




## Predicting Major Depressive Disorder Using Random Forest Models Based on Psychological, Behavioral, and Lifestyle Indicators

Liam. Wooderson<sup>1</sup>, Milena. Fini<sup>2\*</sup>, Mariusz. Szypa<sup>3</sup>

<sup>1</sup> Department of Psychology, University of Toronto, Mississauga, ON L5L 1C6, Canada

<sup>2</sup> Department of Psychology, University of Quebec at Montreal, Montreal, QC, Canada

<sup>3</sup> School of Psychology, Australian Catholic University, Brisbane, Queensland, Australia

\* Corresponding author email address: milena.fini@uqam.ca

### Article Info

#### Article type:

Original Research

#### How to cite this article:

Wooderson, L., Fini, M., & Szypa, M. (2026). Predicting Major Depressive Disorder Using Random Forest Models Based on Psychological, Behavioral, and Lifestyle Indicators. *Journal of Assessment and Research in Applied Counseling*, 8(3), 1-15.

<https://dx.doi.org/10.61838/kman.jarac.5450>



© 2026 the authors. Published by KMAN Publication Inc. (KMANPUB), Ontario, Canada. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

### ABSTRACT

**Objective:** This study aimed to develop and evaluate a Random Forest classification model for predicting Major Depressive Disorder among Canadian adults using integrated psychological, behavioral, and lifestyle indicators.

**Methods and Materials:** This cross-sectional predictive study was conducted among 1,742 adults residing in Canada. Participants completed standardized self-report instruments assessing depressive symptoms, anxiety, stress, emotion regulation difficulties, perceived stress, sleep quality, physical activity, and lifestyle behaviors. Major Depressive Disorder status was determined using the Patient Health Questionnaire-9 cut-off score for clinically significant depressive symptoms. Psychological indicators included DASS depression, anxiety, and stress scores, perceived stress, and emotion regulation difficulties. Behavioral and lifestyle variables included sleep quality, sleep duration, physical activity, screen time, body mass index, alcohol consumption, and demographic characteristics. Data preprocessing included missing-value management, categorical encoding, and feature preparation. The dataset was divided into training and testing subsets using stratified sampling. A Random Forest classification algorithm was trained and optimized through five-fold cross-validation and grid-search hyperparameter tuning. Model performance was evaluated using accuracy, sensitivity, specificity, precision, F1-score, balanced accuracy, ROC-AUC, Cohen's Kappa, Matthews Correlation Coefficient, and feature-importance analysis.

**Findings:** The optimized Random Forest model demonstrated strong predictive performance on the independent test dataset, with accuracy of 91.38%, sensitivity of 89.12%, specificity of 92.31%, precision of 87.64%, F1-score of 88.37%, balanced accuracy of 90.72%, ROC-AUC of 0.957, Cohen's Kappa of 0.804, and Matthews Correlation Coefficient of 0.806. Five-fold cross-validation confirmed model stability, with mean accuracy of 91.38%, mean precision of 87.89%, mean recall of 89.14%, mean F1-score of 88.51%, and mean ROC-AUC of 0.957. Feature-importance analysis identified DASS depression, perceived stress,

emotion regulation difficulties, sleep quality, anxiety, stress, screen time, and physical activity as the strongest predictors.

**Conclusion:** The findings indicate that Random Forest modeling can accurately predict probable Major Depressive Disorder using psychological, behavioral, and lifestyle indicators. The model showed high discrimination, stable validation performance, and clinically interpretable predictor patterns, supporting its potential value as a scalable screening approach for identifying adults at elevated risk of depression.

**Keywords:** *Major Depressive Disorder; Random Forest; Machine Learning; Depression Prediction; Psychological Indicators; Lifestyle Behaviors; Sleep Quality; Mental Health Screening.*

## 1. Introduction

Major depressive disorder (MDD) is one of the most clinically significant psychiatric conditions because of its high prevalence, recurrent course, heterogeneous presentation, and extensive impairment across emotional, cognitive, interpersonal, occupational, and physical domains. Although depressive symptoms are often described through core affective features such as persistent sadness, anhedonia, hopelessness, fatigue, sleep disturbance, appetite change, impaired concentration, psychomotor alterations, and suicidal ideation, contemporary research increasingly conceptualizes MDD as a multidimensional disorder shaped by interacting psychological, behavioral, lifestyle, neurobiological, inflammatory, metabolic, microbial, and social determinants. This shift is important because depression is rarely the product of a single causal pathway; rather, it emerges from cumulative vulnerability and dynamic interactions among predispositional risk factors, stressful experiences, maladaptive cognitive-affective processes, biological dysregulation, and health-related behaviors. General overviews of MDD continue to emphasize its complexity, clinical burden, and diagnostic heterogeneity, while also calling for more integrated models capable of reflecting the disorder's diverse mechanisms and manifestations (Zhang, 2024). In this context, prediction-oriented research has become increasingly important because accurate identification of individuals at elevated risk can support earlier screening, targeted intervention, and more personalized mental health care.

Traditional approaches to MDD assessment have relied primarily on clinical interviews, self-report questionnaires, and symptom-based diagnostic criteria. These methods remain indispensable, but they may not fully capture the nonlinear and interactive risk patterns through which depressive disorder develops and persists. Depression is clinically heterogeneous: two individuals may receive the same diagnosis while sharing only a limited number of symptoms, differing substantially in anxiety, stress

exposure, sleep disruption, behavioral withdrawal, lifestyle patterns, inflammatory status, comorbid physical illness, and treatment response. Such heterogeneity has encouraged researchers to move beyond purely categorical diagnostic frameworks toward more dimensional, network-based, and mechanism-informed classifications. Neuroscience-informed psychiatric classification has therefore been proposed as a way to organize psychiatric conditions through interrelated biological, psychological, and behavioral networks rather than through symptom categories alone (Stoyanov & Maes, 2021). Similarly, precision psychiatry frameworks emphasize the importance of identifying longitudinal trajectories, clinical subtypes, and person-specific risk configurations, because MDD may follow different developmental and clinical pathways in different individuals (Atagün, 2025). These perspectives suggest that predictive modeling of MDD should integrate multiple domains of information rather than focusing exclusively on depressive symptom severity.

Psychological indicators remain central to the prediction of MDD because depressive disorder is strongly associated with negative affect, cognitive vulnerability, emotional dysregulation, perceived stress, and comorbid anxiety. Negative affectivity is particularly relevant because it reflects a broad dispositional tendency toward distress, worry, sadness, irritability, and emotional reactivity. Research using tree-based models has demonstrated the utility of modeling factors associated with trait negative affect in adults with and without major depression, supporting the relevance of machine learning methods for identifying psychological contributors to depressive vulnerability (Cañizares et al., 2023). Anxiety and stress symptoms are also clinically important because MDD frequently co-occurs with anxiety-related distress, and high perceived stress can intensify cognitive rumination, sleep disturbance, fatigue, and avoidance behaviors. Anhedonia, a core feature of depression characterized by diminished pleasure and reduced motivation, has received renewed

attention as both a symptom dimension and a treatment target, particularly because it may be less responsive to conventional interventions than general mood symptoms (Serretti, 2025). Accordingly, a predictive model that includes depression-related psychological distress, anxiety, stress, emotion regulation difficulties, and perceived stress may better approximate the real-world clinical complexity of MDD than a model based on demographic or diagnostic variables alone.

