Salud, Ciencia y Tecnología. 2025; 5:1970
doi: 10.56294/saludcyt20251970

ORIGINAL

Random Forest modeling of bipolar affective disorder in Ecuador

Modelización mediante bosques aleatorios del trastorno afectivo bipolar en Ecuador

Cristhian Ismael Gómez Gaona[1], Andrea del Rocío Mejía Rubio[2], José Rubén León Pérez[3], Jesús Rodríguez[4], Zilma Diago Alfes[5], Laura Esther Muñoz Escobar[5], Cristian Inca[6], Jimmy Yaguana Torres[7], Marco Velasco Arellano[5]

[1]Universidad Estatal de Milagro, Ecuador.
[2]Universidad Técnica de Ambato, Ecuador.
[3]Universidad Autónoma de Tlaxcala. Campus Calpulalpan, México.
[4]Universidad UTE, Ecuador.
[5]Universidad Nacional de Chimborazo, Ecuador.
[6]Escuela Superior Politécnica de Chimborazo (ESPOCH), Ecuador.
[7]Clínica SurHospital, Ecuador.

Cite as: Gómez Gaona CI, Mejía Rubio A del R, León Pérez JR, Rodríguez J, Diago Alfes Z, Muñoz Escobar LE, et al. Random Forest modeling of bipolar affective disorder in Ecuador. Salud, Ciencia y Tecnología. 2025;5:1970. https://doi.org/10.56294/saludcyt20251970

Submitted: 09-02-2025 **Revised:** 27-05-2025 **Accepted:** 07-08-2025 **Published:** 08-08-2025

Editor: Prof. Dr. William Castillo-González

Corresponding author: Cristian Inca

ABSTRACT

Bipolar affective disorder is a mental disorder characterized by depressive and manic or hypomanic episodes. The complexity of the diagnosis of bipolar affective disorder due to the overlapping of its symptoms with other mood disorders led researchers and doctors to search for new and advanced techniques for the precise detection of bipolar affection disorder. One of these methods is the use of advanced machine learning algorithms under a statistical methodology for building logistical regression models, Random Forest. Support vector machines, Decision Tree, K-Nearest Neighbors, and Gradient Boosting, with 146 data collected from the psychiatric services affiliated with the mental health system of Ecuador. At the inferential level, the results suggest that the implementation of automatic algorithms based on the different methodologies for building models enables the successful prediction or classification of individuals with bipolar affective disorders in Ecuador compared to controlled patients who do not profile under this pathological picture. It is the best Random Forest statistical model (89,35 %) that dictates the best performance metrics compared to the Gradient Boosting model. The evolution of the overall prevalence of bipolar affective disorders in Ecuador over the past 22 years has increased by a small differential. However, from 2020 to 2022, there has been a considerable increase in the percentage prevalence of cases of bipolar affective disorders in Ecuador.

Keywords: Algorithms; Machine Learning; Decision Tree Gradient Boosting; Random Forest.

RESUMEN

El trastorno afectivo bipolar representa un trastorno mental caracterizado por episodios depresivos y maníacos o hipomaníacos. La complejidad en el diagnóstico del trastorno afectivo bipolar debido a la superposición de sus síntomas con otros trastornos del estado de ánimo llevó a los investigadores y médicos a buscar nuevas y avanzadas técnicas para la detección precisa del trastorno afectivo bipolar. Uno de estos métodos es el uso de algoritmos avanzados de aprendizaje automático bajo una metodología estadística para la construcción de modelos de Regresión logística, Random Forest. Máquinas de vectores de soporte, Decision Tree, K-Vecinos más cercanos, Gradient Boosting, con datos 146 recopilados de los servicios de psiquiatría adscritos al sistema de salud mental del Ecuador. A nivel inferencial, los resultados sugieren que la implementación de algoritmo automáticos basado en las diferentes metodologías para la construcción de modelos permitan predecir o clasificar con éxito a individuos con trastornos afectivos bipolar en Ecuador en comparación con pacientes controles que no se perfilan bajo este cuadro patológico. Siendo el mejor modelo estadístico Random Forest (89,35 %) que dictamina las mejores métricas de rendimiento en comparación con el modelo Gradient Boosting. La evolución de la prevalencia general en trastornos afectivos bipolar en el Ecuador durante los últimos 22 años, se ha incrementado en un pequeño diferencial. Sin embargo, durante el año 2020 hasta 2022 se ha producido un incremento considerable en la prevalencia porcentual de casos en trastornos afectivos bipolar en el Ecuador.

Palabras clave: Algoritmos; Aprendizaje Automático; Decision Tree Gradient Boosting; Random Forest.

INTRODUCTION

Bipolar disorder (BD) is one of the most serious mental disorders, affecting men and women equally, with estimates of 2 to 3 % of the global population suffering from it, according to figures from the World Health Organization (WHO)[1], which translates into more than 47 million people worldwide who combine depressive episodes with a history of depressive symptoms overlapping with major depressive disorder (MDD) and periods of manic/hypomanic episodes.

