Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, march 2025 Page: 2228−2239
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6532

Optimizing Mental Health Classification on Reddit: A Comparative Study of Adam, RMSProp, and SGD with L2 Regularization

Vander Mulya Putra*, Junta Zeniarja

Faculty of Computer Science, Informatics Engineering, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1,*]111202113343@mhs.dinus.ac.id, [2]junta@dsn.dinus.ac.id
Correspondence Author Email: 111202113343@mhs.dinus.ac.id
Submitted: **26/12/2024**; Accepted: **26/02/2025**; Published: **01/03/2025**

Abstract−The rising prevalence of mental health discussions on social media platforms has created new opportunities for understanding and supporting individuals facing psychological challenges. This study examines the automated classification of mental health content on Reddit, focusing on five clinically significant conditions (ADHD, anxiety, bipolar disorder, depression, and PTSD) and non-clinical discussions. Reddit was selected as the primary data source due to its unique subreddit structure and rich user-generated content in mental health communities, where individuals actively seek support and share experiences. Using a Multi-layer Perceptron (MLP) architecture, the study conducted a comprehensive evaluation of three optimization algorithms (Adam, RMSProp, and SGD) in conjunction with L2 regularization (λ=0.01) for mental health text classification. The study incorporated Easy Data Augmentation (EDA) techniques to enhance model robustness, implementing paraphrase-based augmentation methods that improved classification performance by 3%. Through systematic evaluation across multiple metrics, the study found that the RMSProp optimizer without L2 regularization achieved optimal performance, demonstrating 83% precision and 82% recall across all diagnostic categories. Notably, the application of L2 regularization consistently resulted in decreased model performance across all optimizers, with performance degradation ranging from 3% to 52%. These findings contribute to the development of more accurate automated mental health monitoring systems while highlighting the critical role of optimizer selection in mental health-related Natural Language Processing (NLP) tasks.

Keywords: Classification; Mental Disorder; Multi-layer Perceptron; Reddit; RMSProp

1. INTRODUCTION

Mental health issues have become increasingly prominent across various spheres of society, with manifestations ranging from common conditions like anxiety and depression to complex behavioral challenges. Recent studies have highlighted concerning trends, including rising cases of anxiety disorders, depression, and other mental health conditions that significantly impact quality of life [1], [2]. These disorders often lead to cascading effects on individuals' wellbeing, potentially exacerbating their overall health condition and social functioning [3]. The growing prevalence of mental health challenges necessitates innovative approaches to understanding, detecting, and addressing these issues through various channels, including digital platforms.

Cases of mental disorders have risen significantly during the COVID-19 pandemic period, with global impact reaching concerning levels. Studies indicate that approximately 20% of adults worldwide experienced mental health conditions during the pandemic, representing nearly 450 million individuals [4]. These cases manifested in various forms, including bipolar disorder, mood disorders, eating disorders, burnout [5], and heightened levels of stress and anxiety [6]. The pandemic context has highlighted the importance of maintaining mental wellbeing alongside physical health, including through exercise [7] and maintaining positive mindsets. However, when mental health challenges emerge, they can potentially complicate recovery efforts, necessitating additional medical and psychological interventions.

While seeking professional help remains the most effective approach for addressing mental health issues, many individuals turn to peer support or community-based mental health services [8]. The advancement of technology has enabled these support communities to exist online through various social media platforms, facilitating emotional support [9] and guidance. Among these platforms, Reddit has emerged as a particularly valuable resource, hosting 52 million daily active users and over 138,000 community topics or “subreddits” [10]. Reddit's unique structure, with its unlimited thread length and specialized subreddits, makes it particularly suitable for mental health discussions, covering topics such as ADHD, anxiety, bipolar disorder, depression, and PTSD [11].

Given the extensive mental health discourse on Reddit, Natural Language Processing (NLP) serves as a vital tool for analyzing user-generated content and identifying behavioral and emotional patterns. Models like Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) enhance this capability by capturing contextual nuances in mental health discussions [12]. However, the unregulated nature of social media introduces challenges such as misinformation and harmful content, which can distort analysis and impact vulnerable individuals [13]. Despite these challenges, NLP remains a promising approach for detecting distress signals and facilitating early intervention in mental health research.

The evolution of artificial neural networks (ANNs) over the past two decades has significantly advanced machine learning capabilities, particularly in pattern recognition, prediction, and classification tasks [14]. These networks process information through sophisticated layer structures, generating insights through iterative learning processes [15]. The Multi-layer Perceptron (MLP) architecture, characterized by multiple hidden layers, has demonstrated particular efficacy in addressing complex classification challenges, despite certain computational constraints.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, march 2025 Page: 2228−2239
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6532

Recent advancements in computational analysis of mental health have shown promising directions while revealing significant research opportunities. Jiang et al. developed deep contextualized representations for mental health detection on Reddit, demonstrating potential in disorder identification but noting challenges with emotional expression complexity [11]. Uban et al. advanced this field through emotion and cognitive-based analysis of social media data, though their approach faced limitations in cross-cultural contexts [9]. While Low et al. established comprehensive mental health datasets [16], and Li et al. improved classification accuracy through augmented retrieval models [17], crucial gaps persist in integrating advanced neural architectures with mental health-specific features and optimizing model parameters for specialized content classification.

This research aims to develop and implement an advanced classification framework for mental health-related content using MLP architecture with optimized parameters, while also comparing the effectiveness of various optimizers, including Adaptive Moment Estimation (Adam), Root Mean Square Propagation (RMSProp), and Stochastic Gradient Descent (SGD) [18]. The optimizer plays a crucial role in managing model accuracy during training by adjusting the learning rate, and its impact is further analyzed with and without L2 0.01 regularization to enhance generalization capabilities through parameter constraints and penalty mechanisms [19]. The study focuses on five specific mental health categories, namely ADHD, anxiety, bipolar disorder, depression, and PTSD, alongside a general category for broader discussions. Additionally, the methodology integrates sophisticated text processing techniques such as Term Frequency-Inverse Document Frequency (TF-IDF) vectorization, neural network hyperparameter optimization, and Easy Data Augmentation (EDA) for paraphrase generation. Through a comprehensive evaluation of optimization strategies, this research demonstrates the effectiveness of computational approaches in analyzing and classifying mental health discussions on social media platforms.

