Received: 27 June 2022 | Revised: 6 August 2022 | Accepted: 14 August 2022

DOI: 10.1002/erv.2949

WILEY

RESEARCH ARTICLE

A machine learning investigation into the temporal dynamics of physical activity-mediated emotional regulation in adolescents with anorexia nervosa and healthy controls

Damien Lekkas[1,2] | Joseph A. Gyorda[1,2] | Nicholas C. Jacobson[1,2,3,4]

[1]Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, Lebanon, New Hampshire, USA

[2]Quantitative Biomedical Sciences Program, Dartmouth College, Hanover, New Hampshire, USA

[3]Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Lebanon, New Hampshire, USA

[4]Department of Psychiatry, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire, USA

Correspondence
Damien Lekkas, Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, 46 Centerra Parkway, Suite 300, Office #313S, Lebanon, NH 03766, USA.
Email: Damien.Lekkas.GR@dartmouth.edu

Funding information
National Institute on Drug Abuse

Abstract

Objective: Anorexia nervosa (AN) is commonly experienced alongside difficulties of emotion regulation (ER). Previous works identified physical activity (PA) as a mechanism for AN sufferers to achieve desired affective states, with evidence towards mitigation of negative affect. However, temporal associations of PA with specific emotional state outcomes are unknown.

Method: Using lag-ensemble machine learning and feature importance analyses, 888 affect-based ecological momentary assessments across $N = 75$ adolescents with AN ($N = 44$) and healthy controls ($N = 31$) were analysed to explore significance of past PA, measured through passively collected wrist-worn actigraphy, with subsequent self-report momentary affect change across 9 affect constructs.

Results: Among AN adolescents, later lags (≥2.5 h) were important in predicting change across negative emotions (hostility, sadness, fear, guilt). AN-specific model performance on held-out test data revealed the holistic "negative affect" construct as significantly predictable. Only joviality and self-assurance, both positively-valenced constructs, were significantly predictable among healthy-control-specific models.

Discussion: Results recapitulated previous findings regarding the importance of PA in negative ER for AN individuals. Moreover, PA was found to play a uniquely prominent role in predicting negative affect 4.5–6 h later among AN adolescents. Future research into the PA-ER dynamic will benefit from targeting specific negative emotions across greater temporal scales.

KEYWORDS
affect, anorexia nervosa, ecological momentary assessment, emotion regulation, feature importance, lag, longitudinal data, machine learning, passive sensing, physical activity

Abbreviations: AN, anorexia nervosa; AUC, area under the receiver operating characteristic curve; BMI, body mass index; CI, confidence interval; EMA, ecological momentary assessment; ER, emotion regulation; HC, healthy controls; PA, physical activity; PANAS-X, positive and negative affect schedule – expanded version; ROC, receiver operating characteristic; SMOTE, selective minority oversampling technique.

Key Points

- Physical activity (PA) plays an important role in the regulation of negative emotion for those with anorexia nervosa (AN).
- Compared with healthy controls (HC), negative affect among adolescents with AN is uniquely predicted by PA 4.5–6 h prior.
- Further exploration of delayed associations between PA and specific negative affect states may help to better understand the dynamics of emotion regulation (ER) in those with AN.

1 | INTRODUCTION

Anorexia nervosa is a psychiatric eating disorder characterised by the maintenance of low body weight due to fear of weight gain (Manuel & Wade, 2013). Risk associated with AN is significant: meta analyses suggest almost a six-times-greater risk for mortality relative to HC (van Eeden et al., 2021). Moreover, AN has the highest mortality rate among all psychiatric disorders (Jagielska & Kacperska, 2017). Although AN is prevalent across all age groups (van Eeden et al., 2021), it is particularly and increasingly prevalent among adolescents (Jagielska & Kacperska, 2017). The onset of 85% of AN cases occurs before age 20, with the majority beginning between ages 14 and 18 (Jagielska & Kacperska, 2017). Importantly, AN is commonly experienced alongside ER difficulties (Manuel & Wade, 2013). Studies revealed that high and low levels of emotion dysregulation were associated with increases and decreases in AN symptoms, respectively (Meule et al., 2021; Racine & Wildes, 2015). Emotion regulation difficulties have also been found specifically in adolescents with AN (Nalbant et al., 2019). Given the heightened age-specific vulnerability, health risks, and morbidity associated with AN, investigation into behaviours that impact ER among AN adolescents is worthwhile.

Associations between AN and ER may be partially mediated by PA. Research indicates that those with AN may be more inclined to engage in PA than healthy individuals (El Ghoch et al., 2013; Keyes et al., 2015). Anorexia nervosa individuals also exhibit greater tendencies to exercise compulsively, or partake in driven and rigid exercise patterns with an incapacity to stop (Coniglio et al., 2022). Despite the detriments, people with AN often turn to exercise with hopes to alleviate their negative affect (Vansteelandt et al., 2007). Exercise may also elicit positive emotions since it is used to obtain a desired body image (Coniglio et al., 2022). If used in moderation, PA may be a sustainable way for people with AN to regulate their emotions. For instance, AN adolescents participating in a supervised exercise programme experienced greater levels of positive affect and reduced negative affect following exercise in comparison to HC (Noetel et al., 2016). Another study found higher PA levels associated with higher positive affect within the same hour and in the hour following activity (Karr et al., 2017). Thus, research suggests that PA may be a strategy for individuals with AN to cope with emotional instability.

Given preliminary support for AN, PA, and ER associations, utilising increasingly data-rich and complex methods for further investigation into the underlying mechanisms that drive these associations is worthwhile. Departing from traditional cross-sectional approaches, AN research regarding PA has benefited from the use of ecological momentary assessments (EMAs) (Bourke et al., 2021; Selby et al., 2014). Ecological momentary assessments involve repeated longitudinal sampling of individual behaviours and mental states (Selby et al., 2014) to obtain measurements within a naturalistic setting (Bourke et al., 2021). Ecological momentary assessments are a form of intensive sampling that provide dense, contextual information. Ecological momentary assessments in mental health research are especially useful when coupled with data available via passive sensing technologies (e.g., wearable devices). However, proper integration of EMA with passive sensing data requires robust models capable of handling longitudinal and multivariate complexities.

Recent developments in passive sensing technologies, in conjunction with machine learning techniques, have resulted in the emergence of powerful and novel means of unobtrusive mental health monitoring (Garcia-Ceja et al., 2018). Extensive research has shown the promise of machine learning both generally in the mental health domain (Cho et al., 2019; D'Alfonso, 2020; Le Glaz et al., 2021; Linardon et al., 2020; Shatte et al., 2019), and specifically, through integration of passive sensing data, for the identification of eating behaviours and disorders (Haynos et al., 2021; Meegahapola et al., 2020; Sadeh-Sharvit et al., 2020). Despite a lack of longitudinal machine learning studies within the eating disorder literature, especially those that model PA and emotion, the combination of longitudinal data and machine learning modelling encourages uniquely elucidating time-based feature engineering. For example, prior work used smartphone passive sensing data in conjunction with EMA to investigate time-lagged effects of app use on work fatigue and boredom (Lekkas et al., 2022). In general, models utilising time-lagged features capture the distribution of effects of a phenomenon (e.g., PA) on an outcome (e.g., ER) when measured across different time points. The synergistic potential of passive sensing and machine learning in the context of mental health indicates that this methodology could be extended to AN to distill novel and salient information in both new and previously analysed datasets.

Kolar et al. (2020) is a notable study that investigated the impact of PA on ER in adolescents with AN. This work utilised EMA and wearable sensor data to provide unique empirical support for existing psychological models of AN within a longitudinal framework. Briefly, researchers used linear mixed models to predict momentary EMA-based emotional state in AN and HC adolescents from PA data within the preceding 30 min. They found that increased levels of PA predicted lower levels of negative affect and higher levels of positive affect among AN individuals, with the former association not recapitulated in HC. The authors therefore hypothesised that a down-regulation of negative affect might be an AN-specific effect of PA. In laying initial groundwork for further interrogation, the study did not differentiate among specific positive and negative emotions. They further acknowledged that the study was primarily interested in short-term effects of PA on momentary affect and that future efforts should consider time-lagged effects to more fully investigate associations.

To build upon the analytical efforts of Kolar et al. (2020) while capitalising on the strengths of their published dataset, the current study interrogated a distally-reaching suite of lagged associations linking PA with subsequent self-report momentary affect. This work also used the 60-item Positive and Negative Affect Schedule – expanded version (PANAS-X) as a guided reference to combine and recontextualise EMA item response scores for a more nuanced representation of positive and negative affective state. Use of the PANAS-X is supported by recent research suggesting that a greater differentiation of broad emotional state categories better explains within-person changes in state emotion (Jacobson et al., 2021). Employing an ensemble machine learning modelling approach, predictions of models built from lag-specific derivatives of PA data simultaneously served as features to predict momentary affect change across nine emotional constructs that were then analysed for relative importance in prediction.

In summary, the current endeavour differs from the work of Kolar et al. (2020) in several important methodological and analytical respects, including the (a) use of all available EMA data versus only those from participants who completed greater than 33% of EMAs, (b) framing and quantification of affect change as a classification problem versus as a regression problem, (c) use of an ensemble machine learning modelling approach as opposed to linear mixed modelling, (d) addition of more specific emotional constructs as outcomes (e.g., joviality, self-assurance, fear) as opposed to analysis of only summative negative and positive affect items, (e) consideration of AN-PA associations across temporal lags versus no consideration of temporality, and (f) leveraging of model feature importance to summarise and highlight trends in the data. Thus, this study sought to provide unique insight into the temporal dynamics of PA-mediated ER in AN adolescents and was guided by the following aims and hypotheses:

- Predict subsequent self-report emotional state from passively sensed PA in AN and HC individuals.

