Computational and Structural Biotechnology Journal 28 (2025) 118–127

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Research Article

# Predictive assessment of eating disorder risk and recovery: Uncovering the effectiveness of questionnaires and influencing characteristics

A. Pikatza-Huerga [a,*], C. Las Hayas [b], U. Zulaika [a], A. Almeida [a]

[a] Faculty of Engineering, Universidad de Deusto, Avenida Universidades, Bilbao, Spain
[b] Faculty of Health Sciences, Universidad de Deusto, Avenida Universidades, Bilbao, Spain

ARTICLE INFO

*Keywords:*
Eating disorders
Machine learning
Feature importance
Questionnaires

ABSTRACT

This study aims to assess the predictive capabilities of various questionnaires in determining the risk of Eating Disorders (ED) and predicting the level of recovery among individuals. Employing machine learning models and diverse datasets, the research focuses on understanding the effectiveness of different questionnaires in providing insights into ED symptoms and recovery outcomes. Additionally, the study seeks to identify the characteristics that significantly influence the recovery process. The investigation aims to contribute valuable information to enhance the diagnostic and monitoring tools used in the field of mental health, particularly concerning ED.

## 1. Introduction

Mental disorders have a significant societal impact, affecting individuals across all demographics and cultures. According to the World Health Organization (WHO), approximately one in four people worldwide will experience a mental or neurological disorder at some point in their lives [1]. The lifetime prevalence of eating disorders (EDs) varies. For example, estimates place the prevalence of Anorexia Nervosa (AN) and Bulimia Nervosa (BN) under 1%, while Binge Eating Disorder (BED) and sub-threshold BEDs range from 5.6–6.9% [2]. These variations highlight the complexity of estimating global ED prevalence and the need for further research.

EDs persistently disrupt eating behaviour, causing severe physical, psychological, and social impairments. Medical complications, such as cardiac arrhythmias, osteoporosis, and gastrointestinal issues, are common, and EDs have the highest mortality rate for psychiatric disorders [3]. A study in Spain found that 3.6% of 11-year-olds met ED diagnostic criteria [4,5].

Prognosis studies reveal recovery rates of 45% for AN and BN, while 28–33% show improvement, 23% experience chronicity, and 5% (AN) or 0.32% (BN) succumb to related causes [6]. Mortality rates in young people with AN are 12 times higher than in the general population and twice as high as in other mental disorders [7,8]. Early detection, younger age, and education level influence short-term outcomes, though their long-term predictive value remains uncertain [9].

Despite advancements in ED research, predicting recovery remains challenging. Proposed predictors include demographic and clinical traits, personality, cognition, social support, and treatment variables, but their interactions and relative importance are poorly understood.

This study aims to develop a Machine Learning (ML) model to predict ED recovery within one year based on responses to standardised questionnaires, including the World Health Organization Quality of Life (WHOQOL) and Resilience Scale, alongside ED symptoms, mood, and anxiety measures. Additionally, we will assess the predictive value of each questionnaire. The model will identify key recovery factors and estimate individual recovery probabilities.

This research contributes significantly to ED treatment by identifying recovery factors and tailoring predictive models. Key contributions include:

Dataset preparation: Cleaned and processed data ready for analysis. Predictive power analysis: Comparative evaluation of questionnaires using ML approaches. Identifying recovery factors: Key traits associated with ED recovery identified. The article follows this structure: Section 1 introduces the study context and problem, Section 2 reviews related literature, Section 3 explains the methodology, Section 4 presents and discusses the results, and Section 5 concludes with key findings, practical implications, and future research directions.

* Corresponding author.
*E-mail address:* a.pikatza@deusto.es (A. Pikatza-Huerga).

https://doi.org/10.1016/j.csbj.2025.03.048
Received 16 December 2024; Received in revised form 28 March 2025; Accepted 31 March 2025

Available online 2 April 2025

## 2. Related work

In recent years, Machine Learning (ML) has significantly improved the prediction of eating disorder (ED) outcomes. Studies like Haynos et al. [10] have shown that elastic net models outperform traditional Logistic Regression in predicting ED diagnoses, binge eating, and underweight BMI, demonstrating higher accuracy and revealing important predictors such as initial ED diagnosis and demographic factors. Melisse et al. [11] found that factors like the severity of the disorder predict treatment outcomes, while Lammers et al. [12] highlighted the importance of baseline pathology. Rohrbach et al. [13] and Lydecker and Grilo [14] emphasised the role of psychological and psychiatric factors, such as motivation, body dissatisfaction, and comorbidities, in predicting treatment.

Moreover, Wang [15]'s study discusses the challenges of applying ML in ED research, such as small sample sizes and data imbalances, proposing strategies like multi-site collaborations to improve accuracy. However, it does not implement a specific predictive ML model.

In addition to ML, questionnaires like the Eating Disorder Examination Questionnaire (EDE-Q) have become vital tools for assessing ED symptoms and predicting treatment outcomes. The EDE-Q is a reliable tool, with higher baseline scores predicting poorer prognosis [12,16]. Other tools, like the Eating Attitudes Test (EAT-26) [17] and ChEDE-Q [18], also assess psychological constructs. Psychological assessments, such as the Patient Health Questionnaire (PHQ), enhance the predictive validity of treatment outcomes [13,19].

Krug et al. [20] compared ML methods to predict ED onset, finding that ML techniques like LASSO and Predictive Rule Sets offered concise predictive models. However, the study did not examine patient recovery. In Espel-Huynh et al. [21]'s study, a simple Logistic Regression model outperformed complex ML models in predicting short-term treatment responses, with an AUC of 0.93.

In this study, the measurement of ED symptoms will be based solely on information derived from questionnaire responses. Assessing the clinical utility of using questionnaires is crucial, as their routine application may provide information capable of modifying the treatment approach. It is important to note that the self-perception of individuals in this context plays a central role. Subjective experiences, attitudes and beliefs about the body, eating habits and general well-being significantly influence the development and progression of EDs. By focusing on self-perception as a predictor, this research sheds light on the subjective experiences and internal dynamics contributing to the manifestation and persistence of ED symptoms. Knowing the complex relationship between EDs and self-perception can help with early detection, intervention, and individualised treatment plans that address each patient's particular self-perceptual difficulties during recovery. At the same time, this study will assess the predictive capacity of each questionnaire individuals use to potentially discard some in the future and simplify the process for patients.

## 3. Methodology

This article comprehensively describes the project's development process and technical aspects. The primary objective is to document the methodologies, tools, and techniques utilised during the project's development phase, thereby providing an in-depth understanding of the technical intricacies that underpin the project.

This section consists of distinct sub-sections, each dedicated to an aspect of the project. It begins by exploring the data acquisition process and detailing the methods used for data collection. This section includes discussions of data sources, sampling techniques, and ethical considerations incorporated into the data collection procedure.

Following the acquisition of the dataset, the following subsection focuses on data cleaning. This phase involves preprocessing data and preparing it for analysis. It involves validating data, managing missing values, detecting outliers, and applying other data-cleaning techniques to ensure the dataset's quality and reliability.

The following subsection dives into the prediction phase, where machine learning (ML) algorithms are applied to build a predictive model. This section explores the selected ML techniques, approaches for feature selection or feature engineering, and model training and evaluation processes in depth.

### 3.1. Dataset

The dataset is composed of responses collected from individuals who participated in standardised and non-standardised questionnaires focused on assessing specific aspects of their psychological well-being, such as resilience, anxiety, and quality of life. Below is a detailed description of the process of acquiring this data.

Patients having a current diagnosis of ED, patients with a prior diagnosis of ED, and those without ED made up the study samples. Table 1 displays the three samples' sociodemographic traits and clinical information. From four mental health facilities, psychiatrists identified and enlisted patients with ED. According to the DSM-IV-TR [22] criteria for ED, a psychiatrist's diagnosis of ED was required for inclusion (American Psychiatric Association, 2000). At baseline, 165 patients participated, and 122 (73.9%) reacted after a year. The same mental health facilities also recruited patients who had recovered from an illness. These patients had been free of ED symptoms for at least a year. A total of 57 former ED patients completed the survey at the outset, and 33 (or a response rate of 57.9%) did so during the follow-up after a year. People approaching the University of Deusto (Bilbao, Spain) by any of the three main entrances were given flyers about the study and its inclusion criteria to recruit participants from the general community. The participant has to be a woman over 18 without a history of ED. A total of 148 women (42.41% response rate) out of the 349 who participated in the baseline survey did so at the one-year follow-up. Participants who answered at the follow-up did not differ from those who did not regarding their initial characteristics.

