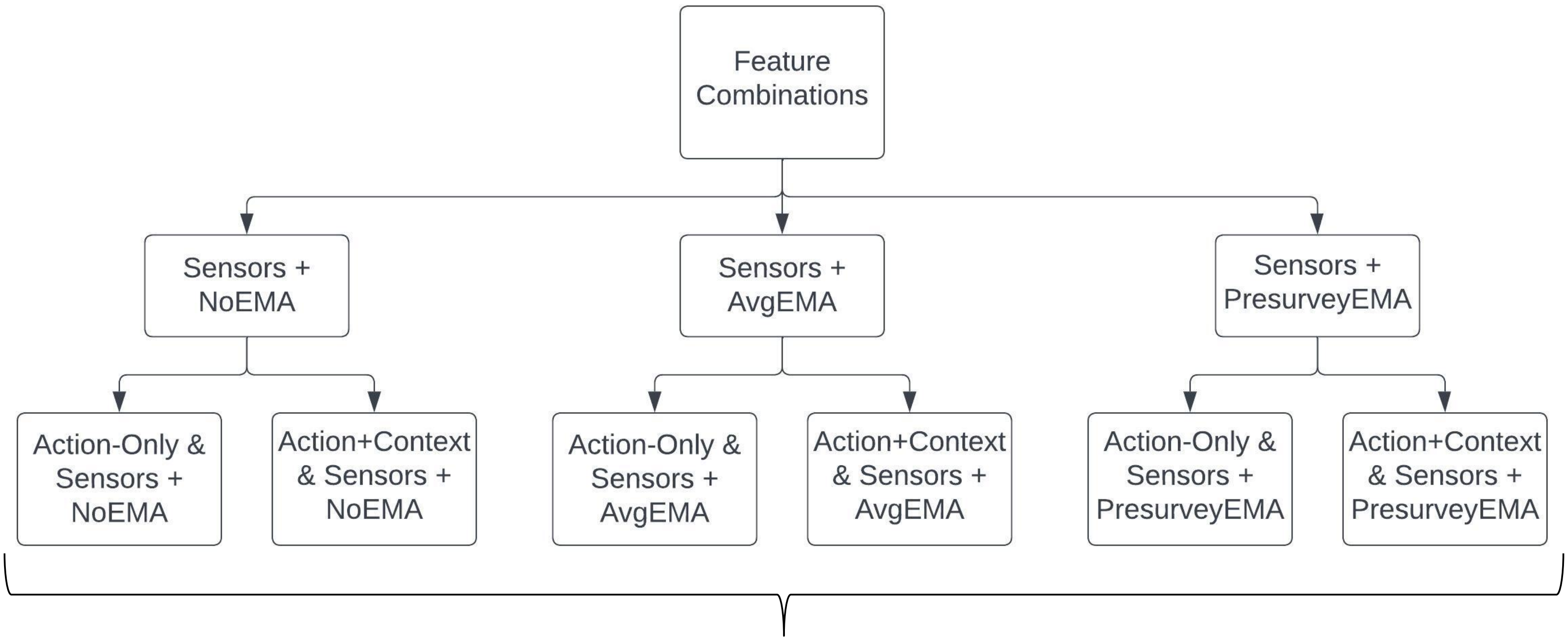


S5: Methodology – Feature Extraction

- For every action and context period, we computed two types of EMAs:
  - Average EMA: For every period, average all EMAs from each of the 15 time slices.
  - Pre-survey EMA: For every period, take the last EMA in the period from each of the 15 time slices.



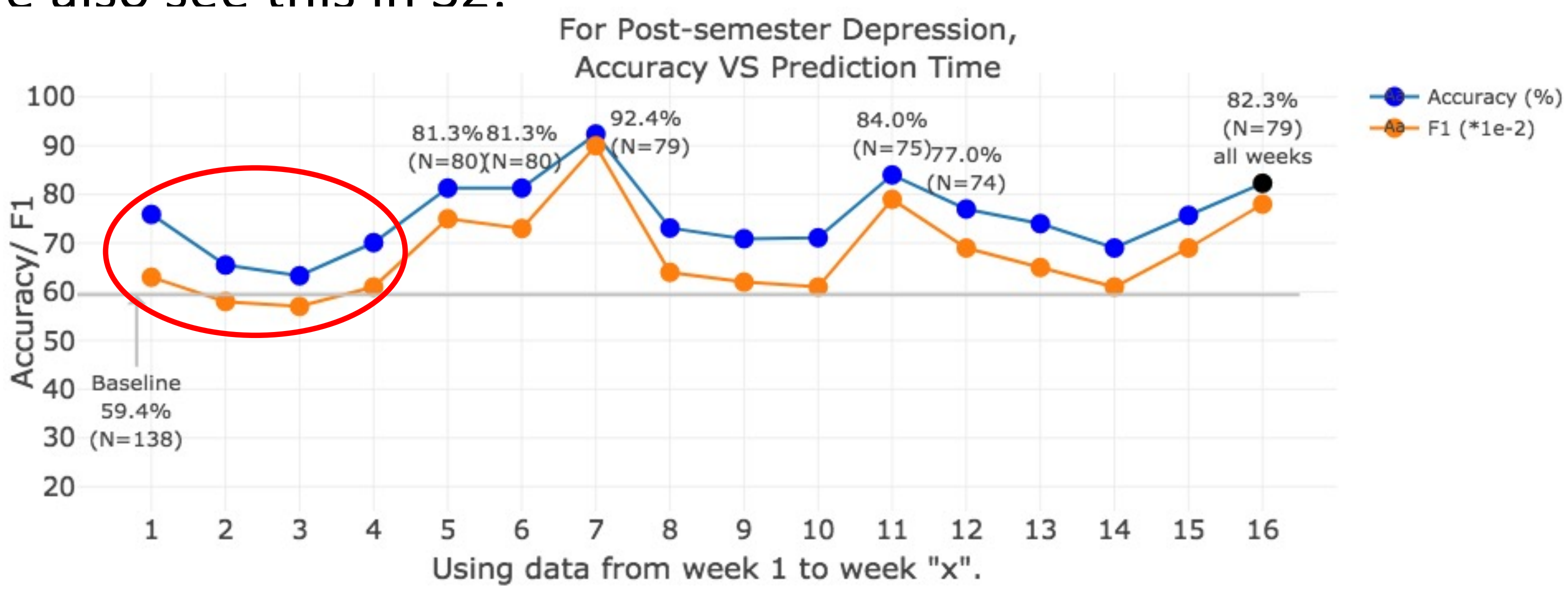
S5: Methodology – Feature Sets to Try



We developed 6 best models per outcome. Each best model combines specific sensors (same as S1-S3).

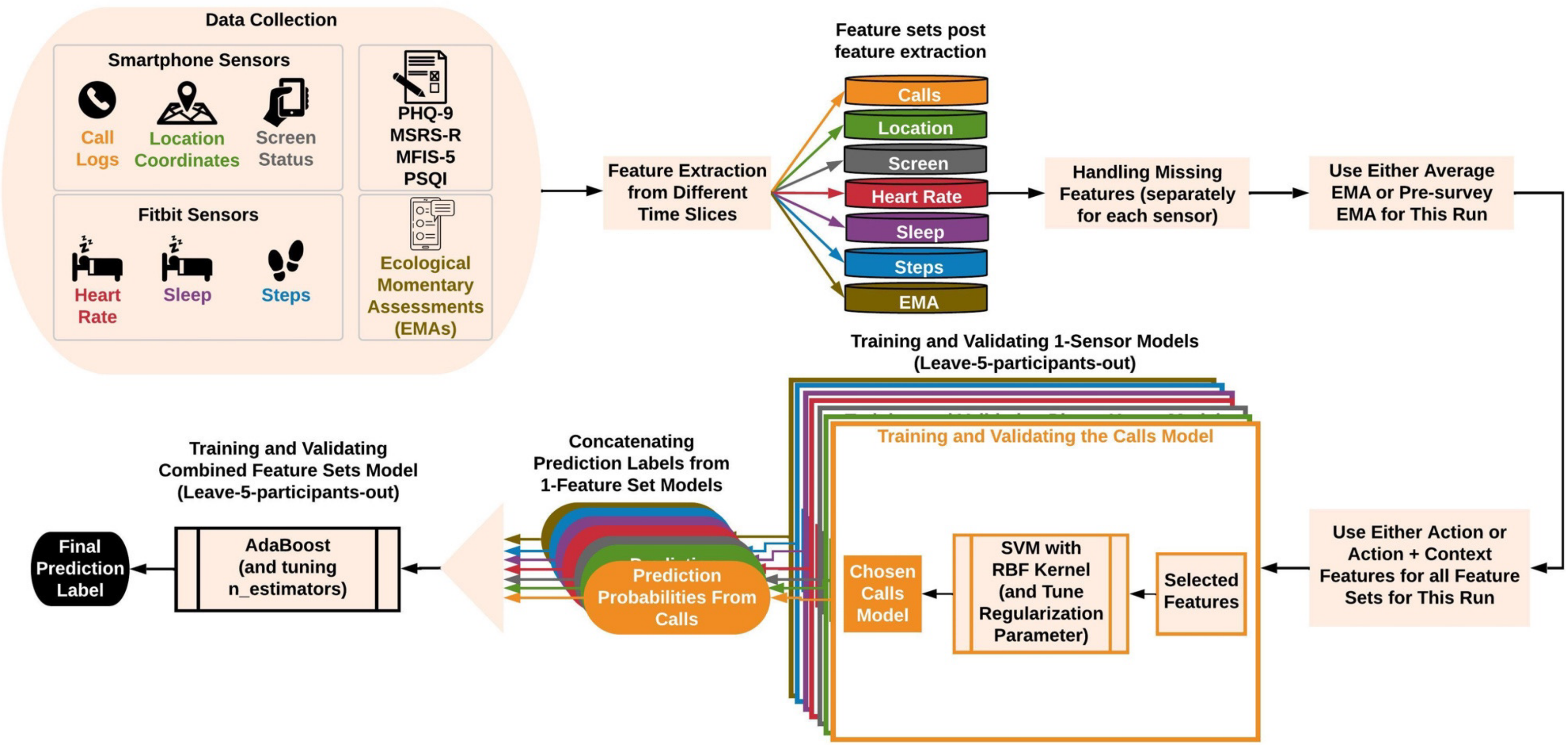
S3: Methodology – Trying the NRLR Approach

- Tried NRLR on feature vectors from every 2 or 4 weeks to detect the corresponding outcomes.
- Accuracy for best models was only 55-70% across all 4 outcomes.
- NRLR needs more data for good model performance and 2-4 weeks may not suffice.
- We also see this in S2:



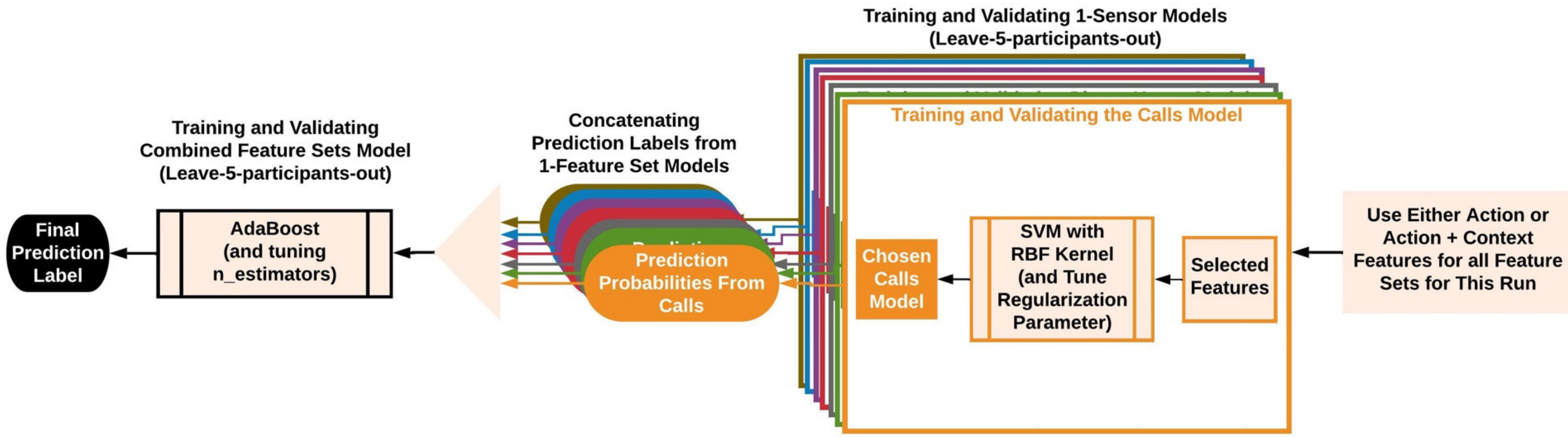
S5: Methodology – Modeling

- Hence, we tried different algorithms and found that SVM with Radial Basis Function (RBF) Kernel performed the best.



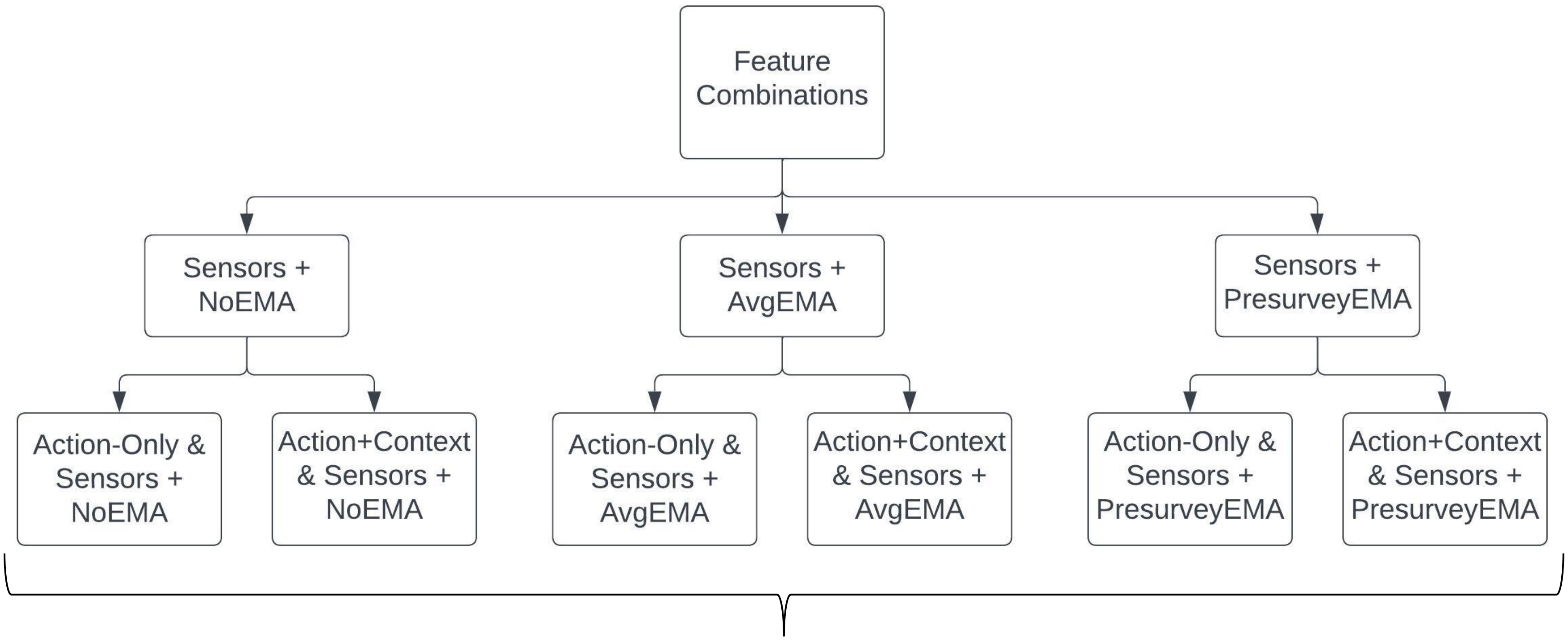
S5: Methodology – Modeling

- Instead of NRLR feature selection, followed by logistic regression or gradient boosting classifier, we used SVM + RBF on our selected feature set.



