



Optimization of a New Adaptive Stacking Ensemble Model Integrated with IoT for Stress Level Detection Based on Physiological Signals

Muhardi¹, Mohd Rinaldi Amarth², Rika Melyanti³, Yuda Irawan⁴

^{1,4}Department of Computer Science, Faculty of Computer Science, Universitas Hang Tuah Pekanbaru, Pekanbaru, Indonesia

²Department of Medical Records and Health Informatics, Faculty of Health, Universitas Hang Tuah Pekanbaru, Pekanbaru, Indonesia

³Department of Information System, Faculty of Computer Science, Universitas Hang Tuah Pekanbaru, Pekanbaru, Indonesia

¹muhardi.yudie@gmail.com, ²amartharc@gmail.com, ³camelya2105@gmail.com, ⁴yudairawan89@gmail.com

Abstract

Mental health issues among college students are receiving increasing attention, particularly because of academic and social pressures and the impact of technology use. This study aims to develop a real-time stress level prediction model using a New Adaptive Stacking Ensemble approach based on physiological data and IoT devices. The data included heart rate, SpO₂, body temperature, and systolic and diastolic blood pressure. Five machine learning algorithms are used as base models: SVM, C4.5, Decision Tree, KNN, and Random Forest. The MLP serves as the meta-model, which is then optimized using Optuna. The model training process begins with pre-processing, feature standardization using StandardScaler, and data balancing using SMOTE. The results showed that the stacking model with the MLP meta-model achieved an accuracy of 90.00% under the individual Random Forest and KNN models, and increased to 97.00% after hyperparameter optimization. This model was then integrated with IoT devices using MAX30102, MLX90614, and digital tensiometer sensors, as well as a Streamlit interface to display real-time stress classification results. The system built not only excels in accuracy but can also be implemented to directly detect stress levels, thereby potentially supporting early intervention and mental health promotion in campus environments.

Keywords: IoT; mental health; optuna; physiological data; stacking ensemble

How to Cite: Muhardi, Mohd Rinaldi Amarth, Rika Melyanti, and Yuda Irawan, "Optimization of a New Adaptive Stacking Ensemble Model Integrated with IoT for Stress Level Detection Based on Physiological Signals", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 1, pp. 102 - 112, Feb. 2026.

Permalink/DOI: <https://doi.org/10.29207/resti.v10i1.6770>

Received: June 13, 2025

Accepted: August 26, 2025

Available Online: [February 23, 2026]

This is an open-access article under the [CC BY 4.0 License](#)

Published by [Ikatan Ahli Informatika Indonesia](#)

1. Introduction

Mental health issues among Indonesian students are becoming increasingly alarming. The 2022 National Adolescent Mental Health Survey (I-NAMHS) revealed that 34.9% of adolescents experienced mental health issues in the past 12 months, and 5.5% of them suffered from mental disorders[1], [2]. Ironically, only 2.6% of them accessed support or counseling services for emotional and behavioral problems[3], [4].

Academic pressure, social adaptation, and technological developments contribute to increased stress and emotional disorders among students[5]. During the COVID-19 pandemic, 1 in 20 adolescents reported an increase in symptoms of depression, anxiety, loneliness, and difficulty concentrating

compared to before the pandemic[6]. Mental health assessments are currently still subjective, relying on questionnaires without the support of objective physiological data. However, IoT technology enables the real-time acquisition of physiological signals such as heart rate, SpO₂, body temperature, and blood pressure, which have been proven to correlate with stress and other emotional conditions[7], [8].