The findings of this study contribute to the growing body of research at the intersection of mental health analysis and computational linguistics, offering insights into the practical application of machine learning techniques for mental health content classification. This research holds significant implications for the development of automated systems capable of understanding and categorizing mental health-related discussions in online communities, potentially facilitating more effective digital mental health support systems.

2. RESEARCH METHODOLOGY

2.1 Research Stages

The first steps of the research approach include gathering the necessary datasets, preprocessing the data, modeling, and evaluation. Figure 1 illustrates the methodological steps employed in the study, detailing the sequential processes from data preprocessing to model evaluation. Each stage is systematically structured to ensure a clear and logical progression in analyzing the Reddit Mental Health Dataset (RMHD).

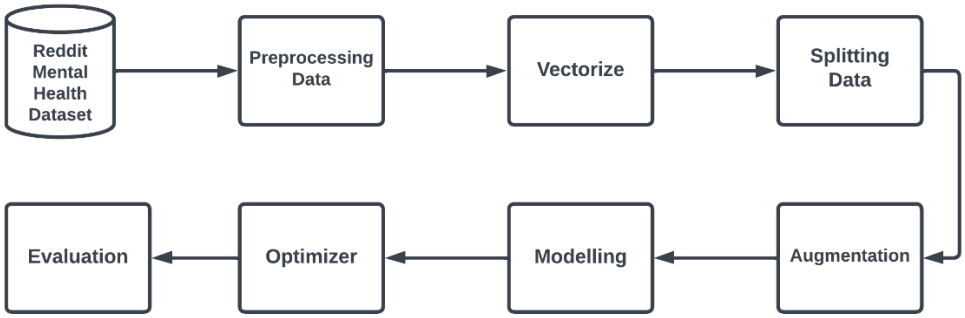


Figure 1. Research Methodology Stages

The first step of the research was gathering mental health data from the Reddit platform. The data then entered the data preprocessing stage to clean up the noise in the data. Next, it moves to the vectorization stage to convert the represented sentiment text into numerical features. Then, the data will be split into training and test data from which data augmentation will be performed. Following that, it goes into the modelling stage accompanied by optimizers to find out their performance. At long last, evaluation is carried out to compare and get the best performance.

2.2 Dataset

Low et al. [16] developed a dataset of postings from the Reddit platform, which spans 28 subreddits and includes topics pertaining to mental health groups in 15 of them. Between 2018 and 2020, data was collected to determine the impact of the pandemic outbreaks on people's mental health before and after. This study specifically covers five mental illnesses (ADHD, anxiety, bipolar disorder, depression, and PTSD) and one non-mental disorder group among the dataset's many mental health groups. The dataset for the non-mental health group was derived from the combined subjects of the subreddits “personal finance” and “fitness”, which were subsequently adjusted to “none”.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, march 2025 Page: 2228−2239
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6532

In conducting this study, we ensured adherence to ethical guidelines by anonymizing all data collected from Reddit, removing personally identifiable information, and utilizing only publicly available posts in compliance with the platform's policies. Ethical considerations in mental health research using social media data encompass issues such as privacy, consent, and the potential for misinterpretation of user-generated content. Researchers have a duty to protect participants' interests and prevent harm, emphasizing the need for vigilant ethical oversight and the development of consistent guidelines for ethical research design [20].

2.3 Vectorization with TF-IDF

In classification-based experiments, textual data must be transformed into numerical representations to be processed by machine learning models. In this study, TF-IDF vectorization is selected over alternative word embedding techniques such as Word2Vec and GloVe due to its interpretability and computational efficiency. TF-IDF effectively highlights important terms by assigning lower weights to frequently occurring words, which helps in distinguishing key textual features. In this research, a vocabulary size of up to 2,500 features is used, meaning that terms are ranked based on their frequency across the entire corpus. Additionally, an $Ngram$ parameter is applied to capture contextual relationships between words within a predefined range.

While TF-IDF provides interpretability and computational efficiency, its reliance on a bag-of-words (BoW) assumption limits its ability to capture semantic relationships and contextual dependencies. Advanced embedding techniques like Word2Vec and BERT address these limitations by generating dense vector representations that encode semantic similarity and contextual meaning. Word2Vec learns word associations through co-occurrence patterns, whereas BERT enhances contextual understanding using bidirectional encoding. Although these methods significantly improve semantic representation, they require substantial computational resources and extensive pre-training. Given the scope of this study, TF-IDF remains the preferred choice due to its simplicity, efficiency, and suitability for analyzing mental health-related discussions while balancing computational feasibility [21].

2.4 Text Augmentation with EDA

Since the distribution of classes in the data set is unbalanced, there is a need for techniques or methods that can handle data imbalances. One method that can be applied is EDA. Five augmentation approaches are used in this study: Back Translation (BT), Random Swap (RS), Random Deletion (RD), Random Insertion (RI), and Synonym Replacement (SR) [17]. The first method, SR, is used to replace words with other word equivalents that do not include stopwords. The RI technique is used to insert a word randomly in a sentence while still paying attention to the context or meaning of the sentence. The RS technique is used to swap the places of two words randomly in a sentence.

Using the RD approach, a random word with a predefined probability p is eliminated from the phrase. Finally, the BT technique is used to translate the sentence into another language which will then be translated back into the original language. The BT employed in this study is the “Helsinki-NLP/opus-mt-en-fr” model for translating from English to French and the “Helsinki-NLP/opus-mt-fr-en” model for translating back to English from French. These methods enrich the dataset by generating variations of existing minority class samples, thereby increasing their representation during model training. Example phrases generated by each EDA approach are displayed in Table 1.

Table 1. Sentence Generated with EDA

Operation	Sentence
None	I get so nervous before social events, my heart races and my palms get sweaty.
SR	I **acquire** so nervous before sociable events, my heart **move** and my palms get sweaty.
RI	I get my heart so **sweaty** nervous get before social events **sweaty**, races and **sweaty** my palms get sweaty.
RS	before **nervous** I so **get** social events, heart my and my palms races sweaty get.
RD	I nervous before events, heart and sweaty.
BT	I get so **anxious** before social events, my heart is racing and my hands are sweating.

