Institute of Research and Publication Indonesia (IRPI)
Public Research Journal of Engineering, Data Technology and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/predatecs
Vol. 3 Iss. 2 January 2026, pp: 150-164
ISSN(P): 3024-921X | ISSN(E): 3024-8043

Depression Classification in University Students Using A Machine Learning Approach Based on Multi-Layer Perceptron

Fatimah Azzahra[1], Muhammad Rafiq Pohan[2*], Ainul Mardhiah Binti Mohammed Rafiq[3], Imran Hazim Bin Abdullah Salim[4], Azwa Nurnisya Binti Ayub[5], Nuralya Medina Binti Mohammad Nizam[6]

[1,2]Department of Information System, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
[3,4]Department of Computer Science, Faculty of Information and Communication Technology, International Islamic University Malaysia, Malaysia
[5,6]Department of Information System, Faculty of Information and Communication Technology, International Islamic University Malaysia, Malaysia

E-Mail:[1]12250324568@students.uin-suska.ac.id, [2]12250311460@students.uin-suska.ac.id, [3]mrdhainul06@gmail.com, [4]imranhazim02@gmail.com, [5]Azwanurnisya612@gmail.com, [6]medinamohdnizam@gmail.com

Received Jun 12th 2025; Revised Jan 15th 2026; Accepted Jan 27th 2026; Available Online Jan 31th 2026
Corresponding Author: Muhammad Rafiq Pohan
Copyright © 2026 by Authors, Published by Institute of Research and Publication Indonesia (IRPI)

Abstract

Depression among university students is a critical mental health concern, often exacerbated by academic pressure and social adaptation. While prior studies have utilized Multi-Layer Perceptron (MLP) models to achieve up to 78% accuracy, the effectiveness of these systems remains highly sensitive to architectural design and optimization strategies. To address this gap, this study systematically evaluates the performance of modern MLP architectural variants including DenseNet, ResMLP, and ResNet paired with SGD, Adam, and RMSprop optimizers. Using a dataset of 1,025 student records, the methodology integrates Chi-Square feature selection and Min-Max normalization, followed by an 80:20 Hold-Out validation. Results demonstrate that the ResNet-RMSprop synergy yields a superior accuracy of 83.86%, significantly outperforming traditional MLP benchmarks . By identifying the optimal combination of deep learning structures and optimization algorithms, this research provides a more robust and precise technical foundation for AI-driven early detection systems in academic settings.

Keywords: Classification, Hold-Out Validation, Mental Health, Multi-Layer Perceptron, Student Depression.

1. INTRODUCTION

In recent years, mental health issues, particularly depression, have garnered growing global concern [1]. According to a WHO report, over 350 million individuals worldwide are affected by depression, with the numbers increasing annually [2]. Depression frequently co-occurs with anxiety disorders, with approximately 40–60% of individuals suffering from major depressive disorder also experiencing anxiety-related symptoms [2]. His condition is often characterized by prolonged sadness, a loss of interest in activities, and diminished emotional and physical functioning [3]. Common physical symptoms include disrupted sleep patterns, appetite loss, and fatigue, while cognitive symptoms may include slowed thinking, suicidal ideation, and excessive guilt [4].

University students are among the groups most vulnerable to mental health issues, including depression. The academic workload, social challenges, and the need to adapt to a new university environment contribute significantly to their mental health struggles [5]. Research by Meda et al. (2023) [6] revealed a high prevalence of depression among students, with about 17–22% exhibiting severe symptoms. More recent studies, by Mumenin et al. (2024) [7] have reported even higher rates of depression in low- and middle-income countries: 43.7% of students in India, 40.9% in Pakistan, and 52.2% in Bangladesh have shown depressive symptoms. Hence, it is crucial to develop technology-based early detection systems to enable timely and targeted interventions [8]

To predict depressive states, various approaches have been implemented, including Machine Learning (ML) and Deep Learning (DL) algorithms. Among the commonly used algorithms for depression classification is the Multi-Layer Perceptron (MLP) [9]. MLP is an artificial neural network composed of several fully connected layers of neurons [10]. According to Chung et al. (2022) [4], MLP has demonstrated an accuracy of up to 78% in predicting mental health disorders among children and an AUC of 88% in detecting prenatal depression [10]. Due to its capability to learn complex data patterns, MLP is considered a strong candidate for symptom-based psychological classification tasks [11]. However, most existing studies focus on standard MLP implementations, without exploring specialized architectural variants or their interactions with different optimization strategies, leaving a significant gap in identifying the most efficient configuration for student-specific data.

However, the performance of MLP models heavily depends on both the architectural design and the choice of optimization algorithms. Poor architectural decisions may lead to overfitting or weak generalization capabilities [12] while unsuitable optimizers can slow down training or prevent optimal performance [13]. A major challenge in building effective MLP models lies in identifying the right combination of architecture and optimizer to ensure accuracy, stability, and interpretability [14]. While previous research has often explored architectural models and optimization techniques separately, a comprehensive evaluation comparing modern architectures such as DenseNet, ResMLP, and ResNet across multiple optimizers (SGD, RMSprop, and Adam) remains largely unexplored in the context of student depression. This study addresses this gap by providing a systematic comparative analysis of these nine combinations. The novelty of this research lies in identifying the optimal synergy between residual-based architectures and adaptive optimization to enhance the sensitivity and accuracy of early mental health screening tools.

Several studies have proposed architectural variants such as Dense Layers, ResMLP, and ResNet. Dense layers are easy to implement but prone to overfitting without regularization techniques such as dropout or L2 regularization [10]. ResMLP incorporates residual learning to mitigate performance degradation in deeper networks [15]. Originally designed for image recognition tasks, ResNet utilizes shortcut connections between layers to enhance learning in very deep networks. Although primarily used for image data, ResNet has also yielded promising results with tabular datasets (Ravid Shwartz-Ziv, Pfister & Sercan, n.d.)

In addition to architectural choices, the selection of optimizers plays a key role in efficient model training. Three widely used optimizers in MLP training are Stochastic Gradient Descent (SGD), RMSprop, and Adam [17], [18]. While SGD is a foundational optimizer, it is sensitive to learning rates and often requires longer training times [13]. RMSprop offers better stability when dealing with data with high variability and is frequently used in DL models that require adaptive learning [19]–[21]. Adam has gained popularity for combining the advantages of both SGD and RMSprop, enabling faster, more stable training [10].

While previous research has explored architectural models and optimization techniques separately, a comprehensive evaluation comparing different architectures (Dense, ResMLP, and ResNet) and optimizers (SGD, RMSprop, and Adam) remains largely unexplored, particularly in the area of student depression classification. Accordingly, this research focuses on assessing how different combinations of MLP architectures and optimization algorithms perform in identifying depression levels among university students. The assessment uses evaluation metrics suitable for imbalanced data classification, such as accuracy, precision, recall, and F1-score. The outcomes are anticipated to provide data-driven insights that could help build AI-powered systems for the early identification of mental health issues.

2. MATERIAL AND METHOD

The methodological process begins by preparing the dataset, including selecting relevant features and applying normalization techniques. Afterward, the data is partitioned using an 80:20 Hold-Out Validation strategy. Distinct from conventional methodologies that typically focus on single-model optimization, the approach illustrated in Figure 1 implements a systematic grid-evaluation framework. The portion allocated for training is employed to build the MLP model using two primary approaches: employing different optimization algorithms (Adam, SGD, RMSprop) and implementing alternative network architectures (ResNet, DenseNet, ResMLP). By integrating specialized architectures like ResNet and ResMLP which are traditionally used for image data into a tabular data classification task, this methodology offers a novel perspective on improving screening accuracy. The performance of each training outcome is evaluated based on accuracy and is further analyzed comparatively to identify the most effective configuration before conclusions.

