scientific reports

Check for updates

OPEN

Gene–environment–brain topology reveals clinical subtypes of depression in UK Biobank

Emma Tassi[1,2], Alessandro Pigoni[1], Nunzio Turtulici[1], Federica Colombo[3,4], Lidia Fortaner-Uyà[3,4], Anna Maria Bianchi[2,7], Francesco Benedetti[3], Chiara Fabbri[5], Benedetta Vai[3,4], Paolo Brambilla[1,6]✉ & Eleonora Maggioni[1,2]

Major depressive disorder (MDD) is a leading cause of disability worldwide, affecting over 300 million people and posing a significant burden on healthcare systems. The heterogeneity of MDD can be attributed to diverse etiologic mechanisms. Characterizing MDD subtypes with distinct clinical manifestations could improve patient care through targeted personalized interventions. Topological Data Analysis (TDA) has emerged as a promising tool for identifying homogeneous subgroups of diverse medical conditions and key disease markers. Our study applied TDA to data from a UK Biobank MDD subcohort comprising 3052 samples, leveraging genetic, environmental, and neuroimaging data to assess their differential capability in predicting clinical outcomes in MDD. TDA graphs were built from unimodal and multimodal feature sets and quantitatively compared based on their capability to predict depression severity, physical comorbidities, and treatment response outcomes. Our findings showed a key role of the environment in determining the severity of depressive symptoms. Comorbid medical conditions of MDD were best predicted by brain imaging characteristics, while brain functional patterns resulted in the best predictors of the treatment response profiles. Our results suggest that considering genetic, environmental, and brain characteristics is essential to characterize the heterogeneity of MDD, providing avenues for the definition of robust markers of health outcomes in MDD.

Keywords Major depressive disorder, Genetic, Environment, Brain imaging, Topological data analysis

Major depressive disorder (MDD) is one of the most prevalent mental illnesses and among the leading causes of disability worldwide, affecting over 300 million people and representing a major burden in health systems[1–3]. MDD is a highly heterogeneous disorder, not only phenotypically (e.g., symptoms and response to treatment), but also in terms of etiologic mechanisms, such as genetic, neural substrates, societal and environmental factors[4,5]. The incomplete understanding of the determinants of this heterogeneity, especially in biological terms, limits drug discovery and treatment personalization for patients with MDD[6,7]. Studies utilizing depressive symptoms to characterize MDD subtypes have generally identified up to five clinical subtypes[8,9]. However, even with encouraging recent efforts to leverage multiple data sources (e.g., symptomatology, genetic, biochemical, and neuroimaging data) for identifying MDD subtypes, these findings have shown substantial heterogeneity and inconsistency, often facing limitations due to narrow clinical relevance, insufficient validation, and a lack of replication. One crucial example is the mixed results of Randomized Controlled Trials (RCTs) with antinflammatory drugs in depression, because only a subgroup shows increased inflammation and can benefit from this treatment. The inability to adequately address this heterogeneity, particularly from a biological standpoint, hinders both drug discovery and the development of personalized approaches for patients with MDD[10,11].

Regarding treatment, no single drug for depression is universally effective and sequential antidepressant trials are often needed, with about one third of patients not showing sufficient symptom relief after the first trial, and about 15% of patients still symptomatic even after multiple trials[7,12]. This trial-and-error approach

[1]Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. [2]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy. [3]Division of Neuroscience, Unit of Psychiatry and Clinical Psychobiology, IRCCS Ospedale San Raffaele, Milan, Italy. [4]University Vita-Salute San Raffaele, Milano, Italy. [5]Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy. [6]Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy. [7]Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, Milan, Italy. ✉email: paolo.brambilla1@unimi.it

increases the time needed to reach remission, therefore greatly increasing the overall burden of the disorder[13]. Treatment-resistant depression (TRD), a condition associated with significant disability and socioeconomic burden[14,15], is defined as a lack of response to at least two antidepressants with appropriate duration and dose. The identification and characterization of homogeneous subtypes of MDD, with distinct patterns of clinical manifestations and outcome trajectories, would be beneficial for the development of personalized treatments and patient care, improving outcomes and quality of life[6,7,16]. The healthcare digital revolution can make this ambitious goal feasible, thanks to the increasing availability of detailed clinical and biological data in large population cohorts, which can be used for stratifying patients and developing predictive models of mental health outcomes. The UK Biobank (UKB) dataset is an example of this type of data, including information on a large cohort of about half a million participants[17]. At the same time, new analysis techniques driven by artificial intelligence (AI) and machine learning (ML) have emerged for handling high-dimensional clinical and biological data, to identify relevant biomarkers of complex disorders such as MDD, where contributions from both genetics and environmental factors interact to determine the disease and its outcomes[18]. In this framework, several unsupervised ML methods were employed to stratify MDD patients based on clinical[19], biological[20] , and imaging data[21,22]. However, the classical clustering algorithms, e.g., k-means clustering, present some limitations. Indeed, these data-driven ML algorithms would fail to stratify patients and thus extract relevant subgroups if the structure of the data is large, complex, and multidimensional[21,23].

In this context, new techniques such as the data-analytics tool of Topological Data Analysis (TDA) emerge as a valid option. TDA is a multidimensional data analysis tool, based on data topology and geometry, which can reduce such high-dimensional data to simple and compact geometric structure (e.g., graphs). These structures, named topological skeleton, represent a simple topological summary of the data from which fundamental characteristics of a dataset can be directly captured[24,25]. TDA is data-driven and has the advantage of being able to capture the topological and geometrical structure of complex and large datasets, and it is stable concerning perturbations and noise in the input multidimensional dataset[26].

Unlike Principal Component Analysis (PCA) or traditional clustering methods, TDA is inherently coordinate-free, as it depends only on distance functions rather than coordinate systems, though the distance metric defines the shape of the resulting topological structure. This approach provides enhanced robustness to noise and perturbations in the input multidimensional dataset while being able to capture the topological and geometrical structure of complex, heterogeneous datasets that traditional clustering methods struggle to process effectively. While dimensionality reduction methods such as PCA and t-SNE primarily capture local structures, TDA detects data structure at multiple scales simultaneously, providing a comprehensive framework incorporating both local and global information, overcoming traditional limitations where methods are either globally accurate but locally unreliable or locally sensitive but globally inadequate[27]. TDA exploits higher-order interactions among multimodal phenotypes, generating novel topological features for ML analysis, providing outputs for rapid and intuitive exploration of the dataset structure. Thus, the TDA approach is sensitive to large- and small-scale patterns that conventional methods like PCA, Multi-dimensional scaling, and traditional clustering often miss, revealing hidden geometric properties and robust subgroups[27,28].

TDA can successfully exploit higher order (more than pairwise) interactions among phenotypes, which are expected in genetic, environmental, clinical, and brain imaging data, and provides outputs for rapid and intuitive exploration of the dataset structure, allowing for investigate how features converge or diverge in the defined clusters[26]. Recently, TDA has been successfully employed in different medicine fields[24,29] to navigate multimodal and high-dimensional biological datasets, including functional brain connectivity[30] and biomolecular structure[31], e.g., for the data-driven investigation of neuropsychiatric disorders such as attention-deficit/hyperactivity disorder[23] employing functional connectome and delirium[26] using electroencephalograph signals. Beyond its applications to previously mentioned fields, TDA has proven to be an efficient tool for augmenting and enhancing classical ML and deep learning methods defined as "topological machine learning"[32]. Of note, the joint application of TDA and Spatial Analysis of Functional Enrichment (SAFE) score allowed for a multidimensional comparative analysis, enabling a quantitative characterization of MDD clinical outcomes stratification patterns based on different multimodal feature sets[33,34]. Specifically, SAFE score analysis has been applied to quantitatively examine the local enrichment created on the graph built on candidate predictive features by several functional outcomes[33]. SAFE score was demonstrated to provide a quantitative measure (e.g., through permutation) of statistical association between the network organization and selected outcomes[33], in contrast to most other methods, which mainly focus on the qualitative identification of subgroups within the TDA-based network by the visual inspection of the color-coded network obtained by mapping each node's outcome[19,32,35].

This quantitative approach represents a critical methodological advancement for clinical research, eliminating subjective interpretation bias while providing statistically robust predictive relationships between outcomes and feature sets. The SAFE score framework enables direct comparison of predictive performance across different multimodal feature combinations, facilitating evidence-based biomarker selection. The local enrichment detection capability identifies specific network regions with significant feature-outcome associations, creating a fingerprint for quantifying the predictive strength of different feature combinations in relation to specific outcomes. This improves upon traditional clustering approaches that obscure local feature-outcome links[27,32], making SAFE-enhanced TDA valuable for precision medicine applications requiring understanding of both global topology and local enrichment patterns.

In this framework, the multidimensional tool of TDA could drive a more precise stratification of MDD, leading to the identification of key biomarkers related to the disease's trajectories, such as TRD. For the first time, our study aimed to apply the novel multidimensional tool of TDA to assess the differential predictive capability of multimodal feature sets for various clinical outcomes in a large cohort of MDD from the UKB.

A robust TDA pipeline was implemented and applied to genetic, environmental, and brain magnetic resonance imaging (MRI) features in UKB participants with MDD. TDA graphs built on each feature type and

on their combinations were compared in terms of their capability to predict different clinical characteristics, measured as their degree of topological resemblance to the outcomes of interest. Based on prior methodological and clinical knowledge, we hypothesized that our TDA pipeline could effectively: (1) unveil intricate associations within complex and high-dimensional data from multiple sources, (2) provide new insights on MDD as a heterogeneous disorder, specifically by identifying which multivariate feature combinations best predict different clinical outcomes.

Materials and methods

The data analysis workflow (Fig. 1) was applied to three outcome groups using multimodal imaging and gene-environmental feature sets as TDA input to assess their capability to stratify MDD individuals based on homogeneous clinical outcomes. Multimodal MRI predictors included T1-weighted structural MRI (sMRI), diffusion-weighted MRI (dMRI), and resting-state (rs-fMRI) and task-based (t-fMRI) functional MRI data. Subsampling (B = 100, bootstrap sampling without replacement) subdivided subjects of each outcome group into 100 subsamples, creating 1000 graphs (100 subsamples × 10 feature sets). SAFE enriched scores were computed and compared across feature sets to identify the best set for each outcome group, which underwent additional TDA analysis on the entire sample to extract top-ranked features and assess their stability using the same subsampling framework. Methodological steps included: (1) TDA Mapper application on subsampled datasets, (2) SAFE score estimation and statistical analysis over subsampled datasets, (3) TDA Mapper application on entire sample using best feature set, including (a) edge-level (EL) metric estimation for the selected feature set and outcome, (b) PCA dimensionality reduction to the EL metric of the best predictive feature set, (c) Linear regression to predict EL metric of each outcome from Principal Components (PCs), identifying the best predictive PC, and (d) feature ranking within the best predictive feature set based on PC weights of the best predictive PC, (4) stability analysis of top-ranked features across subsampled datasets.