The augmented data is not intended to fully replace the original samples; rather, it is merged with them to form a more balanced training set. This integration ensures that the model has exposure to both original and augmented instances, promoting better generalization. To facilitate effective training while maintaining a robust evaluation process, the dataset is initially split into 80% for training and 20% for testing. The augmentation occurs solely within the training data, enhancing its size and diversity without altering the testing set, which remains strictly composed of original samples. This approach allows for a reliable assessment of the model's performance by ensuring that the evaluation metrics reflect its ability to generalize to unseen data.

Table 2. Training Data Per Class

Class	Training Data
None	482
Anxiety	423
ADHD	409
Depression	403

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, march 2025 Page: 2228−2239
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6532

PTSD	402
Bipolar	401

Table 2 outlines the distribution of training data per class before applying EDA techniques to address the imbalance in the dataset. Following the augmentation process, the minority classes were increased in size to match the most populous class, indicated here as the "none" category, which consists of 482 samples. The augmentation was executed until each lower class reached parity with the highest class, ensuring an equitable representation during training. This balanced dataset enhances the model's ability to learn effectively from diverse examples, ultimately improving classification performance.

The careful implementation of these augmentation techniques not only mitigates the effects of class imbalance but also enriches the training dataset, ultimately improving the model's capability to classify mental health-related texts accurately.

2.5 Model Architecture and Hyperparameter Optimization

2.5.1 MLP Model Architecture

The Multi-layer Perceptron (MLP) is a type of feedforward artificial neural network (ANN) trained using supervised learning and the backpropagation algorithm. The network consists of three primary components as shown in Figure 2 below: an input layer, one or more hidden layers, and an output layer. The synaptic weights govern the connections between these layers. The input layer processes and maps the incoming data, while the output layer produces the final predictions. The hidden layers utilize non-linear activation functions, which are crucial for handling complex, non-linear relationships within the data [22].

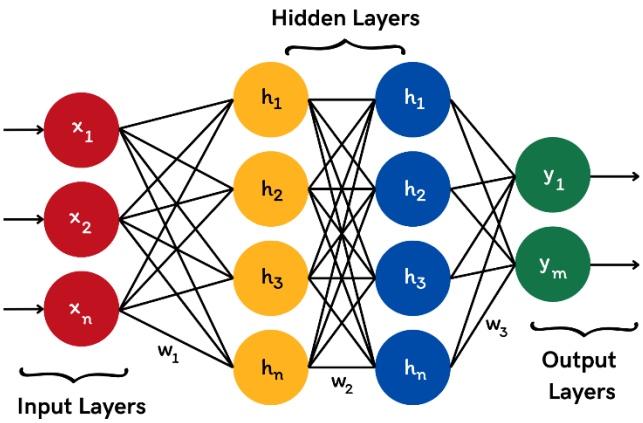


Figure 2. Multi-layer Perceptron Architecture

The training process in MLP follows two main stages: the forward propagation and the backward propagation. During forward propagation, signals are transmitted through the network using fixed synaptic weights, affecting only the activation potential and neuronal outputs. Backward propagation calculates the error signal by comparing predicted outputs with actual labels. The synaptic weights are then adjusted iteratively to minimize the error, optimizing the model's performance over multiple epochs [23].

2.5.2 Hyperparameter Selection and Tuning

To optimize the model's performance, several key hyperparameters were tuned, including the number of hidden layers, the number of neurons per layer, the learning rate, and the batch size. The optimization process was conducted using grid search, which systematically evaluates combinations of predefined hyperparameter values to identify the best configuration. The final model was selected based on the highest validation accuracy and lowest loss.

Table 3. Hyperparameter Settings and Optimal Values

Hyperparameter	Tested Values	Optimal Value
Number of Hidden Layers	{1, 2, 3, 4, 5}	2
Neurons per Layer	{32, 64, 128, 256}	256
Learning Rate	{0.00001, 0.0001, 0.001, 0.01}	0.001
Batch Size	{16, 32, 64}	32
Optimizer	{Adam, RMSProp, SGD}	RMSProp
Number of Epochs	{5, 10, 15, 20}	7

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, march 2025 Page: 2228−2239
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6532

Based on experimental results shown in Table 3 above, the optimal configuration consists of 256 neurons per hidden layer, which provides a balance between model complexity and computational efficiency. Additionally, the RMSProp optimizer was selected due to its adaptive learning rate mechanism, which enhances convergence stability. A batch size of 32 provided a balance between training speed and model stability. The model was trained for 7 epochs, as this number achieved a satisfactory trade-off between performance and overfitting prevention. Additionally, a learning rate of 0.001 was found to be optimal, preventing issues of slow convergence or divergence. These hyperparameter selections were determined through systematic experimentation to ensure optimal classification performance.

2.6 Optimizer

The study utilized three optimizers to enhance model performance, this section briefly highlights the distinguishing features of each optimizer:

a. Adaptive Moment Estimation (Adam)
   Adam's optimizer adjusts the learning rate for each parameter by analyzing the first and second moment estimates of the gradient [24], achieving high computational efficiency with minimal configuration. Its robustness in handling non-stationary objectives ensures stable convergence, making it highly suitable for tasks like natural language processing. This adaptability to varying learning rates aligns with the needs of classifying mental health-related text, where gradient distributions can differ significantly across samples, solidifying Adam as a key candidate for experimentation in this study.
b. Root Mean Square Propagation (RMSProp)
   RMSProp is a gradient-based optimization method that normalizes the gradient using a moving average of squared gradients, effectively balancing step sizes to prevent gradient explosion or vanishing. By employing an adaptive learning rate, it stabilizes the training process, making it particularly suitable for tasks with imbalanced data distributions, such as mental health sentiment classification. Its ability to balance learning dynamics across parameters allowed the model to discern subtle distinctions between classes, like "bipolar" and "depression," establishing RMSProp as a valuable comparative baseline in this study.
c. Stochastic Gradient Descent (SGD)
   SGD is an optimization method that updates model parameters by calculating gradients using a randomly selected data point or mini-batch from the dataset in each iteration. This approach reduces computational complexity compared to processing the entire dataset simultaneously and introduces randomness, which gives SGD its "stochastic" nature [25]. In this study, SGD served as a benchmark to assess the performance of adaptive optimizers like Adam and RMSProp. Its simplicity, computational efficiency, and effectiveness on large datasets made it a relevant choice for comparison, offering valuable insights into the trade-offs between adaptive methods and traditional optimization strategies in text classification tasks.