Previous research has developed a physiological data-based stress detection system using methods such as Fuzzy Logic[9-12] and Support Vector Machine [13], [14]. However, this approach has limitations in handling the complexity of multivariate data and is not optimal in utilizing interactions between physiological signals. In addition, most studies only utilize one or two types of signals without comprehensive integration of

various physiological indicators. Other researchers developed a stacking ensemble model for classifying student stress levels with 95% accuracy. However, the main drawback is that hyperparameter optimization has not been applied, leading to overfitting in prediction results, and the model has not been integrated with real-time IoT-based data, making it suboptimal for direct implementation in stress monitoring using smart devices[15]. The Internet of Things (IoT) has emerged as a transformative technology that connects physical devices and sensors to digital platforms, enabling real-time data acquisition and intelligent decision-making across diverse domains[16-19]. Its applications range from healthcare, agriculture, and smart cities to industrial systems, providing opportunities to develop innovative solutions that enhance efficiency, monitoring, and automation[20-24].

Previous studies have developed stress detection systems using various approaches, such as IoT-based ensemble voting with five ML models and physiological sensors, but their accuracy of 75% is still limited and weaker than stacking because they do not adaptively utilize the relationships between models[25]. Another approach uses body motion data from IMU [26], AI-based HRV [27], or wearable IoT for the elderly [28]. However, they generally do not involve comprehensive physiological signals or apply ensemble learning. Studies using smartphone sensors and typing behavior have also been conducted with KNN and genetic algorithms[29], but without physiological signals and without ensemble model exploration.

The stacking ensemble model offers higher prediction accuracy and stability by combining various learning algorithms[30], [31]. Several studies have applied stacking ensemble for classifying student stress with 95% accuracy[15]. However, it does not yet support IoT, real-time detection, or meta-feature engineering. Another study discusses wearable-based detection with CNN, LSTM, and SVM, but has not yet implemented stacking or adaptive models[32]. Meanwhile, several IoT systems have been designed to monitor physiological parameters such as oxygen saturation and heart rate[7], [8], [33], but have not yet implemented AI and generally use non-industrial sensors. The use of smartwatches for stress detection in autism is also limited to closed systems without AI models [34].

On the other hand, studies on graduation prediction and chronic diseases show that stacking ensembles outperform single models[35-37], including the NSGA-II approach and stacked autoencoder[38], [39]. However, a common weakness of all these studies is that meta-feature engineering and hyperparameter tuning optimization have not been applied, so the potential of stacking has not been maximized optimally.

As an extension of previous studies, this research proposes a new model, the New Adaptive Stacking Ensemble Model, for detecting the mental health of students based on physiological signals and IoT technology. This model is designed to classify mental

conditions such as relaxed, calm, anxious, and tense in real-time and objectively. It combines five basic algorithms (SVM, C4.5, Decision Tree, KNN, Random Forest) and a meta-model MLP.

The model is optimized using hyperparameter tuning (Optuna) and data balancing with SMOTE, and validated using k-fold cross-validation. From a hardware perspective, the system is supported by IoT-based devices. This approach combines artificial intelligence, real-time data acquisition, and advanced predictive processing, which have not been comprehensively applied in previous studies, thereby potentially offering significant improvements in detection accuracy and original contributions to the fields of AI and mental health education. Unlike traditional ensemble models, the adaptive stacking approach dynamically combines base model predictions and meta-feature representations in a layer-wise manner, allowing improved generalization and robustness against class noise. Moreover, the integration of Optuna enables efficient and automated hyperparameter tuning through Bayesian optimization, outperforming grid or random search strategies in both speed and performance reliability.

2. Research Methods

The research flow was systematically designed to integrate IoT devices, machine learning processes, and model optimization to produce an accurate and real-time mental health detection system. Details of the stages in the development process of the New Adaptive Stacking Ensemble Model integrated with IoT are presented in Figure 1.

The stages of development of the New Adaptive Hybrid Stacking Ensemble Model for detecting the mental health of students based on physiological signals and IoT.

2.1 Dataset

The physiological dataset was collected through a controlled experimental setup at Universitas Hang Tuah Pekanbaru over a span of three weeks in early 2024. Participants were asked to sit comfortably while their physiological signals were recorded using certified medical devices connected to an IoT system.