Behavioral and lifestyle indicators are also increasingly recognized as essential components of MDD risk and prognosis. Lifestyle psychiatry emphasizes that sleep, physical activity, nutrition, sedentary behavior, substance use, social rhythms, and daily routines are not merely secondary consequences of depression but may function as modifiable determinants of symptom onset, maintenance, and recovery. Recent expert consensus has specifically highlighted lifestyle interventions for MDD, emphasizing the clinical importance of physical activity, nutrition, sleep, stress management, social connection, and avoidance of harmful substances in depression prevention and treatment (Merlo et al., 2026). Exercise has been described as a therapeutic modality for both prevention and treatment of depression, with proposed mechanisms including improved cardiovascular health, neuroplasticity, inflammatory regulation, self-efficacy, and behavioral activation (Blumenthal & Rozanski, 2023). Sleep disturbance is another major behavioral domain, as poor sleep quality, insomnia, irregular sleep schedules, and daytime dysfunction are closely linked to depressive symptoms and may predict both onset and recurrence. Research on sleep and adolescent depression underscores the bidirectional relationship between sleep disruption and depressive symptoms and highlights sleep as a clinically meaningful intervention target (Castiglione-Fontanellaz & Tarokh, 2023). Seasonal and circadian patterns further illustrate that mood fluctuations may be predictable through self-report and physiological markers, supporting the relevance of repeated behavioral and biological indicators in depression prediction (Höller et al., 2022).

Nutrition-related factors have also become increasingly prominent in depression research. Dietary patterns may influence mood through inflammatory pathways, oxidative stress, metabolic regulation, gut microbiota composition, and neurotransmitter synthesis. Studies examining anthocyanin intake suggest that higher intake of specific dietary bioactives may be associated with lower depressive symptoms in adults with and without MDD, indicating that

nutritional quality can contribute meaningful information to depression risk profiling (Mestrom et al., 2023). Omega-3 fatty acids have similarly been discussed for their potential psychological benefits, with evidence suggesting mood-related relevance even when cognitive effects are less consistent (Kossowska-Wywił & Brzezicka, 2025). Broader conceptualizations of diet, microbiota, and psychological health have framed nutritional exposure as part of an “immune mind” pathway, linking dietary patterns, microbial ecology, immune regulation, and mental health outcomes (Marano, 2025). In addition, natural antidepressant mechanisms have been discussed in relation to neuroinflammation and gut flora, emphasizing the possibility that nutritional and microbiome-related interventions may influence depressive processes through inflammatory and neuroimmune modulation (Frota et al., 2025). These findings support the inclusion of lifestyle indicators such as diet quality, physical activity, sleep, and other health behaviors in predictive models of MDD.

Biological and physiological research further supports the need for multidomain prediction. MDD is associated with alterations in immune activity, inflammatory markers, neuroendocrine function, neural connectivity, gut microbiome composition, and epigenetic regulation. Psychoneuroimmunoendocrinological research has emphasized that immune system dysfunction is embedded within broader mood disorder mechanisms, involving interactions among inflammatory signaling, endocrine regulation, stress response systems, and clinical symptoms (Ortega et al., 2024). Biomarkers such as the neutrophil-to-lymphocyte ratio have been examined as possible predictors for early diagnosis of depression, reflecting growing interest in accessible inflammatory indicators (Ghafari et al., 2024). Blood-based biomarker composites have also been explored for distinguishing MDD from bipolar disorder in adolescents and adults, demonstrating the potential value of biological measures in differential diagnosis and clinical stratification (Huang et al., 2023). Neuroimaging studies have suggested that structural and functional alterations, particularly in fronto-insular and cerebellar systems, may help classify major depression, although translation into routine clinical screening remains challenging (Rizzi et al., 2025). Portable neurophysiological tools, including prefrontal EEG devices, have also been investigated for scalable mental health screening and gender-sensitive biomarker development (Tang et al., 2025). Collectively, these lines of evidence illustrate that depression prediction is moving toward

integrative models that combine symptom, behavioral, lifestyle, and biological data.

The microbiota-gut-brain axis provides another important example of the complexity of depressive pathophysiology. Mechanistic reviews have described how gut microbial composition, intestinal permeability, immune activation, vagal signaling, tryptophan metabolism, and microbial metabolites may contribute to depressive symptomatology (Akif & Islam, 2026). Microbial-based approaches to mental health have also considered the potential role of probiotics in depression treatment, suggesting that modulation of gut microbiota may influence mood through anti-inflammatory, neurochemical, and stress-regulatory mechanisms (Johnson et al., 2023). Exploratory analyses have extended this perspective to suicidal ideation in patients with and without MDD, linking microbiome-related variation to clinically severe outcomes within depressive conditions (Chen & Wu, 2025). Although microbiome data were not directly modeled in the present study, this literature is conceptually relevant because it reinforces the broader biopsychosocial and lifestyle-sensitive nature of depression. Nutrition, sleep, physical activity, inflammation, stress physiology, and gut microbiota may operate as interconnected systems rather than isolated predictors. Therefore, even when a study uses questionnaire-based indicators, behavioral and lifestyle variables can function as accessible proxies for deeper biological and psychosocial regulatory processes.

Comorbidity further complicates depression prediction and strengthens the argument for multidimensional assessment. MDD is frequently comorbid with physical diseases, and this comorbidity may be explained by shared inflammatory, metabolic, behavioral, endocrine, and psychosocial mechanisms (Berk et al., 2023). Depression in the context of chronic physical illness can differ from depression in otherwise healthy populations, particularly in fatigue, pain, disability, inflammation, and functional impairment. Comparative research on depression mechanisms in multiple sclerosis and rheumatoid arthritis illustrates how inflammatory and neurological processes may overlap with mood symptoms in medically complex populations (Wenger & Calabrese, 2021). Cancer-related research has similarly emphasized emotional, inflammatory, and genetic factors of resilience and vulnerability to depression, particularly among patients experiencing major health stressors and biological disruption (Oliveira et al., 2023). These studies show that depressive vulnerability is shaped not only by psychological symptoms but also by bodily condition, immune activity, stress burden, and

adaptive capacity. For predictive modeling, this means that lifestyle and behavioral variables may improve classification because they capture health-relevant patterns linked to both psychological and physical functioning.

Sex and gender differences are also relevant in MDD prediction. Depression prevalence, symptom expression, help-seeking behavior, stress exposure, hormonal influences, sleep patterns, and comorbid anxiety may differ by gender, and such differences may affect both risk estimation and model performance. Reviews of gender differences in MDD have highlighted the need for interventions and assessment approaches that are sensitive to gender-specific vulnerability and expression (Xiao, 2023). The importance of gender-sensitive biomarkers in scalable screening further reinforces the need to consider whether predictors operate similarly across demographic groups (Tang et al., 2025). Digital biomarkers may also contribute to more individualized and continuous monitoring of depression, particularly through passive or active indicators derived from smartphones, wearable devices, online behavior, speech, sleep patterns, and activity profiles (Vignapiano et al., 2023). However, the implementation of digital biomarkers requires careful attention to privacy, validity, interpretability, and clinical integration. In the present study, behavioral indicators such as screen time, sleep duration, and physical activity were included because they are feasible to collect in large samples and may reflect clinically meaningful lifestyle patterns associated with depressive risk.

Machine learning provides a methodological framework for integrating these heterogeneous predictors. Unlike traditional regression models, which often assume linearity, independence, and pre-specified relationships, machine learning algorithms can detect nonlinear effects, interactions, thresholds, and complex combinations of variables. Whole-person modeling approaches have already been reviewed in relation to suicide, demonstrating the value of machine learning for biopsychosocial prediction when multiple domains of risk are involved (Tio et al., 2024). Tree-based algorithms are particularly suitable for mental health prediction because they can accommodate mixed variable types, rank predictor importance, and model interactions without requiring extensive distributional assumptions. Random Forest models are especially useful because they aggregate multiple decision trees, reduce overfitting through bootstrap sampling, and provide robust predictive performance in high-dimensional data. Prior research using tree-based models to identify contributors to

negative affect further supports their applicability to depression-related psychological phenotypes (Cañizares et al., 2023). In addition, biomarker and neurophysiological studies increasingly use classification frameworks to distinguish diagnostic groups and identify predictive signatures, supporting the broader movement toward data-driven mental health screening (Huang et al., 2023; Rizzi et al., 2025; Tang et al., 2025).

