



Published in final edited form as:

*J Affect Disord.* 2021 March 01; 282: 104–111. doi:10.1016/j.jad.2020.12.086.

## Deep Learning Paired with Wearable Passive Sensing Data Predicts Deterioration in Anxiety Disorder Symptoms across 17–18 Years

Nicholas C. Jacobson, PhD<sup>1,2,3,4,\*</sup>, Damien Lekkas, MS<sup>1,4</sup>, Raphael Huang<sup>1</sup>, Natalie Thomas, BS<sup>1</sup>

<sup>1</sup>Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College

<sup>2</sup>Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College

<sup>3</sup>Department of Psychiatry, Geisel School of Medicine, Dartmouth College

<sup>4</sup>Quantitative Biomedical Sciences Program, Dartmouth College

### Abstract

**Background.**—Recent studies have demonstrated that passive smartphone and wearable sensor data collected throughout daily life can predict anxiety symptoms cross-sectionally. However, to date, no research has demonstrated the capacity for these digital biomarkers to predict long-term prognosis.

**Methods.**—We utilized deep learning models based on wearable sensor technology to predict long-term (17–18-year) deterioration in generalized anxiety disorder and panic disorder symptoms from actigraphy data on daytime movement and nighttime sleeping patterns. As part of Midlife in the United States (MIDUS), a national longitudinal study of health and well-being, subjects ( $N = 265$ ) (i) completed a phone-based interview that assessed generalized anxiety disorder and panic disorder symptoms at enrollment, (ii) participated in a one-week actigraphy study 9–14 years later, and (iii) completed a long-term follow-up, phone-based interview to quantify generalized anxiety disorder and panic disorder symptoms 17–18 years from initial enrollment. A deep auto-encoder paired with a multi-layered ensemble deep learning model was leveraged to predict whether participants experienced increased anxiety disorder symptoms across this 17–18 year period.

**Results.**—Out-of-sample cross-validated results suggested that wearable movement data could significantly predict which individuals would experience symptom deterioration (AUC = 0.696, CI [0.598, 0.793], 84.6% sensitivity, 52.7% specificity, balanced accuracy = 68.7%).

\*46 Centerra Parkway, Suite 300, Office # 333S, Lebanon, NH 03766, phone: (603) 646-7037; Nicholas.C.Jacobson@dartmouth.edu. Author Contributions

NCJ contributed to methods development, study design, data analysis and interpretation, and manuscript review; DL contributed to literature searches, figure creation, data interpretation, manuscript write-up, and manuscript review; RH and NT contributed to literature searches and manuscript write-up.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflicts of Interest

The authors do not declare any conflicts of interest.

**Conclusions.**—Passive wearable actigraphy data could be utilized to predict long-term deterioration of anxiety disorder symptoms. Future studies should examine whether these methods could be implemented to prevent deterioration of anxiety disorder symptoms.

Anxiety disorders have a collective lifetime prevalence of 24.9%, making them the most commonly occurring group of disorders after substance use disorders (Bruce et al., 2005). Specifically, generalized anxiety disorder (GAD) and panic disorder (PD) affects 6 million and 8 million Americans each year, respectively (Kessler et al., 2012). Individuals affected by GAD experience persistent, uncontrollable worry, while PD symptoms include sudden episodes of anxiety, often accompanied by cardiorespiratory and otoneurological discomforts such as dizziness and tinnitus (Roy-Byrne et al., 2006). Individuals affected by these illnesses experience deterioration in health-related quality of life (sleep disturbances, headaches, restlessness) and impairment in the ability to complete tasks (Markowitz et al., 1989; Toghiani et al., 2014). These symptoms are disabling and have a significant toll on both national productivity and individual quality of life (Hoffman et al., 2008). Additionally, these anxiety disorders often require expensive primary care services, costing an estimated \$33.71 billion in annual US public health expenditures (Shirneshan et al., 2013; Wittchen, 2002).

Despite the clear costs to individuals and society, GAD and PD suffer high misdiagnosis rates, estimated at 71.0% and 85.8% respectively by one study diagnosing patients with the neuropsychiatric “MINI” scale and comparing the results to their existing clinical records (Vermani et al., 2011). Furthermore, only an estimated 40% of those suffering from anxiety disorders seek treatment in the same year as the first onset of disorder (Christiana et al., 2000). These shortcomings in diagnosis and treatment contribute to a median delay of eight years between onset of anxiety disorders and appropriate treatment (Christiana et al., 2000). During this time period, the disease lowers the quality of life for affected individuals and in some cases continues to grow in severity, thereby increasing risk due to all-cause mortality in older adults (Lenze & Wetherell, 2011).

Early identification of at-risk individuals allows for the potential to offer preemptive, preventative treatments which have proven to be effective in reducing prevalence and lowering disorder-associated disabilities (Essau et al., 2012; Feldner et al., 2008). For example, a study examining the effects a universal anxiety prevention program on school children found a significant decrease in anxiety symptoms for children who were enrolled, and as a result, received psychoeducation, relaxation, and positive-self-talk within a cognitive behavioral therapy (CBT) framework (Lau & Rapee, 2011). The literature also indicates that preventative, internet-delivered, cognitive-based therapy has promising clinical efficacy in the prevention of anxiety symptoms six weeks after intervention (Kenardy et al., 2003). Regardless of methodology, the ability to provide effective preventative care and treatment depends on timely identification of afflicted individuals.

To this end, researchers have sought to predict the trajectory of anxiety disorders using patient characteristics. These include the type of anxiety disorder (e.g. panic disorder with or without agoraphobia), as well as clinical variables like severity, duration, and level of disability (Spinhoven et al., 2016). While most studies have been cross-sectional in nature,

the realization that anxiety disorder symptoms are often recurring and fluctuate dynamically from period to period has prompted the implementation of longitudinal studies (Bruce et al., 2005). For example, a study using latent growth mixture modeling (LGMM) found that clinical variables (i.e. severity and duration of anxiety) better predicted long term anxiety trajectory compared to diagnostic (type of anxiety disorder) and personality trait variables (i.e. neuroticism, extraversion) (Spinoven et al., 2016). A Netherlands longitudinal study followed a cohort over an eight-year period to examine long-term predictors of anxiety (Penninx et al., 2008). Subsequent analysis of the data found that severity and duration of index episode, comorbid depression-anxiety, and earlier onset age were associated with worse symptom course trajectories (Penninx et al., 2011). Additionally, the “late chronotype” defined in this study -- being an evening rather than a morning person -- was found in later analyses of this longitudinal data to be significantly associated with cross-sectional anxiety and major depression symptoms (Antypa et al., 2016). Both the Spinoven and the Penninx studies relied on clinical interviews and detailed assessments to collect baseline and follow-up data from which predictors of anxiety symptoms were derived. While these methods tracked long-term changes, they relied on clinical interviews that lasted up to four hours in some cases (Penninx et al., 2008), thereby making it challenging to scale to larger sample sizes.