Participants

The UKB is a population-based cohort from the United Kingdom, including ~ 500,000 individuals. UKB has collected longitudinal environmental, lifestyle, activity, genetic, multimodal neuroimaging data, and other

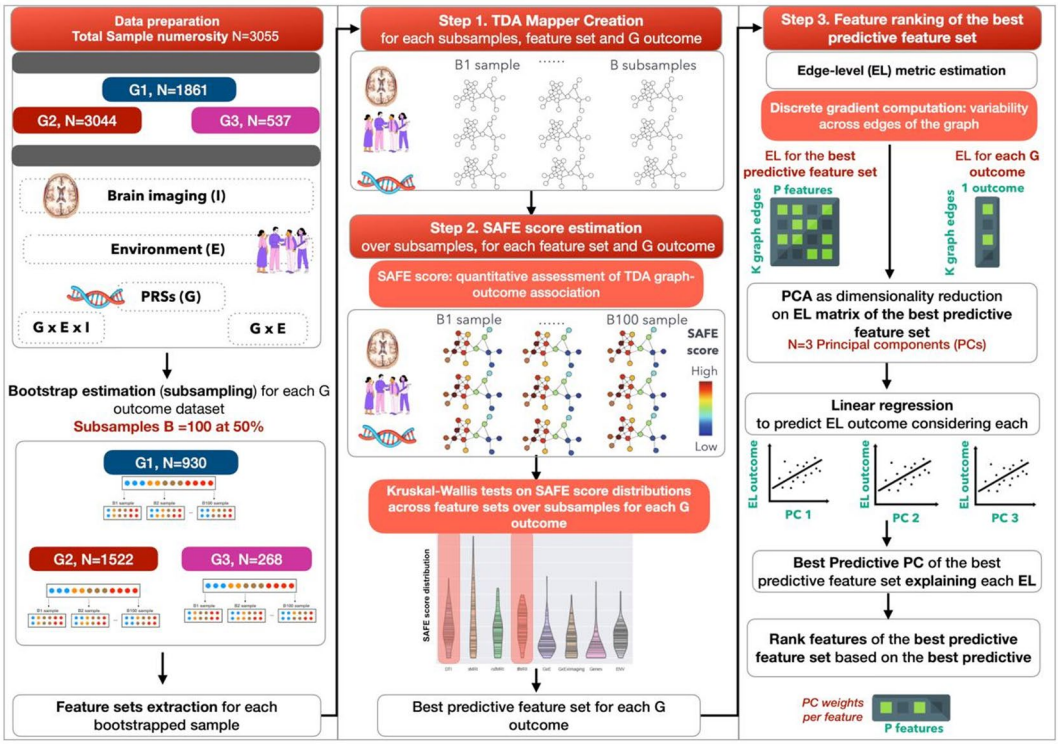

Fig. 1. Data analysis framework. The steps include (from the left to the right panel): Data preparation, comprising division of the MDD individuals in three groups of health-related outcomes and feature sets extraction; Subsampled dataset estimation, subdividing the feature set of each outcome's group into 100 subsamples; Feature sets extraction for the 100 subsamples of each outcome's group; TDA Mapper application (step 1, Methods), creating 100 graphs for the 100 subsamples for the feature set of each outcome's group; SAFE score estimation and statistical analysis (step 2, Methods) to extract the best predictive feature set for each outcome's group; Feature ranking (step 3, Methods) comprising: the estimation of the EL metric, PCA application to the EL metric of the best predictive feature set, linear regression model application for the prediction of outcome's EL metric to identify the best predictive PC, and ranking of the best predictive features based on the best predictive PC's weights.

biomarkers, as well as health-related information[17]. We included in this study participants with a diagnosis of MDD according to any measure/assessments available. In detail, we considered: (1) diagnosis of a depressive disorder according to primary care records, considering diagnostic codes previously described (at least one diagnostic code for depression, n = 37,394)[36]; (2) MDD defined by the Composite International Diagnostic Interview Short Form (CIDI-SF) (n = 34,870), which was part of the Mental Health Questionnaire (MHQ)[37]; (3) hospital diagnoses (ICD-10 codes F32-F33) (n = 17,271); (4) Smith et al. definition (n = 30,876)[38] . Participants with a diagnosis of a bipolar, psychotic, or substance use disorder according to primary care records, hospital ICD-10 codes, and/or MHQ data were excluded, as described also in[36]. In total, we included 95,741 individuals with lifetime MDD according to one or more of the described criteria. Among these individuals, 3052 participants were selected based on the availability of socio-demographic, genetic, environmental, and multimodal imaging information. From this group of participants, different subsets were extracted based on the availability of information on health-related outcomes of interest. Specifically, three groups of health-related outcomes were considered, relative to depression severity (n = 1861), cardio-metabolic and medical conditions (n = 3044), and TRD (n = 537). The reported sample sizes represent subjects with complete data for all outcomes within their respective domains.

Measures

Brain imaging predictors

Multimodal MRI information available from the UKB was extracted, considered as a candidate predictor of MDD characteristics, and included as input for the TDA. Pre-processing steps and quality control analyses described in[39] were applied to extract the following features. The dataset included T1-weighted sMRI, dMRI, and rs-fMRI and t-fMRI data relative to an implicit emotion processing task[40]. The variables and numerosity included in each of these multimodal MRI sets are listed in Table 1. Specifically, sMRI features (P = 200) included 14 subcortical volumes (SubV), 62 cortical gray matter volumes (GMV), 62 cortical thickness (CT), and 62 cortical surface areas (SA) extracted from the Desikan-Killiany atlas regions of interest (ROIs). The dMRI features (P = 384) included fractional anisotropy (FA), mean diffusivity (MD), orientation dispersion index (OD), Intracellular Volume Fraction (ICVF), Isotropic Compartment Volume Fraction (ISOVF), axial diffusivity (AD), radial diffusivity (RD), and L1 direction extracted from the 48 tract ROIs defined using the Johns Hopkins University tract atlas.

The rs-fMRI features (P = 110) consisted of node-level summary measures: the strengths of positive and negative weights extracted from the 55 × 55 adjacency functional connectivity (FC) matrix among pairs of non-artificial group-level spatial ICs. Specifically, functional features of rs-fMRI resulted in 110 total features (55 positive and 55 negative strength measures). Each of the 55 ICs obtained from group ICA was grouped in nine resting-state functional networks (RSNs) reported in Table S1, as in[41]. At last, P = 8 summary measures of t-fMRI activations were extracted, including the 90th percentile of BOLD effect (4 measures) and of the z-statistic (4 measures) for faces-shapes contrast, faces activation, and shapes activation in group-defined mask, and for faces-shapes contrast in group-defined amygdala activation mask. Details on the MRI data pre-processing and feature extraction are reported in the Supplementary Materials.

Genetic-environmental predictors

The following genetic-environmental information was included as input in the TDA.

Environment. Environmental features, listed in Table 1, were either divided into two feature sets (SetA and SetB) or considered together (SetC). A detailed description of each environmental variable within SetA (P = 7) and SetB (P = 27) is reported in Table S3 and Table S4. SetB covered solely the environmental factors "experienced" by participants during their lifetime, including substance use (i.e., alcohol, smoking, and cannabis), lifestyle related to dietary changes, and traumatic or stressful events. SetA included characteristics partly innate and partly resulting from lifetime events, such as personality traits, social support, telomere length (corrected for the confounding factor of total white cell counts[42], chronotype, and physical activity. SetC (P = 34) was defined as the concatenation of SetA and SetB. A detailed description of the variables included in each set is reported in Tables S3 and S4.

Genetics. Genetic features consisted of polygenic risk scores (PRSs), listed in Table 1, estimated using PRS-CS-auto[43] and calculated using the score function in PLINK 2.0[44] in participants of European ancestry (defined by 4-means clustering on the first two genetic principal components[36]. PRSs were adjusted for the first six population principal components, genotyping batch, and centre of recruitment. We considered P = 20 PRSs including both psychiatric and non-psychiatric traits (see Table S2 for a description of the GWAS summary statistics used). In detail, we estimated the PRSs of major psychiatric disorders (bipolar disorder, schizophrenia, MDD, anorexia nervosa, autism, attention deficit hyperactivity disorder, alcohol dependence), related PRSs (neuroticism, smoking, alcohol consumption, years of education), and PRSs of immune-cardiometabolic traits (C-reactive protein, glycated hemoglobin, triglycerides, LDL and HDL cholesterol, coronary artery disease, body mass index, type 2 diabetes mellitus). The PRS of immune-cardiometabolic traits were considered given their clinical and pathogenetic correlation with MDD[45], also in relation to the inclusion of cardio-metabolic comorbidities among the outcomes of interest. Each PRS was adjusted for ancestry-relevant population principal components, genotyping batch and centre of recruitment before inclusion in the TDA.

Confounding variables

Age, sex, ethnicity, and antidepressant medication(s) were considered as confounding input features for the TDA. The variable of ethnicity was converted into a dichotomous variable (1, Caucasian; 0 Other). Socio-demographical variables, including age, sex, and ethnicity, are described in Table S6. Antidepressant medications

	Dimension of the dataset	Overview of the variables included
Environmental Set A	P = 7	Frequency of friend / family visits Leisure / social activities Able to confide Morning/evening person (chronotype) Summed MET minutes per week for all activity Z adjusted T/S log (i.e., Telomeres length) Neuroticism-EPQ RS
Environmental Set B	P = 26	Felt love as a child Physically abused by family as a child Felt hated by family member as a child Sexually molested as a child Someone to take to doctor when needed as a child Been in a confiding relationship as an adult Physical violence by partner or ex-partner as an adult Belittlement by partner or ex-partner as an adult Sexual interference by partner or ex-partner without consent as an adult Able to pay rent/mortgage as an adult Victim of sexual assault Victim of physically violent crime Been in serious accident believed to be life-threatening Witnessed sudden violent death Diagnosed with life-threatening illness Been involved in combat or exposed to war-zone Alcohol intake frequency Ever Taken Cannabis Childhood stressful events Adulthood stress Serious illness, injury, assault to yourself or assault of a close relative in the last 2 years Death of a close relative or death of a spouse or partner in the last 2 years Stress related to marital separation/divorce or to financial difficulties in the last 2 years Smoking status Major dietary changes because of illness in the last 5 years Major dietary changes because other reason in the last 5 years
Environmental Set C	P = 33	Environmental Set A + Environmental Set B
Imaging	Total, P = 606 sMRI, P = 200 dMRI, P = 288 rs-fMRI, P = 110 t-fMRI, P = 8	sMRI included 4 types of variable: 14 Subcortical Volumes, 62 GMV (Desikan-Killiany atlas), 62 CT (Desikan-Killiany atlas), 62 SA (Desikan-Killiany atlas) dMRI included 6 types of variable: 48 FA ROIs, 48 ODI, 48 ICVF, 48 ISOVF, 48 RD, L1 directions, all from JHU atlas rs-fMRI included: 55 strength of positive weights, 55 strength of negative weights t-fMRI included: 4 measures of 90th percentile of BOLD effect, 4 measures z-statistics for faces-shapes contrast, faces activation, and shapes activation in group-defined mask, and for faces-shapes contrast in group-defined amygdala activation mask
Genes (PRSs)	P = 20	Bipolar disorder Schizophrenia MDD Anorexia nervosa Autism Attention deficit hyperactivity disorder Alcohol dependence Neuroticism Smoking Alcohol consumption Years of education College PRS C-reactive protein Glycated hemoglobin Triglycerides LDL cholesterol HDL cholesterol Coronary artery disease Body mass index Type 2 diabetes mellitus

Table 1. Numerosity and specific variables included in each feature sets. MET, metabolic equivalent task; EPQ- RS, Eysenck Personality Questionnaire—Revised; PRS, Polygenetic risk scores; sMRI, structural MRI; rs-fMRI, resting-state functional MRI; t-fMRI, task-based functional MRI; dMRI, diffusion weighted MRI; JHU atlas, Johns Hopkins University atlas; GMV, gray matter volume; CT, cortical thickness; SA, surface area; FA, Fractional Anisotropy; RD, radial diffusivity; ODI, orientation dispersion index; ICVF, Intracellular Volume Fraction; ISOVF, Isotropic Compartment Volume Fraction; L1, L1 directions.

are described in Table S7. A dichotomous treatment indicator was extracted (i.e., 1 for participants taking any antidepressant medications, 0 for the other participants).

Health-related outcomes

Information on physical and mental health was used to characterize MDD subgroups in terms of clinically relevant outcomes, which were divided into three groups, as reported in Table 2: (i) variables related to severity of depression (i.e., type and duration of depression, self-harm behaviors, as well as depression with anxious and neurovegetative symptoms defined as in[46,47])(**G1**); (ii) variables related to cardio-metabolic and general health conditions (i.e., cancer, type 2 diabetes, cardiovascular diseases) (**G2**); (iii) TRD, defined as having at least two

switches between different antidepressant drugs (independently on the class) (**G3**), as detailed in a previous work[36]. A detailed description of each outcome variable is reported in Table S5.

Data analysis

Multimodal imaging and gene-environment feature sets were employed as inputs for the TDA to assess their differential capability to cluster MDD individuals into subgroups with homogeneous clinical outcomes. We used TDA Mapper, provided by the *tmap* library[34]. Most previous TDA applications assessed graph-outcome associations based on a qualitative basis[29,32,35]. On the other hand, the SAFE score allows a quantitative estimation, previously described in the applications of Liao and colleagues[34] and Baryshnikova and colleagues[33]. The SAFE score maps the values of a target variable of interest onto the Mapper graph and extracts significant association patterns, called "subgraphs enrichment", for the specific target variable. This score represents the only quantitative metric used in the literature for assessing the outcome predictive capability of a TDA graph[33,34].

