Journal of Affective Disorders 329 (2023) 293–299

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Leveraging deep learning models to understand the daily experience of anxiety in teenagers over the course of a year

Brian Wang [a], Matthew D. Nemesure [a,c,*], Chloe Park [a], George D. Price [a,c], Michael V. Heinz [a], Nicholas C. Jacobson [a,b,c]

[a] Center for Technology and Behavioral Health, Dartmouth College, Hanover, NH, United States of America
[b] Department of Biomedical Data Science, Geisel School of Medicine, Lebanon, NH, United States of America
[c] Quantitative Biomedical Sciences, Geisel School of Medicine, Lebanon, NH, United States of America

ARTICLE INFO

Keywords:
Anxiety
Machine learning
Long-Short-Term-Memory
Latent features
Generalized Anxiety Disorder-7

ABSTRACT

Introduction: Anxiety disorders are a prevalent and severe problem that are often developed early in life and can disrupt the daily lives of affected individuals for many years into adulthood. Given the persistent negative aspects of anxiety, accurate and early assessment is critical for long term outcomes. Currently, the most common method for anxiety assessment is through point-in-time measures like the GAD-7. Unfortunately, this survey and others like it can be subject to recall bias and do not fully capture the variability in an individual's day-to-day symptom experience. The current work aims to evaluate how point-in-time assessments like the GAD-7 relate to daily measurements of anxiety in a teenage population.

Methods: To evaluate this relationship, we leveraged data collected at four separate three week intervals from 30 teenagers (age 15–17) over the course of a year. The specific items of interest were a single item anxiety severity measure collected three times per day and end-of-month GAD-7 assessments. Within this sample, 40 % of individuals reported clinical levels of generalized anxiety disorder symptoms at some point during the study. The first component of analysis was a visual inspection assessing how daily anxiety severity fluctuated around end-of-month reporting via the GAD-7. The second component was a between-subjects comparison assessing whether individuals with similar GAD-7 scores experienced similar symptom dynamics across the month as represented by latent features derived from a deep learning model. With this approach, similarity was operationalized by hierarchical clustering of the latent features.

Results: The aim clearly indicated that an individual's daily experience of anxiety varied widely around what was captured by the GAD-7. Additionally, when hierarchical clustering was applied to the three latent features derived from the (LSTM) encoder ($r = 0.624$ for feature reconstruction), it was clear that individuals with similar GAD-7 outcomes were experiencing different symptom dynamics. Upon further inspection of the latent features, the LSTM model appeared to rely as much on *anxiety variability* over the course of the month as it did on anxiety severity ($p < 0.05$ for both mean and RMSSD) to represent an individual's experience.

Discussion: This work serves as further evidence for the heterogeneity within the experience of anxiety and that more than just point-in-time assessments are necessary to fully capture an individual's experience.

1. Introduction

A review of epidemiological surveys for all anxiety disorders indicated that 13 % of individuals will experience an anxiety disorder in any given year, and 21 % of individuals will experience an anxiety disorder at some point in their lives (Bandelow & Michaelis, 2015). Unfortunately, for many of these individuals their anxiety will begin early in life (Kessler, 1994). A meta-analysis of 24 studies indicated that the average age of onset for all anxiety disorders fell during the later teenage years with specific anxiety disorders having an age of onset earlier than 15 years old (de Lijster et al., 2017). At such a young age, these disorders have a significant effect on mood, irritability, academic performance, and quality of life (Khesht-Masjedi et al., 2019). Additionally, given that half of all *lifetime* cases of anxiety disorders begin during the teenage

* Corresponding author at: Center for Technology and Behavioral Health, Dartmouth College, 46 Centerra Pkwy, Lebanon, NH 03766, United States of America.
E-mail address: matthew.d.nemesure.gr@dartmouth.edu (M.D. Nemesure).

https://doi.org/10.1016/j.jad.2023.02.084
Received 26 July 2022; Received in revised form 10 February 2023; Accepted 19 February 2023
Available online 27 February 2023

years and that early intervention has been shown to vastly improve outcomes, it is imperative that we better understand the individual experience of anxiety in young teens (Kessler et al., 2005; Mifsud & Rapee, 2005).

Typically, when people are evaluated for anxiety disorders, it is done via a point-in-time assessment. The most common of these assessments is the GAD-7 questionnaire, a 7-item self-report form used to assess the severity to which an individual experiences generalized anxiety disorder over a prior time period before the survey is taken (Spitzer et al., 2006). The results of the GAD-7 questionnaire play a substantial role in evaluating an individual's experience with anxiety and defining potential treatment options (Spitzer et al., 2006). Unfortunately, these point-in-time assessments may not comprehensively capture an individual's daily experience with anxiety and how it varies over time (Frank et al., 2017). This lack of representativeness can be attributed in large part to the fact that these self-reported point-in-time assessments are subject to cognitive biases (Sato & Ichiro, 2011). Specifically, recall bias may cause individuals to heavily weight more recent experiences, misrepresent past experiences as they become distorted with time, and even to forget critical parts of their recent experience with anxiety (Colombo et al., 2020; Hassan, 2005; Sato & Ichiro, 2011).

