A Machine Learning Analysis of Risk Factors for Somatic Symptoms in the PHQ-15 Among Young Adults

Miao Yu & Jessica Tanchone

What is somatic symptoms?

- **Somatic symptoms = physical sensations or complaints that are not primarily caused by a medical condition**
- Often linked to stress, identity, and social factors
- Understudied in non-clinical young adult samples
- Early detection can guide prevention & care
- Prevalence in general population

Background

10% Adults experiencing persistent symptoms

20%-45% primary care visits

69% Depressed Patient

Literature Review

Headache

- Stress *(Duan et al., 2023)*
- mood symptom *(Holroyd et al., 2000)*

Shortness of breath / chest pain (stronger physical associations)

- identifiable medical conditions *(Zachariah et al., 2017;* Gulati et al., 2021; Riley et al., 2022*)*

Fatigue *(Creed et al., 2022)*

- Stress
- chronic illness
- Anxiety
- Neuroticism

Sleep problems *(Meredith et al., 2020)*

- chronic health conditions
- increased by financial strain and reduced by physical activity

What is PHQ - 15 questionnaires?

The most commonly used screening instrument to detect somatization symptoms in the general population.

Summed scores across item responses were used for analysis with **higher scores** indicating more **severe** somatic symptoms

Limitation: PHQ-15 assuming they all reflect the same underlying cause, yet individual symptoms differ widely in their origins

The Patient Health Questionnaire-15

During the past four weeks, how much have you been bothered by the following symptoms?

Symptom	*Not at all*	*A little*	*A lot*
Back pain	0	1	2
Chest pain	0	1	2
Constipation, loose bowels, or diarrhea	0	1	2
Dizziness	0	1	2
Fainting	0	1	2
Feeling tired or having low energy	0	1	2
Feeling your heart pound or race	0	1	2
Headaches	0	1	2
Menstrual cramps or other problems with your periods (women only)	0	1	2
Nausea, gas, or indigestion	0	1	2
Pain in your arms, legs, or joints	0	1	2
Pain or problems during sexual intercourse	0	1	2
Shortness of breath	0	1	2
Stomach pain	0	1	2
Trouble sleeping	0	1	2

Score: _____

Scoring: No somatic symptom disorder (0 to 4), mild (5 to 9), moderate (10 to 14), severe (15 or higher).

OBJECTIVE

Our primary goal is to build and compare **symptom-specific prediction models** and to identify **the most influential predictors** for targeted symptoms.

Data

- Source: EAMMi2 dataset (Grahe et al., 2018)
- 4000+ participants from 30+ universities
- Participants age range from 18 to 29.
- 72.8% of respondents were women (*Nwomen* = 2280; *Nmen* = 771) (*Mage* = 21.10, *SD* = 4.83)

Racial and ethnicity	percent(%)
White/European American	63.5%
Black/African-American	7.6%
Hispanic/Latino/Latina	8.7%
Asian/Pacific Islander	6.5%
Native American (0.4%)	0.4%
“Other” Race (2.2%),	2.2%

Demographic & Attitudinal Predictor

- **Sex** (male, female, other; categorical)
- **Education level**
- **Race/Ethnicity**
- **Household income**
- **School attended**
- **Parental marriage status** (recoded into categorical dummy variables)
- **Siblings** (recoded: -0.5 = no siblings, 0.5 = at least one sibling)
- **Importance of marriage** (single-item rating)

Psychological Predictor

Individual Differences & Personality

- *IDEA-8*: Identity exploration and development
- *NPI-13*: Narcissism
- *Interpersonal Exploitativeness* (3 items)
- *Self-Efficacy / Competence*
- *Mindfulness Scale*

Social & Interpersonal Factor

- *Perceived Social Support* (12 items)
- *Need to Belong* (10 items)
- *Interpersonal Transgressions* (4 items)
- *Social Media Use*: Maintaining / Making connections, Seeking information