At the same time, predictive modeling must remain clinically interpretable. A highly accurate model has limited practical utility if clinicians and researchers cannot understand which predictors drive classification. Random Forest models offer a partial solution through feature importance indices, permutation importance, and complementary interpretability methods such as SHapley Additive exPlanations. These approaches can clarify whether depressive classification is driven mainly by psychological distress, stress perception, sleep quality, physical inactivity, screen time, or other lifestyle factors. This is especially important in MDD because prediction should not merely assign risk; it should also identify modifiable targets for prevention and intervention. Translational frameworks for MDD have called for models that connect biomolecular mechanisms with spiraling risk-factor processes, emphasizing how interacting vulnerabilities may accumulate and reinforce depressive states over time (Baalen et al., 2025). Epigenetic perspectives similarly suggest that psychiatric risk and treatment response may be shaped by gene-environment interactions and dynamic biological embedding, which aligns with precision psychiatry goals (Mason & Miller, 2025). Even research outside depression, such as narrative work on biological, psychosocial, and microbial determinants of childhood-onset obsessive-compulsive disorder, highlights the broader psychiatric relevance of integrating biological, psychological, social, and microbial pathways within explanatory and predictive frameworks (Borrego-Ruiz & Borrego, 2025).

Despite growing evidence that MDD is multidimensional, many predictive studies remain limited by narrow predictor sets, insufficient attention to lifestyle variables, or inadequate interpretability. Some models focus primarily on symptoms, while others emphasize biomarkers that may be difficult to collect at scale. A practical screening model should balance clinical relevance, feasibility, and predictive performance by using variables that can be collected efficiently in community, primary care, educational, or workplace settings. Psychological variables

such as depression, anxiety, stress, perceived stress, and emotion regulation difficulties may identify core internal vulnerability, whereas behavioral and lifestyle indicators such as sleep quality, physical activity, screen time, diet-related patterns, and substance use may capture modifiable daily processes linked to depression onset and maintenance. This approach is consistent with the contemporary movement toward lifestyle-informed, biopsychosocial, and precision-oriented models of depression. It also addresses the clinical need for tools that can distinguish individuals with probable MDD from those without the disorder while simultaneously identifying the most influential contributors to risk.

The present study aimed to develop and evaluate a Random Forest classification model for predicting Major Depressive Disorder among Canadian adults using integrated psychological, behavioral, and lifestyle indicators.

## 2. Methods and Materials

### 2.1. Study Design and Participants

This study employed a cross-sectional, predictive machine learning design to develop and validate a Random Forest classification model for predicting Major Depressive Disorder (MDD) using a comprehensive set of psychological, behavioral, and lifestyle indicators. Unlike conventional inferential statistical approaches that focus primarily on hypothesis testing, the present study emphasized predictive accuracy by integrating multiple domains of risk factors into a supervised machine learning framework. The research was conducted across multiple provinces in Canada between January and September 2025. Participants were recruited through community health centers, university mailing lists, workplace wellness programs, and online advertisements distributed via social media platforms. The recruitment strategy aimed to ensure broad demographic representation across different age groups, educational levels, occupational categories, and socioeconomic backgrounds.

A total of 1,864 Canadian adults initially volunteered to participate in the study. Following data screening, exclusion of incomplete questionnaires, duplicate submissions, and participants who failed embedded attention-check questions, the final analytical sample consisted of 1,742 participants. Participants ranged in age from 18 to 69 years, with a mean age of 38.74 years (SD = 12.31). The sample included individuals from urban, suburban, and rural regions across

Canada, providing substantial diversity in demographic and psychosocial characteristics. Approximately 53% of participants identified as female, 46% as male, and 1% identified as non-binary or preferred not to disclose their gender.

Eligibility criteria included being at least 18 years of age, residing in Canada for a minimum of two years, possessing sufficient English language proficiency to complete the questionnaires, and providing informed consent before participation. Individuals with severe neurological disorders, cognitive impairments that could interfere with questionnaire completion, active psychotic disorders, or incomplete survey responses exceeding 10% missing data were excluded from the final analyses. Missing values for eligible cases were minimal and were handled using multiple imputation procedures prior to model development.

## 2.2. Measures

Major Depressive Disorder symptoms were assessed using the Patient Health Questionnaire-9 (PHQ-9), developed by Kroenke, Spitzer, and Williams in 2001. The PHQ-9 is one of the most widely used screening instruments for depressive symptoms in both clinical practice and epidemiological research. It consists of nine items corresponding directly to the diagnostic criteria for major depressive disorder described in the Diagnostic and Statistical Manual of Mental Disorders (DSM). Participants rate how frequently they have experienced each symptom during the previous two weeks using a four-point Likert scale ranging from 0 (Not at all) to 3 (Nearly every day). Total scores range from 0 to 27, with higher scores indicating greater depressive symptom severity. The PHQ-9 has consistently demonstrated excellent internal consistency, test-retest reliability, criterion validity, and diagnostic accuracy across numerous international studies, including Canadian populations. For the purpose of supervised machine learning classification, participants scoring 10 or higher were categorized as exhibiting clinically significant depressive symptoms indicative of probable Major Depressive Disorder.

Psychological characteristics were assessed using the Depression Anxiety Stress Scales-21 (DASS-21), originally developed by Lovibond and Lovibond in 1995. The DASS-21 contains twenty-one items organized into three seven-item subscales measuring depression, anxiety, and stress. Participants indicate the extent to which each statement applied to them during the previous week using a four-point

response format ranging from 0 (Did not apply to me at all) to 3 (Applied to me very much or most of the time). Subscale scores are multiplied by two to obtain equivalent scores to the original forty-two-item version. Higher scores reflect greater psychological distress. Numerous validation studies have confirmed the instrument's strong psychometric properties, including high internal consistency, convergent validity, discriminant validity, and factorial stability across diverse cultural and clinical populations.

Emotion regulation abilities were measured using the Difficulties in Emotion Regulation Scale-16 (DERS-16), developed by Bjureberg and colleagues in 2016 as a brief version of the original DERS. The instrument consists of sixteen items assessing six dimensions of emotion regulation difficulties, including nonacceptance of emotional responses, difficulties engaging in goal-directed behavior, impulse control difficulties, lack of emotional awareness, limited access to emotion regulation strategies, and lack of emotional clarity. Responses are recorded on a five-point Likert scale ranging from 1 (Almost never) to 5 (Almost always), with higher scores indicating greater emotion regulation difficulties. Previous investigations have consistently demonstrated excellent reliability, construct validity, and measurement invariance across adult populations.

Perceived stress was evaluated using the Perceived Stress Scale-10 (PSS-10), developed by Cohen, Kamarck, and Mermelstein in 1983. The instrument measures the extent to which individuals perceive situations in their lives as stressful, unpredictable, uncontrollable, and overwhelming during the previous month. The questionnaire contains ten items scored on a five-point Likert scale ranging from 0 (Never) to 4 (Very often). Four positively worded items are reverse scored before calculating the total score. Higher scores indicate greater perceived stress. Extensive psychometric evaluations have established excellent reliability, convergent validity, and predictive validity in community and clinical samples.

Sleep quality was measured using the Pittsburgh Sleep Quality Index (PSQI), introduced by Buysse and colleagues in 1989. The PSQI consists of nineteen self-report items organized into seven component scores assessing subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction. Component scores are summed to produce a global score ranging from 0 to 21, with higher scores representing poorer sleep quality. The instrument has demonstrated excellent reliability and validity in numerous

epidemiological and clinical investigations and is considered the international standard for measuring subjective sleep quality.

Physical activity was assessed using the International Physical Activity Questionnaire-Short Form (IPAQ-SF), developed by the International Physical Activity Questionnaire Research Committee in 2003. The questionnaire measures vigorous physical activity, moderate physical activity, walking behavior, and sedentary time during the previous seven days. Activity levels are converted into metabolic equivalent (MET)-minutes per week according to standardized scoring guidelines. The IPAQ-SF has been extensively validated internationally and has demonstrated acceptable reliability and criterion validity for population-based physical activity assessment.