This disorder includes euphoria or irritability, increased activity or energy, and other symptoms such as increased talkativeness, racing thoughts, increased self-esteem, decreased need for sleep, distraction, and impulsive and reckless behavior in the clinical prevalence of vulnerable patients with this disease. However, there are effective treatment options that include psychoeducation, stress reduction, and strengthening social functioning, as well as medication.[2]

People with bipolar disorder are likely to experience some form of disability and have a higher risk of developing suicidal tendencies. Suicide rates among patients with BD are 20 to 30 times higher than in the general population[3], and BD is associated with significant premature death for a variety of reasons.[3] The diagnosis of BD can be challenging and may require a long diagnostic process lasting 6 to 10 years.[4]

Affected individuals often experience a significant depressive episode and are misdiagnosed with unipolar depression. Misdiagnosis leads to inappropriate prescriptions of antidepressants that do not stabilize the mood and increase the likelihood of manic episodes.[5] Consequently, untreated bipolar disorders are associated with more prolonged and more severe mood episodes, as well as a higher frequency of suicide attempts.[2]

Identifying individuals at risk for bipolar disorder could allow for targeted assessment, early intervention, and more appropriate therapies. A recent comprehensive analysis of treatment trials to prevent bipolar disorder indicated a dependence on family history for risk identification.[3] However, given the multifactorial nature of bipolar disorder, most affected individuals would not have a favorable family history, furthermore, unlike schizophrenia, which is characterized by significant disturbances in reality perception and changes in behavior, including persistent delusions, hallucinations, experiences of influence, disorganized thinking, highly disorganized behavior, "negative symptoms," and extreme agitation or slowing of movements.[4] In the context of bipolar disorder, no early warning signs are recognized.[5] It would be helpful to develop new methods for identifying risk that do not rely on existing clinical signs or symptoms.

Reviews of previous research in the field indicate that there is a notable precedent for evaluating longitudinal electronic health records (EHRs). This, together with predictive analytics, may open new avenues for identifying risk factors for the presentation of bipolar disorder. As demonstrated above, these data can produce reliable diagnostic phenotyping of bipolar patients.[6,7]

Clinical practice can now incorporate scientific findings, given the widespread adoption of machine learning in the medical industry. One of the scientific findings involves the creation of a model that allows for the differentiation of various diseases at the individual level, and then applying the model to newly collected data to perform classification or prediction. A supervised machine learning technique called Support Vector Machine (SVM) aimed to detect patterns of spatially dispersed brain changes in several voxels.[8]

Another contribution was based on the early identification of bipolar disorder risk using predictive models trained on various cohorts in the United States, which could improve the specific assessment of high-risk individuals, reduce misdiagnosis, and improve the allocation of limited mental health resources. Valid predictive models were developed with multiple algorithms at each study site, including random forests, gradient boosting machines, penalized regression, and stacked ensemble learning algorithms. The predictors were based on widely available data, independent of a standard data model that included demographic data, diagnostic codes, and medications.

Consequently, the objective of the present study was to describe the phenomenon related to bipolar affective disorder in Ecuador by adopting classification and prediction models appropriate to the available data. Specifically, it focused on classifying patients with bipolar type I (BTI), bipolar type II (BTI), and unipolar depression. This will allow for the selection of a treatment appropriate to the diagnosis without errors and will also enable the identification of characteristic features that allow for an accurate differential diagnosis independent of the states of affective disorders.

THEORETICAL FRAMEWORK

Bipolar affective disorder

Cyclical and extreme mood swings, including episodes of mania or hypomania and sadness, interspersed with periods of stable mood, are the hallmark of bipolar disorder, a serious mental illness. The person experiences euphoria, increased energy, and activity during the manic period, while intense negativity and lack of enthusiasm are among the symptoms of the depressive phase. H The mood swings can occur with varying frequency and duration, depending on the type of bipolar disorder, age, and other external circumstances[9]

Types of bipolar affective disorder

Bipolar disorder (BD) is a severe mood disorder characterized by recurrent episodes of depression and (hypo)mania. Bipolar disorder is defined by two mood swings: emotional highs (mania or hypomania) and lows (depression). Although mania and hypomania are not the same, the Diagnostic and Statistical Manual of Mental Disorders DSM-5 divides bipolar disorder into three main subtypes: bipolar I, bipolar II, and cyclothymic disorder. The two main subtypes are BD-I (manic episodes, typically accompanied by depression) and BD-II (hypomanic episodes, combined with depression). At the same time, cyclothymic disorder is a cyclical condition that produces short periods of hypomania and depression.[10]

Unipolar disorder is a mental illness that significantly impairs daily functioning and is characterized by a persistently low mood or loss of interest in activities. However, when bipolar disorder is identified early, its onset can be prevented or delayed, and those who already have it can have better clinical outcomes.[11] In this regard, diagnosis is essential. Therefore, four main processes were included in the conventional diagnosis of bipolar disorder: suspicion based on depressive symptoms, follow-up of the clinical interview, case analysis using diagnosis-based search methods, and confirmation of the diagnosis.[12]

This is where the difficulty of diagnosing bipolar disorder lies, particularly in distinguishing it from unipolar depression. The IHME[13] proposes protocols to distinguish between major depression and bipolar disorder, emphasizing that bipolar disorder is associated with greater mood lability, motor slowness, and sleep duration, which facilitate its identification and, with it, treatment options.

METHOD

Data was collected from ourworldindata.org, which describes the estimated participation of men and women who presented bipolar disorder in Ecuador, whether diagnosed or not, based on representative surveys, medical data, and statistical modeling.[14] This response variable is represented by the estimated prevalence associated with the age of individuals with bipolar disorder, per 100 people. This metric was constructed following principles of comparability through standardization by age and prevalence ratio records using a uniform methodology with a comparative principle.