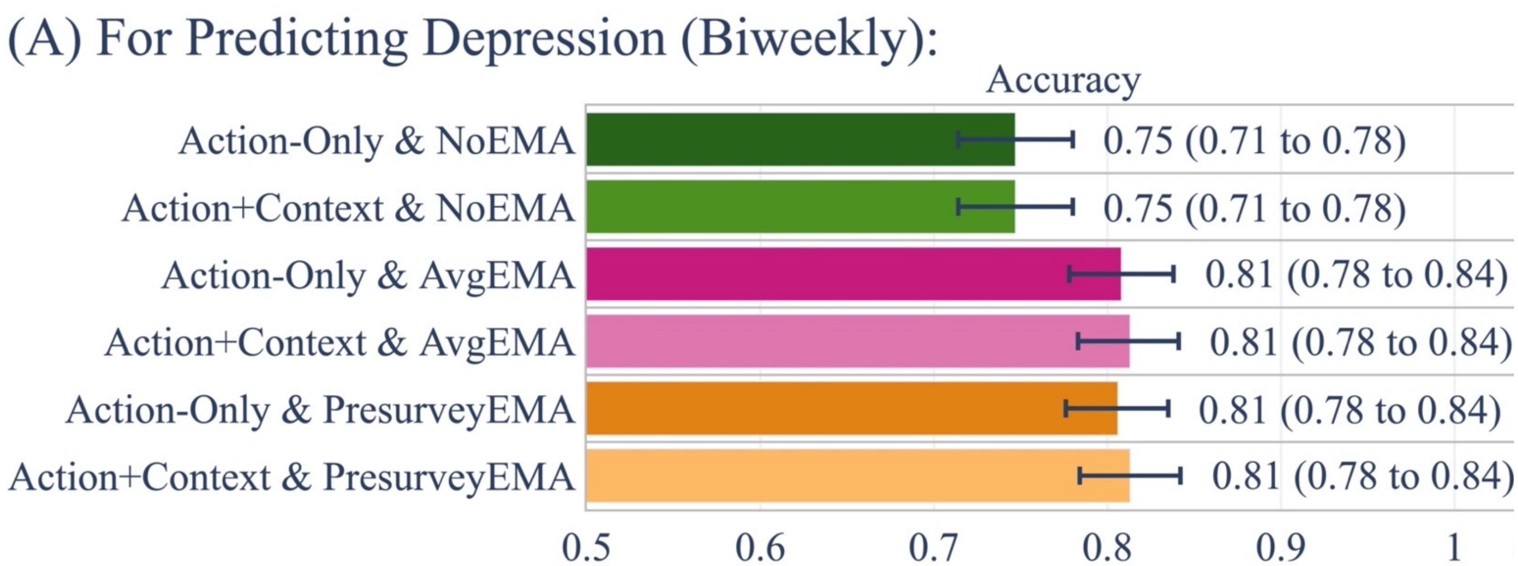
- Similar to S1-S3, we try different combination of sensors to get the best model for each outcome.

S5: Methodology – Comparing models



We compared these models by computing the 95% confidence intervals of differences in their bootstrapped accuracy and F1-scores.

S5: Results – Biweekly Depression

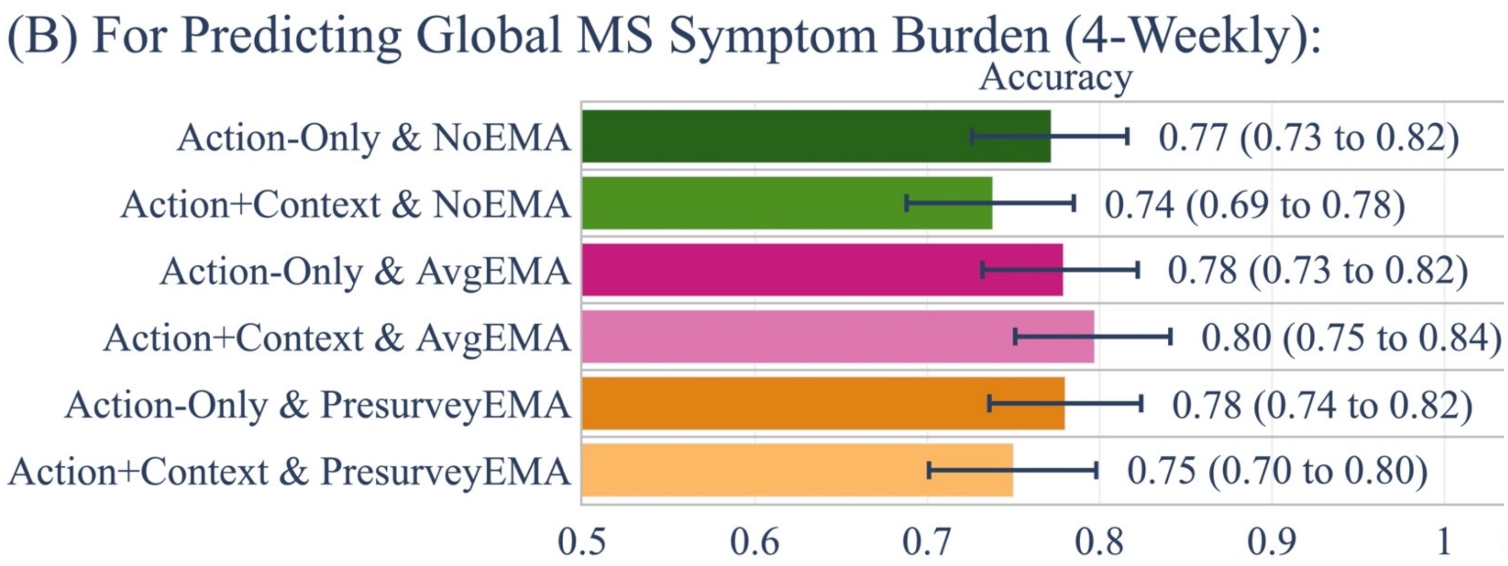


Baseline: 59.5% (majority class – no depression)

Statistically Best Model: Action-Only & PresurveyEMA

- Best performance while requiring the least amount of EMA.
- Accuracy: 80.6% - a 35.5% improvement over baseline.
- Combination: heart rate, steps, and pre-survey EMA

S5: Results – 4-Weekly Global MS Symptom Burden

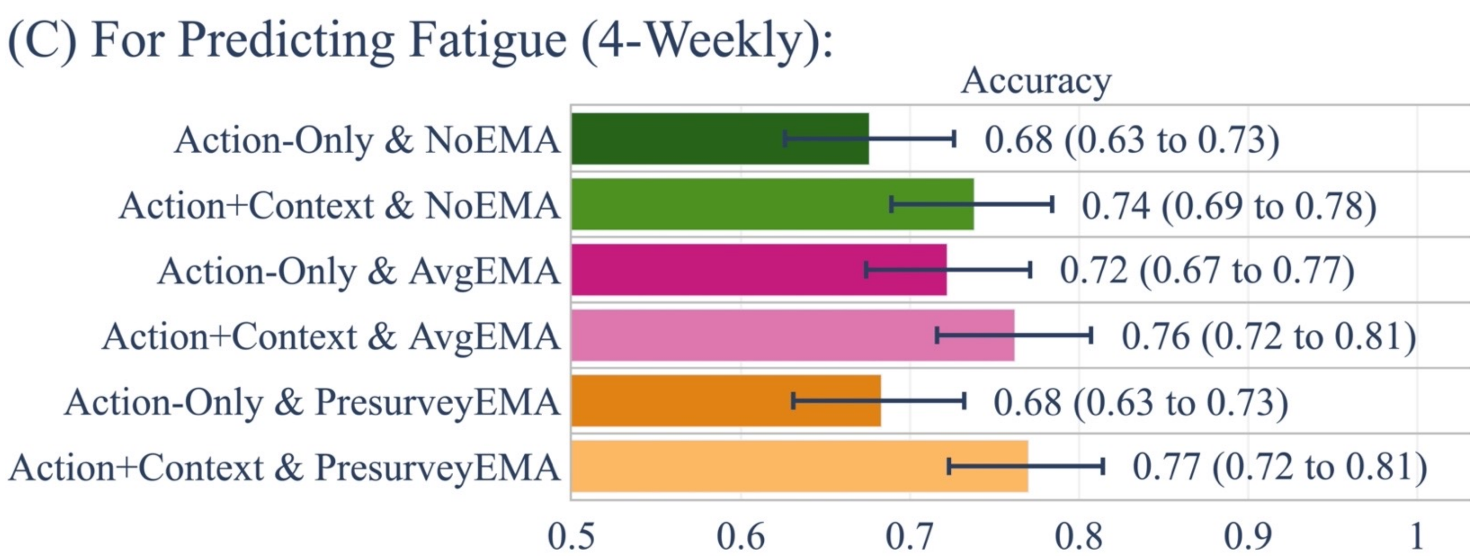


Baseline: 51.1% (majority class – high burden)

Statistically Best Model: Action-Only & NoEMA

- Best performance while requiring no EMA
- Accuracy: 77.3% - a 51.3% improvement over baseline.
- Combination: heart rate, location, sleep, and steps.

S5: Results – 4-Weekly Fatigue

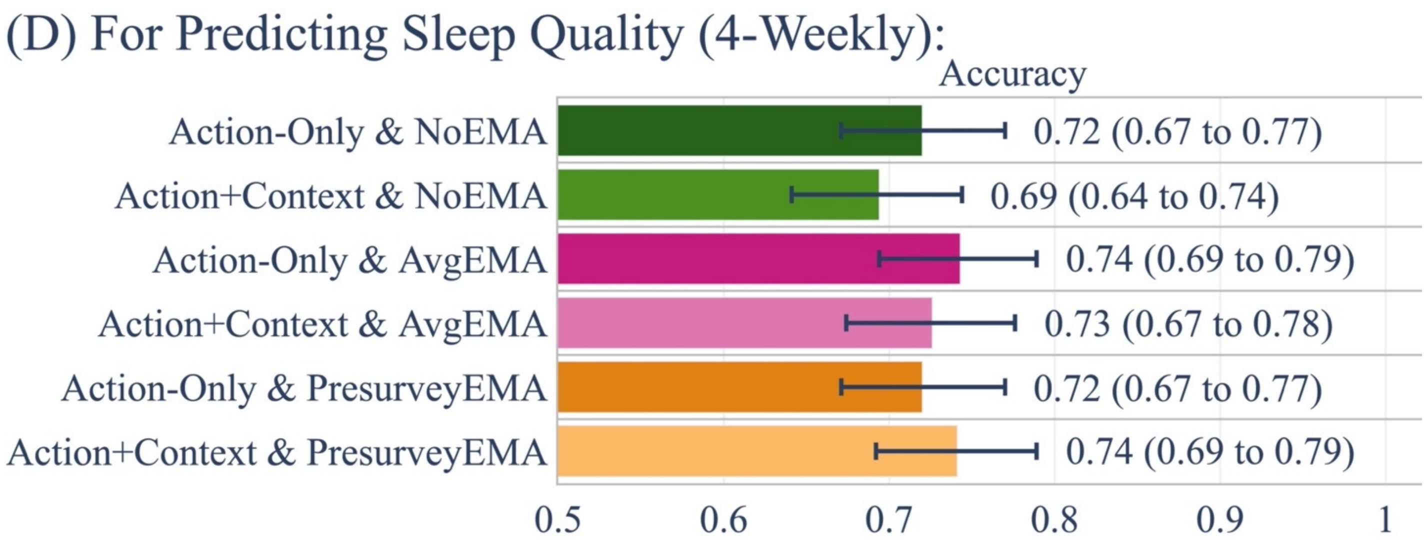


Baseline: 50.9% (majority class – severe fatigue)

Statistically Best Model: Action+Context & NoEMA

- Best performance while requiring no EMA
- Accuracy: 73.8% - a 45% improvement over baseline.
- Combination: heart rate, screen, and steps.