2.7 L2 Regularization

Ridge regularization, also known as L2 regularization, is the process of adding a penalty equal to the square of the feature coefficient, which decreases it toward zero but does not completely eliminate it. In this way, any feature can continue to be added to the model. This method works well for overcoming multicollinearity and yields a model that is less sparse but more stable than L1 regularization. In this experiment, a regularization value of 0.01 is applied.

2.8 Evaluation Metrics

This study uses accuracy, precision, recall, and F1-score to evaluate classification performance, especially for imbalanced datasets. Accuracy reflects overall correctness but can be misleading in skewed distributions. Precision measures the relevance of positive predictions, while recall focuses on capturing all actual positive instances. The F1 score balances precision and recall, making it particularly useful for ensuring effective classification in these scenarios. These metrics are calculated from the confusion matrix, providing a thorough assessment of the model's performance in classifying mental health-related texts.

3. RESULT AND DISCUSSION

3.1 Data Collection and Preprocessing

The study utilized the Reddit Mental Health Dataset (RMHD) from Zenodo's open-access platform, comprising 109,181 records across six classes. A balanced subset of approximately 500 records per class was extracted, yielding 3,150 total records. The preprocessing pipeline included comprehensive text cleaning operations: case normalization, punctuation removal, symbol elimination, numerical data filtering, URL removal, and stopwords exclusion using WordNetLemmatizer.

Out of the total number of subreddits, all of them were scraped down to roughly 500 records ($n \approx 500$). Figure 3 depicts the distribution of classes in the RMHD dataset, indicating the imbalance among different mental health categories. Understanding this distribution is crucial for selecting appropriate data balancing techniques to enhance model performance.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, march 2025 Page: 2228−2239
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6532

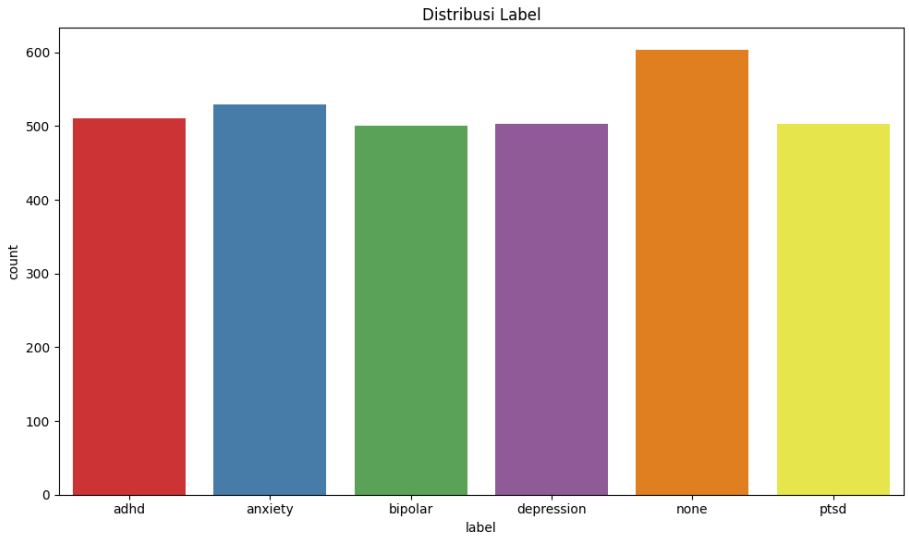


Figure 3. Subreddit Distribution by Category

In the data preprocessing phase, the dataset goes through a cleaning phase which involves converting text to lower case, removing symbols that often do not provide information (e.g., “. ”, “,”, “””, “@”, “#”, “&”, etc.), performing construction mapping expansion which means changing English abbreviations that are often used in conversation (e.g., “k” → “okay”, “u” → “you”, “y” → “why”, “ive” → “i have”, etc.), separating two words that are sometimes connected within one word (e.g., iampossible → i am possible), converting emojis into text representations (e.g., “:D” → ‘joyful’), removing numeric numbers and link URLs, lemmatization and tokenization. All of these cleaning processes are intended to improve accuracy when entering the modelling stage later. In addition, stopwords are also applied in this data preprocessing stage to reduce the number of words in the sentence by using the stopwords “English” for English text classification. Table 4 provides an example of a sample of postings from each class or subreddit.

Table 4. Reddit Mental Health Dataset Example with Modification

Subreddit	Post
ADHD	Does anyone else feel like they wear out their friendships after a few years by being impulsive… (shortened)
Anxiety	Does anyone else feel anxious at the thought of it being 4 months into the year already? … (shortened)
Bipolar	I just got diagnosed, and I need advice. Hey, all, I just got diagnosed with Bipolar... (shortened)
Depression	I need someone to talk to, but I am not sure if this is the right place… (shortened)
PTSD	Best jobs for PTSD or getting back out in the workforce? … (shortened)
None	How to make LISS cardio less boring? … (shortened)

Other NLP studies often use the libraries utilized in this study's cleanup phase. This study makes use of wordnet, punkt, and stopwords because it makes use of the Natural Language Toolkit (NLTK), which has a stronghold in the NLP field. Furthermore, it will be dropped from the dataset to handle redundancy issues and empty or NULL values in the data.

3.2 Feature Engineering and Data Augmentation

The preprocessed data underwent TF-IDF vectorization with 2,500 maximum features, incorporating unigram and bigram feature extraction. The dataset was split using an 80:20 ratio for training and testing. To address class imbalance, the study implemented EDA techniques to match the majority class size (n = 482). The augmentation process combined five strategies with each quantity detailed in Table 5 as follows:

Table 5. Quantity of Each EDA Method

Method	Quantity
Synonym Replacement	n = 2
Random Insertion	n = 1
Random Swap	n = 1
Random Deletion	p = 0.3
Back Translation	p = 0.5

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, march 2025 Page: 2228−2239
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6532

Those techniques significantly increased minority class representation, enhancing the model's ability to generalize across classes. For example, the bipolar class, initially comprising only 2.5% of the dataset, was augmented to achieve parity with majority classes. Quantitative analysis shows a substantial improvement in recall for minority classes post-augmentation. In this regard, the bipolar class's recall improved from 0.58 to 0.73. Similarly, precision metrics across minority classes exhibited noticeable gains. These findings underscore the role of augmentation in addressing imbalanced datasets and enhancing model performance for underrepresented categories.