Advancements in the field of machine learning aim to make informed predictions that replace the burden of lengthy clinical interviews with rapid processing of patient data. A comprehensive literature review of works published between 2005 and November 2017 that interrogated anxiety disorder through the application of machine learning techniques identified twenty studies (Kotsilieris et al., 2018). While the majority of the studies highlighted in this review achieved high prediction accuracy (>80%) with the help of novel predictors such as heart-rate measurements (Chatterjee et al., 2014) and various biomarkers in the brain’s gray matter (Chi et al., 2014), the application of the models were diagnostic of cross-sectional anxiety rather than predictive of long-term anxiety trajectory. One particular study used self-esteem data collected longitudinally (at age 13, 16, and 22) through DSM-IV structural interviews to predict anxiety at age 33 (Chen et al., 2015). The application of a Bayesian joint model to this data yielded an AUC of 0.75. The model was agnostic to the temporality of anxiety predictors since immediate anxiety was diagnosed using both present and past symptomatic features. As such, the clinical utility was diminished given the lack of capacity to predict anxiety trajectory solely from past data.

Thus, most research has only examined the ability of machine learning to cross-sectionally predict anxiety disorders, and the sparse longitudinal machine learning research to date has utilized time-intensive and costly measures (Balogh et al., 2015; Bor, 2015), thereby decreasing the potential to scale this research and utilize it within larger populations. Fortunately, passive, continuous sensing of movement data is one avenue by which this resource burden can be mitigated to yield rich data for predictive purposes. Moreover, research in this domain has indicated that passive data can aid in the identification of digital biomarkers which, when paired with machine learning methodologies, can assess psychiatric symptoms with high accuracy (Jacobson, Weingarden, et al., 2019; Jacobson et al., 2020; Jacobson & O’Cleirigh, 2019) and predict symptom change across time (Jacobson, Weingarden, et al., 2019). This suggests that digital biomarkers formed from movement data

may have the potential to enable scalable assessment of anxiety symptoms, but more research is needed to determine whether these digital biomarkers have longitudinal prognostic value.

Our study applies an ensemble machine learning model to process passively collected actigraphy data and predict deterioration in anxiety symptoms 17–18 years in the future. As discussed above, the temporal component to our analytical procedure is distinct in scope from cross-sectional approaches employed in prior studies. Unlike some prior studies which used different biomarkers, the current study made use of movement data as indexed through an actigraph. Additionally, the passively collected data we leverage is less burdensome, less biased, and more naturalistic than the clinical interview data that is traditionally collected (Chen et al., 2015; Penninx et al., 2008). In a practical sense, such data lends itself as a scalable and potentially wide-reaching tool for the identification of deterioration in at-risk persons suffering from anxiety-related disorders. Based on prior studies indicating a relationship between eye movement, sleep and anxiety (Mellman, 2006; Mogg et al., 2000), we hypothesized we could predict deterioration in anxiety symptoms over 17–18 years with high precision using machine learning models based on digital biomarkers formed from movement data.

## METHODS

### Participants

Participants ( $N = 265$ ; 58.1% female, 41.9% male; age 25 – 72 years, mean 44.3 years; 94.7% White, 1.5% Black, 0.4% Native American/Alaskan/Aleutian Islander, 0.4% Pacific Islander, 0.4% multiracial, and 0.8% other/no response), as shown in Table 1, were part of the MIDUS study completing each of the three phases beginning in 1995 with wave 1 (MIDUS-1), with subsequent wake-sleep actigraphy data collection in wave 2 (MIDUS-2), as well as a wave 3 follow-up (MIDUS-3). The first phase, occurring between 1995 and 1996, involved individuals from across the nation aged 25–75 years who completed telephone interviews to capture mental health symptoms. The second phase, occurring between 2004 and 2009, collected wearable movement data from participants out of the UW-Madison study site. In the third phase, occurring between 2013 and 2019, participants were re-interviewed regarding their mental health symptoms. Five percent of the participants ( $n=13$ ) had seen a psychiatrist and 11.69% ( $n=31$ ) of the participants had seen a psychologist in the past 12 months. At baseline, 6.75% ( $n=18$ ) and 1.88% ( $n=5$ ) of the participants were diagnosed with PD and GAD, respectively. Of those percentages, only 3 persons had both PD and GAD. None of those participants diagnosed with PD or GAD had previously received antidepressants. MIDUS data collection was reviewed and approved by the Education and Social/Behavioral Sciences and the Health Sciences IRBs at the University of Wisconsin-Madison. Subjects were selected for analysis based on their successful completion of all three phases of the study, with participation in the collection of passive sleep-wake actigraphy information in wave 2 as the most limiting inclusion criterion.

## Measures

**Anxiety Disorder Assessment using the Composite International Diagnostic Interview (CIDI).**—Each participant was interviewed using screening versions of the World Health Organizations' CIDI version 10 (Robins et al., 1988; World Health Organization, 1990), assessing symptoms of GAD and PD according to the Diagnostic and Statistical Manual III-R criteria. The CIDI has demonstrated good test-retest reliability and strong inter-rater reliability (Andrews and Peters, 1998) with Kappa scores of 0.41 for GAD and 0.84 for PD between lifetime DSM-III diagnoses using the CIDI (Semler et al., 1987). As the most widely used summary of self-report for depression (Rosenström et al., 2015), the sum score of symptom items were used to create a continuous measure of anxiety symptom severity at MIDUS-1 and MIDUS-3. This approach to quantification has been replicated in other studies with CIDI symptom items (Luutonen et al., 2013; Orlando et al., 2001; Zainal & Newman, 2019). As our study is specifically interested in the ability to predict long term anxiety deterioration from these counts, the outcome metric for the model was whether there was an increase in the number of total GAD and PD symptoms experienced between MIDUS-1 to MIDUS-3.