Lifestyle behaviors were further evaluated through a structured health behavior questionnaire designed to assess smoking status, alcohol consumption, recreational drug use, dietary habits, daily fruit and vegetable intake, caffeine consumption, screen time, social media use, average working hours, and average nightly sleep duration. Body mass index (BMI) was calculated from self-reported height and weight. Demographic variables included age, gender, marital status, educational attainment, employment status, annual household income, province of residence, ethnicity, and history of mental health treatment. All questionnaire instruments employed in the study have demonstrated satisfactory psychometric characteristics in previous investigations, with established evidence supporting their validity and reliability across diverse adult populations.

### 2.3. Data analysis

Data preprocessing and predictive modeling were conducted using Python version 3.12. Descriptive statistics were first calculated to summarize demographic characteristics and variable distributions. Continuous variables were standardized where appropriate, while categorical variables were encoded using one-hot encoding techniques. Missing values representing less than 5% of the dataset were addressed through multiple imputation using chained equations. Highly correlated variables were examined to minimize multicollinearity, and feature engineering procedures were applied to improve predictive performance. The Random Forest algorithm served as the primary machine learning model because of its ability to model complex nonlinear relationships, manage high-dimensional datasets, accommodate mixed variable types,

reduce overfitting through bootstrap aggregation, and provide robust estimates of variable importance. The dataset was randomly divided into training (80%) and testing (20%) subsets while preserving the proportional distribution of depressive and non-depressive cases through stratified sampling. Hyperparameter optimization was performed using five-fold cross-validation combined with grid search to identify the optimal number of decision trees, maximum tree depth, minimum samples required for node splitting, minimum samples per leaf, and maximum number of predictor variables considered at each split.

Model performance was evaluated using multiple classification metrics, including overall accuracy, sensitivity (recall), specificity, precision, F1-score, balanced accuracy, and the area under the receiver operating characteristic curve (AUC-ROC). Confusion matrices were generated to assess classification performance across depressive and non-depressive groups. Feature importance scores based on Gini impurity reduction were calculated to identify the strongest psychological, behavioral, and lifestyle predictors contributing to model performance. To enhance interpretability, permutation importance analysis and SHapley Additive exPlanations (SHAP) values were additionally computed to quantify both global and individual predictor contributions. Statistical analyses for descriptive characteristics were conducted using IBM SPSS Statistics version 29, while all machine learning analyses were implemented using the Scikit-learn, Pandas, NumPy, and SHAP libraries in Python. Statistical significance for descriptive analyses was established at a two-tailed alpha level of 0.05.

### 3. Findings and Results

Data from 1,742 Canadian adults were included in the final analyses after the exclusion of incomplete questionnaires and responses failing quality-control procedures. Participants ranged in age from 18 to 69 years ( $M = 38.74$ ,  $SD = 12.31$ ). The sample consisted of 923 females (53.0%), 801 males (46.0%), and 18 participants (1.0%) identifying as non-binary or preferring not to disclose their gender. Approximately 61.5% of participants were employed full-time, 14.2% were employed part-time, 12.8% were university students, 7.4% were unemployed, and 4.1% were retired. Regarding marital status, 51.8% were married or living with a partner, 39.4% were single, and 8.8% were divorced, separated, or widowed. Educational attainment showed that 67.1% possessed a university degree, while

32.9% had completed secondary education or vocational training. Based on the PHQ-9 classification criterion, 487 participants (28.0%) were classified as presenting probable Major Depressive Disorder, whereas 1,255 participants

(72.0%) were categorized as non-depressed. The prevalence observed in this sample provided sufficient representation of both classes for supervised machine learning classification while minimizing substantial class imbalance.

**Table 1**

*Descriptive Statistics of Psychological, Behavioral, and Lifestyle Variables Included in the Random Forest Model*

| Variable                         | Mean    | SD      | Minimum | Maximum |
|----------------------------------|---------|---------|---------|---------|
| PHQ-9 Total Score                | 8.84    | 6.17    | 0       | 27      |
| DASS Depression                  | 11.46   | 8.32    | 0       | 42      |
| DASS Anxiety                     | 8.51    | 7.24    | 0       | 42      |
| DASS Stress                      | 14.27   | 8.11    | 0       | 42      |
| DERS-16 Total Score              | 42.75   | 12.41   | 16      | 80      |
| Perceived Stress Scale           | 18.61   | 6.84    | 1       | 39      |
| Pittsburgh Sleep Quality Index   | 7.18    | 3.82    | 0       | 20      |
| Physical Activity (MET-min/week) | 2416.38 | 1248.66 | 0       | 7920    |
| Daily Screen Time (hours)        | 6.24    | 2.41    | 1.0     | 15.3    |
| Sleep Duration (hours/night)     | 6.78    | 1.19    | 3.2     | 10.1    |
| Body Mass Index                  | 26.04   | 4.73    | 17.2    | 43.8    |

Table 1 presents the descriptive characteristics of all continuous variables entered into the Random Forest classification model. Average depressive symptom severity remained below the clinical cut-off for the overall sample; however, considerable variability was observed across participants, indicating substantial heterogeneity in mental health status. Indicators of emotional distress, including depression, anxiety, stress, perceived stress, and emotion regulation difficulties, demonstrated relatively wide standard deviations, suggesting that participants represented a broad spectrum of psychological functioning. Sleep quality scores indicated mild-to-moderate sleep disturbances on

average, while the mean nightly sleep duration was below the generally recommended seven to nine hours for adults. Physical activity demonstrated substantial variability, reflecting marked differences in exercise habits among participants. Screen time averaged more than six hours per day, suggesting considerable engagement with digital devices. Collectively, these descriptive statistics indicate adequate variability across predictors, providing a favorable basis for machine learning classification and reducing the likelihood of ceiling or floor effects that could adversely influence predictive performance.

**Table 2**

*Performance of the Random Forest Classification Model on the Independent Test Dataset*

| Performance Metric               | Value  |
|----------------------------------|--------|
| Accuracy                         | 91.38% |
| Sensitivity (Recall)             | 89.12% |
| Specificity                      | 92.31% |
| Precision                        | 87.64% |
| F1-Score                         | 88.37% |
| Balanced Accuracy                | 90.72% |
| ROC-AUC                          | 0.957  |
| Matthews Correlation Coefficient | 0.806  |
| Cohen's Kappa                    | 0.804  |

The Random Forest classifier demonstrated excellent predictive performance when evaluated using the independent testing dataset. Overall classification accuracy reached 91.38%, indicating that the model correctly

classified more than nine out of every ten participants. Sensitivity of 89.12% demonstrated that the algorithm successfully identified the majority of individuals presenting probable Major Depressive Disorder while minimizing

false-negative classifications. Equally important, specificity reached 92.31%, reflecting a high capability for correctly identifying participants without depressive disorder and minimizing false-positive predictions. Precision exceeded 87%, indicating that most individuals classified as depressed by the algorithm genuinely belonged to the depressed category. The F1-score of 88.37% reflected an excellent balance between precision and recall, while the balanced accuracy confirmed consistent performance across both

diagnostic groups despite unequal class frequencies. The receiver operating characteristic analysis produced an area under the curve (ROC-AUC) of 0.957, demonstrating outstanding discrimination between depressed and non-depressed participants. Likewise, Cohen's Kappa and the Matthews Correlation Coefficient exceeded 0.80, indicating excellent agreement beyond chance and confirming that the Random Forest model achieved highly reliable and clinically meaningful classification performance.