2.1. Data collecting

The dataset utilized in this study is the "Student Mental Health Dataset," sourced from the Kaggle platform. It contains information related to mental health issues among university students, with a particular focus on symptoms of depression. The dataset comprises 14 variables, including demographic data (such as age and gender), lifestyle habits, and daily routines (such as sleep patterns and exercise frequency). In total, there are 1,025 data entries. Additionally, the dataset includes mental health indicators that may influence the severity of depression among students.

The target variable in this dataset is a categorical attribute named "Depression," which is divided into two classes: "Yes" (indicating the student is experiencing depression) and "No" (indicating the student is not experiencing depression). This dataset is utilized to develop a predictive model aimed at identifying signs of depression in students through the application of machine learning techniques, specifically focusing on the Multi-Layer Perceptron (MLP) and its architectural variations.

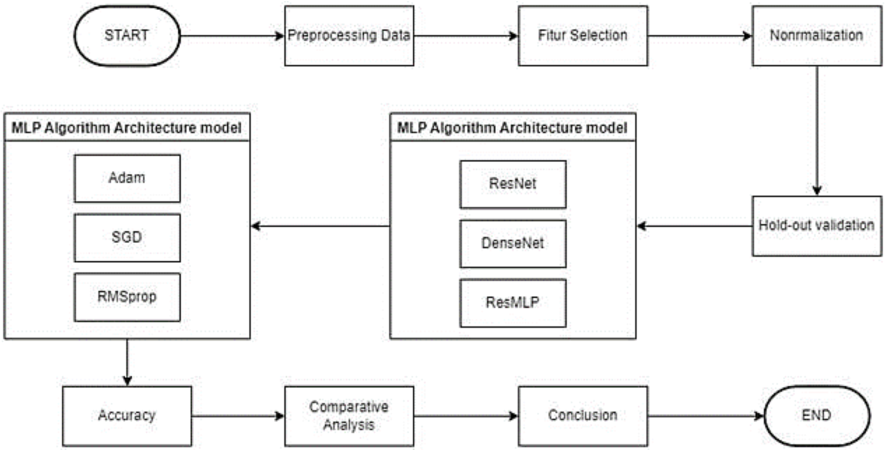

Figure 1. Research Methodology

2.2. Pre-processing Data

Before analyzing the data using machine learning models, a procedure known as data preprocessing is carried out, this step aims to clean and modify the dataset to ensure it is suitable and optimized for the training process [22]. The goal is to minimize potential errors or misinterpretations during data input by removing missing or inconsistent entries and reducing the number of attributes used in the classification process [23].

To improve model performance and reduce data dimensionality, this study applies the Chi-Square (χ^2) method as a feature selection technique. The Chi-Square test assesses the association between each feature and the target class, enabling the selection of statistically significant features [24]. In this research, the fundamental formula used for the Chi-Square calculation is equation 1.

$$\chi^2(t_i, c_j) = \frac{N\,(AD - CB)^2}{(A + C)(B + D)(A + B)(C + D)} \tag{1}$$

In the formula, value A represents the number of data entries in the class c_j that contain the featuret_i, while B is the number of entries outside the classc_j that also contain that feature. Value C denotes the number of entries in the class c_j that do not include feature t_i, and D is the count of entries outside class c_j that lack the feature. The total number of entries in the dataset is represented by N. The output of this computation indicates the strength of association between a given feature and the target class, which is then used to rank and select the most relevant features.

The selection of the Chi-Square method in this study is strategically based on its proficiency in handling the categorical data that dominates this mental health dataset. Unlike dimensionality reduction techniques such as PCA, which transform original features into uninterpretable components, Chi-Square preserves the original attributes like financial stress and family history. This provides a novel contribution in terms of model transparency, ensuring that the most influential features are statistically identified before being processed by complex architectures—a step often overlooked in conventional studies.

Subsequently, a normalization step is performed to harmonize the range of all feature values, preventing any single feature from disproportionately influencing the model's learning [26]. In this study, the Min-Max Scaling technique is adopted, which scales the data values to fall within a defined interval, typically 0 to 1. The transformation using Min-Max Scaling can be represented mathematically as equation 2.

$$N' = \frac{N - \min(n)}{(n) - \min(n)} \tag{2}$$

In the Min-Max normalization method, the adjusted value N' is calculated by taking the original data point N, subtracting $\min(n)$ (the minimum value of the feature), and then dividing the result by the range calculated from (n) minus min (n). This transformation yields values that fall within a defined range, generally from 0 to 1 [27].

2.3. Splitting data

The next stage in the model development process is dataset splitting, which is vital for providing an unbiased assessment of the model's effectiveness. In this study, the Hold-Out method is used as a widely adopted data-splitting approach in machine learning due to its simplicity and effectiveness for providing an initial performance assessment of the model.

The Hold-Out technique splits the dataset into two parts: a majority for training and a smaller portion for testing. In this research, an 80:20 split is used, with 80% dedicated to training and validation, and the remaining 20% reserved for testing. The split is performed randomly while maintaining class-balanced distributions across both subsets.

The Hold-Out approach is chosen for its ability to provide a low-bias performance estimate with minimal computational complexity, making it particularly suitable for datasets of medium to large size [28]. To support efficient training and testing, the partitioned data is organized into separate directories, facilitating further data handling and preventing data leakage between subsets.

2.4. Student Depression

Depression refers to a psychological condition marked by continuous feelings of sadness, loss of interest, and a decline in both emotional and physical functioning [3]. Common physical symptoms include fatigue, decreased appetite, and sleep disturbances. Individuals with depression may also experience cognitive symptoms such as excessive guilt, suicidal thoughts, and difficulty concentrating or making decisions [4].

Mental health is a critical global issue, especially as university students are particularly vulnerable to psychological stress due to the challenges and complexities of academic life. Among students, mental health disorders such as anxiety and depression are prevalent and significantly affect social interaction, academic performance, and overall well-being [7]. If left unaddressed, depression can lead to severe consequences such as academic failure, dropping out, or even self-harm and suicide [9]. A study conducted by Meda et al. [6], involving 1,388 participants, revealed that nearly 20% of students reported experiencing severe depressive symptoms or suicidal ideation, with financial hardship being a significant contributor to their emotional distress.

2.5. Machine Learning

Machine Learning (ML), a branch of Artificial Intelligence (AI), enables machines to learn from data and make predictions or decisions autonomously [29]. It leverages statistical algorithms to detect patterns in datasets and automatically improves its performance over time [30]. Today, ML has been widely implemented across domains such as finance, healthcare, industrial production, and psychological research.

Currently, machine learning techniques are increasingly utilized in fields such as robotics, computer vision, language assistants, medical diagnostics, marketing, industrial operations, and scientific research. Lately, ML approaches have also been used in designing new material alloys and the prediction of their properties [31][32]

In the medical domain, ML algorithms prove valuable in analyzing patient data to support clinical decision-making. ML encompasses several learning paradigms, such as classification, regression, and clustering, each suited to different types of problems and capable of yielding diverse outcomes depending on the context [32].

.