**Passive Movement Data.**—Passive sensing data on wake-sleep rhythms was obtained from the Mini Mitter Actiwatch©-64 wearable device. This information was collected across seven consecutive, uninterrupted days, binned into epochs of thirty second intervals. For each participant, this equates to data across seven rest/sleep periods and six activity periods. Raw data on the *total activity counts*, *maximum activity counts*, *average activity counts/minute*, *wake time*, *percent wake time*, *total number of wake bouts*, *average number of wake bouts*, *sleep time*, *percent sleep time*, and *total number of sleep bouts* defines the sleep/rest and active periods. Additionally, the Actiware 5 software has generated a variety of summary statistics on activity and sleep, including *sleep onset latency (in minutes)*, *time dozing before rising in minutes (i.e. snooze time)*, *percent sleep efficiency*, and *wake after sleep onset (WASO) in minutes*.

*Sleep onset latency* is the total time elapsed in minutes between the start of a rest interval and the initiation of sleep, or the time required for sleep onset from the start of a participant's attempt to fall asleep. *Snooze time* is the total time elapsed in minutes between the termination of sleep and the end of a rest interval, or the time from when a participant wakes up to when they get out of bed. *Sleep efficiency* is a descriptor that combines three additional metrics, *scored total sleep*, *interval duration*, and *total invalid time*. *Scored total sleep* is the number of epochs between the start and end times of an interval scored as "sleep" and multiplied by the epoch length in minutes. The *interval duration* is the total time in minutes between the beginning and end of an interval, and *total invalid time* is a count of the epochs where the activity counts exceed the threshold of possible value for that interval and is therefore a measure of hardware/software error collection/reporting. With these three metrics, *sleep efficiency* is defined as the proportion or percentage of *scored total sleep time* to *interval duration* after correcting for spurious data through *total invalid time*. *WASO* is the total number of wake epochs spanning an interval of sleep multiplied by the epoch interval time in minutes. Taken together, these statistics are representative of circadian activity patterns and irregularities.

**Active Anxiety Symptom Monitoring Comparison.**—At the suggestion of an anonymous reviewer, we also utilized the first observed anxiety data from the following questions during the second wave as a baseline comparison to determine whether conventional measurements using five face-valid indicators of anxiety symptoms prognostically predict GAD and PD symptoms. The current face-valid indicators were: “Did you feel nervous”, “Did you feel jittery”, “Did you feel restless or fidgety”, “Did you feel calm and peaceful?”, and “Did you feel afraid”, each of which were rated on a 0 (“None of the Time”) to 4 (“All of the Time”) Likert scale. These items reflected the first observed measurements from within a daily diary study and included a total of 196 of the participants.

## Planned Analysis

**Feature engineering.**—Summary statistics for each of 19 sleep actigraphy variables were calculated from the MIDUS-2 sleep actigraphy dataset (see Figure 1 for a graphical illustration of the full feature engineering and analysis pipeline). These included the maximum, minimum, mean, median, mode, standard deviation, skewness, kurtosis, root mean square of successive differences (RMSSD), and quantiles ranging from 20th percent up to 80th percent in 20 percent increments. Differential Time-Varying Effect Models (DTVEMs) interrogating lag relationships between 1 and 5 time points later were implemented to return derivative actigraphy features describing potential autoregressive relationships (Jacobson, Chow, et al., 2019). Spectral analysis was also performed on this data to capture additional features reflecting underlying periodicities. The resulting feature set consisted of 800 features. The predictive outcome of interest was defined by a difference of the composite GAD and PD questionnaire response scores between MIDUS-1 and MIDUS-3, representing a change in anxiety across 17–18 years.

**Non-linear dimension reduction.**—An unsupervised deep neural network autoencoder was utilized prior to modeling the outcome data to reduce the 800 features down to 50 features prior to modeling. The deep autoencoder consisted of 10 layers with exponential linear unit activation functions. Layer density varied from 2000 down to 50 units. Implementation of this neural net bottleneck reconstructed a condensed representation of the feature input space that could then be fed into a machine learning framework.

**Model Development and Validation.**—Similar to prior research showing strong predictive performance of ensemble models (Nemesure et al., 2020), we utilized an ensemble approach to predict long-term anxiety symptom deterioration. Specifically, we utilized neural network models, splines, ridge regression, random forests, general linear models, gaussian process, extreme gradient boosting, k-nearest neighbors, and support vector machine models as base-learners. All models were run with four-fold cross validation with maximum correlation between the predicted and actual change in the number of anxiety symptoms between MIDUS-1 and MIDUS-3. Any best-tuned hyperparameter configuration of a model capable of making predictions was retained, and its prediction was saved as a feature within a new feature space. This new ensemble feature vector was then fed into an extreme gradient boosted tree (“xgBoost”) tree model with four-fold cross validation to ultimately predict anxiety deterioration. Model predictions were standardized to be between

0 and 1. Next, performance of this final model was quantified using a confusion matrix and an associated ROC-curve.

**Comparison to traditional assessment approaches.**—Based on suggestions from an anonymous reviewer, we also compared the predictive performance of these deep learning ensembles with passive sensing data against simple logistic regression based on active self-reported anxiety measurements. Thus, we utilized a four-fold cross validation approach to predict symptom anxiety symptom deterioration using active five face-valid anxiety measurements in the middle assessment period.

## RESULTS

Out-of-sample cross-validated results suggested that the features extracted from the deep autoencoder could significantly predict which persons would experience symptom deterioration across 17–18 years (AUC = 0.696, CI [0.598, 0.793], see Figure 2; see appendix 1 for the performance of the lower level models). The sensitivity was 84.6% and specificity was 52.7% based on the point closest to 1 on the AUC curve (i.e. Euclidean distance) (Perkins & Schisterman, 2006). This results in a predicted balanced accuracy of 68.7%.

The results of the deep autoencoder and ensemble pipeline were then compared with traditional approaches within psychology (i.e. logistic regression with active survey measurement data based on symptom reports in the intervening years predicting long-term anxiety symptom deterioration. The results suggested that the conventional approaches did not predict long-term deterioration in anxiety symptoms with greater than chance (AUC = 0.505). This suggests that the deep autoencoder and ensemble pipeline using the passive sensing data holds far greater predictive validity than more conventional approaches.