Hypothesis 1 *Lagged-ensemble machine learning models will predict both positive and negative affect constructs significantly above chance (AUC $>>$ 0.5, Cohen's kappa $>$ 0.1) in both AN and HC individuals. In line with previous research findings (Bratland-Sanda et al., 2010; Coniglio et al., 2022; Kolar et al., 2020; Oldershaw et al., 2015; Vansteelandt et al., 2007), negative affect constructs will be better predicted overall for AN relative to HC.*

- Quantify the relative importance of past PA alongside age and body mass index (BMI) in the prediction of self-report emotional state.

Hypothesis 2 *Given previous research suggesting that PA aids in the ER of those with AN (Karr et al., 2017; Noetel et al., 2016), PA will drive model predictions of self-report emotional state, with age and BMI exhibiting a much lesser degree of relative importance in AN-based models compared with HC-based models.*

- Characterise the relative importance of past PA across a suite of time lags ranging from 30 min to 6 h prior to emotion-specific self-report.

Hypothesis 3 *Lag feature introspection will uncover differences in temporal PA importance between AN individuals and HC. These differences will highlight* *specific temporal contexts and emotional constructs for further research in AN.*

2 | METHODS

2.1 | Procedures

The data used for the current study was originally collected and analysed in a previously published research effort conducted in Germany (Kolar et al., 2020) and made publicly available for use (see 2.2). In the original study, $N = 51$ AN participants were recruited from a specialised treatment centre for eating disorders in Prien, Germany. Additionally, $N = 37$ HC participants were recruited by word-of-mouth invitation and at local youth groups near Mainz, Germany. Anorexia nervosa participants fulfiled DSM-5 criteria for restrictive, binge/purge, and subclinical subtypes via self-report measures on the Eating Disorder Examination Questionnaire (Hilbert et al., 2007) and the Eating Disorder Inventory-2 (EDI-2) (Thiel et al., 1997). Importantly, AN participants were assessed by an experienced clinician at admission to the study site's specialised eating disorders treatment department with the above self-report measures used to further support clinical diagnoses. Also through assessment by an experienced clinician, AN participants were excluded if diagnosed or suspected to be diagnosed with moderate to severe major depressive disorder, attention-deficit hyperactivity disorder, or borderline personality disorder, since these conditions are associated with changes in affective states and/or PA. Healthy controls participants were excluded at baseline if they reported a history of psychiatric treatment within the last 5 years or had current experience of heightened psychological distress that was comparable to AN participants. Psychological distress was assessed for both AN and HC participants via the German version of the Brief Symptom Inventory (Franke et al., 2011).

All procedures which led to the data leveraged in the current study were approved by the ethics committee of the local medical association in Mainz, Germany as well as the ethics board of the Ludwigs-Maximilians-University Munich, Germany.

2.2 | Dataset

This work utilised actigraphy and EMA data collected for a previously published manuscript (Kolar et al., 2020) and deposited for public use in an online open repository, Open Science Framework (https://osf.io/bk4x8/). Data from an all-female sample of $N = 44$ German adolescent (age 12–20 years) inpatients with DSM-5 (American Psychiatric Association, 2013) diagnosed AN and $N = 31$ HC were utilised for all analyses.

2.3 | Measures and metrics

2.3.1 | Physical activity count data

Physical activity was continuously measured over a 1-day period with a triaxial SOMNOwatch™ accelerometer device (SOMNOmedics, Randersacker, Germany) worn on the non-dominant wrist. Movement was recorded at an acceleration sampling rate of 60-Hz and binned as total counts per minute (Kolar et al., 2020). The complete log of per-minute PA counts for $N = 75$ individuals was the basis for the primary predictive variables in the current analysis (see 2.3.3).

2.3.2 | Secondary variables

Age and BMI were included as secondary predictive variables. Age at date of study inclusion (years with day-level resolution) and BMI (kg/m^2) were reported as continuous variables.

2.3.3 | Ecological momentary assessment self-report of emotional state

The dataset consisted of 888 EMA events across $N = 75$ individuals (4–19 EMAs per person; $\mu = 11.84$; $\sigma = 3.66$). Collected across a time span of 1 day and thus concurrent with the passively collected PA data, participants were prompted hourly from 7 AM to 11 PM through the MovisensXS application (movisens GmbH, Karlsruhe, Germany) on their Android smartphone device to answer questions regarding momentary affect. Emotional state was quantified via 5-point Likert scale responses (from "not at all" to "extremely") to a subset of items contained within the German translation of the validated, 60-item PANAS-X (Watson & Clark, 1999). Eight positive affect items – strong, enthusiastic, proud, attentive, happy, energised, confident, and cheerful – and eight negative affect items – nervous, disgusted, distressed, ashamed, angry at self, afraid, sad, and dissatisfied with self – were selected to measure self-report momentary affect based on previous research that found these PANAS-X items to be clinically relevant for patients with eating disorders (Engel et al., 2013).

Along with the Positive Affect and Negative Affect General Dimension Scales of the PANAS-X modelled by

Kolar et al. (2020), the current work applied the Basic Positive and Basic Negative Emotion Scales to quantify and summarise emotion self-report with further nuance. Accordingly, nine constructs were derived as outcomes of interest:

Joviality
Summed EMA responses to four positive affect items: cheerful, happy, enthusiastic, energetic.

Hostility
Ecological momentary assessment response to one negative affect item: disgusted.

Attentiveness
Ecological momentary assessment response to one positive affect item: attentive.

Self-assurance
Summed EMA responses to three positive affect items: strong, proud, confident.

Sadness
Ecological momentary assessment response to one negative affect item: sad.

Fear
Summed EMA responses to three negative affect items: afraid, nervous, distressed.

Guilt
Summed EMA responses to three negative affect items: ashamed, angry at self, dissatisfied with self.

Positive affect
Summed EMA responses to all eight positive affect items.

Negative affect
Summed EMA responses to all eight negative affect items.

2.4 | Preprocessing and lagged data generation

The original, published dataset is a timestamped log of EMA and passively collected, longitudinal PA information for 83 individuals. Entries for seven AN and one HC participant were removed due to complete absence of PA and/or EMA data. It is important to note that this study utilised data from more individuals ($N = 75$) than in Kolar et al. (2020) ($N = 62$) since this work did not remove participants from analysis if they had recorded EMA data in less than one-third of the measurement occasions; to avoid loss of information, all EMA data, regardless of total compliance, was leveraged in the current work. From this subset of 75 individuals, 12 separate datasets were derived. Each derived dataset used the times of EMA responses as anchors to calculate summative counts of past PA within 30-min windows. Past PA was operationalised by examining the summed PA counts within 30-min centred windows at half-an-hour increments prior to each EMA, extending from 30 min prior to each EMA to up to 6 h prior. Thus, time-lag-specific relationships between PA counts and subsequent emotional state for 888 EMA instances were represented across 12 separate tabular datasets, where each dataset contained EMA data and summed PA counts corresponding to one of the 12 time lags under consideration (i.e., 30, 60, 90, ..., 360 min prior to each EMA). A 30-min window for summing PA counts was selected due to the observation that it is sufficient to capture significant changes in activity prior to an EMA (Kolar et al., 2020). Moreover, Kolar et al. (2020) note that a 30-min window prevented possible overlap between successive 1-h EMA observations.

Depending on the time of an EMA (i.e., early morning), longer lags did not have representative windows of PA for association. Consequently, Lag_1 (30 min) had all 888 EMAs with paired summed PA counts, while Lag_2 – Lag_{11} (1–5.5 h) had 886 paired EMAs, and Lag_{12} (6 h) had 873 EMAs. To prepare each dataset for downstream modelling, EMAs without corresponding lagged PA sum counts were imputed as "NA".

2.5 | Outcome derivation

2.5.1 | Establishing participant-specific emotion baselines

This study sought to analyse each EMA instance as a group-level representation of AN or HC emotional state change rather than focussing on individual differences. Accordingly, baseline values were quantified for each participant, defined as the most common self-report score of each emotion experienced across time. As there is variation in what constitutes "normalcy" in the emotional states of individuals, it was important to first within-person standardise this normalcy such that classifications of "improvement" or "deterioration" relative to this baseline (see 2.5.2 below) were comparable across all EMAs regardless of the individual to which they belong. To accomplish this, the distributional mode of the summed response scores for each of the nine affect constructs (Sections Joviality–Negative affect in Section 2.3.3) was calculated per-individual. A recent meta-analysis on emotion dynamics has indicated that using mode, rather than mean and median, to index average individual emotion results in the least confounding with variability compared with mean and median (Ringwald & Wright, 2022).

2.5.2 | Binary classification of ecological momentary assessment responses relative to baseline

Using modal baseline values derived in 2.5.1, each EMA response, if corresponding to a positive affect construct, was labelled as “0” if the score fell at or below baseline, indicating “no change” or “deterioration”, or labelled as “1” if the score fell above baseline, indicating “improvement” with respect to the individual's most persistent magnitude of emotion-specific state. For negative affect constructs, the EMA response was labelled as “0” if the score fell at or above baseline (“no change” or “deterioration”) and “1” if the score fell below baseline (“improvement”). This binarisation represents the outcomes for subsequent classification modelling, namely predicting per-EMA, emotion-specific change as a function of prior PA.

2.6 | Data grouping for comparative modelling and analysis

The 12 lag datasets (see 2.4) were modelled independently at 3 distinct stratifications: (a) “EMA_{All}” considered EMAs belonging to both adolescents with AN and adolescents characterised as HC (N_{All} = 888), (b) “EMA_{AN}” considered EMAs belonging only to adolescents with AN ($N_{AN} = 461$), and (c) “EMA_{HC}” considered EMAs belonging only to HC ($N_{HC} = 427$).