3.3 Model Architecture

The MLP architecture employed in this study was constructed using the Sequential model from the Keras library. The model comprises three primary layers. The first layer is a Dense layer, which receives input with dimensions matching the number of TF-IDF vectorized features. This input is processed through a Rectified Linear Unit (ReLU) activation function, enhancing the model's ability to capture nonlinear patterns in the data. The second layer is a Dropout layer, designed to mitigate overfitting by randomly disabling a fraction of neurons during training. A dropout rate of 0.3 was utilized, striking a balance between regularization and model performance. Lastly, the output layer employs a SoftMax activation function with six neurons, generating probability distributions across the six target classes.

The model was compiled using the sparse categorical_crossentropy loss function, which is particularly suited for multiclass classification tasks. Accuracy was used as the primary evaluation metric to measure the model's effectiveness. Figure 4 illustrates the architecture of the developed model.

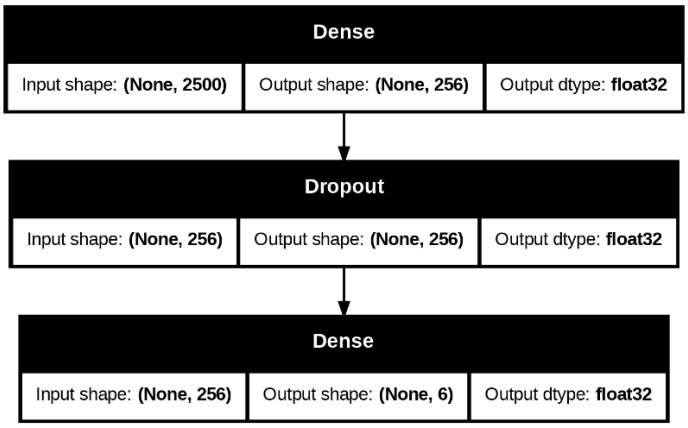


Figure 4. MLP Build Model

Prior to modelling, the target labels were encoded numerically to facilitate processing by the model. The mapping was as follows: ADHD → 0, anxiety → 1, bipolar → 2, depression → 3, PTSD → 4, and none → 5. This transformation ensured consistency in the representation of target classes during training and evaluation.

3.4 Comparative Analysis of Optimizers

3.4.1 Performance Overview

The study evaluated three optimizers, which are Adam, RMSProp, and SGD, with and without L2 regularization (λ = 0.01), which revealed distinct performance patterns as shown in Table 6 below:

Table 6. Comprehensive Performance Metrics Across Optimizers

Optimizer	Regularization	Precision	Recall	F1-score	Accuracy
Adam	None	81%	80%	82%	81%
	L2	76%	75%	76%	75%
RMSProp	None	**83%**	**82%**	**83%**	**82%**
	L2	79%	78%	78%	78%
SGD	None	81%	79%	80%	79%
	L2	55%	27%	29%	27%

RMSProp emerged as the standout performer, achieving an impressive 83% precision and 82% recall without L2 regularization. This success can be attributed to its adaptive learning rate mechanism, which effectively navigates the sparse landscape of TF-IDF features characteristic of mental health text classification [18]. Adam's reliance on momentum-based updates offered competitive performance but fell slightly short of RMSProp due to its less precise adaptability. Conversely, SGD's static learning rate struggled to match the performance of its adaptive counterparts. In addition, the systematic performance analysis revealed consistent degradation with L2 regularization across optimizers, whereby SGD experiencing the most dramatic decline with a 26% precision reduction and 52% recall decrease.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, march 2025 Page: 2228−2239
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6532

3.4.2 Impact of L2 Regularization

The investigation of L2 regularization's impact revealed a nuanced performance degradation across different optimizers. RMSProp and Adam exhibited modest performance reductions, with 4-5% decreases in both precision and recall, suggesting a relatively mild constraint on model learning. In contrast, the SGD optimizer experienced a dramatically more severe performance decline, with precision dropping by 26% and recall plummeting by an alarming 52%.

This differential response to regularization provides critical insights into the underlying mechanisms of neural network optimization. The weight penalty introduced by L2 regularization appears to disproportionately affect the model's capacity to capture subtle semantic distinctions within mental health text classification. The inherent sparsity of TF-IDF features compounds this challenge, creating a particularly challenging environment for model generalization.

The SGD optimizer's extreme sensitivity to regularization constraints is especially noteworthy. This pronounced vulnerability suggests that SGD struggles to maintain meaningful feature representations when constrained by L2 regularization, potentially due to its fundamental optimization approach of randomly sampling gradients. In comparison, more adaptive optimizers like RMSProp demonstrate greater resilience, maintaining a more stable performance profile when subjected to regularization parameters.

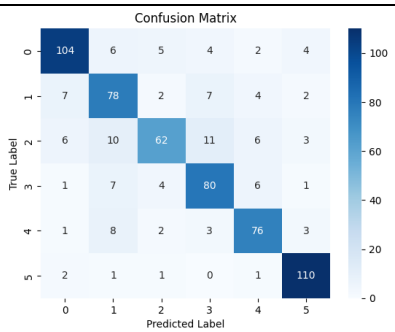
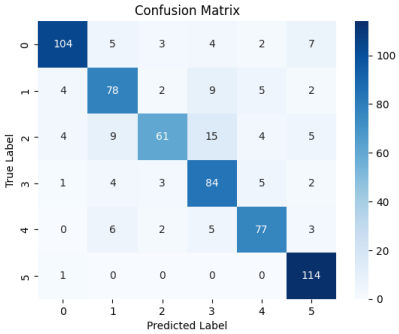
These findings underscore the critical importance of careful optimizer selection and regularization strategy in machine learning models, particularly in complex domains like mental health text classification. The results highlight that a one-size-fits-all approach to regularization can significantly compromise model performance, emphasizing the need for nuanced, context-specific optimization techniques.