Considering these limitations with point-in-time assessments, there is a necessity for evaluation to help understand how well these surveys can reflect day-to-day experiences with anxiety. There has been some prior work in this space assessing the agreement between self-report measures of trait anxiety and ecological momentary assessment (EMA) questionnaire responses, however many studies that capture both don't directly make this comparison (Edmondson et al., 2013; Solhan et al., 2009). The results of these analyses have shown low or low-moderate relationships between the two measures (Edmondson et al., 2013; Solhan et al., 2009). However, to our knowledge, there have been no studies that try to assess between subject similarities across both point-in-time assessment and EMA response. With this in mind, there are two major goals of this work. The first is to investigate, in this sample, how well an end-of-month survey (GAD-7) can capture the daily variation in anxiety severity over the prior month. The second goal is to evaluate the degree to which between-person similarities in GAD-7 reporting associates with reporting of daily anxiety via EMA. To accomplish this second goal, we will employ a deep learning LSTM model to capture data driven components of the daily anxiety experience and map them to a representative latent space. By capturing the data in this way, we can retain the core components that define the daily experience of anxiety while simultaneously allowing for the ease of comparison between individuals. These representations of daily anxiety can then be used to quantitatively assess the second goal: whether individuals with similar levels of anxiety via point-in-time assessment actually experience their daily anxiety similarly. To operationalize this similarity assessment, hierarchical clustering will be employed to group individuals in such a way that those with analogous experiences would be closer together. All analysis will be done based on publicly available data representing daily anxiety measurements and monthly GAD-7 scores of 30 teenagers from a previously conducted larger study investigating stressful life events and within-person development of anxiety and depression (Rodman et al., 2021).

2. Methods

2.1. Participants and data collection

The original investigators recruited female participants ($N = 30$) aged 15–17 from schools, libraries, and other public spaces in Seattle, Washington. The study inclusion criteria ensured that participants were female, age 15–17, fluent in English, and had access to an individual smartphone (Rodman et al., 2020). Twelve monthly in-person assessments were conducted on each participant to assess symptom severity for anxiety and depression over the past month, measured using the GAD-7 and PHQ-9 scales, respectively (Kroenke et al., 2001; Spitzer et al., 2006). Within the same year, participants underwent four separate occasions of three weeks worth of monitoring by daily ecological momentary assessments where they would report their feelings of anxious and depressive affect (each on a 1 to 7 scale) three times a day through the MetricWire app (Metricwire Inc, n.d.). All study procedures were reviewed and approved by the Institutional Review Board at the University of Washington. Written consent to participate in the daily and monthly self-assessments was obtained from both the adolescent participants and their legal guardians (Rodman et al., 2020).

Among the 120 measurements for monthly GAD-7 assessments across the 30 participants, 38 of those measurements would be considered representative of moderate to severe anxiety using a standard cutoff for moderate anxiety of 10 (Spitzer et al., 2006). Pairing this down to assess at the individual level rather than the monthly level, 12 participants had at least one monthly GAD-7 measurement that indicated experiences of moderate to severe anxiety.

2.2. Within-person comparison

The first aim was to understand how daily anxiety affect measurements related to monthly anxiety symptom measurements. Specifically, the goal was to assess how well the GAD-7 captured the magnitude of the daily assessments as well as how widely the daily assessments varied around the end of month reported value. To accomplish this goal, we visualized person-level plots of the normalized monthly GAD-7 scores and daily anxiety scores. The monthly GAD-7 scores were assumed to represent the anxiety level two weeks prior to measurement and the daily anxiety values were layered onto the plot. In this way, a qualitative evaluation could be made about how well the within-person normalized daily ratings compared to the end of month assessment.

2.3. Preprocessing for deep learning

Given that the EMA assessment did not cover the entire year and GAD-7 assessment was given at the end of each month, participant's anxiety information was treated separately on a month-by-month basis. This data relating to a specific person and month of the year will be henceforth referred to as a person-month (e.g. the analysis treats data from month one of the study for person A as separate from data from month 4 of the study for person A). For each person-month, the monthly GAD-7 score and the 28 days of daily anxiety scores leading up to the day of their monthly in-person assessment was extracted. Thus, for each person-month, 84 points of daily anxiety measurements are considered and are broken down into three features based on collection time being in the morning, afternoon or evening over the 28 days. By processing the raw data through this method, 120 person-months worth of anxiety data are constructed. Daily anxiety measurements and monthly GAD-7 scores were normalized via min/max normalization based on the measurement scale to be constrained between zero and one (Pérez-García et al., 2021).

To deal with the sporadic sampling of months as well as some lack of compliance, person-months with over 70 % of daily anxiety measurements missing were excluded from analysis, resulting in 66 of the original 120 person-months remaining eligible for analysis. Importantly, every individual participant in the study was represented by at least one person-month, with a maximum of three person-months for a given participant.

