Wearables Predict Stress in Frontline Healthcare Workers

By: Joshua Alvarado, Ian Gascon, Joanne Kim, Mac Kul, Jessica Nguyen, Alex Franks, Julio Vega

UC SANTA BARBARA

Introduction

Our sponsor, Evidation Health, helps connect individual participants to research studies. They partner with leading healthcare companies to analyze the collected clinical data and publish rigorous studies. We are using data collected in "The Stress and Recovery in Frontline COVID-19 Health Care Workers Study." Daily notifications prompted 365 participants to complete tasks through a REDCap study app including a Demographic Survey and a Daily Stress Measure and Work Shift Details survey.

Our project objectives consist of two parts: The first part is to **identify the main factors of stress** (duration of sleep in different stages, heart rate, respiratory rate, and levels of oxygen saturation); and the second part is to **predict whether a participant is stressed given those factors.**

Data

The Demographic Survey indicates that our sample includes 325 Females and 40 Males and that the sample disproportionately represents individuals who identify as White (82.7%).

The Daily Stress Measure survey includes questions regarding daily stress experienced, the participant's work schedule, and whether they were caring for COVID-19 patients.

Garmin Vivoactive 4 Smartwatch

Oura Ring 2

Semi-continuous biometric data were collected from the Oura 2 smartring worn while off shift for the duration of the study. A subset of participants (95 workers) wore a Garmin smartwatch while on shift for 4 consecutive weeks.

Exploratory Data Analysis

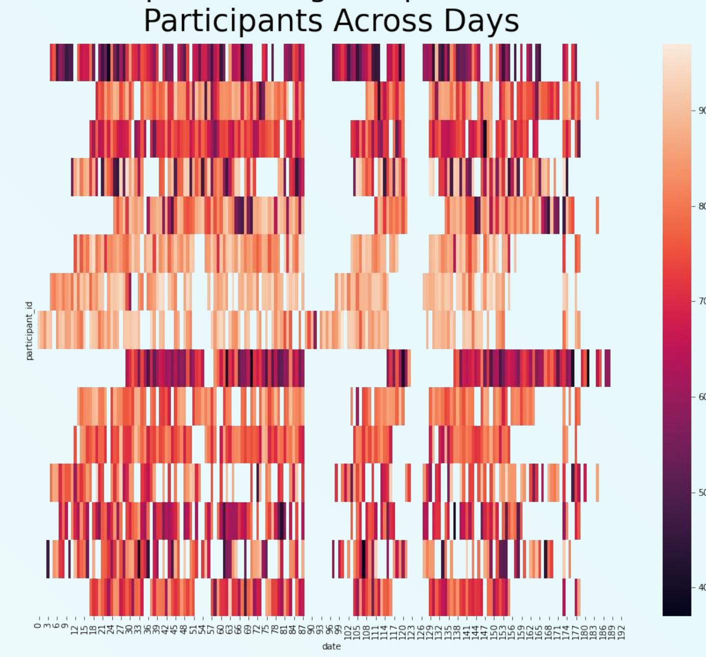

Figure 1: Heatmap containing the sleep scores for the top 20 participants (ranked by number of observations) across time. Darker colors indicate less sleep, highlights how average amount of sleep varies by participant.