## DISCUSSION

The current model utilizes longitudinal passive sensing data collected across 17–18 years to predict symptom deterioration associated with GAD and PD. The results suggest that this model has the capability to predict anxiety symptom deterioration with an above chance accuracy across almost two decades (84.6% sensitivity, 52.5% specificity), which suggests that this approach could hold prognostic value in evaluating the potential to identify GAD and PD symptom deterioration. The results are especially notable because they suggest an ability to predict change in long-term GAD and PD symptoms, whereas previous studies have leveraged predictive models within more restrictive, cross-sectional paradigms (Fukazawa et. al, 2019, Sano et. al, 2013). GAD and PD are heavily burdensome (Maki et. al, 2003) and often go undetected (Fifer et al, 1994), therefore the efforts presented herein offer a potentially promising approach for the early prevention interventions related to GAD and PD symptoms.

Our model demonstrated a balanced accuracy of 68.7%, with an AUC of 0.696. Such results illustrate the prospective viability of using passively collected biomarkers to develop informed predictions on long-term anxiety progression. A sensitivity of 84.6% indicates that

the model is highly accurate in its ability to correctly identify positive results (those with anxiety deterioration); on the other hand, a specificity of 52.7% shows that it is not as effective in correctly categorizing those without anxiety deterioration. The relatively low specificity may lead to a higher number of false negatives and is therefore an area for future model improvement. Nevertheless, given the pernicious nature of anxiety disorders, it is likely preferable to correctly identify susceptible individuals at a higher rate and risk unnecessary treatment on a smaller subset of healthy individuals.

Our model's AUC of 0.696 is on par with comparable machine learning studies that predict anxiety across shorter periods of time (ranging from same-day to two years) using various physical and psychological characteristics (Chen et al., 2015; Kotsilieris et al., 2018). This above-chance predictive capacity suggests a potential for clinical utility. As an additional benefit, the collection of passive actigraphy information for model utility results in a process that is less obtrusive for the patient and less time-consuming for the provider when compared with previous prediction pipelines. Ultimately, this approach has the potential to result in improved quality of care for the patient as well as better cost efficiency in diagnosing disease.

One distinguishing feature of this project is the long term horizon of prediction. Predicting the course of anxiety 17–18 years into the future is significantly longer compared to existing studies involving either cross-sectional diagnostics (Antypa et al., 2016; Chatterjee et al., 2014; Chi et al., 2014), or short term predictive windows of 1–2 years (Månsson et al., 2015; Murphy et al., 2008). Making long term predictions may allow for the opportunity to intervene in less intrusive and more effective ways. Given the relationship of this data to movement, it may also suggest that future research should explore the relative performance of biofeedback from digital devices as a potential prevention intervention strategy. If developed and validated in future work, biofeedback prevention interventions could be delivered via altered suggestions via smartwatches to change small behavioral patterns. The potential to deliver prevention interventions is particularly important, since anxiety disorders are disruptive to many everyday activities and in some cases become more severe over time, posing increased all-cause mortality risk for adults (Lenze & Wetherell, 2011).

The results of our investigation are encouraging, demonstrating the potential of utilizing passive sensor data for machine learning models to make long-term predictions. However, there are several limitations. A low specificity of 52.7% suggests our current model may be oversensitive. To remedy this, future efforts can screen out individuals who may be false positives according to longitudinal re-assessments with subsequent application of traditional measures to detect and treat potential symptom deterioration instead. Given the requirement of participation in all three phases of MIDUS, there may also have been confounding in terms of the level of compliance as well as with age in the analyzed cohort (e.g. older individuals at baseline may not have passed away). Another confounding factor concerns the nature of our predictors. Because sleep-based data was used, it is possible that the captured deterioration reflected behaviors associated with chronic insomnia, rather than those stemming primarily from GAD and PD. Additionally, because the data we used was collected in three distinct phases over a 17–18 year interval, the resulting model may not account for more nuanced temporal changes that may have occurred between these



collection periods. Future studies may address this weakness by adding more frequent outcome assessments to capture the finer dynamics of symptom change and potentially increase the accuracy of the model. At the same time, the collection of more data can mitigate issues of oversensitivity and ultimately lead to a more direct and detailed prediction of GAD. This could also allow for more fine-grained assessments of potential sequential comorbidities, as anxiety and depressive disorder symptoms are often dynamically linked across time (de Graaf et al., 2003; Hek et al., 2011; Lamers et al., 2011; Jacobson et al., 2017; Jacobson & Newman, 2014, 2016). Indeed, while the study has shown promise in the ability to predict GAD and PD symptom deterioration over long periods of time using passive sensing information, it does not explore the dynamics and impacts of related, co-occurring psychological processes on this deterioration. Although the current research examined predictive performance within a research setting, further work is needed to determine whether similar methods would have performance in real world applications (e.g. where participants might have very irregular and sporadic wear patterns). By extension, the low prevalence of GAD and PD in the analyzed cohort limits generalizability to clinical populations. Subsequent research will benefit from deployment in these settings. Taken together, such efforts will prove invaluable for informing development and refinement of evidence-based interventions. It is also important to note that MIDUS did not assess anxiety disorder outside of GAD and PD diagnoses, and therefore this analysis did not fully capture other anxiety symptom trajectories.

Overall, this research demonstrates significant promise in the ability to preemptively predict GAD and PD symptom exacerbation. Early identification of potential symptom deterioration has the potential to offset disease burden and/or related comorbidities (“Effective Recognition and Treatment of Generalized Anxiety Disorder in Primary Care,” 2004). Studies have indicated that first presentation of GAD and PD symptoms for many individuals is in a primary care setting and is oftentimes incorrectly diagnosed, thereby limiting subsequent treatment opportunities (Allgulander, 2006). Although more research is needed to determine whether this pipeline would work well in production in primary care settings, this research may suggest that wearable movement data might have the potential to add high prognostic value and could help to narrow the longstanding wait between symptom deterioration and treatment initiation.

### Financial Support:

The MIDUS study has been funded by the following: John D. and Catherine T. MacArthur Foundation Research Network, National Institute on Aging (P01-AG020166), National Institute on Aging (U19-AG051426). MIDUS Biomarker data collection was further supported by the National Institutes of Health National Center for Advancing Translational Sciences (NCATS) Clinical and Translational Science Award (CTSA) program as follows: UL1TR001409 (Georgetown), UL1TR001881 (UCLA) and 1UL1RR025011 (UW). The current investigation was funded by the National Institute of Drug Abuse (P30 DA02992610) and the National Institute of Mental Health (R01 MH123482).