2.4. Deep learning based feature reduction for EMA

The goal of the between-subjects analysis was to examine whether individuals who scored similarly on the GAD-7 had an analogous experience with their daily anxiety. Given the dimensionality of time series data, making these comparisons required dimensional reduction that maintained the variation within the raw assessment data. To accomplish this, we leveraged a Long-Short-Term-Memory (LSTM) deep

learning model within an encoder-decoder framework. LSTM nodes within a deep learning framework are uniquely poised to handle time-series data. This is due to their ability to model time-dependency, which essentially allows for the model to handle a given time point without losing the context of the prior time points. This was particularly suitable for our framework as the reduced feature space could still have some latent representation of the time-varying components in the original data. This helped to ensure that the encoded (reduced) feature space maintained enough variance to then be decoded to the original feature space of daily anxiety assessment. By approaching the data in this way, we could directly test whether or not the latent variables in the reduced feature space were capturing the variability in the original feature space. If this was the case, it served as proof that the reduced feature space is a comprehensive representation of a person's day-to-day anxiety experience over the course of a month.

2.4.1. Model development

All model development was done in Python 3.8 using Tensorflow (Drake et al., 2010; TensorFlow Developers, 2022). Inputs to the LSTM model are 66 person-months of anxiety data separated into three features (morning, afternoon and evening anxiety) and 28 time steps. The encoder section of the model is responsible for reducing this feature space into three features per person. The decoder portion of the model then reconstructs the daily anxiety data back from these three features. The model specifically consists of three layers for the encoder and three layers for the decoder. These layers include 64, 32 and 3 LSTM nodes, respectively, and this structure is mirrored in the decoder framework. Each of these nodes use the Exponential Linear Units (ELU) activation function due to its ability to incorporate negative values in addition to faster run times (Clevert et al., 2015; Jacobson et al., 2021). This allows the model to train faster while also being able to penalize varying degrees of wrongness for inaccurate decoded sequences. Prior to these layers, we implemented a masking layer to mask all missing values throughout the person-month. After the encoder-decoder layers, the final layer was a TimeDistributed layer wrapped on a Dense Layer for the output of the model. The dense layer has three units, which allows for the prediction of three anxiety sequence features (morning, afternoon, evening) at a time (i.e. one day at a time). The model used the ADAM optimization algorithm for training and uses the mean absolute error as the loss function (Keras, n.d.).

2.4.2. Model tuning

To ensure the model was not being overfit and an appropriate epoch was selected, the model was trained on 46 person-months with 20 person-months acting as the validation set. Layer sizes were optimized for the encoder model (with the decoder model always having the same layer sizes as the encoder model) to minimize the output feature space without sacrificing captured variability. To measure the accuracy of each model's decoding of the feature layer back to the original sequence, each time point was represented as having an x-value of its respective anxiety value in the original sequence and a y-value of its respective anxiety value in the predicted sequence.

2.4.3. Model evaluation

To evaluate the model, Pearson correlations were calculated for each individual between the true EMA anxiety response and the predicted EMA response produced by the decoder portion of the model. The main idea was that if the decoder was able to accurately reproduce the original input space, we could be confident that the encoded latent features captured the majority of the variability of the EMA items over time. These correlations were calculated both including and excluding missing data in the input space to ensure that the model was picking up on true changes in daily severity and not just missing vs. non-missing data.

2.5. Evaluating the encoded features

With the encoded features that represented the daily anxiety variation in the original data set, hierarchical clustering was implemented to organize person-months in such a way that those with a similar overall daily experience of anxiety were close together. Following this clustering, we evaluated whether those person-months that were clustering together had similar reporting via the end-of-month GAD-7. In this way, we could directly investigate if individuals with similar daily experiences (as represented by the encoded feature-space of their raw anxiety reporting) were also reporting analogously via the point-in-time assessment.

To further explore what aspects of daily anxiety the LSTM encoder was leveraging to create the three representative features, three hierarchical regressions were implemented to evaluate how summary statistics of daily anxiety data predicted each of the three embedded features. We hypothesized that the most important aspects of daily anxiety that the deep learning model was picking up on were related to variability, severity, and missingness. To capture variability, we calculated the variance, crossing points and root mean square of successive differences for the 84 anxiety values measured throughout the person-month (Hovenkamp-Hermelink et al., 2019; Shaffer & Ginsberg, 2017). To capture severity of anxiety, we calculated the mean reported anxiety across the 84 values measured throughout the month (Cox et al., 2018). Related to missing surveys, as these were left as missing and masked in the model, we wanted to see if the percent missingness was contributing to any of the latent features and thus was included as a predictor. Finally, related to the main outcome of interest, we wanted to see if any of the daily anxiety based latent features could be mapped to the reported monthly GAD7 and thus this was included in the model. The general version of the mixed model is shown below:

$$latent\ feature = \beta_0 + percent.missing^*\beta_1 + anxiety.variance^*\beta_2 + anxiety.mean^*\beta_3 + crossing.points^*\beta_4 + rmssd^*\beta_5 + GAD.monthly^*\beta_6 + \theta_i + \epsilon_i$$
$$\theta \sim N(0,1)$$

3. Results

3.1. Data visualization/visual inspection

Qualitative assessment mapping daily anxiety to monthly reporting was done by evaluating a visualization of both normalized values together. The plots (Fig. 1) show that average normalized anxiety severity over the course of the month tends to be similar to the normalized monthly assessment, although this is not always the case. Additionally, the figures show that more often than not the daily anxiety values vary widely around the mean daily value as well as the monthly severity assessment. These differences in variability lead to cases where individuals have very similar monthly severity assessment scores but vastly different experiences over the course of that month.