The research and ethical committees of the participating health centres supported data gathering, which had received approval from the University of Deusto's ethics committee. All individuals taking part in the study gave their informed consent.

Participants voluntarily provided their contact information (email address, phone number, or postal address) for follow-up. Between April 2013 and June 2014, current ED patients completed the baseline test battery. Former ED patients finished the baseline test battery between March 2013 and February 2014. Between April 2013 and November 2013, participants in the normative group completed the battery of baseline tests. The researchers approached all participants to complete the surveys once more after a year, either online, over the phone, or on paper, sent to their postal address. The study protocol for the analysis presented in this research was reviewed and approved in 2023.

Researchers administered an identical set of tests to participants at both the initial and follow-up visits, consisting of the following questionnaires: WHOQOL, which measures subjective well-being and satisfaction; Hospital Anxiety and Depression (HAD); EAT-26, Eating Attitude Test; The Resilence Scale (RS) - 25 [23], which measures a person's ability to cope and adapt positively in the face of adversity; Resilence in Eating Disorders (RED), which measures resilience but specific to EDs; RED5 (a reduced version of the specific resilience questionnaire for eating disorders) [24]; and SEIQoL, which measures personal satisfaction in various areas such as health, relationships, work, hobbies, and individual achievements. A more detailed description of each questionnaire follows:

#### 3.1.1. WHOQOL

The WHOQOL-BREF (Skevington, Lotfy, & O'Connell, 2004) was translated into Spanish (Espinoza, Osorio, Torrejón, Lucas-Carrasco, &

**Table 1**
Demographic Statistics for Age, Age at Onset, and Years Receiving Psychiatric Treatment.

| | | **DSM-IV-TR** Diagnosis | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **AN** | **BN** | **EDNOS** | **BED** | **AN and BN** | **Female** | **Age** | **Age at Onset** | **Treatment Years** |
| Baseline (T0) | CP (n = 165) | 37.8% | 28.8% | 13.5% | 3.2% | 16.7% | 95.1% | 30 (9.5) | 19.1 (6.8) | 5.9 (6.29) |
| | FP (n = 57) | 57.1% | 17.9% | 12.5% | 10.7% | 0% | 95% | 29 (7.42) | 17.2 (3.5) | 5.2 (4.39) |
| | GP (n = 349) | - | - | - | - | - | 100% | 27.81 (7.9) | - | - |
| **Follow-up (T1)** | **CP (n = 116)** | 46.9% | 22.1% | 12.4% | 4.4% | 12.4% | 94.2% | 30.75 (9.8) | 18.6 (6.4) | 5.8 (5.8) |
| | **FP (n = 29)** | 58.6% | 20.7% | 6.9% | 3.4% | 6.9% | 93.3% | 30.6 (8.1) | 17.3 (3.5) | 5.5 (4.9) |
| | **GP (n = 148)** | - | - | - | - | - | 100% | 29.59 (8.4) | - | - |

Notes: EDNOS - Eating disorder not otherwise specified, CP - Current Patients, FP - Former Patients, GP - General Population. Values represent medians with standard deviations in parentheses.

Bunout, 2011) to measure participants' overall quality of life. The Spanish version's confirmatory factor analysis revealed appropriate fit indices. The physical, psychological, social, and environmental domains are all measured by the 26 items that make up the WHOQOL-BREF. Each domain's WHOQOL-BREF ratings typically fall between 0 and 100, representing excellent health. The calculation of a WHOQOL-BREF total score involves directly summing its four domains, with a range from 0 (worst health) to 400 (highest health) (Cronbach's $\alpha = 0.93$), to compare WHOQOL-BREF scores with the SEIQoL total index directly. The relative Cronbach's $\alpha$ values for physical health, psychological health, environment, and social relationships were 0.82, 0.91, 0.78, and 0.76. According to Webster, Nicholas, Velacott, Cridland, and Fawcett (2010), the WHOQOL-BREF distinguishes between depressed and non-depressed women and has shown convergence with a well-being measure ($r \geq 0.45$).

### 3.1.2. HAD

The HAD is a widely used questionnaire to assess anxiety and depression in hospitalised patients. Developed by Zigmond and Snaith in 1983, it consists of 14 questions divided into two subscales. Each question has four response options, with scores ranging from 0 to 3. Scores for the anxiety and depression subscales range from 0 to 21, with higher scores indicating more severe symptoms.

### 3.1.3. EAT-26

The Eating Attitudes Test-26 measures ED patients' behavioural and cognitive traits (Garner, Olmsted, Bohr, & Garfinkel, 1982). Higher scores suggest ED symptoms at higher levels. Scores above 20 on this test indicate the existence of ED symptoms. The total score ranges from 0 to 78. Additionally, the test has been approved for use with Spanish-speaking individuals (Castro, Toro, Salamero, and Guimerá, 1991). For the whole scale in the current study, Cronbach's $\alpha$ was 0.94.

### 3.1.4. RS-25

This questionnaire assesses a person's ability to cope with and adapt to difficult situations, such as stress or trauma. It contains 25 questions that explore aspects such as self-confidence, stress management, and goal-setting skills by 7-point Likert scales for each item, ranging from '1' (Strongly Disagree) to '7' (Strongly Agree). [23]

### 3.1.5. RED

This questionnaire focuses on assessing the ability of these individuals to cope with and recover from the unique challenges associated with ED conditions. It includes 44 questions tailored to measure emotional strength, self-acceptance, ability to handle social pressure related to body image, and coping skills for problematic eating patterns. This questionnaire was created by psychologists at the University of Deusto and was the extensive draft version of RED5 explained below.

### 3.1.6. RED5

A comprehensive validation process of the RED questionnaire was carried out, including predictive tests, validity assessments and convergence analysis. As a result of this process, the reduced version RED5, consisting of 5 questions, was obtained. [24]

### 3.1.7. SEIQOL

Spanish translation [25] of the SEIQoL was permitted by Dr Hickney, one of the SEIQoL's creators, in November 2012. The administration manual [26] was made available in Spanish thanks to a professional translation firm that was hired. The paper and online SEIQoL self-report versions were pilot-tested with 15 women from the University of Deusto (staff and students) [mean (M) age = 35.13; standard deviation (SD) = 14]. Self-completion of the online SEIQoL took participants 10.58 minutes (SD = 4.21; range 5–20 minutes). 16% of the sample assessed the SEIQoL's length as suitable, 66% as short, and 16% as too short; three people did not comment. 66% of the pilot sample said it was simple to grasp the directions for completing the questionnaire, and 33% thought it was elementary. Considering which aspects of their lives were essential, the task's complexity was scored as easy or nearly easy by 75% of the sample and tough or quite tricky by the remaining 25%. It was determined that the paper and digital versions of the SEIQoL were suitable for use.

The administration manual's description of the SEIQoL [26] was utilised to evaluate the subject's idiographic QoL in four steps. In the first stage, the participant was asked to list the top five aspects of their lives (keys) influencing their quality of life. The second step included two additional steps: The participant was asked to rank the five areas (2a) in order of importance and (2b) to give each area a weight by dividing a total of 100 virtual points among the five areas, giving the most points to the sections they believed to be the most crucial. The participant was asked to rate each region as the last stage. The individual was then asked to rate each location on a visual analogue scale (VAS) scale from 0 (extremely poor) to 100 (excellent). Each weighting is divided by 100 to provide a SEIQoL score, with weighting scores ranging from 0 to 1. The VAS (0 to 100) score for each region is multiplied by the weight (0 to 1) for that area, and the sum of these products in the five areas is added up to obtain the overall SEIQoL index: $\sum$(VAS ratings $\times$ weights). An evaluation of the SEIQoL's performance in clinical research by Wettergren et al. (2009) reveals that this instrument seems workable and reliable. The convergent and discriminant validity findings are consistent with the assumption that the SEIQoL has a weak to moderate association with functional status and physical health and a moderate to high correlation with global QoL, life satisfaction, and mental health measures.