2.6. Multi-Layer Perceptron (MLP)

The Multi-Layer Perceptron (MLP), a form of artificial neural network (ANN), is often labeled a basic or 'vanilla' neural network due to its simple structure, modeled after the human brain's neural systems, which are composed of billions of linked neurons [33]. An MLP comprises three main types of layers: an input layer, several hidden layers, and an output layer. Each of these layers includes neurons interconnected through weighted links, and each neuron applies an activation function to process incoming signals [34]

The prediction process in an MLP is carried out through forward propagation, in which signals flow progressively from the input to the output layer. Neurons in the hidden layers use non-linear functions such as ReLU or sigmoid to transform inputs. Key parameters that impact how effectively the MLP model performs include the count of hidden nodes the selection of activation functions, and the training techniques like backpropagation and the gradient descent method. The mathematical expression for the output of the MLP model is presented below [35], which can be seen as equation 3.

$$y = \delta_2 \left(\sum_{i=1}^{m} \left(w_i^{(2)} \delta_1 hgjh(x) \right) + b^{(2)} \right) X = \sum_{j=1}^{n} \left(X_j W_{xj}^{(1)} \right) + b^{(1)} \tag{3}$$

In the given equation, y denotes the prediction vector produced by the MLP model, which is computed based on the input vector x_j, representing the feature set of a data sample. The weights connecting the input layer to the hidden layer are represented as $w^{(1)}$, while the weights linking the hidden layer to the output layer are denoted as $w^{(2)}$. The activation function applied in the hidden layer is symbolized by δ_1, and the activation function used in the output layer is represented by δ_2.

Bias terms are included for each layer, which $b^{(1)}$ is assigned to the hidden layer and $b^{(2)}$ is associated with the output layer. The parameters m and n indicate the number of observations (examples) and the corresponding quantity of input attributes in the dataset, respectively. This architecture enables the MLP to capture intricate non-linear patterns linking the input data to the output results, making it widely applicable for use in solving either classification or regression problems.

Through training, MLP adjusts its weights and biases to minimize prediction error, allowing it to extract meaningful patterns from data. Due to its effectiveness in modeling complex and non-linear data, MLP has been extensively used in diverse domains, including tabular data processing, image analysis, and natural language processing [35].

2.7. Dense Layer

The dense layer, also known as a fully connected layer, is a key element in neural networks, where each unit connects to all neurons in both the preceding and succeeding layers. This structure facilitates learning complex feature representations through a process of linear mapping followed by a non-linear function such as ReLU or sigmoid. Dense layers are widely employed in tasks such as pattern recognition, image categorization, and Natural Language Processing (NLP).

In medical prediction tasks, such as diagnosing diabetes, incorporating dense layers into neural network architectures significantly enhances prediction accuracy and model efficiency. When combined with validation techniques such as K-fold cross-validation and hyperparameter tuning, models can learn more effectively from data, leading to improved predictive performance [36].

To further enhance generalization, dense layers are often integrated with regularization techniques such as dropout. As demonstrated by the DenseNet architecture, dense connectivity strengthens feature propagation, reduces parameter counts, and addresses vanishing gradients in deep networks [37].

2.8 Residual Multi-Layer Perceptron (ResMLP)

Residual Multi-Layer Perceptron (ResMLP) is a deep learning architecture composed entirely of MLP layers and residual connections, without convolutional or attention mechanisms. Originally proposed by Touvron et al. (2021) [38], ResMLP offers a streamlined yet competitive alternative for image classification tasks. The model processes input data as patches using linear transformations and feedforward layers, with residual connections to enhance training stability. In this study, ResMLP is adapted for tabular data to explore its feature-extraction capabilities without requiring complex attention mechanisms, thereby offering higher computational efficiency for depression detection while mitigating the vanishing gradient problem through its residual blocks.

An advancement of this architecture is E-ResMLP+, a hybrid model combining the strengths of ResMLP and EfficientNetV2b0. Designed specifically for wheat species classification, E-ResMLP+ leverages EfficientNet's robust feature extraction and ResMLP's straightforward feedforward structure. According to the study by Dönmez et al. (2024) [39], this model achieved 98.33% classification accuracy without requiring data preprocessing or specialized hardware, demonstrating its effectiveness for high-accuracy image classification with minimal preprocessing.

2.9 Residual Neural Network (ResNet)

Residual Neural Network (ResNet) is a deep learning architecture introduced by Kaiming He et al. in 2015 to address the problem of accuracy degradation in very deep networks through skip connections. Research has extensively explored optimizing ResNets to enhance both model efficiency and accuracy, particularly for deployment on resource-constrained devices. In the context of this study, the implementation of ResNet on tabular student mental health data represents a significant methodological shift. By utilizing skip connections, the model ensures that the nuanced psychological indicators identified during preprocessing are preserved across deeper layers, preventing the loss of critical information and stabilizing the gradient flow during training a common challenge in traditional MLP models.

One such optimization approach involves reducing start-up latency by pruning channels and residual blocks. In [40], a two-stage optimization method was proposed, consisting of approaches for Global

Constraints and for reducing start-up latency. This method achieved a latency reduction of 70.40%, surpassing the previous method by 13.63% when using only channel pruning, thus allowing immediate computation on edge platforms such as desktop CPUs, FPGAs, ARM CPUs, and PULP platforms.ResNet has proven highly effective in image recognition tasks and is widely employed for applications like object detection in photos, quality assessment, and even signal classification [41].

2.10 Stochastic Gradient Descent (SGD)

Among the most popular optimizers in machine learning is Stochastic Gradient Descent (SGD), particularly effective for large-scale datasets due to its ability to optimize efficiently over multiple epochs, or full passes through the training data [42].

A recent study, by Alharbi et al. (2025) [43], applied SGD to evaluate sentiment in both Arabic and English film critiques, demonstrating its capability to handle linguistic diversity and cultural nuances. It yielded accuracy scores of 84.89% on Arabic data and 87.44% on English data, highlighting the algorithm's effectiveness and adaptability. Thanks to its flexibility in adjusting model complexity, SGD is well-suited for various types of textual data, including content with spoilers or stylistic variations.SGD is utilized as a baseline optimizer to evaluate the performance of non-adaptive learning rate methods against more complex adaptive algorithms like Adam and RMSprop. By implementing SGD within residual and dense architectures, this study aims to identify whether a consistent, manual learning rate schedule can provide better generalization for mental health classification compared to automated scaling, an analysis that is often missing in previous student depression studies.

2.11 Root Mean Squared Propagation (RMSprop)

RMSprop, short for Root Mean Squared Propagation, functions as an optimization algorithm designed to address the vanishing and exploding gradient problems commonly encountered in deep learning. It adaptively modifies each parameter's learning rate by normalizing it using the exponential moving average of the squared derivatives. This approach stabilizes and improves the efficiency of parameter updates, such as weights (W) and biases (B) [44].

For instance, this approach computes a weighted average of the parameter squares [45], as shown in equations 4-5.

$$\Delta \mathrm{Wi} = \beta \times \Delta \mathrm{Wi} + (1 - \beta) \times \mathrm{Wi}^2 \tag{4}$$

$$\Delta \mathrm{Wj} = \beta \times \Delta \mathrm{Wj} + (1 - \beta) \times \mathrm{Wj}^2 \tag{5}$$

In this case, the momentum hyperparameter β, is a value between 0 and 1. The formula changes the parameter once the average value has been updated [45], as shown in equations 6-7.

$$\mathrm{W} = \mathrm{W} - \text{learning rate} \times \frac{\mathrm{Wi}}{\sqrt{(\Delta \mathrm{Wi})}} \tag{6}$$

$$\mathrm{B} = \mathrm{B} - \text{learning rate} \times \frac{\mathrm{Wj}}{\sqrt{(\Delta \mathrm{Wj})}} \tag{7}$$

Since smaller gradient updates (ΔWi) lead to more significant weight changes, while larger updates (ΔWj) limit drastic parameter modifications, RMSProp effectively balances the step sizes during training. RMSprop was strategically selected for its ability to stabilize gradient fluctuations in the student mental health dataset. By normalizing the moving average of squared gradients, the model is expected to achieve more consistent convergence and potentially improve classification accuracy beyond the conventional optimization methods used in earlier studies.