**Table 3**

*Relative Importance of Predictors in the Random Forest Model*

| Rank | Predictor                                 | Relative Importance |
|------|---|---------------------|
| 1    | DASS Depression                           | 0.184               |
| 2    | Perceived Stress                          | 0.152               |
| 3    | Emotion Regulation Difficulties (DERS-16) | 0.137               |
| 4    | Sleep Quality (PSQI)                      | 0.122               |
| 5    | DASS Anxiety                              | 0.094               |
| 6    | DASS Stress                               | 0.086               |
| 7    | Daily Screen Time                         | 0.061               |
| 8    | Physical Activity                         | 0.055               |
| 9    | Sleep Duration                            | 0.042               |
| 10   | Body Mass Index                           | 0.033               |
| 11   | Alcohol Consumption                       | 0.021               |
| 12   | Age                                       | 0.013               |

The feature importance analysis demonstrated that psychological variables contributed most substantially to model performance. Depression symptom severity measured by the DASS emerged as the strongest predictor, accounting for 18.4% of the overall model importance. Perceived stress represented the second most influential predictor, followed closely by difficulties in emotion regulation and subjective sleep quality. Together, these four variables explained nearly 60% of the predictive information extracted by the Random Forest algorithm, emphasizing the central role of emotional distress and adaptive self-regulation in the identification of Major Depressive Disorder. Anxiety and stress symptoms also contributed meaningful predictive information, although to a lesser extent than depressive symptoms and

perceived stress. Lifestyle indicators, including screen time, physical activity, sleep duration, body mass index, and alcohol consumption, demonstrated moderate yet meaningful contributions, suggesting that behavioral habits enhanced predictive accuracy when combined with psychological variables. Age contributed relatively little additional predictive information, indicating that depressive symptom prediction depended primarily upon modifiable psychological and behavioral characteristics rather than demographic factors. Overall, the ranking of feature importance supports the multidimensional conceptualization of depression and demonstrates that integrating psychological, behavioral, and lifestyle indicators substantially strengthens predictive modeling.

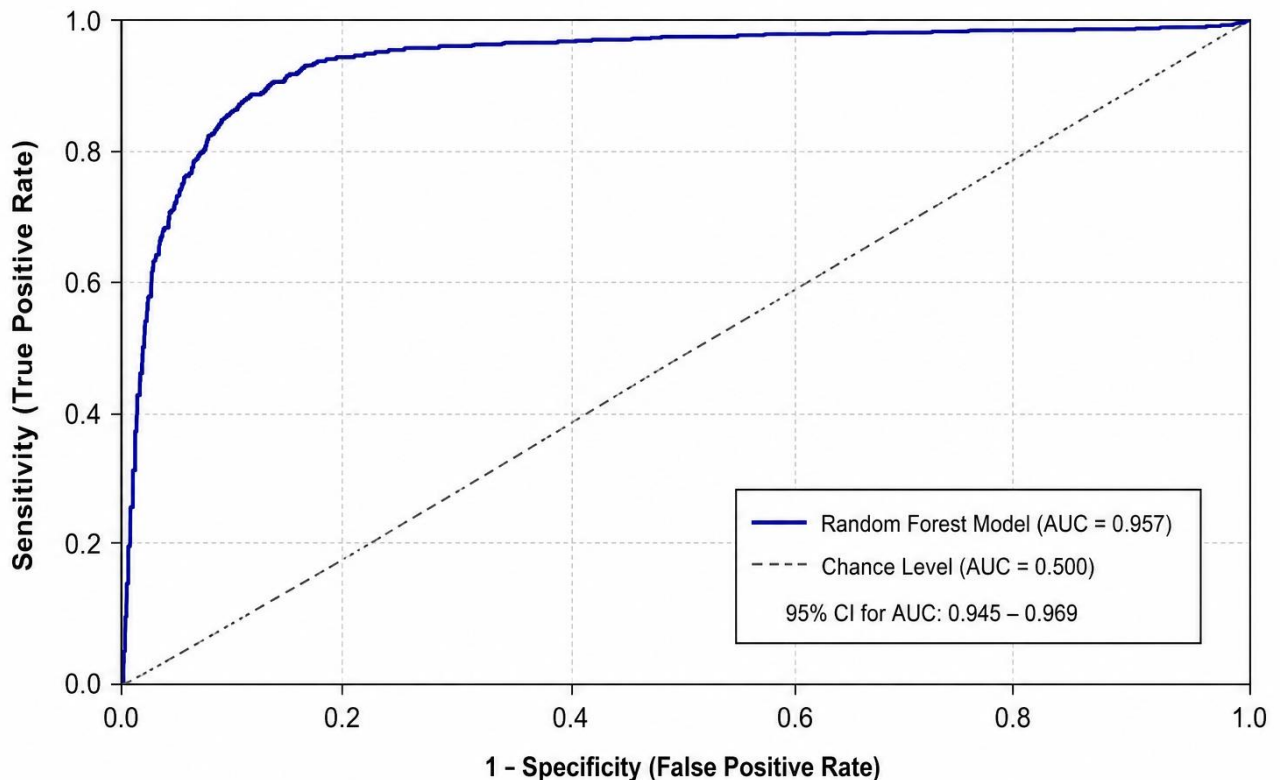
**Table 4**

*Five-Fold Cross-Validation Performance of the Random Forest Model*

| Fold               | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) | ROC-AUC |
|--------------------|--------------|---------------|------------|--------------|---------|
| Fold 1             | 90.87        | 87.16         | 88.42      | 87.79        | 0.951   |
| Fold 2             | 91.62        | 88.11         | 89.46      | 88.78        | 0.960   |
| Fold 3             | 91.03        | 87.52         | 88.94      | 88.22        | 0.954   |
| Fold 4             | 92.11        | 88.93         | 89.81      | 89.37        | 0.963   |
| Fold 5             | 91.26        | 87.71         | 89.05      | 88.37        | 0.958   |
| Mean               | 91.38        | 87.89         | 89.14      | 88.51        | 0.957   |
| Standard Deviation | 0.49         | 0.69          | 0.53       | 0.61         | 0.005   |

Figure 1

Receiver Operating Characteristic (ROC) Curve of the Optimized Random Forest Model for Predicting Major Depressive Disorder



Five-fold cross-validation further demonstrated the robustness and generalizability of the Random Forest classifier. Accuracy remained remarkably stable across all validation folds, varying by less than 1.3 percentage points. Similarly, precision, recall, and F1-score showed only minimal fluctuations, reflecting consistent model performance regardless of the subset used for validation. The mean ROC-AUC remained at 0.957 with an extremely small standard deviation of only 0.005, indicating exceptional discrimination that generalized reliably across independent validation samples. Such consistency suggests that the model did not exhibit substantial overfitting during training and maintained stable predictive capability when exposed to previously unseen observations. The low variability observed across all evaluation metrics supports the reliability of the optimized Random Forest algorithm and provides strong evidence that the predictive performance achieved in the testing dataset is likely to be reproducible in comparable Canadian adult populations.

The receiver operating characteristic curve illustrated excellent discriminatory performance of the optimized Random Forest classifier. The ROC curve closely followed the upper-left corner of the coordinate system, reflecting

simultaneously high sensitivity and specificity across a wide range of classification thresholds. The area under the curve of 0.957 indicated outstanding diagnostic accuracy, substantially exceeding the conventional threshold of 0.90 considered indicative of excellent classification performance. Visual inspection of the ROC curve further demonstrated that the classifier maintained strong sensitivity without sacrificing specificity, thereby reducing both false-positive and false-negative classifications. This finding supports the clinical applicability of the proposed machine learning model as a reliable screening tool capable of accurately distinguishing individuals at elevated risk for Major Depressive Disorder using an integrated combination of psychological, behavioral, and lifestyle characteristics.