On a VAS scale from 0 (not at all recovered from ED) to 100 (fully recovered from ED), participants with a current or prior diagnosis of ED were asked for the clinical data listed: the age at ED onset, years with the disorder, years in treatment, date of discharge from treatment, and self-perceived level of recovery.

## 3.2. Preprocessing

First, data was divided into three groups: one for time T1, one for time T2, and one with general information about the individuals. Responses specific to T1 and T2 were selected for times T1 and T2. In addition, a dataset with personal details such as status and gender was created.

Then, we focused on predicting two things at time T2: the EAT-26 score and the level of recovery. To do this, we first created a dataset to predict the EAT-26 score. First, data was cleaned, removing unnecessary information such as identifiers and dates. Then, the scores were converted into binary labels: 1 for risk for ED (scores above 20) and 0 for no risk (scores equal to or below 20).

This dataset was then matched with general information about the individuals. Given the limited availability of data, removing rows with missing values was not a viable option, so several methods were tested to fill in the missing values, such as K-Nearest Neighbors (KNN) [27], iterative imputation, and filling with the median or mean. However, in terms of the F1 score, using the mean yielded the best results. The F1 score is a crucial metric in binary classification, as it balances precision and recall, providing a better evaluation of models when imbalanced classes. Ultimately, to preserve the integrity of the dataset and ensure the best performance metric, the decision was made to fill in missing values using the mean. The data was then prepared for analysis by separating the features from the labels, normalising the feature values, and dividing the data into training (80%) and test (20%) sets.

On the other hand, we followed a similar process for the recovery level prediction dataset with some differences in the target variable and labels. The first steps involved loading the overall dataset and the datasets for Time Points T1 and T2.

Instead of predicting a binary classification as in the EAT-26 dataset, the target variable for this dataset was the recovery level at T2. Recovery level values were extracted from the T2 dataset using the 'T2 RECOVERY LEVEL' column. These values represented the VAS scale range from 0 to 100, signifying the self-perceived level of recovery. This column was used as the label column in this dataset. No further transformations or conversions were necessary, as the values were already in numerical format, representing the self-perceived level of recovery on a continuous scale.

### 3.3. Predicting the risk of ED

As explained above, the Eating Attitudes Test (EAT) is a widely used assessment tool to evaluate an individual's attitudes, beliefs and behaviours related to eating habits and body image. It is a valuable instrument in mental health, especially for identifying the presence of ED. The selection of appropriate machine learning models for this task is influenced by the need to accommodate the unique characteristics of questionnaire-based data and track symptom changes over time. Here, we explore the rationale for employing various ML techniques to predict ED symptoms at a one-year interval.

- **Logistic Regression [28].** Logistic Regression is well suited for predicting outcomes based on multiple-choice questionnaire responses. It effectively models the relationship between categorical predictor variables and the probability of experiencing ED symptoms. Its interpretability allows for identifying which specific questionnaire items contribute most to symptom progression or recovery.
- **Decision Trees [29].** Decision Trees are valuable for identifying patterns in structured questionnaire responses. They capture nonlinear relationships between multiple-choice answers and symptom changes, providing interpretable insights into how different response combinations influence ED symptom trajectories.
- **Support Vector Machine (SVM) [30].** SVM is useful for classifying individuals based on their questionnaire responses by finding an optimal decision boundary. It is particularly effective when the relationship between multiple-choice answers and symptom categories is complex, allowing it to separate different symptom progression groups even in high-dimensional feature spaces.
- **Neural Networks [31].** Neural Networks can model intricate patterns in structured questionnaire data, capturing subtle relationships between multiple-choice responses. Their ability to learn from past assessments helps detect evolving trends in ED symptoms, making them particularly useful for long-term symptom prediction.
- **K-Nearest Neighbors (KNN) [32].** KNN effectively clusters individuals with similar symptom patterns based on their multiple-choice questionnaire responses. Comparing response profiles across different time points helps group individuals with worsening, improving, or stable symptoms, aiding in the identification of typical recovery trajectories.
- **Random Forest [33].** Random Forest enhances prediction accuracy by aggregating multiple Decision Trees trained on structured questionnaire data. It efficiently captures complex interactions between categorical response patterns while reducing the risk of overfitting, making it a robust choice for symptom progression modelling.
- **Gradient Boosting [34].** Gradient Boosting models, such as XGBoost and LightGBM, sequentially refine predictions by focusing on misclassified questionnaire response patterns. Their ability to capture nuanced trends in structured response data improves the accuracy of symptom progression predictions.
- **Naive Bayes [35].** Naive Bayes is particularly effective when multiple-choice questionnaire items can be treated as independent predictors. Its efficiency and scalability make it well-suited for analysing significant categorical responses, providing interpretable insights into the likelihood of symptom progression or recovery.

To maximise the effectiveness of these models, we conducted a comprehensive hyperparameter tuning process. This critical phase of optimisation allowed us to carefully adjust the settings of each algorithm, fine-tuning their performance and ensuring they were adequately calibrated for the specific data we were analysing. The focus on finding the best hyperparameters proved essential in achieving the most accurate and reliable results in our research.

### 3.4. Predicting recovery level

Similar to the prediction of risk of ED, the prediction of the level of recovery at time point 2 will be made using data from time point 1. As explained above, the values of the VAS scale range from 0 to 100, signifying the self-perceived level of recovery. The target variable to be predicted will be the recovery level itself. It is crucial to note that the level of recovery is assessed subjectively, as it is based on the answers provided by the people who completed the questionnaires. By exploiting the available data and using predictive models, the aim is to estimate and predict people's level of recovery based on their responses to the questionnaires.

In this case, the focus is on predicting patients' level of recovery based on their responses to questionnaires. To achieve this, a series of regression models specifically tailored to capture the continuous nature of the level of recovery, which ranges from 0 to 100, are employed. These models offer diverse algorithms and techniques to analyse the relationships between various input characteristics and the target variable. Using these regression models aims to understand better the factors that influence the subjective measure of recovery. In the following sections, each of the regression models used in this study will be examined, providing an overview of their underlying principles and demonstrating their effectiveness in predicting the level of recovery.

- **Linear Regression [36].** Linear Regression provides a straightforward and interpretable baseline for predicting recovery levels based on multiple-choice questionnaire responses. Assuming a linear relationship between predictor variables and recovery helps identify general trends in the data. Its simplicity helps to understand how individual questionnaire items contribute to recovery predictions.

- **Random Sample Consensus (RANSAC) Regression [37].** Given that real-world questionnaire data may contain inconsistencies or outliers, RANSAC is a robust alternative. This model iteratively fits regression models to subsets of the data, effectively mitigating the influence of outliers. Its ability to focus on the most representative responses ensures more stable predictions, making it well-suited for handling potential noise in recovery assessments.
- **Theil-Sen Regression [38].** This non-parametric regression method is particularly effective when dealing with questionnaire datasets that exhibit skewed distributions or extreme values. Its robustness to outliers ensures that anomalous responses do not disproportionately influence recovery predictions, making it a reliable option when recovery trends may not follow traditional linear assumptions.
- **Huber Regressor [39].** The Huber Regressor balances sensitivity to variations in multiple-choice responses and robustness against outliers. Its adaptive loss function ensures that extreme questionnaire responses do not overly distort the predictions while capturing meaningful trends. That makes it particularly useful when analysing questionnaire-based recovery scores, which may contain consistent and anomalous responses.
- **Support Vector Regression (SVR) [40].** When recovery trends exhibit non-linear relationships with questionnaire responses, SVR provides a flexible approach. Mapping responses into a higher-dimensional space identifies complex patterns that simpler regression models might miss. Its ability to handle both linear and non-linear relationships makes it valuable for capturing subtle variations in recovery trajectories based on structured questionnaire data.
- **XGBoost Regression [41].** As an advanced gradient boosting algorithm, XGBoost effectively models non-linear interactions between questionnaire responses and recovery outcomes. Its ability to sequentially improve predictions ensures that it captures complex response patterns, making it particularly useful for high-dimensional datasets where recovery depends on multiple interacting questionnaire variables.

As was done for the classification models, the regression models were also searched for the best hyperparameters to maximise their efficiency.