S5: Results – 4-Weekly Sleep Quality



Baseline: 56.2% (majority class – better sleep quality)

Statistically Best Model: Action-Only & NoEMA

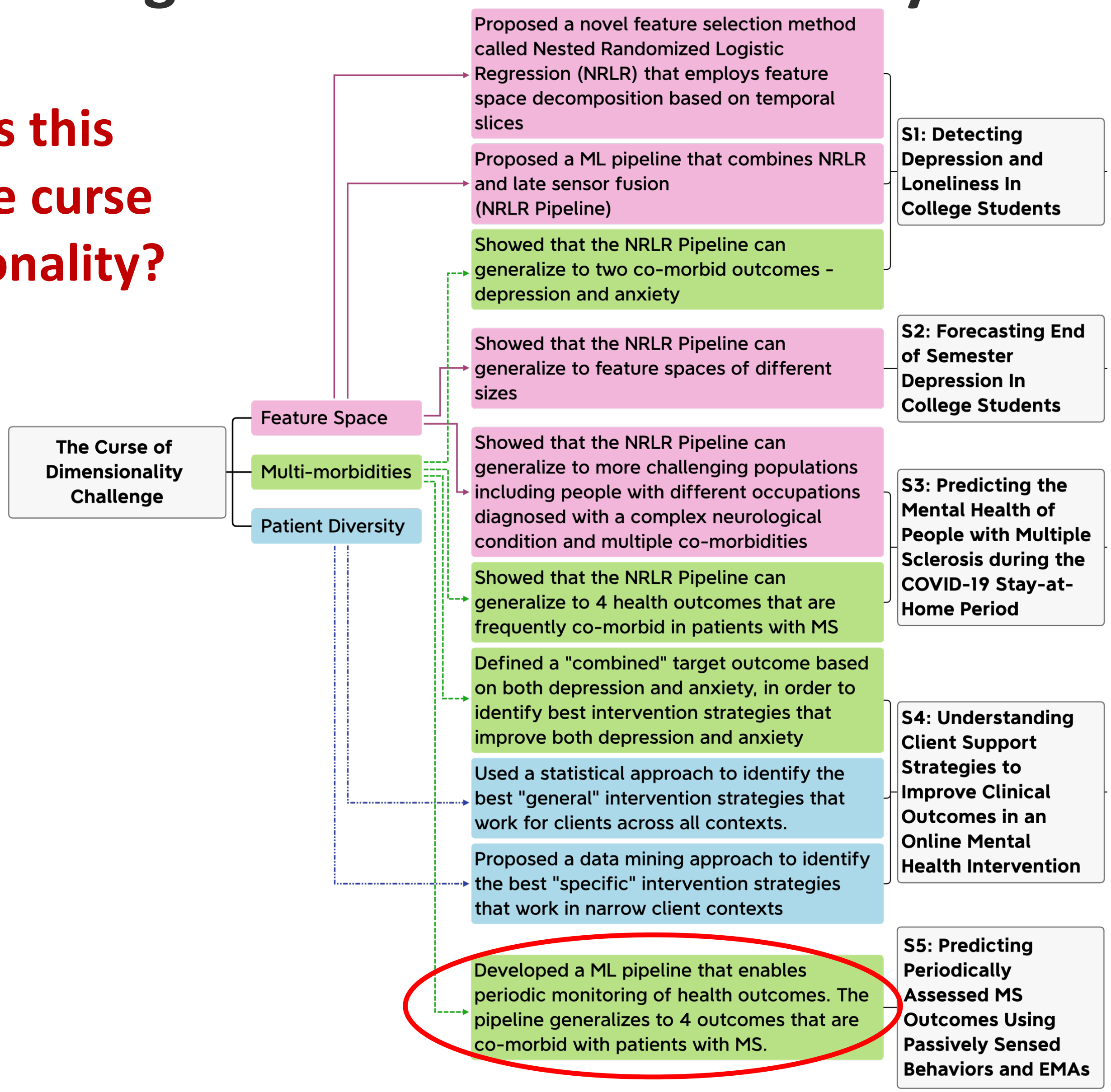
- Best performance while requiring no EMA
- Accuracy: 72.0% - a 28.1% improvement over baseline.
- Combination: heart rate, location, sleep, and steps.

S5: Discussion

- EMA improved performance for biweekly depression.
  - But we only needed pre-survey EMA.
- EMA did not improve model performance for other outcomes.
- Context or past features helped in prediction of fatigue.
  - Hence, action + context features should be considered.
- Similar sensors were selected for best combinations for multiple outcomes.
  - Heart rate and steps were selected for all outcomes.
  - Best for Global MS symptom burden and sleep quality: heart rate, location, sleep, and steps.
  - May not need to include all 6 sensors.

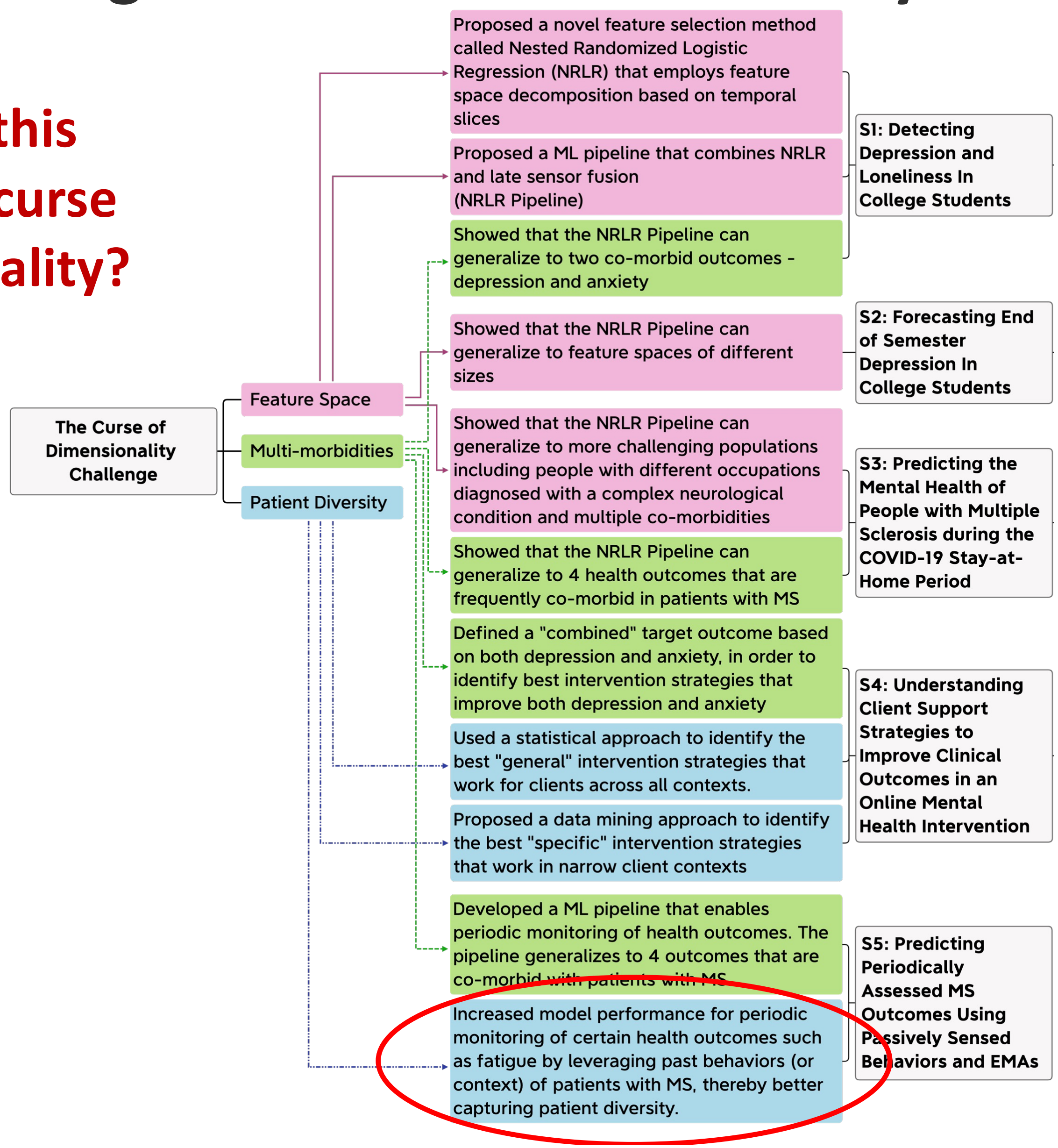
S5: Addressing the Curse of Dimensionality

How does this address the curse of dimensionality?



S5: Addressing the Curse of Dimensionality

How does this address the curse of dimensionality?



Thesis Outline

- Introduction
- The Curse of Dimensionality Challenge
- S1: Detecting Depression and Loneliness in College Students
- S2: Forecasting End of Semester Depression in College Students
- S3: Predicting the Mental Health of People with Multiple Sclerosis (MS) during the Covid-19 Stay-at-Home Period
- S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention
- S5: Predicting Periodically Assessed MS Outcomes Using Passively Sensed Behaviors and Ecological Momentary Assessments
- Thesis Contributions and Future Work

Thesis Contributions

1. Contributed a feature selection approach (NRLR) that mitigates the curse of dimensionality in the feature space by decomposing the feature space during feature selection. Enables detection and early prediction with limited ground truth, thus reducing survey burden on the user.
2. Presented a ML pipeline (the NRLR Pipeline) that can be used to detect multiple co-morbid health outcomes in homogenous and complex populations.
3. Presented an approach that combines multiple outcomes into one final outcome, that can then be used to understand or personalize an intervention.

Thesis Contributions

4. Presented an approach that leverages the patient's context to analyze and identify "general" intervention strategies that work across multiple contexts, and "specific" intervention strategies that work in certain contexts.
5. Developed a machine learning pipeline that leverages recent and past behaviors to periodically predict health outcomes that are frequently comorbid in patients with MS.
6. Demonstrated that adding EMAs to models using passively sensed behaviors to predict health outcomes does not significantly improve performance for most outcomes.

Future Work

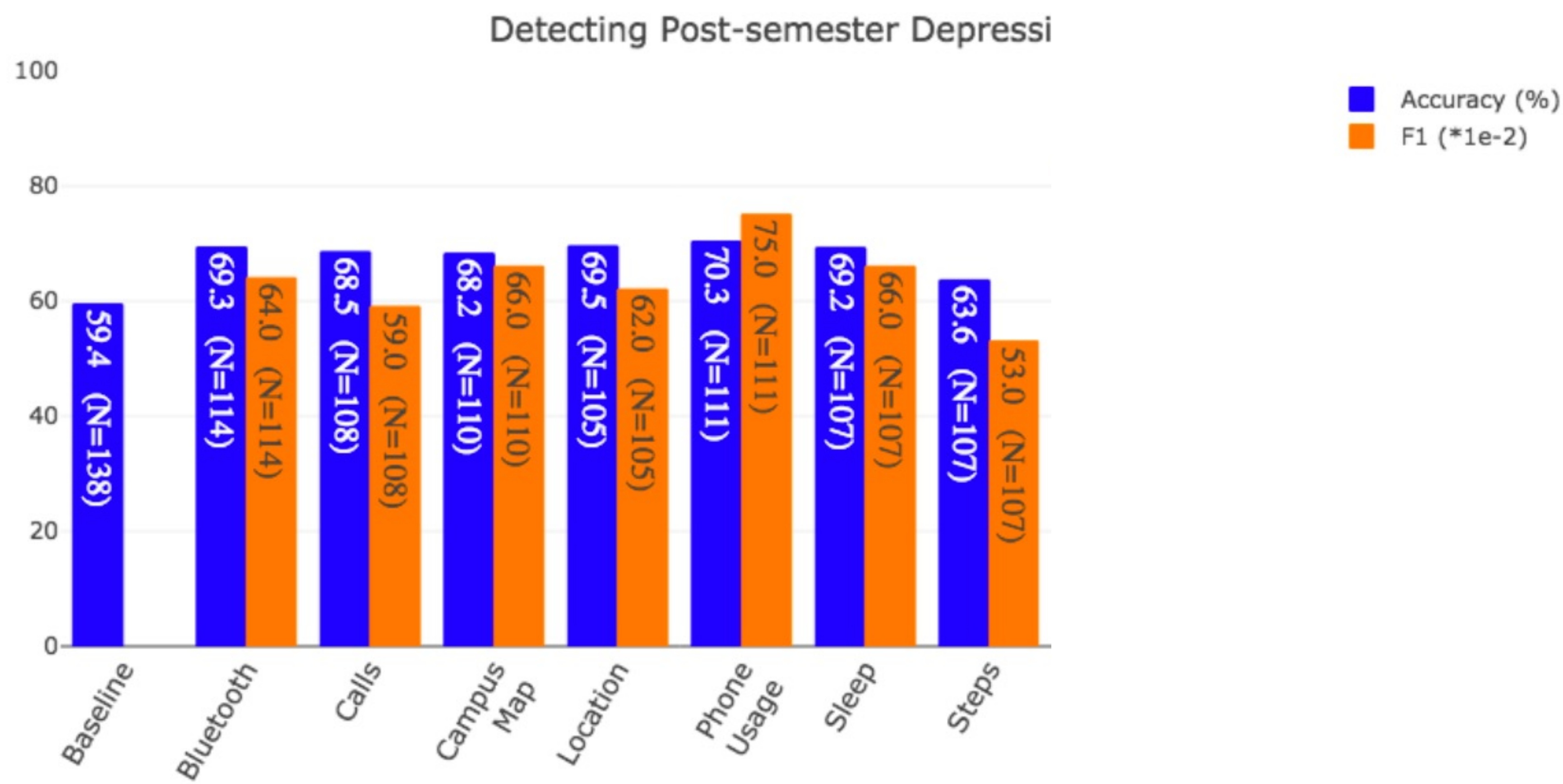
1. Evaluate the generalizability and robustness of our methods on larger sample size.
2. Deploy "live" models to assess the acceptability of predictions amongst patients and clinicians, and address barriers.
3. Use patient feedback on predictions from "live" models to further improve them and understand predictions.
4. Enable the detection of specific patient phenotypes (e.g. moderate depression and insomnia vs. mild depression and hypersomnia) with the goal of personalizing interventions.
5. Create visual interfaces to enable user self-experimentation and smarter decision making.

Q&A?

