Machine learning algorithms developed from daily wearable and smartphone use characteristics identify clinically significant changes in anxiety and mood

Ji Y[1,2], An X[1,2], Kee S[1,2*], Li Q[1,2], Liu M[2], Vizer L[2], Clifford G[3], Ressler K[4], Kessler R[5], Alvarado M[6], Neylan T[7], and McLean SA[1,2]

From the [1]Institute for Trauma Recovery; [2]Department of Psychiatry, University of North Carolina, Chapel Hill, NC; [3]Emory University School of Medicine; [4]McLean Hospital/Harvard Medical School; [5]Harvard School of Medicine; [6]Mindstrong Health; [7]University of California San Francisco

*Presenting author

UNC Institute for Trauma Recovery

UNC SCHOOL OF MEDICINE

AURORA

Introduction

- This analysis assessed the ability of different types of daily wrist wearable and smartphone data to identify changes in self-report symptoms.
- If wearable and smartphone data can identify changes in health status, they may serve useful as useful clinical decision support tools for patients and their caregivers.

Methods

- The severity of common mental health (Depressive, Anxiety, Re-experiencing, Avoidance, Hyperarousal, and Nightmares) and physical health (Pain, Somatic Symptoms, Sleep Discontinuity, and Concentration/Fatigue) were collected from a large longitudinal cohort of traumatic stress exposure survivors (AURORA cohort, n=2,267), including at week 4 and week 12.
- Daily wrist wearable (activity, heart rate) and passive smartphone data (language, GPS, keystroke) were collected, including between weeks 4 and 12.
- This analysis examined the ability of features derived from wearable and smartphone biomarkers to identify clinically significant change the severity of each symptom was assessed using machine learning models.
- Based on the literature, a change in symptom severity (derived variable) of ± .35 SD between week 4 and week 12 was used to define clinically significant symptom worsening or improvement.