TDA mapper application: step 1

Mapper algorithm overview. The Mapper algorithm transforms high-dimensional data into a simplified topological graph through systematic processing of input 2D matrices (subjects × features)[32]. The algorithm comprises: (1) Standardization of input matrix ensuring equal feature contribution to distance computation; (2) Distance matrix computation calculating pairwise distances across samples; (3) Filtering applying dimensionality reduction for lower-dimensional embedding (dimension D); (4) Covering dividing data into D-overlapping subsets; (5) Pullback identifying original data points within subsets; (6) Clustering data points within each subset to identify distinct groups; (7) Output generating a graph where nodes represent subject clusters and connections indicate shared data points between clusters.

Application to UKB sample. Different graphs were built by separately employing ten feature sets: (i) sMRI, (ii) dMRI, (iii) t-fMRI; (iv) rs-fMRI; (v) genetic (G); environmental (SetA, SetB, SetC); (ix) genetic-environmental (G–E); (x) genetic-environmental-imaging (G-E-Imaging). Confounding features were added to all sets, to estimate their predictive capability while accounting for these confounding variables. For each feature set, three graphs were built for the three outcomes' groups identified before (G1 = 1861, G2 = 3044, G3 = 537 subjects with complete data within their respective domains).

The subjects' space S was defined as the selected feature set. Euclidean distance was chosen as the similarity metric between data points in the subjects' space S, following established approaches for multimodal feature integration in TDA[24,27,29].

All features underwent robust scaling using median and interquartile range to mitigate outlier influence and ensure equal contribution across heterogeneous feature types to the Euclidean distance computation. This approach preserves underlying data structure and subject relationships better than standard normalization for multimodal biological measurements, following established TDA literature[48].

Uniform Manifold Approximation and Projection for dimensionality reduction (UMAP) with 2-dimensional embedding was applied for dimensionality reduction, chosen for its ability to preserve both local neighborhood structures and global topological properties essential for multimodal data analysis, maintaining strong mathematical foundations for high-dimensional biological data[49,50]. The mean and variance of the projected data matrix were computed to assess the variance-normalized Euclidean distance metric between projected data points.

The filtered data were divided into overlapping intervals using cover parameters of Resolution (R, number bins) and Gain (G, overlap between bins). Since we employed 2-dimensional UMAP filtering, the covering step generates overlapping rectangles for 2-dimensional binning, enabling data compression and noise reduction. Cover parameters were selected by systematically testing R-G pairs on the G-E-Imaging feature set through visual inspection of topological stability. Parameters maintaining key topological features (connected components, hub nodes, densely connected subnetworks) across variations were selected (R = 15, G = 80%) and used in all TDA subsampled runs, following established TDA practices for identifying stable configurations less likely to represent artifacts[28,51,52]. This choice captured variance across all modalities while ensuring consistent analysis across different data types and computational efficiency.

G1	G2	G3
O1: Duration of worst depression	O1: Cancer diagnosed by doctor	O1:Treatment-resistant depression
O2: Frequency of depressed days during worst episode of depression	O2: Vascular heart problems	
O3: Impact on normal roles during worst period of depression	O3: Diabetes diagnosed by doctor	
O4: Thoughts of death during worst depression		
O5: Ever self-harmed		
O6: Belief that owns life is meaningful		
O7: Depression possibly related to stressful or traumatic event		
O8: Depression with atypical neurovegetative symptoms		
O9: Depression with anxious features		

Table 2. Description of clinically relevant outcomes considered in each group. G, group clinical outcome; O, clinical outcome.

Density-based spatial clustering (DBSCAN) clustering was performed to cluster subjects within bins, with edges traced between overlapping clusters. DBSCAN was selected for its robustness to outliers, effectiveness with varying density distributions, and ability to identify clusters without requiring predetermined cluster numbers, making it suitable for TDA applications with complex multimodal data[52,53].

Subsampling (B = 100, bootstrap without replacement) was applied at 50%, selecting half of the subjects from each original group (n = 930 for G1, n = 1522 for G2, n = 268 for G3). This percentage balances computational efficiency with robust topological analysis given the high complexity of large-scale graph construction[54]. For each outcome group, 100 subsamples × 10 feature sets = 1000 graphs were constructed, totaling 3000 graphs across all groups (G1, G2, G3). Subjects were subdivided into 100 subsampled datasets, with subsamples drawn separately within each outcome group to maintain distinct clinical domains. The subsampling number balanced graph construction complexity, SAFE score estimation, and computational time. SAFE scores were calculated independently for each group, addressing sample size imbalances across clinical domains.

Graph-based outcome prediction: step 2

SAFE score. We compared the subsampled TDA graphs built on the different feature sets in terms of outcome predictive capability. A quantitative assessment of the associations between the TDA-graph and the three health-related outcome groups was assessed by extracting the SAFE scores. The SAFE score represents a systematic quantitative framework for measuring statistical associations between network topology and target outcomes within TDA applications. SAFE objectively quantifies functional organization within complex network structures by measuring continuous distributions of functional enrichment across local neighborhoods and producing maps of their relative positioning.

The SAFE algorithm operates by first defining local neighbourhoods around each network node, identifying neighbouring nodes within a maximum distance threshold [33,34]. For each neighbourhood, the target variable values across neighbour nodes are summed to generate an observed neighbourhood score, which is then statistically validated against a null distribution created through random permutation of target variables among network nodes. The enrichment score is calculated using logarithmic transformation of permutation-based p-values, yielding normalized scores ranging from 0 to 1[33]. SAFE score applied in TDA Mapper application enables obtaining a quantification of the network-outcome relationship, by detecting local enrichment patterns revealing significant spatially concentrated associations of the network's features and outcome. Higher SAFE scores indicate a stronger statistical association between a specific feature and the target outcome within the network topology, with scores approaching 1 representing the most significant enrichment patterns. Features with elevated SAFE scores demonstrate the strongest predictive relationships with the outcome variable while accounting for the complex interconnections present in the input feature set.

Statistical analysis

For each outcome group, Kruskal–Wallis (KW) tests were performed to test whether the SAFE score distributions were significantly different among the different sets of features over subsamples. If significant KW differences among feature sets were found ($p < 0.05$), post-hoc pairwise comparisons using Dunn's test were performed to identify specific differences between each pair of feature sets. Dunn's test p-values were automatically corrected for multiple comparisons across all 45 pairwise tests. From the pairwise comparison, for each outcome, we identified the best predictor as the feature set associated with the highest median value of SAFE score distribution with respect to the other significant pairwise test. Indeed, the best predictor has the highest median value of SAFE score distribution among pairs and for the highest number of pairs.

Feature ranking: step 3

Edge-level metric definition and relevance. We ranked features starting from the best predictive feature sets. For each outcome group, Mapper was applied to the best predictive feature set considering the entire sample (n = 3052).

Node-wise weighted average computation. For a target variable function f, where $f(s)$ is the target variable function value for subject s, we computed its **node-wise weighted average** f_{node}. The subject's weight reflects the total number of nodes in which the subject fell. For each node v in the graph, $f_{node}(v)$ is defined as:

$$f_{node}(v) = \frac{\sum_{s \in v} \frac{f(s)}{w_s}}{\sum_{s \in v} \frac{1}{w_s}} \tag{1}$$

where v is a node (i.e., vertex) in the graph, we consider the sums over all subjects s in a node v, with w_s is the subject's weight indicating the number of nodes that contain subject s.

Discrete gradient computation: edge-level metric. The EL metric, which corresponds to the discrete gradient on a graph[55], quantifies the variability of a selected variable across all edges of the graph.

For a pair of nodes v and w connected by an edge in the graph, the EL metric corresponds to the discrete gradient of the node-wise weighted average f_{node} across the edge. It is defined as:

$$f_{node}(vw) = |f_{node}(v) - f_{node}(w)| \tag{2}$$

The EL metric serves as a local boundary detection mechanism within the topological network, capturing local transition patterns between neighboring nodes. Edges with large EL metric values indicate transitions between nodes with substantially different values in the selected characteristic; the EL metric of an outcome identifies boundaries between distinct clinical states. When an edge connects two nodes containing subjects with markedly

different values of the clinical outcome variable, the EL metric has large absolute values, signaling a meaningful transition zone between outcome states. Conversely, edges connecting nodes with similar outcome values have small EL metrics, indicating homogeneous network regions. In this study, these transitions were analyzed across depression severity measures, cardio-metabolic comorbidities, and the TRD group of outcomes (Table 2).

PCA-based feature selection procedure. EL scores were extracted for each predictive feature and organized in a matrix of K x P (K graph edges, P features), while outcome EL scores were extracted for the outcome variables, obtaining a vector K × 1. To rank feature relevance, we fitted linear regression models using the PCs of predictor ELs (independent variables) and outcome ELs (dependent variables). This regression analysis identifies when boundaries in clinical outcome space align with feature-based boundaries in predictor space, revealing the features that contribute to the local graph topology separating different clinical states of a specific outcome.

PCA was applied to the EL score matrix of the best predictive feature set, capturing edge-related variation in all best predicting features from SAFE score graph-outcome prediction analysis. Variance of the best predictive features' EL score matrix was decomposed into the first three PCs representing highest variance directions in the data. Each PC was separately fed into a linear regression as an independent variable to predict the edge-related variation of the outcome. Inference on feature-based PC effects on outcome EL scores was made via t-statistics on beta coefficients. For each outcome, features of the best predictive feature set were ranked by the descending magnitude of PC coefficients from the best predictive PC in the linear regression model.

To evaluate the robustness and consistency of the top-ranked features, we conducted a sensitivity analysis using the same 100-iteration subsampling framework, focusing on the best predictive feature set and best predictive PC identified in the entire dataset. For each subsampling iteration, we assessed statistical significance persistence of the best predictive PC-EL outcome associations and quantified feature weight stability and ranking consistency within the selected PC.

Results

Participants characterization

After selecting MDD participants with genetic, environmental, and brain data (n = 3052) (Methods), different subsets were extracted based on the availability of information on health-related outcomes. In detail, three groups of outcomes were considered, relative to depression severity (Group 1, G1) (n = 1861), cardio-metabolic and medical conditions (Group 2, G2) (n = 3044), and TRD (Group 3, G3) (n = 537). The sample characteristics are detailed in Table 3 (considering the sample entire and split based on the availability of outcome group) and in Fig. 2.

Graph-based outcome prediction

The subsampled TDA graphs built on the different feature sets were compared in terms of outcome predictive capability. The latter was quantified through the SAFE score, representing the relevant association between input features and outcome (Methods). After multiple comparison corrections applied to KW tests, we showed differences among feature sets based on the SAFE score distributions over subsamples for all the outcomes ($p < 0.05$). Depending on the outcomes, different feature sets resulted as best predictors according to the distribution of SAFE scores, which are shown in Figs. 3 and 4. Table 4 reports the KW statistics, which were significant for all outcomes ($p < 0.05$), as well as the best predictive feature set and any pairwise differences between the latter and the other feature sets ($p < 0.05$, Dunn's post-hoc test corrected for 45 pairwise comparisons).

For each of these best-performing feature sets, we additionally report the number of nodes, edges, modularity values, and additional graph-level metrics of the Mapper graphs computed on the full sample in Supplementary Table S8. Representative examples of TDA graph visualizations for the best-performing feature sets are shown in Figure S3.

Furthermore, Fig. S1–S2 show the results of post-hoc pairwise comparisons performed for all the outcomes in the form of heatmaps of SAFE score differences among feature sets. For each outcome, the heatmap shows the significant differences in the mean ranks of SAFE score distribution between each pair of feature sets surviving multiple corrections applied on the KW test. The feature set (row) with the highest median value

	Whole dataset	G1	G2	G3
N subjects	3052	1861	3044	537
Sex[a]				
Male	1076 (35.25%)	611 (32.83%)	1074 (35.28%)	160 (29.79%)
Female	1976 (64.45%)	1250 (67.17%)	1970 (64.71%)	377 (70.20%)
Age at baseline (years)[b]	53.66 ± 7.18	53.27 ± 7.05	53.66 ± 7.19	53.63 ± 7.27
BMI (kg/m^2)[b]	26.71 ± 4.51	26.90 ± 4.68	26.71 ± 4.51	26.76 ± 4.54
WC (mm)[b]	86.91 ± 12.64	87.14 ± 12.97	86.93 ± 12.75	86.54 ± 13
Ethnicity[a]				
White	3037 (99.50%)	1851 (99.46%)	3029 (99.50%)	536 (99.81%)
Other ethnic background	15(0.50%)	10 (0.54%)	15 (0.50%)	1 (0.19%)

Table 3. Main clinical and socio-demographic characteristics of the sample. G1-G3 indicate each of the three considered groups of outcomes (see Table 2). Abbreviations: BMI, Body mass index; WC, waist circumference. [a] Data expressed as count (percentage %). [b] Data expressed as mean ± standard deviation.