3.4.3 Analysis of Confusion Matrix

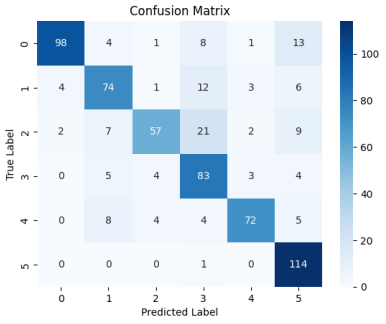
The confusion matrix analysis revealed key challenges in classifying mental health conditions, uncovering nuanced misclassification patterns tied to the linguistic complexity of mental health descriptions. A notable finding was the 15% misclassification rate between bipolar disorder and depression, highlighting semantic overlaps likely caused by shared emotional and cognitive descriptors. Similarly, the 12% confusion rate between anxiety and PTSD underscores the difficulty of distinguishing these conditions due to shared anxiety-related terminology, reflecting the intricate linguistic commonalities in psychological expression.

In contrast, the model excelled in identifying ADHD, achieving 89% accuracy in distinguishing it from non-mental health categories. This suggests that ADHD-related language contains more structured and recognizable patterns, providing a clear demarcation from other classes. These findings not only expose current limitations in mental health text classification but also emphasize the need for incorporating advanced embeddings, such as BERT or GPT, to capture the contextual nuances of mental health diagnosis. These embeddings could help differentiate subtle linguistic variations that traditional TF-IDF features fail to capture.

Table 7. Comparison of Confusion Matrix Results Across Optimizers

Optimizer	Confusion Matrix
Adam	
RMSProp	

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, march 2025 Page: 2228−2239
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6532

SGD	

As an overview, Table 7 presents a comparison of confusion matrix results across three optimizers highlighting variations in classification performance for mental health conditions. Adam and RMSProp demonstrate relatively balanced misclassification patterns, with notable challenges in differentiating bipolar disorder from depression and anxiety from PTSD, as reflected in their higher confusion rates for these classes. On the other hand, SGD struggles significantly, with higher misclassification rates across most categories, underscoring its sensitivity to optimization constraints. This table reinforces the earlier discussion on the nuanced linguistic overlaps within mental health text and the critical role of optimizer selection in mitigating these classification challenges.

3.5 Model Performance Evaluation

Evaluation metrics play a critical role in assessing the performance of classification models, especially in domains involving imbalanced datasets, such as mental health diagnoses. In such scenarios, metrics like recall and precision are often prioritized over overall accuracy due to the critical consequences of misclassifications. Recall ensures that the model effectively identifies positive cases, minimizing false negatives, which is vital in applications like mental health and disease diagnosis where undetected conditions could lead to severe outcomes [26]. Precision, on the other hand, helps in reducing false positives, ensuring that only individuals truly needing intervention are flagged.

This balance is particularly emphasized in medical applications, as demonstrated by Setiadi et al. in their study on diabetes detection models, where precision is crucial to avoid unnecessary treatments while maintaining high recall to identify all at-risk patients [27]. Similarly, in mental health contexts, prioritizing these metrics allows the model to both identify patients requiring help and avoid misdiagnosing healthy individuals. For instance, if the model fails to detect a depressed individual, the lack of intervention could lead to major psychological risks. Conversely, falsely labeling a healthy individual as depressed may lead to unwarranted treatment or social stigma.

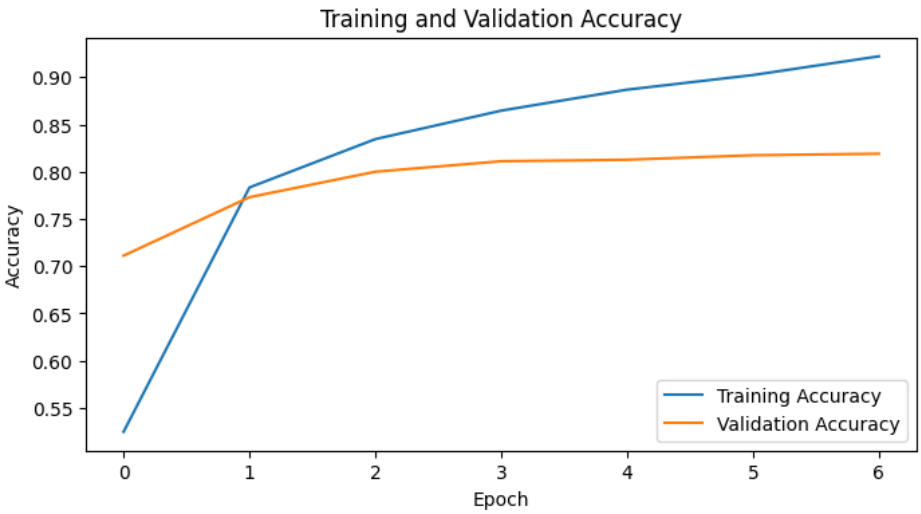


Figure 5. Model Accuracy Graph

Figures 5 and 6 illustrate the learning dynamics of the implemented model, showcasing its performance in terms of accuracy and loss. The accuracy graph on Figure 5 above reveals consistent improvements, with training accuracy rising from 53% in the first epoch to 92% by the sixth epoch. Validation accuracy similarly increases from 71% to stabilize around 82% during the fifth and sixth epochs. This stability in validation accuracy indicates the model's ability to generalize effectively, avoiding significant overfitting despite the complexity of the dataset.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, march 2025 Page: 2228−2239
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6532