### 3.5. Analysis of the importance of features in prediction

In the analysis to predict the level of recovery and the risk of ED, it is crucial to determine which characteristics significantly impact the lack of recovery and the persistence of symptoms. Understanding the directly influential factors allows us to intervene accurately and effectively, improving the accuracy of our predictions. Furthermore, by analysing these aspects meticulously, we can identify which personal factors, how they perceive themselves, and their social environment most significantly affect their recovery.

The permutation feature importance technique has been used for that analysis; a model analysis method applied to any trained estimator when the data are organised in tabular form. This technique is beneficial for complex or non-linear estimators and involves assessing how the score of a model changes when the value of a single feature is randomly altered. This process breaks the relationship between that feature and the outcome to be predicted, allowing us to measure how much that feature influences the model's performance. Importantly, this technique is independent of the type of model used and can be applied multiple times with different permutations of the feature in question.

$$i_j = s - \frac{1}{K}\sum_{k=1}^{K} s_{k,j} \tag{1}$$

Using a previously trained predictive model ($m$) and a tabular dataset ($D$), the model's baseline score ($s$) is calculated on the original data.

**Table 2**
Comparison of the comorbidities between the different groups.

| | General (N = 350) | Recovered (N = 61) | Active (N = 168) |
|---|---|---|---|
| None | 296 (84,6) | 22 (36,1) | 30 (17,9) |
| Anxiety | 37 (10.6) | 42 (68.9) | 79 (47.0) |
| Depression | 3 (0.9) | 32 (52.5) | 81 (48.2) |
| Personality disorder | 1 (0.3) | 0 (0.0) | 22 (13.1) |
| Bipolar disorder | 1 (0.3) | 0 (0.0) | 1 (0.6) |
| Psychotic disorder | 0 (0.0) | 1 (1.6) | 1 (0.6) |
| Other | 5 (1.4) | 3 (4.9) | 24 (14.3) |

Then, for each feature $j$ in the dataset, $K$ repetitions of the following process are performed: the column $j$ of the data $D$ is randomly rearranged to create a corrupted version of the data ($\tilde{D}_{k,j}$). Subsequently, the score ($s_{k,j}$) of the model $m$ is calculated on this corrupted data. The importance ($i_j$) of feature $f_j$ is determined through this procedure by assessing how the model scores vary when the specific feature is randomly corrupted.

## 4. Results and discussion

This results section will present a comprehensive analysis to assess the prediction models' performance. It will also assess each questionnaire's predictive ability and analyze the importance of specific characteristics in prediction. The training code, along with the application that allows the use of the trained models, can be found on our GitHub repository (https://github.com/apikatza/ED-Outcome-Predictions). This repository includes the best models and synthetic data to test the application.

### 4.1. Participants

A total of 583 individuals were recruited for the study. However, after one year, 4 of them did not respond again, slightly reducing the final sample. See Table 2 for information on the participants' comorbidities.

### 4.2. Predictors

In the present study, the selection of predictor variables was informed by a conceptual model of recovery that integrates both clinical and experiential aspects of eating disorder recovery. This model is grounded in a previous study by Las Hayas et al. [42], which provided a rich, qualitative understanding of resilience and its pivotal role in the recovery process. This comprehensive approach allows us to encompass a broad spectrum of recovery dimensions, from symptom remission to the enhancement of psychosocial well-being, reflecting the complex and multifaceted nature of recovery.

Moreover, the theoretical underpinning of our predictors aligns with Antonovsky's [43] concept of recovery, which emphasises an empowerment-centred approach. This perspective was corroborated by the narratives from our earlier study, reinforcing the significance of individual agency and the construction of a positive, future-oriented personal vision. Additionally, this aligns with the literature that suggests diverse conceptualisations of recovery, acknowledging that the clinical and functional recovery models may coexist and interact in complex ways [44]. By integrating these models, our study captures a holistic view of recovery, incorporating both the reduction of clinical symptoms and the enhancement of quality of life through resilience and empowerment.

### 4.3. Predicting the risk of ED

The dataset is divided into training and test sets using a training and test split approach. The training-to-test ratio is set at 0.8, indicating that 80% of the data will be used for training and 20% for testing. The splitting is done while keeping the data balanced, ensuring the class distribution is preserved in both sets. This partitioning allows the models to be trained on a portion of the data and to evaluate their performance on unseen data, providing information on their generalisation capabilities.

Afterwards, the predictive models were trained and evaluated using appropriate evaluation metrics. The models' performance was assessed using ROC AUC, specificity, sensitivity, and F1-score metrics. ROC AUC measures the model's ability to discriminate between individuals at risk or not of Eating Disorder using the formula:

$$AUC - ROC = \frac{TP \times TN}{(TP + FN) \times (TN + FP)} \tag{2}$$

Meanwhile, specificity quantifies the model's ability to identify those without symptoms with the formula correctly:

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

Sensitivity measures its ability to identify those with symptoms using the formula correctly:

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

The F1-score provides an overall measure of accuracy, considering both false positives and false negatives:

$$F1 - score = \frac{2 \times TP}{2 \times TP + FN} \tag{5}$$

Where:

- $TP$ is a True Positive
- $TN$ is a True Negative
- $FP$ is a False Positive
- $FN$ is a False Negative

To comprehensively assess the prediction models' performance, all models discussed in the previous section were applied to each questionnaire separately. This approach ensured that the models were tested on various datasets, capturing the nuances and variations present in each questionnaire.

To further refine the accuracy of the predictions, a specific approach was employed that considered the scores obtained in each questionnaire. Considering the interconnections between the answers, the results of each questionnaire were derived by summing the scores on each question. This approach aimed to increase the accuracy of the predictions by incorporating the specific context and relevance of each response within its respective questionnaire.

By applying different models to different inputs, which included the complete set of responses, the responses to each questionnaire individually, and the scores for each questionnaire, we could evaluate the performance of the questionnaires and the models used. This comprehensive approach allowed us to obtain valuable information on the applicability and effectiveness of the questionnaires and models in detecting the risk of EDs.

The evaluation of the prediction models revealed that Naive Bayes achieved the highest ROC AUC score among all the models. Notably, the RED questionnaire showed excellent predictive performance when combined with the Naive Bayes model, as evidenced by the results in Table 3.

Considering the collective questionnaires, the models achieved a commendable ROC AUC score of 0.8. However, when focusing solely on the RED questionnaire, an even higher ROC AUC score of 0.83 was attained. This observation suggests that utilising only the RED questionnaire instead of having individuals respond to multiple questionnaires could provide a cost-effective alternative for determining the presence or absence of symptoms after a year. The higher ROC AUC score obtained with RED indicates its efficacy in accurately predicting the likelihood of individuals continuing to experience Eating Disorder symptoms. This finding highlights the potential efficiency and practicality of using a single, well-designed questionnaire to gain insights into the long-term prognosis of individuals regarding Eating Disorder symptoms.

**Table 3**
ROC AUC scores predicting EAT-26.

| | LR | DT | SVM | NN | KNN | RF | GB | NB |
|---|---|---|---|---|---|---|---|---|
| All | 0,68 | 0,71 | 0,50 | 0,70 | 0,68 | 0,69 | 0,67 | **0,80** |
| QS | 0,68 | 0,73 | 0,52 | 0,62 | 0,70 | 0,71 | **0,77** | 0,75 |
| WHOQOL | 0,69 | 0,62 | 0,72 | 0,72 | 0,70 | 0,69 | 0,66 | **0,73** |
| HAD | 0,70 | 0,60 | 0,70 | 0,58 | 0,62 | 0,69 | 0,64 | **0,77** |
| EAT | 0,65 | 0,68 | 0,71 | 0,66 | 0,72 | 0,70 | 0,59 | **0,73** |
| RED | 0,61 | 0,64 | 0,62 | 0,60 | 0,64 | 0,62 | 0,70 | **0,83** |
| RED5 | 0,50 | 0,50 | 0,50 | 0,49 | 0,61 | **0,67** | 0,61 | 0,63 |
| SEIQoL | 0,50 | 0,58 | 0,50 | 0,50 | 0,45 | **0,61** | 0,58 | 0,50 |
| RS25 | 0,62 | **0,69** | 0,55 | 0,64 | 0,57 | 0,68 | 0,54 | 0,68 |

Notes: LR - Logistic Regression, DT - Decision Tree, NN - Neural Network, RF - Random Forest, GB - Gradient Boosting, NB - Naive Bayes, QS - Questionaries' scores.