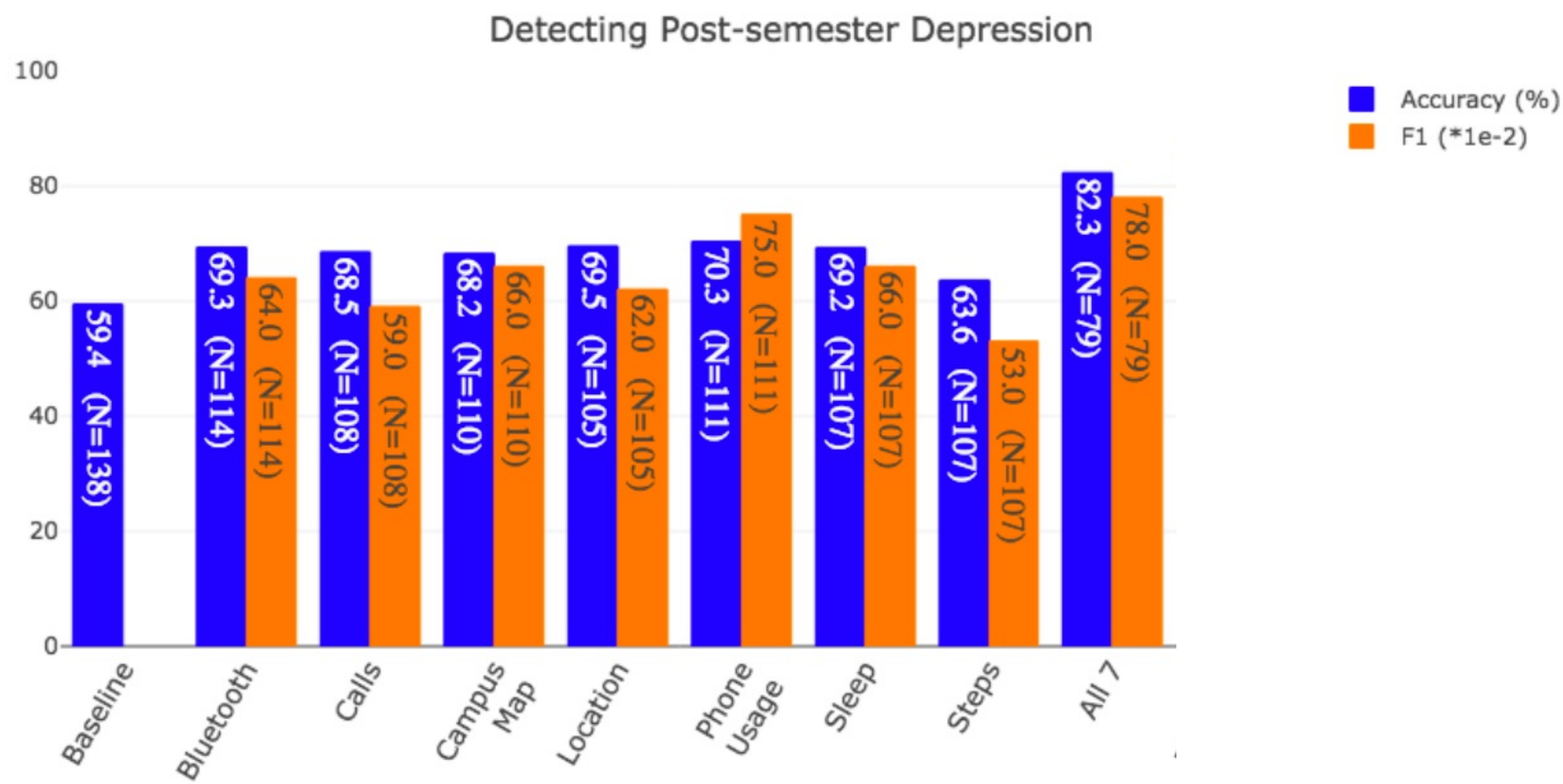
Prerna Chikersal
prerna@cmu.edu

S1: Results – Post-semester Depression

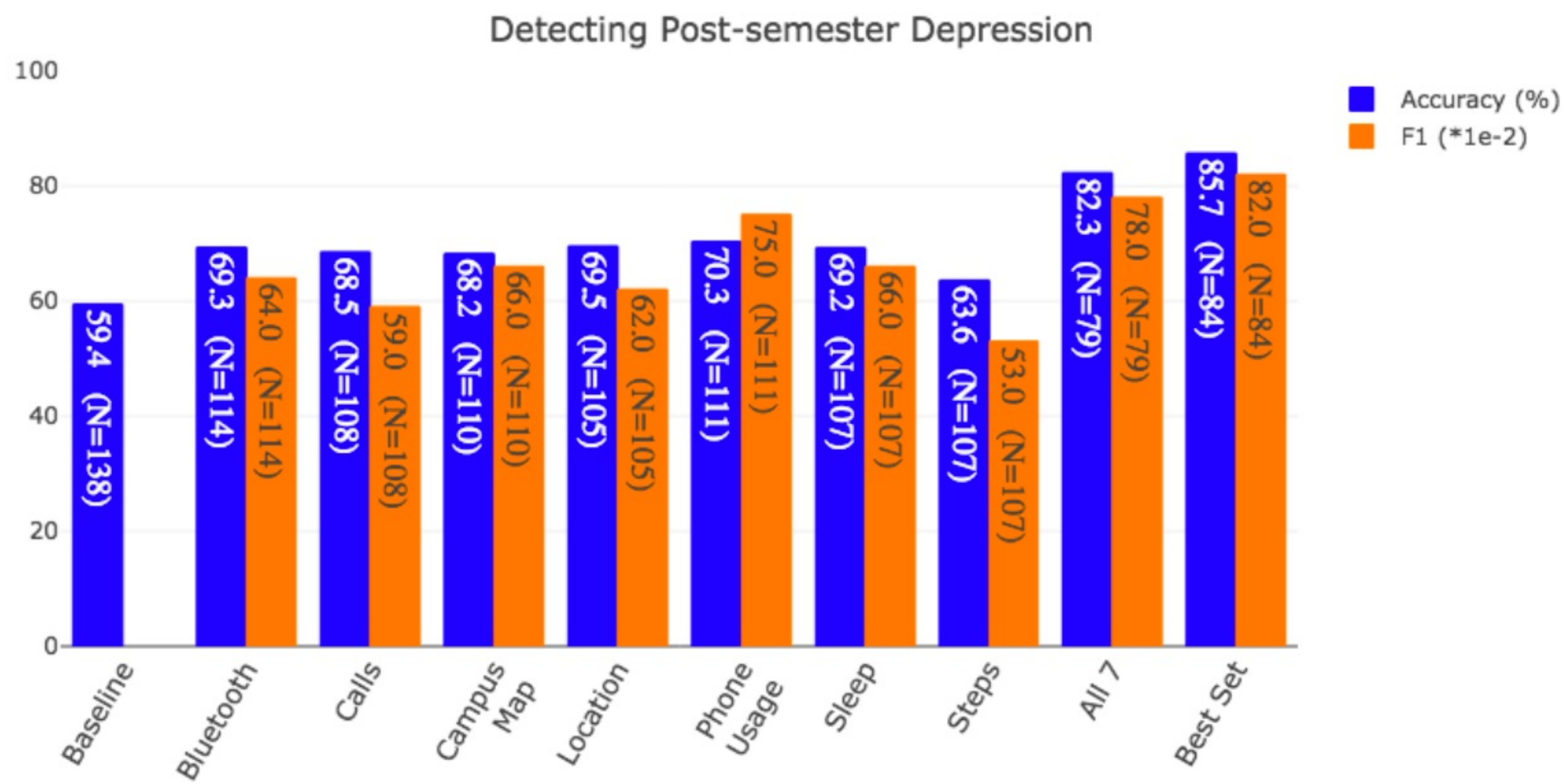
S1: Results – Post-semester Depression Contd.



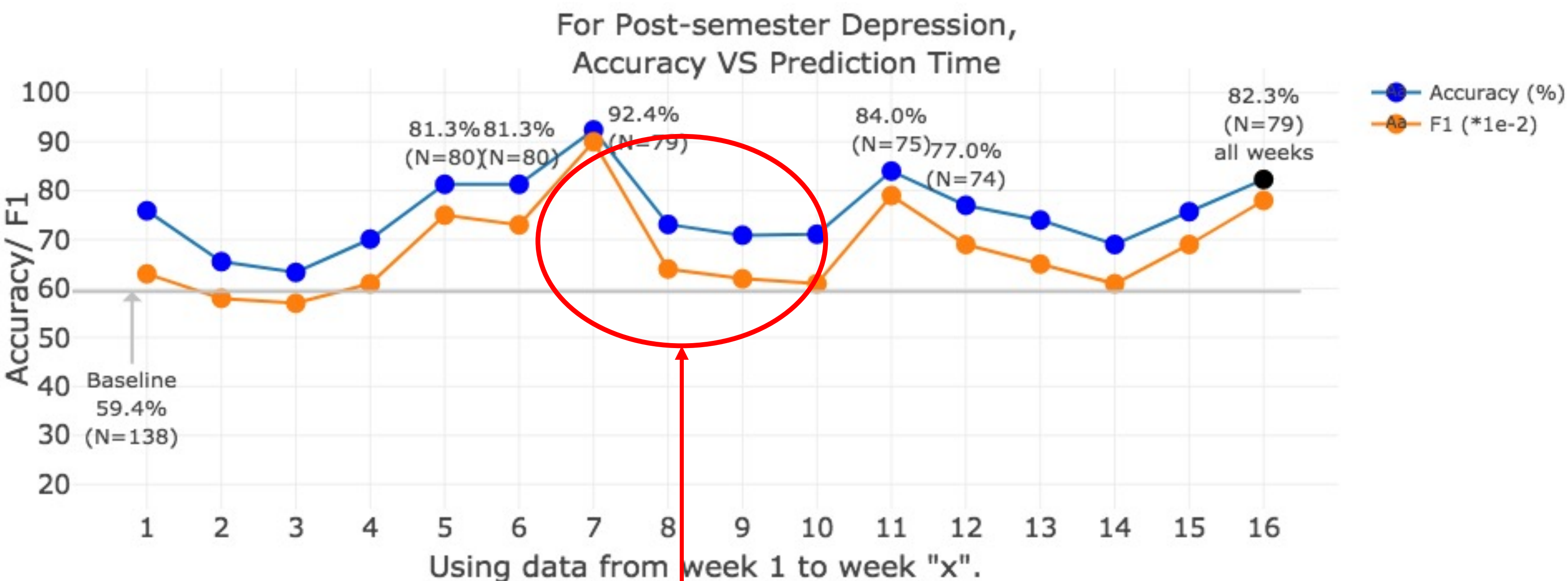
S1: Results – Post-semester Depression Contd.



S1: Results – Post-semester Depression Contd.

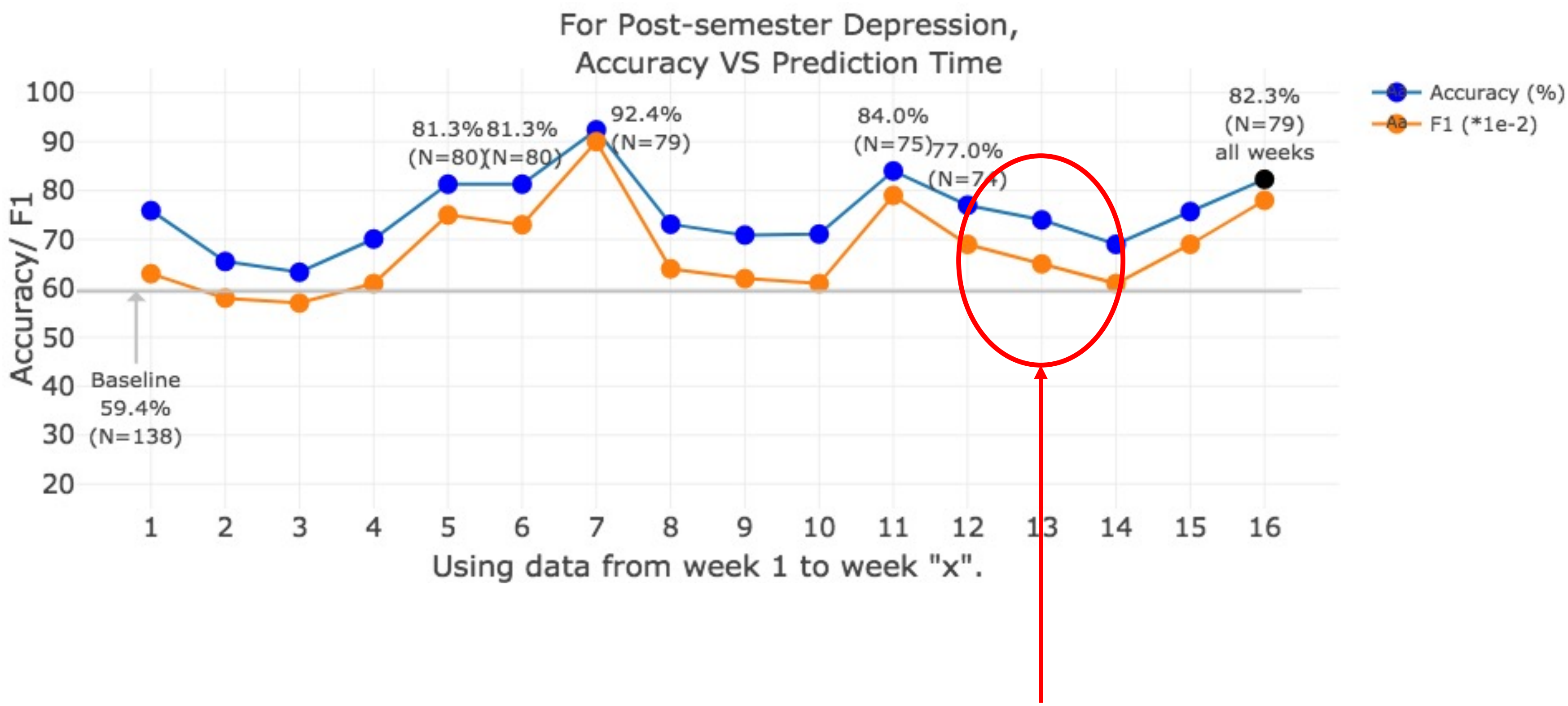


S2: Results



- Drop in accuracy during the spring break and midterms.