The data collection was supervised by a multidisciplinary team consisting of computer scientists and health professionals. Each session lasted approximately 10 - 15 minutes per participant, during which five physiological parameters heart rate, SpO₂, body temperature, systolic, and diastolic blood pressure were recorded. The mental condition of participants during the session was concurrently assessed and labeled by licensed psychologists based on observable indicators and short semi-structured interviews. A sample of the physiological dataset used in this study is presented in Table 1.

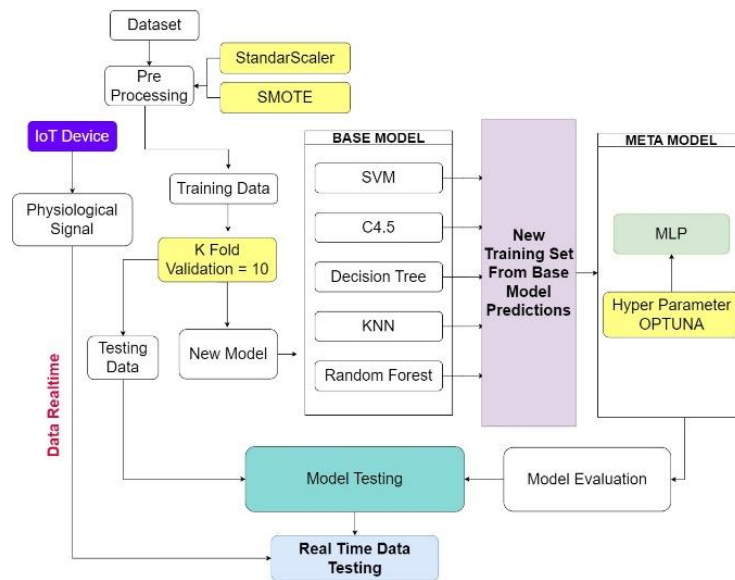


Figure 1. Research Flow

Table 1. Physiological Dataset

SpO ₂	Heart Rate	Temperature	Systolic	Diastolic	Condition
0.19	0.44	-0.06	0.33	0.40	Anxious
0.37	0.68	-0.15	0.07	0.64	Anxious
-0.69	-0.33	0.68	-0.69	-1.18	Calm
0.01	-0.39	0.90	-0.26	-0.55	Calm
-0.33	-0.63	0.68	0.86	-0.07	Calm

The complete dataset consists of 4,700 physiological records. A total of 1,200 records were collected directly from 40 university students aged between 18 and 24 years (mean age: 20.2 ± 1.4 years). These participants were full-time undergraduate students at Hang Tuah University Pekanbaru, recruited voluntarily from the Faculty of Computer Science. Inclusion criteria included being in good physical health and not undergoing psychiatric treatment at the time of participation. All participants completed an informed consent form prior to data collection.

To increase the diversity and volume of the dataset, an additional 3,500 physiological records were sourced from publicly available datasets on Kaggle, focusing on stress-related biometric readings with similar features: heart rate, SpO₂, temperature, and blood pressure. This additional dataset was carefully filtered, cleaned, and standardized to match the structure and labeling scheme of the primary dataset.

To ensure compatibility, the Kaggle dataset underwent several preprocessing steps to align it with the primary dataset. The records were filtered to retain only features consistent with the physiological variables used in this study. Incomplete and outlier entries were removed to improve data quality. Column labels were then renamed and aligned with the primary dataset structure, followed by normalization using the StandardScaler procedure identical to the main preprocessing pipeline.

2.2 IoT Data Acquisition for Real-Time Testing

To test the model's performance in real time, an IoT system was developed that is capable of capturing physiological data from users directly. IoT technology provides a framework for connecting sensors and devices to digital platforms, enabling seamless real-time data exchange and intelligent monitoring in various application domains[40], [41].

Figure 2 presents the block diagram of the IoT system developed for real-time data acquisition.