**Table 4**
Sensitivity scores predicting EAT-26.

| | LR | DT | SVM | NN | KNN | RF | GB | NB |
|---|---|---|---|---|---|---|---|---|
| All | 0,41 | 0,53 | 0,00 | 0,47 | 0,41 | 0,47 | 0,41 | **0,82** |
| QS | 0,47 | 0,59 | 0,12 | 0,35 | 0,53 | 0,53 | **0,71** | **0,71** |
| WHOQOL | 0,47 | 0,35 | 0,53 | 0,53 | 0,47 | 0,47 | 0,35 | **0,71** |
| HAD | 0,41 | 0,24 | 0,41 | 0,18 | 0,24 | 0,41 | 0,35 | **0,76** |
| EAT | 0,35 | 0,47 | 0,53 | 0,35 | 0,53 | 0,47 | 0,24 | **0,71** |
| RED | 0,24 | 0,35 | 0,29 | 0,24 | 0,35 | 0,29 | 0,47 | **0,88** |
| RED5 | 0,00 | 0,00 | 0,00 | 0,00 | 0,29 | **0,35** | 0,24 | 0,29 |
| SEIQoL | 0,00 | 0,18 | 0,00 | 0,00 | 0,00 | **0,24** | 0,18 | 0,00 |
| RS25 | 0,29 | 0,47 | 0,12 | 0,35 | 0,18 | 0,41 | 0,12 | **0,59** |

Additionally, Naive Bayes exhibited strong sensitivity, outperforming other models in accurately identifying individuals with ED symptoms, as highlighted in Table 4. That indicates that Naive Bayes effectively captures the patterns and characteristics of the dataset, making it a promising model for predicting the presence of symptoms.

Given the notable disproportion in the data set, with only 84 cases indicating risk in contrast to 496 cases that do not, it is evident that there are a more significant number of individuals who, after one year, are not at risk of developing eating disorders (ED). So, the Synthetic Minority Over-sampling Technique (SMOTE) Fernández et al. [45] was employed to address the class imbalance. SMOTE is a technique to create synthetic minority class samples by interpolating between existing instances. This process helps to balance the class distribution and improve the performance of ML models.

However, after applying SMOTE to the dataset, the results failed to outperform the models trained without using SMOTE. The best-achieved ROC AUC score of 0.81 was obtained using Naive Bayes with questionnaire scores, but it was still lower than the performance achieved without employing SMOTE. This outcome indicated that SMOTE did not effectively enhance the predictive accuracy of the models for the given dataset. As a result, the decision was made to discard the use of SMOTE in the final analysis.

In evaluating prediction models, the specificity metric consistently demonstrated strong performance across most models and subsets of data [Table 5]. That implies that the models could identify individuals without Eating Disorder symptoms accurately. However, it is necessary to consider the results of both specificity and sensitivity simultaneously. Seeing that some models scored 1 for one of the metrics and 0 for the other, we can extract that in those cases, the classification is not good and that the model predicts in all cases that the individual will be symptom-free after one year.

On the other hand, the range of F1 scores obtained [Table 6] reflects the complexity of predicting these symptoms and the associated challenges. While some models achieved relatively higher F1 scores, indicating a better balance between true positives and negatives, others obtained lower scores, highlighting the need for further refinement.

**Table 5**
Specificity scores predicting EAT-26.

| | LR | DT | SVM | NN | KNN | RF | GB | NB |
|---|---|---|---|---|---|---|---|---|
| All | 0,94 | 0,89 | 1,00 | 0,93 | 0,94 | 0,91 | 0,93 | 0,77 |
| QS | 0,90 | 0,87 | **0,93** | 0,89 | 0,88 | 0,90 | 0,84 | 0,80 |
| WHOQOL | 0,92 | 0,88 | 0,91 | 0,92 | 0,93 | 0,91 | **0,96** | 0,76 |
| HAD | 0,99 | 0,96 | 0,99 | 0,99 | 1,00 | 0,96 | 0,93 | 0,78 |
| EAT | 0,94 | 0,90 | 0,90 | **0,96** | 0,91 | 0,93 | 0,95 | 0,76 |
| RED | **0,98** | 0,92 | 0,94 | 0,97 | 0,93 | 0,95 | 0,93 | 0,78 |
| RED5 | 1,00 | 1,00 | 1,00 | 0,99 | 0,92 | 0,98 | 0,99 | 0,97 |
| SEIQoL | 1,00 | 0,99 | 1,00 | 1,00 | 0,91 | 0,98 | 0,99 | 1,00 |
| RS25 | 0,95 | 0,92 | **0,98** | 0,93 | 0,97 | 0,94 | 0,96 | 0,78 |

**Table 6**
F1 scores predicting EAT-26.

| | LR | DT | SVM | NN | KNN | RF | GB | NB |
|---|---|---|---|---|---|---|---|---|
| All | 0,47 | 0,49 | 0,00 | 0,50 | 0,47 | 0,47 | 0,45 | **0,52** |
| QS | 0,46 | 0,50 | 0,15 | 0,35 | 0,47 | 0,50 | **0,53** | 0,49 |
| WHOQOL | 0,48 | 0,34 | 0,51 | **0,53** | 0,50 | 0,47 | 0,44 | 0,45 |
| HAD | 0,56 | 0,32 | 0,56 | 0,29 | 0,38 | 0,50 | 0,40 | 0,50 |
| EAT | 0,41 | 0,46 | 0,50 | 0,44 | **0,51** | 0,50 | 0,31 | 0,45 |
| RED | 0,35 | 0,39 | 0,36 | 0,33 | 0,40 | 0,37 | 0,50 | 0,56 |
| RED5 | 0,00 | 0,00 | 0,00 | 0,00 | 0,33 | **0,48** | 0,36 | 0,40 |
| SEIQoL | 0,00 | 0,29 | 0,00 | 0,00 | 0,00 | **0,35** | 0,29 | 0,00 |
| RS25 | 0,37 | **0,48** | 0,19 | 0,40 | 0,26 | 0,47 | 0,17 | 0,41 |

### *4.4. Predicting recovery level*

The dataset for predicting the level of recovery is also divided into training and test sets using a training-test split approach. The training-to-test ratio is set to 0.8, indicating that 80% of the data will be used for training and 20% for testing. However, unlike the previous scenario, it is not necessary to maintain a balance based on the recovery level variable, as the outcome is continuous. Therefore, the splitting into train and test is performed, allowing us to assess the degree of generalisability of the models to unseen data and providing us with information about their effectiveness in predicting the level of recovery.

Following the data analysis, the prediction models were trained and tested using an appropriate evaluation metric: the R2 score. The R2 score, also known as the coefficient of determination, measures the proportion of variance in the recovery level that the prediction models explain. A higher R2 score indicates a better fit of the models to the data and suggests higher predictive accuracy.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

Where:

- $n$ is the number of observations in the data set.
- $y_i$ is the actual value of the dependent variable.
- $\hat{y}_i$ are the values predicted by the model.
- $\bar{y}$ is the mean of the actual values of the dependent variable.

In predicting the level of recall, the best R2 score was obtained using the SVR model with questionnaire scores, as seen in Table 7. The questionnaire scores process, as explained above, consisted of creating a single feature for each questionnaire that represented the overall score obtained in that questionnaire rather than considering the individual responses to all questions.

By achieving the highest R2 score, the SVR model with questionnaire scores proved effective in capturing and explaining variability in the level of recall. That indicates that the model successfully used the questionnaire summary scores as informative features to make accurate recall predictions.

The questionnaire scores approach of representing each questionnaire with a single score simplified the model input and reduced noise and dimensionality. This improved the predictive power of the SVR model, allowing it to capture patterns and relationships between questionnaire scores and recall better.