The Ecuadorian Social Security Institute (IESS) offers psychiatric services with beds allocated in four hospitals throughout the country, where outpatient care is also provided. In the cities of Quito and Guayaquil, psychiatric services are available in outpatient clinics. In Ecuador, there are five psychiatric hospitals in operation, three in Quito, one in Guayaquil, and one in Cuenca, with a total of 1 635 beds, equivalent to 12 beds per 100 000 inhabitants. Administrative records have been managed with global indicators to protect doctor-patient privacy.

The World Health Organization's Instrument for the Evaluation of Mental Health Systems (WHO-IESM) was used to collect information on Ecuador's mental health system. In this regard, a total of 146 patients under the monitoring of the psychiatry department attached to the IESS were included in this study, in which different metrics were evaluated, such as: sex (1: female, 2: male), age (age groups), type of condition (1: bipolar II, 2: unipolar depression, 3: bipolar I). Based on the related conditions, symptoms of melancholy (1: melancholy, 2: no melancholy), hospitalization (1: hospitalized, 2: outpatient), education (education grouped in years), marital status (1: married or cohabiting, 2: single), employment (1: patient in work or studying, 2: unemployed/retired due to illness). In this context, the detection of bipolar affective disorder and its subsequent classification into three types requires the management of patient information for adjustment to the model and subsequent prediction.

Given that the research was based exclusively on previously collected and anonymized secondary data, it was not necessary to obtain direct informed consent from the patients. However, confidentiality, the protection of the subjects' identity, and the exclusive use of the information for scientific purposes were guaranteed at all times. Ethical validation of the protocol ensures the integrity of the study, considering the sensitive nature of bipolar affective disorder as the subject of analysis.

Methods based on artificial intelligence algorithms

The choice of classification models used in this study on bipolar affective disorder in Ecuador responds to the need to address a complex clinical problem from a robust, flexible, and locally adapted computational approach. Widely validated algorithms such as logistic regression, Random Forest, SVM, decision trees, K-nearest neighbors, and Gradient Boosting were chosen due to their specific strengths for clinical data analysis, which include characteristics such as the handling of small samples, the presence of missing data, the noise inherent in medical records, and the need to interpret models for application in mental health settings.

Logistic regression offers a traditional statistical approach that is useful for its interpretability, while Random Forest and Gradient Boosting stand out for their accuracy and ability to avoid overfitting. SVMs were selected for their high performance with little data and their ability to identify nonlinear patterns, while decision trees and KNN provide transparency and simplicity in decision-making. These models allow for the identification of risk factors and the classification of patient clinical states, which is essential for improving early detection and treatment personalization in vulnerable populations. Likewise, rigorous cross-validation was applied to compare model performance and ensure their generalization ability, thus guaranteeing that the results of the analysis can be used with confidence in future clinical and research applications in the Ecuadorian context.

Some of the classification learning algorithms are listed below:

1) Logistic regression: a statistical technique for binary classification tasks that calculates the probability of an outcome based on input attributes. The use of the Scikit-learn, TensorFlow, and PyTorch libraries supported this logistic regression algorithm.

2) Random Forest: represents a joint learning technique that reduces overfitting and increases overall accuracy by constructing multiple decision trees and combining their predictions. The Random Forest method can handle missing data and is resistant to noise and outliers under specific TensorFlow and PyTorch libraries.

3) Support vector machines: This is a powerful technique that seeks to identify the ideal hyperplane that divides two classes for classification and regression applications. SVMs are recognized for their superior generalization abilities and can handle nonlinear decision boundaries using kernel functions. The SVM algorithm was applied as a classifier due to its excellent performance on a small sample size[15] by adopting the Python-based Scikit-learn package.[16]

4) Decision Tree: A decision tree is a type of non-parametric supervised learning method used for classification and regression tasks. It is characterized by a hierarchical tree structure consisting of a root node, branches, internal nodes, and leaf nodes. Python uses Scikit-learn, an open-source machine learning package that includes several machine learning methods, such as decision trees.

5) K-Nearest Neighbors: represents a non-parametric approach to classifying instances using the majority vote of their k nearest neighbors. KNN is easy to understand and can be successful with small data sets. However, it can be computationally expensive for large data sets. In Python, this algorithm is implemented using the scikit-learn library.

6) Gradient Boosting: In the context of minimizing a loss function, gradient boosting represents a functional gradient method that continuously chooses a function that points in the direction of a weak hypothesis or a negative gradient. A robust prediction model is created by combining many weak learning models with a gradient-boosting classifier. This ensemble learning technique builds decision trees sequentially and combines their predictions to improve overall performance under the Gradient Boosting algorithm through its GradientBoostingClassifier class. Its excellent predictive accuracy has made it a popular choice in machine learning implementations.

Finally, cross-validation was performed to produce a correct classification procedure, followed by statistical tests to evaluate which estimation technique works best. The cross-variation above compares the proportion of proper internal and external classification to assess the generalization of each approach used in model estimation.

Model evaluation

Relevant research in this field reveals a wide range of metrics that can reflect various synonyms in the domains of medicine and data science (information retrieval). We include here a summary of the most significant metrics based on the percentage of correct classification, accuracy, and precisión.[5] Metrics in specific risk areas included sensitivity, specificity, and positive predictive value.[5] To facilitate indirect comparisons with previous research, it is advisable to provide as many measurements as possible for future comparisons in the same field of study.

RESULTS

Exploratory analysis of preliminary data

The exploratory analysis of the annual prevalence of bipolar affective disorders in Ecuador reveals an approximate proportion of 0,94 % within the time coverage, with a variability in the yearly prevalence parameter of 0,036. This metric was obtained by manipulating information from Our World in Data, which allows working with age-standardized values to enable comparisons between countries and over time.[17,18]

Conversely, it is possible to work at the percentage level on the evolution of the overall prevalence of bipolar affective disorders in Ecuador over the last 22 years (figure 1). The trend has been to increase by a small margin. However, between 2020 and 2022, there was a considerable increase in the percentage prevalence of cases of bipolar affective disorders in Ecuador.

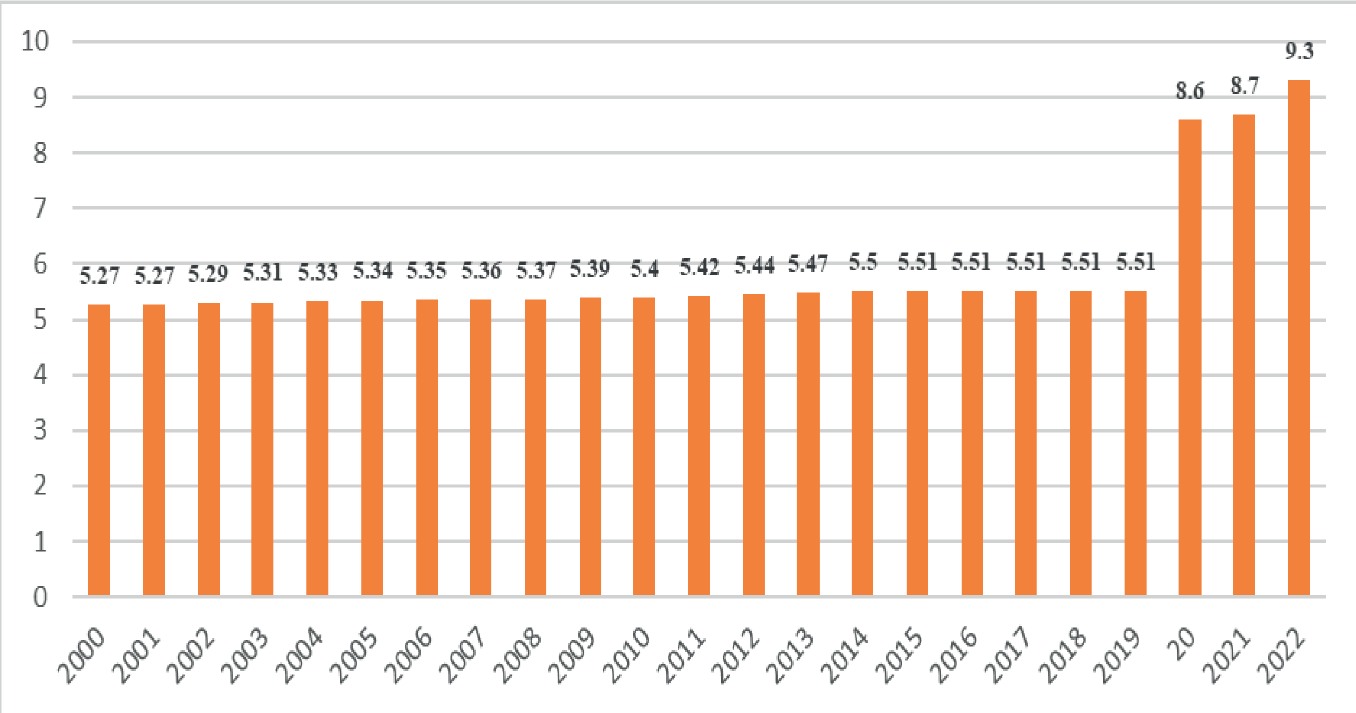

Figure 1. Percentage prevalence of bipolar affective disorders in Ecuador. Period 2000-2022.
Source: National Institute of Statistics and Census - Presidency of the Republic of Ecuador

Comparative Analysis

In this section, a comparative analysis was developed between the models specified using a machine learning algorithm to classify and predict bipolar affective disorder in Ecuador (table 1). These models were constructed with cross-sectional data corresponding to a total of 146 patients under monitoring by the psychiatry departments affiliated with the IESS.

Table 1. Representation of comparative accuracy in the identified models

Model Estimated with cross-validation	Classification accuracy
Logistic Regression (LR)	66,09 % (+/- 4,58)
Random Forest (RF)	89,35 % (+/- 3,85)
SVM	77,39 % (+/- 5,21)
Decision Tree (DT)	88,70 % (+/- 3,74)
K-Nearest Neighbors (KNN)	84,35 % (+/- 1,11)
Gradient Boosting (GB)	90,65 % (+/- 3,93)

Source: Python 3.11

Table 1 shows the performance of the models built using the precision metric as a measure calculated with the model's average score, which reflects the model's performance. At this point, two possible models that optimize performance stand out: Random Forest (89,35 %) and Gradient Boosting (90,65 %).

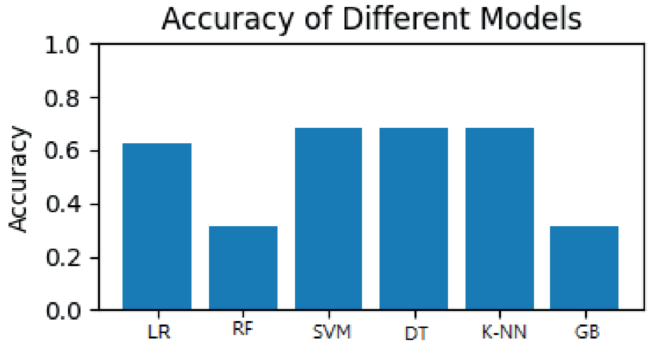

Figure 1. Representation of accuracy in different models

According to the results obtained, the SVM, DT, K-NN, and LR models perform equally well when discriminating patients with bipolar affective disorders. In addition, the GB and RF models represent the best option for modeling and explaining the presence of this phenomenon (Figure 1). These results are provided by implementing the algorithms in Python 3.11 under the PyCharm 2024.1 development environment.

In addition, the standard deviation of the scores multiplied by 100 and formatted with two decimal places can be used to compare the variability inherent in the dataset and to evaluate the accuracy of a given model. In this sense, among the two models identified in Table 1 that tend to optimize performance, we highlight the findings in the choice of Random Forest estimation, which present a lower standard deviation of the scores (+/- 3,85) compared to those obtained by the Gradient Boosting model (+/- 3,93).

Regarding the estimation of the best parameters for the Random Forest model, the best parameters are described by activating ('bootstrap': True), which represents sampling with replacement, helping to improve the accuracy of the model. Then, the maximum depth in the Random Forest model is limited to 10, avoiding overfitting ('max_depth': 10), defining the minimum number of samples that must be present in a leaf for the tree to split ('min_samples_leaf': 4); as well as stipulating the minimum number of four samples that must be present in a node for the tree to split ('min_samples_split': 10) to obtain 10 estimators as the best parameters of the Random Forest model (table 2).

Table 2. Future importance in classification with the Random Forest model

Variable	Contribution
Age	0,432777
Marriage	0,207413
Days	0,1308
Work	0,074981

Source: Python 3.11

The findings indicate that age and marital status in the cohort under study for classifying bipolar affective disorder in Ecuador have the highest incidence or weight in explaining the prevalence of this type of disease using the Random Forest model (table 2), which allows sensitivity, precision, and precision scores to be estimated from the data. Sensitivity, defined by the Recall metric, represents the probability frequencies around 0,89 for the detection of the disorder in conjunction with the actual positive rate. This is the proportion of samples that are correctly identified as positive among all existing positive samples. The precision metric, also known as positive predictive value, indicates the proportion of 0,81 in the returned samples that discriminate positive precision. The f1_score of 0,85 represents a test accuracy metric achieved by calculating the harmonic mean between accuracy and recall. Based on the above, the output criterion in terms of the construction of the confusion matrix (figure 2).

According to the results shown in the confusion matrix, the estimated value in cases that achieve a correct classification of bipolar affective disorders within the cohort under study is 17 % (actual positive cases 25). In contrast, the estimated value correctly presented for actual negative cases or cases without bipolar affective disorders is 61 % (actual negative cases 89). Essentially, there is a 78 % correct classification, which translates into the representation of the number of samples correctly identified within the study population. In this context, the above classifications shown in figure 3 are differentiated according to the defined bipolar affective types (unipolar depression, bipolar I, and bipolar II).

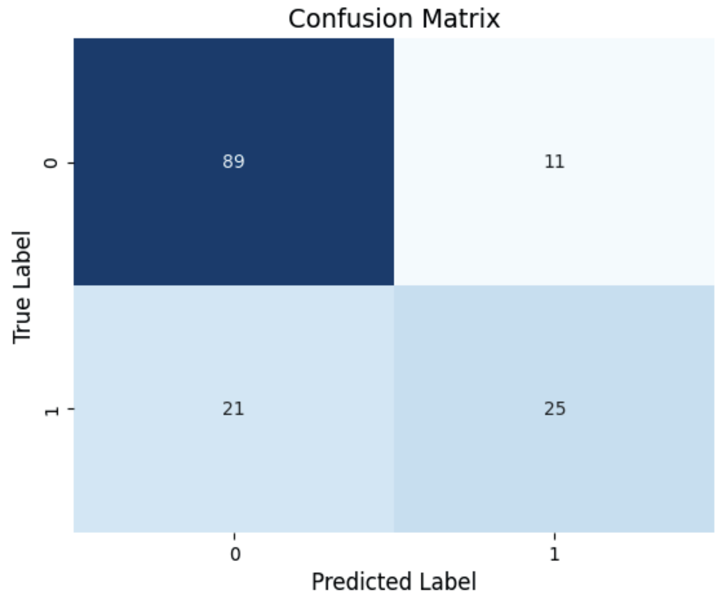

Figure 2. Random Forest confusion matrix

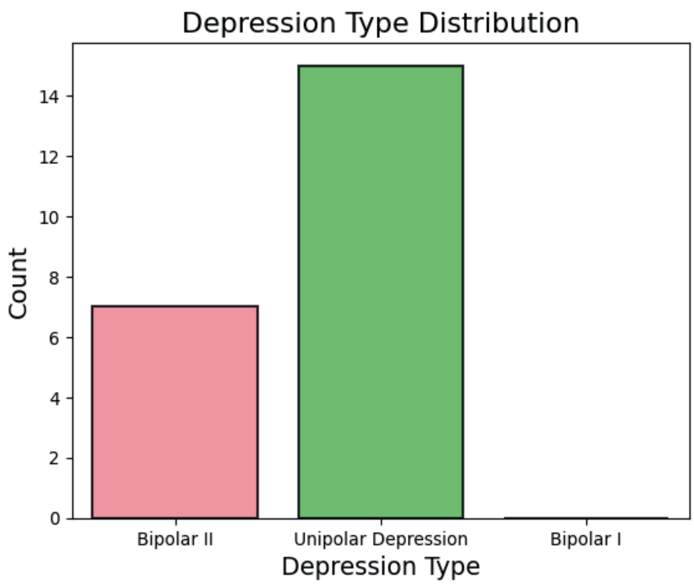

Figure 3. Classification of bipolar affective disorders

As shown in figure 3, in terms of the total number of patients studied (146), 15 fall into the classification of unipolar depressive disorders, reflecting a disease incidence of 10 % according to the sample population. In addition, seven (07) individuals fit the criteria for bipolar II disorder, equivalent to a prevalence of 5 %, and only three (03) individuals were defined as having bipolar type I disorder (prevalence of 2 %).

DISCUSSION

This study, which focuses on the mathematical modeling of bipolar affective disorder in Ecuador, has several limitations that should be considered when interpreting its results. First, the availability and quality of clinical data are a challenge, as Ecuador does not have a standardized national database that systematically records cases of affective disorders. This makes it necessary to use secondary sources or institutional data with possible underreporting or access biases, especially in rural or hard-to-reach areas. In addition, the mathematical model requires simplifying complex dynamics associated with bipolar disorder, such as psychiatric comorbidities, family environment impact, therapeutic adherence, or exposure to stressful events, which cannot be fully modeled due to their multifactorial and subjective nature. Furthermore, the epidemiological parameters used remain static over time, without considering social, economic, or health fluctuations that could significantly alter the incidence, clinical course, or therapeutic response of the population. Another significant limitation is the partial geographical representation of the model, given that most of the data comes from urban areas with greater access to mental health services, which excludes the realities of indigenous populations or marginalized communities.

Despite its focus on the Ecuadorian population, the mathematical model developed has strong potential for replication in other Latin American contexts, especially in countries with similar structural conditions in mental health. The modeling framework used, based on dynamic systems and transition rates between clinical states, can be easily adapted to different populations by updating specific epidemiological parameters, such as prevalence, episode duration, relapse rates, or therapeutic adherence. This feature makes it a helpful tool not only for local analysis but also for regional comparative studies.

The Random Forest model exhibits high sensitivity, indicating a high degree of ability to detect significant cases. This shows that it has excellent skills for discriminating negative cases. In other words, the generation of false positives is unlikely. As a result, this model is much more sensitive than specific. When our goal is to avoid false positives at all costs, this is the scenario that catches our attention.

The Random Forest (RF) model represents an integrated feature selection technique that is often used in data mining[19], has been shown to increase model performance[20,21] significantly, and identifies key genes.[22] One way to measure the importance of a feature in the RF model is to use the mean impurity decay (MDI). This is the average reduction in the impurity of the nodes in all trees, weighted by the probability of reaching that node. In this work, we use a supervised learning methodology to build a random forest model on the training set using the default configuration.

In agreement, Llamocca et al.[23] demonstrated that enhanced negative affect amplifies positive emotional contrasts when faced with unexpected positive events. At the same time, people sensitive to negative emotional changes use previous increases in negative affect to avoid further escalation in response to adverse changes, according to a contrast avoidance model that was developed. People with bipolar spectrum disorders, which are characterized by oscillations between episodes of hypomania and depression, may experience more subtle differences in their feelings.

A marginal impact was identified in bipolar subtype II and showed a greater propensity to avoid unpleasant emotional contrasts compared to bipolar I disorder according to the contrast avoidance model. These results imply that a psychological process related to bipolar spectrum disorders may be contrast avoidance, whose early detection in people at risk of bipolar affective disorder allows for examination and targeted prevention. Although a variety of risk factors for bipolar disorder have been identified, such as family history[24] and stressful life events[25], quantitative and scalable risk prediction remains a challenge. Therefore, previous research has mainly focused on those who have had depression in the past and/or used small samples.

Previous studies have primarily focused on individuals with a history of depression and/or have included relatively small samples.[26] This study began by evaluating the general prevalence of bipolar affective disorder in the Republic of Ecuador during the period 2000-2022. In a specific context, a cohort of people with the characteristic of interest in the present study was managed to validate multiple approaches and describe the phenomenon under a diagnosis of bipolar affective disorder.

In practice, advances in clinical and precision medicine could lead to the development of generalizable bipolar disease prediction models that have been trained and validated in various health systems. While the accuracy of the models is strong, it does not take into account the innate social and environmental factors that influence bipolar affective behavior. Instead, it describes and predicts the physical and psychological components of quality of life.

Orsolini used digital phenotyping to classify bipolar disorder and unipolar disorder: exploratory findings using machine learning models under the following objectives were to investigate 1) disparities between bipolar disorder and unipolar disorder in smartphone-based data on phone use; and 2) intelligent categorization of bipolar disorder and unipolar disorder using machine learning models, sensitivity, specificity, and AUC of phone data. For six months, daily smartphone-based mood self-reports were accessed, and at the same time, smartphone usage data was passively collected. In summary, digital phenotyping highlights the difficulty of generalizing to people with few tangible elements, even if it shows promise in distinguishing between patients with bipolar disorder and unipolar disorder. It should be used in conjunction with comprehensive clinical assessments by professionals.[27]

The development of generalizable prediction models for bipolar disease risk in a variety of empirical settings is ultimately the path toward precision medicine. A combination strategy proved to be the most effective overall when comparing a variety of machine learning algorithms; however, it required local retraining, in that different algorithms (random forests, gradient boosting machines, penalized regression, and stacked ensemble learning algorithms) were used to create valid prediction models.

In this context, a recent study addressed predictors that were limited to attributes derived from publicly available electronic medical records, without reference to a shared data model encompassing demographic data, disease codes, and prescribed medications, demonstrating that early detection of bipolar disorder risk using generalizable prediction models trained on diverse cohorts in Ecuadorian populations should improve targeted assessment of high-risk patients, prevent misdiagnosis, and optimize the allocation of scarce mental health resources.[28]

CONCLUSIONS

The exploratory analysis of the annual prevalence of bipolar affective disorders in Ecuador is described with an approximate proportion of 0,94 % within the time frame covered. This metric was obtained by manipulating information from Our World in Data, which allows working with age-standardized values to enable comparisons between countries and over time. Regarding the comparative analysis between the models specified by the machine learning algorithm to classify and predict bipolar affective disorder in Ecuador, the SVM, DT, K-NN, and LR models perform equally well when discriminating patients with bipolar affective disorders. In addition, the GB and RF models represent the best option for modeling and explaining the presence of this phenomenon.

The future importance of classification with the Random Forest model lies in its ability to explain the prevalence of this type of disease more effectively. This Random Forest model allows the estimation of sensitivity, precision, and accuracy scores. Sensitivity, defined by the Recall metric, represents the probabilistic frequencies around 0,89 for the detection of the disorder in conjunction with the actual positive rate. Accuracy, also known as positive predictive value, indicates the proportion of 0,81 in the returned samples that are correctly identified as positive. The f1_score of 0,85 represents a test accuracy metric achieved by calculating the harmonic mean between accuracy and recall. Or, indeed, consider expanding the study by using artificial intelligence to evaluate actigraphic records of motor activity, which represent an objective way of observing depression, even if psychiatric research on this topic has not analyzed it in great detail.

REFERENCES

1. OMS. Psychosis and bipolar disorders. Who.int. Disponible en: https://www.who.int/teams/mental-health-and-substance-use/treatment-care/mental-health-gap-action-programme/evidence-centre/psychosis-

2. Weye N, Momen NC, Christensen MK, Iburg KM, Dalsgaard S, Laursen TM, et al. Association of specific mental disorders with premature mortality in the danish population using alternative measurement methods. JAMA Netw Open. 2020;3(6):e206646. http://dx.doi.org/10.1001/jamanetworkopen.2020.6646

3. Biazus TB, Beraldi GH, Tokeshi L, Rotenberg L de S, Dragioti E, Carvalho AF, et al. All-cause and cause-specific mortality among people with bipolar disorder: a large-scale systematic review and meta-analysis. Mol Psychiatry. 2023;28(6):2508-24. http://dx.doi.org/10.1038/s41380-023-02109-9

4. McIntyre RS, Alda M, Baldessarini RJ, Bauer M, Berk M, Correll CU, et al. The clinical characterization of the adult patient with bipolar disorder aimed at personalization of management. World Psychiatry. 2022;21(3):364-87. http://dx.doi.org/10.1002/wps.20997

5. Chakrabarti S, Singh N. Psychotic symptoms in bipolar disorder and their impact on the illness: A systematic review. World J Psychiatry. 2022;12(9):1204-32. http://dx.doi.org/10.5498/wjp.v12.i9.1204

6. Holm M, Taipale H, Tanskanen A, Tiihonen J, Mitterdorfer-Rutz E. Employment among people with schizophrenia or bipolar disorder: A population-based study using nationwide registers. Acta Psychiatr Scand. 2021;143(1):61-71. http://dx.doi.org/10.1111/acps.13254

7. Joas E, Bäckman K, Karanti A, Sparding T, Colom F, Pålsson E, et al. Psychoeducation for bipolar disorder and risk of recurrence and hospitalization - a within-individual analysis using registry data. Psychol Med. 2020;50(6):1043-9. http://dx.doi.org/10.1017/s0033291719001053

8. Walsh CG, Ripperger MA, Hu Y, Sheu Y-H, Lee H, Wilimitis D, et al. Development and multi-site external validation of a generalizable risk prediction model for bipolar disorder. Transl Psychiatry. 2024;14(1). http://dx.doi.org/10.1038/s41398-023-02720-y

9. Rantala MJ, Luoto S, Borráz-León JI, Krams I. Bipolar disorder: An evolutionary psychoneuroimmunological approach. Neurosci Biobehav Rev. 2021;122:28-37. http://dx.doi.org/10.1016/j.neubiorev.2020.12.031

10. Cheon E-J, Bearden CE, Sun D, Ching CRK, Andreassen OA, Schmaal L, et al. Cross disorder comparisons of brain structure in schizophrenia, bipolar disorder, major depressive disorder, and 22q11.2 deletion syndrome: A review of ENIGMA findings. Psychiatry Clin Neurosci. 2022;76(5):140-61. http://dx.doi.org/10.1111/pcn.13337

11. Malhi GS, Bell E, Bassett D, Boyce P, Bryant R, Hazell P, et al. The 2020 Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. Aust N Z J Psychiatry. 2021;55(1):7-117. http://dx.doi.org/10.1177/0004867420979353

12. Di Salvo G, Porceddu G, Albert U, Maina G, Rosso G. Correlates of long duration of untreated illness (DUI) in patients with bipolar disorder: results of an observational study. Ann Gen Psychiatry. 2023;22(1). http://dx.doi.org/10.1186/s12991-023-00442-5

13. IHME Bipolar disorder prevalence. Our World in Data. Disponible en: https://ourworldindata.org/grapher/bipolar-disorder-prevalence?tab=chart&time=earliest..latest&country=~ECU

14. INEC Ecuador - Estadísticas Hospitalarias Camas y Egresos 2022. Gob.ec. Disponible en: https://anda.inec.gob.ec/anda/index.php/catalog/976/datafile/F47/V1715

15. Carruthers SP, Rossell SL, Murray G, Karantonis J, Furlong LS, Van Rheenen TE. Mindfulness, mood symptom tendencies and quality of life in bipolar disorder: An examination of the mediating influence of emotion regulation difficulties. J Affect Disord. 2022;298:166-72. http://dx.doi.org/10.1016/j.jad.2021.10.107

16. Lalli M, Brouillette K, Kapczinski F, de Azevedo Cardoso T. Substance use as a risk factor for bipolar disorder: A systematic review. J Psychiatr Res. 2021;144:285-95. http://dx.doi.org/10.1016/j.jpsychires.2021.10.012

17. Rodriguez V, Alameda L, Trotta G, Spinazzola E, Marino P, Matheson SL, et al. Environmental risk factors in Bipolar disorder and Psychotic depression: A systematic review and meta-analysis of prospective studies. Schizophr Bull. 2021;47(4):959-74. http://dx.doi.org/10.1093/schbul/sbaa197

18. He H, Hu C, Ren Z, Bai L, Gao F, Lyu J. Trends in the incidence and DALYs of bipolar disorder at global, regional, and national levels: Results from the global burden of Disease Study 2017. J Psychiatr Res. 2020;125:96-105. http://dx.doi.org/10.1016/j.jpsychires.2020.03.015

19. Palacios-Ariza MA, Morales-Mendoza E, Murcia J, Arias-Duarte R, Lara-Castellanos G, Cely-Jiménez A, et al. Prediction of patient admission and readmission in adults from a Colombian cohort with bipolar disorder using artificial intelligence. Front Psychiatry. 2023;14. http://dx.doi.org/10.3389/fpsyt.2023.1266548

20. Montazeri M, Montazeri M, Bahaadinbeigy K, Montazeri M, Afraz A. Application of machine learning methods in predicting schizophrenia and bipolar disorders: A systematic review. Health Sci Rep. 2023;6(1). http://dx.doi.org/10.1002/hsr2.962

21. Wanderley Espinola C, Gomes JC, Mônica Silva Pereira J, dos Santos WP. Detection of major depressive disorder, bipolar disorder, schizophrenia and generalized anxiety disorder using vocal acoustic analysis and machine learning: an exploratory study. Res Biomed Eng. 2022;38(3):813-29. http://dx.doi.org/10.1007/s42600-022-00222-2

22. Cearns M, Amare AT, Schubert KO, Thalamuthu A, Frank J, Streit F, et al. Using polygenic scores and clinical data for bipolar disorder patient stratification and lithium response prediction: machine learning approach. Br J Psychiatry. 2022;220(4):219-28. http://dx.doi.org/10.1192/bjp.2022.28

23. Llamocca P, López V, Santos M, Čukić M. Personalized characterization of emotional states in patients with bipolar disorder. Mathematics. 2021;9(11):1174. http://dx.doi.org/10.3390/math9111174

24. Gordovez FJA, McMahon FJ. The genetics of bipolar disorder. Mol Psychiatry. 2020;25(3):544-59. http://dx.doi.org/10.1038/s41380-019-0634-7

25. Mikolas P, Bröckel K, Vogelbacher C, Müller DK, Marxen M, Berndt C, et al. Individuals at increased risk for development of bipolar disorder display structural alterations similar to people with manifest disease. Transl Psychiatry. 2021;11(1). http://dx.doi.org/10.1038/s41398-021-01598-y

26. Comsa M, Anderson KN, Sharma A, Yadav VC, Watson S. The relationship between sleep and depression and bipolar disorder in children and young people. BJPsych Open. 2022;8(1). http://dx.doi.org/10.1192/bjo.2021.1076

27. Orsolini L, Fiorani M, Volpe U. Digital phenotyping in bipolar disorder: Which integration with clinical endophenotypes and biomarkers? Int J Mol Sci. 2020;21(20):7684. http://dx.doi.org/10.3390/ijms21207684

28. Sonkurt HO, Altınöz AE, Çimen E, Köşger F, Öztürk G. The role of cognitive functions in the diagnosis of bipolar disorder: A machine learning model. Int J Med Inform. 2021;145(104311):104311. http://dx.doi.org/10.1016/j.ijmedinf.2020.104311

FUNDING
The authors did not receive funding for the development of this research.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTION
Conceptualization: Cristhian Ismael Gómez Gaona.
Data curation: Andrea del Rocío Mejía Rubio.
Formal analysis: José Rubén León Pérez.
Research: Jesús Rodríguez.
Methodology: Zilma Diago Alfes.
Project management: Laura Esther Muñoz Escobar.
Resources: Cristian Inca.
Software: Jimmy Yaguana Torres.
Supervision: Marco Velasco Arellano.
Validation: Jesús Rodríguez.
Visualization: Cristian Gómez.
Writing - original draft: Cristian Inca.
Writing - revision and editing: Marco Velasco Arellano.