Table 1. Composite biomarker analysis summary – Smartphone and Watch Data

RDoc Outcome			Pain*	Somatic Symptoms	Concentration/ Think/Fatigue	Depressive symptoms	Anxiety	Hyperarousal	Sleep Discontinuity	Nightmares	Avoidance	Re-experiencing
Percentage		Improvement	17%	21%	30%	31%	34%	35%	40%	25%	43%	43%
		Worsening	7%	12%	13%	16%	13%	11%	16%	12%	13%	9%
Best AUC	Smartphone + Watch + Initial Symptom Severity (n=709)	Improvement	0.849 (RF)	0.907 (GB)	0.867 (SVC)	0.848 (GB)	0.819 (GB)	0.871 (GB)	0.867 (SVC)	0.915 (GB)	0.882 (GB)	0.850 (GB)
		Worsening	0.809 (RF)	0.848 (NN	0.770 (NN)	0.815 (SVC)	0.771 (GB)	0.785 (SVC)	0.843 (NN)	0.796 (RF)	0.852 (SVC)	0.713 (KNN)
	Initial Symptom Severity (n=2267)	Improvement	0.748 (KNN)	0.788 (GB)	0.716 (KNN)	0.724 (RF)	0.657 (GB)	0.650 (NN)	0.596 (GB)	0.786 (NN)	0.708 (NN)	0.662 (KNN)
		Worsening	0.663 (KNN)	0.690 (NN)	0.622 (GB)	0.612 (GB)	0.611 (RF)	0.599 (KNN)	0.616 (GB)	0.624 (KNN)	0.607 (NN)	0.553 (NN)
	Smartphone + Watch (n=709)	Improvement	0.836 (RF)	0.814 (GB)	0.858 (SVC)	0.802 (SVC)	0.780 (GB)	0.844 (SVC)	0.867 (SVC)	0.831 (NN)	0.869 (GB)	0.801 (GB)
		Worsening	0.789 (RF)	0.827 (NN)	0.763 (NN)	0.815 (SVC)	0.774 (NN)	0.785 (SVC)	0.836 (NN)	0.790 (SVC)	0.852 (SVC)	0.703 (NN)
	Smartphone (n=950)	Improvement	0.783 (RF)	0.829 (GB)	0.802 (SVC)	0.825 (SVC)	0.774 (GB)	0.850 (SVC)	0.862 (GB)	0.851 (RF)	0.866 (GB)	0.830 (GB)
		Worsening	0.789 (KNN)	0.798 (KNN)	0.791 (NN)	0.813 (NN)	0.725 (GB)	0.823 (KNN)	0.822 (SVC)	0.847 (RF)	0.850 (SVC)	0.701 (KNN)
	Keystroke (n=982)	Improvement	0.791 (RF)	0.817 (RF)	0.815 (SVC)	0.791 (SVC)	0.755 (GB)	0.793 (GB)	0.836 (RF)	0.832 (SVC)	0.840 (GB)	0.787 (GB)
		Worsening	0.721 (SVC)	0.787 (NN)	0.735 (NN)	0.829 (NN)	0.705 (RF)	0.760 (KNN)	0.819 (KNN)	0.847 (RF)	0.834 (SVC)	0.757 (NN)
	Language (n=1113)	Improvement	0.781 (RF)	0.765 (RF)	0.792 (SVC)	0.748 (SVC)	0.762 (GB)	0.821 (SVC)	0.823 (GB)	0.710 (NN)	0.806 (SVC)	0.803 (GB)
		Worsening	0.761 (SVC)	0.792 (NN)	0.738 (RF)	0.654 (KNN)	0.696 (NN)	0.755 (NN)	0.759 (KNN)	0.683 (KNN)	0.730 (NN)	0.730 (GB)
	Location (n=689)	Improvement	0.699 (SVC)	0.718 (RF)	0.772 (GB)	0.743 (NN)	0.739 (SVC)	0.793 (SVC)	0.746 (GB)	0.691 (SVC)	0.722 (GB)	0.653 (GB)
		Worsening	0.753 (SVC)	0.721 (NN)	0.737 (KNN)	0.708 (KNN)	0.714 (SVC)	0.653 (KNN)	0.665 (NN)	0.713 (NN)	0.694 (NN)	0.669 (KNN)
	Watch (n=1563)	Improvement	0.715 (GB)	0.745 (RF)	0.725 (SVC)	0.710 (SVC)	0.750 (SVC)	0.733 (GB)	0.707 (GB)	0.741 (RF)	0.710 (GB)	0.756 (GB)
		Worsening	0.679 (GB)	0.703 (SVC)	0.738 (RF)	0.682 (NN)	0.708 (KNN)	0.669 (RF)	0.693 (KNN)	0.697 (SVC)	0.657 (NN)	0.717 (NN)
	Activity (n=1790)	Improvement	0.654 (NN)	0.666 (SVC)	0.666 (SVC)	0.633 (SVC)	0.692 (RF)	0.665 (GB)	0.702 (SVC)	0.674 (RF)	0.633 (GB)	0.657 (SVC)
		Worsening	0.655 (SVC)	0.665 (NN)	0.659 (NN)	0.607 (RF)	0.670 (NN)	0.634 (SVC)	0.664 (SVC)	0.635 (NN)	0.601 (NB)	0.601 (KNN)
	Heart rate variability (n=1593)	Improvement	0.589 (NN)	0.708 (NN)	0.728 (SVC)	0.661 (SVC)	0.682 (GB)	0.707 (GB)	0.688 (GB)	0.703 (RF)	0.693 (GB)	0.681 (GB)
		Worsening	0.634 (GB)	0.687 (RF)	0.698 (SVC)	0.644 (NN)	0.652 (NN)	0.680 (RF)	0.645 (RF)	0.684 (RF)	0.675 (NN)	0.671 (KNN)

** SMOTE was used for oversampling to address the imbalanced issue in the prediction of pain worsening.*

GB: Gradient Boosting.

KNN: N Nearest Neighbors

NN: Neural Network.

RF: Random Forest.

SVC: Support Vector Classification.