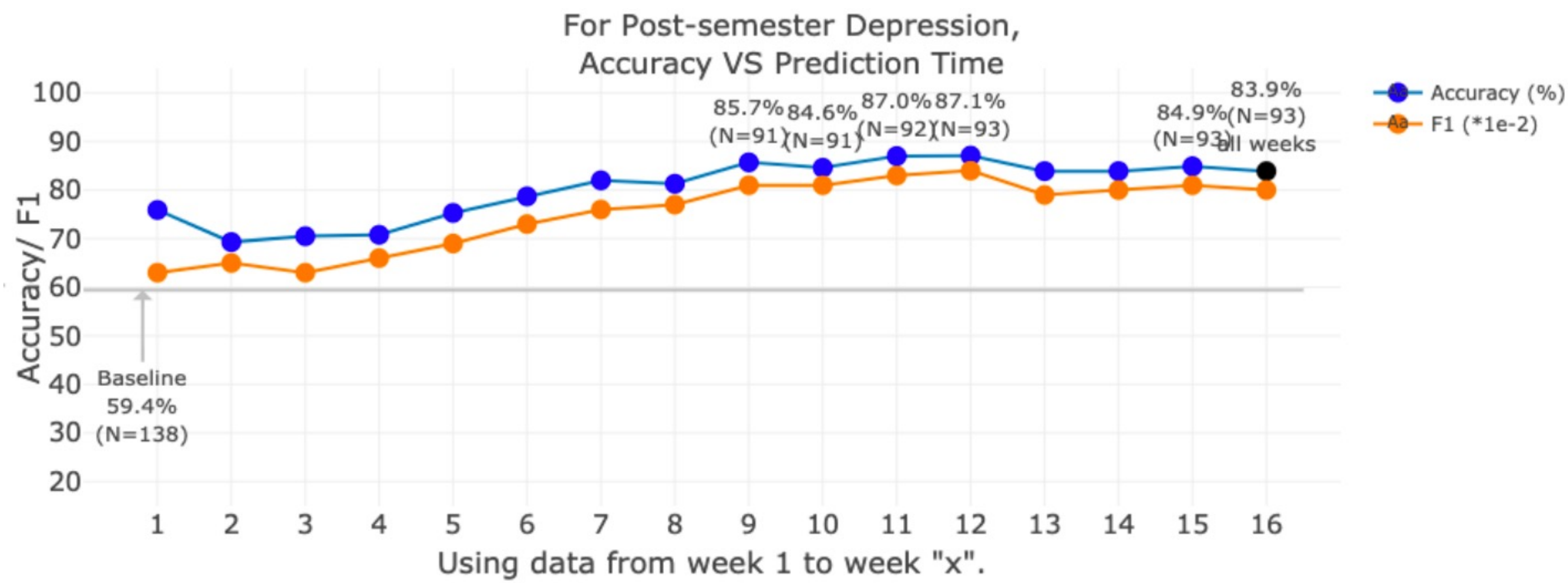
S2: Results



- Hard to say what's happening here without more background information → majority voting

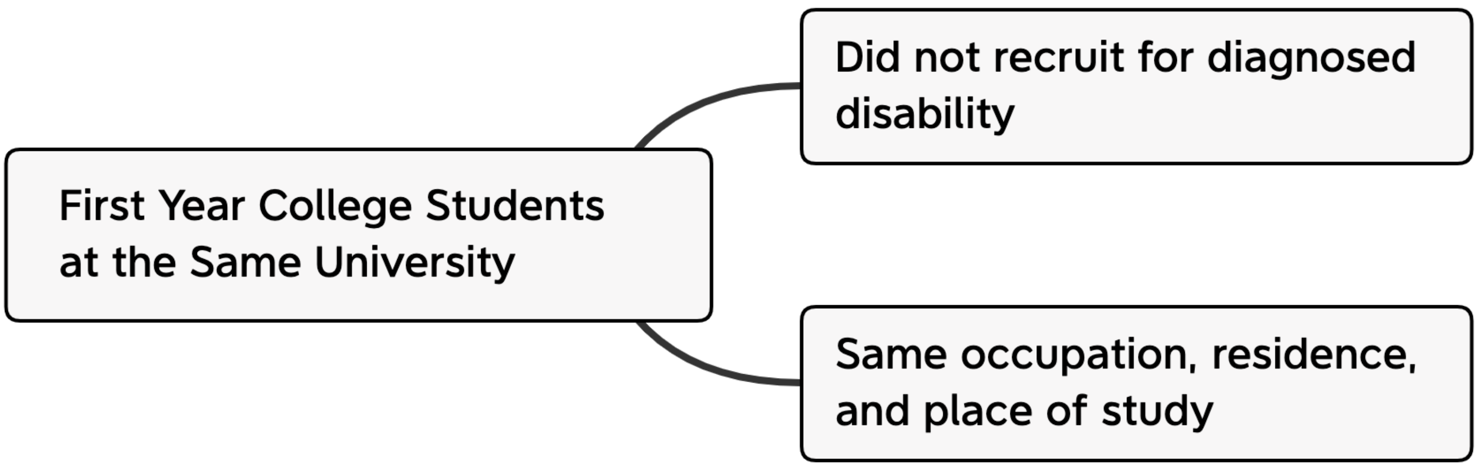
S2: Results

- After majority class voting:



S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period

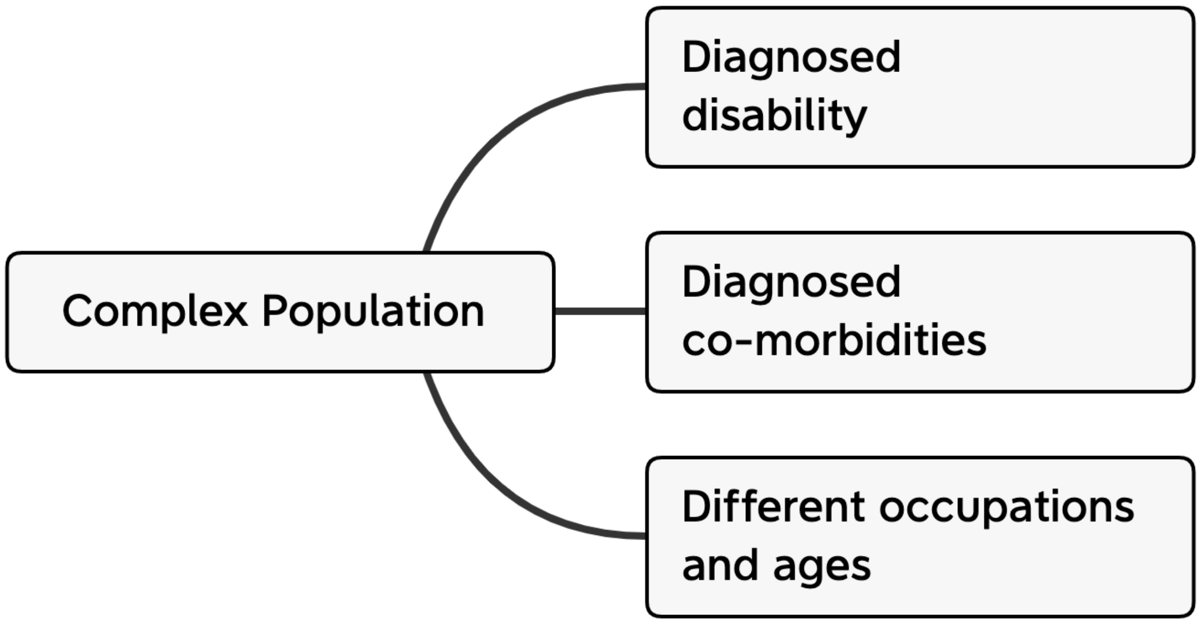
- Study 1 and 2:
  - Participants are likely to have similar behaviors.
  - → it makes sense for our population model (NRLR) to work.



- Q) Would NRLR generalize to a more complex population?

S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period

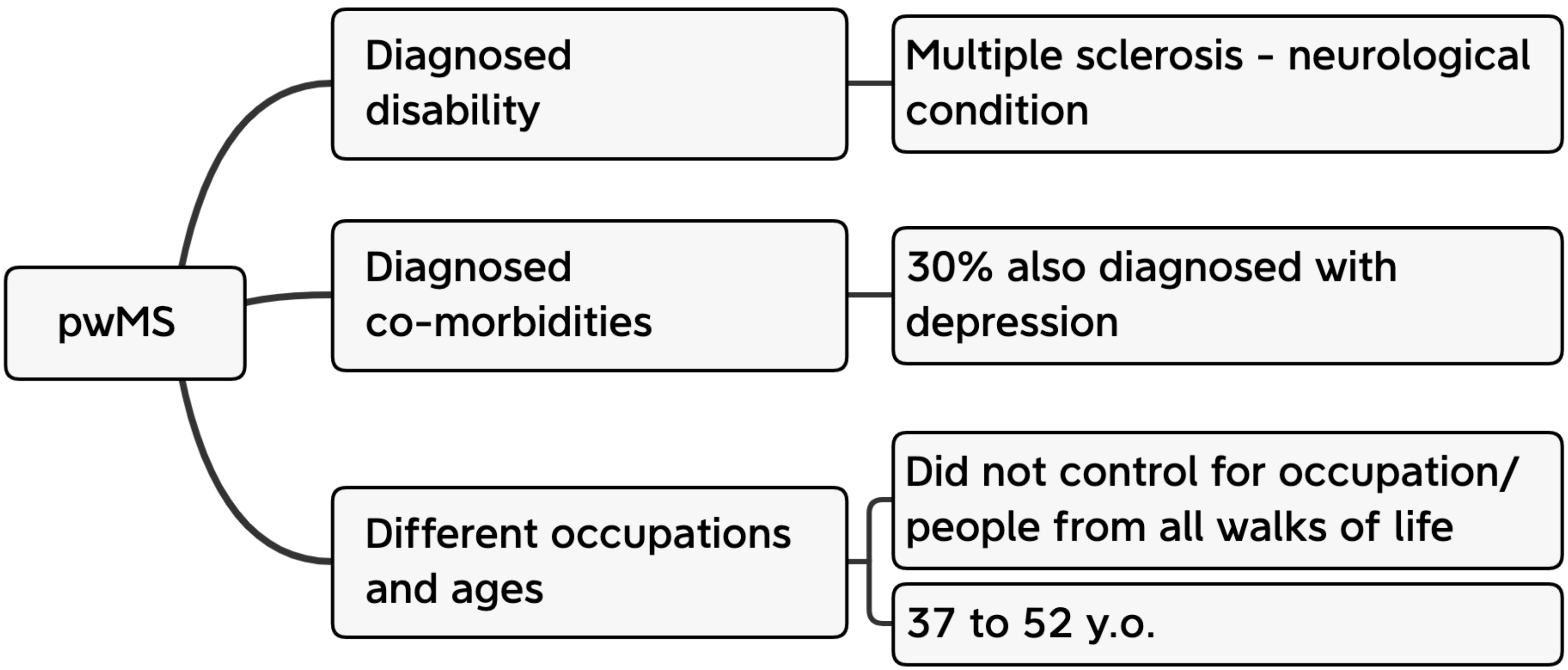
- What factors would make a population "complex"?



- Would NRLR generalize to such a population?
  - One such population is patients with Multiple Sclerosis.

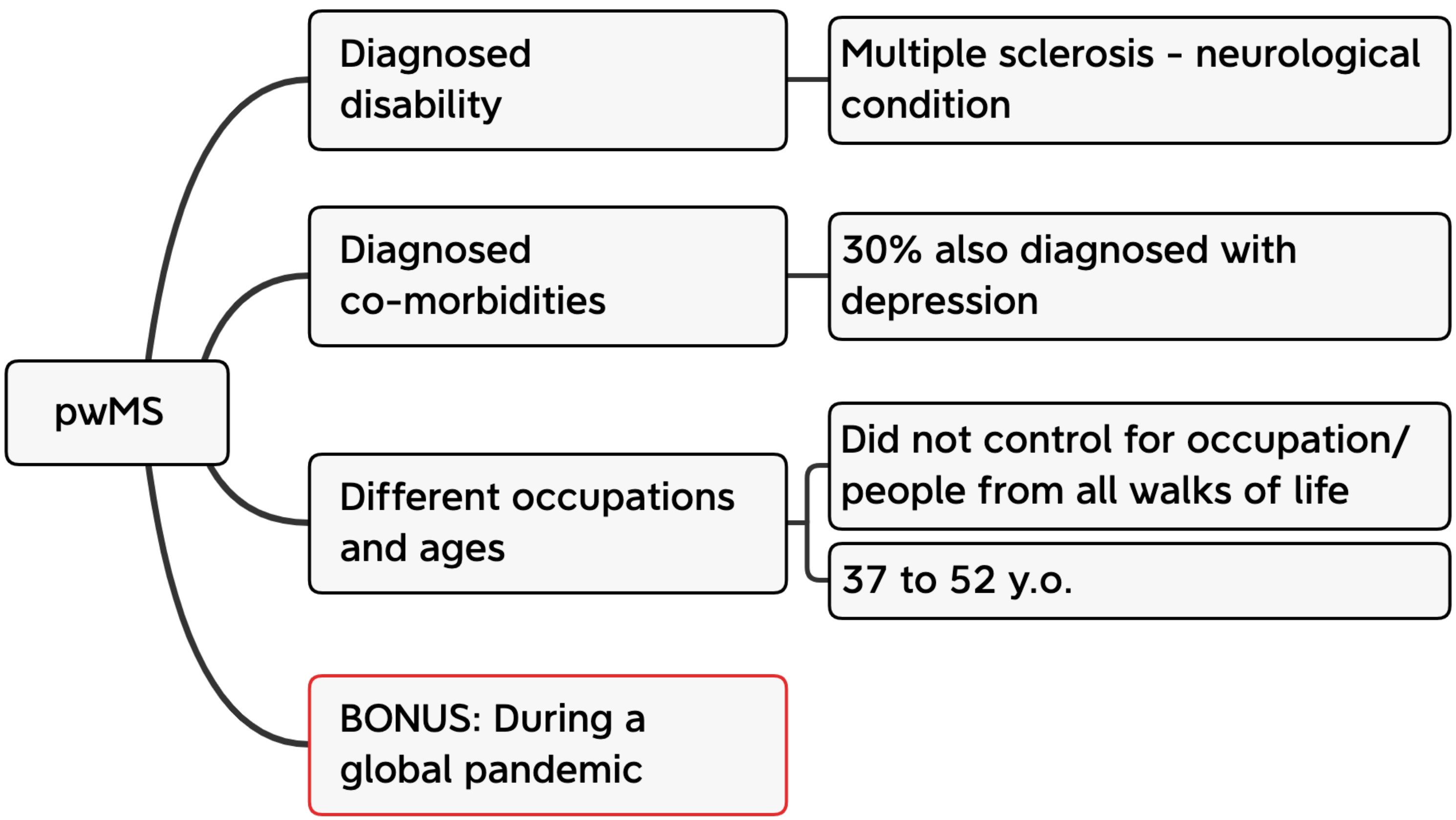
S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period

- Patients with Multiple Sclerosis (pwMS)