2.7 | Descriptive statistics

Mean and range of individual baseline scores for each of nine affect constructs were calculated per group (EMA_{All}, EMA_{AN}, EMA_{HC}) along with associated binary class prevalence (count and percentage). Unequal variances two sample *t*-tests for difference in mean baseline score were performed between EMA_{AN} and EMA_{HC} for each affect construct to quantify significant differences. The *t*-value (*t*), degrees of freedom (*df*) for unequal variance, and the *p*-value for each *t*-test were reported.

2.8 | Machine learning predictive modelling

2.8.1 | Introduction to the modelling framework

The current work addresses the defined binary classification task by taking advantage of one special type of machine learning framework called an “ensemble.” Broadly speaking, ensemble frameworks have shown to yield higher classification performance than individual machine learning models (Hsieh et al., 2012). For more information on machine learning and the ensemble approach, interested readers are encouraged to consult this study's supplemental S1. In addition, readers may consider previously published primers and tutorials which offer an accessible and clinically relevant introduction to machine learning principles and their associated decision making processes (Lekkas et al., 2021; Rosenbusch et al., 2021; Scott, 2021).

2.8.2 | Internal cross-validation and external testing

The selected ensemble machine learning framework for this study consisted of 3 parallel and independent modelling pipelines on each of EMA_{All}, EMA_{AN}, and EMA_{HC} groupings of the 12 lag datasets. To perform internal cross-validation and external testing on held-out samples, each stratification's EMAs were randomly split such that 80% ($Ntrain_{All} = 712$; $Ntrain_{AN} = 369$; $Ntrain_{HC} = 343$) were used for 5-times-repeated, 5-fold cross-validation with grid-search hyperparameter tuning, and 20% ($Ntest_{All} = 176$; $Ntest_{AN} = 92$; $Ntest_{HC} = 84$) were used for external testing with the trained models. Data were split to ensure identical EMAs across lags for a given group. For more information on cross-validation, readers are encouraged to consult a recent tutorial by de Rooji and Weeda (de Rooij & Weeda, 2020).

2.8.3 | Lag-specific models – Cross-validation

The *caret* package (Kuhn, 2008) in R (v4.03) was used to build eXtreme Gradient BOOSTed tree (xgbtree) (Chen & Guestrin, 2016) models that each utilised one lag-specific training dataset (Lag_1–Lag_{12}) containing nine EMA outcomes (binarised change for nine affect constructs) and the predictor variables of PA count, age, and BMI. Unlike PA count, age and BMI were both static variables and therefore did not differ among EMAs belonging to the same individual nor across lags for any single EMA. In total, 324 models were trained. This included 12 lag-specific datasets to predict each of 9 affect outcomes for 3 groups. All trained models were saved for testing on the held-out set.

2.8.4 | Lag-based ensemble models – Cross-validation

Using binary class predictions from each set of 12 lag-specific, cross-validated models as features in a derived prediction space, 27 (9 affect outcomes per All/AN/HC grouping) ensemble xgbtree models were trained using 5-times-repeated, 5-fold cross-validation to derive final EMA-specific class predictions. All trained models were saved for testing on the held-out set.

2.8.5 | Model application to held-out test set

For testing, each of the initial 324 trained models were first applied to the held-out test data to derive the 27 corresponding lag-based prediction spaces for external testing of the previously trained ensemble models. Then, each of these 27 derived prediction spaces were applied as input to their respective ensemble model to arrive at the final class predictions of emotion-specific change on the held-out test set.

2.8.6 | Synthetic minority oversampling technique

To mitigate impact of class imbalance on model performance, selective minority oversampling technique (SMOTE) was applied on each lag-specific model in cross-validation and in testing using *caret*'s built-in functionality (Chawla et al., 2002).

2.9 | Model performance assessment

All ensemble models, in cross-validation and testing, were evaluated for overall accuracy, sensitivity, specificity, Cohen's kappa, area under the Receiver Operating Characteristic (ROC) curve (AUC), and the 95% confidence interval (CI) for the AUC using *caret*'s default model performance output alongside the *MLeval* package (John, 2020) in R. Models with a 95% CI that included 0.5 and/or had near-zero (<0.1) kappa scores were considered to be uninformative. Receiver Operating Characteristic curves were also constructed using base R and *MLeval* to plot the sensitivity/1–specificity trade-off for EMA_{All}, EMA_{AN}, and EMA_{HC} models across each emotional state-based outcome.

2.10 | Variable importance

To assess general relative importance of prior PA in predicting emotion-state change across time lags, the *varImp()* function in *caret* was applied to each cross-validated, lag-specific model (Kuhn, 2008). The average and standard deviation of scaled feature importance across models for each of PA counts, age, and BMI were reported for EMA_{AN} and EMA_{HC}. To assess relative importance of each time lag in the prediction of emotional construct-specific change, *varImp()* was also applied to each ensemble model. The average scaled lag importance across emotion-specific ensemble models for each group were reported. Additionally, the top five lags, along with their associated scaled importance by group (EMA_{AN} and EMA_{HC}) for each affect construct, were reported. From this, the average relative importance of the most proximal (Lag_1-Lag_4), intermediate (Lag_5-Lag_8), and most distal (Lag_9-Lag_{12}) lags were also compared between EMA_{AN} and EMA_{HC}. Scaled feature importance is dimensionless and ranges from 0 to 100.

3 | RESULTS

3.1 | Descriptive statistics of class representation

Statistics of self-report baseline emotion scores across affect constructs are provided in Table 1. EMA_{AN} and EMA_{HC} differed significantly across all measured constructs ($p < 0.05$). For a subset of constructs (i.e., hostility, sadness, guilt), the positive class comprised <10% of data across one or more stratifications. This distribution was prior to train/cross-validation and test set splits for modelling, thus SMOTE was employed to mitigate imbalance (see 2.8.6).

3.2 | Ensemble model performance on cross-validated data

Performance results of cross-validated ensemble models are presented in Table 2 with associated ROC curves illustrated in Figure 1. Of the 27 models, 26 performed significantly above chance (95% CI AUC does not contain 0.5) and were more informative than a completely random model (kappa ≥ 0.1). Ranked by kappa, guilt

TABLE 1 Descriptive statistics of outcome class representation across groups

		Emotion self-report baseline score					Prevalence of prediction classes	
	Grouping	Mean	Range	t-value	df	*p*-value	No change/ Deterioration from baseline	Improvement from baseline
JOV (4 items)	All	9.09	4.0–19.0				544 (61%)	333 (38%)
	AN	7.76	4.0–19.0	12.04	872.05	<0.0001	277 (60%)	173 (38%)
	HC	10.52	4.0–19.0				267 (63%)	160 (37%)
HOS (1 item)	All	1.61	1.0–5.0				806 (91%)	71 (8%)
	AN	2.08	1.0–5.0	−15.91	536.75	<0.0001	389 (84%)	61 (13%)
	HC	1.10	1.0–3.0				417 (98%)	10 (2%)
ATT (1 item)	All	2.54	1.0–5.0				677 (76%)	200 (23%)
	AN	2.46	1.0–5.0	2.47	884.84	0.014	363 (78%)	87 (19%)
	HC	2.63	1.0–5.0				314 (74%)	113 (26%)
SAS (3 items)	All	5.80	3.0–15.0				594 (67%)	283 (32%)
	AN	4.72	3.0–13.0	13.81	819.13	<0.0001	304 (66%)	146 (32%)
	HC	6.96	3.0–15.0				290 (68%)	137 (32%)
SAD (1 item)	All	2.11	1.0–5.0				757 (85%)	120 (14%)
	AN	2.88	1.0–5.0	−23.04	654.36	<0.0001	339 (74%)	111 (24%)
	HC	1.28	1.0–4.0				418 (98%)	9 (2%)
FER (3 items)	All	6.54	3.0–15.0				673 (76%)	204 (23%)
	AN	8.52	3.0–15.0	−22.44	684.43	<0.0001	303 (66%)	147 (32%)
	HC	4.41	3.0–12.0				370 (87%)	57 (13%)
GLT (3 items)	All	6.31	3.0–15.0				696 (78%)	173 (19%)
	AN	8.25	3.0–15.0	−21.53	659.86	<0.0001	304 (66%)	138 (30%)
	HC	4.26	3.0–12.0				392 (92%)	35 (8%)
POS (8 items)	All	17.43	8.0–38.0				515 (58%)	362 (41%)
	AN	14.93	8.0–37.0	13.2	848.47	<0.0001	264 (57%)	186 (40%)
	HC	20.11	8.0–38.0				251 (59%)	176 (41%)
NEG (8 items)	All	16.44	8.0–40.0				655 (74%)	214 (24%)
	AN	21.53	8.0–40.0	−24.59	621.66	<0.0001	296 (64%)	146 (32%)
	HC	11.05	8.0–26.0				359 (84%)	68 (16%)

Note: Metrics of *t*-value, *df*, and *p*-value represent results of unequal variances two-sample *t*-tests for EMA_{AN} and EMA_{HC}. Designations of "improvement" and "deterioration" for prediction classes are relative to whether the emotional construct is positively- or negatively valenced. If positively valenced, improvement corresponds to a self-report score above baseline and deterioration corresponds to a self-report score below baseline; if negatively valenced, improvement corresponds to a self-report score below baseline and deterioration corresponds to a self-report score above baseline. Total counts of each class per group and affect construct combination, along with percentages in parentheses, are provided. $N_{All} = 888$, $N_{AN} = 461$, $N_{HC} = 427$ logged EMA events.