## Appendix 1

### Lower Level Model AUCs

Lower Level Model	AUCs
avNNet	0.51384
bagEarth	0.56767
bagEarthGCV	0.627615
blasso	0.505632
blassoAveraged	0.502092
bridge	0.537496
cforest	0.629224
ctree	0.464355
ctree2	0.55037
cubist	0.518828
earth	0.493804
gamboost	0.558899
gaussprLinear	0.660927
gaussprPoly	0.508851
gaussprRadial	0.625201
gbm	0.57821
gcvEarth	0.588268
glm	0.550853
glmboost	0.57982
glmnet	0.516575
glmStepAIC	0.652398
icr	0.538622
kernelpls	0.53122
kknn	0.622063
knn	0.492758
lars	0.610396
lars2	0.497103
leapBackward	0.51046
leapForward	0.504345
leapSeq	0.592855
lm	0.643869
lmStepAIC	0.67364
mlp	0.523656
mlpML	0.543772
mlpWeightDecay	0.487287
mlpWeightDecayML	0.535726
nnet	0.515127

Lower Level Model	AUCs
npls	0.478919
parRF	0.512955
pcaNNet	0.579176
pcr	0.542002
penalized	0.543772
pls	0.507242
ppr	0.581107
ranger	0.50354
rbfDDA	0.506759
rf	0.612005
rlm	0.523495
rpart	0.526955
rpart1SE	0.436112
rpart2	0.517058
RRF	0.521242
RRFglobal	0.50177
rvmLinear	0.476505
rvmRadial	0.525587
simpls	0.541036
spls	0.484551
svmLinear	0.672192
svmLinear2	0.550048
svmLinear3	0.600418
svmPoly	0.477148
svmRadial	0.590924
svmRadialCost	0.658674
svmRadialSigma	0.58674
treebag	0.574992
widekernelpls	0.534599
WM	0.56767
xgbDART	0.507
xgbLinear	0.52607
xgbTree	0.55375

## References

- Allgulander C (2006). Generalized anxiety disorder: What are we missing? *European Neuropsychopharmacology*, 16, S101–S108. 10.1016/j.euroneuro.2006.04.002 [PubMed: 16730165]
- Antypa N, Vogelzangs N, Meesters Y, Schoevers R, & Penninx BWJH (2016). Chronotype Associations with Depression and Anxiety Disorders in a Large Cohort Study. *Depression and Anxiety*, 33(1), 75–83. 10.1002/da.22422 [PubMed: 26367018]