2.12 Adaptive Moment Estimation (Adam)

Adam (Adaptive Moment Estimation) is an adaptive optimization algorithm developed to address the challenges of optimizing stochastic gradient-based objective functions [46]. It combines momentum methods with an adaptive learning rate adjustment mechanism, enhancing both the stability and the speed of convergence during training [47]. By leveraging estimates of the mean (first moment) and the variance (second moment), Adam effectively adjusts a unique learning rate per parameter based on recent and historical gradient values [46].

The algorithm combines RMSProp and momentum to reduce vertical oscillations during training, resulting in faster convergence on minibatch gradients [48]. In practice, Adam introduces two key parameters, β_1, and β_2, which control the exponential decay rates of the first and second moment estimates, respectively [46][49]. Common default values for these parameters are 0.9 for β_1 and 0.999 for β_2 [46].

The parameter update formula for Adam can be expressed as follows, as shown in equations 8-12.

1. Calculation of the first and second moments:

$$m_t = \beta_{1,t} \cdot m_t + (1 - \beta_{1,t}) \cdot g_t \quad (8)$$

$$v_t = \beta_{1,t} \cdot v_t + (1 - \beta_{1,t}) \cdot {g^2}_t \quad (9)$$

2. Bias correction for the first and second moments:

$$\widehat{m}_t = \frac{m_t}{1 - \beta_{1,t}^t} \quad (10)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_{2,t}^t} \quad (11)$$

3. Parameter update:

$$\theta_{t+1} = \theta_t - \eta_t \cdot \frac{\widehat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \cdot e_t \quad (12)$$

θ_t denotes the variable being tuned during optimization, while m_t and v_t denote the gradient's estimated mean and variance, in that order. Corrected versions of the first and second moment estimates are represented by $\widehat{m}_t$ and $\hat{v}_t$. Here, η_t is the learning rate, and e_t acts as a minor constant to ensure numerical stability in updating weights [48]. In this research, Adam is incorporated into a systematic evaluation framework to provide a robust comparison against other adaptive optimizers. Unlike previous studies that often rely on a single default optimizer, the inclusion of Adam aims to verify whether its momentum-based stability is more effective than RMSprop or SGD when applied to modern residual MLP architectures for mental health classification.

2.13 Confusion Matrix

A crucial part of building a successful machine learning (ML) model is evaluating its performance. Predictive models are commonly assessed using performance metrics derived from the confusion matrix [50]. This matrix serves as a standard evaluation tool that reflects the classification results, helping to determine the effectiveness of the machine learning model employed [51]. Performance metrics such as accuracy, precision, and recall are calculated from the confusion matrix for each tested method to obtain reliable evaluation results. Additionally, the confusion matrix is widely used to assess the performance of ML models. Figure 2 illustrates the confusion matrix diagram [52].

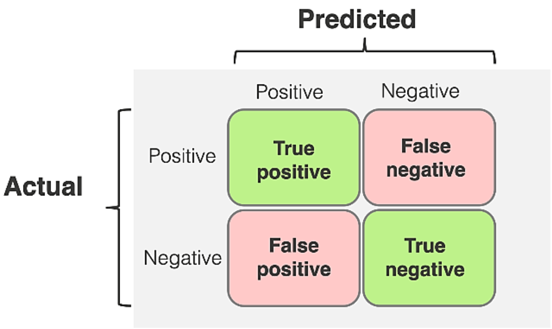

Figure 2. Confusion Matrix Diagram [42]

Accuracy, precision, recall, specificity, and F1 score are metrics derived from the confusion matrix. The formulas for these metrics are derived during the calculation process [52] [54], can be seen in equations 13-17.

1. *Accuracy*: A measure of the proportion of correct predictions out of the total data.

$$Accuracy = \frac{(TP+TN)}{(Tp+TN+FP+FN)} \quad (13)$$

2. *Precision*: Indicates how accurate the model is in making positive predictions, representing the percentage of true positive predictions among all positive predictions.

$$Precision = \frac{\mathrm{TP}}{(\mathrm{Tp+FP})} \tag{14}$$

3. *Recal*: Measures the model's ability to identify all actual positive cases in the data.

$$Recal = \frac{\mathrm{TP}}{(\mathrm{Tp+FN})} \tag{15}$$

4. *Specificity*: Measures the model's ability to correctly identify all actual negative cases in the data.

$$Specificity = \frac{\mathrm{TN}}{(\mathrm{TN+FP})} \tag{16}$$

5. $F1 - Score$: The harmonic mean of precision and recall, providing a balanced measure of the model's accuracy in terms of both precision and positive detection.

$$F1 - Score = \frac{2\mathrm{TP}}{(2\mathrm{Tp+FP+FN})} \tag{17}$$

3. RESULTS AND DISCUSSION

To identify the best model combination for depression classification, six configurations of Multi-Layer Perceptron (MLP) models were implemented, comprising three architectures (DenseNet, ResMLP, and ResNet) and three optimizers (Adam, SGD, and RMSprop), on a dataset of students exhibiting depressive symptoms. After undergoing preprocessing and data splitting, each model was trained and tested, then evaluated using four key performance metrics: accuracy, precision, recall, and F1-score. The selection of these metrics ensures a comprehensive assessment, where recall is particularly critical in a mental health context to minimize false negatives ensuring that students who are actually depressed do not go undetected.

3.1. Preprocessing Data

The initial step in this study involved data preprocessing, including feature selection using the Chi-Square method and normalization via Min-Max Scaling. Feature selection aimed to identify the most relevant attributes related to depression status, such as age, gender, financial stress, sleep patterns, and family history of mental disorders. Normalization was performed to scale all features to the range [0, 1].

The processed dataset was split using the Hold-Out method with an 80:20 ratio, with 80% allocated to training and 20% to testing. This split was done randomly while maintaining balanced class proportions. The models employed were Multi-Layer Perceptrons (MLP) with three architectures: DenseNet, ResMLP, and ResNet. Each architecture was combined with three optimizers, Adam, SGD, and RMSprop, resulting in nine model combinations ready for evaluation. The evaluation aimed to assess model performance in classifying students' depression status based on the prepared data. Table 1 is the data used.

Table 1. Data

Age	Gender	Have you ever had suicidal thoughts?	Work/Study Hours	…	Financial Stress	Family History of Mental Illness	Depression
33	Male	Yes	3.0	…	1.0	No	1
24	Female	No	4.0	…	2.0	Yes	0
31	Male	No	9.0	…	1.0	Yes	0
28	Female	Yes	4.0	…	5.0	Yes	1
25	Female	Yes	1.0	…	1.0	No	0
29	Male	No	4.0	…	1.0	No	0
30	Male	No	1.0	…	2.0	No	0

3.2. DenseNet

The evaluation results for the MLP model with the DenseNet architecture demonstrated strong performance across the various optimizers used in this classification task (see Table 2).