**Table 7**
R2 scores predicting recovery level.

| | Linear Regression | RANSAC | Theil Sen | Huber Regressor | SVR | XGBoost Regression |
|---|---|---|---|---|---|---|
| All | -4,73 | -0,05 | -4,73 | 0,11 | **0,62** | 0,47 |
| QS | 0,63 | 0,56 | 0,63 | 0,65 | 0,68 | 0,64 |
| WHOQOL | 0,30 | 0,24 | 0,29 | 0,13 | **0,55** | 0,46 |
| HAD | 0,47 | -3,26 | **0,54** | 0,49 | 0,50 | 0,29 |
| EAT | 0,12 | -0,95 | 0,17 | 0,12 | **0,62** | 0,12 |
| RED | 0,21 | -3,37 | 0,20 | 0,24 | **0,48** | -0,10 |
| RED5 | **0,23** | -0,24 | -0,01 | 0,05 | 0,12 | -1,00 |
| SEIQoL | 0,50 | 0,58 | 0,50 | 0,50 | 0,45 | **0,61** |
| RS25 | 0,27 | 0,07 | 0,12 | 0,23 | **0,39** | 0,13 |

### *4.5. Analysis of the importance of features in prediction*

In the following, we present our findings after a detailed analysis of the features' importance in the context of our study. This meticulous analysis has allowed us to identify the crucial aspects that influence the results of our predictive models. By understanding which features significantly impact the results, we have gained a deeper and more accurate understanding of the data. Through these findings, we gain a clearer picture of the determinants in the context of our study.

#### *4.5.1. When predicting the risk of ED*

It is worth noting that the RED5 questionnaire, although considered a refined subset of the RED questionnaire, exhibited inferior results compared to the entire questionnaire. The RED5 questionnaire specifically includes questions RESI16, RESI27, RESI28, RESI30, and RESI31, while the complete questionnaire contains an additional question, RESI33, which is not present in RED5.

When utilising the entire questionnaire, the prediction models placed a greater emphasis on question RESI33, as evidenced by the feature importance plot in Fig. 1. That suggests that RESI33 may hold significant predictive power in determining the presence or absence of Eating Disorder symptoms. However, since RESI33 is not included in the RED5 questionnaire, the models trained on RED5 alone may lack important information, leading to inferior performance.

The question "RESI33: *Me acepto con mis defectos y mis virtudes*" ("I accept myself with my flaws and virtues") is a key component in understanding individuals' self-perception, understood as a person's view of their self or of any of the mental or physical attributes that constitute the self [46]. This question allows individuals to express their level of self-acceptance and how they view themselves holistically with their ED in mind. Interestingly, this question has proven to be significant in predicting the occurrence of symptoms in the future.

The fact that the self-acceptance question, RESI33, plays a crucial role in predicting symptoms suggests that an individual's perception of themselves substantially impacts their overall well-being. The degree to which individuals accept themselves, including their flaws and virtues, can influence their mental health and susceptibility to Eating Disorder symptoms. This finding emphasises the importance of self-perception and self-acceptance in the development and progression of EDs.

The questions RESI32, "I accept myself as I am"; RESI34, "I am generally satisfied with myself"; and RESI35, "I believe I am a valuable person even if people disapprove of me", delve into the self-acceptance and self-esteem aspects of individuals, as does RESI33. They explore the individual's perception of themselves and their overall satisfaction with who they are. These questions reflect the importance of self-acceptance, self-esteem, and self-worth concerning one's mental well-being.

RESI14, which states, "I have an idea of my past and how it has influenced my behaviour and life choices", is the third most important question for predicting future symptoms. This question is about self-

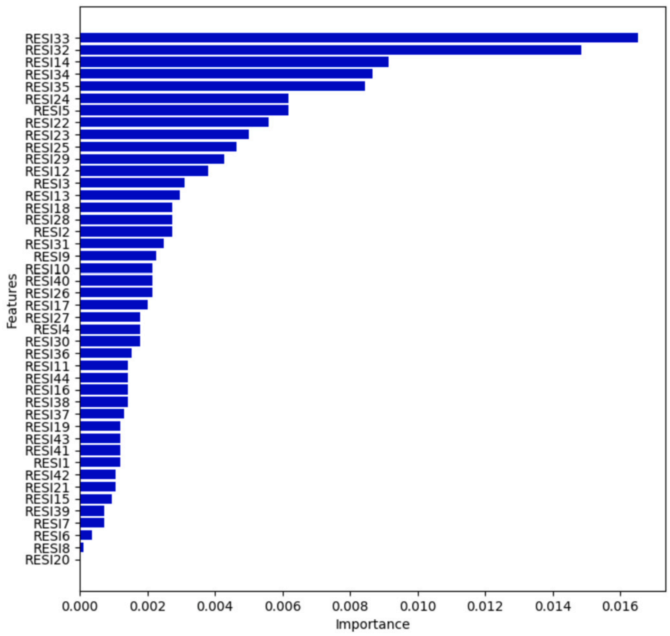


**Fig. 1.** Naive Bayes Feature Importance. RESIn (e.g., RESI1, RESI6, RESI44) corresponds to item *n* of the RED questionnaire.

knowledge, which is a person's understanding and awareness of themselves, including their beliefs, values, emotions, strengths, weaknesses and motivations, and in this case, understanding of their past experiences and how these have influenced their behaviour and life choices.

RESI14 highlights the importance of self-reflection and awareness in determining the likelihood of developing symptoms. Individuals who clearly understand the impact of their past on their current behaviour are more likely to show a conscious awareness of their actions and possible consequences. This self-perception and awareness play a crucial role in predicting whether an individual will experience symptoms in the future.

#### *4.5.2. When predicting the level of recovery*

The World Health Organization Quality of Life (WHOQOL) questionnaire emerges as one of the most important questionnaires for predicting the recovery level from Eating Disorder symptoms. (See Fig. 2.) That can be attributed to the strong association between an individual's recovery and perceived quality of life. The WHOQOL questionnaire assesses various dimensions of an individual's well-being, including physical health, psychological well-being, social relationships, and environmental factors.

When individuals are on their path to recovery, their perception of their quality of life plays a significant role. The questionnaire captures their subjective evaluation of different aspects of their life, providing insights into their overall satisfaction and contentment. People who perceive a higher quality of life may exhibit greater motivation, resilience, and engagement in recovery-related activities, leading to improved recovery outcomes. Conversely, individuals with a lower perceived quality of life may face additional challenges and may require additional support and interventions to enhance their recovery.

On the other hand, the EAT questionnaire emerges as the second most important questionnaire. It is designed to assess an individual's attitudes, beliefs, and behaviours related to eating and body image. It explores several dimensions, such as food restriction, bulimia, and oral control, providing information about the individual's relationship with food and their body.

The EAT questionnaire's relevance for predicting the recovery level lies in its ability to capture the severity and presence of ED symptoms. Individuals who score higher on the EAT questionnaire may exhibit more

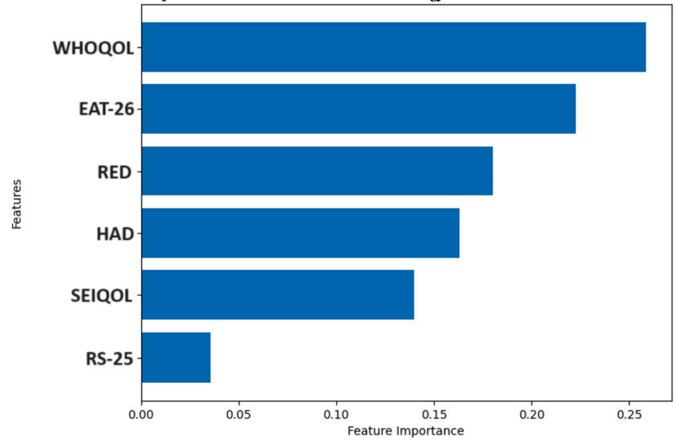


**Fig. 2.** SVR Feature Importance.

pronounced eating-disordered behaviours and beliefs, which may impact their recovery process. By assessing these attitudes and behaviours, the EAT questionnaire provides valuable information about an individual's struggle with EDs and their progress towards recovery.

Multiple dimensions of a positive self-concept, including self-esteem, self-acceptance, and self-compassion, have been shown to enhance treatment outcomes and alleviate eating disorder (ED) symptoms. Building on Bardone-Cone et al. [47], who identified self-concept, personality traits, and negative affect as critical predictors of ED recovery, we emphasise the consistent predictive value of self-esteem and self-efficacy—both tied to perceived worth and competence—across various ED diagnoses (e.g., anorexia nervosa, bulimia nervosa). Conversely, perfectionism can impede recovery, whereas higher self-efficacy and lower impulsivity under distress appear to facilitate positive treatment trajectories.