#### 4. Discussion

The present study developed and evaluated a Random Forest classification model for predicting Major Depressive Disorder among Canadian adults using psychological, behavioral, and lifestyle indicators. The findings demonstrated that the optimized model achieved excellent classification performance, with an accuracy of 91.38%,

sensitivity of 89.12%, specificity of 92.31%, precision of 87.64%, F1-score of 88.37%, balanced accuracy of 90.72%, and ROC-AUC of 0.957. These results indicate that the Random Forest model was highly capable of distinguishing participants with probable Major Depressive Disorder from those without depressive disorder. The high sensitivity suggests that the model successfully identified most individuals with clinically significant depressive symptoms, whereas the high specificity indicates that it also minimized false-positive classifications among non-depressed participants. This balance is particularly important for mental health screening because false negatives may delay intervention for individuals in need of care, while false positives may create unnecessary concern or inefficient use of clinical resources. The strong Matthews Correlation Coefficient and Cohen's Kappa values further confirmed that the model performed well beyond chance and maintained robust agreement between predicted and observed classifications. These findings are consistent with the growing view that depression prediction benefits from data-driven and multidimensional approaches capable of capturing complex patterns that may be missed by conventional linear models (Cañizares et al., 2023; Tio et al., 2024).

The feature importance results showed that psychological variables were the most influential predictors of Major Depressive Disorder. DASS depression scores emerged as the strongest predictor, followed by perceived stress, emotion regulation difficulties, sleep quality, anxiety, and stress. This pattern is theoretically coherent because MDD is fundamentally characterized by disturbances in mood, cognition, motivation, stress regulation, and affective control. The dominant role of depressive symptom severity was expected, but the strong contributions of perceived stress and emotion regulation difficulties suggest that depression prediction cannot be reduced to symptom severity alone. Perceived stress may intensify depressive vulnerability by increasing cognitive load, emotional exhaustion, rumination, hopelessness, and physiological arousal. Similarly, emotion regulation difficulties may undermine adaptive coping, increase affective instability, and prolong negative mood states. These findings align with contemporary perspectives that conceptualize MDD as a multidimensional disorder involving psychological distress, maladaptive emotional processing, and complex interactions among cognitive-affective risk factors (Stoyanov & Maes, 2021; Zhang, 2024). They also correspond with research emphasizing the importance of negative affect and tree-

based modeling for identifying psychological contributors to depression-related outcomes (Cañizares et al., 2023).

The high predictive contribution of perceived stress, anxiety, and general stress also supports a transdiagnostic understanding of depression. Major Depressive Disorder often overlaps with anxiety symptoms, chronic stress exposure, irritability, sleep disturbance, fatigue, and motivational impairment. These overlapping dimensions may reflect shared mechanisms such as hyperactivation of stress-response systems, threat sensitivity, inflammatory signaling, and impaired regulation of emotional salience. The present model's ability to incorporate these interrelated predictors is a key strength of the Random Forest approach, because tree-based models are able to identify nonlinear and interactive patterns among variables without assuming that each predictor contributes independently. This is consistent with precision psychiatry perspectives suggesting that MDD should be understood through longitudinal trajectories, heterogeneous symptom profiles, and individualized combinations of vulnerability factors rather than as a uniform diagnostic entity (Atagün, 2025). It also reflects broader calls for neuroscience-informed and nomothetic network models of psychiatric classification, in which depressive disorder is conceptualized as an interconnected system of symptoms, biological processes, behavioral patterns, and environmental influences (Stoyanov & Maes, 2021).

Sleep quality was the strongest behavioral predictor in the model and ranked among the most important variables overall. This finding is consistent with extensive evidence that sleep disturbance is both a symptom and a risk factor for depression. Poor sleep may increase vulnerability to depressive disorder through impaired emotional regulation, reduced reward responsiveness, increased inflammatory activity, circadian disruption, daytime fatigue, and reduced coping capacity. Conversely, depressive symptoms can worsen insomnia, reduce sleep efficiency, and disturb daily rhythms, creating a self-reinforcing cycle. The present findings support the view that sleep quality should be considered a central indicator in depression screening rather than a secondary or peripheral symptom. This is consistent with research showing that sleep and depression are strongly linked across developmental stages, including adolescence, and that sleep-related factors can meaningfully inform depression risk assessment (Castiglione-Fontanellaz & Tarokh, 2023). The importance of sleep in the present model is also compatible with evidence that seasonal and mood fluctuations may be predicted from self-report and

physiological indicators, reinforcing the relevance of daily rhythm and sleep-related variables in mood prediction (Höller et al., 2022).

Lifestyle variables, including screen time, physical activity, sleep duration, body mass index, and alcohol consumption, contributed less strongly than psychological indicators but still added meaningful predictive value. This finding supports the biopsychosocial and lifestyle-informed interpretation of MDD. Depression is not only expressed through subjective distress but also through behavioral withdrawal, sedentary routines, disrupted sleep-wake cycles, changes in appetite and weight, reduced social engagement, and altered health behaviors. Physical activity is particularly important because exercise has been identified as a therapeutic modality for both prevention and treatment of depression, with potential effects on neuroplasticity, inflammation, self-efficacy, behavioral activation, and cardiovascular functioning (Blumenthal & Rozanski, 2023). The present finding that physical activity contributed to classification supports lifestyle psychiatry frameworks emphasizing movement, sleep, nutrition, stress management, and social connection as modifiable determinants of depression risk and recovery (Merlo et al., 2026). Although lifestyle variables did not surpass psychological distress indicators in importance, their contribution suggests that screening models may become more clinically useful when they incorporate both internal psychological states and observable daily behaviors.

The inclusion and predictive relevance of lifestyle indicators also correspond with growing evidence linking nutrition, inflammation, microbiota, and mood. Although the present study did not directly measure dietary biomarkers or microbiome composition, lifestyle behaviors such as physical activity, sleep, body mass index, and alcohol consumption may partly reflect broader regulatory systems associated with metabolic, inflammatory, and microbial pathways. Recent work has emphasized the microbiota-gut-brain axis as a mechanistic pathway in MDD, linking gut microbial composition, immune activity, neurotransmitter metabolism, and stress regulation to depressive symptoms (Akif & Islam, 2026). Microbial-based approaches, including probiotics, have been discussed as potential adjunctive strategies for depression treatment (Johnson et al., 2023), and exploratory findings have suggested associations between microbiome-related variation and suicidal ideation among individuals with and without MDD (Chen & Wu, 2025). Nutritional research similarly suggests that higher intake of specific bioactive compounds, such as

anthocyanins, may be associated with lower depressive symptoms (Mestrom et al., 2023), while omega-3 fatty acids have been discussed as potentially more relevant for psychological health than for cognitive outcomes (Kossowska-Wywiiał & Brzezicka, 2025). These findings help explain why lifestyle indicators may improve MDD prediction even when the model is not directly biological.

The model's strong ROC-AUC of 0.957 indicates excellent discrimination and suggests that the combined predictor set captured clinically meaningful differences between depressed and non-depressed participants. This result is consistent with the increasing use of machine learning for whole-person modeling in psychiatry, where predictive algorithms are used to integrate psychological, behavioral, biological, and contextual risk indicators (Tio et al., 2024). It also aligns with research using digital biomarkers, neuroimaging, EEG, and blood-based measures to improve classification of MDD and related psychiatric conditions (Huang et al., 2023; Rizzi et al., 2025; Tang et al., 2025; Vignapiano et al., 2023). However, the present study contributes to this literature by demonstrating that high predictive performance can be achieved using feasible, questionnaire-based psychological, behavioral, and lifestyle indicators. This is clinically important because many biomarker-based systems are costly, technically demanding, or not yet ready for routine screening, whereas self-report and behavioral variables can be collected efficiently in community, university, workplace, and primary care settings. The findings therefore support the potential value of scalable screening models that combine clinical relevance with practical feasibility.

The importance of psychological and lifestyle predictors also supports translational frameworks that conceptualize MDD as a spiraling process in which biological vulnerability, stress exposure, emotional dysregulation, behavioral disruption, and lifestyle deterioration reinforce one another over time (Baalen et al., 2025). In this interpretation, depressive symptoms may lead to poorer sleep, less physical activity, more screen time, social withdrawal, and unhealthy routines, while these behaviors may in turn intensify mood symptoms and reduce resilience. Immune and inflammatory mechanisms may further connect psychological stress and lifestyle disruption with depressive disorder. Psychoneuroimmunoendocrinology research has emphasized the role of immune dysfunction in mood disorders (Ortega et al., 2024), and inflammatory biomarkers such as the neutrophil-to-lymphocyte ratio have been discussed as potential indicators for early depression

detection (Ghafori et al., 2024). Similarly, research on natural antidepressants, neuroinflammation, diet, and flora suggests that psychological health may be shaped by interactions among immune activity, nutrition, and microbial balance (Frota et al., 2025; Marano, 2025). The present findings are consistent with this integrated view because they show that both psychological distress and modifiable behavioral patterns are relevant to predicting probable MDD.