Table 2. DenseNet Architecture Result

Architecture	Optimizer	Evaluation Results	
DenseNet	Adam optimizer	Precision	83.36

Architecture	Optimizer	Evaluation Results	
		Recall	82.57
		F1-Score	82.89
		Accuracy	83,52
	SGD optimizer	Precision	83.24
		Recall	82.83
		F1-Score	83.01
		Accuracy	83,55
	RMSprop optimizer	Precision	81.34
		Recall	82.18
		F1-Score	81.32
		Accuracy	81,47

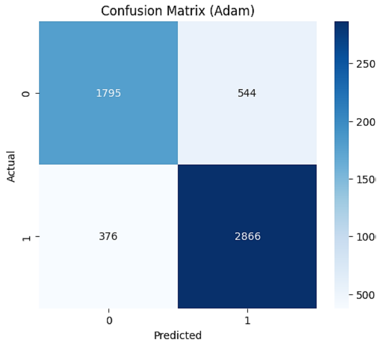

Figure 3. Confusion Matrix DenseNet-Adam

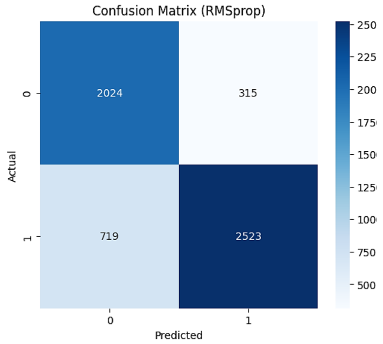

Figure 4. Confusion Matrix DenseNet-RMSprop

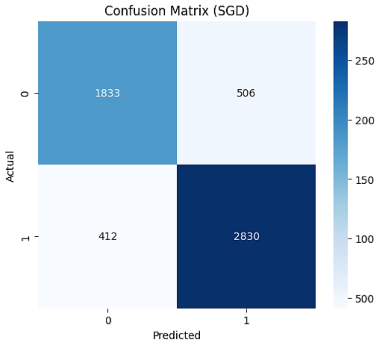

Figure 5. Confusion Matrix DenseNet-SGD

Table 2 shows that the SGD optimizer achieves the highest accuracy (83.55%) and F1-Score (83.01%) for this architecture. Figures 3,4,5 are confusion matrices, specifically, the Confusion Matrix in Figure 5 reveals that DenseNet-SGD correctly identified 2,830 depressed instances, outperforming DenseNet-RMSprop, which showed a significant drop in accuracy to 81.47%. This 2.08% performance gap suggests that, while DenseNet is flexible, standard gradient descent is more effective at navigating its loss landscape for this specific tabular dataset than adaptive RMSprop.

3.3. ResMLP

The evaluation results of the ResMLP architecture demonstrated consistent performance and slightly outperformed DenseNet across most metrics.

Table 3. ResMLP Architecture Result

Architecture	Optimizer	Evaluation Results	
		Precision	82.42
	Adam optimizer	Recall	82.12
		F1-Score	82.26
		Accuracy	82,80
		Precision	83.38
ResMLP	SGD optimizer	Recall	82.94
		F1-Score	83.13
		Accuracy	83,68
		Precision	82.27
	RMSprop optimizer	Recall	82.75
		F1-Score	82.44
		Accuracy	82,76

Table 3 shows that ResMLP achieves its best performance when combined with the SGD optimizer, attaining the highest F1-score and accuracy values of 83.68%, outperforming other combinations. The performance differences among optimizers within this architecture are relatively small, indicating the model's stability against changes in optimization methods. Nevertheless, SGD remains the most optimal choice for ResMLP in this classification task. Figures 6, 7, and 8 show that the ResMLP model with the SGD optimizer achieves the highest number of correct predictions among Adam and RMSprop. This demonstrates that the combination of ResMLP and SGD delivers the best and most consistent classification performance.

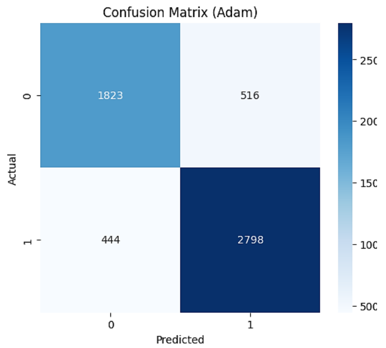

Figure 6. Confusion Matrix ResMLP-Adam

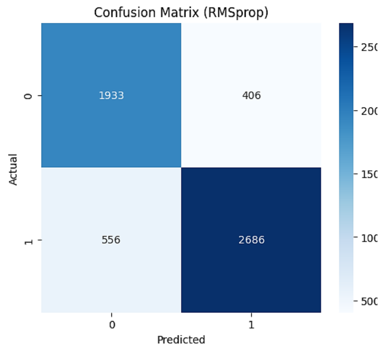

Figure 7. Confusion Matrix ResMLP-RMSprop

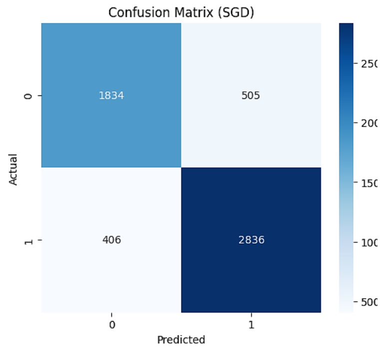

Figure 8. Confusion Matrix ResMLP-SGD

3.4. ResNet

The ResNet architecture within the MLP model also demonstrated competitive evaluation results across the three optimizers tested.

Table 4. ResNet Architecture Result

Architecture	Optimizer	Evaluation Results	
ResNet	Adam optimizer	Precision	82.84
		Recall	82.32
		F1-Score	82.54
		Accuracy	83,12
	SGD optimizer	Precision	83.39
		Recall	83.08
		F1-Score	83.22
		Accuracy	83,73
	RMSprop optimizer	Precision	83.98
		Recall	82.67
		F1-Score	83.13
		Accuracy	83,86

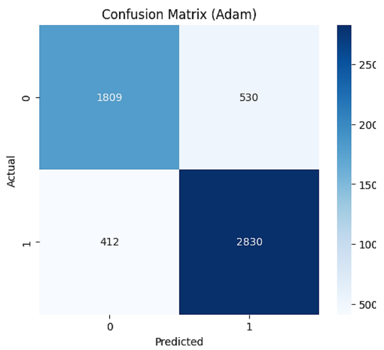

Figure 9. Confusion Matrix ResNet-Adam

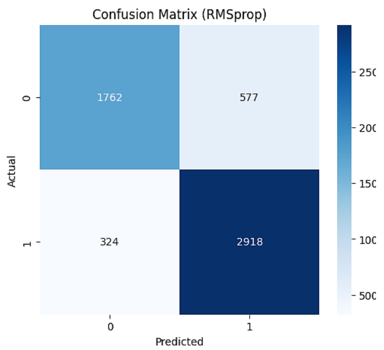

Figure 10. Confusion Matrix ResNet-RMSprop

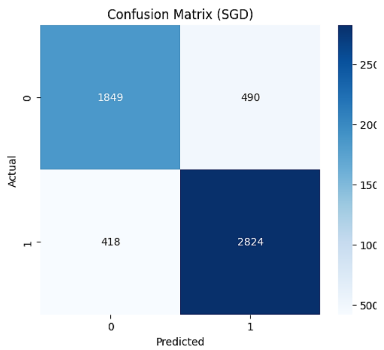

Figure 11. Confusion Matrix ResNet-SGD

Table 4 illustrates that the combination of ResNet with RMSprop yields the highest accuracy of 83.86%, although its recall score is slightly lower than that of SGD. Overall, both SGD and RMSprop outperform Adam. The consistently high performance of ResNet indicates the effectiveness of the residual architecture in enhancing classification accuracy. Furthermore, Figures 9, 10, and 11 show that ResNet paired with RMSprop achieves the highest number of correct predictions compared to Adam and SGD. This demonstrates that the combination of ResNet and RMSprop delivers the best and most accurate classification performance on the test data.