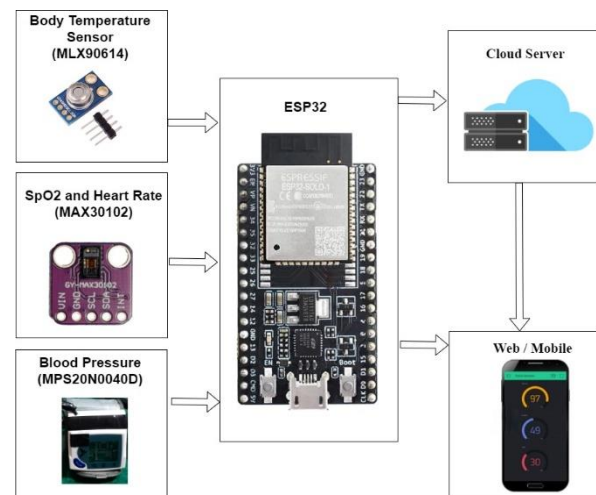


Figure 2. IoT System Sketch Block

This system consists of several main components: MLX90614 Sensor: used to measure body temperature without contact, based on infrared technology. MAX30102 Sensor: measures heart rate and blood oxygen saturation (SpO₂). IoT Digital Tensimeter: measures systolic and diastolic blood pressure; data is automatically transmitted via wireless connection. ESP32: serves as the main microcontroller coordinating sensor readings and sending data to the server via Wi-Fi. This device is only used in the final testing phase of

the system to demonstrate its ability to predict mental states in real time based on sensor input.

2.3 Data Pre-processing

Data preprocessing aims to prepare the dataset so that it is ready for use in model training[42]. This stage includes:

Since features have different value scales, standardization is performed using StandardScaler. This technique converts each feature value into a distribution with a mean of 0 and a standard deviation of 1. This is important to ensure that the learning model is not biased toward features with larger values. Features after the standard scaling process are presented in Table 2.

Table 2. Features after Standard Scaling

SpO ₂	Heart Rate	Temperature	Systolic	Diastolic	Condition
96	96	34,35	125	95	Anxious
98	100	34,20	122	98	Anxious
91	83	35,60	113	75	Calm
95	82	35,95	118	83	Calm
93	78	35,60	111	89	Calm
101	125	31,50	138	116	Tense
87	62	36,80	108	75	Relaxed

The distribution of labels in the dataset is uneven. To address this imbalance, SMOTE (Synthetic Minority Over-sampling Technique) is used to generate new synthetic data for minority classes based on nearest neighbor interpolation[43],[44]. This technique effectively prevents overfitting to the majority class. Table 3 presents the class distribution before and after SMOTE was applied.

Table 3. Class Distribution Before and After SMOTE

Stress Level	Before SMOTE	After SMOTE
Anxious	1103	1103
Calm	869	1103
Relaxed	535	1103
Tense	486	1103

Before SMOTE, there was an imbalance between classes, with the Anxious class dominating with 1,103 samples, while the Tense class only had 486 samples. This imbalance had the potential to cause bias in the model training process. After SMOTE was applied, all classes were balanced to 1,103 samples per class, so that the model had a fair training distribution and could improve classification performance for minority classes.

2.4 Model Architecture

The prediction model was built using the stacking ensemble approach, which is a technique that combines several base models to produce more accurate predictions through a meta model.

Each base model generates independent predictions on the test data, which are then fed into the meta model. The base models used are SVM, Decision Tree, C4.5, KNN, and Random Forest.

Multi-Layer Perceptron (MLP) is trained using prediction results from the entire base model to learn combined patterns and improve classification performance.

The stacking ensemble formulation used in this study is expressed in Equation 1.

$$\hat{y}_{meta} = f_{meta}(f_1(X), f_2(X), \dots, f_n(X)) \quad (1)$$

X represented original input feature matrix. $f_1, f_2, \dots, f_n$: base model. f_{meta} : meta model. $\hat{y}_{meta}$ = final prediction results from the stacking model.