The findings also have implications for precision psychiatry. The model did not rely on a single predictor but instead used a constellation of factors, which reflects the heterogeneity of MDD. Some participants may present primarily with emotional distress and stress reactivity, while others may show dominant sleep disturbance, inactivity, increased screen time, or lifestyle dysregulation. Precision psychiatry requires identifying these individual risk profiles and linking them to tailored interventions. This aligns with recent work on epigenetics and psychiatry, which emphasizes dynamic gene-environment interactions, candidate mechanisms, and the future possibility of more personalized psychiatric care (Mason & Miller, 2025). It also corresponds with research showing that depression may be shaped by sex and gender differences in prevalence, presentation, and intervention needs (Xiao, 2023). Although gender was not among the strongest predictors in the present model, future model development should examine whether predictor importance differs across gender groups. This is especially relevant given emerging work on gender-sensitive biomarkers and scalable screening using portable neurophysiological devices (Tang et al., 2025).

Another important interpretation concerns the relationship between MDD and broader physical health. Depression is highly comorbid with physical diseases, and this comorbidity is shaped by overlapping behavioral, inflammatory, endocrine, metabolic, and psychosocial mechanisms (Berk et al., 2023). The predictive relevance of body mass index, physical activity, sleep, and alcohol consumption in the present model may therefore reflect the embeddedness of depression within whole-person health. Studies comparing depression mechanisms in multiple sclerosis and rheumatoid arthritis also suggest that inflammatory and disease-related processes may contribute to depressive symptoms in medically complex populations (Wenger & Calabrese, 2021). Similarly, research on breast cancer has emphasized emotional, inflammatory, and genetic factors associated with resilience and vulnerability to depression (Oliveira et al., 2023). These studies support

the interpretation that depression prediction should include variables reflecting both mental and physical functioning. While the present study focused on adults from the general Canadian population, the findings suggest that future predictive models may benefit from incorporating chronic disease status, inflammatory markers, medication use, and functional impairment.

## 5. Conclusion

Finally, the present results support the methodological suitability of Random Forest modeling for MDD prediction. Random Forest algorithms are particularly useful when predictors are numerous, interrelated, and potentially nonlinear. The stable five-fold cross-validation results indicated that the model was not overly dependent on a single data split and that predictive performance generalized across validation folds. This stability is important because mental health datasets often contain heterogeneous individuals and complex covariance patterns. The present findings are also compatible with the broader movement toward integrating digital biomarkers, biological indicators, psychological questionnaires, and behavioral data into clinically interpretable prediction systems (Tio et al., 2024; Vignapiano et al., 2023). At the same time, the results suggest that even before advanced biomarker systems become widely available, clinically meaningful prediction can be achieved by systematically modeling psychological distress, stress, emotion regulation, sleep, and lifestyle behavior. Therefore, the study contributes to the literature by demonstrating a feasible and accurate approach to MDD screening in a Canadian adult sample.

## 6. Limitations & Suggestions

This study has several limitations that should be acknowledged. First, the cross-sectional design prevents causal interpretation, and the model should therefore be understood as a classification tool rather than evidence of temporal or etiological pathways. Second, Major Depressive Disorder classification was based on questionnaire scores rather than structured clinical interviews, which may have introduced misclassification in some cases. Third, all psychological, behavioral, and lifestyle variables were self-reported, making the data vulnerable to recall bias, social desirability bias, and subjective interpretation. Fourth, although the sample was diverse and Canadian, it may not fully represent all cultural, linguistic, Indigenous, rural, immigrant, or clinically diagnosed populations in Canada.

Fifth, the model was internally validated using cross-validation and an independent test subset, but external validation in a completely independent sample was not conducted.

Future studies should use longitudinal designs to examine whether psychological, behavioral, and lifestyle indicators predict the future onset, persistence, recurrence, or remission of Major Depressive Disorder. Researchers should also validate the present model in independent clinical and community samples, including primary care patients, adolescents, older adults, and individuals with chronic physical illness. Future models may be strengthened by incorporating objective behavioral indicators from wearable devices, sleep trackers, smartphones, or ecological momentary assessment, as well as biological markers related to inflammation, endocrine function, microbiome composition, and metabolic health. Additional work should examine model fairness across gender, age, socioeconomic status, ethnicity, and geographic region to ensure that machine learning tools do not reproduce hidden bias or perform unequally across population subgroups.

The findings suggest that mental health screening may be improved by combining psychological indicators with behavioral and lifestyle information rather than relying only on depressive symptom severity. Clinicians, counselors, and primary care providers should pay particular attention to perceived stress, emotion regulation difficulties, sleep quality, anxiety symptoms, physical inactivity, excessive screen time, and disrupted sleep duration when assessing depression risk. Random Forest-based screening tools may be useful as decision-support instruments in community health, university counseling centers, workplace wellness programs, and primary care settings, provided that they are used to support rather than replace clinical judgment. Preventive and therapeutic programs should also address modifiable lifestyle factors, especially sleep, exercise, daily routine, and stress management, because these variables may represent practical intervention targets for individuals at elevated risk of Major Depressive Disorder.

### Acknowledgments

We would like to express our appreciation and gratitude to all those who cooperated in carrying out this study.

### Declaration of Interest

The authors of this article declared no conflict of interest.

### Ethical Considerations

The study protocol adhered to the principles outlined in the Helsinki Declaration, which provides guidelines for ethical research involving human participants.

### Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

### Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

### Authors' Contributions

All authors equally contributed in this article.

### Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

### References

- Akif, A., & Islam, M. R. (2026). The Microbiota-Gut-Brain Axis in the Pathophysiology of Major Depressive Disorder: A Mechanistic Review. *Comprehensive Physiology*, 16(1). <https://doi.org/10.1002/cph4.70100>
- Atagün, M. İ. (2025). Longitudinal Trajectories of Major Depressive Disorder Provide Further Clinical Perspectives for Precision Psychiatry. *Psychiatry and Clinical Neurosciences Reports*, 4(4). <https://doi.org/10.1002/pcn5.70221>
- Baalen, M. v., Velden, L. v. d., Gronde, T. v. d., & Pieters, T. (2025). Developing a Translational Research Framework for MDD: Combining Biomolecular Mechanisms With a Spiraling Risk Factor Model. *Frontiers in Psychiatry*, 15. <https://doi.org/10.3389/fpsyt.2024.1463929>
- Berk, M., Köhler-Forsberg, O., Turner, M., Penninx, B. W., Wrobel, A., Firth, J., Loughman, A., Reavley, N., McGrath, J. J., Momen, N. C., Plana-Ripoll, O., O'Neil, A., Siskind, D., Williams, L. J., Carvalho, A. F., Schmaal, L., Walker, A. J., Dean, O., Walder, K., . . . Marx, W. (2023). Comorbidity Between Major Depressive Disorder and Physical Diseases: A Comprehensive Review of Epidemiology, Mechanisms and Management. *World Psychiatry*, 22(3), 366-387. <https://doi.org/10.1002/wps.21110>
- Blumenthal, J. A., & Rozanski, A. (2023). Exercise as a Therapeutic Modality for the Prevention and Treatment of Depression. *Progress in Cardiovascular Diseases*, 77, 50-58. <https://doi.org/10.1016/j.pcad.2023.02.008>
- Borrego-Ruiz, A., & Borrego, J. J. (2025). Biological, Psychosocial, and Microbial Determinants of Childhood-Onset Obsessive-Compulsive Disorder: A Narrative Review.