3.5. Architecture Comparison

Figure 12 presents the accuracy results of nine Multi-Layer Perceptron (MLP) model combinations formed from three network architectures: DenseNet, ResMLP, and ResNet, paired with three types of optimizers: Adam, SGD, and RMSprop. The combination of ResNet architecture with the RMSprop optimizer achieved the highest accuracy of 83.86%, followed by ResNet + SGD with 83.73%, and ResMLP + SGD with 83.68%. These results indicate that RMSprop and SGD optimizers can deliver competitive classification performance when paired with the appropriate architecture. Meanwhile, other combinations, such as DenseNet + SGD (83.55%) and DenseNet + Adam (83.52%), also demonstrated fairly good accuracy. The DenseNet with RMSprop combination achieved the lowest accuracy of 81.47%. This variation in outcomes highlights the significant impact of optimizer choice on model accuracy, underscoring the importance of selecting an appropriate optimizer to enhance classification performance for student depression detection.

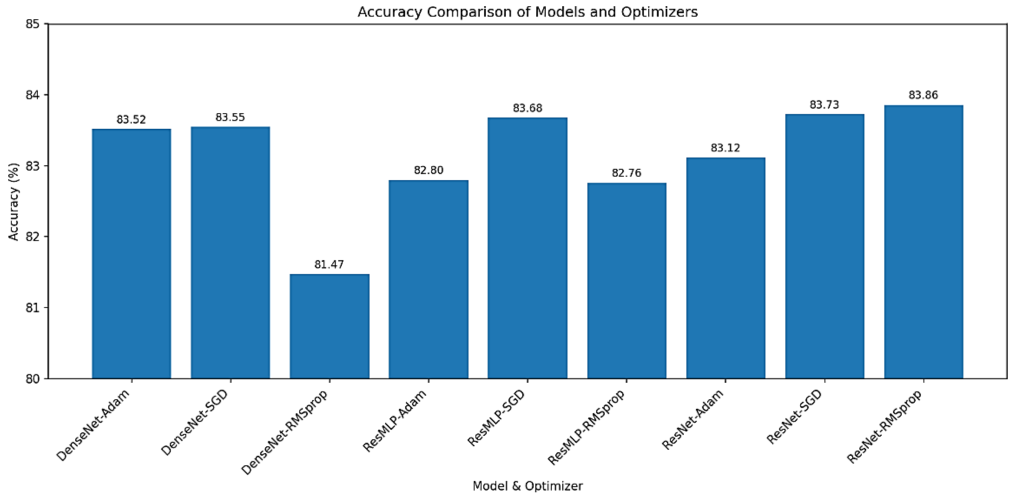

Figure 12. Comparison Result

This study demonstrates that the combination of architecture and optimizer significantly influences model performance in classifying depression among students [10][11][17]. Although all tested Multi-Layer Perceptron (MLP) models showed good performance, the ResNet architecture combined with the RMSprop optimizer yielded the highest accuracy, followed by ResNet + SGD and ResMLP + SGD combinations. Similarly, previous research by Desai (2020) [17] found that SGD is more effective than other optimizers in managing large and complex datasets, despite requiring higher computational resources. This indicates that the consistent use of the SGD optimizer delivers superior results compared to Adam and RMSprop, especially in handling data with complex characteristics such as student depression symptoms [14][17].

The superior performance of ResNet can be attributed to its "skip connections," which allow the network to learn identity mappings, effectively mitigating the vanishing gradient problem. This allows the model to refine its weights more deeply without losing information from earlier layers, a crucial advantage when dealing with complex psychological features. Regarding optimizers, the consistent success of SGD and RMSprop over Adam in this study suggests that for tabular mental health data, simpler or specifically tuned adaptive rates may be more effective. While Adam is often preferred for speed, it can sometimes converge to suboptimal local minima in non-image datasets. This study highlights that the synergy between a residual structure and an adaptive learning rate (RMSprop) or a steady momentum-based descent (SGD) is key to maximizing accuracy.

The study also emphasizes the importance of selecting the appropriate model configuration, not only in terms of network architecture but also in the optimization algorithm used [11][17]. These findings align with prior theories and studies, which state that optimal model structure and training methods are crucial for the success of machine learning-based classification systems, particularly in mental health issues [10][14]. Future research could focus on testing models on larger or multi-class datasets, as well as applying data balancing techniques to improve model sensitivity in detecting cases of depression ranging from mild to severe. For future research, these findings can be refined by incorporating data balancing techniques such as SMOTE to improve model sensitivity to severe cases, which are often underrepresented. Additionally, expanding the dataset to include longitudinal data would allow the MLP models to move beyond static classification into time-series prediction of mental health trends.

The results of this study demonstrate that the architectural configuration and optimization strategy are pivotal in achieving high accuracy for mental health classification . The peak accuracy of 83.86% achieved by the ResNet architecture with the RMSprop optimizer represents a significant improvement over the 78% accuracy reported in prior studies utilizing standard MLP models for mental health prediction . This improvement can be attributed to the structural advantages of ResNet; specifically, its use of shortcut connections allows the model to learn identity mappings, which effectively mitigates the vanishing gradient problem often encountered in deep networks . Furthermore, the success of the RMSprop optimizer in this combination suggests that its ability to adaptively adjust learning rates based on an exponential moving average of squared derivatives is highly effective for the high variability observed in student psychological datasets.

Interestingly, the SGD optimizer showed the most consistent performance across all tested architectures, yielding results above 83.5% for DenseNet, ResMLP, and ResNet. This finding aligns with Desai's (2020) theories, which posit that SGD is particularly robust for managing complex datasets despite requiring more computational iterations. In contrast, the DenseNet architecture combined with RMSprop achieved the lowest accuracy of 81.47%, indicating that, without residual connections, the adaptive nature of RMSprop may lead to suboptimal convergence on tabular datasets with specific feature distributions.

The practical implication of these findings is that institutions can leverage specific residual-based MLP configurations to build more reliable early-detection systems. A strength of this research is the systematic grid evaluation of nine distinct combinations, providing a clear benchmark for future AI-powered mental health screening tools. However, this study is limited by the small size of the dataset (1,025 entries) and its focus on binary classification. To further refine these models, future research should implement data balancing techniques such as SMOTE to better detect severe but underrepresented cases of depression. Additionally, transitioning from static datasets to longitudinal data would allow for time-series predictions, enabling proactive interventions as a student’s mental health state fluctuates over time.

4. CONCLUSION

This study successfully achieved its primary objective of evaluating the performance of various MLP architectures and optimization algorithms to identify the most effective configuration for student depression classification. Based on the evaluation results of nine Multi-Layer Perceptron (MLP) model combinations, formed from three network architectures (DenseNet, ResMLP, and ResNet) and three optimizers (Adam, SGD, and RMSprop), it can be concluded that the combination of ResNet with the RMSprop optimizer delivers the best performance in classifying depression among students. This combination achieved the highest accuracy of 83.86%. The results directly address the research question by demonstrating that residual-based architectures, such as ResNet and ResMLP, provide superior stability and accuracy when paired with RMSprop or SGD compared to standard DenseNet structures. These findings imply that the synergy between architectural design and optimization is a critical factor in developing reliable AI-powered early-detection systems in academic settings. Despite these meaningful results, a notable weakness of this study is its reliance on a dataset of 1,025 records and its focus on binary classification (Yes/No), which may not fully capture the clinical nuances of varying depression severity levels. To address these limitations, future research should focus on testing models on larger, multi-class datasets to improve the detection of depression across mild to severe levels. Furthermore, incorporating data-balancing techniques such as SMOTE and using longitudinal data are recommended to move beyond static classification toward more dynamic, sensitive time-series predictions of mental health trends. These improvements will be essential for refining machine learning-based screening tools into more proactive and accurate intervention platforms for university students.