Abbreviations: All, anorexia nervosa and healthy control EMAs; AN, anorexia nervosa EMAs; ATT, attentiveness; *df*, degrees of freedom; FER, fear; GLT, guilt; HC, healthy control EMAs; HOS, hostility; JOV, joviality; NEG, negative affect; POS, positive affect; SAD, sadness; SAS, self-assurance.

(AUC = 0.75; kappa = 0.33), sadness (AUC = 0.76; kappa = 0.32), and negative affect (AUC = 0.69; kappa = 0.32) were among the best predicted emotional constructs for ensemble models trained and validated on EMA_{AN} time-lagged data, while positive affect (AUC = 0.76; kappa = 0.39), sadness (AUC = 0.85; kappa = 0.34), and guilt (AUC = 0.84; kappa = 0.29) were among the best predicted emotional constructs for EMA_{HC}. This significant performance in cross-validation suggested a detectable overall association between PA and the majority of positive and negative affect dimensions in both AN individuals and HC, thereby

TABLE 2 Emotion-specific ensemble model performance summary on the cross-validated training set data

	Grouping	Accuracy	Sensitivity	Specificity	AUC	95% CI	Kappa
JOV	All	0.68	0.74	0.54	0.69	0.65–0.73	0.24
	AN	0.67	0.64	0.7	0.7	0.64–0.76	0.28
	HC	0.62	0.62	0.62	0.65	0.59–0.71	0.17
HOS	All	0.91	0.86	0.72	0.85	0.79–0.91	0.19
	AN	0.85	0.76	0.75	0.8	0.72–0.88	0.29
	HC	0.99	1	0.91	0.96	0.83–1.00	0.05
ATT	All	0.74	0.8	0.51	0.68	0.63–0.73	0.18
	AN	0.8	0.58	0.72	0.69	0.61–0.77	0.23
	HC	0.72	0.66	0.72	0.71	0.65–0.77	0.23
SAS	All	0.68	0.56	0.74	0.67	0.63–0.71	0.2
	AN	0.67	0.49	0.75	0.63	0.57–0.69	0.21
	HC	0.69	0.58	0.69	0.67	0.6–0.74	0.18
SAD	All	0.83	0.84	0.53	0.75	0.69–0.81	0.2
	AN	0.75	0.69	0.76	0.76	0.7–0.82	0.32
	HC	0.98	0.88	0.82	0.85	0.68–1.02	0.34
FER	All	0.74	0.87	0.49	0.72	0.67–0.77	0.15
	AN	0.69	0.48	0.86	0.71	0.65–0.77	0.3
	HC	0.85	0.94	0.54	0.79	0.71–0.87	0.24
GLT	All	0.81	0.8	0.64	0.78	0.73–0.83	0.28
	AN	0.72	0.57	0.81	0.75	0.69–0.81	0.33
	HC	0.93	0.91	0.67	0.84	0.74–0.94	0.29
POS	All	0.63	0.6	0.65	0.66	0.62–0.7	0.21
	AN	0.67	0.54	0.8	0.71	0.65–0.77	0.31
	HC	0.7	0.71	0.71	0.76	0.71–0.81	0.39
NEG	All	0.73	0.89	0.44	0.71	0.66–0.76	0.2
	AN	0.72	0.56	0.82	0.69	0.63–0.75	0.32
	HC	0.84	0.79	0.7	0.8	0.73–0.87	0.28

Note: The binary classification performance metrics for all 27 cross-validated lagged-ensemble models. Note that the EMA_{HC} results for HOS, SAD, and GLT, as well as the EMA_{All} results for HOS, should be interpreted with caution given the extremely low positive class representation (<10%). Kappa = Cohen's kappa.

Abbreviations: All, anorexia nervosa and healthy control EMAs; AN, anorexia nervosa EMAs; ATT, attentiveness; AUC, area under the (receiver operating characteristic) curve; FER, fear; GLT, guilt; HC, healthy control EMAs; HOS, hostility; JOV, joviality; NEG, negative affect; POS, positive affect; SAD, sadness; SAS, self-assurance; 95% CI, 95% confidence interval of the AUC.

supporting subsequent interrogation of lag importance structure.

3.3 | Lag-specific model variable importance

Among EMA_{AN} models, PA had the highest average relative importance (64.2 ± 42.6) followed by age (45.1 ± 40.8) and BMI (39.9 ± 45.3). By contrast, EMA_{HC} models saw PA as least important (41.9 ± 47.5) relative to age (47.1 ± 39.1) and BMI (55.9 ± 44.4). Moreover, PA was appreciably more important in predicting emotional state change in EMA_{AN} models compared with EMA_{HC} models. Standard deviations were high in both model groups, indicating that there was a large degree of variation among lag-specific, emotion-specific models despite the observed trends.

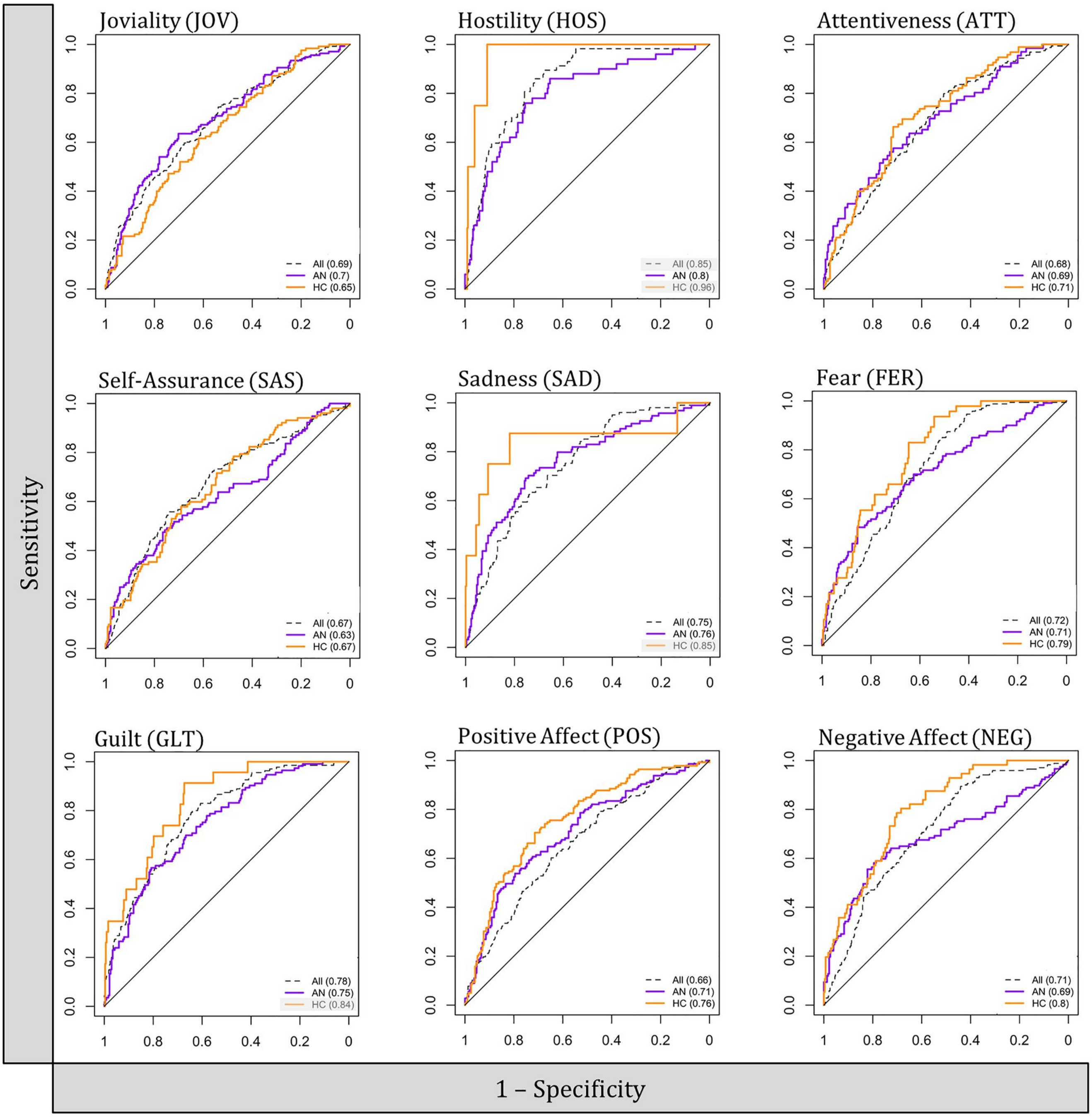

FIGURE 1 Emotion-Specific Ensemble Model Receiver Operating Characteristic (ROC) Curves on the Cross-Validated Training Data Set. ROC curves are plotted in triplets (EMA_{All}, EMA_{AN}, EMA_{HC}) according to the binary classification ensemble model's outcome of interest. AUC values for each curve are provided in parentheses next to the group names in each plot's legend. Note that the EMA_{HC} results for HOS, SAD, and GLT as well as the EMA_{All} results for HOS (grey highlight in legend) should be interpreted with caution given the extremely low positive class representation (<10%). AUC = area under the (ROC) curve; All = Anorexia nervosa (AN) and healthy control EMAs; AN = anorexia nervosa EMAs; healthy controls (HC) = healthy control EMAs

3.4 | Ensemble model variable importance

Broadly, feature importance analyses implicated Lag_9, Lag_{11}, and Lag_{12} – all distal lags – as driving predictions of emotion state change for EMA_{AN}, while Lag_1, Lag_4, and Lag_{11} were most influential for EMA_{HC} (Table S1). Table 3 compares the most important lag-based features between EMA_{AN} and EMA_{HC} for each affect construct-specific cross-validated model. On average, distal lags (Lag_9-Lag_{12}) were more important in predicting self-report emotion change from baseline in EMA_{AN} relative to EMA_{HC} models; this difference was more appreciable for negatively valenced construct models (80.57 vs. 57.09) compared with positively valenced construct models (70.77 vs. 64.53). Moreover, within EMA_{AN} for negatively valenced construct models specifically, distal lags were more important on average than proximal lags (80.57 vs. 68.84); this difference was more appreciable for negatively valenced, relative to positively valenced (70.77 vs. 67.92), construct models. Accordingly, proximal lags were marginally more significant on average for EMA_{HC} model predictions relative to distal lags regardless of construct valence (71.80 vs. 64.53 for positive; 68.66 vs. 57.09 for negative).