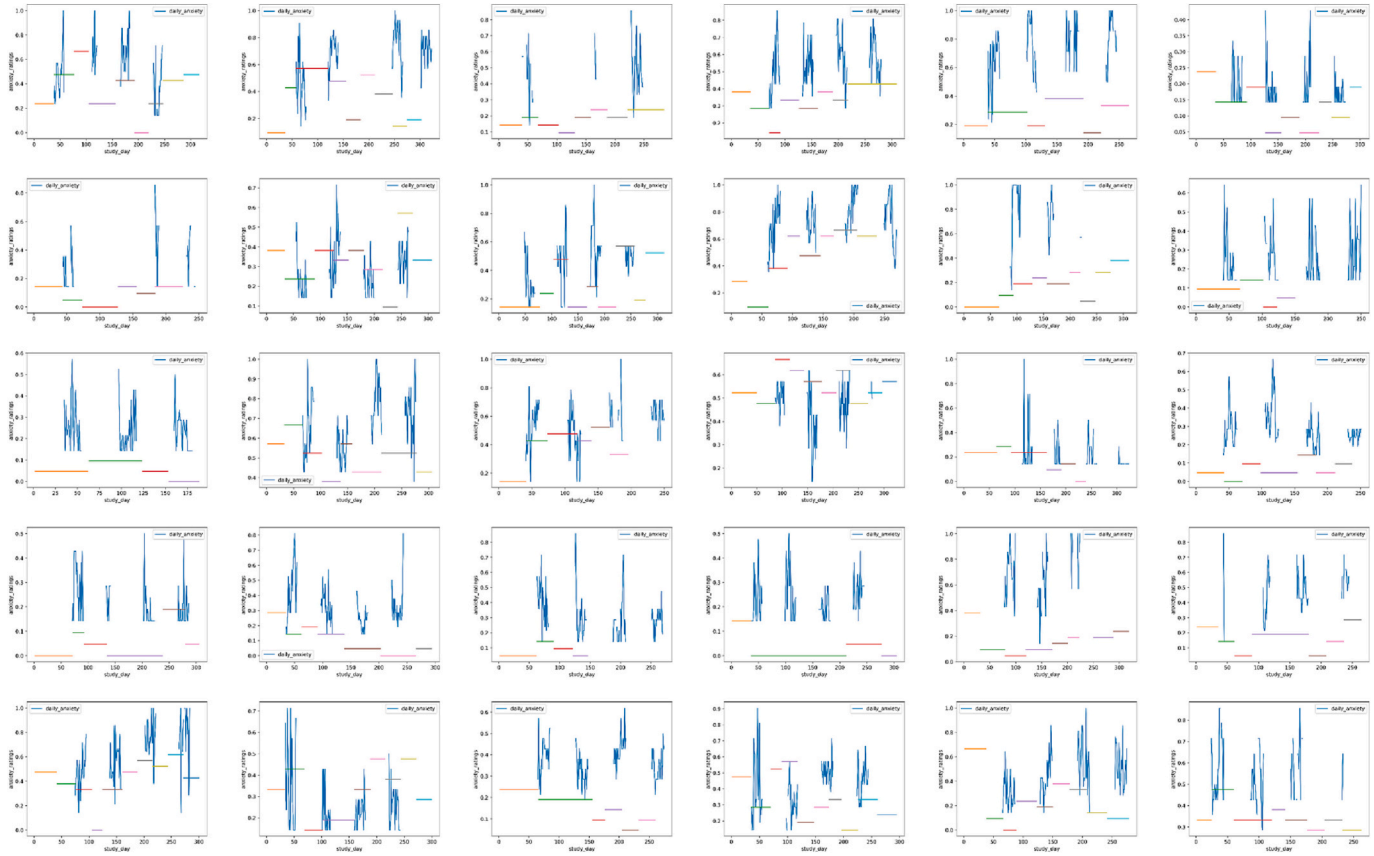

Fig. 1. Idiographic Plots representing individual's normalized daily anxiety EMA reports and end-of-month GAD7.

3.2. LSTM encoder-decoder

The training/validation loss was plotted for 1500 training epochs and the point at which the validation loss stopped decreasing (epoch 450) was taken as the final model (Fig. 2). With the final model, the accuracy of the decoder model was calculated by assessing the correlation between true anxiety values and predicted anxiety values over the course of the month for each individual. These correlations were then averaged across individuals to get an overall sense of how well the encoded feature space represented the original measurements of anxiety over the month. Correlation coefficients were computed when missing data was both included and excluded. With missing values excluded, the model was able to reproduce the original feature space with an average correlation of 0.624 in the training set and 0.602 in the testing set. With missing data included the average correlation in the training set was 0.555 and in the testing set was 0.480. The distribution of the correlation coefficients across persons can be seen in Fig. 3.