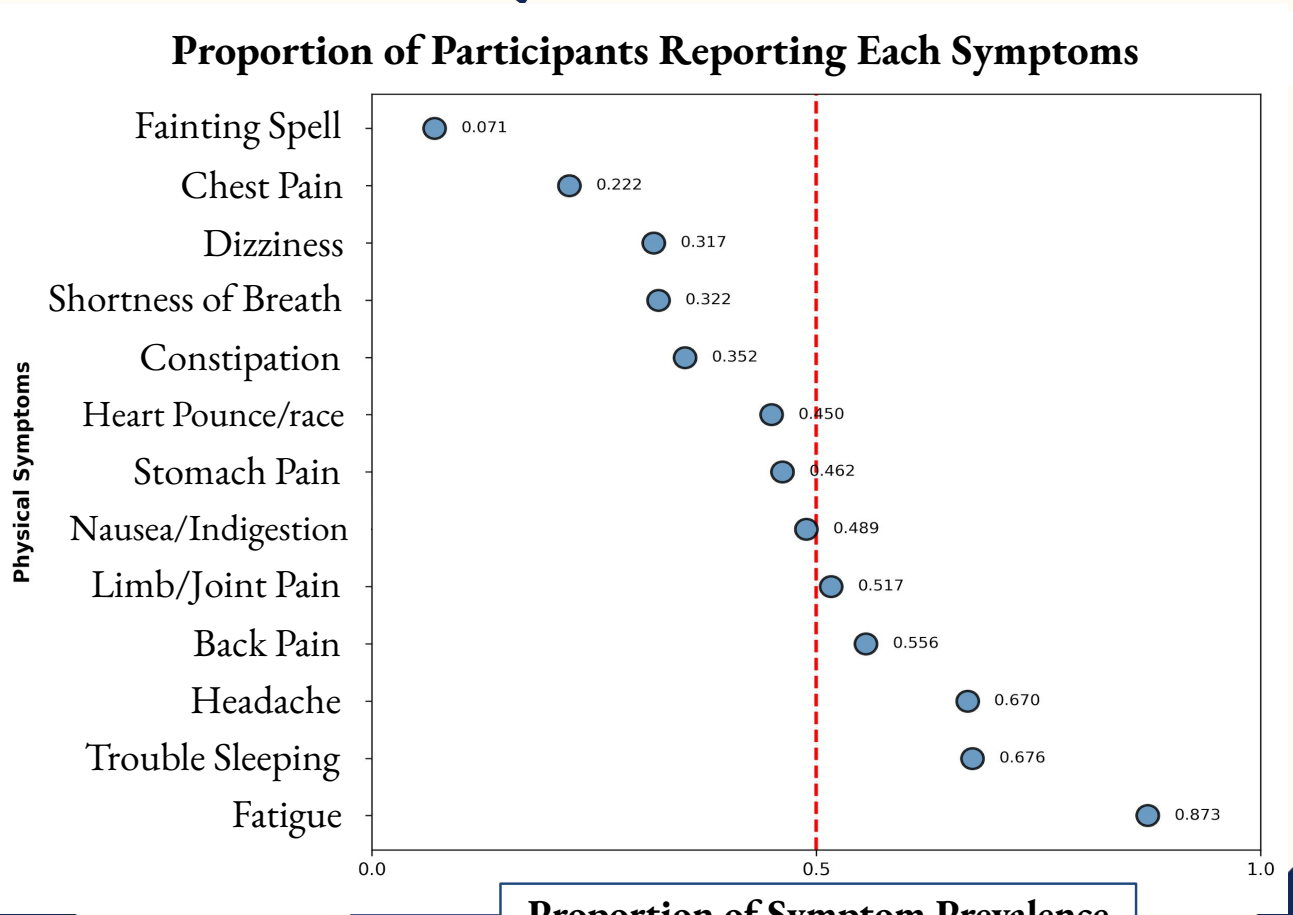
Stress & Well-Being

- *Perceived Stress Scale*
- *Subjective Well-Being* (6 items)
- *Disability Identity and Status* (22 items combined from Q10–Q14)
- *Belief in the "American Dream"* (2 items)

Markers of Adulthood

- *Achievement* subscale (20 items)
- Importance subscale (20 item)