3.5 | Ensemble model performance on test data

Performance results of ensemble models on held-out test data are presented in Table 4 with associated ROC curves available as Figure S1. Of the 27 models, seven performed significantly above chance and were more informative than a completely random model. Only the negative affect model performed significantly above chance (AUC = 0.71; kappa = 0.34) on test EMA_{AN} time-lagged PA data, while only joviality (AUC = 0.67; kappa = 0.23) and self-assurance (AUC = 0.69; kappa = 0.34) performed significantly above chance on test EMA_{HC} data. Holistically, these results suggested a uniquely detectable association between PA and negative affect among AN individuals as well as a uniquely detectable association between PA and positive affect-related constructs among HC.

4 | DISCUSSION

This study extended analysis of a previously published dataset to investigate lagged associations of longitudinally and passively measured PA across a broad range of self-report positive and negative emotional states in $N = 75$ adolescents with AN and HC. Through employment of ensemble machine learning modelling and feature importance analyses, results bolstered previous findings regarding the general role of PA in altering momentary affect in those with AN. Results also extended current conceptualizations of this association by implicating a uniquely prominent role of distal lagged PA (four to six h) in the prediction of negatively valenced emotional states in AN adolescents. Model performance across affect constructs on a held-out test set revealed negative affect as the only significantly predictable construct among EMA_{AN}, with joviality and self-assurance—both positively valenced constructs—as significantly predictable among EMA_{HC}. As several models did not perform significantly on their respective held-out test set, and, by extension, not all negative constructs were better predicted for EMA_{AN} relative to EMA_{HC} models, Hypothesis 1 was only partially supported. Nonetheless, results holistically suggested uniquely significant, longer term effects of PA on negative affect-specific states in those with AN compared to HC.

When compared with age and BMI, analyses found PA to have the most important role in influencing prediction of AN emotional state change on average. By comparison, PA was found to be least influential (and comparable in influence to age and BMI) in the prediction of HC emotional state change. These results are in support of Hypothesis 2, which predicted a higher overall importance of PA for the prediction of emotional state change in AN-based models relative to HC-based models. Furthermore, introspection of EMA_{AN} ensemble model lags showed increased importance of distal lag associations among negatively valenced emotions – a pattern not observed among negatively valenced EMA_{HC} models nor among positively valenced models of either EMA_{AN} or EMA_{HC} (Table 3). Taken together, the observed informative differences in lag importance structure between EMA_{AN} and EMA_{HC} models are in support of Hypothesis 3.

Current AN theory places PA in two potential roles: as a pathogenic factor (Naumann et al., 2014; Vayalapalli et al., 2018) and as a coping mechanism (Haynos & Fruzzetti, 2011; Vayalapalli et al., 2018). Naturally, this study explored the latter. Empirical work within the coping literature indicates that AN patients with chronically negative affect have the strongest urge to be physically active. Herein, differences in association between PA and negative affect were found between EMA_{AN} and EMA_{HC}; however, the temporal dynamics of this association were surprizing. Current theory posits that PA is negatively reinforced by its ability to mitigate adverse emotional states in those with AN (Coniglio et al., 2022; Vansteelandt et al., 2007), suggesting more immediate temporal influence of PA on subsequent negative emotion state change. This was not entirely supported by the results, as the higher relative importance of later lags (>2.0 h) compared with proximal lags (≤2.0 h) in the prediction of AN-specific emotional state change indicates a uniquely delayed regulatory capacity of PA among those with AN. While proximal PA did also have some impact on predictions, the emphasis on delay may point to the influence of a potential intermediary variable or step in the PA-ER association, such as cognitive mechanisms related to the attenuation of attention bias or psychobiological mechanisms that modulate hormone

TABLE 3 Top lags by scaled model importance for positively valenced and negatively valenced emotions

	AN			HC		
	Lag	**Δt (hours)**	**Importance**	**Lag**	**Δt (hours)**	**Importance**
	POSITIVE VALENCE CONSTRUCTS					
JOV	7	3.5	100.00	7	3.5	100.00
	9	4.5	74.43	11	5.5	81.92
	3	1.5	66.23	4	2.0	77.59
	12	6.0	60.97	9	4.5	69.14
	8	4.0	49.22	12	6.0	68.87
ATT	1	0.5	100.00	8	4.0	100.00
	5	2.5	92.61	5	2.5	74.45
	2	1.0	73.69	4	2.0	68.69
	8	4.0	66.98	10	5.0	42.29
	12	6.0	63.54	1	0.5	35.67
SAS	11	5.5	100.00	2	1.0	100.00
	1	0.5	76.45	6	3.0	81.75
	6	3.0	59.69	10	5.0	53.94
	4	2.0	52.01	7	3.5	45.53
	12	6.0	49.09	9	4.5	44.52
POS	9	4.5	100.00	11	5.5	100.00
	2	1.0	63.74	1	0.5	92.61
	7	3.5	58.04	4	2.0	56.25
	11	5.5	47.33	12	6.0	55.58
	1	0.5	43.33	6	3.0	49.33
Prox Avg	*67.92*			*71.80*		
Intr Avg	*71.09*			*75.17*		
Dist Avg	*70.77*			*64.53*		
	NEGATIVE VALENCE CONSTRUCTS					
HOS	9	4.5	100.00	1	0.5	100.00
	11	5.5	92.30	11	5.5	42.61
	6	3.0	81.26	2	1.0	32.47
	12	6.0	78.97	10	5.0	19.72
	4	2.0	67.85	9	4.5	12.04
SAD	11	5.5	100.00	4	2.0	100.00
	6	3.0	67.02	9	4.5	88.11
	4	2.0	55.42	3	1.5	53.42
	12	6.0	46.44	12	6.0	41.88
	9	4.5	40.27	5	2.5	39.31
FER	9	4.5	100.00	11	5.5	100.00
	12	6.0	73.29	1	0.5	84.46
	2	1.0	69.59	6	3.0	81.79
	10	5.0	61.74	9	4.5	65.18

TABLE 3 (Continued)

	AN			HC		
	Lag	**Δ*t* (hours)**	**Importance**	**Lag**	**Δ*t* (hours)**	**Importance**
	4	2.0	55.59	12	6.0	45.34
GLT	10	5.0	100.00	5	2.5	100.00
	3	1.5	93.31	11	5.5	98.90
	5	2.5	88.19	8	4.0	63.50
	7	3.5	83.56	4	2.0	60.32
	4	2.0	59.22	3	1.5	58.29
NEG	10	5.0	100.00	2	1.0	100.00
	2	1.0	80.89	1	0.5	70.01
	9	4.5	80.44	6	3.0	63.37
	6	3.0	76.84	3	1.5	51.20
	11	5.5	73.97	4	2.0	45.06
Prox Avg	*68.84*			*68.66*		
Intr Avg	*79.37*			*69.59*		
Dist Avg	*80.57*			*57.09*		

Note: Proximal (*Prox*) lags were defined as lags representing no greater than 2 h prior to an EMA (Lag_1-Lag_4). Intermediate (*Intr*) lags were defined as lags between 2.5 and 4 h prior to an EMA (Lag_5-Lag_8). Distal (*Dist*) lags were defined as lags at least 4.5 h prior to an EMA (Lag_9-Lag_{12}). These delineations were selected to maintain consistency of representation (i.e., four lag variables for each of proximal, intermediate, and distal). Averages were calculated to summarise scaled importance beyond a simple rank order since, for some models, one or two lags dominated the prediction dynamics, while for others, several lags drove predictions with more equal influence. The time difference (Δt) represented by each lag is shown in hours.

Abbreviations: AN, anorexia nervosa EMAs; ATT, attentiveness; Avg, average; FER, fear; GLT, guilt; HC, healthy control EMAs; HOS, hostility; JOV, joviality; NEG, negative affect; POS, positive affect; SAD, sadness; SAS, self-assurance.

levels (Kolar & Gorrell, 2021). Temporality notwithstanding, the ability of both cross-validated (hostility, sadness, fear, guilt, negative affect; Table 2) and test-set (negative affect; Table 4) EMA_{AN} models to significantly predict negatively valenced emotion state change from baseline supports potentially meaningful associations between PA and negative affect regulation. The insignificance of the Negative Affect EMA_{HC} model evaluated on the test-set (Table 4) further highlights PA-mediated negative affect coping as a potentially defining behavioural mechanism of AN.