3.3. Mapping latent feature space to original data

Using the three latent features (representing daily GAD) derived from the encoder portion of the LSTM model, hierarchical clustering revealed that individuals with similar monthly GAD-7 scores did not cluster together (Fig. 4). This lack of clustering extended to other individual summary statistics including missingness, anxiety variability and mean daily anxiety. This indicated that the clustering may be relying on some combination of these features or some other characteristics of the daily anxiety measurements.

3.4. Understanding latent feature representations

To evaluate what each of the embedded features was representing and thus determine what items were important in representing daily anxiety, the mixed modeling approach was implemented. The models indicated that measures of day to day instability in anxiety (root mean squared successive differences) and mean daily anxiety severity significantly ($p < 0.05$) associated with latent variable one and latent variable three. Ultimately, this is evidence that individuals with similar *overall* levels of anxiety (as measured by the GAD-7) may experience their anxiety in different ways. This is due to variability in their daily anxiety as well as mixed severity reporting between daily and monthly surveys.

4. Discussion

This project aimed to assess how and the degree to which monthly GAD-7 measurements are related to an individual's day-to-day experience with anxiety. Furthermore, novel latent representations of were explored to try to capture the daily patterns in each person's anxiety.

When considering how monthly GAD-7 measurements relate to day-to-day experiences with anxiety, it is evident that participants with similar GAD-7 scores for a certain month may have experienced anxiety in a substantially different way over the course of the month. To exemplify this, we can dive into an instance where two individuals (1001, 1029) have roughly the same monthly GAD-7 score. Yet, for one participant, they experienced high anxiety at the beginning of the month, low anxiety towards the middle of the month, and high anxiety again at the end of the month. In contrast, for the other participant, they experienced low anxiety at the beginning of the month, high anxiety towards the middle of the month, and low anxiety again towards the end of the month. Conversely, participants with similar experiences of anxiety over the course of a month may have substantially different GAD-7 scores. For example, take an instance where two participants (1013, 1027) both seem to have relatively low daily anxiety ratings towards the beginning of the month with spikes in anxiety towards the end of the