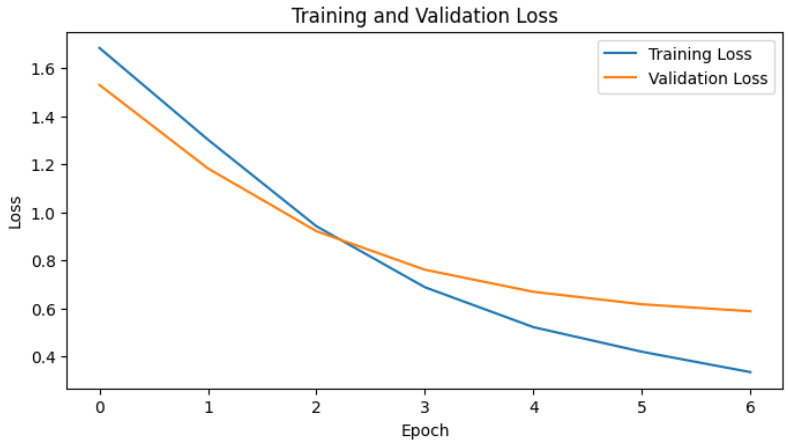


Figure 6. Model Loss Graph

Furthermore, the loss graph on Figure 6 above demonstrates a steady decline, with training loss decreasing from 1.67 to 0.38 and validation loss reducing from 1.55 to 0.62 by the sixth epoch. The relatively small gap between training and validation losses underscores the model's balance between optimization and generalization. These patterns collectively suggest that the model has successfully converged, achieving a robust performance while maintaining stability across both training and validation datasets.

By prioritizing recall and precision alongside accuracy and loss, the model achieves an optimal trade-off between minimizing misclassification risks and ensuring reliable performance for mental health diagnosis tasks.

3.6 Final Result

The model with the best performance, developed using the RMSProp optimizer without L2 regularization, was tested to classify mental disorder sentiment texts using the same dataset. This evaluation aimed to measure the model's accuracy and its ability to generalize across diverse data instances effectively.

	text	label	label_pred
103	treatment change officially diagnosed depressi...	3	3
424	sick living live reason know mum heart broken ...	3	3
125	know manic well know manic going super stable ...	2	2
320	deal pay attention conversation affect sociall...	0	0
4	sent friend believe ptsd hey thinking conversa...	4	4
507	hypomania thinking normal mood real problem ge...	2	2
528	anxiety attack talking color start question di...	4	1
266	bet others arent constantly playing immanent d...	1	0
491	feel worried feel alone scared people talk wai...	4	1
438	withdraw hello experience pax il withdrawal long	1	2
8	cost fun stayed late last night celebrating bi...	4	3
3	take med certain amount calorie long wait eati...	2	0

Figure 7. Sample Final Classification Model Result

As illustrated in Figure 7, the model demonstrates a strong capability to correctly classify mental disorder sentiment text. However, anomalies persist, particularly in cases where semantic ambiguity and contextual nuances complicate classification. For instance, bipolar disorder sentiments are sometimes misclassified as depression due to overlapping representations between the two classes. Additionally, preprocessing steps, such as the removal of punctuation marks, can inadvertently alter the semantic meaning of certain sentences, further challenging the model's performance.

Beyond quantitative performance, this study highlights critical challenges in mental health text classification, including the inherent complexity of natural language processing, the risk of information loss during preprocessing, and the influence of contextual variations. Addressing these limitations requires advanced techniques, such as integrating contextual embeddings, such as BERT or RoBERTa, employing enhanced preprocessing strategies, and exploring hierarchical classification approaches tailored to the domain of mental health analytics.

Despite these challenges, the optimized model's robust performance underscores the potential of machine learning in advancing mental health research and interventions. By providing a reliable framework for automated text classification, this research bridges the intersection of natural language processing and mental health analytics. It

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, march 2025 Page: 2228−2239
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6532

offers promising opportunities for early detection, personalized support mechanisms, and a deeper understanding of mental health discourse in online communities.

4. CONCLUSION

This study demonstrates the significant impact of optimizer selection and parameter tuning on mental health sentiment classification performance, with RMSProp outperforming other optimizers due to its adaptive learning rate and effective gradient normalization capabilities, particularly in handling sparse TF-IDF features. The implemented Multi-layer Perceptron model achieved superior results without L2 regularization, reaching 82% accuracy, 83% precision, and 82% recall across six mental health categories. The integration of Easy Data Augmentation (EDA) techniques, particularly Synonym Replacement and Back Translation, enhanced the model's generalization capacity and improved classification performance for minority classes. However, our analysis revealed challenges in preserving semantic context during text preprocessing, suggesting future research should focus on developing more sophisticated tokenization and stopwords removal techniques that better preserve domain-specific mental health terminology. Additionally, future work could explore advanced deep learning architectures, implement multilabel classification approaches, and investigate the integration of contextual embeddings to better capture the nuanced language patterns in mental health discussions on social media platforms.

REFERENCES

[1] X. Man, J. Liu, and Z. Xue, "Effects of Bullying Forms on Adolescent Mental Health and Protective Factors: A Global Cross-Regional Research Based on 65 Countries," *Int. J. Environ. Res. Public Health*, vol. 19, no. 4, 2022, doi: 10.3390/ijerph19042374.

[2] N. C. Momen *et al.*, "Association between Mental Disorders and Subsequent Medical Conditions," *N. Engl. J. Med.*, vol. 382, no. 18, pp. 1721–1731, 2020, doi: 10.1056/nejmoa1915784.

[3] C. Arango *et al.*, "Risk and protective factors for mental disorders beyond genetics: an evidence-based atlas," *World Psychiatry*, vol. 20, no. 3, pp. 417–436, 2021, doi: 10.1002/wps.20894.

[4] Aschbrenner, J. A. Naslund, A. Bondre, J. Torous, and K. A., "Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice," *J. Technol. Behav. Sci.*, vol. 5, no. 3, pp. 245–257, 2020, [Online]. Available: https://doi.org/10.1007/s41347-020-00134-x.

[5] V. J. Clemente-Suárez *et al.*, "The impact of the covid-19 pandemic on mental disorders. A critical review," *Int. J. Environ. Res. Public Health*, vol. 18, no. 19, 2021, doi: 10.3390/ijerph181910041.

[6] J. A. D. B. Campos, B. G. Martins, L. A. Campos, F. de Fátima Valadão-Dias, and J. Marôco, "Symptoms related to mental disorder in healthcare workers during the COVID-19 pandemic in Brazil," *Int. Arch. Occup. Environ. Health*, vol. 94, no. 5, pp. 1023–1032, 2021, doi: 10.1007/s00420-021-01656-4.