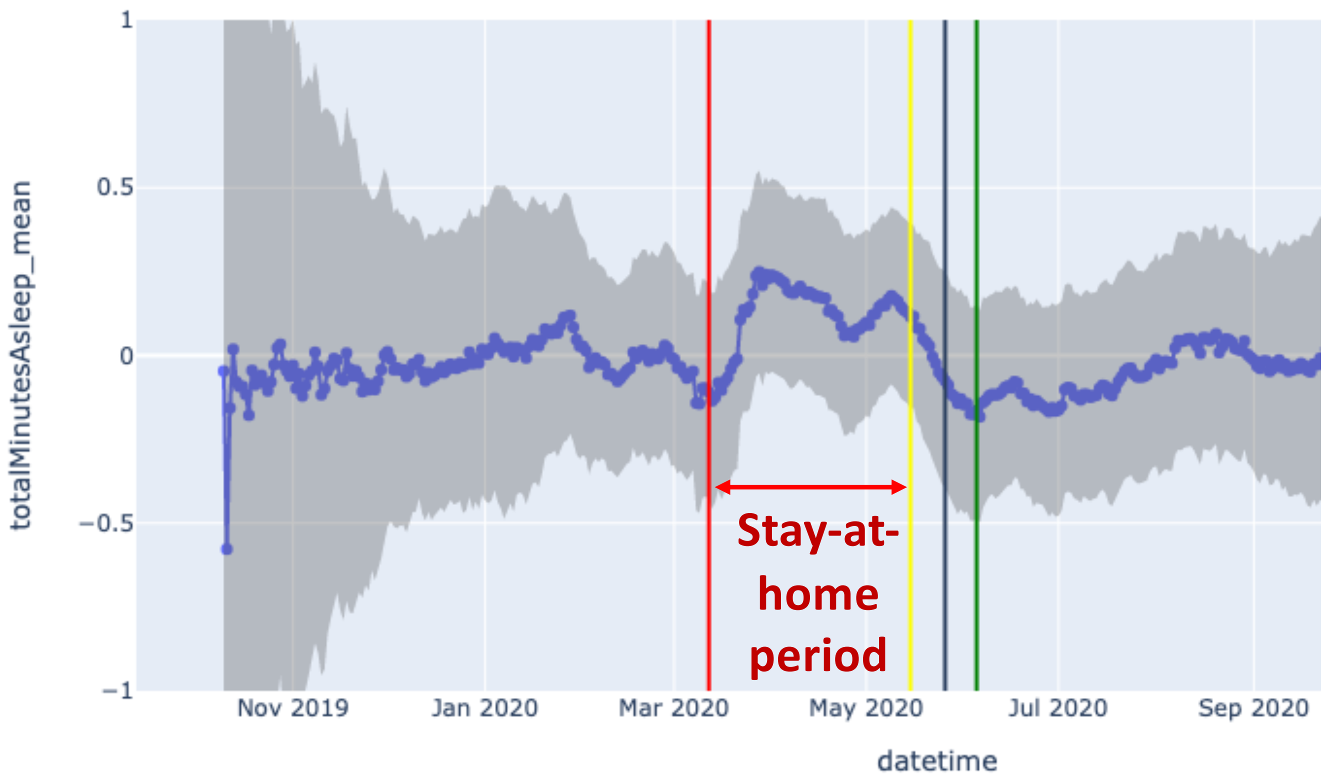
S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period

- Patients with Multiple Sclerosis (pwMS)



S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period

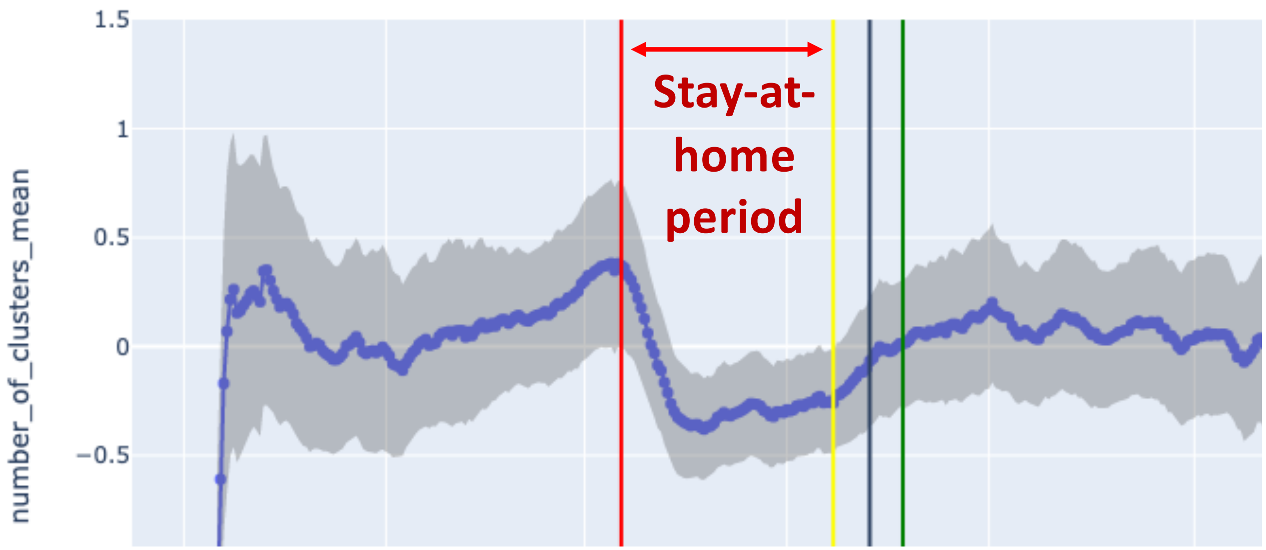
- We found that the stay-at-home period imposed during the pandemic had a major effect on people's behaviors.



Time spent asleep per day (normalized per person)

S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period

- We found that the stay-at-home period imposed during the pandemic had a major effect on people's behaviors.



Number of significant locations (normalized per person)

What does this mean for Multimodal Behavioral Sensing research?

S3: Methodology – Feature Extraction

- Collected data from 6 sensors.
- For every sensor, extracted features from 15 time slices from the pre-covid-19 and stay-at-home periods.

Final Feature Matrix = Stay-at-home Feature Matrix − Pre-Covid-19 Feature Matrix

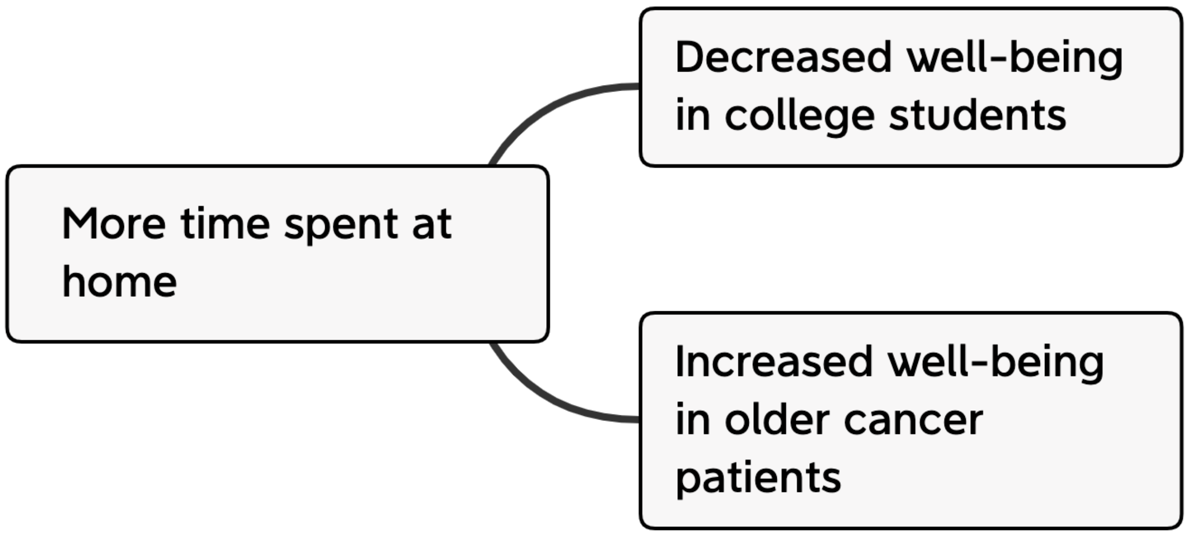
- E.g., for person A:
  - Average Steps Pre-Covid-19 = 7000
  - Average Steps during the Stay-at-Home Period = 4000
  - Final Feature = -3000

Thesis Outline

- Introduction
- The Curse of Dimensionality Challenge
- S1: Detecting Depression and Loneliness in College Students
- S2: Forecasting End of Semester Depression in College Students
- S3: Predicting the Mental Health of People with Multiple Sclerosis (MS) during the Covid-19 Stay-at-Home Period
- S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention
- S5: Predicting Periodically Assessed MS Outcomes Using Passively Sensed Behaviors and Ecological Momentary Assessments
- Thesis Contributions and Future Work

S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention

- **RECAP:** The relationship between behaviors and outcome may be dependent on patient context and characteristics.



- We have not yet addressed the curse of dimensionality in the diversity in patient context and characteristics.

S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention

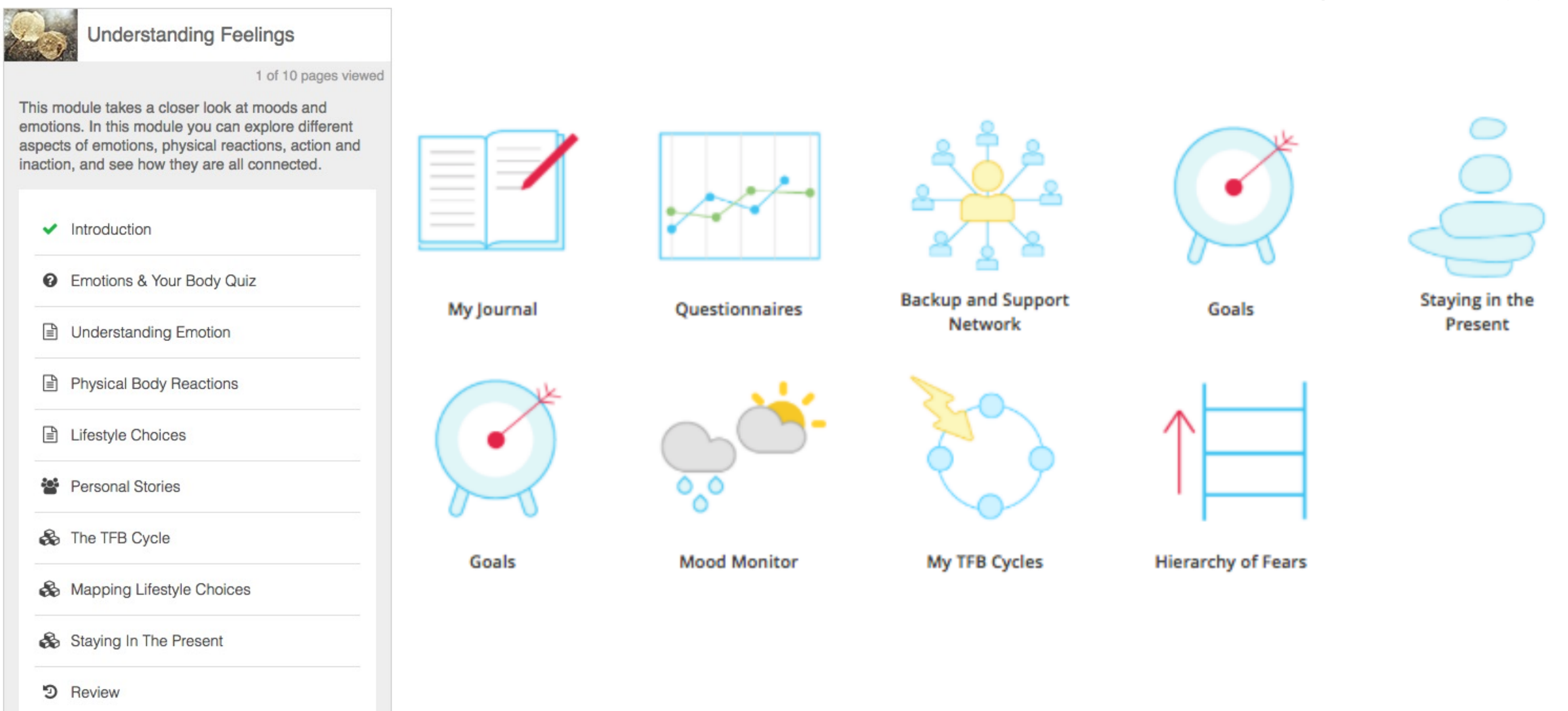
- Further, interventions are an important part of this kind of work.
- Hence, I’d like to switch gears and talk about interventions

- Analyzing behaviors in the context in which they occur, can lead to more interpretable insights,
  - Interventions require more trust and transparency.
  - So, interpretability is even more important for interventions.

- **S4’s GOAL:** Analyze the users’ interaction with a MH intervention app and the human supporters on the app to understand how supporter behaviors correlate with patient outcomes for patients in different contexts or situations.

S4: Background

- Patient is engaged with an online mental health intervention based on Cognitive Behavioral Therapy (CBT).
- Patient has access to course content and tools through the app.



S4: Background Contd.

- A human supporter reviews the patient’s clinical scores, progress on course material and tools usage, and sends a personalized message offering feedback, **each week for 6-8 weeks**.
- Supporters use their experience and discretion to employ a wide variety of strategies in their messages.

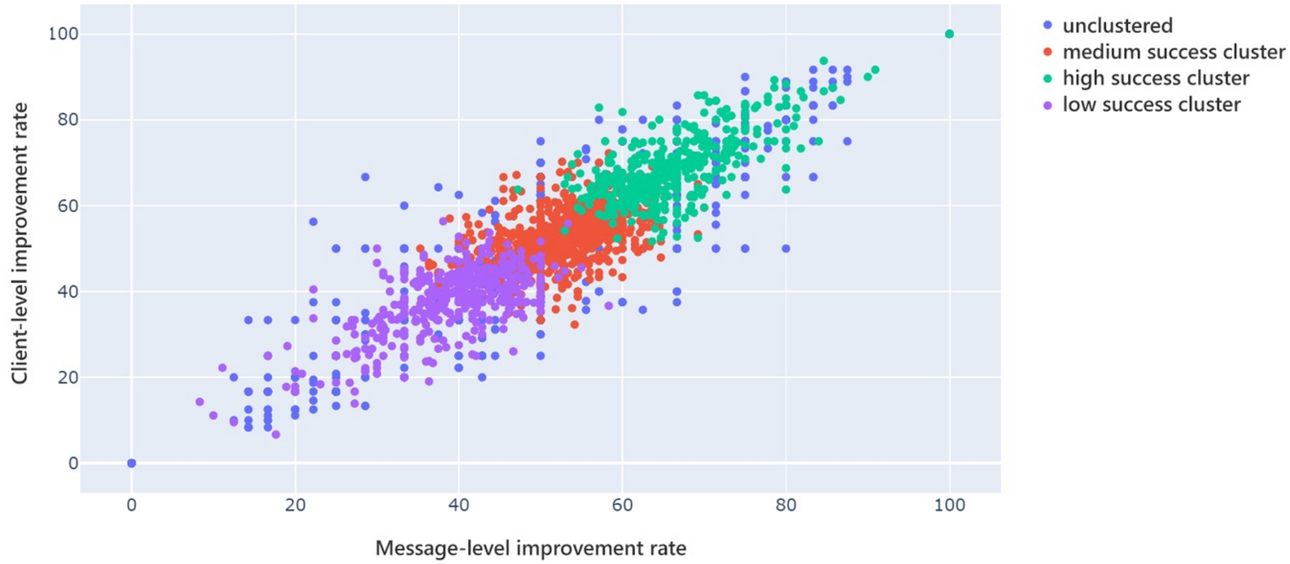
~3,500 supporters → ~200,000 messages → ~50,000 patients

S4: Background

- Patients also fill out **weekly surveys** that measure symptoms of depression and anxiety.
- Goal of the app-based intervention is to improve **both depression and anxiety**, which are often co-morbid.
  - Deriving insights for both outcomes separately would make it harder to identify best support strategies
  - Hence, **need to combine** multiple measures of **depression and anxiety** into ONE target outcome.