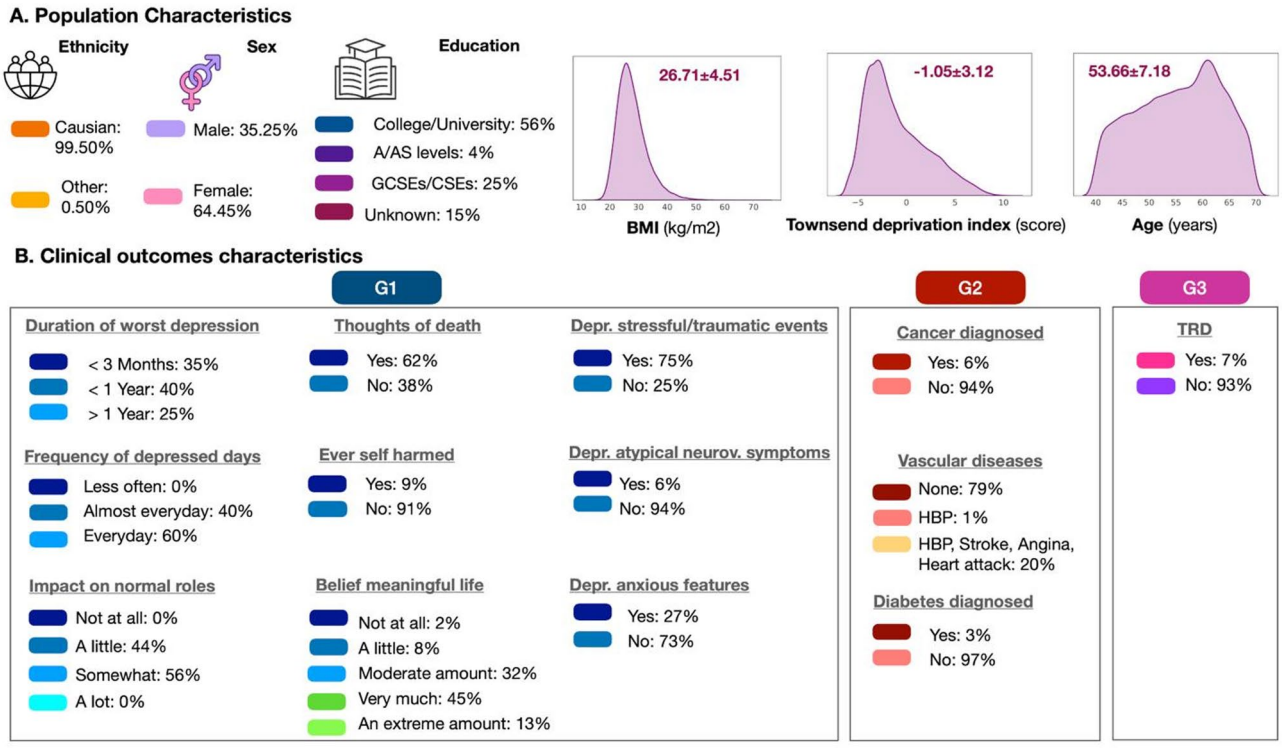

Fig. 2. Socio-demographic (**A**) and clinical outcomes (**B**) characteristics for the participants included in the study. G1, clinical outcome group n. 1; G2, clinical outcome group n. 2; G3, clinical outcome group n. 3 (see also Table 2). Note that the number of individuals included in G1-3 varies as reported in Table 3, and that percentages reported in this figure were calculated considering the number of individuals in each outcome group and not the whole sample.

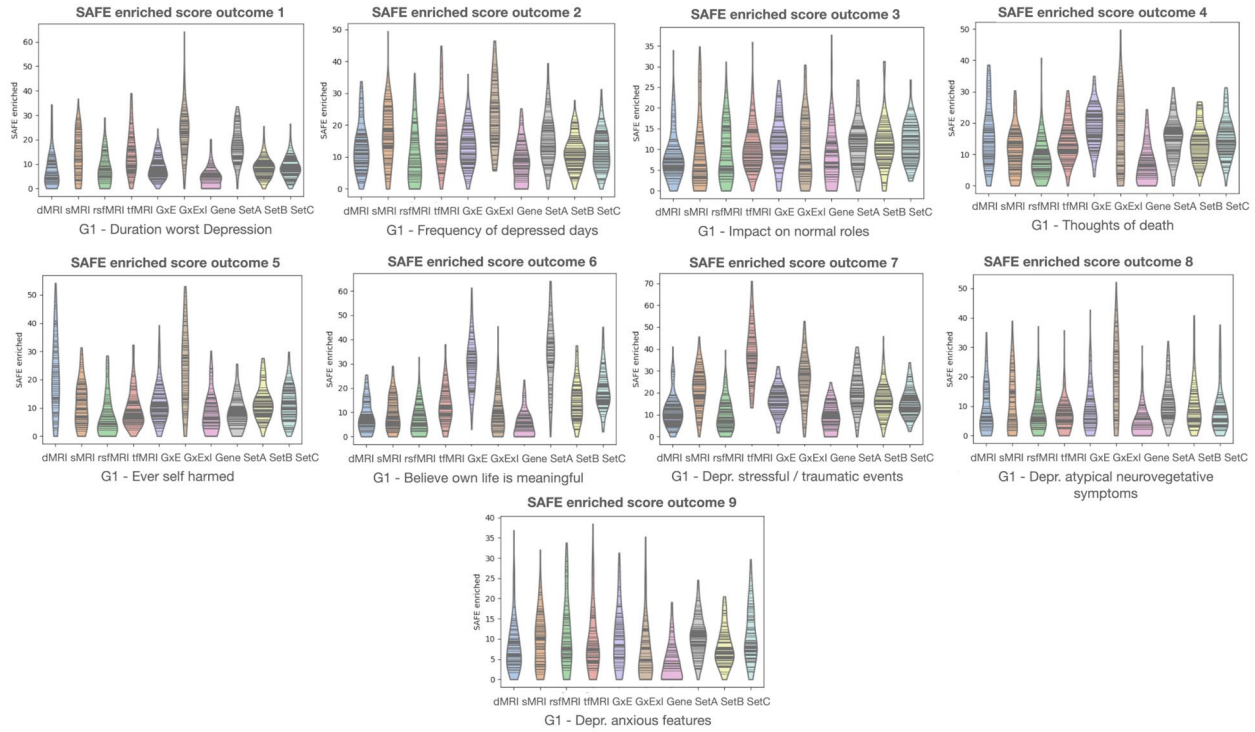

Fig. 3. Violin plots of SAFE score distributions of all feature sets for G1 outcomes.

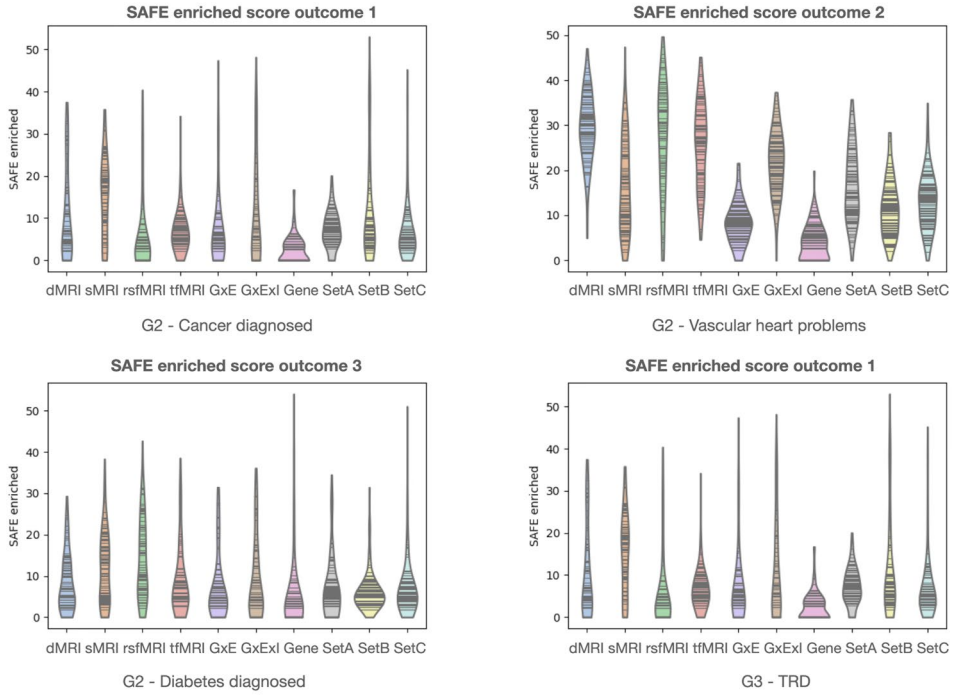

Fig. 4. Violin plots of SAFE score distributions of all feature sets for G2 and G3 outcomes.

	Kruskal–Wallis		Best predictor	
G outcomes	Statistics	p value	Median value	Dunn's test pairwise differences ($p < 0.001$)
G1 outcomes				
O1	H = 349	p = 9.38e-70	G–E-Imaging	Genes, SetB Env, SetC Env, t-fMRI, rs-fMRI, sMRI, dMRI, G–E (8/9 feature sets)
O2	H = 187	p = 1.45e-35	G–E-Imaging	Genes, SetA Env, SetB Env, SetC Env, rs-fMRI, dMRI, G–E (7/9 feature sets)
O3	H = 67	p = 5.03e-11	SetC Environment	Genes, sMRI, dMRI (3/9 feature sets)
O4	H = 209	p = 4.28e-40	G–E	Genes, SetB Env, t-fMRI, rs-fMRI, sMRI (5/9 feature sets)
O5	H = 206	p = 1.27e-39	G–E-Imaging	Genes, SetA Env, SetB Env, SetC Env, t-fMRI, rs-fMRI, sMRI, G–E (8/9 feature sets)
O6	H = 488	p = 2.11e-99	SetA Environment	G–E-Imaging, Genes, SetB Env, SetC Env, t-fMRI, rs-fMRI, sMRI, dMRI (8/9 feature sets)
O7	H = 418	p = 1.56e-84	t-fMRI	G–E-Imaging, Genes, SetA Env SetB Env, SetC Env, rs-fMRI, sMRI, dMRI, G–E (9/9 feature sets)
O8	H = 134	p = 1.11e-24	G–E-Imaging	Genes, SetB Env, SetC Env, t-fMRI rs-fMRI, sMRI, dMRI, G–E (8/9 feature sets)
O9	H = 105	p = 1.05e-18	SetA Environment	G–E-Imaging, Genes (2/9 feature sets)
G2 outcomes				
O1	H = 140	p = 9.20e-26	sMRI	Genes, SetC Env, t-fMRI, rs-fMRI, G–E, (5/9 feature sets)
O2	H = 516	p = 1.47e-105	dMRI	Gene, SetA Env, SetB Env, SetC Env, sMRI, G–E (6/9 feature sets)
O3	H = 93	p = 3.8e-16	rs-fMRI	G–E-Imaging, Gene, SetA Env, SetB Env, SetC Env, G–E (6/9 feature sets)
G3 outcomes				
O1	H = 114	p = 1.63e-20	t-fMRI	SetA Env, SetB Env, SetC Env, sMRI (4/9 feature sets)

Table 4. KW statistics reported for the best predictor obtained for each outcome and Dunn's test pairwise differences between the best predictor and the other feature sets. G, group of clinical outcome; O, clinical outcome; G, genetics; E, environment; sMRI, structural MRI; rs-fMRI, resting-state fMRI; dMRI, diffusion weighted MRI; t-fMRI, task-based fMRI.

of SAFE score was considered the best predictor for the corresponding outcome. For most outcomes in G1, feature sets including environmental variables were the best predictors (G–E-Imaging, G–E concatenation and environmental sets), while imaging-based feature sets were the best predictors for outcomes in G2 and G3. Post-hoc pairwise comparisons showed that for outcomes measuring the severity of MDD (worst depressive episodes, self-harming, and neurovegetative symptoms (O1, O2, O5, O8) the G–E-Imaging feature set emerged as the best predictor (i.e., the combination of imaging, genetic, and environmental variables). Differently, for the outcomes measuring perceived life meaning and anxious symptoms (O6 and O9), the environmental SetA (features that

were partly innate and partly affected by experience) was found as the best predictor. Environmental features (set C) reached the greatest SAFE score for the outcome referring to death thoughts (O4); further, the concatenation of these with genetic features (G–E feature set) resulted as the best predictors from SAFE score analysis for the outcome measuring the disease impact on normal roles (O3). Notably, stress-related depressive symptoms (O7) were best predicted by brain functional characteristics through the t-fMRI feature set. For G2 outcomes, reflecting cardiometabolic and general health conditions, brain characteristics were more predictive than genetic and environmental features. Specifically, post-hoc pairwise comparison analysis showed that sMRI, dMRI, and rs-fMRI feature sets resulted in the best predictors, associated with significantly higher SAFE scores than the sets composed of genetic and environmental features, for the outcomes cancer, vascular heart problems, and type 2 diabetes, respectively. Last, the t-fMRI feature set achieved the highest SAFE score for G3 TRD outcome, significantly higher with respect to the environmental and sMRI feature sets.