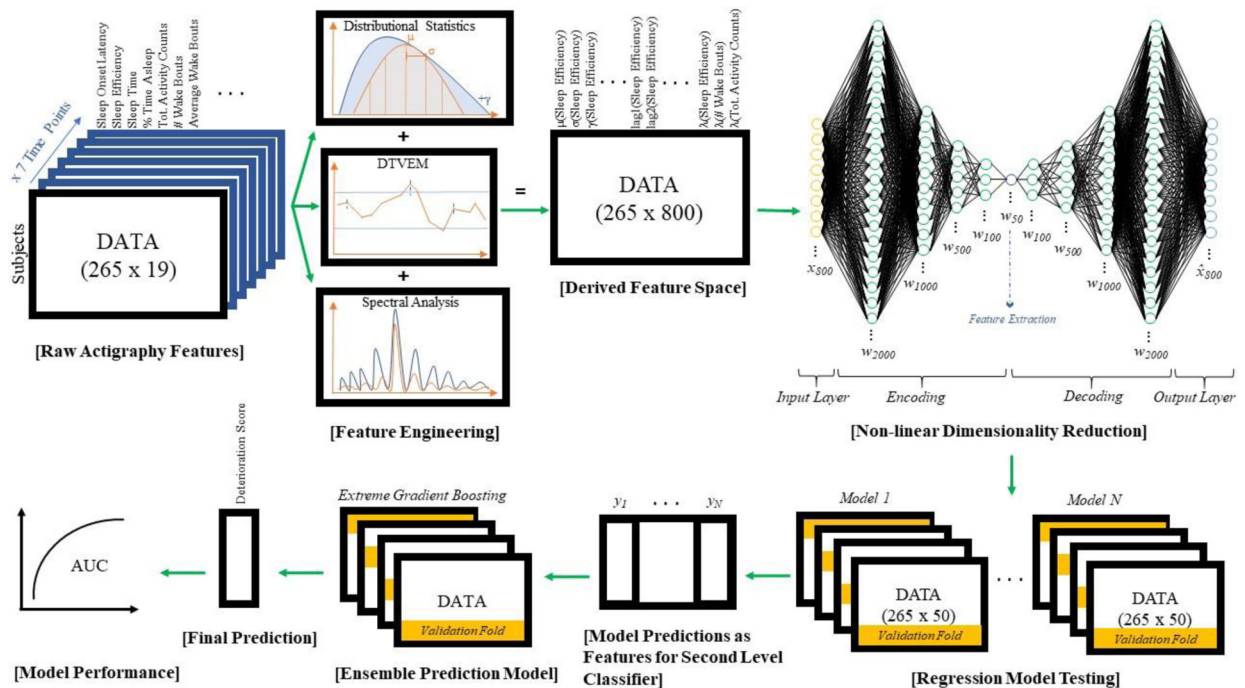
- Balogh EP, Miller BT, Ball JR, Care, C. on D. E. in H., Services, B. on H. C., Medicine, I. of., & The National Academies of Sciences, E. (2015). *The Diagnostic Process In Improving Diagnosis in Health Care*. National Academies Press (US) <https://www.ncbi.nlm.nih.gov/books/NBK338593/>
- Bor JS (2015). Among The Elderly, Many Mental Illnesses Go Undiagnosed. *Health Affairs*, 34(5), 727–731. 10.1377/hlthaff.2015.0314 [PubMed: 25941272]
- Bruce SE, Yonkers KA, Otto MW, Eisen JL, Weisberg RB, Pagano M, Shea MT, & Keller MB (2005). Influence of Psychiatric Comorbidity on Recovery and Recurrence in Generalized Anxiety Disorder, Social Phobia, and Panic Disorder: A 12-Year Prospective Study. *American Journal of Psychiatry*, 162(6), 1179–1187. 10.1176/appi.ajp.162.6.1179
- Chatterjee M, Stratou G, Scherer S, & Morency L-P (2014). Context-based signal descriptors of heart-rate variability for anxiety assessment. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 3631–3635. 10.1109/ICASSP.2014.6854278
- Chen H, Huang Y, & Zhang N (2015). Joint modeling of a linear mixed effects model for self-esteem from mean ages 13 to 22 and a generalized linear model for anxiety disorder at mean age 33. *Journal of Medical Statistics and Informatics*, 3(1), 1 10.7243/2053-7662-3-1
- Chi M, Guo S, Ning Y, Li J, Qi H, Gao M, Wang J, Hu X, Guo Y, Yang Y, Peng H, & Wu K (2014). Using Support Vector Machine to Identify Imaging Biomarkers of Major Depressive Disorder and Anxious Depression In Pan L, P un, Pérez-Jiménez MJ, & Song T (Eds.), *Bio-Inspired Computing—Theories and Applications* (pp. 63–67). Springer 10.1007/978-3-662-45049-9\_10
- Christiana JM, Gilman SE, Guardino M, Mickelson K, Morselli PL, Olfson M, & Kessler RC (2000). Duration between onset and time of obtaining initial treatment among people with anxiety and mood disorders: An international survey of members of mental health patient advocate groups. *Psychological Medicine*, 30(3), 693–703. 10.1017/s0033291799002093 [PubMed: 10883723]
- de Graaf R, Bijl RV, Spijker J, Beekman ATF, & Vollebergh W. a. M. (2003). Temporal sequencing of lifetime mood disorders in relation to comorbid anxiety and substance use disorders—Findings from the Netherlands Mental Health Survey and Incidence Study. *Social Psychiatry and Psychiatric Epidemiology*, 38(1), 1–11. 10.1007/s00127-003-0597-4 [PubMed: 12563553]
- Recognition Effective and Treatment of Generalized Anxiety Disorder in Primary Care. (2004). *Primary Care Companion to The Journal of Clinical Psychiatry*, 6(1), 35–41.
- Essau CA, Conradt J, Sasagawa S, & Ollendick TH (2012). Prevention of anxiety symptoms in children: Results from a universal school-based trial. *Behavior Therapy*, 43(2), 450–464. 10.1016/j.beth.2011.08.003 [PubMed: 22440079]
- Feldner MT, Zvolensky MJ, Babson K, Leen-Feldner EW, & Schmidt NB (2008). An Integrated Approach to Panic Prevention Targeting the Empirically-Supported Risk Factors of Smoking and Anxiety Sensitivity: Theoretical Basis and Evidence from a Pilot Project Evaluating Feasibility and Short-term Efficacy. *Journal of Anxiety Disorders*, 22(7), 1227–1243. 10.1016/j.janxdis.2008.01.005 [PubMed: 18281190]
- Hek K, Tiemeier H, Newson RS, Luijendijk HJ, Hofman A, & Mulder CL (2011). Anxiety disorders and comorbid depression in community dwelling older adults. *International Journal of Methods in Psychiatric Research*, 20(3), 157–168. 10.1002/mpr.344 [PubMed: 22547298]
- Hoffman DL, Dukes EM, & Wittchen H-U (2008). Human and economic burden of generalized anxiety disorder. *Depression and Anxiety*, 25(1), 72–90. 10.1002/da.20257 [PubMed: 17146763]
- Jacobson NC, Chow S-M, & Newman MG (2019). The Differential Time-Varying Effect Model (DTVEM): A tool for diagnosing and modeling time lags in intensive longitudinal data. *Behavior Research Methods*, 51(1), 295–315. 10.3758/s13428-018-1101-0 [PubMed: 30120682]
- Jacobson NC, Lord KA, & Newman MG (2017). Perceived emotional social support in bereaved spouses mediates the relationship between anxiety and depression. *Journal of Affective Disorders*, 211, 83–91. 10.1016/j.jad.2017.01.011 [PubMed: 28103522]
- Jacobson NC, & Newman MG (2014). Avoidance mediates the relationship between anxiety and depression over a decade later. *Journal of Anxiety Disorders*, 28(5), 437–445. 10.1016/j.janxdis.2014.03.007 [PubMed: 24858656]
- Jacobson NC, & Newman MG (2016). Perceptions of Close and Group Relationships Mediate the Relationship Between Anxiety and Depression Over a Decade Later. *Depression and Anxiety*, 33(1), 66–74. 10.1002/da.22402 [PubMed: 26290461]