In clinical practice [48,49], ED treatment often requires an integrative approach, combining psychotherapeutic interventions (e.g., Cognitive-Behavioural Therapy, Cognitive-Behavioural Therapy-Enhanced, Family-Based Treatment (FBT) / Maudsley Approach, Dialectical Behaviour Therapy, Interpersonal Psychotherapy), medical oversight, nutritional counselling, and, when necessary, pharmacotherapy. Resilience, self-acceptance, and quality of life are systematically incorporated into these modalities by:

1. Increasing awareness of personal values and fostering hope.
2. Encouraging social support and accountability.
3. Building coping skills that enhance emotional regulation and problem-solving.
4. Promoting self-knowledge and self-acceptance.
5. Emphasising meaningful goals and overall well-being, rather than exclusively focusing on weight or shape.

Through these strategies, patients experience not only a reduction in disordered eating behaviours but also a fortified capacity to manage life challenges, ultimately improving both immediate and long-term recovery outcomes.

### *4.6. Limitations*

Despite its valuable contributions, the study in question has limitations that require careful consideration when interpreting its results. One of the most prominent limitations lies in the relatively small sample size used in the research. This restriction in sample size may have implications for the generalisability of the findings on a broader scale, as the characteristics and dynamics present in larger samples may vary considerably. Consequently, the importance of exercising caution in extrapolating the findings of this study to larger populations is stressed, and future research with larger sample sizes could provide a more robust basis for more robust and reliable generalisations.

Another important limitation relates to the unique geographic background of the participants, who were concentrated in a specific geographic region. This restricted geographical focus may limit the applicability of the results to other geographical areas and cultural contexts, as the diversity of experiences and demographic characteristics may vary substantially between locations. In this regard, it is important to consider the geographical context when interpreting and applying the results of this study, recognising the possible variations that may arise in different settings.

Additionally, it is relevant to note that over 90% of the participants were women. While this gender bias may raise concerns about representativeness in the study, it is imperative to recognise that, in the field of EDs, the prevalence of EDs in women exceeds 90%. Consequently, in this particular context, gender bias may not significantly impact the understanding of EDs. However, this limitation is urged to be considered when interpreting the results and applying them to more gender-diverse populations, recognising the importance of addressing representativeness comprehensively for more robust and applicable conclusions.

Furthermore, another limitation of this study is the lack of data on the educational level of participants. Educational background can be a relevant factor influencing health-related behaviours, access to treatment, and overall understanding of health conditions. Indeed, previous research has suggested that higher educational attainment may correlate with improved recovery in individuals with EDs [50]. The absence of this information prevents a deeper analysis of potential associations between educational level and the study variables. Future research should aim to incorporate this variable to provide a more comprehensive understanding of its potential impact.

## 5. Conclusions

In the current context of mental health, EDs represent a significant challenge for health professionals and scientists. The complexity of these conditions demands a multidisciplinary approach that combines clinical expertise with the latest innovations in technology and data analysis. In this context, research into identifying key factors for recovery and evaluating diagnostic and monitoring tools has become essential.

In our research, we have explored the role of questionnaires as information-gathering tools, finding that they go beyond their conventional function by suggesting potential predictive value. In addition to capturing relevant data, we have identified indications that these instruments could play a crucial role as predictors, providing advanced insight into possible future outcomes and trends.

This finding expands the perspective on questionnaires' versatility, highlighting their ability to offer both retrospective and prospective insights. Our research has deepened our understanding of how these questionnaires could act as leading indicators, thus contributing to informed decision-making and anticipating relevant developments.

In addition, comparative analysis of the predictive ability of existing questionnaires has revealed essential information for health professionals. Identifying which questionnaires are most effective in assessing patient progression improves the efficiency of diagnostic processes. It allows for early interventions tailored to patients' needs, which can make a significant difference in their lives.

This study has identified relevant factors that have provided a deeper understanding of the elements that influence the recovery process of patients with EDs. This information is essential for formulating personalised intervention strategies and represents a substantial advance in the scientific understanding of these complex conditions.

In terms of future lines, this study suggests several promising areas. Applying more advanced machine learning techniques, such as deep learning, could offer new insights and further increase the accuracy of predictions. Furthermore, exploring how environmental and cultural factors impact patients' recovery could open up new avenues for research, providing a more complete and holistic understanding of these disorders. Additionally, an area of study that emerges as fundamental is the application of Natural Lenguaje Processing (NLP) analysis to examine and understand responses to open-ended questionnaire questions. That would allow a deeper understanding of patients' experiences and perspectives, contributing to improved treatment and support strategies.

## CRediT authorship contribution statement

**A. Pikatza-Huerga:** Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation. **C. Las Hayas:** Writing – review & editing, Supervision, Resources, Investigation. **U. Zulaika:** Writing – review & editing, Supervision, Methodology. **A. Almeida:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Conceptualization.

## Declaration of competing interest

The authors declare that they have no conflicts of interest to disclose. They affirm that the research was conducted without any commercial or financial relationships that could be perceived as influencing the outcomes or interpretations of the study. All findings and conclusions presented in this work are based solely on the research data and analysis.

## Acknowledgement

The authors thank the Deustek5 research team at the University of Deusto for their invaluable support.

## References

[1] Härtl G. The world health report 2001: mental disorders affect one in four people — who.int. https://www.who.int/news/item/28-09-2001-the-world-health-report-2001-mental-disorders-affect-one-in-four-people, 2021. [Accessed 3 April 2025].
[2] Hay P, Girosi F, Mond J. Prevalence and sociodemographic correlates of dsm-5 eating disorders in the Australian population. J Eat Disord 2015;3. https://doi.org/10.1186/s40337-015-0056-0.
[3] Iwajomo T, Bondy SJ, de Oliveira C, Colton P, Trottier K, Kurdyak P. Excess mortality associated with eating disorders: population-based cohort study. Br J Psychiatry 2020;219:487–93. https://doi.org/10.1192/bjp.2020.197.
[4] Sancho C, Arija MV, Asorey O, Canals J. Epidemiology of eating disorders. Eur Child Adolesc Psychiatry 2007;16:495–504. https://doi.org/10.1007/s00787-007-0625-0.
[5] Gualandi M. Medical complications in eating disorders. Berlin Heidelberg: Springer; 2012. p. 17–30.
[6] Steinhausen H-C. The outcome of anorexia nervosa in the 20th century. Am J Psychiatr 2002;159:1284–93. https://doi.org/10.1176/appi.ajp.159.8.1284.
[7] Harris C, Barraclough B. Excess mortality of mental disorder. Br J Psychiatry 1998;173:11–53. https://doi.org/10.1192/bjp.173.1.11.
[8] Tortorella A, Brambilla F, Fabrazzo M, Volpe U, Monteleone AM, Mastromo D, et al. Central and peripheral peptides regulating eating behaviour and energy homeostasis in anorexia nervosa and bulimia nervosa: a literature review. Eur Eat Disord Rev 2014;22:307–20. https://doi.org/10.1002/erv.2303.
[9] Sáiz PA, Gonzalez MA, Bascarán MT, Fernandez JM, Bousoño M, Bobes J. Prevalencia de trastornos de la conducta alimentaria en jóvenes de enseñanza secundaria: un estudio preliminar. Actas Esp Psiquiatr 1999;27:367–74.
[10] Haynos AF, Wang SB, Lipson S, Peterson CB, Mitchell JE, Halmi KA, et al. Machine learning enhances prediction of illness course: a longitudinal study in eating disorders. Psychol Med 2021;51:1392–402. https://doi.org/10.1017/S0033291720000227.
[11] Melisse B, Dekker J, Berg Evd, Jonge Md, Furth EFv, Peen J, et al. Comparing the effectiveness and predictors of cognitive behavioural therapy-enhanced between patients with various eating disorder diagnoses: a naturalistic study. The Cognitive Behaviour Therapist 2022. https://doi.org/10.1017/s1754470x22000174.
[12] Lammers M, Vroling M, Ouwens M, Engels R, Strien T. Predictors of outcome for cognitive behaviour therapy in binge eating disorder. Eur Eat Disord Rev 2015;23:219–28. https://doi.org/10.1002/erv.2356.
[13] Rohrbach PJ, Fokkema M, Spinhoven P, Furth EFv, Dingemans AE. Predictors and moderators of three online interventions for eating disorder symptoms in a randomized controlled trial. Int J Eat Disord 2023;56:1909–18. https://doi.org/10.1002/eat.24021.