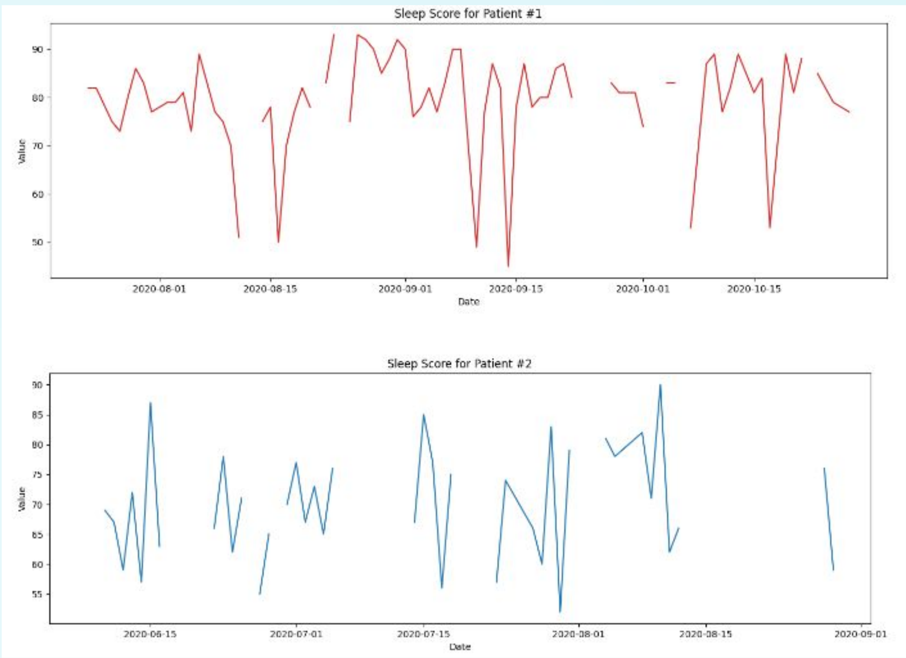

Figure 2: These two plots show the range of missing values within the participants we had to work with. Some have very little missing values while others are missing quite a few.

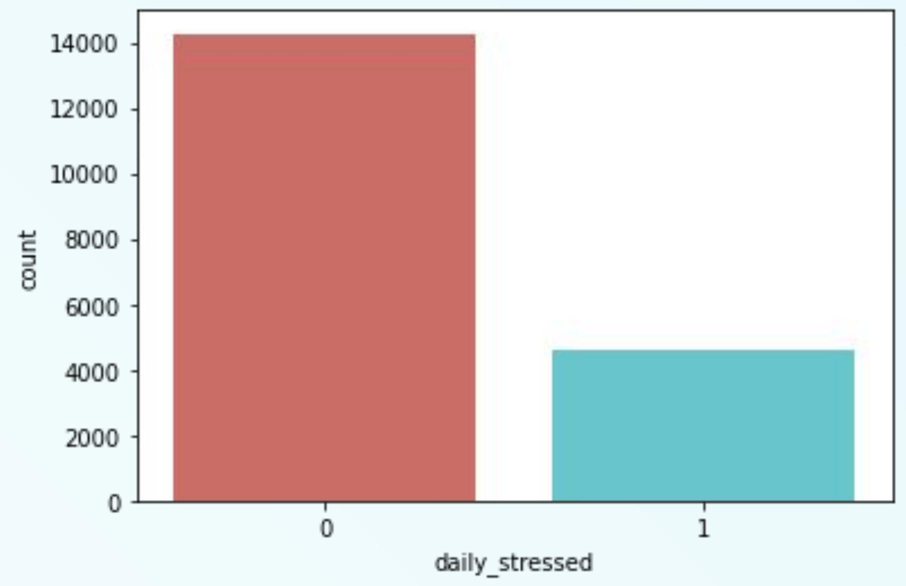

Figure 3: This figure shows the amount of 0's (no stress) dominates over the amount of 1's (stress) from the redCAP data. Data imbalanced presents a problem because the trained models will have a bias towards 0's.