[7] F. B. Schuch and D. Vancampfort, "Physical activity, exercise, and mental disorders: it is time to move on," *Trends Psychiatry Psychother.*, vol. 43, no. 3, pp. 177–184, 2021, doi: 10.47626/2237-6089-2021-0237.

[8] S. H. Aarestad *et al.*, "Clinical Characteristics of Patients Seeking Treatment for Common Mental Disorders Presenting With Workplace Bullying Experiences," *Front. Psychol.*, vol. 11, no. November, pp. 1–12, 2020, doi: 10.3389/fpsyg.2020.583324.

[9] A. S. Uban, B. Chulvi, and P. Rosso, "An emotion and cognitive based analysis of mental health disorders from social media data," *Futur. Gener. Comput. Syst.*, vol. 124, pp. 480–494, 2021, doi: 10.1016/j.future.2021.05.032.

[10] N. Proferes, N. Jones, S. Gilbert, C. Fiesler, and M. Zimmer, "Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics," *Soc. Media Soc.*, vol. 7, no. 2, 2021, doi: 10.1177/20563051211019004.

[11] Z. Jiang, S. I. Levitan, J. Zomick, and J. Hirschberg, "Detection of mental health conditions from Reddit via deep contextualized representations," *EMNLP 2020 - 11th Int. Work. Heal. Text Min. Inf. Anal. LOUHI 2020, Proc. Work.*, pp. 147–156, 2020, doi: 10.18653/v1/2020.louhi-1.16.

[12] T. S. Adekunle, O. O. Alabi, M. O. Lawrence, G. N. Ebong, G. O. Ajiboye, and T. A. Bamisaye, "The Use of AI to Analyze Social Media Attacks for Predictive Analytics," *J. Comput. Theor. Appl.*, vol. 1, no. 4, pp. 386–395, 2024, doi: 10.62411/jcta.10120.

[13] A. Iorliam and J. A. Ingio, "A Comparative Analysis of Generative Artificial Intelligence Tools for Natural Language Processing," *J. Comput. Theor. Appl.*, vol. 1, no. 3, pp. 311–325, 2024, doi: 10.62411/jcta.9447.

[14] B. Turkoglu and E. Kaya, "Training multi-layer perceptron with artificial algae algorithm," *Eng. Sci. Technol. an Int. J.*, vol. 23, no. 6, pp. 1342–1350, 2020, doi: 10.1016/j.jestch.2020.07.001.

[15] A. C. Cinar, "Training Feed-Forward Multi-Layer Perceptron Artificial Neural Networks with a Tree-Seed Algorithm," *Arab. J. Sci. Eng.*, vol. 45, no. 12, pp. 10915–10938, 2020, doi: 10.1007/s13369-020-04872-1.

[16] D. M. Low, L. Rumker, T. Talker, J. Torous, G. Cecchi, and S. S. Ghosh, "Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study," *J. Med. Internet Res.*, vol. 22, no. 10, p. e22635, 2020, doi: 10.17605/OSF.IO/7PEYQ.

[17] J. Li *et al.*, "KRA: K-Nearest Neighbor Retrieval Augmented Model for Text Classification," *Electron.*, vol. 13, no. 16, pp. 1–16, 2024, doi: 10.3390/electronics13163237.

[18] J. I. T. Krisna, A. Luthfiarta, L. D. Cahya, S. Winarno, and A. Nugraha, "Comparing Optimizer Strategies For Enhancing Emotion Classification In IndoBERT Models," *Adv. Sustain. Sci. Eng. Technol.*, vol. 6, no. 2, pp. 1–8, 2024, doi: 10.26877/asset.v6i2.18228.

[19] X. Ni, L. Fang, and H. Huttunen, "Adaptive L2 regularization in person Re-identification," *Proc. - Int. Conf. Pattern Recognit.*, pp. 9601–9607, 2020, doi: 10.1109/ICPR48806.2021.9412481.

[20] Y. Zhang *et al.*, "Reporting of Ethical Considerations in Qualitative Research Utilizing Social Media Data on Public Health Care: Scoping Review," *J. Med. Internet Res.*, vol. 26, no. 1, 2024, doi: 10.2196/51496.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, march 2025 Page: 2228−2239
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6532

[21] R. Patil, S. Boit, V. Gudivada, and J. Nandigam, "A Survey of Text Representation and Embedding Techniques in NLP," *IEEE Access*, vol. 11, no. March, pp. 36120–36146, 2023, doi: 10.1109/ACCESS.2023.3266377.
[22] A. Al Bataineh, D. Kaur, and S. M. J. Jalali, "Multi-Layer Perceptron Training Optimization Using Nature Inspired Computing," *IEEE Access*, vol. 10, pp. 36963–36977, 2022, doi: 10.1109/ACCESS.2022.3164669.
[23] I. Lorencin, N. Anđelić, J. Španjol, and Z. Car, "Using multi-layer perceptron with Laplacian edge detector for bladder cancer diagnosis," *Artif. Intell. Med.*, vol. 102, 2020, doi: 10.1016/j.artmed.2019.101746.
[24] J. Wu, K. Hu, Y. Cheng, H. Zhu, X. Shao, and Y. Wang, "Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network," *ISA Trans.*, vol. 97, no. xxxx, pp. 241–250, 2020, doi: 10.1016/j.isatra.2019.07.004.
[25] 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," *Multimed. Tools Appl.*, vol. 82, no. 11, pp. 16591–16633, 2023, doi: 10.1007/s11042-022-13820-0.
[26] M. Ahmad, N. Wahid, R. A. Hamid, S. Sadiq, A. Mehmood, and G. S. Choi, "Decision Level Fusion Using Hybrid Classifier for Mental Disease Classification," *Comput. Mater. Contin.*, vol. 72, no. 3, pp. 5041–5058, 2022, doi: 10.32604/cmc.2022.026077.
[27] D. R. I. M. Setiadi, K. Nugroho, A. R. Muslikh, S. W. Iriananda, and A. A. Ojugo, "Integrating SMOTE-Tomek and Fusion Learning with XGBoost Meta-Learner for Robust Diabetes Recognition," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 23–38, 2024, doi: 10.62411/faith.2024-11.