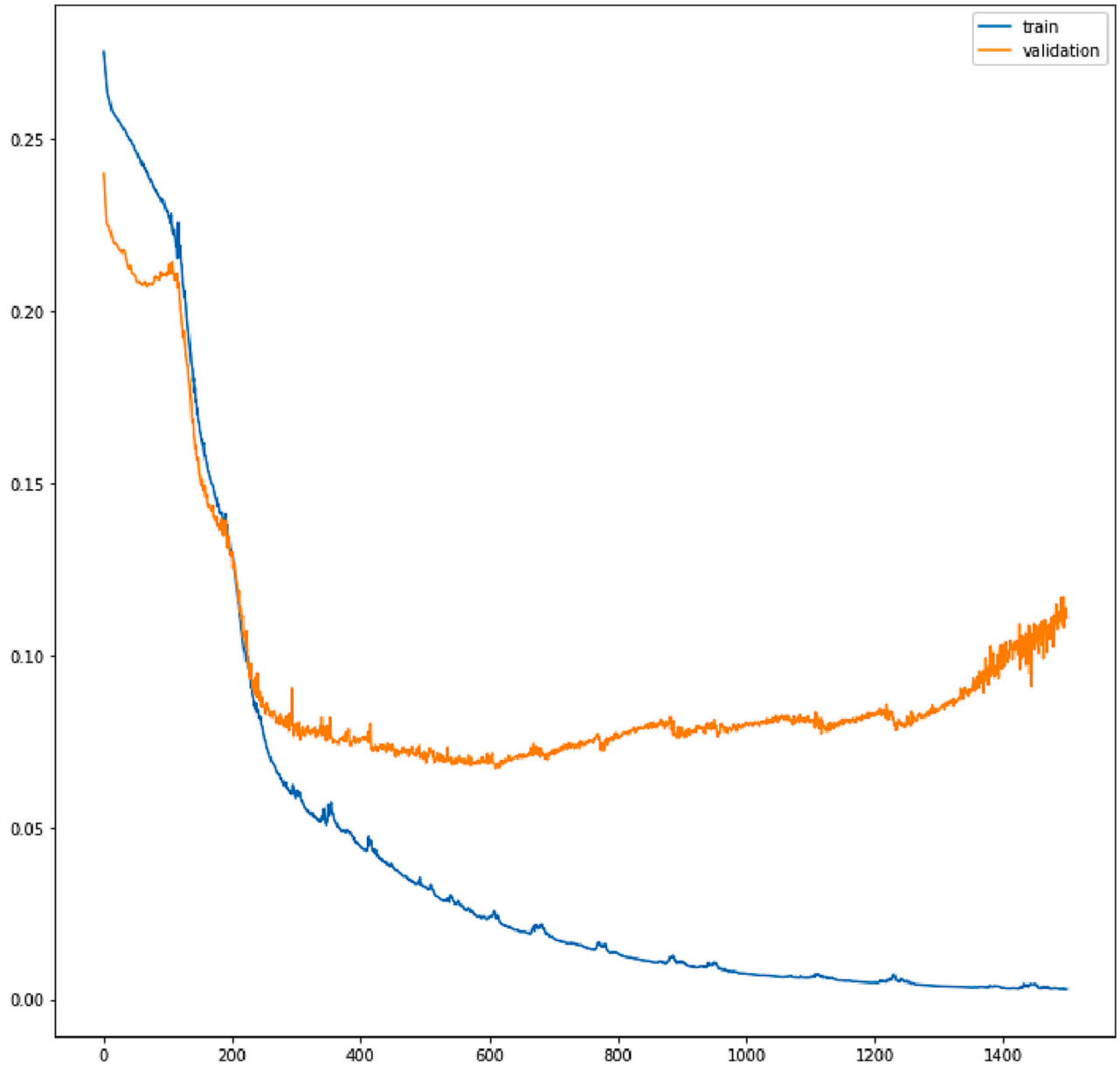

Fig. 2. Training and validation loss over 1500 epochs of training for the LSTM encoder-decoder framework.

month. However, after normalization, the monthly GAD-7 measurement for one of those subjects is almost three times the measurement given for the other subject. It appears as if GAD-7 scores are often unable to comprehensively capture a person's day-to-day experience with anxiety over the course of a month.

Previous works have shown the potential limitations of point-in-time assessments in capturing a person's day-to-day mental health experience. Studies that have directly assessed the comparison between these items have reported low levels of association (Edmondson et al., 2013; Solhan et al., 2009). Given this, there has been a large shift in the field towards using EMA as a measure of anxiety rather than standard surveys as noted in this review (Walz et al., 2014). This work serves as further evidence that point-in-time assessments are not always indicative of a person's daily anxiety experience. People with similar GAD-7 scores may experience anxiety in different ways on a day-to-day basis while people who experience anxiety in similar ways on a day-to-day basis may have different GAD-7 scores. Furthermore, when reducing anxiety sequences to a latent feature space of that represents the dynamic changes in severity over the course of a month, it is evident that people who report similarly on point-in-time surveys may have very different experiences over the course of the month.

There are, however, some limitations with respect to this work that need to be noted. The particular sample used for this study was limited in both scope and size given that it consisted of only thirty 15–17 year old females. Further limiting the sample size, only 66 of the 120 potential person-months were able to be included based on missingness cutoffs. Even with the cutoff, of the included person-months there was an average missingness of 48.74 % of daily anxiety measurements. Given these limitations, it is difficult to generalize the results of this project to people of different ages and genders. Related to modeling, the average correlation for reproducing the original feature space was around 0.6. There is always the opportunity to continue developing the deep learning architecture to capture more variability within the latent features, however, we chose the current model as a stopping point to avoid overdeveloping the model to a point of overfitting. Additionally, our mixed model approach allowed us to explain some components of the latent features, however, this did not completely explain what aspects of the data the model relied on to generate the features.

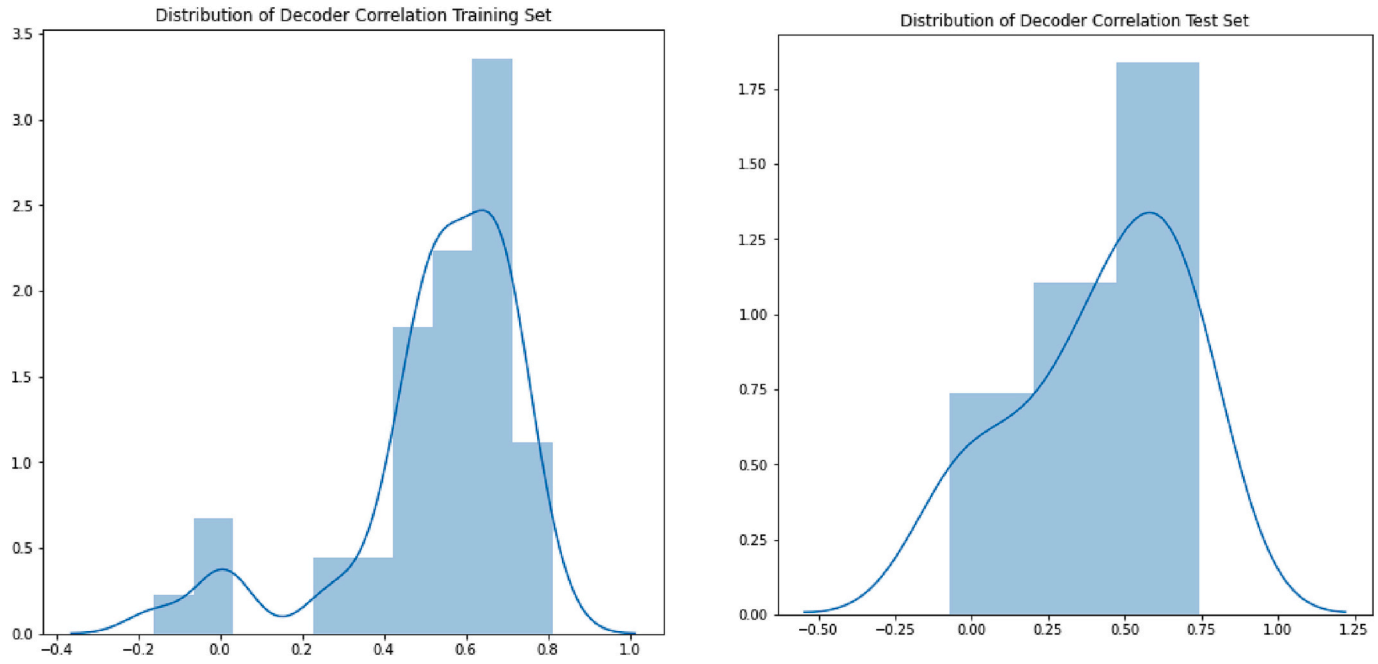

Fig. 3. Distribution of training and testing correlations between predicted within-person daily anxiety reporting via the encoder-decoder model and the actual reported value.