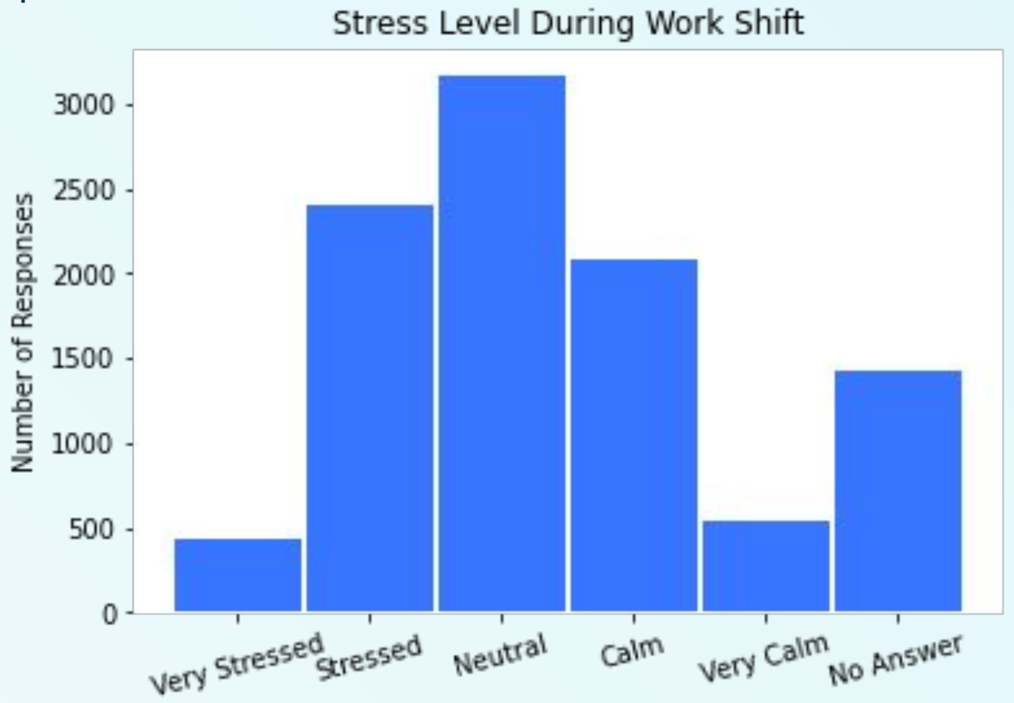

Figure 4: Bar chart showing the cumulative responses for daily stress levels from all 365 healthcare workers surveyed daily.

Method

- **Data Preprocessing**
  - Data was missing for varying reasons: participants may not have worn their devices everyday; participants may have only had certain variables measured thus having more or less observations in some variables than others.
  - Missing data was imputed with the average of the last three rows to allow the most recent values to have the most impact on filling in a missing value.
  - The residual values for each feature were calculated in order to account for varying baselines of sleep, heart rate, etc. among participants.
  - The scikit package SMOTE was used to help correct for the imbalance data.
- **Modeling**
  - A logistic regression model was used to measure which variables predicted stress strongly.
  - When predicting whether a participant was stressed or not, a variety of models were used including Logistic Regression, LightGBM, XGBoost, and a Decision Tree Classifier.
- **Evaluation**
  - The precision, accuracy, and recall of the models were pitted against a baseline dummy model.

Results

Oura Ring Data			
Model	**Accuracy**	**Precision**	**Recall**
Stratified KFold (Baseline)	0.635	0.245	0.245
Logistic Regression	0.550	0.548	0.550
DT	0.577	0.564	0.685
XGBoost	0.577	0.564	0.686
LightGBM	0.626	0.621	0.646

Garmin Watch Data			
Model	**Accuracy**	**Precision**	**Recall**
Stratified KFold (Baseline)	0.600	0.268	0.268
Logistic Regression	0.529	0.523	0.534
DT	0.601	0.563	0.904
XGBoost	0.591	0.569	0.728
LightGBM	0.746	0.738	0.765

Discussion

When looking at the recall for each model listed above, XGBoost and the Decision Tree had the highest values for the Oura and Garmin data, respectively. On the other hand, it cannot be concluded that wearable device data can be reliable predictors of stress, since the accuracy results for the models fit on both the Oura Ring and Garmin Watch datasets were suboptimal. However, we believe it is possible to expand on and further examine this topic. Some next steps would include: focus on stress throughout the week instead of holistically; curate a larger Garmin Watch dataset; perform rigorous feature selection to collect the most effective variables; improve cross-validation techniques to be better suited for time series data; and further explore the performance of LightGBM.

References

Goodday S, Karlin E, Alfarano A, Brooks A, Chapman C, Desille R, Rangwala S, Karlin D, Emami H, Woods N, Boch A, Foschini L, Wildman M, Cormack F, Taptiklis N, Pratap A, Ghassemi M, Goldenberg A, Nagaraj S, Walsh E, Stress And Recovery Participants , Friend S. "An Alternative to the Light Touch Digital Health Remote Study: The Stress and Recovery in Frontline COVID-19 Health Care Workers Study". DOI: 10.2196/32165

Acknowledgements

We would like to thank our mentor Professor Franks, our project sponsor Julio Vega, and the team at Evidation for their guidance and support. We would also like to thank the Central Coast Data Science Partnership and all of the staff involved for giving us this opportunity.