Feature ranking

We ranked features starting from the best predictive feature sets extracted from the SAFE score analysis, by firstly extracting the EL metric (Methods) for the best predictive feature sets and for each outcome, and as follows applying PCA to the EL metric of the best predictive feature set. Then, linear regression models were applied to extract the best predictive PC by considering as independent variables each PC to predict the EL metric of each outcome (Methods). Last, the feature ranking within the best predictive feature set was performed by considering the weights of the best predictive PC previously extracted (Methods).

The PCA results and the corresponding feature ranking for all the outcomes are summarized in Tables 5 and 6, including the best predictive PCs and related linear regression statistics for all the outcomes of interest. For all outcomes and the best PC, the top-five predicting features are reported in Table 5, sorted based on the magnitude of the PCA coefficient. Group-specific Bonferroni corrections were applied on the p-values of Table 5 (G1: $p < 0.005$, Bonferroni corrected for 9 outcomes; G2: $p < 0.01$, Bonferroni corrected for 3 outcomes; G3: $p < 0.05$, no Bonferroni correction needed for 1 outcome in G3 group). PC1 resulted as the best predictive one for all the outcomes, except for the G1 outcome relative to perceived life meaning (O6), for which PC3 was associated with the highest T-stat in the linear regression. G–E-Imaging feature set resulted in the best predictor for the majority of G1 outcomes (40%), especially the ones measuring worst depressive episodes, self-harming, and neurovegetative symptoms (O1, O2, O5, O8). Within the G–E-Imaging feature sets, the dMRI feature of mean ODI in the fornix resulted in the most important, maintaining the persistence of statistical significance of PC-EL outcome associations in linear regression models despite showing a more unstable feature ranking pattern. As follows, environmental sets (i.e., SetA and SetC) contributed to 30% of G1 outcomes, especially for the ones related to perceived life meaning, anxious symptoms, and outcomes concerning disease impact on normal roles (O3, O6, O9). Precisely, the feature telomere length from environmental SetA demonstrated exceptional robustness and ranking stability (Top3 rate = 91%), and the variable "felt hated by family members as a child" from environmental SetC resulted in the most relevant, showing moderate ranking consistency. Further, brain imaging predictors were reported as the most important feature sets for medical comorbidities outcomes,

		PC1		PC2		PC3	
G outcomes	**Best predictor**	**T-stat**	**p value**	**T-stat**	**p-value**	**T-stat**	**p value**
G1 outcomes							
O1	G–E-Imaging	T = 18.82	**p < 0.005**	T = −1.60	p = 0.1092	T = 4.07	**p < 0.005**
O2	G–E-Imaging	T = 31.37	**p < 0.005**	T = 0.32	p = 0.75	T = −3.02	**p < 0.005**
O3	SetC Environment	T = 16.66	**p < 0.005**	T = −1.00	p = 0.31	T = 6.18	**p < 0.005**
O4	G–E	T = 28.11	**p < 0.005**	T = 1.48	p = 0.14	T = 1.52	p = 0.13
O5	G–E-Imaging	T = 11.79	**p < 0.005**	T = 1.29	p = 0.19	T = −1.78	p = 0.07
O6	SetA Environment	T = 14.93	**p < 0.005**	T = −2.18	p < 0.05	T = −17.78	**p < 0.005**
O7	t-fMRI	T = 17.00	**p < 0.005**	T = −7.99	**p < 0.005**	T = 15.88	**p < 0.005**
O8	G–E-Imaging	T = 7.98	**p < 0.005**	T = −1.15	p = 0.25	T = 3.08	**p < 0.005**
O9	SetA Environment	T = 24.98	**p < 0.005**	T = −8.89	**p < 0.005**	T = −19.17	**p < 0.005**
G2 outcomes							
O1	sMRI	T = 12.49	**p < 0.005**	T = 2.03	p < 0.05	T = −0.95	p = 0.34
O2	dMRI	T = 26.74	**p < 0.005**	T = −2.59	**p < 0.005**	T = 1.90	p = 0.06
O3	rs-fMRI	T = 9.14	**p < 0.005**	T = −1.28	p = 0.20	T = −0.71	p = 0.47
G3 outcome							
O1	t-fMRI	T = 4.89	**p < 0.005**	T = −0.50	p = 0.62	T = 4.23	**p < 0.005**

Table 5. Best predictive PC, T-statistics and p values from linear regressions considering as independent variable edge-related variation of the features of the best predictor to predict the edge-related variation of the outcome. Bold for p value surviving to group-specific Bonferroni correction (corrected for 9 outcomes in G1, corrected for 3 outcomes in G2, no Bonferroni correction needed for 1 outcome in G3). G, group of clinical outcome; O, clinical outcome; PC, principal components; G, genetics; E, environment; sMRI, structural MRI; rs-fMRI, resting-state fMRI; dMRI, diffusion weighted MRI; t-fMRI, task-based fMRI.

Predictive feature set for G outcomes	Feature ranking	PC coefficient
G1 outcomes		
G–E-Imaging for O1, O2, O5, O8	**Mean OD in Fornix**	0.1477
	Mean ISOVF in left external capsule	0.0908
	Mean RD in left superior fronto-occipital fasciculus	0.0886
	Mean ISOVF in right external capsule	0.0825
	Mean ISOVF in left superior fronto-occipital fasciculus	0.0818
G–E for O4	**Belittlement by partner or ex-partner as an adult**	0.3142
	Felt hated by family member as a child	0.2800
	Someone to take to doctor when needed as a child	0.2442
	Physical violence by partner or ex-partner as an adult	0.2385
	Childhood stressful events	0.2354
SetA for O6 (PC3)	**Z adjusted T/S log (i.e., Telomeres length)**	0.4784
	Summed MET minutes per week for all activity	0.0626
	Ethnicity	0.0070
	Able to confide	−0.0296
	Frequency of friend or family visits	−0.0817
SetA for O9 (PC1)	**Z adjusted T/S log (i.e., Telomeres length)**	0.71
	Summed MET minutes per week for all activity	0.46
	Neuroticism EPQ-RS	0.35
	Age	0.32
	Sex	0.15
t-fMRI for O7	**Median z-statistic (in group-defined mask) for shapes activation**	0.3882
	90th percentile of z-statistic (in group-defined mask) for shapes activation	0.3849
	90th percentile of z-statistic (in group-defined amygdala activation mask) for faces-shapes contrast	0.3729
	Median z-statistic (in group-defined amygdala activation mask) for faces-shapes contrast	0.3579
	Median z-statistic (in group-defined mask) for faces-shapes contrast	0.3112
SetC for O3	**Felt hated by family member as a child**	0.4419
	Be littlement by partner or ex-partner	0.4036
	Childhood stressful events	0.3799
	Someone to take to the doctor when needed as a child	0.3169
	Physically abused by family as a child	0.2898
Predictive feature set for G2 outcomes	**Feature ranking**	**PC coefficient**
sMRI for O1	**Volume of parsopercularis (right emisphere)**	0.1437
	Area of parsoeprcularis (right emisphere)	0.1348
	Area of supramarginal (right emisphere)	0.1180
	Mean thickness of precentral (right emisphere)	0.1073
	Volume of middle temporal (left emisphere)	0.1059
dMRI for O2	**Mean OD in fornix**	0.3516
	Mean RD in left superior fronto-occipital fasciculus	0.1874
	Mean RD in fornix	0.1405
	Mean RD in right superior fronto-occipital fasciculus	0.1347
	Mean ICVF in left superior fronto-occipital fasciculus	0.1166
Brs-fMRI for O3	**Strength of negative weights in IC 54 (Sub&Cereb)**	0.1496
	Strength of negative weights in IC 30 (SMN)	0.1384
	Strength of positive weights in IC 11(DMNp)	0.1347
	Strength of positive weights in IC 2 (VAN)	0.1328
	Strength of positive weights in IC 53 (Sub&Cereb)	0.1306
Predictive feature set for G3 outcome	**Feature ranking**	**PC coefficient**
Continued		

Predictive feature set for G outcomes	Feature ranking	PC coefficient
t-fMRI for TRD	**Median z-statistic (in group-defined mask) for faces-shapes contrast**	0.4657
	90th percentile of z-statistic (in group-defined amygdala activation mask) for faces-shapes contrast	0.4362
	Median z-statistic (in group-defined amygdala activation mask) for faces-shapes contrast	0.3744
	90th percentile of z-statistic (in group-defined mask) for faces-shapes contrast	0.3285
	90th percentile of z-statistic (in group-defined mask) for shapes activation	0.3149

Table 6. Ranking of the five most important features extracted from the best predictive feature set for all outcomes. The most important feature in each predictive feature set is reported in bold. G, group of clinical outcome; O, clinical outcome; PC, principal components; G, genetics; E, environment; sMRI, structural MRI; rs-fMRI, resting-state fMRI; dMRI, diffusion weighted MRI; t-fMRI, task-based fMRI; OD, orientation dispersion index; ISOVF, Isotropic Compartment Volume Fraction; RD, radial diffusivity; MET, metabolic equivalent task; ICVF, Intracellular Volume Fraction; SMN, somatomotor network; DMNp, default mode network posterior; VAN, ventral attention network; Sub&Cereb, subcortical and cerebellar network.

and TRD. Specifically, the structure brain characteristic of volume of pars opercularis (right hemisphere) was relevant for cancer (O1 for G2), maintaining PC-EL statistical significance despite poor individual feature ranking stability across subsampling iterations. For vascular problems (O2 for G2), dMRI feature of mean OD in the fornix emerged as the most robust predictor, demonstrating superior ranking consistency across subsampling iterations. Rs-fMRI connectivity strength in subcortical and cerebellar networks (IC 54) predicted diabetes (O3 for G2), showing intermediate ranking stability in sensitivity analysis. Among functional brain characteristics relevant for TRD (G3), the t-fMRI median z-statistic for faces-shapes contrast resulted in the most important features, demonstrating moderate robustness with strong weight preservation but variable top-tier positioning.

Overall, our findings related to sensitivity analysis conducted on the top-ranked features (Table 5) support the stability of top-ranked features linked with the different outcomes across subsampling iterations, with some exceptions in the high-dimensional feature sets that demonstrated ranking instability despite preserved statistical significance for PC-EL outcome association (see Supplementary Materials for detailed stability metrics results in Tables S9, S10, S11).

Discussion

In the present work, we reported the novel application of TDA to assess the differential performance of multimodal datasets including genetic, environmental, and brain characteristics to stratify individuals with MDD based on different health outcomes, embracing disease severity, medical comorbidities, and TRD. For the first time, TDA was employed to perform a multivariate stratification of MDD based on diverse sets of predictors on the large UKB cohort, enabling a comprehensive, robust, and data-driven exploration of this heterogeneous disease. The data-driven TDA approach revealed a complex picture of predictors, unraveling the multitude of domains that contribute to stratifying MDD based on the selected clinical outcomes. Consistent with previous knowledge[18,56–58], we found a key role for the environment, alone or integrated with biological factors, in determining the severity of depressive symptoms. Conversely, brain characteristics emerged as the most relevant predictors for medical comorbidities and TRD. Overall, these findings support the hypothesis that the consideration of genetic, environmental, and brain characteristics is essential to characterize the heterogeneity of MDD phenotypes. TDA has emerged as a promising unsupervised tool for identifying candidate biological and environmental markers of disease outcomes based on different multimodal feature sets, that could be used to predict disease-related trajectories. If replicated on independent cohorts, our results will set the ground for a meaningful dimensional stratification of MDD, for clarifying the corresponding biological and environmental underpinnings, and potentially for developing clinical applications.