PHQ-15 outcomes

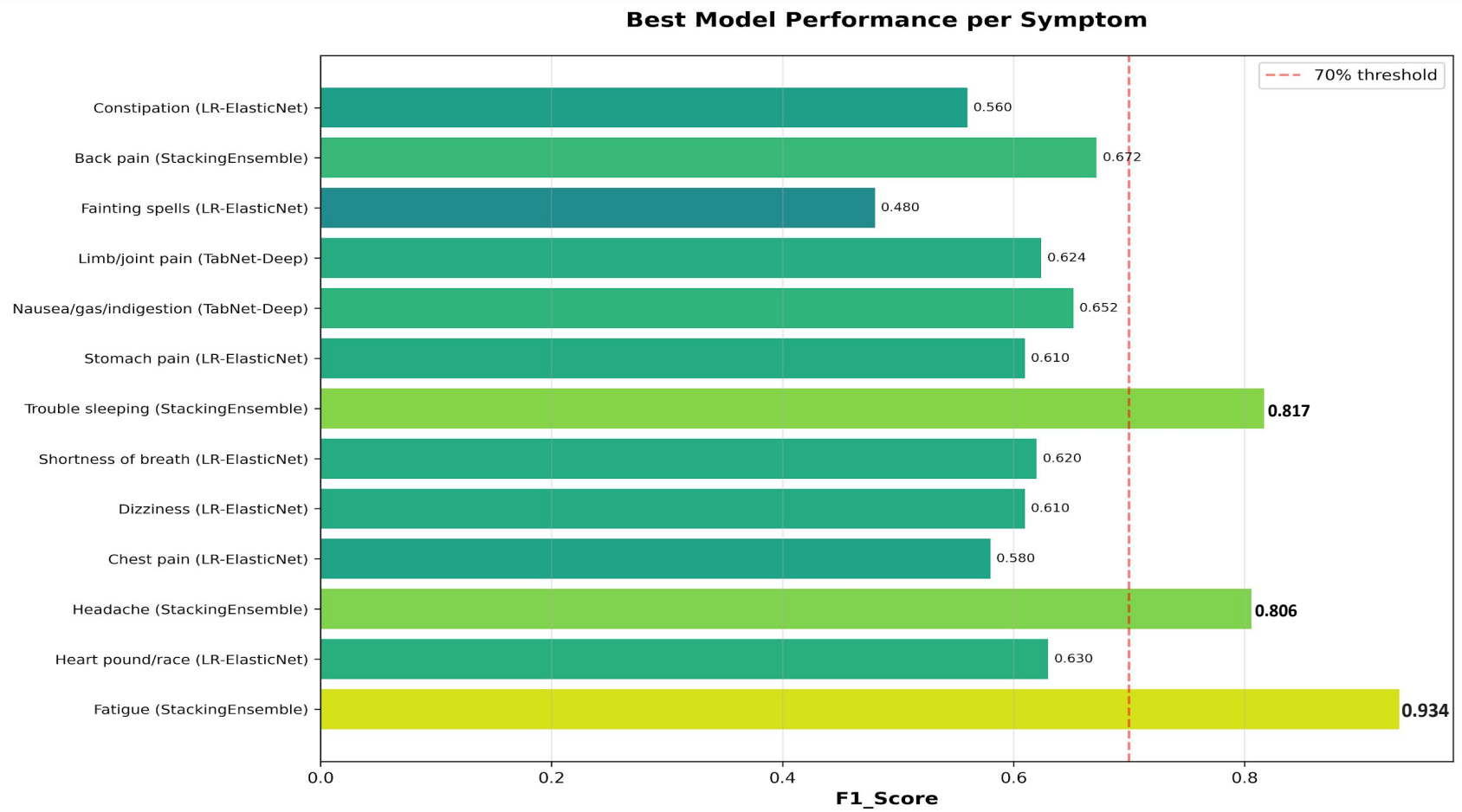


Models & Evaluation Metrics

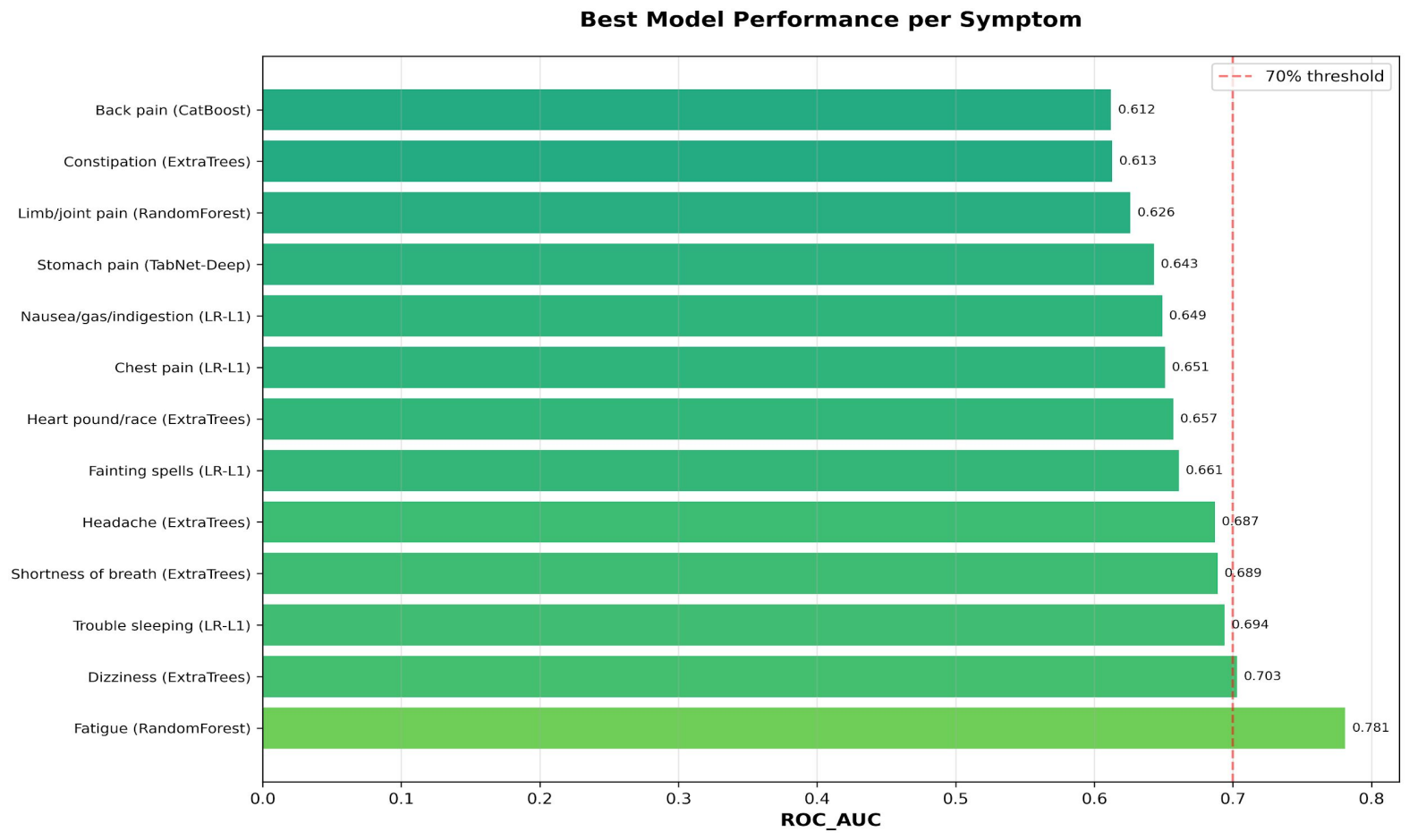
- 3x repeated 10 fold for best hyperparameter
- Logistic regression (GLM net)
- K-nearest neighbors
- Tree-based models:
  - XGboost
  - Catboost
  - Light Gradient Boosting
  - Random forest
  - Extratree
- Voting Ensemble (combines predictions from multiple algorithms)
- Neural network (Tabnet-deep)

- Due to Class Imbalanced:
  - Balanced Accuracy
  - **ROC-AUC**
  - **F1 Score**
- Feature Importance
  - Shapley Value