The above modelling results highlighted differences among negative affect constructs between AN and HC data – importantly, differences that were not observed among positive affect constructs. Accordingly, a focus on the significance of PA within the negative affect space of those with AN is warranted. One detailed qualitative study leveraged interpretive phenomenological analysis to explore how individual women deal with and make sense of their AN and revealed consistent narrative themes of negative feeling avoidance and regulatory control via compulsive exercise (Kolnes & Rodriguez-Morales, 2016). Anger, guilt, shame, and sadness were noted as negative emotions targeted for suppression or escape through vigorous exercise, with anger, sadness, and guilt also specifically implicated in relation to exercise-mediated regulatory control (Kolnes & Rodriguez-Morales, 2016). An array of methodologically distinct studies on PA rationale, meaning, and statistical modelling have also cited sadness/depression (Bratland-Sanda et al., 2010; Naumann et al., 2014; Vansteelandt et al., 2004), anger and disgust (Aruguete et al., 2012; Espeset et al., 2012; Moncrieff-Boyd et al., 2014; Vansteelandt et al., 2004), and guilt/shame (Bewell-Weiss & Carter, 2010; Oldershaw et al., 2015; Vansteelandt et al., 2004; Zunker et al., 2011) as particularly relevant negative emotions for targeted regulation in those with AN.

Looking closely at the modelling results for some of these emotions, the EMA_{AN} sadness state change model was both informative and significant (AUC = 0.76, kappa = 0.32) (Table 2) in cross-validation. Physical activity 5.5 h prior to EMA (Lag_{11}) was the most important predictor of EMA_{AN} state change with no other time differences exhibiting a relative importance score that exceeded 75.0 (Table 3). This pattern is in contrast to the

TABLE 4 Emotion-specific ensemble model performance on the held-out testing set data

	Grouping	Accuracy	Sensitivity	Specificity	AUC	95% CI	Kappa
JOV	All	0.63	0.5	0.76	0.65	0.57–0.73	0.19
	AN	0.46	0.09	0.93	0.41	0.29–0.53	−0.13
	HC	0.63	0.69	0.65	0.67	0.55–0.79	0.23
HOS	All	0.9	0.75	0.57	0.65	0.48–0.82	0.05
	AN	0.87	1	0.13	0.43	0.26–0.6	0.1
	HC	0.99	1	0.98	0.98	0.79–1.00	0
ATT	All	0.72	0.79	0.32	0.52	0.42–0.62	0.05
	AN	0.72	0.25	0.9	0.48	0.34–0.62	0
	HC	0.62	0.56	0.56	0.48	0.33–0.63	0.03
SAS	All	0.66	0.6	0.61	0.63	0.54–0.72	0.15
	AN	0.6	0.63	0.62	0.61	0.48–0.74	0.12
	HC	0.74	0.42	0.94	0.69	0.57–0.81	0.34
SAD	All	0.84	0.94	0.58	0.79	0.66–0.92	0.1
	AN	0.69	0.82	0.62	0.67	0.52–0.82	0.07
	HC	--	--	--	--	--	--
FER	All	0.75	0.97	0.32	0.63	0.52–0.74	−0.01
	AN	0.63	0.8	0.4	0.59	0.46–0.72	0.07
	HC	0.89	0.56	0.93	0.68	0.48–0.88	0
GLT	All	0.75	1	0.41	0.7	0.59–0.81	0.07
	AN	0.58	0.83	0.41	0.54	0.4–0.68	−0.03
	HC	0.89	0.88	0.65	0.72	0.51–0.93	−0.02
POS	All	0.54	0.85	0.36	0.56	0.47–0.65	0.03
	AN	0.56	0.8	0.47	0.57	0.45–0.69	0.09
	HC	0.58	0.62	0.56	0.58	0.45–0.71	0.05
NEG	All	0.73	0.79	0.65	0.72	0.62–0.82	0.23
	AN	0.73	0.44	0.94	0.71	0.59–0.83	0.34
	HC	0.52	0.91	0.38	0.55	0.36–0.74	−0.03

Note: The binary classification performance metrics for all 27 lagged-ensemble models on the held-out test set. Note that HC and All results for HOS should be interpreted with caution given the extremely low positive class representation (<10%). In addition, there was no positive class representation in the HC model for SAD, thus an assessment of performance could not be calculated (labelled as two dashes in the table). Kappa = Cohen's kappa.

Abbreviations: All, anorexia nervosa and healthy control EMAs; AN, anorexia nervosa EMAs; ATT, attentiveness; AUC, area under the (receiver operating characteristic) curve; FER, fear; GLT, guilt; HC, healthy control EMAs; HOS, hostility; JOV, joviality; NEG, negative affect; POS, positive affect; SAD, sadness; SAS, self-assurance; 95% CI, 95% confidence interval of the AUC.

EMA_{HC} model where a two-hour time difference (Lag_4) dominated model prediction. The lag structure governing AN sadness prediction adds perspective to a previous study whose findings were inconsistent with ER theory (Naumann et al., 2014). The causal pathway under scrutiny represented the first half (increased sadness → PA) of a two-stage association (increased sadness → PA → decreased sadness) implied by ER theory. Therein, desire to exercise *decreased* in both AN and HC women immediately after exposure to a sadness-eliciting stimulus, suggesting that PA may not be leveraged as an ER strategy to decrease negative affect in the short-term. Importantly, the work did not explore distal influences and pointed to the possibility that individuals with AN may use PA to cope with depressive mood at a later stage of the sadness generation process (Naumann et al., 2014). While the current study does not address the same component of the ER pathway, present results (germane to the PA → decreased sadness stage) nonetheless speak to the potential significance of delayed emotional regulatory influence of PA when the counteracting effects of physical passivity in depressive states (Rucker & Petty, 2004) are concerned.

The current analysis did not directly model "anger"; however, "disgust" (an item comprising the "hostility" construct) has been theoretically coupled (Fox & Power, 2009; Munro et al., 2017) and linked empirically (Fox et al., 2013; Fox & Harrison, 2008) with feelings of anger in those with AN and other eating disorders. Under this coupling paradigm, one thought is that disgust serves to suppress expressions of anger; however, further work is needed to support this association. Where anger is commonly cited as a highly prevalent emotion among eating disorder patients and hypothesised to be more closely regulated by PA than sadness (Aruguete et al., 2012; Naumann et al., 2014), there is no research (for either AN or in general) directly interrogating regulation of emotional disgust alongside PA. Naturally, the temporal scope of PA's influence on emotional disgust or anger regulation for those with AN is also unknown. The results of the hostility state change EMA_{AN} model offered some preliminary insight. The model was informative and significant (AUC = 0.8, kappa = 0.29) in cross-validation (Table 2). Three distal lags, reflecting PA counts at 4.5 (Lag_9), 5.5 (Lag_{11}), and 6 h prior (Lag_{12}) to EMA, and one intermediate lag representing counts of PA 3 h prior (Lag_6), drove the model's predictive dynamics of hostility state change with relative importance scores exceeding 75.0 (Table 3). This was in stark contrast to the EMA_{HC} hostility model, where only PA counts from the past half hour (Lag_1) dominated influence over prediction. Results holistically suggest that PA may serve to subsequently regulate hostility (disgust) across several hours in AN adolescents.

It is noteworthy that of all emotion constructs considered, sadness, guilt, and hostility were characterised by the lowest incidence of "improvement" among HC (Table 1). By contrast, AN participants had a higher proportion of EMAs with lower negative affect relative to baseline (sadness: 24% for AN vs. 2% for HC; guilt: 30% for AN vs. 8% for HC; hostility: 13% of AN vs. 2% for HC). Moreover, sadness, guilt, and hostility among AN participants were significantly higher than HC at baseline. These patterns give empirical credence to the potential increased utility of PA as a regulator for these negative emotions. In consideration of the holistic negative affect construct cross-validation models for both EMA_{AN} and EMA_{HC}, lag importance differed considerably in composition (Table 3). For EMA_{AN}, 4 of 5 lags with the highest relative importance (>70.00) corresponded to PA counts that occurred >2.5 h previously (Lag_{10}, Lag_9, Lag_6, Lag_{11}). For EMA_{HC}, both lags with the highest relative importance corresponded to PA that occurred <1.5 h previously (Lag_2, Lag_1). This indicated that negative affect change was broadly predicted by more recent PA in HC and less recent PA in AN adolescents.

The holistic positive affect change models (Positive Affect) were not found to be significantly predicted on held-out test data from either EMA_{AN} or EMA_{HC}. Moreover, no negatively valenced affect construct models performed significantly on EMA_{HC} held-out test data. Instead, for EMA_{HC}, only two positively valenced affect construct models, self-assurance and joviality, made significant predictions of affect change (AUC = 0.69, kappa = 0.34 and AUC = 0.67, kappa = 0.23, respectively; Table 4) on the held-out test data. For AN individuals, current results align with empirically derived hypotheses placing PA in a prominent functional role of negative emotion mitigation. However, null results for Positive Affect are contrary to the findings of Kolar et al. (2020), where higher levels of preceding PA were associated with higher levels of positive affect in AN individuals and HC. Several reasons could explain this difference, including the binary, "summative", and mode-derived nature of emotion outcome classification, and the more robust assessment of model performance on held-out data afforded through a machine learning approach when compared with the linear mixed effects model previously employed.