In the stacking ensemble framework, each base model (SVM, C4.5, Decision Tree, KNN, Random Forest) is trained independently using the same input features. Once trained, these models generate class probability predictions for each instance in the dataset. These predicted probabilities rather than raw labels are then used as new features that form a secondary dataset. This secondary dataset becomes the input for the meta-model (MLP), which learns how to best combine the outputs of the base learners to improve overall prediction accuracy and reduce individual model biases.

2.5 Hyperparameter Optimization

The MLP meta model was optimized using Optuna, a hyperparameter search framework based on Bayesian Optimization. Parameters such as the number of hidden neurons, learning rate, and dropout were explored to find the best configuration based on validation accuracy. Optuna builds a surrogate model (typically a Tree-structured Parzen Estimator) to approximate the objective function and selects hyperparameter values that are likely to improve performance in future trials. It evaluates each trial by training the model with a sampled configuration and uses the result to guide subsequent sampling. This process continues iteratively to find the optimal set of hyperparameters efficiently with minimal computation cost.

The parameters selected for optimization include the number of hidden units, learning rate, dropout rate, and activation function. These were chosen because they significantly affect the learning capacity, generalization, and regularization of the MLP, and have been shown in literature to influence ensemble performance in high-dimensional physiological data.

To avoid overfitting and measure model generalization, 10-Fold Cross Validation is used. The dataset is divided into 10 subsets; each subset is used as test data alternately, while the rest is used as training data. The evaluation results are calculated as the average of 10 tests.

2.6 Evaluation

The model was evaluated using the following multi-class classification metrics:

Accuracy: Measure the percentage of correct predictions from all data. The accuracy metric used to

evaluate overall classification performance is defined in Equation 2.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

In this context, true positive (TP) refers to correctly predicted instances belonging to the target class, while true negative (TN) denotes correctly predicted instances outside the class. False positive (FP) represents incorrect predictions assigning samples to the class, whereas false negative (FN) corresponds to missed detections where samples belonging to the class were predicted otherwise.

Precision: Measuring the accuracy of positive predictions. Precision measures the correctness of positive predictions and is formulated in Equation 3.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Recall (Sensitivity): Measures the success of the model in finding all positive cases.

F1-Score: Harmonic mean of precision and recall. The F1-score evaluates the balance between precision and recall and is expressed in Equation 4.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

Confusion Matrix: A 4x4 matrix showing the distribution of predictions for each target class.

ROC Curve and AUC: For each class, the Receiver Operating Characteristic (ROC) curve and Area Under Curve (AUC) value are displayed, illustrating the trade-off between TPR and FPR[45].

This evaluation provides a comprehensive overview of the model's performance in multi-class classification and its reliability when implemented in real-time conditions.

2.7 Ethical Approval

This study was conducted following ethical standards outlined in the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee of the Faculty of Health, Universitas Hang Tuah Pekanbaru.

All participants were provided with detailed information about the study's purpose, procedures, and data usage. Written informed consent was obtained prior to any data collection, and participants were allowed to withdraw at any time without penalty. All data were anonymized to protect participants' identities and confidentiality.

3. Results and Discussions

3.1 Application of SMOTE for Data Balancing

An unbalanced class distribution in a dataset can cause machine learning models to be biased toward the majority class and ignore the minority class. In this study, before oversampling was performed, the mental condition data showed an imbalance in distribution, with the Anxious class having 1,103 samples, while the Tense class had only 486 samples. This imbalance has the potential to reduce the accuracy of predictions for the minority class. To address this issue, the SMOTE technique was used. SMOTE works by synthesizing new data for the minority class based on interpolation from existing data. Figure 3 illustrates the class distribution before and after applying SMOTE.