Multivariate data analysis of TDA and application of SAFE score

In this study, we applied TDA to different multidimensional feature spaces, for the first time in a large cohort of MDD individuals from the UKB. TDA applies mathematical concepts from geometric topology to characterize the shape of a multidimensional dataset into a simple and compact geometric structure, named topological skeleton, representing a simple topological summary of the data. This approach allows the investigation of multidimensional data, and further the characterization of the dataset in terms of geometrical and topological information extracted from the topological skeleton[32]. Examples of TDA in precision medicine context are beginning to appear in the literature, including the definition of clinically meaningful phenotypes for diverse medical conditions[23,24,29]. However, TDA was never applied before to multimodal (genetic, environmental, neuroimaging) and high-dimensional datasets in an MDD population. Therefore, the availability of quantitative functional maps created by the SAFE score allows an increase in the interpretability of the association between features on which the network is constructed and the target outcome analyzed, especially in the case of high-dimensional complex networks for which the network's visual inspection could be computationally challenging. Given that clinical applications are high-stakes, we require understandability from the prediction tools or they will grow in distrust[59]. This is especially true in the analysis of the neurobiological underpinnings of psychiatric disorders, where innovative statistical tools should assist clinicians without introducing further complexity, by proving to be trustworthy, therefore not only valid and reliable, but also easily understandable. Therefore, the TDA technique, coupled with the extraction of SAFE score functional maps, provides the simultaneous possibility

of multifactorial data aggregation, capturing the variability of multiple feature sets, and an improvement in interpreting the associations between TDA networks and target outcome variables.

Environment's interactions as a key determinant of depressive symptoms

We observed a key influence of the environment on depressive symptoms, as it represented the most predictive feature sets for most G1 outcomes, either alone or in combination with genetic and brain features. Specifically, intertwining complex effects of biological and environmental features were observed for the outcomes of self-harm behaviors, as well as the frequency of depressed days and the duration of worst depressive episode. These results are very important as we move further into the practice of precision medicine. Instead of suggesting a single (either genetic or environmental) predictor of mental health, our results point to an interaction between one's biological makeup and how they adapt to their experiences and environment, strengthening the idea of depression as a complex disorder, characterized by biological and environmental aspects. It is important to notice that only multimodal studies, assessing both the biological (genetic, imaging) and the environmental part, could capture these nuances. Therefore, this represents a major strength of the current study.

In this sense, it is established that the pathogenesis of MDD can be traced back to the synergic interaction between biological aspects (e.g., genes and neurobiology) and environmental risk factors[43], although the relative weight of the single factors is unclear. Apart from the etiopathogenesis, recent studies suggested that outcome measures, such as global functioning, could be predicted using a combination of baseline clinical, brain morphological, and environmental information[18].

A direct link has also been established between various environmental factors and brain structural and functional changes, highly correlated with severe depressive symptoms and poor prognosis[57,58].

Brain characteristics might therefore act as mediators of the environment. Consistently with this hypothesis, within the G–E-Imaging feature set of predictors, the microstructural integrity of the fornix was one of the most relevant predicting features for G1 (severity-related) outcomes. However, sensitivity analysis revealed ranking instability for G–E-Imaging features, highlighting the need for independent validation of high-dimensional multimodal datasets to confirm brain-environment associations.

This is in line with previous studies suggesting that the fornix and stria terminalis are involved in the pathophysiology of mood and psychotic disorders[60,61]. White matter (WM) abnormalities in the fornix have been described in patients with early-onset MDD[62], as well as during late-life depression[63]. Further, alterations in the fornix in patients with MDD might impair connections between brain regions such as the hippocampus and prefrontal cortex, which are important for depression and psychological functioning[64].

Depression-related outcomes reflecting anxiety, life meaning, and suicidal ideation, were best explained by environmental factors, with a key role of telomere length, reflecting both innate and environmental characteristics. Conversely, childhood stressful events, especially the feeling of being hated by family members as a child, were associated with the impact of the disease on normal roles. Previous studies have investigated the impact of the environment on the severity of depression, identifying early life adversities, stressful life events, socioeconomic status, and exposure to traumatic events as determinants of the progression and intensity of depressive symptoms[57,58,65,66].

Finally, regarding telomere length, this feature has been previously associated with psychiatric disorder and especially MDD[67,68]. Telomere length can be considered a cellular clock, affecting how quickly cells reach senescence[69]. A compromised telomere biology has been associated with different medical conditions, and recent studies suggested telomere shortening as a potential mechanism by which MDD may increase the risk of morbidity and mortality[67]. The causal nature of this association is not known; recent literature suggested an interaction between inflammation and telomeres, with the possible mediation of gut microbiome[68]. Telomere shortening is known to result from repeated mitotic divisions and exposure to a variety of cellular stress mechanisms[70]. Thus, activation of telomerase activity during stress may represent one of the compensatory mechanisms to withstand stress-related disorders. In this sense, it has been speculated that MDD is associated with increased cellular stress and replication, resulting therefore in accelerated telomere shortening[71]. However, more studies are needed to understand the prospective importance of telomere length in MDD. Importantly, sensitivity analysis revealed an intermediate-to-robust pattern of stability for environmental features across G1 outcomes, with telomere length demonstrating exceptional ranking consistency and childhood stressful events showing more variable performance. These findings help support the reliability of environmental predictors, though validation in independent samples remains essential for establishing robust biomarkers.

Particularly, mixed results are reported regarding the association between telomere shortening and severity of MDD, with some studies reporting a link between telomeres and severity measures[72], while others reporting negative results[73], confirming it as a potential source of heterogeneity. Taken together, these findings highlight the importance of considering the role of the environment in the assessment of MDD, as it can shape disease severity and potentially progression. Our results confirm that the integration of environmental factors within a predictive stratification framework might improve outcome prediction.

Brain imaging predictors for medical comorbidities of MDD

Altered connectivity in selective brain regions, abnormal structural brain measures, in combination with lifestyle factors and chronic stress events, were previously associated with clinical outcomes of medical comorbidities conditions[74]. Our results suggest that comorbid medical conditions in MDD could be better predicted by brain imaging features, in some cases along with environmental aspects, than by genetic factors. Specifically, the best predictors for cancer, type 2 diabetes, and vascular diseases were sMRI, dMRI, and fMRI feature sets, respectively.

For the outcome "vascular problems", also environmental factors and their interaction with genes and imaging seem to have a role, further indicating the complex intertwining set of factors that are implicated in MDD. MDD is associated with several comorbidities, including vascular and cardiac problems, risk of diabetes,

autoimmune and inflammatory disorders, and cancer[75,76]. Although the clear mechanisms of this increased risk of comorbidities are still unknown, studies have suggested a bidirectional relationship between cardiovascular problems and major psychiatric conditions, involving different pathophysiological mechanisms, such as metabolic dysfunction, inflammation, oxidative stress, neurohormonal dysregulation, and shared genetic factors[76]. Our results are in line with this hypothesis of shared risk with both environmental and biological factors involved, although they also suggest an important role of brain imaging.

Previous evidence of a role of brain features in shaping cardiometabolic risk factors is limited. Although different studies analyzed the link between depressive disorders and cardiometabolic diseases[77], the brain underpinnings of this link are less known. The autonomic regulation and specifically the dynamics related to central-autonomic network (CAN) is one of the most studied mechanisms that mediates the link between depression and cardiovascular health[78,79].

However, other studies provided insights on the association between alterations in neural substrates and cardiometabolic conditions in the general population (e.g., diabetes, cardiovascular diseases/events, cancer, autoimmune disease), as well as between altered brain patterns and the risk of these conditions. Independent associations between cardio- and cerebrovascular risk factors and brain imaging changes were found to anticipate the disease manifestation[80]. On the other hand, only a few studies suggest an association of neuropsychiatric disorders, including MDD, with medical conditions and concomitant alterations in brain features[74,81,82]. Indeed, the specific role of brain patterns, including predictors, moderators, or mediators for such clinical factors, remains unexplored. Our results suggest different brain imaging predictors for the different comorbidities examined, with possible roles of frontal morphology in cancer, fornix microstructure in cardio-vascular problems, and functional connectivity in the subcortical and cerebellar network in type 2 diabetes. Sensitivity analysis revealed that dMRI features for vascular problems demonstrated superior stability, while structural and functional predictors showed more variable patterns, suggesting that independent validation is particularly needed for biomarkers related to morphometric and connectivity-based associations.

In conclusion, our understanding of the role of brain imaging features in general medical comorbidities of MDD remains largely incomplete, and mostly limited to the study of CAN dysfunctions[83,84]. Further studies are needed to clarify the pathways connecting depressive disorders, general health diseases, and brain characteristics.

fMRI features as predictors of TRD

Previous insights highlighted the complex interplay between environmental factors and brain features in defining other challenging clinical outcomes, such as TRD[85]. Indeed, specific MRI features related to brain structure and function have been reported as possible markers of TRD, demonstrating structural abnormalities and disrupted connectivity within critical brain regions of frontolimbic areas, including prefrontal, anterior cingulate cortex, hippocampus, amygdala, and insula observed in patients with MDD and poor treatment outcomes[86,87]. Our data showed that brain responses to emotional-cognitive tasks were the best predictors for TRD in the UKB. Specifically, the most important feature associated with TRD was the brain activation elicited by faces with negative emotions. Recent literature[88] suggested that brain function captured by fMRI might differentiate TRD when compared both to healthy controls and to MDD patients who respond to treatment, especially in emotional and reward brain areas. In particular, an alteration in amygdala response to emotional processing has been reported[89]. Consistently, a decreased ventromedial and ventrolateral prefrontal-amygdala connectivity during face processing seems to be reversed by the amelioration of symptoms following treatment with psilocybin or administration of selective serotonin reuptake inhibitors[90]. Although the brain imaging features used in our study were not analyzed in relation to TRD by previous studies, our findings suggest that fMRI aspects might be important to discriminate TRD and predict treatment response. Notably, t-fMRI demonstrated moderate stability in the sensitivity analysis, suggesting a potential of this biomarker for TRD condition, though replication in independent cohorts is essential given the heterogeneity of this condition.

However, the understanding of the risk factors of TRD remains limited for several reasons. First, it is an uncommon outcome, and therefore, the number of subjects enrolled is often small. Even in our analyses, the proportion of patients suffering from TRD is small, although in line with previously reported rates[91]. Therefore, all the results must be regarded with caution. However, compared to previous ones[88], our study was able to analyze a larger sample of TRD patients with fMRI. Second, TRD samples may have an increased heterogeneity vs overall MDD, due to the long-term treatment of these patients with multiple medications and the different pathways that might lead to TRD. In this context, TDA provides the advantage of simultaneously handling all the variables in a common multifactorial space that reflects the structure of the underlying dataset.

Study limitations and future perspectives

This study presents potential limitations that need to be discussed. First, TDA application is not straightforward and requires the tuning of multiple parameters. In our study, the resolution and gain TDA parameters have been defined by varying each metric within a range and, through visual inspection, by choosing the ones that ensure that the majority of subjects is included in a connected node, and all the nodes are connected. In future applications, a methodological pipeline for defining TDA parameters is desirable. The graph that is obtained from the high-dimensional raw dataset is highly sensitive to the definition of gain and resolution, as the size and the overlap between bins in the filtered space are responsible for the definition of a coarse-grained network. Similarly, the choice of the filter function, number of filter projections, and clustering method might have influenced the results.

Furthermore, within this study, we did not apply a quantitative comparison for parameter selection. Instead, we applied the classical TDA approach by manually varying cover parameters and using visual inspection to identify optimal combinations that preserve topological structures while avoiding artifacts[28,51,52]. Despite being well-established, this approach relies on subjective visual assessment, and testing a wide range of R-G pairs

can be highly time-consuming. While a comprehensive exploration of covering parameters' combinations, as well as filtering and clustering methods, might be performed using stability and connectivity measures as quality metrics, hyperparameter tuning would still partly depend on arbitrary user's choices or selections based on specific applications (i.e., range of parameters to be explored), indicating an inherently arguable solution not completely free from user's choice. Future studies should address the trade-off between automatic hyperparameter tuning (which can retrieve dataset-specific parameters and reduce generalizability) and manual approaches that, while more generalizable, are computationally intensive and subjective. Developing robust parameter selection frameworks that balance effectiveness with generalizability also across independent samples represents an important area for future TDA methodological advancement. However, our approach currently represents the most common methodology for TDA parameters setup[24,27,51,52].