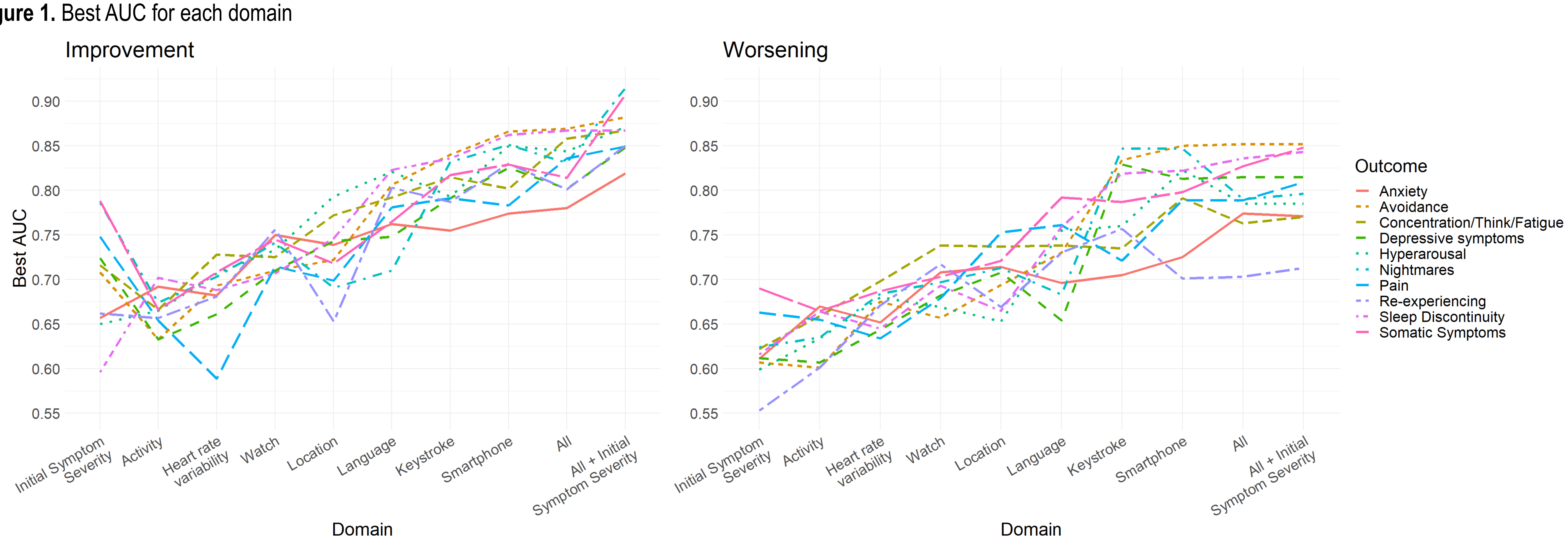
Results

- Within the cohort (n=2,267) and across symptom categories, 6.9%-16.3% and 17.2%-43.0% of individuals experienced clinically significant worsening and improvement of symptoms, respectively.
- In general, across outcomes, features from keystroke had the most classification utility, followed by language, location, heart rate, and activity (Table 1).
- The most useful domain for classification varied across symptoms (Table 1). Prediction was generally more accurate for worsening, perhaps because in this trauma cohort worsening was a more rare/extreme phenotype
- Passive measures models over weeks 4-12 exceeded prediction based on baseline symptom severity (Table 1). Baseline symptoms added to model prediction (Table 1). Predictive utility for many symptom outcomes were excellent (AUC .8-.9).
- The relative ability of different domains to predict clinically significant changes according to symptom is shown in Figure 1.

Conclusions

- Passively collected wearable and smartphone data have the potential to help individuals and their providers identify significant changes in key health status domains.

Figure 1. Best AUC for each domain



References

1. Kessler RC. Posttraumatic stress disorder: the burden to the individual and to society. *The Journal of clinical psychiatry*. 2000;

Funding

Funding for the study was provided by NIMH U01MH110925, the US Army Medical Research and Material Command, The One Mind Foundation, and The Mayday Fund. Verily Life Sciences and Mindstrong Health provided some of the hardware and software used to perform study assessments.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

U.S. Department of Health and Human Services

Supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases