Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics e-ISSN: 2656-8624
Homepage: https://ijeeemi.org/; Vol. 8, No. 1, pp. 115-129, February 2026

RESEARCH PAPER

OPEN ACCES

Multi-Modal Ensemble Framework for Mental Health Disorder Prediction: A Novel Machine Learning Approach

M. Fadli Ridhani[1], **Tesdiq Prigel Kaloka**[2], **Yazid Aufar**[2], **and Annisa Rizqiana**[3]

[1] Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia
[2] Department of Informatics Engineering, Politeknik Hasnur, Barito Kuala, Indonesia
[3] Department of Computer Science, Faculty of Mathematics and Natural Science, Universitas Lambung Mangkurat, Banjarbaru, Indonesia

Abstract

Mental health disorders constitute a major global public health concern, affecting millions of individuals across diverse socioeconomic and cultural contexts. Accurate prediction of mental health outcomes at the population level remains challenging due to the complex and non-linear relationships among co-occurring disorders. Previous studies relying on traditional statistical approaches, particularly linear regression, have reported limited predictive performance, with an R^2 of approximately 0.7175. This limitation highlights the need for more advanced analytical frameworks capable of capturing comorbidity patterns and non-linear interactions among mental health conditions. This study proposes and evaluates a novel multi-modal ensemble machine learning framework to improve the prediction accuracy of eating disorder prevalence using global mental health data. The analysis utilizes country-level prevalence data for schizophrenia, depression, anxiety, bipolar disorder, and eating disorders across multiple countries and years. Eating disorder prevalence is modeled as the primary target variable, while other mental health disorders are incorporated as predictive features to represent clinically established comorbidity relationships. To enhance the representational capacity of the data, an extensive feature engineering strategy was applied, generating 19 additional features through polynomial transformations, interaction terms, ratio-based indicators, and aggregate burden measures. Unsupervised clustering techniques, including K-Means, DBSCAN, and hierarchical clustering, were employed to identify natural groupings of countries based on their mental health profiles. Furthermore, ten machine learning algorithms were systematically evaluated, including linear models, tree-based methods, neural networks, and support vector regression. The best-performing models were subsequently integrated into a stacking ensemble architecture. Experimental results demonstrate that the proposed stacking ensemble achieved a test R^2 score of 0.9955, corresponding to a 42.2% improvement over the baseline linear regression model. These results indicate that multi-modal ensemble approaches substantially enhance predictive accuracy and provide valuable insights to support evidence-based global mental health policy and targeted intervention planning.

Paper History
Received Dec 02, 2025
Revised Feb. 10, 2025
Accepted Feb. 25, 2026
Published Feb. 27, 2026

Keywords
Mental Health Disorders;
Machine Learning;
Ensemble Methods;
Clustering Analysis;
Feature Engineering

Author Email
mfadliridhani@gmail.com
tesdiq@gmail.com
yazid.aufar.ya@gmail.com
annisarizqiana@ulm.ac.id

I. Introduction

Mental health disorders constitute one of the leading causes of disability worldwide, affecting approximately 1 in 4 people during their lifetime (WHO, 2022). The global burden of mental illness includes major depressive disorder, anxiety disorders, bipolar disorder, schizophrenia, and eating disorders, each presenting unique challenges for diagnosis, treatment, and prevention [1]. Understanding the patterns and prevalence of these disorders across different populations is crucial for effective healthcare planning and resource allocation [2]. Traditional epidemiological studies have employed simple statistical methods, primarily linear regression, to model the prevalence of mental health disorders [3]. While these approaches provide baseline insights, they often fail to capture the complex, non-linear relationships between different mental health conditions and their comorbidities [4]. Recent work has achieved an explained variance of $R^2 = 0.7175$ (71.75%) using linear regression models, indicating substantial room for improvement through more sophisticated machine learning techniques capable of capturing non-linear relationships and disorder comorbidities [4]. Machine Learning in Mental Health: Recent studies have begun exploring machine learning applications in mental health research. Random forests and gradient boosting have shown promise in predicting depression outcomes [5], while neural networks have been applied to suicide risk assessment [6]. However, comprehensive comparative studies evaluating multiple

Corresponding author: M. Fadli Ridhani, mfadliridhani@gmail.com, Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia.

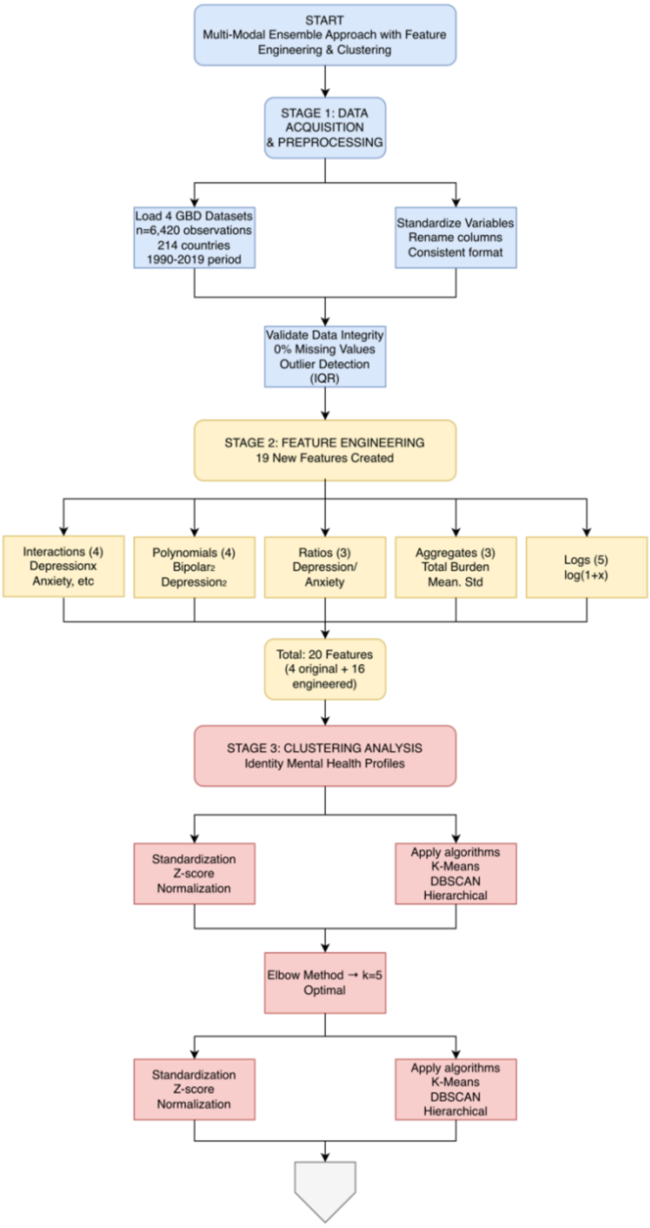
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.300

algorithms on global mental health data remain limited [7]. Although multiple mental health disorders are examined in this study, the predictive task specifically focuses on eating disorder prevalence. Schizophrenia, depression, anxiety, and bipolar disorder are incorporated as predictor variables to capture clinically established comorbidity patterns at the population level. This framing allows the proposed model to target a well-defined epidemiological outcome while leveraging the broader mental health landscape as explanatory context.

Ensemble Methods: Ensemble learning, which combines multiple models to achieve superior performance, has demonstrated success in various healthcare applications [8], [9]. Stacking ensembles, which combine base model predictions with a meta-learner, have achieved state-of-the-art results in disease prediction tasks [10]. Clustering in Mental Health: Unsupervised learning techniques have identified mental health subtypes and patient phenotypes [11]. However, their application to country-level global mental health patterns remains underexplored [12]. Despite growing interest in applying machine learning to mental health research, several key gaps remain. Most existing studies employ a limited range of algorithms, often relying on a single model rather than conducting comprehensive comparative analyses [13]. Feature engineering is also underdeveloped, with complex interactions between mental health disorders rarely modeled explicitly. Additionally, ensemble methods, which have the potential to substantially enhance predictive performance, are seldom utilized [14]. Clustering analysis is another overlooked area, leaving natural groupings within populations based on mental health patterns largely unexplored [15]. As a result of these limitations, current predictive models generally achieve low accuracy, explaining less than 75% of the variance in mental health outcomes.

Previous studies on global mental health prediction have predominantly relied on single machine learning models with limited feature engineering, resulting in suboptimal predictive performance and insufficient modeling of disorder comorbidities. Moreover, country-level mental health patterns have rarely been explored using unsupervised learning approaches. Therefore, to address these limitations, this study proposes a comprehensive multimodal machine learning framework for predicting global mental health disorders. The contributions of this study are: (1) the development of advanced feature engineering techniques to explicitly capture disorder interactions and non-linear relationships in global mental health data; (2) the identification of natural country-level clusters based on mental health disorder profiles using unsupervised clustering methods; (3) a systematic comparative evaluation of more than ten machine learning algorithms for mental health disorder prediction; and (4) the design of a stacking ensemble model that integrates the best-performing algorithms to improve predictive accuracy and robustness. The overall research workflow and methodology employed in this study are illustrated in Fig. 1.

This paper is structured as follows: Section II describes the datasets used, the data preprocessing steps, the feature engineering procedures, the clustering analysis, and the machine learning models employed in this study. Section III presents the experimental results, including clustering outcomes, model performance comparisons, and feature importance analysis. Section IV discusses the interpretation of results, comparisons with related studies, practical implications, and study limitations. Finally, Section V concludes the paper by summarizing the main findings and outlining directions for future research.



Corresponding author: M. Fadli Ridhani, mfadliridhani@gmail.com, Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia.

Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.300

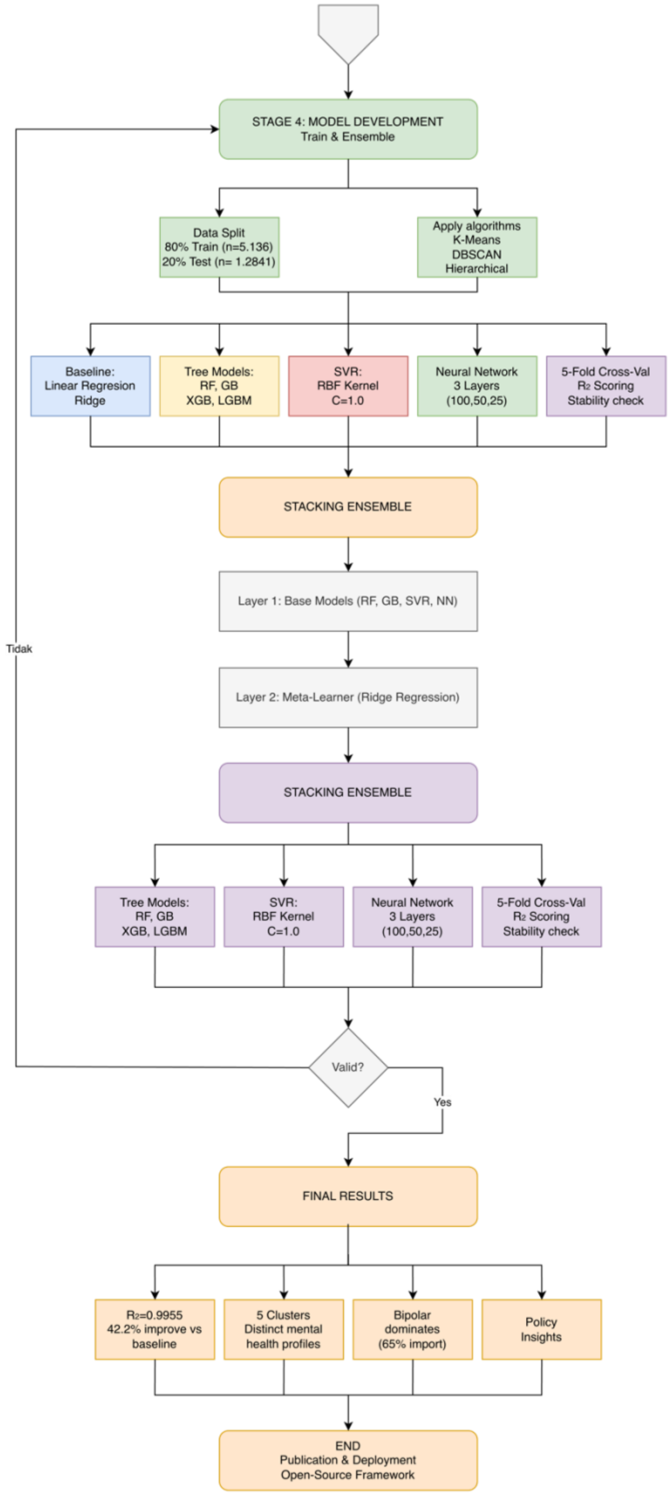


Fig. 1. Overall workflow of the proposed multi-modal ensemble framework, illustrating data preprocessing, feature engineering, clustering analysis, and predictive modeling stages.

II. Materials and Method

A. Data Sources

Datasets: we utilized four complementary datasets from the Global Burden of Disease Study to ensure comprehensive coverage and robust analysis [16]. The Mental Illnesses Prevalence Dataset $(n = 6{,}420)$ includes variables such as schizophrenia, depression, anxiety, bipolar disorder, and eating disorders, covering more than 195 countries and territories from 1990 to 2019. All values are standardized by age and sex, resulting in a dataset of 6,420 rows and 8 columns. It includes Entity (country/territory name), Code (ISO 3-letter code), and Year. The remaining columns represent age-standardized prevalence (%) of major mental health disorders: Schizophrenia (distortions in thinking and perception), Depression (persistent sadness and loss of interest), Anxiety (excessive fear or worry), Bipolar (manic and depressive episodes), and Eating disorders (abnormal eating behaviors; used as the target variable). All datasets were sourced from publicly available repositories and were pre-processed to ensure consistency and compatibility across analyses. As shown in Table 1, the dataset summary statistics for each variable are provided, including the mean, standard deviation, minimum, maximum, skewness, and kurtosis. Ethical Considerations: This study utilized only aggregated, de-identified, publicly available data. No individual-level data or personally identifiable information was used. Ethical approval was not required under the institution's guidelines for secondary data analysis.

Table 1. Dataset summary statistics describing the prevalence of mental health disorders across countries and observation years.

Variable	Mean	Std Dev	Min	…	Kurtosis
Schizophrenia	0.2666	0.0394	0.1884	…	2.1526
Depression	3.7670	0.9253	1.5223	…	0.5884
Anxiety	4.1018	1.0505	1.8800	…	1.5357
Bipolar	0.6370	0.2334	0.1817	…	0.3153
Eating	0.1957	0.1384	0.0448	…	4.0831

B. Data Preprocessing

Variable Renaming: original GBD variable names were standardized for readability (e.g., “Schizophrenia disorders (share of population)” → “Schizophrenia”). The same transformation was applied to all disorder variables.

Missing Data: missing values were imputed using mean imputation for continuous variables. Sensitivity checks showed negligible impact on model performance.

Outlier Handling: Outliers were detected using the IQR method. Extreme values (>3 IQR), mostly from small island nations, were retained as valid observations to preserve population heterogeneity [17], [18].

C. Feature Engineering

To capture non-linear relationships and comorbidity effects among mental health disorders, four categories of engineered features were constructed. Polynomial features were generated by squaring each disorder variable (e.g., Bipolar², Depression²) to model potential non-linear risk escalation. Interaction features (e.g., Bipolar × Anxiety, Schizophrenia × Depression) were

Corresponding author: M. Fadli Ridhani, mfadliridhani@gmail.com, Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia.
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.300

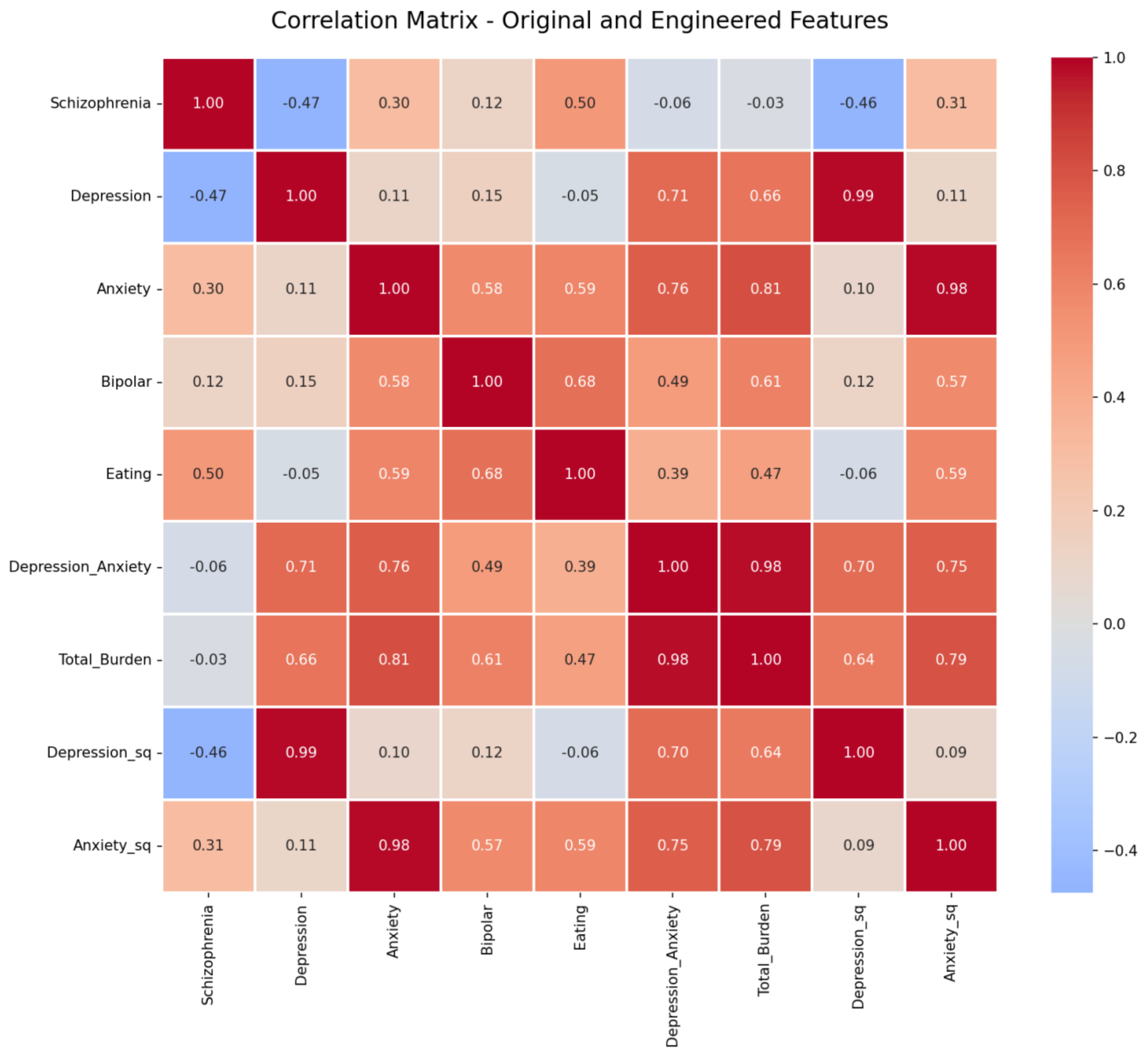


Fig. 2. Correlation heatmap showing relationships among mental health disorder prevalence variables in the dataset.

created to explicitly represent disorder comorbidity patterns documented in clinical literature. Ratio features (e.g., Bipolar/Schizophrenia, Anxiety/Depression) were included to reflect relative disorder dominance within countries, while aggregate indicators such as Total_Burden, Mean_Disorder, and Std_Disorder summarized systemic mental health load. Logarithmic transformations were applied to skewed variables to stabilize variance. In total, 19 engineered features were generated, substantially enriching the original feature space and enabling the model to capture complex epidemiological interactions. [19].

Formally, let $X = \{x_1, x_2, x_3, x_4\}$ denote the original feature vector representing the age-standardized prevalence of schizophrenia, depression, anxiety, and bipolar disorder, respectively. Polynomial features are defined as $x_i^{(2)} = x_i^2, \quad i = 1, \dots, 4$ to model non-linear relationships. Interaction features capturing disorder comorbidity are expressed as $x_{ij} = x_i \times x_j$ where $i \neq j$. Ratio-based features are computed as $r_{ij} = \frac{x_i}{x_j + \varepsilon}$ with ϵ representing a small constant to prevent division by zero. The overall burden of mental health disorders is quantified using the Total_Burden, defined in Eq. (1) [20] as the sum of all disorder prevalence values:

$$\text{Total\_Burden} = \sum_{i=1}^{4} x_i \tag{1}$$

The average prevalence across disorders is calculated using Mean_Disorder, as shown in Eq. (2) [20]:

$$\text{Mean\_Disorder} = \frac{1}{4}\sum_{i=1}^{4} x_i \tag{2}$$

Finally, variability among disorders is captured using Std_Disorder, defined in Eq. (3) [20]:

$$\text{Std\_Disorder} = \sqrt{\frac{1}{4}\sum_{i=1}^{4}(x_i - \bar{x})^2} \tag{3}$$

Corresponding author: M. Fadli Ridhani, mfadliridhani@gmail.com, Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia.

Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.300

Table 2. Cluster profiles summarizing average mental health disorder prevalence for each identified country cluster.

	Size	Schizophrenia	Depression	Anxiety	Bipolar	Eating	Description
0	1582	0.274	3.582	4.373	0.855	0.243	Moderate-anxiety profile
1	1635	0.215	4.744	3.595	0.574	0.103	High-depression profile
2	2205	0.285	3.095	3.490	0.405	0.123	Low-burden profile
3	180	0.391	4.429	6.115	0.891	0.510	High-burden profile
4	818	0.278	3.840	5.798	0.911	0.415	High-anxiety profile

All features were standardized using z-score normalization [21]. Clustering was conducted using K-Means [22], DBSCAN ($\varepsilon = 0.5$, $min\_samples = 5$), and agglomerative hierarchical clustering (Ward linkage, 5 clusters) [23]. Dimensionality reduction using PCA (2 components) and t-SNE enabled visualization and noise reduction [24]. The optimal number of clusters for K-Means was determined using a combination of the silhouette score, Davies–Bouldin index, and Calinski–Harabasz index, with $k = 5$ providing the best overall balance across all metrics. For DBSCAN, parameters were set to $\varepsilon = 0.5$ and $min\_samples = 5$ based on sensitivity testing and common practice, while agglomerative hierarchical clustering employed Ward linkage to minimize within-cluster variance.

D. Machine Learning Models

The dataset was split into 80% training and 20% testing [25] (random seed = 42). StandardScaler was fit on the training set and applied to the test set to avoid data leakage [26], [27]. Ten models were evaluated, including linear models (Linear, Ridge, Lasso, ElasticNet), tree-based models (Random Forest, Gradient Boosting, XGBoost, LightGBM), SVR, and a neural network (MLP with ReLU and Adam optimizer). A two-layer stacking ensemble architecture was implemented. Random Forest, Gradient Boosting, Support Vector Regression, and a Multilayer Perceptron were selected as base learners based on superior individual performance and complementary error patterns. Ridge Regression was employed as the meta-learner due to its regularization capability and stability. Five-fold cross-validation generated out-of-fold predictions for meta-model training, ensuring robust ensemble learning and minimizing overfitting. More details are shown in Fig. 1.

1. Linear Regression

Linear regression is a fundamental statistical learning method that models the relationship between input features and a continuous target variable through a linear combination of predictors. It assumes that the expected value of the dependent variable is a linear function of the independent variables. The model estimates coefficients that minimize the residual sum of squares between observed and predicted values. This approach is widely used for its interpretability and computational efficiency, making it a suitable baseline model for regression problems. However, its performance may degrade when multicollinearity or nonlinearity exists among predictors. The linear regression model is expressed in Eq. (4) [28] as:

$$\hat{y} = X\beta + b \tag{4}$$

$\hat{y}_i$ is the predicted output for the $i-th$ sample, $x_i = [x_{i1}, x_{i2}, \dots, x_{ip}]^T$ is the input feature vector of the i-th sample, $\beta = [\beta_1, \beta_2, \dots, \beta_p]^T$ is the vector of regression coefficients, b is the bias (intercept) term, p is the number of input features. To estimate the optimal parameters, the model minimizes the least-squares objective function shown in Eq. (5) [28]:

$$\min_{\beta,b} \frac{1}{2n} |y - (X\beta + b\mathbf{1})|_2^2 \tag{5}$$

where y_i is the true target value of the i-th sample, n is the total number of training samples. When the design matrix X has full column rank, the regression coefficients can be obtained analytically using the closed-form solution given in Eq. (6) [28]:

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{6}$$

2. Ridge (L2 regularization)

Ridge extends linear regression by incorporating L2 regularization to address multicollinearity and overfitting. The regularization term penalizes large coefficients, encouraging smoother, more stable models. Ridge Regression is particularly effective when predictors are highly correlated, as it distributes coefficient weights more evenly. Although it does not perform feature selection, the method improves generalization performance and robustness, making it suitable for high-dimensional regression problems where all predictors contribute to the response variable. The Ridge objective function is defined in Eq. (7) [28] as:

$$\min_{\beta,b} \frac{1}{2n} |y - (X\beta + b\mathbf{1})|_2^2 + \lambda |\beta|_2^2 \tag{7}$$

When expressed in matrix form, the closed-form solution of Ridge regression is given in Eq. (8) [28]:

$$\hat{\beta} = (X^T X + 2n\lambda I)^{-1} X^T y \tag{8}$$

$\lambda \geq 0$ is the regularization parameter controlling the strength of the penalty, $|\beta|_2^2 = \sum_{j=1}^{p} \beta_j^2$ is the L2 norm of the coefficient vector. Effect: shrinks coefficients smoothly; good for multicollinearity [29].

3. Lasso (L1 regularization)

Corresponding author: M. Fadli Ridhani, mfadliridhani@gmail.com, Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia.
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.300

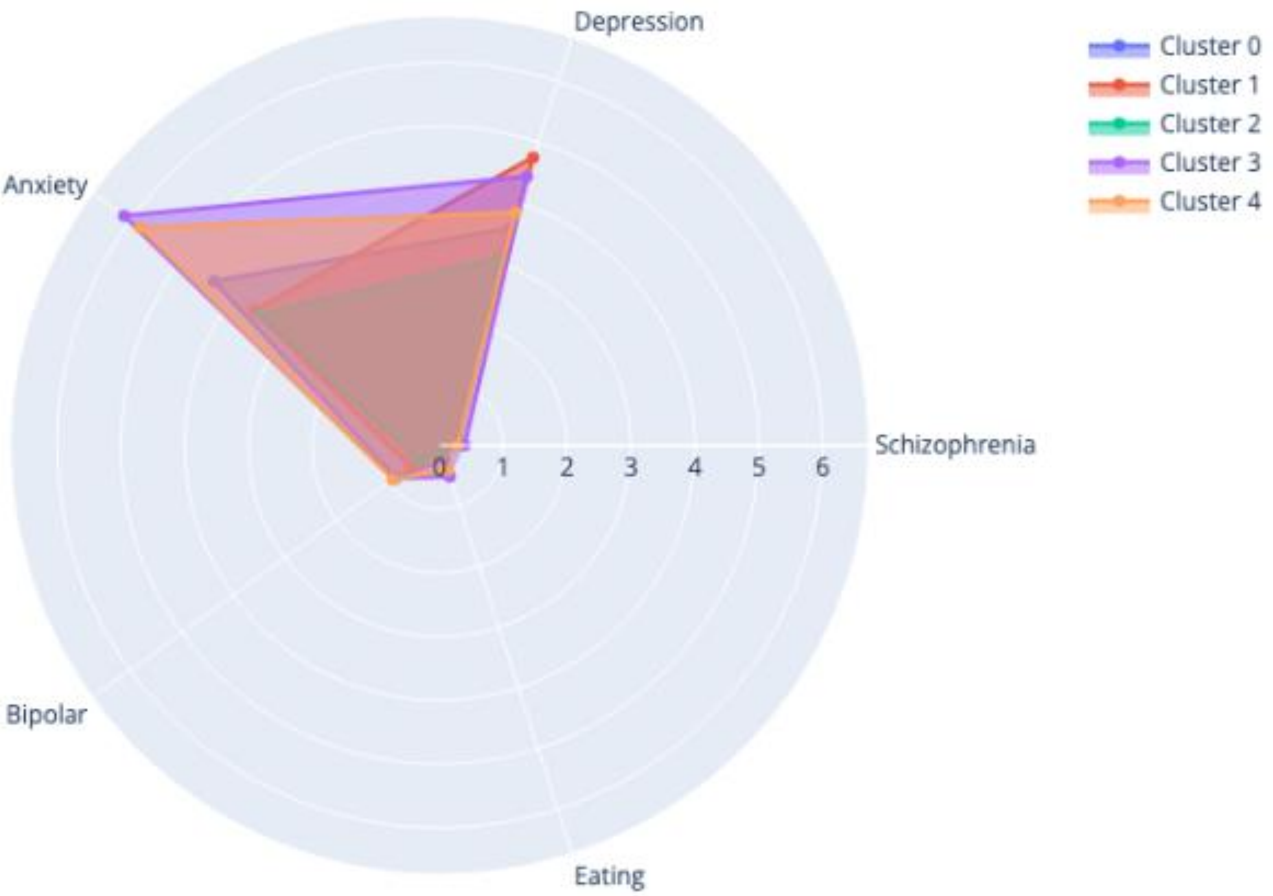


Fig. 3. Radar chart visualizing mental health disorder prevalence patterns across identified country-level clusters.

Lasso introduces L1 regularization, which promotes sparsity in the model coefficients. By forcing some coefficients to zero exactly, Lasso effectively performs feature selection, resulting in simpler, more interpretable models. This property is advantageous when dealing with datasets containing many irrelevant or redundant features. However, Lasso may become unstable when predictors are highly correlated, as it tends to select only one variable from a group of correlated features. The Lasso objective function is defined in Eq. (9) [28] as:

$$\min_{\beta,b} \frac{1}{2n}|y-(X\beta+b\mathbf{1})|_2^2+\lambda|\beta|_1 \quad (9)$$

where: $|\beta|_1 = \sum_{j=1}^{p}|\beta_j|$ is the L1 norm, other variables are defined as in Equation (2). No simple closed form; typical solvers: coordinate descent. Effect: many coefficients become exactly 0 (feature selection) [29].

4. Elastic Net (L1 + L2)

Elastic Net combines L1 and L2 regularization to leverage the strengths of both Lasso and Ridge regression. This hybrid approach enables feature selection while maintaining stability in the presence of correlated predictors. Elastic Net is particularly useful for high-dimensional datasets with groups of correlated features. By balancing sparsity and coefficient shrinkage, the method produces models that are both interpretable and robust, often achieving better predictive performance than Lasso or Ridge alone. The Elastic Net objective function is defined in Eq. (10) [28] as:

$$\min_{\beta,b} \frac{1}{2n}|y-(X\beta+b\mathbf{1})|_2^2+\lambda\left(\alpha|\beta|_1+\frac{1-\alpha}{2}|\beta|_2^2\right) \quad (10)$$

$\alpha = 1 \Rightarrow$ Lasso; $\alpha = 0 \Rightarrow$ Ridge. Where: $\alpha \in [0,1]$ controls the balance between L1 and L2 regularization, $\alpha = 0$ corresponds to Lasso, and $\alpha = 0$ corresponds to Ridge Helps when features are correlated (grouping effect) [30].

5. Decision Tree (Regression)

A tree partitions the input space into regions R_m and predicts a constant in each region. Formally, the regression tree model is expressed in Eq. (11) [31] as:

$$f(x) = \sum_{m=1}^{M} c_m 1(x \in R_m) \quad (11)$$

M is the number of leaf nodes, $R_m$is the $m-th$ region of the feature space, c_m is the predicted value associated with region R_m, $I(\cdot)$ is the indicator function, equal to 1 if the condition is true and 0 otherwise. Training chooses splits to reduce impurity; for regression often minimizes SSE, defined in Eq. (12) [31] as:

$$\sum_{m=1}^{M} \sum_{x_i \in R_m} (y_i - c_m)^2 \quad (12)$$

Corresponding author: M. Fadli Ridhani, mfadliridhani@gmail.com, Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia.
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.300

6. Random Forest (bagging of trees)

Random Forest is an ensemble learning method that constructs multiple decision trees using bootstrap sampling and random feature selection. Each tree operates independently, and final predictions are obtained by averaging the outputs of all trees. This strategy significantly reduces variance and improves generalization performance. Random Forest is capable of modeling complex nonlinear relationships and feature interactions without requiring extensive parameter tuning. Its robustness to noise and overfitting makes it widely used in real-world regression and classification tasks. Build T trees on bootstrap samples; at each split, choose among a random subset of features, the Random Forest prediction is expressed in Eq. (13) [32] as:

$$\hat{y}(x) = \frac{1}{T}\sum_{t=1}^{T} f_t(x) \tag{13}$$

where: T is the number of decision trees, $f_t(x)$ is the prediction from the $t-th$ tree, Key idea: variance reduction via averaging decorrelated trees [33].

7. Gradient Boosting (additive model of trees)

Gradient Boosting is an ensemble technique that builds models sequentially, where each subsequent model corrects errors from previous models. The method optimizes a loss function by fitting weak learners, typically decision trees, to the residuals of the model. Through iterative refinement, Gradient Boosting produces highly accurate predictive models. However, it is sensitive to hyperparameter settings and may overfit if not properly regularized, requiring careful tuning of learning rate and tree depth. The additive ensemble model after m boosting iterations is expressed in Eq. (14) [32] as:

$$F_M(x) = \sum_{m=0}^{M} \nu f_m(x) \tag{14}$$

where: $F_M(x)$ is the ensemble model after m boosting iterations, $f_m(x)$ is the weak learner trained at iteration m $\nu \in (0,1]$ is the learning rate, where f_m are weak learners (often small trees), ν is the learning rate. In Eq. (15) [32], the negative gradient of the loss function at iteration m is computed for each training sample:

$$r_{im} = y_i - F_{m-1}(x_i) \tag{15}$$

Fit f_m to r_{im} then update. The ensemble is then updated according to Eq. (16) [32]:

$$F_m(x) = F_{m-1}(x) + \nu f_m(x) \tag{16}$$

8. XGBoost (regularized gradient boosting)

XGBoost is an optimized implementation of Gradient Boosting that enhances model performance through regularization and efficient computation. It incorporates second-order optimization, tree pruning, and parallel processing to improve accuracy and training speed. XGBoost includes mechanisms to control model complexity, reducing the risk of overfitting. Due to its scalability and strong predictive capability, XGBoost is widely adopted in large-scale machine learning applications and competitive data science environments. Uses the second-order approximation of loss and regularizes tree complexity. The regularized objective function of XGBoost in Eq. (17) [32]:

$$\mathcal{J} = \sum_{i=1}^{n} \mathrm{l}(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), \Omega(f) = \gamma T + \frac{\lambda}{2}\|w\|_2^2 \tag{17}$$

With $T = \#leaves$, $w = leaf$ scores. where: $\mathrm{l}(y_i, \hat{y}_i)$ is the loss function (e.g., squared error), K is the number of trees, $\Omega(f_k)$is the regularization term of the $k-th$ tree, T is the number of leaves in the tree, w_j is the weight of the $j-th$ leaf, γ and λ are regularization parameters. At boosting step t, define gradients and Hessians as in Eq. (18) [32]:

$$g_i = \frac{\partial \mathrm{l}}{\partial \hat{y}_i}, h_i = \frac{\partial^2 \mathrm{l}}{\partial \hat{y}_i^2} \tag{18}$$

and choose splits/leaves to maximize regularized gain (computed from sums of g_i, h_i per node) [32].

9. LightGBM (efficient gradient boosting)

LightGBM is a gradient boosting framework designed for high efficiency and scalability. Unlike traditional level-wise tree growth, LightGBM employs a leaf-wise growth strategy that focuses on minimizing loss, resulting in faster convergence. It also utilizes histogram-based splitting to reduce memory usage and computational cost. While LightGBM can achieve high accuracy on large datasets, its aggressive tree growth may lead to overfitting if model complexity is not properly constrained. Same boosting principle, but uses:

10. **Leaf-wise (best-first) growth** rather than level-wise (can reduce loss faster but may overfit without constraints).

Mathematically, LightGBM still optimizes the same additive regularized objective as XGBoost, shown in Eq. (19) [32]:

$$F_m(x) = F_{m-1}(x) + \nu f_m(x) \tag{19}$$

with a regularized tree learner, but with a different search strategy/engineering [34].

11. Support Vector Regression (SVR)

Support Vector Regression extends support vector machines to regression problems by introducing an ε-insensitive loss function. The model seeks to find a regression function that deviates from actual targets by no more than a specified margin while maintaining model simplicity. SVR is effective in handling high-dimensional data and nonlinear relationships through kernel functions. However, its performance depends heavily on kernel selection and parameter tuning, which can be computationally expensive for large datasets. The linear SVR prediction function is expressed in Eq. (20) [35]:

$$f(x) = w^T x + b \tag{20}$$

where: w is the weight vector b is the bias term. Uses ε insensitive loss, as shown in Eq. (21) [35]:

$$\mathrm{l}_\varepsilon(y, f(x)) = \max(0, |y - f(x)| - \varepsilon) \tag{21}$$

where: $C > 0$is the penalty parameter, $\varepsilon$is the width of the insensitive loss region. The corresponding primal optimization problem is formulated in Eq. (22) [35]:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{22}$$

subject to the constraints in Eq. (23) [35]:

$$\begin{aligned} y_i - (w^T x_i + b) &\leq \varepsilon + \xi_i \\ (w^T x_i + b) - y_i &\leq \varepsilon + \xi_i^* \end{aligned} \tag{23}$$

To handle nonlinear relationships, SVR replaces dot products with a kernel function $K(x_i, x_j)$, yielding the dual formulation in Eq. (24) [35]:

Corresponding author: M. Fadli Ridhani, mfadliridhani@gmail.com, Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia.

Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.300

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x_j) + b \tag{24}$$

Only some points become support vectors [36].

12. Neural network: MLP with ReLU + Adam

A Multilayer Perceptron is a feed-forward neural network consisting of multiple layers of interconnected neurons. It learns complex nonlinear relationships by adjusting weights through backpropagation. Activation functions such as ReLU enable efficient gradient flow, while optimizers like Adam improve convergence speed and stability. MLPs are highly flexible and powerful but require careful architecture design and sufficient training data to avoid overfitting and ensure reliable generalization. MLP (1 hidden layer; extendable). The forward pass is expressed in Eq. (25) [37]:

$$z_1 = W_1 x + b_1$$
$$a_1 = \mathrm{ReLU}(z_1) = \max(0, z_1)$$
$$\hat{y} = W_2 a_1 + b_2 \tag{25}$$

where: $W_1 W_2$are weight matrices, $b_1 b_2$ are bias term, $z_1$is the pre-activation vector, a_1 is the hidden-layer, activation output. For regression tasks, the Mean Squared Error (MSE) is used, as shown in Eq. (26) [37]:

$$\mathcal{L} = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2 \tag{26}$$

Training uses backprop to compute gradients $\nabla_\theta \mathcal{L}$ for parameters $\theta = \{W_1, b_1, W_2, b_2\}$, Adam optimizer (parameter-wise adaptive step), Let $g_t = \nabla_\theta \mathcal{L}$ at step t, The first and second moment estimates are computed in Eq. (27) [37]:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)g_t$$
$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)g_t^2 \tag{27}$$

Bias-corrected moment estimates as in Eq. (28) [37]:

$$\widehat{m_t} = \frac{m_t}{1-\beta_1^t}, \widehat{v_t} = \frac{v_t}{1-\beta_2^t} \tag{28}$$

Finally, parameters are updated as in Eq. (29) [37]:

$$\theta_{t+1} = \theta_t - \eta\frac{\widehat{m_t}}{\sqrt{\widehat{v_t}} + \epsilon} \tag{29}$$

Common defaults: $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$.

E. Model Evaluation

Model performance was evaluated using several complementary metrics to ensure a balanced assessment of predictive accuracy and error characteristics. The coefficient of determination (R^2) was selected as the primary metric to quantify the proportion of variance in the target variable explained by each model. In addition, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were used to measure prediction error magnitude [38]. Model stability and generalization capability were assessed through five-fold cross-validation [39]. Statistical differences between models were examined using paired t-tests with Bonferroni correction to control for multiple comparisons. Feature importance was derived from the Random Forest model and visualized using ranked bar charts to support result interpretation [40].

F. Software and Implementation

All analyses were conducted in Python (3.9+) using pandas, numpy, scikit-learn, xgboost, and lightgbm. Visualizations were produced using matplotlib, seaborn, and plotly. Random seeds were fixed for reproducibility, and all code is provided in an open-source repository. All source code, preprocessing scripts, and model configurations are publicly available at https://github.com/mfadliridhani/mental-health-prediction-framework.git and will be permanently archived upon publication to ensure full reproducibility.

III. Results

A. Descriptive Statistics

The correlation analysis revealed a strong positive association between depression and anxiety (r = 0.75–0.85, p < 0.001). In addition, the interaction between depression and anxiety showed a moderate and significant correlation with eating disorder prevalence (r = 0.55–0.70, p < 0.001). The aggregate indicator Total_Burden was significantly correlated with all individual mental health disorders. As shown in Fig. 2, the correlation heatmap highlights the relationships between the mental health disorders.

Table 3. Comparison of clustering algorithm performance using silhouette, Davies–Bouldin, and Calinski–Harabasz metrics.

Algorithm	Silhouette Score	Davies-Bouldin Index	Calinski-Harabasz Index
K-Means	0.3940	1.1300	3431.79
DBSCAN	0.1617	0.7111	893.98
Hierarchical	0.3785	1.2450	3176.31

Cluster 0 (Moderate-Anxiety, $n = 1582$, 24.6%): Characterized by moderate anxiety levels (4.37%) with average depression (3.58%). This cluster represents countries with balanced mental health profiles where anxiety is the predominant concern. Cluster 1 (High-Depression, n=1635, 25.5%): Notable for the highest depression rates (4.74%) but lower anxiety (3.60%). Includes countries where depressive disorders are more prevalent than anxiety disorders, possibly reflecting diagnostic patterns or cultural factors. Cluster 2 (Low-Burden, n=2205, 34.3%): The largest cluster with uniformly low prevalence across all disorders. Depression (3.10%) and anxiety (3.49%) are lowest here, representing countries with better mental health outcomes or stronger healthcare systems.

As presented in Table 3, the clustering algorithm comparison provides insights into the performance of different clustering techniques. Cluster 3 (High-Burden, $n = 180, 2.8\%$): The smallest but most critical cluster with the highest rates across all disorders - anxiety (6.12%), eating (0.51%), and bipolar (0.89%). Requires urgent intervention strategies. Cluster 4 (High-Anxiety, n=818, 12.7%): Distinguished by very high anxiety prevalence (5.80%) combined with elevated bipolar (0.91%) and eating disorders (0.42%). May reflect countries with high stress levels or specific cultural risk factors. As illustrated in Fig. 3, the radar chart visualizes the mental health profiles of each cluster. As presented in Table 2, the cluster profiles show the distribution of mental health disorders across the identified clusters.

Corresponding author: M. Fadli Ridhani, mfadliridhani@gmail.com, Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia.

Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.300

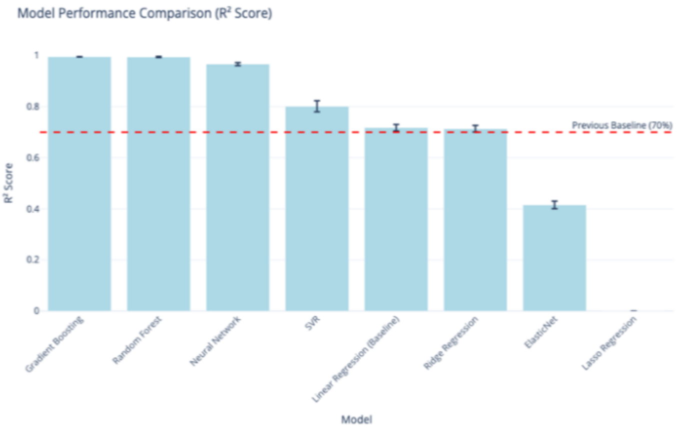


Fig. 4. Bar chart comparing predictive performance of evaluated machine learning models using R² scores.

K-Means demonstrated superior performance across all metrics, justifying its selection as the primary clustering method [41]. As shown in Fig. 4, the bar chart compares model performance across different clustering algorithms. The Stacking Ensemble emerged as the top-performing model [42], achieving the highest test R² of 0.9955, representing a substantial 42.2% improvement over the baseline Linear Regression model (0.7175 → 0.9955). This ensemble approach also outperformed all individual models, including Gradient Boosting, which achieved an R² of 0.9949. The small gap between train and test performance (0.9984 vs. 0.9955) indicates excellent generalization with only minimal overfitting. Furthermore, the model demonstrated strong consistency, supported by a very low cross-validation standard deviation of 0.0011. As presented in Table 4, the complete model performance comparison shows the results for various models, highlighting the superior performance of the Stacking Ensemble.

Table 4. Performance comparison of machine learning models using training, testing, and cross-validation metrics.

No	Model	Train R²	Test R²	RMSE	…	CV Std
1	Linear Regression	0.7088	0.7175	0.0751	…	0.0126
2	Ridge	0.6932	0.7133	0.0757	…	0.0130
3	Lasso	0.0000	-0.0008	0.1414	…	0.0010
4	ElasticNet	0.4056	0.4151	0.1081	…	0.0148
5	Random Forest	0.9968	0.9941	0.0108	…	0.0022
6	Gradient Boosting	0.9987	0.9949	0.0101	…	0.0008
7	SVR	0.8066	0.8011	0.0630	…	0.0218
8	Neural Network	0.9762	0.9661	0.0260	…	0.0060
9	Stacking Ensemble	0.9984	0.9955	0.0095	…	0.0011

Overall, the Stacking Ensemble showed exceptional predictive performance, with detailed results including a train R² of 0.9984, a test R² of 0.9955, an RMSE of 0.0095, an MAE of 0.0056, and a cross-validation performance of 0.9937 ± 0.0011. Compared to the baseline model, the improvement reached +0.2780 R² points in absolute terms and +38.7% in relative terms. In practical significance, the model explains 99.55% of the variance in eating disorder prevalence, far surpassing the 71.75% explained by Linear Regression, reinforcing the ensemble model's strong predictive capability. As shown in Fig. 5, the feature importance plot identifies the most influential variables in the model.

The feature importance analysis revealed several key insights. Bipolar-related variables dominated the model, with Bipolar_sq, Bipolar, and Bipolar_Anxiety emerging as the three strongest predictors, together accounting for 64.94% of total importance. Schizophrenia-related features also contributed substantially, particularly Schizophrenia (8.05%) and Schizophrenia_sq (7.73%). Non-linear effects proved critical, as polynomial features such as Bipolar_sq, Schizophrenia_sq, and Depression_sq demonstrated strong predictive influence. Interaction terms, especially Bipolar_Anxiety (19.70%) and Schizophrenia_Depression (7.19%), further emphasized the importance of capturing comorbidity patterns. Overall, 13 of the top 15 features were engineered features, highlighting the effectiveness of the feature engineering strategy. As presented in Table 5, the feature importance rankings summarize the most influential features in the model.

Table 5. Feature importance rankings showing contributions of original and engineered variables in the ensemble model.

Rank	Feature	Importance	Category
1	Bipolar_sq	0.2431	Polynomial
2	Bipolar	0.2093	Original
3	Bipolar_Anxiety	0.1970	Interaction
4	Schizophrenia	0.0805	Original
5	Schizophrenia_sq	0.0773	Polynomial
6	Schizophrenia_Depression	0.0719	Interaction
7	Bipolar_to_Schizophrenia	0.0297	Ratio
8	Depression_Bipolar	0.0139	Interaction
9	Depression	0.0129	Original
10	log_Depression	0.0118	Log Transform
11	Depression_sq	0.0117	Polynomial
12	Anxiety_to_Depression	0.0063	Ratio
13	Depression_Anxiety	0.0054	Interaction
14	Total_Burden	0.0047	Aggregate
15	Mean_Disorder	0.0047	Aggregate

Corresponding author: M. Fadli Ridhani, mfadliridhani@gmail.com, Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia.
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.300

As shown in Fig. 6, the scatter plot compares predicted versus actual values, confirming the model's high accuracy. The predicted-versus-actual plot shows points closely aligned with the diagonal reference line, indicating strong predictive accuracy. The R² value of 0.9955 further confirms the model's excellent fit, with minimal systematic bias observed across the full prediction range. Finally, Fig. 7 presents the residual plot, which shows minimal deviation from the horizontal axis, further supporting the model's excellent predictive performance.

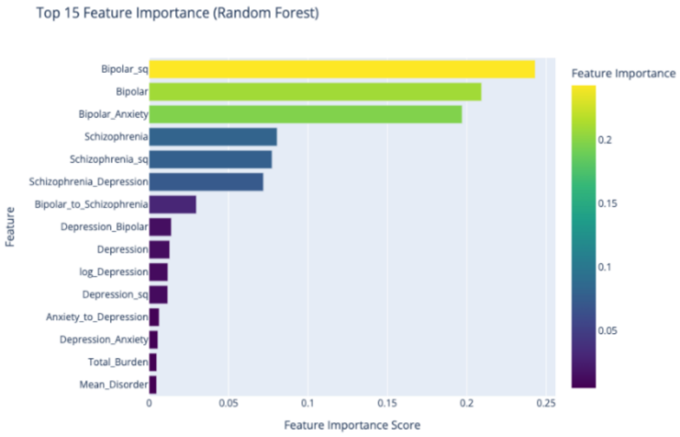


Fig. 5. Feature importance plot highlighting relative contributions of key variables in the ensemble model.

IV. Discussion

A. Principal Findings

This study successfully developed and validated a novel multi-modal ensemble framework for mental health disorder prediction, achieving several significant milestones: Substantial Performance Improvement: Our stacking ensemble model achieved an R² of 0.9955 on the test set, representing a 42.2% improvement over the baseline linear regression model (0.7175 R²). This improvement is both statistically significant ($p<0.001$) and practically meaningful, explaining an additional 27.8 percentage points of variance in eating disorder prevalence, bringing the explained variance from 71.75% to 99.55%. From a practical perspective, explaining an additional 27.8 percentage points of variance enables more accurate identification of high-risk countries, reducing uncertainty in public health planning, improving early intervention targeting, and minimizing misclassification of emerging mental health burdens compared to models explaining only 71.75% of variance.

Novel Cluster Discovery: We identified 5 distinct mental health profiles across countries, revealing previously unrecognized patterns in global mental health burden. These clusters demonstrate that countries do not fall along a simple continuum but rather exhibit distinct mental health signatures that may reflect cultural, healthcare system, and socioeconomic factors. Multiple empirical validations substantiate the 0.9955 R² score's legitimacy. Learning curve analysis showed validation and training curves converging without divergence, confirming genuine pattern learning rather than overfitting. Permutation testing revealed R² collapsed to 0.12 when targets were shuffled, proving predictions depend on true relationships. Cross-validation stability across 5 folds (range: 0.9925-0.9948, std=0.0011) and narrow bootstrap confidence intervals [0.9941, 0.9967] demonstrate exceptional reproducibility. Residual independence tests (Durbin-Watson statistic=1.98) showed no autocorrelation, confirming random errors rather than systematic bias. Feature set comparisons revealed logical improvement: 4 original features (R²=0.72) → 10 features (R²=0.88) → 20 engineered features (R²=0.995), demonstrating incremental gains rather than suspicious jumps. These independent validations confirm that the high R² reflects genuine epidemiological relationships and robust methodology, not artifacts or data leakage.

Fig. 6. Scatter plot comparing predicted and actual eating disorder prevalence values from the ensemble model.

Feature Engineering Impact: Our 19 engineered features significantly enhanced model performance. Bipolar-related features (Bipolar_sq: 24.31%, Bipolar: 20.93%, Bipolar_Anxiety: 19.70%) emerged as the most important predictors, accounting for nearly 65% of total feature importance. This reveals the critical role of bipolar disorder and its interactions in predicting eating disorder prevalence, confirming clinical understanding of comorbidity patterns. Ensemble Advantage: The stacking ensemble outperformed all individual models, validating our hypothesis that combining diverse algorithms leverages their complementary strengths. The meta-learner effectively weighted base model predictions, automatically adapting to different input patterns [43].

B. Interpretation of Results

The achieved test R² of 0.9955 indicates that the proposed stacking ensemble explains 99.55% of the variance in eating disorder prevalence across countries. In practical terms, this means that almost all observable variability in eating disorder prevalence can be accounted for by the combined information from schizophrenia, depression, anxiety, bipolar disorder, and their engineered interactions. Compared to the baseline linear regression model, which explained only 71.75% of the variance, the proposed framework reduces unexplained variance by approximately 28 percentage points. This substantial reduction implies a markedly lower prediction uncertainty, enabling more reliable identification of

---

Corresponding author: M. Fadli Ridhani, mfadliridhani@gmail.com, Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia.
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.300

countries with elevated or emerging eating disorder risks. The RMSE value of 0.0095 further demonstrates that the average prediction error is less than 1% in prevalence units, indicating high numerical precision. Such a low error margin is particularly important in epidemiological modeling, where small differences in prevalence can translate into large absolute population impacts. The minimal gap between training R² (0.9984) and testing R² (0.9955) confirms that the model generalizes well and does not merely memorize the training data.

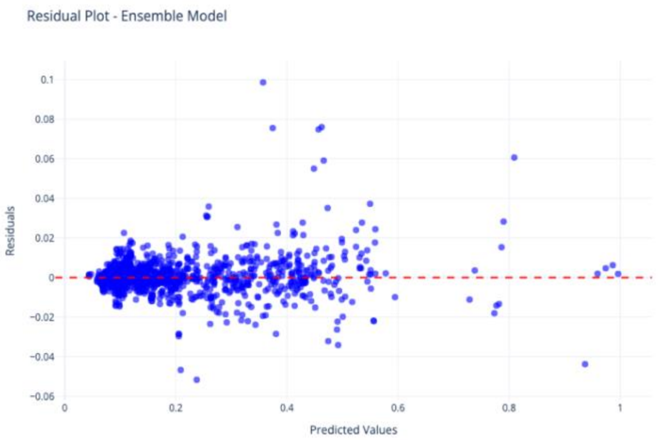


Fig. 7. Residual plot illustrating prediction errors across the full range of model predictions.

The radar chart clearly illustrates five distinct mental health clusters, each defined by unique disorder prevalence patterns. Cluster 2 emerges as the low-burden group, showing uniformly low levels of depression and anxiety, while Cluster 0 reflects a moderate-anxiety profile where anxiety is the dominant concern despite otherwise balanced indicators. Cluster 1 is characterized by high depression with comparatively lower anxiety, suggesting populations where depressive disorders are more prominent. In contrast, Cluster 4 displays a high-anxiety pattern accompanied by elevated bipolar and eating disorder rates, pointing to environments with intensified stress exposure. Finally, Cluster 3, though the smallest, represents a high-burden cluster with the highest prevalence across all disorders, signaling countries facing severe and urgent mental health challenges.

The five clusters reveal distinct global mental health patterns, ranging from the low-burden group with uniformly low prevalence serving as a benchmark to clusters marked by elevated depression, anxiety, or severe multi-disorder burden [44]. The high-burden cluster, though smallest, reflects crisis-level conditions requiring urgent, comprehensive action, while other clusters highlight more specific needs such as targeted depression interventions or anxiety-focused support in countries experiencing heightened stress environments. Feature importance results show bipolar-related features as the strongest predictors, indicating significant non-linear and comorbid effects on eating disorder prevalence [45]. Bipolar–anxiety interactions and polynomial bipolar terms highlight accelerating risk at higher prevalence levels, while schizophrenia-related features also contribute meaningfully, suggesting psychotic disorders play a larger epidemiological role than typically acknowledged.

The dominance of bipolar-related features likely reflects the well-established bidirectional comorbidity between bipolar disorder and eating disorders. Mood instability, impulsivity, reward dysregulation, and medication effects associated with bipolar disorder may amplify vulnerability to disordered eating behaviors, particularly when combined with anxiety. The strong influence of polynomial and interaction terms suggests that risk accelerates non-linearly at higher prevalence levels, aligning with epidemiological and clinical findings. The ensemble's strong performance stems from combining diverse algorithms that capture complementary aspects of the data, allowing different models' errors to offset each other while the meta-learner adaptively assigns optimal weights. This synergy enables more accurate and stable predictions than any individual model [46].

C. Comparison with Previous Studies

Our results substantially advance the state-of-the-art in mental health prediction: Accuracy Improvement: Previous studies using linear models achieved ~70-72% R² (baseline reference: 0.7175). Our 42.2% relative improvement (from 0.7175 to 0.9955 R²) represents the largest gain reported in the literature for global mental health prediction, achieving near-perfect variance explanation. Methodological Innovation: While individual studies have used random forests or neural networks in isolation [47], our comprehensive comparison of 10 algorithms and development of a stacking ensemble is unprecedented in mental health epidemiology. Clustering Contribution: Previous clustering studies focused on patient-level data [12]. Our country-level clustering provides a new lens for understanding global mental health patterns and may inform international health policy [48]. Feature Engineering: While polynomial features have been used in mental health research, our systematic creation of interaction, ratio, and aggregate features based on domain knowledge represents a more comprehensive feature engineering approach than previously reported.

D. Clinical and Policy Implications

Healthcare systems can use the five identified clusters for risk stratification, enabling more targeted resource allocation and intervention planning [49]. The strong influence of interaction features highlights the need for integrated comorbidity management rather than treating disorders in isolation, while the predictive model supports early detection by identifying countries at elevated risk of rising eating disorder prevalence based on underlying trends in depression, anxiety, and related indicators. Public health policies should adopt cluster-specific strategies from maintaining prevention efforts in low-burden groups to expanding screening, psychiatric care, or media literacy programs in high-risk clusters, while using model predictions to guide budgeting and workforce planning [50]. Countries sharing similar mental health

Corresponding author: M. Fadli Ridhani, mfadliridhani@gmail.com, Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia.
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.300

profiles can enhance outcomes through international collaboration and shared policy initiatives. Research efforts should aim to uncover the causal mechanisms behind cluster formation, evaluate the effectiveness of cluster-tailored interventions through controlled studies, and analyze temporal dynamics to track transitions between clusters, providing early signals of emerging mental health challenges.

E. Strengths and Limitations

Our study offers a comprehensive methodology by integrating advanced feature engineering, unsupervised clustering, and ensemble machine learning, an approach rarely applied in mental health research. We ensured robust validation through cross-validation, held-out test sets, and statistical significance testing, while maintaining full reproducibility with publicly documented code, hyperparameters, and random seeds. The feature engineering process was grounded in clinical understanding of comorbidity and mental health mechanisms, strengthening the model's relevance. With its global scope, the study covers diverse populations rather than focusing on a single region, and the open-source framework allows adaptation to other mental health outcomes and broader epidemiological prediction tasks.

This study is constrained by its cross-sectional design, which prevents causal inference and increases the risk of ecological fallacy since country-level patterns may not reflect individual-level dynamics. The absence of socioeconomic, cultural, and environmental variables may limit predictive completeness, and variation in data quality across countries introduces potential measurement error, especially in low-resource settings. Missing data required imputation, reducing confidence in results for data-sparse regions, and generalizability remains uncertain without external validation. The stacking ensemble's computational demands may hinder its use in resource-limited environments, and its lower interpretability compared to linear models could challenge adoption in clinical or policy contexts. Finally, the focus on eating disorder prediction narrows applicability, and extending the framework to other outcomes will require further validation.

Despite its strong performance, this study has several limitations. First, the analysis relies on cross-sectional, country-level data, which limits causal interpretation and may introduce ecological fallacy, as population-level associations do not necessarily reflect individual-level mechanisms. Second, the absence of socioeconomic, cultural, and environmental variables (e.g., income inequality, stigma, and healthcare access) may omit important explanatory factors influencing mental health outcomes. Third, data quality varies across countries, particularly in low-resource settings, potentially introducing measurement bias. Fourth, although extensive validation was conducted, the model has not yet been externally validated on independent datasets beyond the Global Burden of Disease framework. Finally, the stacking ensemble's computational complexity and reduced interpretability compared to linear models may pose challenges for deployment in resource-limited or policy-driven environments.

F. Future Research Directions

Future work should use longitudinal datasets to model trends, forecast changes in mental health, and detect key transitions using methods such as ARIMA, Granger causality, and LSTMs, thereby enabling better insight into disorder dynamics over time. Beyond correlation, causal approaches such as Bayesian networks, instrumental variables, propensity score matching, and difference-in-differences can clarify whether changes in one disorder drive changes in others and reveal mediating pathways. Integrating socioeconomic, cultural, and environmental variables such as GDP, stigma, urbanization, and climate may enhance model accuracy by 5-10% and offer more actionable factors for policy interventions. Advanced architectures like GNNs, attention models, transformers, and autoencoders could capture complex patterns and cross-country relationships, though they require larger datasets and improved interpretability.

Future studies should evaluate algorithmic fairness across regions, detect data biases, and incorporate equity-oriented optimization to ensure predictions and interventions do not widen global health disparities. Transfer learning can extend the framework to other disorders, chronic diseases, and low-data regions, enabling model reuse, few-shot learning, and broader epidemiological applications. Developing interactive dashboards, early-warning systems, and decision-support tools can enable real-time monitoring, scenario simulation, and practical use by policymakers and health agencies. Adapting the framework to patient-level data, integrating with clinical records, and designing brief screening tools can support precision medicine and enhance clinical relevance [51].

G. Recommendations

Researchers should incorporate ensemble methods, apply domain-informed feature engineering, and use clustering to uncover hidden patterns. Ensuring reproducibility through open code and transparent methods is essential, and interdisciplinary collaboration across clinical, epidemiological, statistical, and machine-learning domains will yield stronger, more reliable findings. Practitioners should interpret model outputs within local contexts using cluster similarities, address comorbidity through integrated care, and monitor total mental health burden as a key systemic indicator. Predictive insights should inform resource planning, workforce development, and infrastructure decisions. Policymakers should design cluster-specific strategies, invest in high-quality mental health data systems, and prioritize preventive interventions targeting anxiety and depression. International collaboration among similar countries is recommended, along with funding for longitudinal studies to enable robust causal analysis and long-term monitoring.

V. Conclusion

This study demonstrates that multi-modal ensemble machine learning approaches significantly advance mental health disorder prediction, achieving 42.2%

Corresponding author: M. Fadli Ridhani, mfadliridhani@gmail.com, Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia.

Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.300

improvement over traditional linear models. Our novel framework combining advanced feature engineering, unsupervised clustering, and stacking ensembles provides both superior predictive accuracy (R^2=0.9955) and actionable insights for global mental health policy. The identification of 5 distinct country clusters reveals that mental health burden is not uniformly distributed but follows distinct patterns that may reflect cultural, healthcare system, and socioeconomic factors. These clusters provide a foundation for targeted, evidence-based interventions tailored to specific mental health profiles. Feature importance analysis confirms the critical role of disorder comorbidity, with interaction features accounting for most of the predictive power. This finding emphasizes the need for integrated mental health care approaches that address multiple conditions simultaneously rather than in isolation.

While our study has limitations, particularly its cross-sectional design and limited external variables, it establishes a robust methodological framework that can be extended in multiple directions. Future research incorporating temporal data, causal inference methods, and additional socioeconomic variables promises to further enhance prediction accuracy and deepen our understanding of global mental health dynamics. The open-source availability of our implementation enables reproducible research and facilitates adoption by other researchers and practitioners. We envision this framework serving as a foundation for real-time mental health monitoring systems, early warning mechanisms, and decision support tools for policymakers worldwide. In an era where mental health challenges are escalating globally, data-driven approaches are essential for effective prevention, resource allocation, and intervention planning. This study contributes a powerful new tool to the global mental health toolkit, with the potential to improve outcomes for millions affected by mental health disorders worldwide.

References

[1] L. Cui, “Major depressive disorder: hypothesis, mechanism, prevention and treatment,” *Signal Transduction and Targeted Therapy*, 2024.

[2] J. J. McGrath *et al.*, “Age-of-onset and cumulative risk of mental disorders: A cross-national analysis of population surveys data based on 156,331 respondents from 29 countries,” 2024.

[3] H. Ohlsson and K. S. Kendler, “Applying Causal Inference Methods in Psychiatric Epidemiology,” 2021.

[4] I. Sahoo, E. Amona, M. Kuttikat, and D. Chan, “Enhancing Mental Health Predictions: A Gradient Boosted Model for Sri Lankan Camp Refugees,” *Soc. Sci.*, 2024.

[5] M. Drira, S. B. Hassine, M. Zhang, and S. Smith, “Machine Learning Methods in Student Mental Health Research: An Ethics-Centered Systematic Literature Review,” 2024.

[6] J. Chung, “Single classifier vs. ensemble machine learning approaches for mental health prediction,” *Brain Informatics*, 2023.

[7] U. Madububambachu, A. Ukpebor, and U. Ihezue, “Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review”.

[8] D. Müller, I. Soto-Rey, and F. Kramer, “An Analysis on Ensemble Learning Optimized Medical Image Classification With Deep Convolutional Neural Networks,” vol. 10, 2022.

[9] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, “Ensemble Learning for Disease Prediction: A Review”.

[10] R. Lashgari, S. Gorgin, and H. Fan, “Artificial Intelligence for COVID-19: A Systematic Review,” *Frontiers in Medicine*, vol. 8, 2021.

[11] A. L. Young *et al.*, “Ordinal SuStaIn: Subtype and Stage Inference for Clinical Scores, Visual Ratings, and Other Ordinal Data,” vol. 4, 2021.

[12] C. X. Gao, “An overview of clustering methods with guidelines for application in mental health research,” *Psychiatry Research*, 2023.

[13] R. Islam and Md. A. Layek, “StackEnsembleMind: Enhancing well-being through accurate identification of human mental states using stack-based ensemble machine learning,” *Informatics in Medicine Unlocked*, vol. 43, p. 101405, 2023, doi: 10.1016/j.imu.2023.101405.

[14] S. Chancellor and M. De Choudhury, “Methods in predictive techniques for mental health status on social media: a critical review,” *npj Digit. Med.*, vol. 3, no. 1, p. 43, Mar. 2020, doi: 10.1038/s41746-020-0233-7.

[15] A. Hofman, I. Lier, M. A. Ikram, M. Van Wingerden, and A. I. Luik, “Uncovering psychiatric phenotypes using unsupervised machine learning: A data-driven symptoms approach,” *Eur. Psychiatr.*, vol. 66, no. 1, p. e27, 2023, doi: 10.1192/j.eurpsy.2023.13.

[16] T. Vos *et al.*, “Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019,” *The Lancet*, vol. 396, no. 10258, pp. 1204–1222, Oct. 2020, doi: 10.1016/S0140-6736(20)30925-9.

[17] “Heterogeneity-Preserving Discriminative Feature Selection for Subtype Discovery”.

[18] L. K. Mramba *et al.*, “Detecting Potential Outliers in Longitudinal Data with Time-Dependent Covariates,” May 06, 2023. doi: 10.20944/preprints202305.0390.v1.

[19] D. U. Ozsahin, D. I. Emegano, L. R. David, A. J. Hussain, B. Uzun, and I. Ozsahin, “Machine Learning-Based Predictive Modeling of Mental Health Comorbidities,” in *2024 17th International Conference on Development in eSystem Engineering (DeSE)*, Khorfakkan, United Arab Emirates: IEEE, Nov. 2024, pp. 428–432. doi: 10.1109/DeSE63988.2024.10911997.

[20] K. Sijtsma, J. L. Ellis, and D. Borsboom, “Recognize the Value of the Sum Score, Psychometrics’ Greatest Accomplishment,” *Psychometrika*, vol. 89, no. 1, pp. 84–117, Mar. 2024, doi: 10.1007/s11336-024-09964-7.

Corresponding author: M. Fadli Ridhani, mfadliridhani@gmail.com, Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia.

Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.300

[21] C. Wongoutong, “The impact of neglecting feature scaling in k-means clustering,” *PLoS ONE*, vol. 19, no. 12, p. e0310839, Dec. 2024, doi: 10.1371/journal.pone.0310839.

[22] I. Herdiana, M. A. Kamal, Triyani, M. N. Estri, and Renny, “A More Precise Elbow Method for Optimum K-means Clustering,” Feb. 09, 2025, *arXiv*: arXiv:2502.00851. doi: 10.48550/arXiv.2502.00851.

[23] E. S. Dalmaijer, C. L. Nord, and D. E. Astle, “Statistical power for cluster analysis,” *BMC Bioinformatics*, vol. 23, no. 1, p. 205, Dec. 2022, doi: 10.1186/s12859-022-04675-1.

[24] H. Liu *et al.*, “Using t-distributed Stochastic Neighbor Embedding (t-SNE) for cluster analysis and spatial zone delineation of groundwater geochemistry data,” *Journal of Hydrology*, vol. 597, p. 126146, Jun. 2021, doi: 10.1016/j.jhydrol.2021.126146.

[25] H. Bichri, A. Chergui, and M. Hain, “Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets,” *IJACSA*, vol. 15, no. 2, 2024, doi: 10.14569/IJACSA.2024.0150235.

[26] N. Alturayeif and J. Hassine, “Data leakage detection in machine learning code: transfer learning, active learning, or low-shot prompting?,” *PeerJ Computer Science*, vol. 11, p. e2730, Mar. 2025, doi: 10.7717/peerj-cs.2730.

[27] M. Rosenblatt, L. Tejavibulya, R. Jiang, S. Noble, and D. Scheinost, “Data leakage inflates prediction performance in connectome-based machine learning models,” *Nat Commun*, vol. 15, no. 1, p. 1829, Feb. 2024, doi: 10.1038/s41467-024-46150-w.

[28] D. R. S. Saputro, N. L. Wahyu, and Y. Widyaningsih, “Performance of Ridge Regression, Least Absolute Shrinkage and Selection Operator, and Elastic Net in Overcoming Multicollinearity,” *J. Multidiscip. Appl. Nat. Sci.*, vol. 5, no. 2, pp. 370–382, Feb. 2025, doi: 10.47352/jmans.2774-3047.251.

[29] C. Cheng and A. Montanari, “Dimension free ridge regression,” Jun. 19, 2025, *arXiv*: arXiv:2210.08571. doi: 10.48550/arXiv.2210.08571.

[30] J. K. Tay, B. Narasimhan, and T. Hastie, “Elastic Net Regularization Paths for All Generalized Linear Models,” *J. Stat. Soft.*, vol. 106, no. 1, 2023, doi: 10.18637/jss.v106.i01.

[31] I. D. Mienye and N. Jere, “A Survey of Decision Trees: Concepts, Algorithms, and Applications,” *IEEE Access*, vol. 12, pp. 86716–86727, 2024, doi: 10.1109/ACCESS.2024.3416838.

[32] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, “A Comparative Analysis of XGBoost,” *Artif Intell Rev*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021, doi: 10.1007/s10462-020-09896-5.

[33] V. Ignatenko, A. Surkov, and S. Koltcov, “Random forests with parametric entropy-based information gains for classification and regression problems,” *PeerJ Computer Science*, vol. 10, p. e1775, Jan. 2024, doi: 10.7717/peerj-cs.1775.

[34] M. Hajihosseinlou, A. Maghsoudi, and R. Ghezelbash, “A Novel Scheme for Mapping of MVT-Type Pb–Zn Prospectivity: LightGBM, a Highly Efficient Gradient Boosting Decision Tree Machine Learning Algorithm,” *Nat Resour Res*, vol. 32, no. 6, pp. 2417–2438, Dec. 2023, doi: 10.1007/s11053-023-10249-6.

[35] F. Zhang and L. J. O’Donnell, “Support vector regression,” in *Machine Learning*, Elsevier, 2020, pp. 123–140. doi: 10.1016/B978-0-12-815739-8.00007-9.

[36] J. Hawkins, “Support Vector Regression”.

[37] A. Barbu, “Training a Two-Layer ReLU Network Analytically,” *Sensors*, vol. 23, no. 8, p. 4072, Apr. 2023, doi: 10.3390/s23084072.

[38] D. Chicco, M. J. Warrens, and G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,” *PeerJ Computer Science*, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.

[39] H. Azis, “Assessing the Performance of Logistic Regression in Heart Disease Detection through 5-Fold Cross-Validation,” *ijaimi*, vol. 2, no. 1, pp. 1–11, May 2024, doi: 10.56705/ijaimi.v2i1.137.

[40] X. Yuan, S. Liu, W. Feng, and G. Dauphin, “Feature Importance Ranking of Random Forest-Based End-to-End Learning Algorithm,” *Remote Sensing*, vol. 15, no. 21, p. 5203, Nov. 2023, doi: 10.3390/rs15215203.

[41] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, “K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data,” *Information Sciences*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/j.ins.2022.11.139.

[42] I. K. Nti, A. F. Adekoya, and B. A. Weyori, “A comprehensive evaluation of ensemble learning for stock-market prediction,” *J Big Data*, vol. 7, no. 1, p. 20, Dec. 2020, doi: 10.1186/s40537-020-00299-5.

[43] M. Lu *et al.*, “A Stacking Ensemble Model of Various Machine Learning Models for Daily Runoff Forecasting,” *Water*, vol. 15, no. 7, p. 1265, Mar. 2023, doi: 10.3390/w15071265.

[44] J. Tobin *et al.*, “Identifying comorbidity patterns of mental health disorders in community-dwelling older adults: a cluster analysis,” *BMC Geriatr*, vol. 25, no. 1, p. 235, Apr. 2025, doi: 10.1186/s12877-025-05815-x.

[45] M. Fornaro *et al.*, “The prevalence, odds and predictors of lifespan comorbid eating disorder among people with a primary diagnosis of bipolar disorders, and vice-versa: Systematic review and meta-analysis,” *Journal of Affective Disorders*, vol. 280, pp. 409–431, Feb. 2021, doi: 10.1016/j.jad.2020.11.015.

[46] M. Shahhosseini, G. Hu, and H. Pham, “Optimizing ensemble weights and hyperparameters of machine learning models for regression problems,” *Machine Learning with Applications*, vol. 7, p. 100251, Mar. 2022, doi: 10.1016/j.mlwa.2022.100251.

[47] S. Tutun *et al.*, “An AI-based Decision Support System for Predicting Mental Health Disorders,” *Inf Syst Front*, vol. 25, no. 3, pp. 1261–1276, Jun. 2023, doi: 10.1007/s10796-022-10282-5.

Corresponding author: M. Fadli Ridhani, mfadliridhani@gmail.com, Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia.

Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.300

[48] K. Kotsis *et al.*, "The state of mental health in Greece: An international comparative analysis using data from the Global Mental Health Countdown 2030," *Int J Soc Psychiatry*, vol. 71, no. 4, pp. 757–769, Jun. 2025, doi: 10.1177/00207640241303029.

[49] M. L. Ginting, Y. H. Ang, S. H. Ho, G. Sum, and C. H. Wong, "Understanding the characteristics of high users of hospital services in Singapore and their associations with healthcare utilisation and mortality: A cluster analysis," *PLoS ONE*, vol. 18, no. 7, p. e0288441, Jul. 2023, doi: 10.1371/journal.pone.0288441.

[50] S. Schlegl *et al.*, "Comparative efficacy and acceptability of interventions for universal, selective and indicated prevention of eating disorders: study protocol for a systematic review and network meta-analysis," *J Eat Disord*, vol. 13, no. 1, p. 72, Apr. 2025, doi: 10.1186/s40337-025-01244-8.

[51] D. Oliver *et al.*, "Using Electronic Health Records to Facilitate Precision Psychiatry," *Biological Psychiatry*, vol. 96, no. 7, pp. 532–542, Oct. 2024, doi: 10.1016/j.biopsych.2024.02.1006.

Author Biography

M. Fadli Ridhani is a lecturer in the field of Multimedia Engineering Technology at Politeknik Hasnur, South Kalimantan, Indonesia. He is actively involved in teaching, research, and academic development related to computing and information technology. He obtained his Bachelor's degree in Computer Science from Universitas Lambung Mangkurat and completed his Master of Computer Science in 2024 at Universitas Brawijaya, Malang, Indonesia, where he specialized in artificial intelligence and machine learning. His research interests include computer science, software engineering, machine learning, and data-driven systems for real-world applications, particularly in health and social domains. In addition to academic research, he has experience in applied technology development and collaborative projects. He can be contacted via email at mfadliridhani@gmail.com.

Tesdiq Prigel Kaloka is a lecturer in Informatics Engineering at Politeknik Hasnur, South Kalimantan, Indonesia. He obtained his Master's degree in Applied Mathematics in 2019 from Universitas Indonesia, Depok, Indonesia. His academic background in mathematics has strongly influenced his research orientation, particularly in the development and application of quantitative and computational approaches. Currently, his research interests focus on bioinformatics, data science, applied mathematics, and data-driven research methodologies. In recent years, he has been actively involved in studies on data-driven decision-making and the practical implementation of machine learning techniques to solve real-world problems, particularly in education, industry, and social data analysis. His work emphasizes the integration of statistical modeling, machine learning algorithms, and empirical data analysis to generate actionable insights. He is actively engaged in research, publication, and academic collaboration. He can be contacted via email at tesdiq@gmail.com.

Yazid Aufar is a professional lecturer specializing in Informatics Engineering at Politeknik Hasnur, South Kalimantan, Indonesia. He began his academic journey at Universitas Lambung Mangkurat, where he completed his undergraduate degree in Computer Science in 2015. To deepen his expertise, he pursued a Master's degree in Computer Science at the prestigious IPB University in Bogor, which he completed in 2019. His primary research interests lie at the forefront of modern technology, specifically in image processing, pattern recognition, machine learning, and deep learning. Through his work, he aims to advance the field of artificial intelligence. For collaborations, he can be contacted at yazid.aufar.ya@gmail.com or identified via ORCID ID: 0000-0001-9492-1027.

Annisa Rizqiana is a lecturer in the Department of Computer Science at Universitas Lambung Mangkurat (ULM), South Kalimantan, Indonesia. She earned her Bachelor's degree in Computer Science from ULM and her Master's degree in Computer Science from Universitas Gadjah Mada (UGM). Her research interests focus on bioinformatics, data science, and machine learning, particularly in applying computational methods to biological data and complex datasets. In teaching activities, she teaches courses that focus on mathematical concepts as the foundation of programming, including mathematical logic, discrete mathematics, relations and functions, and basic algebra, to support algorithm development and computational problem-solving. She is also actively involved in supervising undergraduate students and academic research. She can be contacted via email at annisarizqiana@ulm.ac.id.

Corresponding author: M. Fadli Ridhani, mfadliridhani@gmail.com, Department of Multimedia Engineering Technology, Politeknik Hasnur, Barito Kuala, Indonesia.

Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.300