[14] Lydecker JA, Grilo CM. Psychiatric comorbidity as predictor and moderator of binge-eating disorder treatment outcomes: an analysis of aggregated randomized controlled trials. Psychol Med 2021;52:4085–93. https://doi.org/10.1017/s0033291721001045.

[15] Wang SB. Machine learning to advance the prediction, prevention and treatment of eating disorders. Eur Eat Disord Rev 2021;29:683–91. https://doi.org/10.1002/erv.2850. https://onlinelibrary.wiley.com/doi/abs/10.1002/erv.2850.

[16] Lichtenstein MB, Haastrup L, Johansen KK, Bindzus JB, Larsen PV, Støving RK, et al. Validation of the eating disorder examination questionnaire in Danish eating disorder patients and athletes. J Clin Med 2021;10:3976. https://doi.org/10.3390/jcm10173976.

[17] Junior AA, Ferro T, Anunciação L, Landeira-Fernández J. Aspects related to body image and eating behaviors in healthy Brazilian undergraduate students. Global Journal of Educational Studies 2018;4:43. https://doi.org/10.5296/gjes.v4i1.12541.

[18] Schmidt R, Hiemisch A, Kieß W, Hilbert A. Interaction effects of child weight status and parental feeding practices on children's eating disorder symptomatology. Nutrients 2019;11:2433. https://doi.org/10.3390/nu11102433.

[19] Dingemans AE, Son GEv, Vanhaelen CB, Furth EFv. Depressive symptoms rather than executive functioning predict group cognitive behavioural therapy outcome in binge eating disorder. Eur Eat Disord Rev 2020;28:620–32. https://doi.org/10.1002/erv.2768.

[20] Krug I, Linardon J, Greenwood C, Youssef G, Treasure J, Fernandez-Aranda F, et al. A proof-of-concept study applying machine learning methods to putative risk factors for eating disorders: results from the multi-centre European project on healthy eating. Psychol Med 2023;53:2913–22. https://doi.org/10.1017/S003329172100489X.

[21] Espel-Huynh H, Zhang F, Thomas JG, Boswell JF, Thompson-Brenner H, Juarascio AS, et al. Prediction of eating disorder treatment response trajectories via machine learning does not improve performance versus a simpler regression approach. Int J Eat Disord 2021;54:1250–9. https://doi.org/10.1002/eat.23510. https://onlinelibrary.wiley.com/doi/abs/10.1002/eat.23510.

[22] López-Ibor Aliño JJ, Valdés Miyar M. DSM-IV-TR: manual diagnóstico y estadístico de los trastornos mentales, Biblioteca del DSM-IV-TR. 1ª ed. Barcelona: Elsevier Masson; 2002.

[23] Wagnild GM, Young HM. Development and psychometric evaluation of the resilience scale. J Nurs Meas 1993;1:165–78.

[24] Las Hayas C, Hjemdal O, Muñoz P-J, Padierna J-A, Beato L, Gómez-del Barrio A. Resilience in eating disorders (red-5) scale. https://doi.org/10.5281/zenodo.14234933, 2025.

[25] Las Hayas C, Padilla P, del Barrio AG, Beato-Fernandez L, Muñoz P, Gámez-Guadix M. Individualised versus standardised assessment of quality of life in eating disorders. Eur Eat Disord Rev 2015;24:147–56. https://doi.org/10.1002/erv.2411.

[26] O'Boyle C, McGee H, Hickey A, Joyce C, Browne J, O'Malley K, et al. The schedule for the evaluation of individual quality of life (SEIQoL). Administration Manual; 1993.

[27] KNNImputer — scikit-learn.org, https://scikit-learn.org/stable/modules/generated/sklearn.impute.KNNImputer.html, 2025. [Accessed 3 June 2025].

[28] Connelly L. Logistic regression. Medsurg Nurs 2020;29:353–4. https://www.proquest.com/scholarly-journals/logistic-regression/docview/2451174951/se-2. Anthony J. Jannetti, Inc. Sep/Oct 2020; Última actualización - 2020-11-11.

[29] Charbuty B, Abdulazeez A. Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends 2021;2:20–8. https://doi.org/10.38094/jastt20165. https://www.jastt.org/index.php/jasttpath/article/view/65.

[30] Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A. A comprehensive survey on support vector machine classification: applications, challenges and trends. Neurocomputing 2020;408:189–215. https://doi.org/10.1016/j.neucom.2019.10.118. https://www.sciencedirect.com/science/article/pii/S0925231220307153.

[31] Abiodun OI, Jantan A, Omolara AE, Dada KV, Umar AM, Linus OU, et al. Comprehensive review of artificial neural network applications to pattern recognition. IEEE Access 2019;7:158820–46. https://doi.org/10.1109/ACCESS.2019.2945545.

[32] Taunk K, De S, Verma S, Swetapadma A. A brief review of nearest neighbor algorithm for learning and classification. In: 2019 international conference on intelligent computing and control systems (ICCS); 2019. p. 1255–60.

[33] Schonlau M, Zou RY. The random forest algorithm for statistical learning. Stata J 2020;20:3–29. https://doi.org/10.1177/1536867X20909688.

[34] Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. Artif Intell Rev 2020;54:1937–67. https://doi.org/10.1007/s10462-020-09896-5.

[35] Wickramasinghe I, Kalutarage H. Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. Soft Comput 2020;25:2277–93. https://doi.org/10.1007/s00500-020-05297-.

[36] Maulud D, Abdulazeez AM. A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends 2020;1:140–7. https://doi.org/10.38094/jastt1457. https://jastt.org/index.php/jasttpath/article/view/57.

[37] Barath D, Cavalli L, Pollefeys M. Learning to find good models in ransac. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR); 2022. p. 15744–53.

[38] Öztaş C, Erilli NA. Contributions to Theil-Sen regression analysis parameter estimation with weighted median. Alphanumeric Journal 2021;9:259–68. https://doi.org/10.17093/alphanumeric.998384.

[39] Sun Q, Zhou W-X, Fan J. Adaptive Huber regression. J Am Stat Assoc 2020;115:254–65.

[40] Awad M, Khanna R. Support vector regression. Berkeley, CA: Apress; 2015. p. 67–80.

[41] Mitchell R, Frank E. Accelerating the xgboost algorithm using gpu computing. PeerJ Comput Sci 2017;3:e127.

[42] Las Hayas C, Padierna JA, Muñoz P, Aguirre M, Gómez del Barrio A, Beato-Fernández L, et al. Resilience in eating disorders: a qualitative study. Women & Health 2015;56:576–94. https://doi.org/10.1080/03630242.2015.1101744.

[43] Griffiths CA. Sense of coherence and mental health rehabilitation. Clin Rehabil 2009;23:72–8. https://doi.org/10.1177/0269215508095360.

[44] Bardone-Cone AM, Hunt RA, Watson HJ. An overview of conceptualizations of eating disorder recovery, recent findings, and future directions. Curr Psychiatry Rep 2018;20. https://doi.org/10.1007/s11920-018-0932-9.

[45] Fernández A, Garcia S, Herrera F, Chawla NV. Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. J Artif Intell Res 2018;61:863–905.

[46] APA Dictionary of Psychology — dictionary.apa.org, https://dictionary.apa.org/self-perception, 2023. [Accessed 3 April 2025].

[47] Bardone-Cone AM, Miller AJ, Thompson KA, Walsh EC. Predicting a comprehensive operationalization of eating disorder recovery: examining <scp>self-concept</scp>, personality, and negative affect. Int J Eat Disord 2020;53:987–96. https://doi.org/10.1002/eat.23281.

[48] Overview | Eating disorders: recognition and treatment | Guidance | NICE — nice.org.uk. https://www.nice.org.uk/guidance/ng69. [Accessed 27 March 2025].

[49] The Oxford Handbook of Child and Adolescent Eating Disorders: Developmental Perspectives. Oxford University Press; 2011.

[50] Keski-Rahkonen A, Tozzi F. The process of recovery in eating disorder sufferers' own words: an Internet-based study. Int J Eat Disord 2005;37:S80–6. https://doi.org/10.1002/eat.20123.