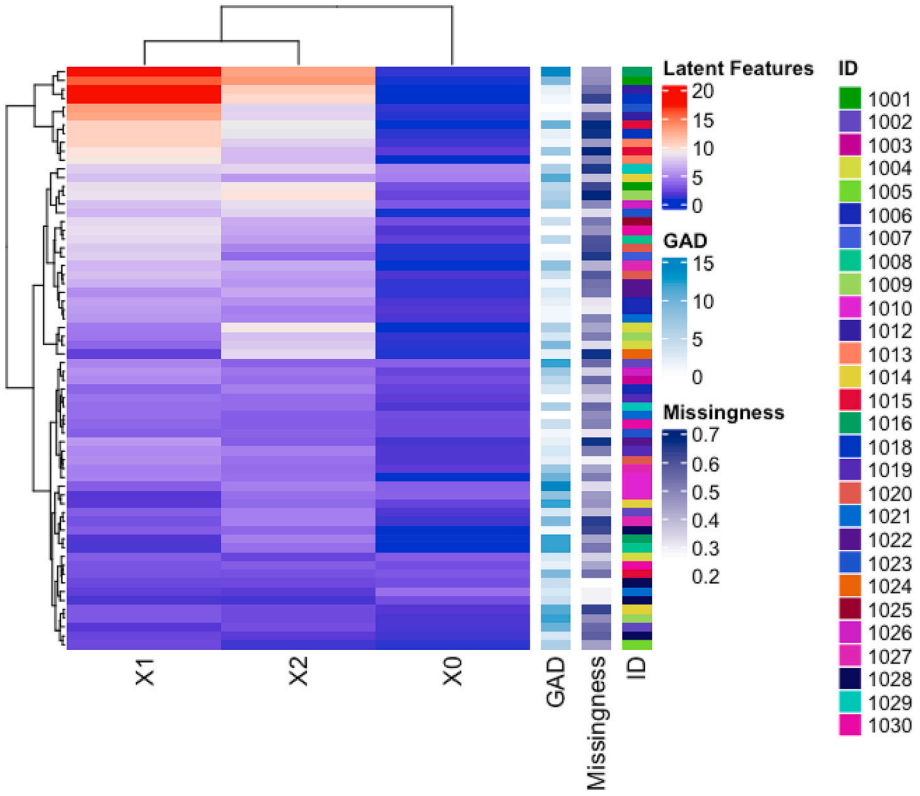

Fig. 4. Heatmap of latent features with GAD7, missingness and ID overlays.

Overall, this project can be used to demonstrate the potential limitations for the ability of point-in-time assessments like GAD-7 scores to capture daily variability in anxiety. Through the discontinuous anxiety graphs that plot each person's daily anxiety against their monthly anxiety measurements, it can be shown that GAD-7 scores often do not signify similarities or differences between how people experience anxiety. Although the GAD-7 did not reflect daily anxiety symptom experiences, the current results demonstrate that it is possible to predict a large proportion in the variation in daily anxiety symptoms using data driven approaches to longitudinal EMA data. By having the values for the three latent features derived from encoding a person's daily anxiety affect in a month, the daily anxiety experience can be somewhat

accurately reconstructed.

CRediT authorship contribution statement

All listed authors have contributed significantly to the manuscript, have agreed to this authorship order, and consent to their names on the manuscript.

Conflict of interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. This research is supported in part by T32 (T32DA037202-07) and P30 (P30DA029926) grants provided by the National Institute on Drug Abuse.

Acknowledgements

We are grateful to Katie A. McLaughlin, PhD, Alexandra M. Rodman, PhD, and the Stress and Development Lab at Harvard University for their efforts in collecting, processing, and publicly posting the original dataset, making this work possible.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2023.02.084.

References

Bandelow, B., Michaelis, S., 2015. Epidemiology of anxiety disorders in the 21st century. Dialogues Clin. Neurosci. 17 (3), 327–335.

Clevert, D.A., Unterthiner, T., Hochreiter, S., 2015. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). https://doi.org/10.48550/ARXIV.1511.07289. Published online.

Colombo, D., Suso-Ribera, C., Fernández-Álvarez, J., et al., 2020. Affect recall bias: being resilient by distorting reality. Cogn. Ther. Res. 44 (5), 906–918. https://doi.org/10.1007/s10608-020-10122-3.