Additionally, the selection of the predictive and outcome features inevitably entailed choices that might be questioned, as well as the definition of outcomes such as TRD, which was derived from the number of antidepressant switches within certain time frames according to primary care records, as previously described[36]. Among the candidate predictors, a delicate choice regarded the inclusion of pharmacological therapy as a candidate confounding predictive factor in all feature sets, which was motivated by the hypothesized interaction of therapy with all the other sets. Further, the reason for taking medications (i.e., disease diagnosis), the duration, the dosage, the response, and/or adverse effects related to antidepressant medications were not considered, thus impeding further analysis with actual treatment response and dosage of medications.

Further, the methodological choice related to the number of subsampling iterations (B = 100) used for evaluating the significant association patterns among graph features-outcomes could have represented a limit for the analysis. Thus, further studies might employ a higher number of subsampled iterations, trying to balance the identification of a stable SAFE score estimation and computational resources to create a complex graph on a large dataset. Moreover, an external set for validation is required, and further studies are needed to replicate our results within an external validation set.

Despite conducting a comprehensive sensitivity analysis for top-ranked features across B = 100 subsamples, potential biases in the results may depend on sample composition effects inherent to our 50% subsampling strategy. Future studies should explore alternative subsampling proportions to further validate the best predictive features' generalizability. Additionally, considering that we did not apply a robust framework for parameter selection but used the classical approach with visual inspection of topological stability, an important future methodological advancement should be the systematic exploration of feature stability concerning parameter variations.

Furthermore, while our study focused on comparing the predictive capability of different feature sets for clinical outcomes without performing any patients' subtyping, future studies could build upon our TDA-based framework to develop data-driven pipelines specifically designed for MDD stratification into distinct clinical subgroups, thus leveraging the identified optimal feature combinations to characterize homogeneous subgroups with distinct clinical trajectories.

Moreover, although SAFE score application within Mapper context introduced by *Liao and colleagues* represents a valuable first metric to quantify statistical associations between network organization and selected outcome variables, future studies introducing new metrics to investigate Mapper graph's feature-outcome association are needed. Indeed, besides defining the graph's feature-outcome association at the level of the node, as is done by SAFE analysis, a full characterization of the graph's topological properties might be useful for further exploitation, with the aim to develop a new integrative framework based on TDA for several disorders' stratification.

Despite maintaining the integrity of distinct clinical domains and preserving the UKB population distribution of each group of outcomes, our study presents two main limitations related to sample characteristics. First, unbalanced sample sizes across clinical groups (Group 1: n = 1861, Group 2: n = 3044, Group 3: n = 537) may have limited our ability to equally compare cross-clinical domain patterns. Topological data analysis methods benefit from larger sample sizes to construct more robust network representations with sufficient connectivity[34]. While our subsampling approach successfully addressed within-group outcome dimensionality, the substantial sample size differences may have limited our ability to equally compare cross-clinical domain patterns and detect subtle topological relationships spanning across different groups of clinical outcomes. Although our group numerosity reflects real-world clinical representativeness of UKB sample, balanced inter-group sampling could potentially enhance detection of integrated biomarker patterns across clinical domains. Second, label imbalance within specific outcomes (e.g., low prevalence of treatment non-response) corresponds to the actual prevalence of these clinical conditions within the UKB general population. However, this imbalance may limit statistical power for detecting associations with rare clinical phenotypes. Future studies considering clinical populations with homogeneous sample sizes and balanced outcome distributions could provide complementary insights into cross-domain topological structures.

Specifically regarding TRD outcomes, the performance related to t-fMRI features must be interpreted with caution given the smallest sample size (n = 537) and the relatively low prevalence of TRD cases (~ 7%) within G3. While this prevalence reflects the expected real-world distribution in general population cohorts[36,92], the limited statistical power associated with the sample size may affect the robustness and generalizability of TRD-specific findings. Thus, the observed t-fMRI patterns for TRD prediction should be considered preliminary and require validation in larger, dedicated clinical populations with enhanced representation of TRD cases to confirm the topological relationships extracted. Future studies should focus on the validation of these neuroimaging-based predictive patterns in independent samples to establish their clinical utility and generalizability across diverse TRD populations.

In addition, the classification of environmental factors in two main classes used in this study was meant to simplify the input data and increase interpretability; however, other classifications could have been applied.

Among brain features, our study considered as functional features of resting-state the strength of positive and negative connections of the brain networks identified through ICA. Alternative topological measures of the functional brain networks, such as degree, centrality, clustering coefficients, and modularity, could have been considered. To conclude, the MDD individuals considered in this study may not be representative of the general MDD population. Indeed, UKB is known to be enriched in females, elderly, wealthier, and more educated individuals vs the general UK population[37].

Conclusions

Our study has exploited TDA as a powerful approach to investigate the biological and environmental markers of several outcome domains in a large MDD population, spanning from disease severity to medical comorbidities and TRD. We highlighted the importance of multimodal studies in psychiatric disorders and how TDA could be a viable solution to handle multimodal datasets. Our findings suggested that multivariate data analysis based on data-driven TDA enables the handling of high-dimensional datasets and the extraction of hidden relationships among multiple types of features. Of note, the application of SAFE score analysis within a TDA pipeline enables a quantitative understanding of how TDA networks are functionally organized with respect to a specific target outcome. Despite the need to test our findings on independent samples, this study provides avenues for the robust definition of biologically- and environmentally-determined dimensions affecting relevant health outcomes in MDD.

Data availability

There are restrictions prohibiting the provision of data in this manuscript. The data in this manuscript were obtained from a third party, UK Biobank, upon application. The genetic, environmental, brain imaging and clinical outcomes' data have been deposited with the UK Biobank and are freely available to approved researchers, as has been done with other datasets to date. Interested parties can apply for data from UK Biobank directly, at http://www.ukbiobank.ac.uk. UK Biobank will consider data applications from bona fide researchers for health-related research that is in the public interest. By accessing data from UK Biobank, readers will be obtaining it in the same manner as we did.

Code availability

The central code of the main work is available upon request at https://osf.io/2se7r/?view_only=9cc533ad4c05449dabfd8613ad9dfdf7.

Received: 12 February 2025; Accepted: 9 September 2025
Published online: 10 October 2025

References

1. Ferrari, A. J. et al. Burden of depressive disorders by country, sex, age, and year: Findings from the global burden of disease study 2010. *PLoS Med.* **10**, e1001547 (2013).
2. Depression, WHO. *Other common mental disorders: global health estimates*. Geneva: World Health Organization 24, (2017).
3. Gutiérrez-Rojas, L., Porras-Segovia, A., Dunne, H., Andrade-González, N. & Cervilla, J. A. Prevalence and correlates of major depressive disorder: A systematic review. *Braz. J. Psychiatry* **42**, 657–672 (2020).
4. Chahal, R., Gotlib, I. H. & Guyer, A. E. Research review: Brain network connectivity and the heterogeneity of depression in adolescence—A precision mental health perspective. *J. Child Psychol. Psychiatry* **61**, 1282–1298 (2020).
5. Buch, A. M. & Liston, C. Dissecting diagnostic heterogeneity in depression by integrating neuroimaging and genetics. *Neuropsychopharmacol* **46**, 156–175 (2021).
6. Jentsch, M. C. et al. Biomarker approaches in major depressive disorder evaluated in the context of current hypotheses. *Biomark. Med.* **9**, 277–297 (2015).
7. Paul, R. et al. Treatment response classes in major depressive disorder identified by model-based clustering and validated by clinical prediction models. *Transl. Psychiatry* **9**, 187 (2019).
8. Trombello, J. M. et al. Characterizing anxiety subtypes and the relationship to behavioral phenotyping in major depression: Results from the EMBARC study. *J. Psychiatr. Res. luglio* **102**, 207–215 (2018).
9. Wang, X. et al. Unveiling diverse clinical symptom patterns and neural activity profiles in major depressive disorder subtypes. *eBiomedicine. giugno* **116**, 105756 (2025).
10. Jentsch, M. C. et al. Biomarker approaches in major depressive disorder evaluated in the context of current hypotheses. *Biomark Med.* **9**(3), 277–297 (2015).
11. Paul, R. et al. Treatment response classes in major depressive disorder identified by model-based clustering and validated by clinical prediction models. *Transl. Psychiatry* **9**(1), 187 (2019).
12. Otte, C. et al. Major depressive disorder. *Nat. Rev. Dis. Primers* **2**, 16065 (2016).
13. Kautzky, A. et al. Combining machine learning algorithms for prediction of antidepressant treatment response. *Acta Psychiatr. Scand.* **143**, 36–49 (2021).
14. Akil, H. et al. Treatment resistant depression: A multi-scale, systems biology approach. *Neurosci. Biobehav. Rev.* **84**, 272–288 (2018).
15. Pigoni, A. et al. Can machine learning help us in dealing with treatment resistant depression?. *Rev. J. Affect. Disord.* **259**, 21–26 (2019).
16. Maggioni, E. et al. Common and distinct structural features of schizophrenia and bipolar disorder: The European network on psychosis, affective disorders and cognitive trajectory (ENPACT) study. *PLoS ONE* **12**, e0188000 (2017).
17. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
18. Antonucci, L. A. et al. Using combined environmental–clinical classification models to predict role functioning outcome in clinical high-risk states for psychosis and recent-onset depression. *Br. J. Psychiatry* **220**, 229–245 (2022).
19. Cotrena, C., Damiani Branco, L., Ponsoni, A., Milman Shansis, F. & Paz Fonseca, R. Neuropsychological clustering in bipolar and major depressive disorder. *J. Int. Neuropsychol. Soc.* **23**, 584–593 (2017).
20. Beijers, L., Wardenaar, K. J., van Loo, H. M. & Schoevers, R. A. Data-driven biological subtypes of depression: Systematic review of biological approaches to depression subtyping. *Mol. Psychiatry* **24**, 888–900 (2019).