S4: Method – Combining Co-morbid Outcomes

- For each supporter, we compute 8 measures based on their patients' depression and anxiety scores → cluster



- **Combined outcome: Success of the supporter**
- High success clusters → "more successful supporters"
- Low success cluster → "less successful supporters"

S4: Method – Combining Co-morbid Outcomes

How does this address the curse of dimensionality?

The Curse of Dimensionality Challenge

- Feature Space
- Multi-morbidities
- Patient Diversity

Proposed a novel feature selection method called Nested Randomized Logistic Regression (NRLR) that employs feature space decomposition based on temporal slices

Proposed a ML pipeline that combines NRLR and late sensor fusion (NRLR Pipeline)

Showed that the NRLR Pipeline can generalize to two co-morbid outcomes - depression and anxiety

S1: Detecting Depression and Loneliness In College Students

Showed that the NRLR Pipeline can generalize to feature spaces of different sizes

S2: Forecasting End of Semester Depression In College Students

Showed that the NRLR Pipeline can generalize to more challenging populations including people with different occupations diagnosed with a complex neurological condition and multiple co-morbidities

Showed that the NRLR Pipeline can generalize to 4 health outcomes that are frequently co-morbid in patients with MS

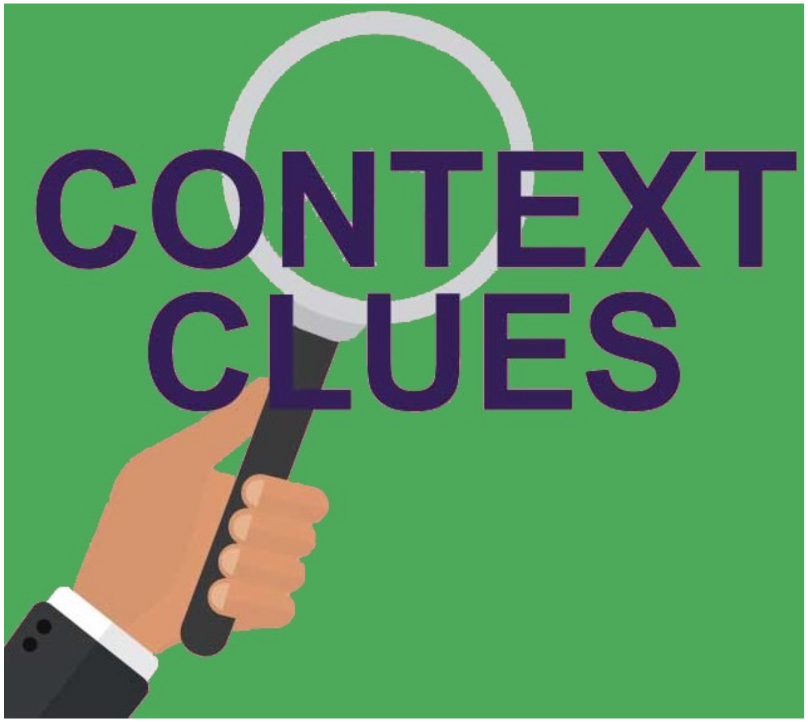
S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period

Defined a "combined" target outcome based on both depression and anxiety, in order to identify best intervention strategies that improve both depression and anxiety

S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention

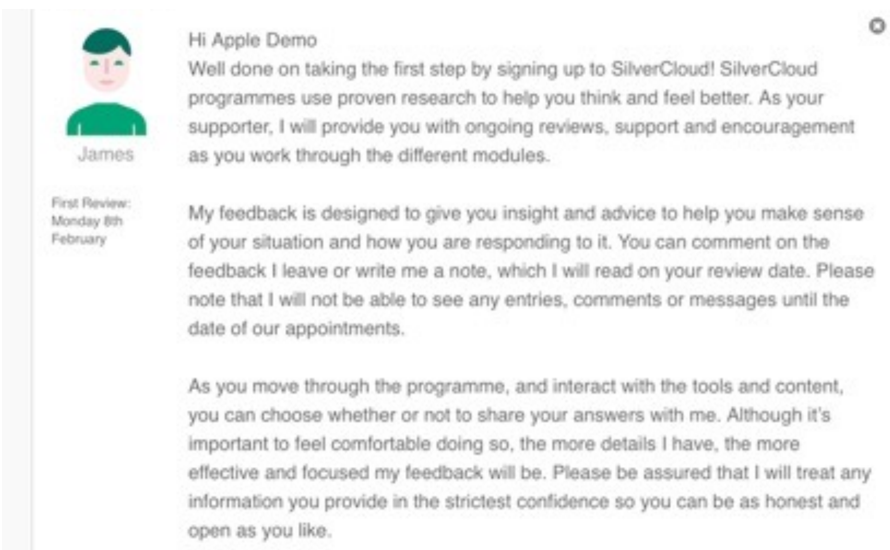
S4: Method – Feature Extraction

- Patient <Context> Variables:
  - 5 Variables:
    - ContentViews,
    - Shared,
    - MessageNumber,
    - CurrentDepression, and
    - CurrentAnxiety



S4: Method – Feature Extraction Contd.

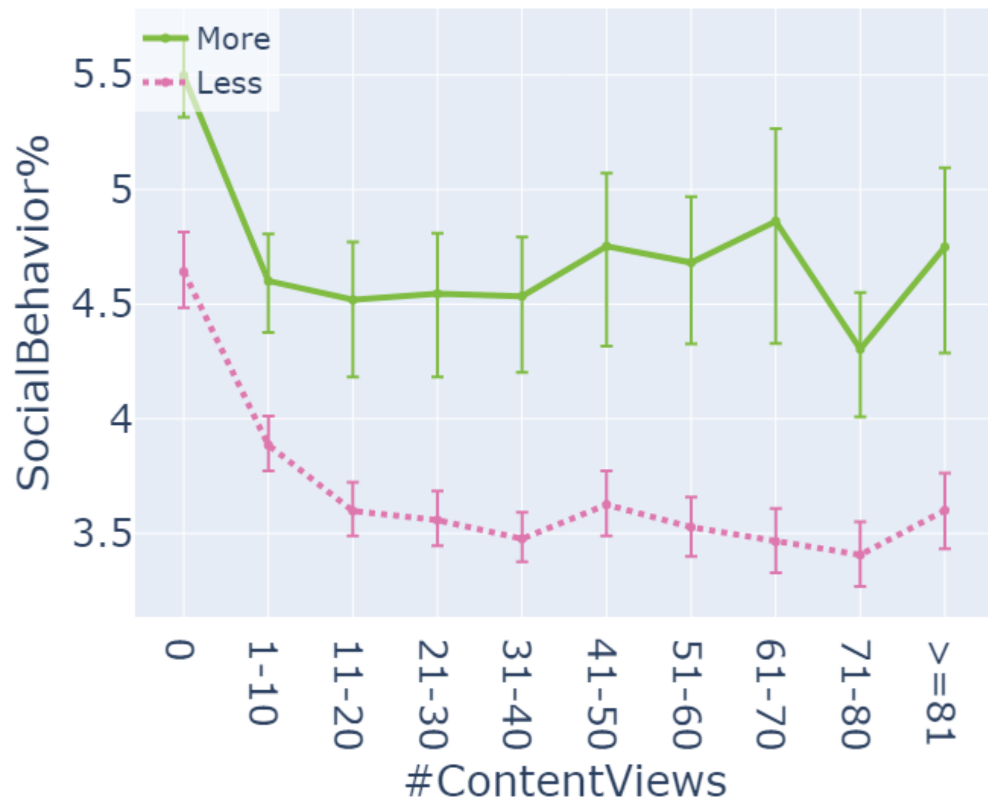
- Support <Strategy> Variables:
  - 23 variables calculated using validated lexicons and NLP techniques.



E.g., positive or negative sentiment, message length, types of words used

S4: Results – Successful Support Strategies Contd.

- More successful messages had:
  - Used more words associated with social behavior (E.g. help, call, discuss, and share.)



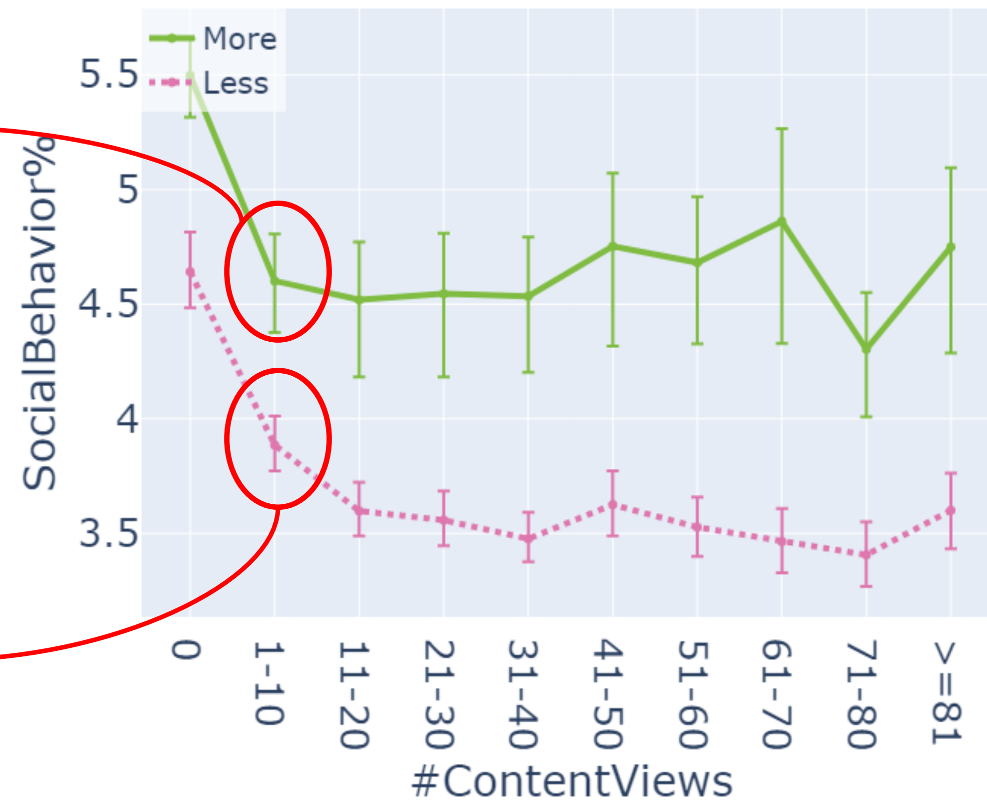
S4: Results – Successful Support Strategies Contd.

- Q) What strategies are best **independent** of the patient's context?
- Only ONE context variable and ONE strategy at a time

Messages from MORE successful supporters and clients with 1-10 ContentViews

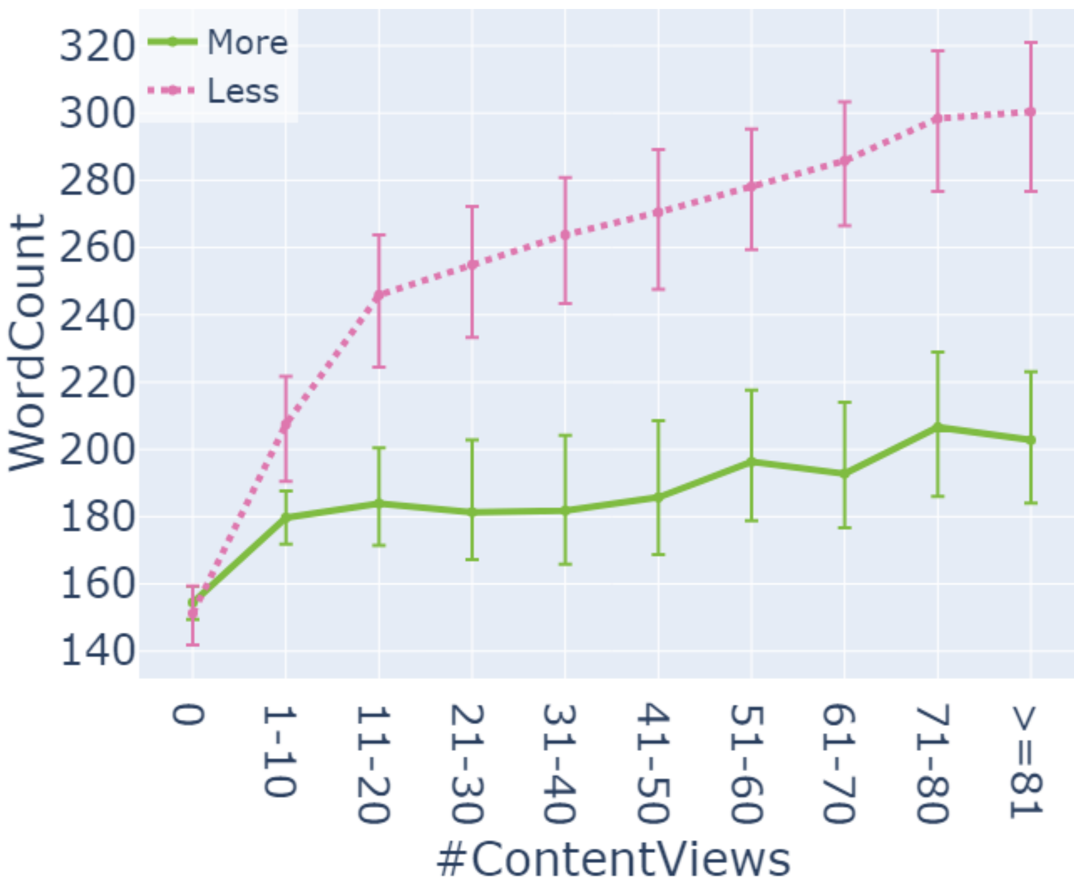
Hierarchical Bootstrapping to compare means (samples not independent)

Messages from LESS successful supporters and clients with 1-10 ContentViews



S4: Results – Successful Support Strategies Contd.

- More successful messages were shorter!



S4: Methods – Context-Specific Support Strategies

- So far, we've found strategies that work well independent of context or across all contexts → "general" strategies
- Q) Can we find strategies that work better in specific narrow contexts? → "specific" strategies
  - Do general strategies "flip" in specific narrow contexts?
  - Has interesting implications for personalization!
- So far, we've only considered ONE context and ONE strategy variable at a time.
  - For this, we will consider multiple context variables *i.e.*, multidimensional client contexts

S4: Results – Context-Specific Support Strategies

S4: Results – Context-Specific Support Strategies Contd.