- Jacobson NC, & O’Cleirigh C (2019). Objective digital phenotypes of worry severity, pain severity and pain chronicity in persons living with HIV. *The British Journal of Psychiatry: The Journal of Mental Science*, 1–3. 10.1192/bjp.2019.168
- Jacobson NC, Summers B, & Wilhelm S (2020). Digital Biomarkers of Social Anxiety Symptom Severity: Digital Phenotyping using Passive Smartphone Sensors. *Journal of Medical Internet Research*. 10.2196/16875
- Jacobson NC, Weingarden H, & Wilhelm S (2019). Digital biomarkers of mood disorders and symptom change. *Npj Digital Medicine*, 2(1), 1–3. 10.1038/s41746-019-0078-0 [PubMed: 31304351]
- Kenardy J, McCafferty K, & Rosa V (2003). Internet-Delivered Indicated Prevention For Anxiety Disorders: A Randomized Controlled Trial. *Behavioural and Cognitive Psychotherapy*, 31(3), 279–289. 10.1017/S1352465803003047
- Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, & Wittchen H-U (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *International Journal of Methods in Psychiatric Research*, 21(3), 169–184. 10.1002/mpr.1359 [PubMed: 22865617]
- Kotsilieris T, Pintelas E, Livieris IE, & Pintelas P (2018). Reviewing machine learning techniques for predicting anxiety disorders. 22.
- Lamers F, van Oppen P, Comijs HC, Smit JH, Spinhoven P, van Balkom AJLM, Nolen WA, Zitman FG, Beekman ATF, & Penninx BWJH (2011). Comorbidity patterns of anxiety and depressive disorders in a large cohort study: The Netherlands Study of Depression and Anxiety (NESDA). *The Journal of Clinical Psychiatry*, 72(3), 341–348. 10.4088/JCP.10m06176blu [PubMed: 21294994]
- Lau EX, & Rapee RM (2011). Prevention of Anxiety Disorders. *Current Psychiatry Reports*, 13(4), 258–266. 10.1007/s11920-011-0199-x [PubMed: 21484451]
- Lenze EJ, & Wetherell JL (2011). A lifespan view of anxiety disorders. *Dialogues in Clinical Neuroscience*, 13(4), 381–399. [PubMed: 22275845]
- Luutonen S, Tikka M, Karlsson H, & Salokangas RKR (2013). Childhood trauma and distress experiences associate with psychotic symptoms in patients attending primary and psychiatric outpatient care. Results of the RADEP study. *European Psychiatry*, 28(3), 154–160. 10.1016/j.eurpsy.2011.11.005 [PubMed: 22520956]
- Månsson KNT, Frick A, Boraxbekk C-J, Marquand AF, Williams SCR, Carlbring P, Andersson G, & Furmark T (2015). Predicting long-term outcome of Internet-delivered cognitive behavior therapy for social anxiety disorder using fMRI and support vector machine learning. *Translational Psychiatry*, 5(3), e530–e530. 10.1038/tp.2015.22 [PubMed: 25781229]
- Markowitz JS, Weissman MM, Ouellette R, Lish JD, & Klerman GL (1989). Quality of Life in Panic Disorder. *Archives of General Psychiatry*, 46(11), 984–992. 10.1001/archpsyc.1989.01810110026004 [PubMed: 2684086]
- Mellman TA (2006). Sleep and Anxiety Disorders. *Psychiatric Clinics*, 29(4), 1047–1058. 10.1016/j.psc.2006.08.005 [PubMed: 17118281]
- Mogg K, Millar N, & Bradley BP (2000). Biases in eye movements to threatening facial expressions in generalized anxiety disorder and depressive disorder. *Journal of Abnormal Psychology*, 109(4), 695–704. 10.1037/0021-843X.109.4.695 [PubMed: 11195993]
- Murphy BM, Elliott PC, Worcester MUC, Higgins RO, Le Grande MR, Roberts SB, & Goble AJ (2008). Trajectories and predictors of anxiety and depression in women during the 12 months following an acute cardiac event. *British Journal of Health Psychology*, 13(1), 135–153. 10.1348/135910707X173312 [PubMed: 17535492]
- Nemesure MD, Heinz M, Huang R, & Jacobson NC (2020). Predictive Modeling of Psychiatric Illness using Electronic Health Records and a Novel Machine Learning Approach with Artificial Intelligence [Preprint]. *PsyArXiv*. 10.31234/osf.io/fhdr5
- Orlando M, Burnam MA, Sherbourne CD, Morton SC, London AS, Hays RD, & Bing EG (2001). Brief screening of psychiatric disorders among a national sample of HIV-positive adults: Concordance between the Composite International Diagnostic Interview (CIDI) and the CIDI

- Short Form (CIDI-SF). *International Journal of Methods in Psychiatric Research*, 10(2), 97–107. 10.1002/mpr.104
- Penninx BWJH, Beekman ATF, Smit JH, Zitman FG, Nolen WA, Spinhoven P, Cuijpers P, Jong PJD, Marwijk HWJV, Assendelft WJJ, Meer KVD, Verhaak P, Wensing M, Graaf RD, Hoogendijk WJ, Ormel J, & Dyck RV (2008). The Netherlands Study of Depression and Anxiety (NESDA): Rationale, objectives and methods. *International Journal of Methods in Psychiatric Research*, 17(3), 121–140. 10.1002/mpr.256 [PubMed: 18763692]
- Penninx BWJH, Nolen WA, Lamers F, Zitman FG, Smit JH, Spinhoven P, Cuijpers P, de Jong PJ, van Marwijk HWJ, der Meer K. van, Verhaak P, Laurant MGH, de Graaf R, Hoogendijk WJ, der Wee N. van, Ormel J, van Dyck R, & Beekman ATF (2011). Two-year course of depressive and anxiety disorders: Results from the Netherlands Study of Depression and Anxiety (NESDA). *Journal of Affective Disorders*, 133(1), 76–85. 10.1016/j.jad.2011.03.027 [PubMed: 21496929]
- Perkins NJ, & Schisterman EF (2006). The Inconsistency of “Optimal” Cut-points Using Two ROC Based Criteria. *American Journal of Epidemiology*, 163(7), 670–675. 10.1093/aje/kwj063 [PubMed: 16410346]
- Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, & Regier DA (1988). The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry*, 45(12), 1069–1077. 10.1001/archpsyc.1988.01800360017003 [PubMed: 2848472]
- Rosenström T, Elovainio M, Jokela M, Pirkola S, Koskinen S, Lindfors O, & Keltikangas- Järvinen L (2015). Concordance between Composite International Diagnostic Interview and self-reports of depressive symptoms: A re-analysis. *International Journal of Methods in Psychiatric Research*, 24(3), 213–225. 10.1002/mpr.1478 [PubMed: 26140369]
- Roy-Byrne PP, Craske MG, & Stein MB (2006). Panic disorder. *The Lancet*, 368(9540), 1023–1032. 10.1016/S0140-6736(06)69418-X
- Shirneshan E, Bailey J, Relyea G, Franklin BE, Solomon DK, & Brown LM (2013). Incremental direct medical expenditures associated with anxiety disorders for the U.S. adult population: Evidence from the Medical Expenditure Panel Survey. *Journal of Anxiety Disorders*, 27(7), 720–727. 10.1016/j.janxdis.2013.09.009 [PubMed: 24135257]
- Spinhoven P, Batelaan N, Rhebergen D, van Balkom A, Schoevers R, & Penninx BW (2016). Prediction of 6-yr symptom course trajectories of anxiety disorders by diagnostic, clinical and psychological variables. *Journal of Anxiety Disorders*, 44, 92–101. 10.1016/j.janxdis.2016.10.011 [PubMed: 27842240]
- Toghiani S, DiBonaventura M, Järbrink K, & Locklear JC (2014, 4 2). Economic and humanistic burden of illness in generalized anxiety disorder: An analysis of patient survey data in Europe. *ClinicoEconomics and Outcomes Research*; Dove Press. 10.2147/CEOR.S55429
- Vermani M, Marcus M, & Katzman MA (2011). Rates of Detection of Mood and Anxiety Disorders in Primary Care: A Descriptive, Cross-Sectional Study. *The Primary Care Companion to CNS Disorders*, 13(2). 10.4088/PCC.10m01013
- Wittchen H-U (2002). Generalized anxiety disorder: Prevalence, burden, and cost to society. *Depression and Anxiety*, 16(4), 162–171. 10.1002/da.10065 [PubMed: 12497648]
- World Health Organization. (1990). *Composite International Diagnostic Interview, CIDI, Version 10*. Geneva: World Health Organization.
- Zainal NH, & Newman MG (2019). Relation between cognitive and behavioral strategies and future change in common mental health problems across 18 years. *Journal of Abnormal Psychology*, 128(4), 295–304. 10.1037/abn0000428 [PubMed: 31045412]