This study has limitations that are important to review. Although analysis was carried out on 888 unique EMA instances, and utility of the dataset was extended by making EMA outcome change classifications between individuals comparable, data nevertheless were derived from a cohort of $N = 75$ individuals. Relatedly, PA and EMA data were collected across a single day; longer time scales resulting in larger numbers of representative lag associations would have been preferred to more confidently distill patterns of association. Reactivity analyses, including ascertainment of time-of-day effects, could not be carried out with the data provided. There was also heterogeneity in AN subtype representation among the participants (with an 81% restrictive subtype majority). Coupled with a dearth of available information about the clinical presentation of AN within the cohort, subtypes and associated descriptors of AN were not analysed in the current study. More research is needed to explore the generalisability of the current findings to distinct subtypes of AN. Recruited by "word-of-mouth" (Kolar et al., 2020), HC participants potentially represented a sample of convenience which limits the generalisability of the findings. Most notably, however, the smaller study sample, in conjunction with a limited number of representative EMAs per person, led to severe class imbalance after stratification in some instances – a phenomenon that was especially true in the held-out test set across several affect outcomes for HC. Despite this, most models (26/27) were found to be informative (kappa ≥ 0.1) in cross-validation, allowing for statistically justifiable introspection of underlying model-specific relative feature importance. When the held-out test set was applied to models, a minority (7/27) performed significantly. While these results highlighted affect constructs for future research, they reflected a degree of model overfitting during training which ultimately suggested idiosyncrasy and variability in the nature of PA-ER associations. Coupled with a homogeneous sample population, results are not generalisable to other AN study populations. Methodologically, quantification of additional PANAS-X constructs from the originally selected AN-specific items resulted in constructs with incomplete item representation (e.g., "attentiveness" was only represented by "attentive", without "alert", "concentrating", and "determined"). While not an issue for modelling affect outcomes per se, interpretation must be tempered since the full complement of validated items comprising constructs could not be considered.

Despite these limitations, use of machine learning with cross-validation ensured that each model's performance was entirely quantified through data not used during training. Feature importance analyses on these same models thus presented as a robust means to explore the relative impact of PA across several independent time lags. This work has strength in its extended use of an existing dataset through temporal feature and theory-driven-outcome engineering. Results not only supported previous findings in the literature but also offered insights for new research directions.

The above findings, while preliminary, encourage further exploratory emphasis into the temporal influence of PA on emotional states in those with AN. Special attention should be given to the role of PA as a management strategy for negative emotions, and future research should explore the trajectories of additional negative affective states (e.g., "anger", "disgust", "nervousness") as well as probe the associations of PA with affect state change across greater time scales. Methodologically, researchers are encouraged to apply machine learning frameworks for further exploration and analysis. With more comprehensive, generalisable, and increasingly longitudinal datasets, machine learning can serve as a powerful means to maximise the synergistic utility of EMA and passive sensing data towards an increased understanding of eating disorders and their associated behaviours.

AUTHOR CONTRIBUTIONS

The following reflects individual contributions per the nomenclature of the Contributor Roles Taxonomy (CRediT): Damien Lekkas was responsible for conceptualisation, methodology, formal analysis, writing–original draft preparation, writing–review and editing, and visualisation. Joseph A. Gyorda was responsible for software, writing–original draft preparation, and writing–review and editing. Nicholas C. Jacobson contributed to formal analysis and writing—review and editing.

ACKNOWLEDGEMENT

This work was funded by an institutional grant from the National Institute on Drug Abuse (NIDA-5P30DA02992610).

CONFLICT OF INTEREST

None.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Open Science Framework at https://osf.io/bk4x8/.

ORCID

Damien Lekkas https://orcid.org/0000-0002-6995-9223
Joseph A. Gyorda https://orcid.org/0000-0003-3554-8546
Nicholas C. Jacobson https://orcid.org/0000-0002-8832-4741

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Aruguete, M. S., Edman, J. L., & Yates, A. (2012). The relationship between anger and other correlates of eating disorders in women. *North American Journal of Psychology*, *14*(1), 139–148.

Bewell-Weiss, C. V., & Carter, J. C. (2010). Predictors of excessive exercise in anorexia nervosa. *Comprehensive Psychiatry*, *51*(6), 566–571. https://doi.org/10.1016/j.comppsych.2010.03.002

Bourke, M., Hilland, T. A., & Craike, M. (2021). A systematic review of the within-person association between physical activity and affect in children's and adolescents' daily lives. *Psychology of Sport and Exercise*, *52*, 101825. https://doi.org/10.1016/j.psychsport.2020.101825

Bratland-Sanda, S., Sundgot-Borgen, J., Rø, Ø., Rosenvinge, J. H., Hoffart, A., & Martinsen, E. W. (2010). "I'm not physically active - I only go for walks": Physical activity in patients with longstanding eating disorders. *International Journal of Eating Disorders*, *43*(1), 88–92. https://doi.org/10.1002/eat.20753

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Cho, G., Yim, J., Choi, Y., Ko, J., & Lee, S.-H. (2019). Review of machine learning algorithms for diagnosing mental illness. *Psychiatry Investigation*, *16*(4), 262–269. https://doi.org/10.30773/pi.2018.12.21.2

Coniglio, K. A., Cooper, M., & Selby, E. A. (2022). Behavioral reinforcement of pathological exercise in anorexia nervosa. *International Journal of Eating Disorders*, *55*(2), 184–192. https://doi.org/10.1002/eat.23626

D'Alfonso, S. (2020). AI in mental health. *Current Opinion in Psychology*, *36*, 112–117. https://doi.org/10.1016/j.copsyc.2020.04.005

de Rooij, M., & Weeda, W. (2020). Cross-validation: A method every psychologist should know. *Advances in Methods and Practices in Psychological Science*, *3*(2), 248–263. https://doi.org/10.1177/2515245919898466

El Ghoch, M., Calugi, S., Pellegrini, M., Milanese, C., Busacchi, M., Battistini, N. C., Bernabè, J., & Dalle Grave, R. (2013). Measured physical activity in anorexia nervosa: Features and treatment outcome: Measured physical activity in anorexia nervosa. *International Journal of Eating Disorders*, *46*(7), 709–712. https://doi.org/10.1002/eat.22140

Engel, S. G., Wonderlich, S. A., Crosby, R. D., Mitchell, J. E., Crow, S., Peterson, C. B., Le Grange, D., Simonich, H. K., Cao, L., Lavender, J. M., & Gordon, K. H. (2013). The role of affect in the maintenance of anorexia nervosa: Evidence from a naturalistic assessment of momentary behaviors and emotion. *Journal of Abnormal Psychology*, *122*(3), 709–719. https://doi.org/10.1037/a0034010

Espeset, E. M. S., Gulliksen, K. S., Nordbø, R. H. S., Skårderud, F., & Holte, A. (2012). The link between negative emotions and eating disorder behaviour in patients with anorexia nervosa: Negative emotions in anorexia nervosa. *European Eating Disorders Review*, *20*(6), 451–460. https://doi.org/10.1002/erv.2183

Fox, J. R. E., & Harrison, A. (2008). The relation of anger to disgust: The potential role of coupled emotions within eating pathology. *Clinical Psychology & Psychotherapy*, *15*(2), 86–95. https://doi.org/10.1002/cpp.565

Fox, J. R. E., & Power, M. J. (2009). Eating disorders and multi-level models of emotion: An integrated model. *Clinical Psychology & Psychotherapy*, *16*(4), 240–267. https://doi.org/10.1002/cpp.626

Fox, J. R. E., Smithson, E., Baillie, S., Ferreira, N., Mayr, I., & Power, M. J. (2013). Emotion coupling and regulation in anorexia nervosa. *Clinical Psychology & Psychotherapy*, *20*(4), 319–333. https://doi.org/10.1002/cpp.1823

Franke, G. H., Ankerhold, A., Haase, M., Jäger, S., Tögel, C., Ulrich, C., & Frommer, J. (2011). Der einsatz des Brief Symptom Inventory 18 (BSI-18) bei psychotherapiepatienten. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, *61*(2), 82–86. https://doi.org/10.1055/s-0030-1270518

Garcia-Ceja, E., Riegler, M., Nordgreen, T., Jakobsen, P., Oedegaard, K. J., & Tørresen, J. (2018). Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing*, *51*, 1–26. https://doi.org/10.1016/j.pmcj.2018.09.003

Haynos, A. F., & Fruzzetti, A. E. (2011). Anorexia nervosa as a disorder of emotion dysregulation: Evidence and treatment implications. *Clinical Psychology: Science and Practice*, *18*(3), 183–202. https://doi.org/10.1111/j.1468-2850.2011.01250.x

Haynos, A. F., Wang, S. B., Lipson, S., Peterson, C. B., Mitchell, J. E., Halmi, K. A., Agras, W. S., & Crow, S. J. (2021). Machine learning enhances prediction of illness course: A longitudinal study in eating disorders. *Psychological Medicine*, *51*(8), 1392–1402. https://doi.org/10.1017/S0033291720000227

Hilbert, A., Tuschen-Caffier, B., Karwautz, A., Niederhofer, H., & Munsch, S. (2007). Eating disorder examination-questionnaire. *Diagnostica*, *53*(3), 144–154. https://doi.org/10.1026/0012-1924.53.3.144

Hsieh, S.-L., Hsieh, S.-H., Cheng, P.-H., Chen, C.-H., Hsu, K.-P., Lee, I.-S., Wang, Z., & Lai, F. (2012). Design ensemble machine learning model for breast cancer diagnosis. *Journal of Medical Systems*, *36*(5), 2841–2847. https://doi.org/10.1007/s10916-011-9762-6

Jacobson, N. C., Evey, K. J., Wright, A. G. C., & Newman, M. G. (2021). Integration of discrete and global structures of affect across three large samples: Specific emotions within-persons and global affect between-persons. *Emotion*. https://doi.org/10.1037/emo0001022

Jagielska, G., & Kacperska, I. (2017). Outcome, comorbidity and prognosis in anorexia nervosa. *Psychiatria Polska*, *51*(2), 205–218. https://doi.org/10.12740/PP/64580

John, C. R. (2020). MLeval: Machine learning model evaluation (R package version 0.3) [Computer software]. https://CRAN.R-project.org/package=MLeval

Karr, T. M., Cook, B., Zunker, C., Cao, L., Crosby, R. D., Wonderlich, S. A., & Mitchell, J. E. (2017). Examining physical activity and affect using objective measures: A pilot study of anorexia nervosa. *The Sport Journal*, *24*, 13.