REFERENCES

[1] M. D. Collaborators, “Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic,” vol. 398, 2021, doi: 10.1016/S0140-6736(21)02143-7.

[2] T. Wang, C. Xue, Z. Zhang, T. Cheng, and G. Yang, “Unraveling the distinction between depression and anxiety : A machine learning exploration of causal relationships,” Comput. Biol. Med., vol. 174, no. March, p. 108446, 2024, doi: 10.1016/j.compbiomed.2024.108446.

[3] M. Bader, M. Abdelwanis, M. Maalouf, and H. F. Jelinek, “Detecting depression severity using weighted random forest and oxidative stress biomarkers,” Sci. Rep., no. 0123456789, pp. 1–19, 2024, doi: 10.1038/s41598-024-67251-y.

[4] J. Chung and J. Teo, “Mental Health Prediction Using Machine Learning : Taxonomy, Applications, and Challenges,” vol. 2022, 2022.

[5] M. Andreu et al., “The Effect of Traumatic Events on the Longitudinal Course and Outcomes of Youth with Bipolar Disorder,” J. Affect. Disord., vol. 274, no. December 2019, pp. 126–135, 2020, doi: 10.1016/j.jad.2020.05.131.

[6] N. Meda, S. Pardini, P. Rigobello, and F. Visioli, “Frequency and machine learning predictors of severe depressive symptoms and suicidal ideation among university students,” no. July 2023, doi: 10.1017/S2045796023000550.

[7] N. Mumenin et al., “Heliyon Screening depression among university students utilizing GHQ-12 and machine learning,” Heliyon, vol. 10, no. 17, p. e37182, 2024, doi: 10.1016/j.heliyon.2024.e37182.

[8] N. M. Monteiro, “Addressing mental illness in Africa : Global health challenges and local

opportunities Addressing Mental Illness In Africa : Global Health Challenges And Local Opportunities A Case of Mental Illness in Africa 27-year- old Fatim lives in a rural village in the eastern region of Guinea and has 3 small," no. October 2015, doi: 10.1285/i24212113v1i2p78.
[9] N. Mumenin, M. A. Yousuf, M. O. Alassafi, M. M. Monowar, and A. Hamid, "DDNet : A Robust, and Reliable Hybrid Machine Learning Model for Effective Detection of Depression Among University Students," IEEE Access, vol. 13, no. February, pp. 49334–49353, 2025, doi: 10.1109/ACCESS.2025.3552041.
[10] F. Javed, S. O. Gilani, S. Latif, A. Waris, and M. Jamil, "Predicting Risk of Antenatal Depression and Anxiety Using Multi-Layer Perceptrons and Support Vector Machines," 2021.
[11] E. S. Mohamed, T. Ahmad, S. Ahmad, C. Bukhari, and I. Rauf, "Healthcare Analytics A hybrid mental health prediction model using Support Vector Machine, Multi-Layer Perceptron, and Random Forest algorithms," Healthc. Anal., vol. 3, no. July 2022, p. 100185, 2023, doi: 10.1016/j.health.2023.100185.
[12] D. Imans, T. Abuhmed, and M. Alharbi, "Explainable Multi-Layer Dynamic Ensemble Framework Optimized for Depression Detection and Severity Assessment," pp. 1–41, 2024.
[13] F. Kabir, A. Hossain, A. F. M. M. Rahman, and S. Z. Mishu, "Depression Detection From Social Media Textual Data Using Natural Language Processing and Machine Learning Techniques," no. December, pp. 13–15, 2023.
[14] G. Hasan, A. Masud, R. Islam, I. Sakin, and M. Rafsan, "Effective depression detection and interpretation : Integrating machine learning, deep learning, language models, and explainable AI," Array, vol. 25, no. February, p. 100375, 2025, doi: 10.1016/j.array.2025.100375.
[15] T. Hasan, A. Rahim, J. Shin, and S. Member, "Dynamics of Digital Pen-Tablet : Handwriting Analysis for Person Identification Using Machine and Deep Learning Techniques," IEEE Access, vol. 12, no. January, pp. 8154–8177, 2024, doi: 10.1109/ACCESS.2024.3352070.
[16] T. Pfister and O. Sercan, "TabNet : Attentive Interpretable Tabular Learning".
[17] C. Desai, "Comparative Analysis of Optimizers in Deep Neural Networks," International Journal of Innovative Science and Research Technology vol. 5, no. 10, pp. 959–962, 2020.
[18] N. Landro, I. Gallo, and R. La Grassa, "Mixing ADAM and SGD : a Combined Optimization Method".
[19] E. Hassan, M. Y. Shams, N. A. Hikal, and S. Elmougy, The effect of choosing optimizer algorithms to improve computer vision tasks : a comparative study. Multimedia Tools and Applications, 2023.
[20] S. Hamid, H. Madni, H. Muhammad, F. Shahzad, S. Shah, and M. Faheem, "Exploring optimizer efficiency for facial expression recognition with convolutional neural networks," no. September 2024, pp. 1–29, 2025, doi: 10.1049/tje2.70060.
[21] M. Maulidah and H. Ferdinandus, "Prediction Of Myers-Briggs Type Indicator Personality Using Long Short-Term Memory," vol. 21, no. 2, pp. 104–111, 2021, doi: 10.14203/jet.v21.104-111.
[22] A. Dhani, D. Lestari, and M. P. Ningrum, "Applying A Supervised Model for Diabetes Type 2 Risk Level Classification," Public Res. J. Eng. Data Technol. Comput. Sci., vol. 2, no. January, pp. 60–67, 2025, doi: 10.57152/predatecs.v2i2.1105.
[23] A. F. Lubis et al., "Classification of Diabetes Mellitus Sufferers Eating Patterns Using K-Nearest Neighbors, Naïve Bayes and Decision Tree," Public Res. J. Eng. Data Technol. Comput. Sci., vol. 2, no. 1, pp. 44–51, 2024, doi: 10.57152/predatecs.v2i1.1103.
[24] S. Bahassine, A. Madani, M. Al-salem, and M. Kissi, "Feature selection using an improved Chi-square for Arabic text classification," J. King Saud Univ. - Comput. Inf. Sci., vol. 32, no. 2, pp. 225–231, 2020, doi: 10.1016/j.jksuci.2018.05.010.
[25] X. Liu et al., "Adapting Feature Selection Algorithms for the Classification of Chinese Texts," Systems, vol. 11, no. 9, p. 483, 2023, doi: 10.3390/systems11090483.
[26] C. R. May et al., "Translational framework for implementation evaluation and research : a normalization process theory coding manual for qualitative research and instrument development," Implement. Sci., pp. 1–15, 2022, doi: 10.1186/s13012-022-01191-x.
[27] A. Rahmah, N. Sepriyanti, M. H. Zikri, I. Ambarani, and M. Y. bin Shahar, "Implementation of Support Vector Machine and Random Forest for Heart Failure Disease Classification," Public Res. J. Eng. Data Technol. Comput. Sci., vol. 1, no. 1, pp. 34–40, 2023, doi: 10.57152/predatecs.v1i1.816.
[28] F. Deng, J. Huang, X. Yuan, C. Cheng, and L. Zhang, "Performance and efficiency of machine learning algorithms for analyzing rectangular biomedical data Performance and efficiency of machine learning algorithms for analyzing rectangular biomedical data," Lab. Investig., no. August, 2022, doi: 10.1038/s41374-020-00525-x.
[29] B. Mahesh, "Machine Learning Algorithms - A Review," no. January 2019, 2020, doi: 10.21275/ART20203995.
[30] E. Mosqueira, R. Elena, H. Pereira, and D. Alonso, Human-in-the-loop machine learning : a state of the art, vol. 56, no. 4. Springer Netherlands, 2023. doi: 10.1007/s10462-022-10246-w.