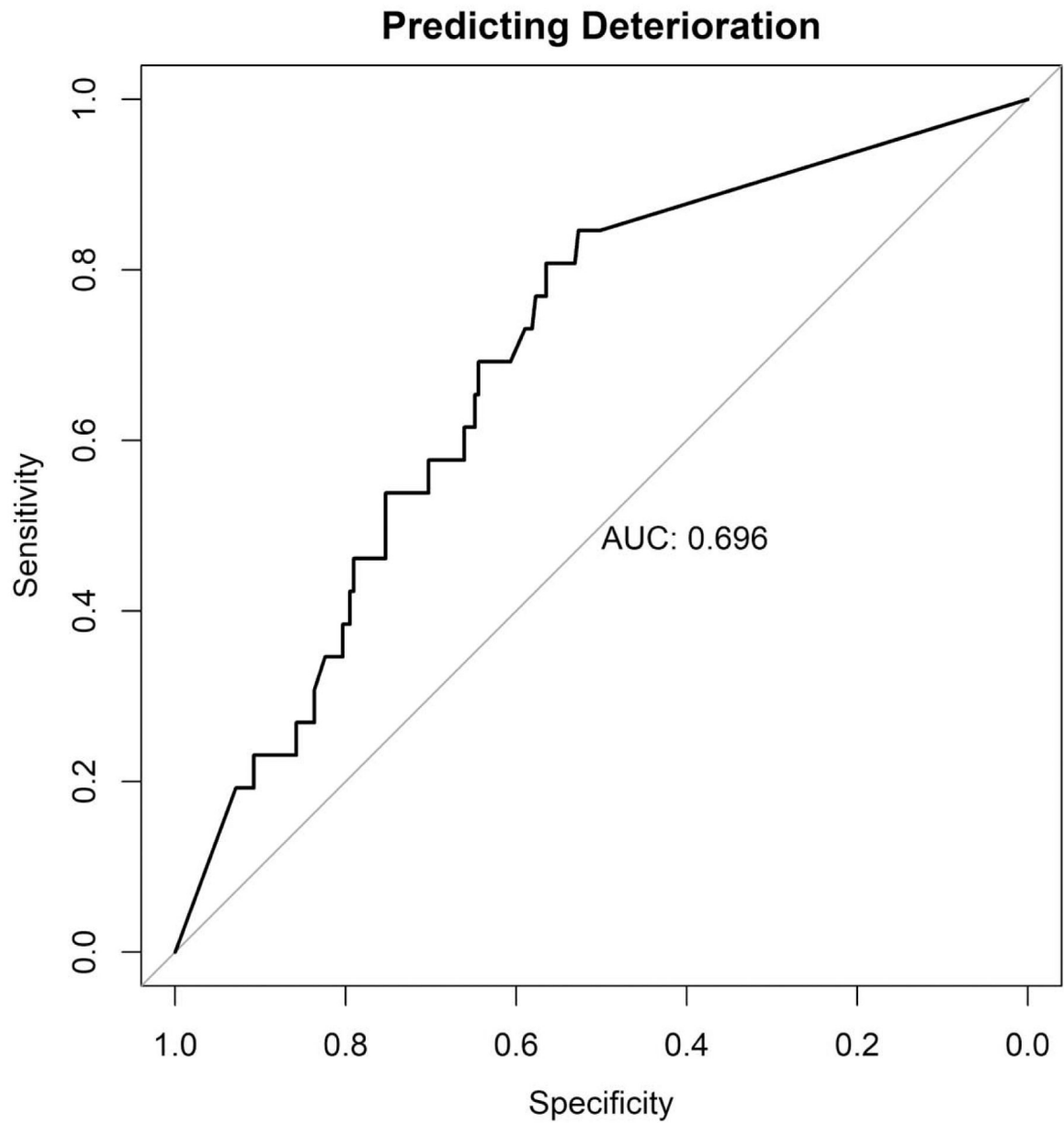
### Highlights

- Passive, wearable sensor data collected longitudinally has potential to predict deterioration in long term anxiety disorder symptoms
- The developed ensemble deep learning model has the capability to predict anxiety symptom deterioration with an above chance accuracy across almost two decades
- This prediction pipeline, in conjunction with passive movement data, may help to narrow the longstanding wait between symptom deterioration and treatment initiation



**Figure 1.** Data Analysis Pipeline. Nineteen raw actigraphy variables for  $N=265$  individuals across 7 distinct time points were processed using basic descriptive statistics, differential time varying effect modeling (DTVEM) and spectral analysis approaches to derive 800 features. This new feature space was processed through a deep autoencoder to compress the information space into a summative subset of 50 features for ensemble machine learning regression modeling. Several lower level models were applied to predict deterioration within a cross-validated framework. The resulting predictions of these lower level models were used as features for the ensemble (Extreme Gradient Boosting) model. The predictions of this ensemble model were analyzed for accuracy, sensitivity, and specificity in the construction of an AUC curve for assessment.





**Figure 2.** Plot of machine learning predictive performance with AUC as the primary outcome metric. The curve demonstrates the trade-off in true-positive (sensitivity) and true-negative (specificity) predictive accuracy. Here, an area under the ROC of 0.696 corresponds with an ability to correctly predict deterioration in 84.6% who deteriorate and an ability to correctly predict non-deterioration in 52.7% of individuals who do not deteriorate.

**Table 1**

## Participant Demographics

Variables		Frequency	Percentage (%)
Gender	Male	111	41.9
	Female	154	58.1
Ethnicity	White	249	94.7
	Black	4	1.5
	Native American/Alaskan, Aleutian Islander	1	.4
	Pacific	1	.4
	Multiracial	1	.4
	Other/no	2	.8
Income	\$0–9,999	172	64.9
	10,000–24,999	26	9.8
	25,000–49,999	19	7.2
	50,000–74,999	20	7.5
	75,000–99,999	8	3
	100,000+	9	3.4
	N/A	11	4.2

Note. Participant(n=265) sociodemographic characteristics