S4: Results – Context-Specific Support Strategies Contd.

- For less engaged clients, writing longer, more positive and more supportive messages is linked with greater outcomes.
- More engaged clients appear to benefit more from messages with less negative words, less abstraction, and more references to social behaviors.

S4: Addressing the Curse of Dimensionality

How does this address the curse of dimensionality?

The Curse of Dimensionality Challenge

- Feature Space
- Multi-morbidities
- Patient Diversity

Proposed a novel feature selection method called Nested Randomized Logistic Regression (NRLR) that employs feature space decomposition based on temporal slices

Proposed a ML pipeline that combines NRLR and late sensor fusion (NRLR Pipeline)

Showed that the NRLR Pipeline can generalize to two co-morbid outcomes - depression and anxiety

S1: Detecting Depression and Loneliness In College Students

Showed that the NRLR Pipeline can generalize to feature spaces of different sizes

S2: Forecasting End of Semester Depression In College Students

Showed that the NRLR Pipeline can generalize to more challenging populations including people with different occupations diagnosed with a complex neurological condition and multiple co-morbidities

Showed that the NRLR Pipeline can generalize to 4 health outcomes that are frequently co-morbid in patients with MS

S3: Predicting the Mental Health of People with Multiple Sclerosis during the COVID-19 Stay-at-Home Period

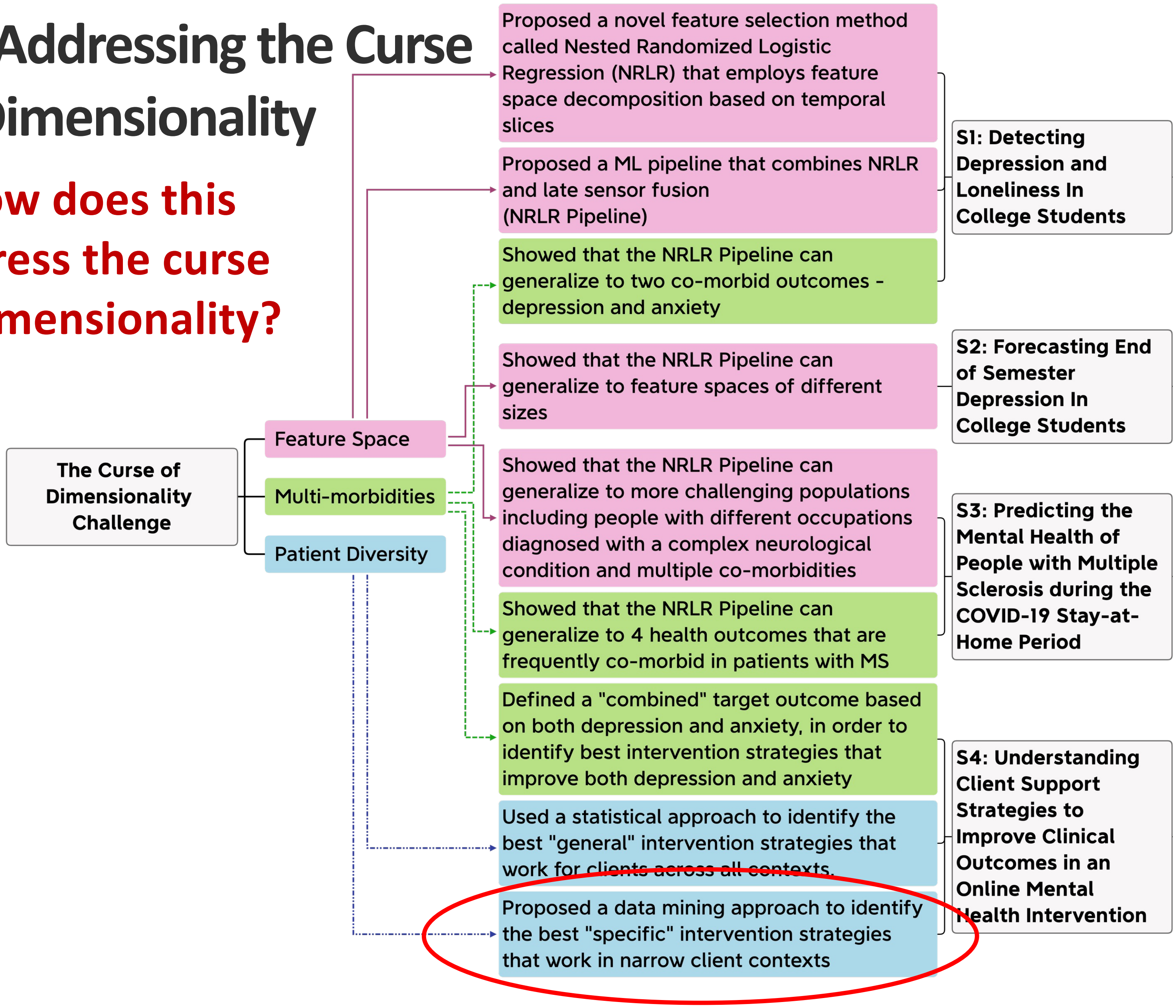
Defined a "combined" target outcome based on both depression and anxiety, in order to identify best intervention strategies that improve both depression and anxiety

Used a statistical approach to identify the best "general" intervention strategies that work for clients across all contexts.

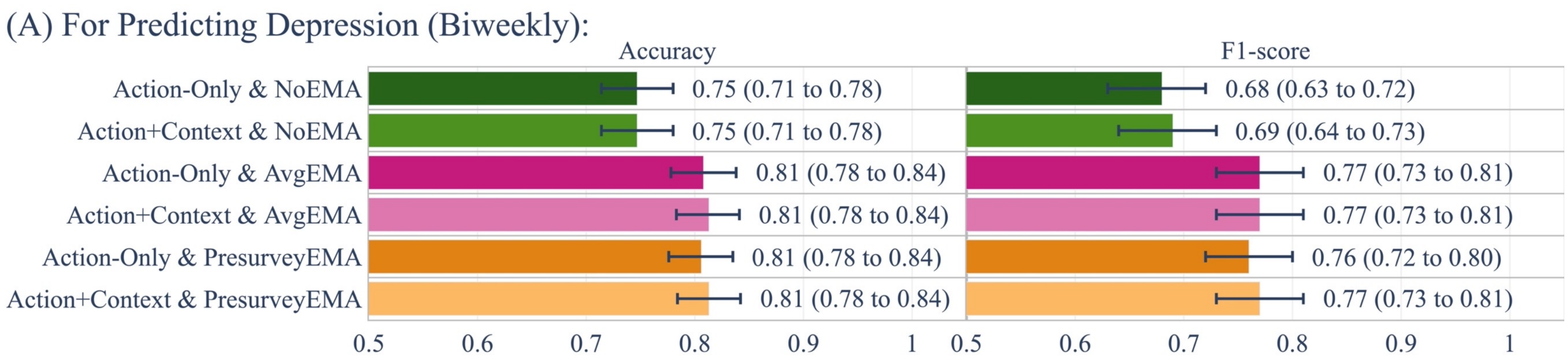
S4: Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention

S4: Addressing the Curse of Dimensionality

How does this address the curse of dimensionality?



S5: Results – Biweekly Depression

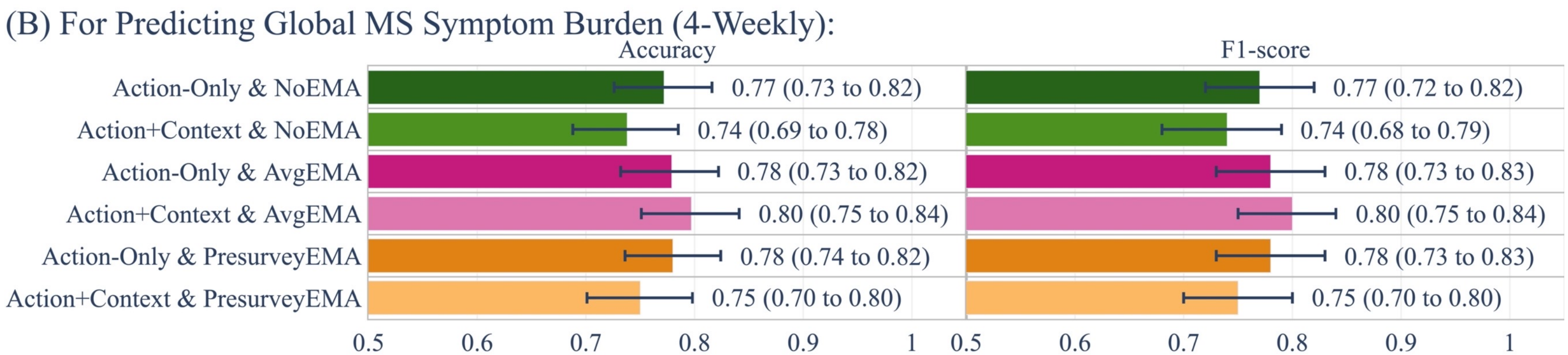


Baseline: 59.5% (majority class – no depression)

Statistically Best Model: Action-Only & PresurveyEMA

- Best performance while requiring the least amount of EMA.
- Accuracy: 80.6% - a 35.5% improvement over baseline.
- F1: 0.76.
- Combination: heart rate, steps, and pre-survey EMA

S5: Results – 4-Weekly Global MS Symptom Burden

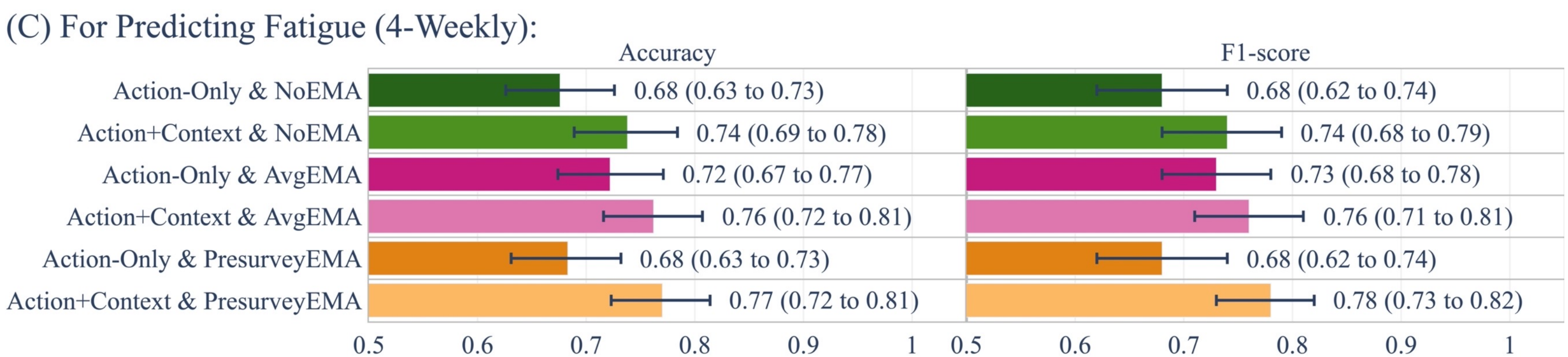


Baseline: 51.1% (majority class – high burden)

Statistically Best Model: Action-Only & NoEMA

- Best performance while requiring no EMA
- Accuracy: 77.3% - a 51.3% improvement over baseline.
- F1: 0.77.
- Combination: heart rate, location, sleep, and steps.

S5: Results – 4-Weekly Fatigue

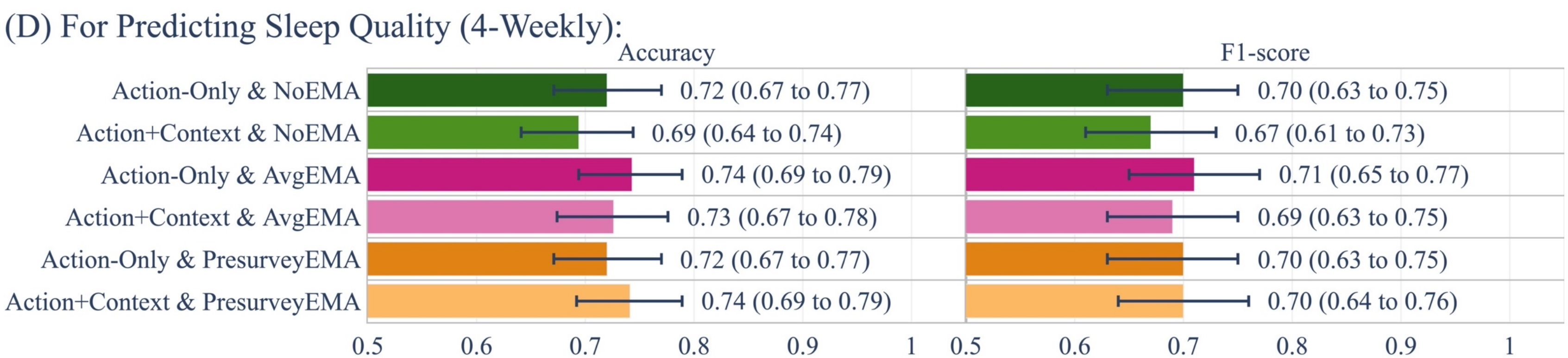


Baseline: 50.9% (majority class – severe fatigue)

Statistically Best Model: Action+Context & NoEMA

- Best performance while requiring no EMA
- Accuracy: 73.8% - a 45% improvement over baseline.
- F1: 0.74.
- Combination: heart rate, screen, and steps.

S5: Results – 4-Weekly Sleep Quality



Baseline: 56.2% (majority class – better sleep quality)

Statistically Best Model: Action-Only & NoEMA

- Best performance while requiring no EMA
- Accuracy: 72.0% - a 28.1% improvement over baseline.
- F1: 0.70.
- Combination: heart rate, location, sleep, and steps.

Key Takeaways

- Multimodal behavioral sensing for precision mental health care is a wicked problem due to the curse of dimensionality w.r.t. the feature space, the existence of co-morbidities, and the diversity in patient characteristics.
    - Collect potentially confounding health measures and patient characteristics.
    - Consider strategies like decomposing the feature space or computationally combining outcomes to mitigate curse of dimensionality.
- Features from multiple time slices, followed by feature space decomposition is effective, though not always.

Key Takeaways

- MH outcomes are complex and frequently comorbid. Demonstrate two ways to deal with comorbid:
  - Combining them
  - Predicting them separately → can also assess generalizability
- Using recent behaviors to predict outcomes isn't sufficient.
  - Contextual or past behaviors can improve model performance by capturing patient diversity more holistically.