[31] T. Ivan, K. Pavel, I. Ivan, M. Anton, and T. Allen, "Graphical abstract Machine learning assisted design of reactor steels with high long-term," Mater. Des., p. 114014, 2025, doi: 10.1016/j.matdes.2025.114014.
[32] H. El Massari, Z. Sabouri, and S. Mohammadi, "Diabetes Prediction Using Machine Learning Algorithms and Ontology," vol. 10, pp. 319–337, 2022, doi: 10.13052/jicts2245-800X.10212.
[33] R. Nosratpour and L. Tafakori, "Satellite-based extreme precipitation estimation using convolution neural networks and ant colony optimized multi-layers perceptron," Atmos. Res., vol. 320, no. November 2024, p. 108037, 2025, doi: 10.1016/j.atmosres.2025.108037.
[34] W. Kamal et al., "Development of multiple machine-learning computational techniques for optimization of heterogenous catalytic biodiesel production from waste vegetable oil," Arab. J. Chem., vol. 15, no. 6, p. 103843, 2022, doi: 10.1016/j.arabjc.2022.103843.
[35] A. Sumayli, "Development of advanced machine learning models for optimization of methyl ester biofuel production from papaya oil : Gaussian process regression (GPR), Multi-Layer Perceptron (MLP), and K-nearest neighbor (KNN) regression models," Arab. J. Chem., vol. 16, no. 7, p. 104833, 2023, doi: 10.1016/j.arabjc.2023.104833.
[36] N. Gupta, B. Kaushik, M. Khalid, and I. Rahmani, "Performance Evaluation of Deep Dense Layer Neural Network for Diabetes," 2023, doi: 10.32604/cmc.2023.038864.
[37] G. Huang, Z. Liu, G. Pleiss, L. Van Der Maaten, and K. Q. Weinberger, "Convolutional Networks with Dense Connectivity," 2019, doi: 10.1109/TPAMI.2019.2918284.
[38] H. Touvron, M. Cord, A. Joulin, P. Bojanowski, M. Caron, and A. El-nouby, "ResMLP : Feedforward networks for image classification with data-efficient training".
[39] E. Dönmez, "Hybrid convolutional neural network and Multi-Layer Perceptron vision transformer model for wheat species classification task : E - ResMLP +," Eur. Food Res. Technol., vol. 250, no. 5, pp. 1379–1388, 2024, doi: 10.1007/s00217-024-04469-0.
[40] Q. Li, H. Li, and L. Meng, "A generic deep learning architecture optimization method for edge device based on start-up latency reduction," 2024, doi: 10.1007/s11554-024-01496-8.
[41] M. Razavi, S. Mavaddati, Z. Kobti, and H. Koohi, "Rice-ResNet : Rice classification and quality detection by transferred ResNet deep model," Softw. Impacts, vol. 20, no. May, p. 100654, 2024, doi: 10.1016/j.simpa.2024.100654.
[42] T. Haber, T. J. Ashby, T. J. Ashby, and T. J. Ashby, "ScienceDirect SW-SGD : The Sliding Window Stochastic Gradient SW-SGD : The Sliding Window Stochastic Gradient Descent Algorithm SW-SGD : The Sliding Window Stochastic Gradient Descent Algorithm Descent Algorithm," Procedia Comput. Sci., vol. 108, pp. 2318–2322, 2017, doi: 10.1016/j.procs.2017.05.082.
[43] Y. Alharbi and S. S. Khan, "Classifying Multi-Lingual Reviews Sentiment Analysis in Arabic and English Languages Using the Stochastic Gradient Descent Model" Computers, Materials & Continua., vol. 8, no. 1, pp. 1275-1290, doi: 10.32604/cmc.2025.061490.
[44] R. K. P and N. C. Naveen, "Op-RMSprop (Optimized-Root Mean Square Propagation) Classification for Prediction of Polycystic Ovary Syndrome (PCOS) using Hybrid Machine Learning Technique," no. January 2022, 2024, doi: 10.14569/IJACSA.2022.0130671.
[45] C. G. Tekkali and K. Natarajan, "ScienceDirect Conference on Assessing CNN's Performance with Multiple Optimization Assessing Functions for Credit Card Fraud Detection Functions," Procedia Comput. Sci., vol. 235, pp. 2035–2042, 2024, doi: 10.1016/j.procs.2024.04.193.
[46] S. Biswas and S. Dey, "Learning Rate Tuner with Relative Adaptation (LRT-RA): Road to Sustainable Computing," AppliedMath, vol. 5, no. 1, pp. 1–39, 2025, doi: 10.3390/appliedmath5010008.
[47] V. Krutikov and E. Tovbis, "Adam Algorithm with Step Adaptation," Algorithms, vol. 18, no. 5, p. 268, 2025, doi: doi.org/10.3390/a18050268.
[48] Y. Shao et al., "An Improved BGE-Adam Optimization Algorithm Based on Entropy Weighting and Adaptive Gradient Strategy," Symmetry (Basel)., vol. 16, no. 5, pp. 1–16, 2024, doi: doi.org/10.3390/sym16050623.
[49] M. Reyad and A. M. Sarhan, "A modified Adam algorithm for deep neural network optimization," Neural Comput. Appl., vol. 35, no. 23, pp. 17095–17112, 2023, doi: 10.1007/s00521-023-08568-z.
[50] K. Moulaei, M. Shanbehzadeh, Z. M. Taghiabad, and H. K. Arpanahi, "Comparing machine learning algorithms for predicting COVID-19 mortality," BMC Med. Inform. Decis. Mak., vol. 0, pp. 1–12, 2022, doi: 10.1186/s12911-021-01742-0.
[51] P. F. Pratama, D. Rahmadani, R. S. Nahampun, D. Harmutika, A. Rahmadeyan, and M. F. Evizal, "Random Forest Optimization Using Particle Swarm Optimization for Diabetes Classification," Public Res. J. Eng. Data Technol. Comput. Sci., vol. 1, no. 1, pp. 41–46, 2023, doi: 10.57152/predatecs.v1i1.809.
[52] E. Helmud, E. Helmud, and P. Romadiana, "Classification Comparison Performance of Supervised Machine Learning Random Forest and Decision Tree Algorithms Using Confusion Matrix," vol. 13,

pp. 92–97, 2024.

[53] Z. C. Dwinnie, L. Khairani, M. A. M. Putri, J. Adhiva, and M. I. F. Tsamarah, “Application of the Supervised Learning Algorithm for Classification of Pregnancy Risk Levels,” Public Res. J. Eng. Data Technol. Comput. Sci., vol. 1, no. 1, pp. 26–33, 2023, doi: 10.57152/predatecs.v1i1.806.

[54] Mustakim, P. Khairunnisa, Sepriano and A. Sari, "Classification of Hypertension Disease Using Backpropagation Algorithm Based on Medical Record Data," 2025 International Conference on Smart Computing, IoT and Machine Learning (SIML), Surakarta, Indonesia, 2025, pp. 1-6, doi: 10.1109/SIML65326.2025.11081146.