Keyes, A., Woerwag-Mehta, S., Bartholdy, S., Koskina, A., Middleton, B., Connan, F., Webster, P., Schmidt, U., & Campbell, I. C. (2015). Physical activity and the drive to exercise in anorexia nervosa. *International Journal of Eating Disorders*, *48*(1), 46–54. https://doi.org/10.1002/eat.22354

Kolar, D. R., & Gorrell, S. (2021). A call to experimentally study acute affect-regulation mechanisms specific to driven exercise in eating disorders. *International Journal of Eating Disorders*, *54*(3), 280–286. https://doi.org/10.1002/eat.23427

Kolar, D. R., Neumayr, C., Roth, M., Voderholzer, U., Perthes, K., & Schlegl, S. (2020). Testing an emotion regulation model of physical activity in adolescents with anorexia nervosa: A pilot ecological momentary assessment. *European Eating Disorders Review*, *28*(2), 170–183. https://doi.org/10.1002/erv.2706

Kolnes, L.-J., & Rodriguez-Morales, L. (2016). The meaning of compulsive exercise in women with anorexia nervosa: An interpretative phenomenological analysis. *Mental Health and*

Physical Activity, *10*, 48–61. https://doi.org/10.1016/j.mhpa.2015.12.002

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Le Glaz, A., Haralambous, Y., Kim-Dufor, D.-H., Lenca, P., Billot, R., Ryan, T. C., Marsh, J., DeVylder, J., Walter, M., Berrouiguet, S., & Lemey, C. (2021). Machine learning and natural language processing in mental health: Systematic review. *Journal of Medical Internet Research*, *23*(5), e15708. https://doi.org/10.2196/15708

Lekkas, D., Price, G., McFadden, J., & Jacobson, N. C. (2021). The application of machine learning to online mindfulness intervention data: A primer and empirical example in compliance assessment. *Mindfulness*, *12*(10), 2519–2534. https://doi.org/10.1007/s12671-021-01723-4

Lekkas, D., Price, G. D., & Jacobson, N. C. (2022). Using smartphone app use and lagged-ensemble machine learning for the prediction of work fatigue and boredom. *Computers in Human Behavior*, *127*, 107029. https://doi.org/10.1016/j.chb.2021.107029

Linardon, J., Messer, M., Helms, E. R., McLean, C., Incerti, L., & Fuller-Tyszkiewicz, M. (2020). Interactions between different eating patterns on recurrent binge-eating behavior: A machine learning approach. *International Journal of Eating Disorders*, *53*(4), 533–540. https://doi.org/10.1002/eat.23232

Manuel, A., & Wade, T. D. (2013). Emotion regulation in broadly defined anorexia nervosa: Association with negative affective memory bias. *Behaviour Research and Therapy*, *51*(8), 417–424. https://doi.org/10.1016/j.brat.2013.04.005

Meegahapola, L., Ruiz-Correa, S., & Gatica-Perez, D. (2020). Alone or with others? Understanding eating episodes of college students with mobile sensing. *19th International Conference on Mobile and Ubiquitous Multimedia*, 162–166. https://doi.org/10.1145/3428361.3428463

Meule, A., Richard, A., Schnepper, R., Reichenberger, J., Georgii, C., Naab, S., Voderholzer, U., & Blechert, J. (2021). Emotion regulation and emotional eating in anorexia nervosa and bulimia nervosa. *Eating Disorders*, *29*(2), 175–191. https://doi.org/10.1080/10640266.2019.1642036

Moncrieff-Boyd, J., Byrne, S., & Nunn, K. (2014). Disgust and anorexia nervosa: Confusion between self and non-self. *Advances in Eating Disorders*, *2*(1), 4–18. https://doi.org/10.1080/21662630.2013.820376

Munro, C., Randell, L., & Lawrie, S. M. (2017). An integrative bio-psycho-social theory of anorexia nervosa. *Clinical Psychology & Psychotherapy*, *24*(1), 1–21. https://doi.org/10.1002/cpp.2047

Nalbant, K., Kalaycı, B. M., Akdemir, D., Akgül, S., & Kanbur, N. (2019). Emotion regulation, emotion recognition, and empathy in adolescents with anorexia nervosa. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity*, *24*(5), 825–834. https://doi.org/10.1007/s40519-019-00768-8

Naumann, E., Tuschen-Caffier, B., Voderholzer, U., & Svaldi, J. (2014). On the role of sadness in the psychopathology of anorexia nervosa. *Psychiatry Research*, *215*(3), 711–717. https://doi.org/10.1016/j.psychres.2013.12.043

Noetel, M., Miskovic-Wheatley, J., Costa, D., Crosby, R. D., Hay, P., Kohn, M., Madden, S., & Touyz, S. (2016). Exercise for the compulsive exercisers? An exploratory study in adolescent inpatients with anorexia nervosa. *Advances in Eating Disorders*, *4*(3), 264–276. https://doi.org/10.1080/21662630.2016.1202123

Oldershaw, A., Lavender, T., Sallis, H., Stahl, D., & Schmidt, U. (2015). Emotion generation and regulation in anorexia nervosa: A systematic review and meta-analysis of self-report data. *Clinical Psychology Review*, *39*, 83–95. https://doi.org/10.1016/j.cpr.2015.04.005

Racine, S. E., & Wildes, J. E. (2015). Dynamic longitudinal relations between emotion regulation difficulties and anorexia nervosa symptoms over the year following intensive treatment. *Journal of Consulting and Clinical Psychology*, *83*(4), 785–795. https://doi.org/10.1037/ccp0000011

Ringwald, W. R., & Wright, A. G. C. (2022). Overcoming the confound of means and variability for measuring everyday emotion dynamics related to neuroticism. *PsyArXiv*. https://doi.org/10.31234/osf.io/nxbyd

Rosenbusch, H., Soldner, F., Evans, A. M., & Zeelenberg, M. (2021). Supervised machine learning methods in psychology: A practical introduction with annotated R code. *Social and Personality Psychology Compass*, *15*(2), e12579. https://doi.org/10.1111/spc3.12579

Rucker, D. D., & Petty, R. E. (2004). Emotion specificity and consumer behavior: Anger, sadness, and preference for activity. *Motivation and Emotion*, *28*(1), 3–21. https://doi.org/10.1023/B:MOEM.0000027275.95071.82

Sadeh-Sharvit, S., Fitzsimmons-Craft, E. E., Taylor, C. B., & Yom-Tov, E. (2020). Predicting eating disorders from internet activity. *International Journal of Eating Disorders*, *53*(9), 1526–1533. https://doi.org/10.1002/eat.23338

Scott, I. A. (2021). Demystifying machine learning: A primer for physicians. *Internal Medicine Journal*, *51*(9), 1388–1400. https://doi.org/10.1111/imj.15200

Selby, E. A., Wonderlich, S. A., Crosby, R. D., Engel, S. G., Panza, E., Mitchell, J. E., Crow, S. J., Peterson, C. B., & Le Grange, D. (2014). Nothing tastes as good as thin feels: Low positive emotion differentiation and weight-loss activities in anorexia nervosa. *Clinical Psychological Science*, *2*(4), 514–531. https://doi.org/10.1177/2167702613512794

Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine*, *49*(09), 1426–1448. https://doi.org/10.1017/S0033291719000151

Thiel, A., Jacobi, C., Horstmann, S., Paul, T., Nutzinger, D. O., & Schüßler, G. (1997). Eine deutschsprachige Version des Eating Disorder Inventory EDI-2 [German translation of the Eating Disorder Inventory EDI-2]. *PPmP: Psychotherapie Psychosomatik Medizinische Psychologie*, *47*(9–10), 365–376.

van Eeden, A. E., van Hoeken, D., & Hoek, H. W. (2021). Incidence, prevalence and mortality of anorexia nervosa and bulimia nervosa. *Current Opinion in Psychiatry*, *34*(6), 515–524. https://doi.org/10.1097/YCO.0000000000000739

Vansteelandt, K., Pieters, G., Vandereycken, W., Claes, L., Probst, M., & Van Mechelen, I. (2004). Hyperactivity in anorexia nervosa: A case study using experience sampling methodology. *Eating Behaviors*, *5*(1), 67–74. https://doi.org/10.1016/S1471-0153(03)00062-X

Vansteelandt, K., Rijmen, F., Pieters, G., Probst, M., & Vanderlinden, J. (2007). Drive for thinness, affect regulation and physical activity in eating disorders: A daily life study. *Behaviour Research and Therapy*, *45*(8), 1717–1734. https://doi.org/10.1016/j.brat.2006.12.005

Vayalapalli, A., Deshpande, R., & Vayalapalli, A. (2018). The relationship between exercise and athletes to the pathogenesis and recovery from anorexia nervosa. *International Journal of Neuroscience and Behavioral Science*, *6*(2), 17–21. https://doi.org/10.13189/ijnbs.2018.060201

Watson, D., & Clark, L. A. (1999). The PANAS-X: Manual for the positive and negative affect schedule-expanded form.

Zunker, C., Mitchell, J. E., & Wonderlich, S. A. (2011). Exercise interventions for women with anorexia nervosa: A review of the literature. *International Journal of Eating Disorders*, *44*(7), 579–584. https://doi.org/10.1002/eat.20862

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lekkas, D., Gyorda, J. A., & Jacobson, N. C. (2022). A machine learning investigation into the temporal dynamics of physical activity-mediated emotional regulation in adolescents with anorexia nervosa and healthy controls. *European Eating Disorders Review*, 1–19. https://doi.org/10.1002/erv.2949