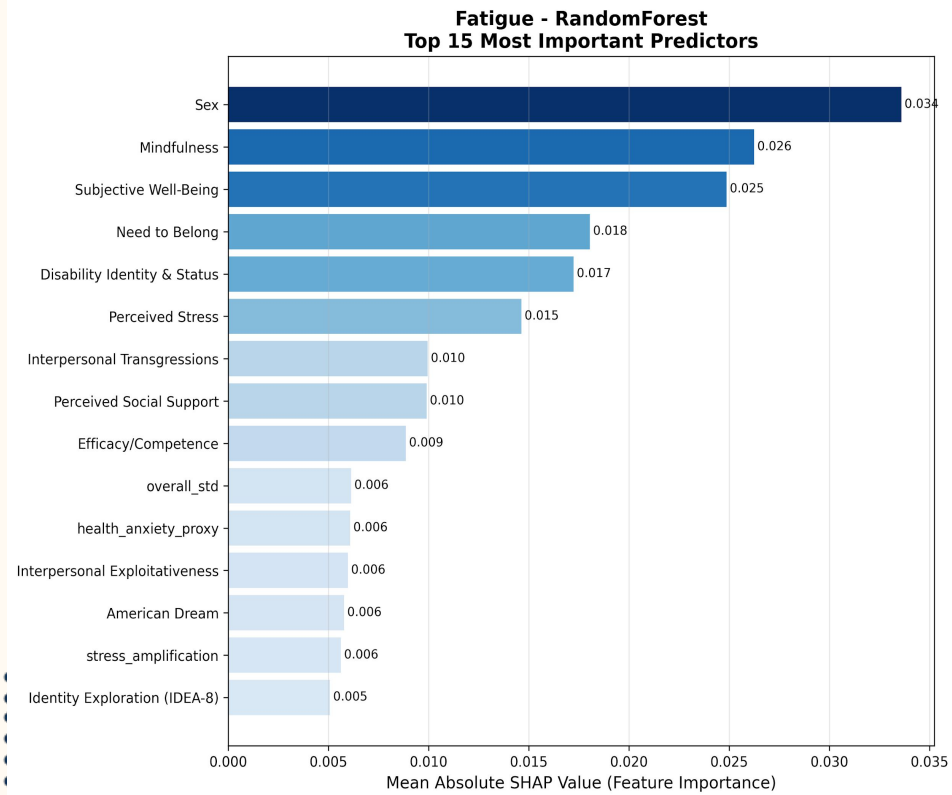
Results: Symptom-Specific Findings

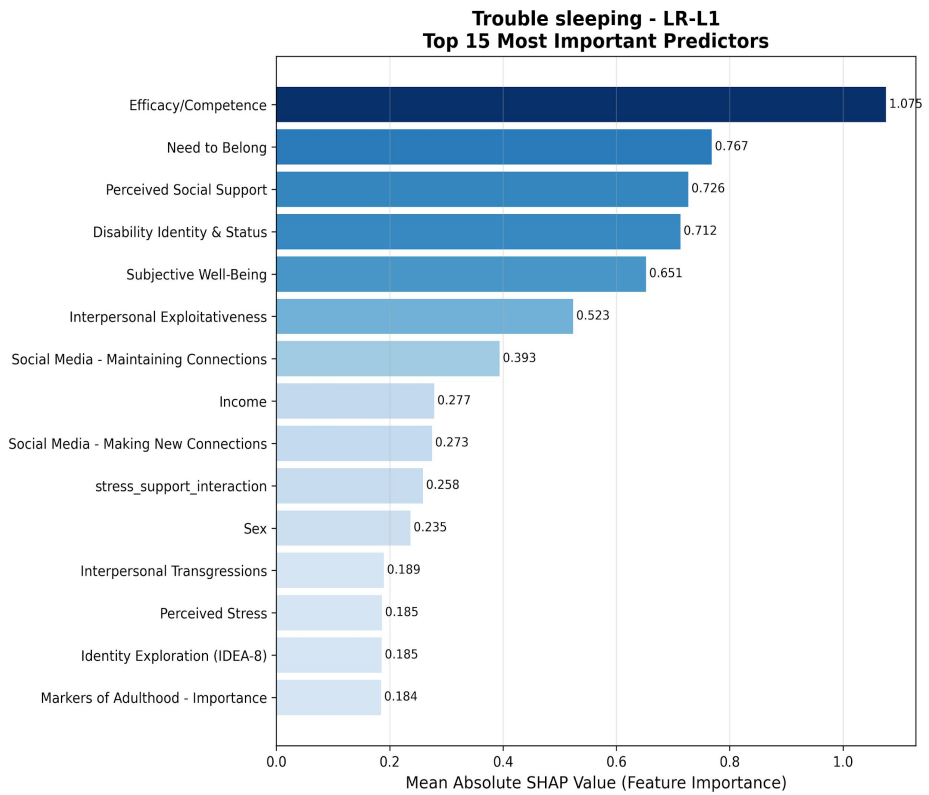


Results: Symptom-Specific Findings



Results: Feature Importance





Discussion & Limitations

Conclusion:

- Different symptoms have different predictors.
- Psychosocial factors matter more for some symptoms (fatigue, sleep problem, headache, dizziness).
- Physiological factors likely play a larger role for others (chest pain, shortness of breath).

Limitation:

- Young adult sample only
- Self-reported symptoms
- Generalizability

Practical & Research Implications

1. **PHQ-15 should not be treated as a single homogeneous construct**; item-level variation is meaningful.
2. Psychological predictors show **symptom-specific relevance**, suggesting **researchers studying somatic symptoms should look at items individually**
3. Symptoms with poor predictability (e.g., chest pain, constipation) may require **biomedical or contextual variables** not present in this dataset.
4. Demonstrates the value of combining **machine learning + explainability tools** (SHAP) in psychosomatic research.

Future Direction

- **Expand to additional somatic symptoms**
  (include more PHQ items and symptom-specific models)
- **Incorporate predictors related to psychopathology**
  (e.g., anxiety, depression, trauma-related symptoms)
- **Validate our models using a national dataset (Add Health)**
  – Test model performance in a large, representative U.S. sample
  – Examine whether relationships replicate across populations
- **Build a more generalized prediction framework**
  – Compare cross-dataset performance
  – Identify stable predictors across multiple symptoms
  – Improve interpretability and potential clinical relevance

THANK YOU