Cox, R.C., Sterba, S.K., Cole, D.A., Upender, R.P., Olatunji, B.O., 2018. Time of day effects on the relationship between daily sleep and anxiety: an ecological momentary assessment approach. Behav. Res. Ther. 111, 44–51. https://doi.org/10.1016/j.brat.2018.09.008.

de Lijster, J.M., Dierckx, B., Utens, E.M.W.J., et al., 2017. The age of onset of anxiety disorders: a meta-analysis. Can. J. Psychiatr. 62 (4), 237–246. https://doi.org/10.1177/0706743716640757.

Drake, F.L., Van Rossum, G., Rossum, G.van, 2010. The Python Language Reference. Release 3.0.1 [Repr.]. Python Software Foundation.

Edmondson, D., Shaffer, J.A., Chaplin, W.F., Burg, M.M., Stone, A.A., Schwartz, J.E., 2013. Trait anxiety and trait anger measured by ecological momentary assessment and their correspondence with traditional trait questionnaires. J. Res. Pers. 47 (6) https://doi.org/10.1016/j.jrp.2013.08.005.

Frank, B., Jacobson, N.C., Hurley, L., McKay, D., 2017. A theoretical and empirical modeling of anxiety integrated with RDoC and temporal dynamics. Journal of Anxiety Disorders. 51, 39–46. https://doi.org/10.1016/j.janxdis.2017.09.002.

Hassan, E., 2005. Recall bias can be a threat to retrospective and prospective research designs. internetJournal of Epidemiology 3 (2).

Hovenkamp-Hermelink, J.H.M., van der Veen, D.C., Oude Voshaar, R.C., et al., 2019. Anxiety sensitivity, its stability and longitudinal association with severity of anxiety symptoms. Sci. Rep. 9 (1), 4314. https://doi.org/10.1038/s41598-019-39931-7.

Jacobson, N.C., Lekkas, D., Huang, R., Thomas, N., 2021. Deep learning paired with wearable passive sensing data predicts deterioration in anxiety disorder symptoms across 17–18 years. J. Affect. Disord. 282, 104–111. https://doi.org/10.1016/j.jad.2020.12.086.

Keras. https://keras.io/about/.

Kessler, R.C., 1994. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. Arch. Gen. Psychiatry 51 (1), 8. https://doi.org/10.1001/archpsyc.1994.03950010008002.

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Merikangas, K.R., Walters, E.E., 2005. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch. Gen. Psychiatry 62 (6), 593. https://doi.org/10.1001/archpsyc.62.6.593.

Khesht-Masjedi, M., Shokrgozar, S., Abdollahi, E., et al., 2019. The relationship between gender, age, anxiety, depression, and academic achievement among teenagers. J. Fam. Med. Prim. Care 8 (3), 799. https://doi.org/10.4103/jfmpc.jfmpc_103_18.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16 (9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Metricwire Inc. https://www.metricwire.com/.

Mifsud, C., Rapee, R.M., 2005. Early intervention for childhood anxiety in a school setting: outcomes for an economically disadvantaged population. J. Am. Acad. Child Adolesc. Psychiatry 44 (10), 996–1004. https://doi.org/10.1097/01.chi.0000173294.13441.87.

Pérez-García, F., Sparks, R., Ourselin, S., 2021. TorchIO: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Comput. Methods Prog. Biomed. 208, 106236 https://doi.org/10.1016/j.cmpb.2021.106236.

Rodman, A.M., Vidal Bustamante, C.M., Dennison, M., et al., 2020. A year in the social life of a teenager: within-person fluctuations in stress, phone communication, and anxiety and depression. PsyArXiv. https://doi.org/10.31234/osf.io/aekjt.

Rodman, A.M., Vidal Bustamante, C.M., Dennison, M.J., et al., 2021. A year in the social life of a teenager: within-persons fluctuations in stress, phone communication, and anxiety and depression. Clin. Psychol. Sci. 9 (5), 791–809. https://doi.org/10.1177/2167702621991804.

Sato, H., Ichiro, Kawahara J., 2011. Selective bias in retrospective self-reports of negative mood states. Anxiety Stress Coping 24 (4), 359–367. https://doi.org/10.1080/10615806.2010.543132.

Shaffer, F., Ginsberg, J.P., 2017. An overview of heart rate variability metrics and norms. Front. Public Health 5, 258. https://doi.org/10.3389/fpubh.2017.00258.

Solhan, M.B., Trull, T.J., Jahng, S., Wood, P.K., 2009. Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. Psychol. Assess. 21 (3), 425–436. https://doi.org/10.1037/a0016869.

Spitzer, R.L., Kroenke, K., Williams, J.B.W., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch. Intern. Med. 166 (10), 1092. https://doi.org/10.1001/archinte.166.10.1092.

TensorFlow Developers, 2022. TensorFlow. https://doi.org/10.5281/ZENODO.5949169. Published online February 2.

Walz, L.C., Nauta, M.H., aan het Rot, M., 2014. Experience sampling and ecological momentary assessment for studying the daily lives of patients with anxiety disorders: a systematic review. Journal of Anxiety Disorders 28 (8), 925–937. https://doi.org/10.1016/j.janxdis.2014.09.022.