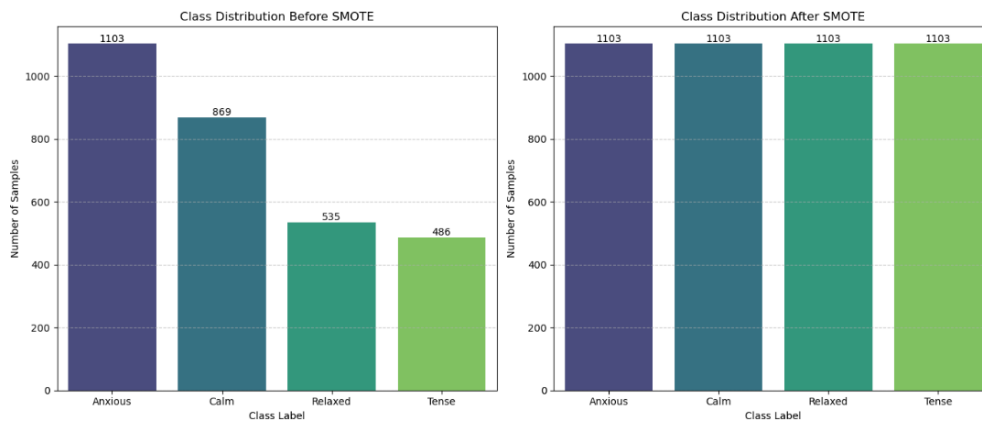


Figure 3. Class Distribution Before and After SMOTE Implementation

After implementation, the number of samples for each class became balanced, with 1,103 samples each. This equalization of sample numbers resulted in a more balanced training dataset, allowing the model to recognize patterns from all classes fairly. The impact was seen in improved classification performance, especially for minority classes such as Relaxed and Tense in the final evaluation.

3.2 Model Evaluation Results

In this study, the term new model refers to the New Adaptive Stacking Ensemble framework, which combines five machine learning algorithms (SVM, C4.5, Decision Tree, KNN, and Random Forest) as base models. The predictions from these models are then used as input for the meta-model, namely Multi-Layer Perceptron (MLP), which learns to improve classification accuracy through combined decision patterns. To measure the performance of each basic

machine learning algorithm used in this study, five models were evaluated: SVM, C4.5, Decision Tree, KNN, and Random Forest. The evaluation was based on four main metrics for multi-class classification, namely accuracy, precision, recall, and F1-score. A summary of the evaluation results is presented in Table 4 and Figure 4.

Table 4. Classification Report Results on Base Model

Model	Accuracy	Precision	Recall	F1-Score
SVM	0.89	0.90	0.89	0.89
C4.5	0.89	0.89	0.89	0.89
DT	0.89	0.88	0.89	0.89
KNN	0.91	0.91	0.91	0.91
RF	0.91	0.91	0.91	0.91

The accuracy comparison results of the five base models are shown in Figure 4.

Table 3 and Figure 4 show the performance of five basic machine learning algorithms used in mental health prediction systems, namely SVM, C4.5, Decision Tree, KNN, and Random Forest. The evaluation was based on four main metrics: accuracy, precision, recall, and F1-score. In general, KNN and Random Forest showed the best performance with consistent average scores across all four metrics at 0.91. This indicates that both models are capable of providing stable and balanced predictions in recognizing all mental condition classes (Anxious, Calm, Relaxed, Tense).

The SVM and C4.5 models have identical accuracy and F1-scores of 0.89, indicating good performance but still slightly below KNN and RF. Meanwhile, the Decision Tree (DT) has the lowest precision score (0.88), although it still has a recall and F1-score of 0.89, indicating that this model tends to produce more false positives than other models. These results indicate that while all base models have adequate performance for multi-class classification, models like KNN and RF are superior and more suitable as primary candidates in an ensemble due to their stability in predicting all classes with balanced precision and sensitivity.