- Children*, 12(8), 1063. <https://doi.org/10.3390/children12081063>
- Cañizares, C., Gómez, Y., Ferro, E., Torres, C. A., Agudelo, D., & Odom, G. (2023). Using Tree-Based Models to Identify Factors Contributing to Trait Negative Affect in Adults With and Without Major Depression. <https://doi.org/10.21203/rs.3.rs-2978274/v1>
- Castiglione-Fontanellaz, C. E. G., & Tarokh, L. (2023). Sleep and Adolescent Depression. *Clinical and Translational Neuroscience*, 8(1), 3. <https://doi.org/10.3390/ctn8010003>
- Chen, V. C., & Wu, S. I. (2025). An Exploratory Analysis on the Association Between Suicidal Ideation and the Microbiome in Patients With or Without Major Depressive Disorder. *Journal of affective disorders*, 370, 362-372. <https://doi.org/10.1016/j.jad.2024.10.120>
- Frota, F. F., Araújo, L. P., Valenti, V. E., Eliana de Souza Bastos Mazuqueli, P., Detregiach, C. R. P., Galhardi, C. M., Caracio, F. C. C., Laurindo, L. F., Tanaka, M., & Barbalho, S. M. (2025). Neuroinflammation and Natural Antidepressants: Balancing Fire With Flora. *Biomedicines*, 13(5), 1129. <https://doi.org/10.3390/biomedicines13051129>
- Ghafari, S. S., Yousefi, Z., Bakhtiari, E., mohammad hossein mohammadi mahdiabadi, h., & Hassanzadeh, G. (2024). Neutrophil-to-Lymphocyte Ratio as a Predictive Biomarker for Early Diagnosis of Depression: A Narrative Review. *Brain Behavior & Immunity - Health*, 36, 100734. <https://doi.org/10.1016/j.bbih.2024.100734>
- Höller, Y., Urbschat, M. M., Kristófersson, G. K., & Olafsson, R. (2022). Predictability of Seasonal Mood Fluctuations Based on Self-Report Questionnaires and EEG Biomarkers in a Non-Clinical Sample. *Frontiers in Psychiatry*, 13. <https://doi.org/10.3389/fpsy.2022.870079>
- Huang, J., Hou, X., Li, M., Xue, Y., An, J., Wen, S., Wang, Z., Cheng, M., & Yue, J. (2023). A Composite of Blood-Based Biomarkers to Distinguish Major Depressive Disorder and Bipolar Disorder in Adolescents and Adults. <https://doi.org/10.21203/rs.3.rs-3058571/v1>
- Johnson, D., Letchumanan, V., Thum, C., Thurairajasingam, S., & Lee, L. H. (2023). A Microbial-Based Approach to Mental Health: The Potential of Probiotics in the Treatment of Depression. *Nutrients*, 15(6), 1382. <https://doi.org/10.3390/nu15061382>
- Kossowska-Wywił, M., & Brzezicka, A. (2025). Nourishing the Brain or the Mood? Dietary Omega-3s for Psychological, but Not Cognitive Health. *Nutrients*, 18(1), 50. <https://doi.org/10.3390/nu18010050>
- Marano, G. (2025). The Immune Mind: Linking Dietary Patterns, Microbiota, and Psychological Health. *Nutrients*, 18(1), 96. <https://doi.org/10.3390/nu18010096>
- Mason, C. E., & Miller, J. J. (2025). The Clinical Use of Epigenetics in Psychiatry: A Narrative Review of Epigenetic Mechanisms, Key Candidate Genes, and Precision Psychiatry. *Frontiers in Psychiatry*, 16. <https://doi.org/10.3389/fpsy.2025.1671122>
- Merlo, G., Sugden, S., Rosenfeld, R. M., Baron, D., Karlsen, M., Keyes, S. A., McHugh, J., Miller, L. A., Nemeroff, C. B., Ramas, M.-E., Livingston, K. A., Williams, K. A., Wilson, K. P., Wong, W., & Viswanathan, R. (2026). Lifestyle Interventions for Major Depressive Disorder (MDD): An Expert Consensus Statement From the American College of Lifestyle Medicine. *American Journal of Lifestyle Medicine*, 20(4), 608-627. <https://doi.org/10.1177/15598276251408353>
- Mestrom, A., Charlton, K., Thomas, S. J., Larkin, T., Walton, K., Elgellaie, A., & Kent, K. (2023). Higher Anthocyanin Intake Is Associated With Lower Depressive Symptoms in Adults With and Without Major Depressive Disorder. *Food Science & Nutrition*, 12(3), 2202-2209. <https://doi.org/10.1002/fsn3.3850>
- Oliveira, M. A., Medeiros, R., Guerra, M. P., Pariente, C. M., & Fernandes, L. (2023). Emotional, Inflammatory, and Genetic Factors of Resilience and Vulnerability to Depression in Patients With Premenopausal Breast Cancer: A Longitudinal Study Protocol. *PLoS One*, 18(2), e0279344. <https://doi.org/10.1371/journal.pone.0279344>
- Ortega, M. Á., Fraile-Martinez, O., García-Montero, C., Díaz, R., López-González, L., Monserrat, J., Barrena-Blázquez, S., Álvarez-Mon, M. Á., Lahera, G., & Álvarez-Mon, M. (2024). Understanding Immune System Dysfunction and Its Context in Mood Disorders: Psychoneuroimmunoenocrinology and Clinical Interventions. *Military Medical Research*, 11(1). <https://doi.org/10.1186/s40779-024-00577-w>
- Rizzi, R., Jornkokgoud, K., Ghomroudi, P. A., Stella, M., & Grecucci, A. (2025). The Dark Side of the Mood: Structural and Functional Fronto-Insular and Cerebellar Alterations Classify Major Depression. <https://doi.org/10.1101/2025.04.09.25325506>
- Serretti, A. (2025). Anhedonia: Current and Future Treatments. *Psychiatry and Clinical Neurosciences Reports*, 4(1). <https://doi.org/10.1002/pcn5.70088>
- Stoyanov, D., & Maes, M. H. (2021). How to Construct Neuroscience-Informed Psychiatric Classification? Towards Nomothetic Networks Psychiatry. *World journal of psychiatry*, 11(1), 1-12. <https://doi.org/10.5498/wjp.v11.i1.1>
- Tang, S., Han, C., & Li, X. (2025). Toward Scalable Mental Health Screening: Gender-Sensitive Biomarkers From Portable Prefrontal EEG Devices. <https://doi.org/10.21203/rs.3.rs-7197377/v1>
- Tio, E. S., Misztal, M., & Felsky, D. (2024). Evidence for the Biopsychosocial Model of Suicide: A Review of Whole Person Modeling Studies Using Machine Learning. *Frontiers in Psychiatry*, 14. <https://doi.org/10.3389/fpsy.2023.1294666>
- Vignapiano, A., Monaco, F., Pagano, C., Piacente, M., Farina, F., Petrillo, G., Sica, R., Marenga, A., Shin, J. I., Solmi, M., & Corrivetti, G. (2023). A Narrative Review of Digital Biomarkers in the Management of Major Depressive Disorder and Treatment-Resistant Forms. *Frontiers in Psychiatry*, 14. <https://doi.org/10.3389/fpsy.2023.1321345>
- Wenger, A., & Calabrese, P. (2021). Comparing Underlying Mechanisms of Depression in Multiple Sclerosis and Rheumatoid Arthritis. *Journal of Integrative Neuroscience*, 20(3). <https://doi.org/10.31083/jjin2003081>
- Xiao, J. (2023). Gender Differences in Major Depressive Disorder and Relevant Interventions. *Lecture Notes in Education Psychology and Public Media*, 3(1), 356-361. <https://doi.org/10.54254/2753-7048/3/2022502>
- Zhang, J. (2024). The Analysis of Major Depressive Disorder. *Theoretical and Natural Science*, 63(1), 133-137. <https://doi.org/10.54254/2753-8818/2024.17937>