21. Wang, Y. et al. Data-driven clustering differentiates subtypes of major depressive disorder with distinct brain connectivity and symptom features. *Br. J. Psychiatry* **219**, 606–613 (2021).
22. Chen, X., Dai, Z. & Lin, Y. Biotypes of major depressive disorder identified by a multiview clustering framework. *J. Affect Disord.* **329**, 257–272 (2023).
23. Kyeong, S., Kim, J.-J. & Kim, E. Novel subgroups of attention-deficit/hyperactivity disorder identified by topological data analysis and their functional network modular organizations. *PLoS ONE* **12**, e0182603 (2017).
24. Romano, D. et al. Topological methods reveal high and low functioning neuro-phenotypes within fragile X syndrome. *Hum. Brain Mapp.* **35**, 4904–4915 (2014).
25. Li, L. et al. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci. Transl. Med.* **7**, 311ra174 (2015).
26. Yamanashi, T. et al. Topological data analysis (TDA) enhances bispectral EEG (BSEEG) algorithm for detection of delirium. *Sci. Rep.* **11**, 304 (2021).
27. Lum, P. Y. et al. Extracting insights from the shape of complex data using topology. *Sci. Rep.* **3**(1), 1236 (2013).
28. Chazal F, Michel B. An introduction to topological data analysis: Fundamental and practical aspects for data scientists. Front Artif Intell. 29 settembre 2021.
29. Nielson, J. L. et al. Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. *Nat. Commun.* **6**, 8581 (2015).
30. Lee, H., Kang, H., Chung, M. K., Kim, B.-N. & Lee, D. S. Persistent brain network homology from the perspective of dendrogram. *IEEE Trans. Med. Imaging* **31**, 2267–2277 (2012).
31. Yao, Y. et al. Topological methods for exploring low-density states in biomolecular folding pathways. *J. Chem. Phys.* **130**, 144115 (2009).
32. Skaf, Y. & Laubenbacher, R. Topological data analysis in biomedicine: A review. *J. Biomed. Inform.* **130**, 104082 (2022).
33. Baryshnikova, A. Systematic functional annotation and visualization of biological networks. *Cell Syst.* **2**, 412–421 (2016).
34. Liao, T., Wei, Y., Luo, M., Zhao, G.-P. & Zhou, H. tmap: An integrative framework based on topological data analysis for population-scale microbiome stratification and association studies. *Genome Biol.* **20**, 293 (2019).
35. Casaclang-Verzosa, G. et al. Network tomography for understanding phenotypic presentations in aortic stenosis. *JACC Cardiovasc. Imaging* **12**, 236–248 (2019).
36. Fabbri, C. et al. Genetic and clinical characteristics of treatment-resistant depression using primary care records in two UK cohorts. *Mol. Psychiatry* **26**, 3363–3373 (2021).
37. Fry, A. et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
38. Davis, K. A. S. et al. Mental health in UK Biobank—development, implementation and results from an online questionnaire completed by 157 366 participants: A reanalysis. *BJPsych open* **6**, e18 (2020).
39. Alfaro-Almagro, F. et al. Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* **166**, 400–424 (2018).
40. Hariri, A. R., Tessitore, A., Mattay, V. S., Fera, F. & Weinberger, D. R. The amygdala response to emotional stimuli: A comparison of faces and scenes. *Neuroimage* **17**, 317–323 (2002).
41. Jiang, R. et al. A functional connectome signature of blood pressure in >30 000 participants from the UK Biobank. *Cardiovasc. Res.* **119**, 1427–1440 (2023).
42. Codd, V. et al. Measurement and initial characterization of leukocyte telomere length in 474,074 participants in UK Biobank. *Nat. Aging.* **2**, 170–179 (2022).
43. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C.A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
44. Chang, C. C. et al. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaSci* **4**, 7 (2015).
45. Milaneschi, Y., Lamers, F., Berk, M. & Penninx, B. W. J. H. Depression heterogeneity and its biological underpinnings: Toward immunometabolic depression. *Biol. Psychiat.* **88**, 369–380 (2020).
46. Fabbri, C., Mutz, J., Lewis, C. M. & Serretti, A. Depressive symptoms and neuroticism-related traits are the main factors associated with wellbeing independent of the history of lifetime depression in the UK Biobank. *Psychol. Med.* **53**, 3000–3008 (2023).
47. Brailean, A., Curtis, J., Davis, K., Dregan, A. & Hotopf, M. Characteristics, comorbidities, and correlates of atypical depression: Evidence from the UK Biobank mental health survey. *Psychol. Med.* **50**, 1129–1138 (2020).
48. Reddy KVA, Ambati SR, Rithik Reddy YS, Reddy AN. *AdaBoost for Parkinson's Disease Detection using Robust Scaler and SFS from Acoustic Features*. In: 2021 Smart Technologies, Communication and Robotics (STCR).
49. McInnes L, Healy J, Melville J. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction* [Internet]. arXiv; 2020.
50. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol. Gennaio* **37**(1), 38–44 (2019).
51. Singh G, Memoli F, Carlsson G. *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition* [Internet]. Eurographics Symposium on Point-Based Graphics. The Eurographics Association; 2007.
52. Palande, S. et al. Topological data analysis reveals a core gene expression backbone that defines form and function across flowering plants. *Plos Biol.* **21**(12), e3002397 (2023).
53. Zhang, T. T. & Yuan, B. Density-based multiscale analysis for clustering in strong noise settings with varying densities. *IEEE Access.* **6**, 25861–25873 (2018).
54. Wills, P. & Meyer, F. G. Metrics for graph comparison: A practitioner's guide. *PLoS ONE* **15**, e0228728 (2020).
55. Grady, L. J. & Polimeni, J. *Discrete Calculus: Applied Analysis on Graphs for Computational Science* 366 (Springer, New York, 2010).
56. Mullins, N. et al. Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychol. Med.* **46**, 759–770 (2016).
57. Frodl, T. et al. Effect of hippocampal and amygdala volumes on clinical outcomes in major depression: A 3-year prospective magnetic resonance imaging study. *J. Psychiatry Neurosci.* **33**, 423–430 (2008).
58. Liston, C., McEwen, B. S. & Casey, B. J. Psychosocial stress reversibly disrupts prefrontal processing and attentional control. *Proc. Natl. Acad. Sci. USA* **106**, 912–917 (2009).
59. Joyce, D. W., Kormilitzin, A., Smith, K. A. & Cipriani, A. Explainable artificial intelligence for mental health through transparency and interpretability for understandability. *npj Digit. Med.* **6**, 6 (2023).
60. Koshiyama, D. et al. White matter microstructural alterations across four major psychiatric disorders: Mega-analysis study in 2937 individuals. *Mol. Psychiatry* **25**, 883–895 (2020).
61. Zhang, X. et al. Association of visual health with depressive symptoms and brain imaging phenotypes among middle-aged and older adults. *JAMA Netw. Open* **5**, e2235017 (2022).
62. Geng, H. et al. Disrupted structural and functional connectivity in prefrontal-hippocampus circuitry in first-episode medication-naïve adolescent depression. *PLoS ONE* **11**, e0148345 (2016).
63. Li, W. et al. Effects of the coexistence of late-life depression and mild cognitive impairment on white matter microstructure. *J. Neurol. Sci.* **338**, 46–56 (2014).
64. Shim, J.-M., Cho, S.-E., Kang, C.-K. & Kang, S.-G. Low myelin-related values in the fornix and thalamus of 7 Tesla MRI of major depressive disorder patients. *Front. Mol. Neurosci.* **16**, 1214738 (2023).

65. Gianaros, P. J. et al. Potential neural embedding of parental social standing. *So. Cogn. Affect Neurosci.* **3**, 91–96 (2008).
66. Dannlowski, U. et al. Childhood maltreatment is associated with an automatic negative emotion processing bias in the amygdala. *Hum. Brain Mapp.* **34**, 2899–2909 (2013).
67. Ridout, K. K., Ridout, S. J., Price, L. H., Sen, S. & Tyrka, A. R. Depression and telomere length: A meta-analysis. *J. Affect. Disord.* **191**, 237–247 (2016).
68. Bazaz, M. R., Balasubramanian, R., Monroy-Jaramillo, N. & Dandekar, M. P. Linking the triad of telomere length, inflammation, and gut dysbiosis in the manifestation of depression. *ACS Chem. Neurosci.* **12**, 3516–3526 (2021).
69. Price, L. H., Kao, H.-T., Burgers, D. E., Carpenter, L. L. & Tyrka, A. R. Telomeres and early-life stress: An overview. *Biol. Psychiat.* **73**, 15–23 (2013).
70. Xu, L. & Blackburn, E. H. Human cancer cells Harbort-stumps, a distinct class of extremely short telomeres. *Mol. Cell* **28**, 315–327 (2007).
71. Wolkowitz, O. M., Epel, E. S., Reus, V. I. & Mellon, S. H. Depression gets old fast: Do stress and depression accelerate cell aging?. *Depress. Anxiety* **27**, 327–338 (2010).
72. Mendes-Silva, A. P. et al. Telomere shortening in late-life depression: A potential marker of depression severity. *Brain and Behavior* **11**, e2255 (2021).
73. Hartmann, N., Boehner, M., Groenen, F. & Kalb, R. Telomere length of patients with major depression is shortened but independent from therapy and severity of the disease. *Depress. Anxiety* **27**, 1111–1116 (2010).
74. Berk, M. et al. Comorbidity between major depressive disorder and physical diseases: A comprehensive review of epidemiology, mechanisms and management. *World Psychiatry* **22**, 366–387 (2023).
75. Massironi, S. et al. The Burden of psychiatric manifestations in inflammatory bowel diseases: A systematic review with meta-analysis. *Inflamm. Bowel Dis.* **31**(5), 1441–1459 (2025).
76. Ong, L. T. & Sia, C. H. Pathophysiological links between myocardial infarction and anxiety disorder, major depressive disorder. *Bipolar Disord. Schizophr. Biol.* **14**(4), 336 (2025).
77. Hare, D. L., Toukhsati, S. R., Johansson, P. & Jaarsma, T. Depression and cardiovascular disease: A clinical review. *Eur. Heart J.* **35**, 1365–1372 (2014).
78. Grippo, A. J. & Johnson, A. K. Stress, depression and cardiovascular dysregulation: A review of neurobiological mechanisms and the integration of research from preclinical disease models: Review. *Stress* **12**, 1–21 (2009).
79. Penninx, B. W. J. H. Depression and cardiovascular disease: Epidemiological evidence on their linking mechanisms. *Neurosci. Biobehav. Rev.* **74**, 277–286 (2017).
80. Friedman, J. I. et al. Brain imaging changes associated with risk factors for cardiovascular and cerebrovascular disease in asymptomatic patients. *JACC Cardiovasc. Imaging* **7**, 1039–1053 (2014).
81. Sheline, Y. I. 3D MRI studies of neuroanatomic changes in unipolar major depression: The role of stress and medical comorbidity. *Biol. Psychiat.* **48**, 791–800 (2000).
82. Gold, S. M. et al. Comorbid depression in medical diseases. *Nat. Rev. Dis. Primers* **6**, 69 (2020).
83. Valenza, G. Depression as a cardiovascular disorder: Central-autonomic network, brain-heart axis, and vagal perspectives of low mood. *Front. Netw. Physiol.* **3**, 1125495 (2023).
84. Tonhajzerova, I. et al. Major depressive disorder at adolescent age is associated with impaired cardiovascular autonomic regulation and vasculature functioning. *Int. J. Psychophysiol.* **181**, 14–22 (2022).
85. Mayberg, H. S. et al. Cingulate function in depression: A potential predictor of treatment response. *NeuroReport* **8**, 1057–1061 (1997).
86. Fonseka, T. M., MacQueen, G. M. & Kennedy, S. H. Neuroimaging biomarkers as predictors of treatment outcome in major depressive disorder. *J. Affect. Disord.* **233**, 21–35 (2018).
87. Klok, M. P. C. et al. Structural brain characteristics in treatment-resistant depression: Review of magnetic resonance imaging studies. *BJPsych. Open* **5**, e76 (2019).
88. Kotoula, V., Evans, J. W., Punturieri, C., Johnson, S. C. & Zarate, C. A. Functional MRI markers for treatment-resistant depression: Insights and challenges. in *Progress in Brain Research* vol. 278 117–148 (Elsevier, 2023).
89. Ferri, J. et al. Blunted amygdala activity is associated with depression severity in treatment-resistant depression. *Cogn. Affect. Behav. Neurosci.* **17**, 1221–1231 (2017).
90. Vai, B. et al. Fronto-limbic effective connectivity as possible predictor of antidepressant response to SSRI administration. *Eur. Neuropsychopharmacol.* **26**, 2000–2010 (2016).
91. McIntyre, R. S. et al. Treatment-resistant depression: Definition, prevalence, detection, management, and investigational interventions. *World Psychiatry* **22**(3), 394–412 (2023).
92. Sheehan, J. J. et al. The prevalence and national burden of treatment-resistant depression and major depressive disorder in the United States. *J. Clin. Psychiatry* **82** (2), 20m13699. https://doi.org/10.4088/JCP.20m13699 (2021).

Acknowledgments

This research has been conducted using the UK Biobank Resource under Application Number 56514 "Stratification of health outcomes in mood disorders". The study was supported by the Italian Ministry of Health (DEPTYPE project, grant n. GR-2019-12370616). PB was partially supported by grants from the Italian Ministry of Education and Research - MUR ('Dipartimenti di Eccellenza' Programme 2023–27 - Dept. of Pathophysiology and Transplantation, Università degli Studi di Milano), the Italian Ministry of Health (Hub Life Science- Diagnostica Avanzata, HLS-DA, PNC-E3-2022-23683266– CUP: C43C22001630001 / MI-0117; Ricerca Corrente 2025), and by the ERANET Neuron JTC 2023 (ERP-2023-23684211 - ERP-2023-Neuron-ResilNet). EM was partly supported by the Italian Ministry of University and Research (PRIN 2022 PNRR, grant n. P20229MFRC).

Author contributions

ET: Conceptualization, Methodology, Formal analysis, Software, Data curation, Writing—original draft, Writing—review & editing; AP: Writing—original draft, Writing—review & editing; NT: Methodology, Writing—review & editing; FC: Data curation, Methodology, Writing—review & editing; LF: Data curation, Methodology, Writing—review & editing; AMB: Methodology, Writing—review & editing; FB: Writing—review & editing; CF: Conceptualization, Project administration, Supervision, Data curation, Methodology, Writing—review & editing, Funding acquisition; BV: Conceptualization, Project administration, Supervision, Data curation, Methodology, Writing—review & editing, Funding acquisition; PB: Supervision, Writing—review & editing; EM: Conceptualization, Project administration; Supervision; Methodology; Writing—original draft, Writing—review & editing, Funding acquisition.

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval
UK Biobank had obtained ethics approval from the North West Multi-centre Research Ethics Committee (MREC) with approval number 11/NW/0382; participants provided written informed consent before inclusion.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-19624-0.

Correspondence and requests for materials should be addressed to P.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.