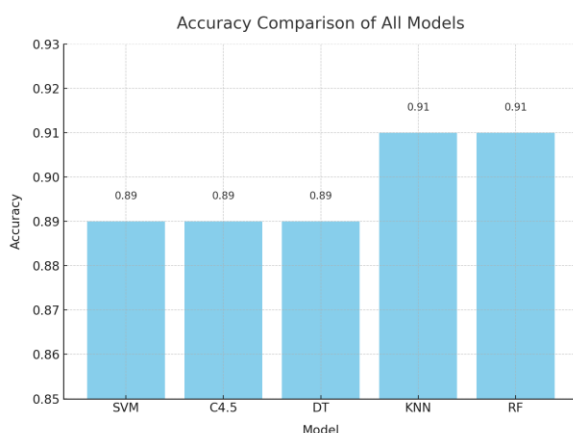


Figure 4. Base Model Accuracy Comparison Chart

3.3 Stacking Model Evaluation Results

After testing each base model independently, the next step is to evaluate the performance of the stacking model that combines all base models (SVM, C4.5, DT, KNN, RF) using Multi-Layer Perceptron (MLP) as a meta-learner. This stacking model is trained using prediction data from the base model and then evaluated against test data to measure the effectiveness of integration between models. The classification results of the stacking model are summarized in Table 5.

Table 5. Classification Report Results on the MLP Meta Model

Class	Precision	Recall	F1-Score	Support
Anxious	0.95	0.89	0.92	245
Calm	0.90	0.92	0.91	245
Relaxed	0.89	0.90	0.89	245
Tense	0.85	0.88	0.87	245
Accuracy	-	-	0.90	980
Macro Avg	0.90	0.90	0.90	980
Weighted Avg	0.90	0.90	0.90	980

Based on Table 5, it can be concluded that the META MLP model shows excellent overall classification performance with consistent average accuracy, precision, recall, and F1-score values of 0.90. The Calm and Anxious classes have the highest F1-scores of 0.91 and 0.92, respectively, indicating that the model is able to recognize these two mental states very well. Meanwhile, the Tense class has the lowest F1-score of 0.87, suggesting that predictions for this class are relatively more complex and may be ambiguous with other classes. The highest precision is achieved in the Anxious class at 0.95, indicating that most positive predictions for this class are correct. The consistency between the macro average and weighted average values also shows a balanced performance distribution across all classes. Overall, the META MLP model achieves an accuracy of 0.90, though it is lower than the individual KNN and Random Forest models.

3.4 Results of Stacking Model Evaluation with OPTUNA

To improve the performance of the stacking model, optimization of the hyperparameters of the MLP meta-model was performed using the Optuna framework. The goal of this process was to find the best combination of parameters, such as the number of hidden neurons, alpha value, and learning rate, to produce a more efficient and accurate model. The results of the model evaluation after hyperparameter tuning are presented as a confusion matrix in Figure 5.

Based on Figure 5, it can be seen that the Meta MLP model that has been optimized using Optuna shows a significant improvement in classification performance compared to before tuning. The model is able to classify the four classes with high precision and a very low error rate. For example, the Anxious class was correctly recognized in 216 out of 221 data points, with only 5 data points incorrectly classified as Tense. The Calm class even has very accurate prediction results, with 215 correct classifications and only 5 minor errors spread across the other three classes. The Relaxed and Tense

classes also show classification stability with minimal errors. The increasingly focused prediction distribution along the main diagonal of the confusion matrix indicates an improvement in the model's generalization ability and the effectiveness of the tuning strategy applied. These results demonstrate that integrating Optuna into the meta-model training process enhances prediction accuracy and strengthens the model's overall reliability.

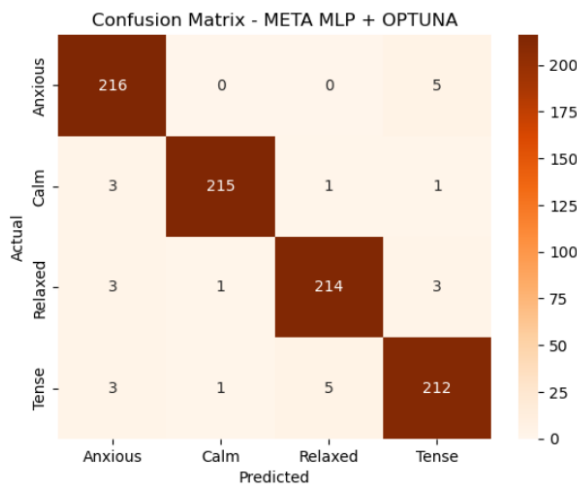


Figure 5. Confusion Matrix Meta MLP + OPTUNA

To provide a comprehensive overview of the performance of the META MLP model that has been optimized using Optuna, an evaluation was conducted on classification metrics, namely precision, recall, and F1-score for each class as well as the overall average. The evaluation results are summarized in Table 6.

Table 6. Classification Report Results on the MLP+OPTUNA Meta Model

Class	Precision	Recall	F1-Score	Support
Anxious	0.96	0.98	0.97	221
Calm	0.99	0.98	0.98	220
Relaxed	0.97	0.97	0.97	221
Tense	0.96	0.96	0.96	221
Accuracy	-	-	0.97	883
Macro Avg	0.97	0.97	0.97	883
Weighted Avg	0.97	0.97	0.97	883

The classification report results show that the META MLP + OPTUNA model achieved very high and stable classification performance across all evaluation metrics. The overall accuracy reached 0.97, with consistent macro average and weighted average values, indicating consistent performance across all classes. The Calm class had the highest precision of 0.99, indicating that the model rarely misclassified this condition. Meanwhile, the Anxious class has the highest recall of 0.98, meaning that most Anxious cases are successfully recognized by the model. All classes show F1-scores above 0.96, reflecting a balance between precision and recall. This improvement compared to the META MLP model without tuning confirms that the hyperparameter optimization process with Optuna can significantly enhance the model's accuracy and reliability.

As part of the model performance evaluation, ROC curves are used to measure the model's ability to distinguish between classes. Each ROC curve represents the relationship between the true positive rate (TPR) and false positive rate (FPR) for each class. The AUC value is used to assess the quality of the model's probabilistic predictions. The ROC curve visualization is presented in Figure 6.

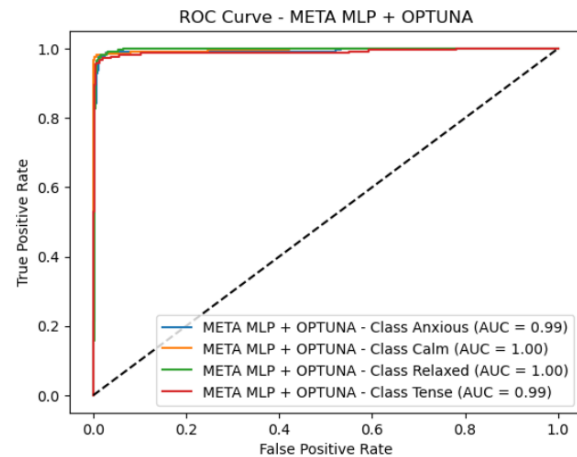


Figure 6. ROC Curve Meta MLP + OPTUNA

Based on Figure 6, the META MLP model optimized with Optuna shows very high predictive performance across all classes. The Calm and Relaxed classes each recorded a perfect AUC of 1.00, while the Anxious and Tense classes each obtained an AUC of 0.99. The curve approaching the upper left corner indicates that the model has very high sensitivity and specificity. These results reflect that the model is able to distinguish each class very well. Although very high AUC values can generally be an early indication of overfitting, in this study, this risk can be mitigated because the evaluation was performed using 10-Fold Cross Validation. This technique ensures that the model is tested on data not seen during training in each fold, so the results obtained can be considered generalizable and not merely the result of training on overly similar data.

3.5 Final Evaluation with 10-Fold Cross Validation

As a final validation of the reliability of the Meta MLP + OPTUNA model, an evaluation was conducted using the 10-Fold Cross Validation technique. This approach divides the training data into 10 parts (folds), where each fold is alternately used as test data while the other folds are used for training. This technique is effective in assessing the stability and generalization ability of the model as a whole. The accuracy results of each fold are presented in Figure 7.

Figure 7 shows the variation in model accuracy across each fold, with the highest value being 0.9796 (in Folds 6 and 9) and the lowest being 0.9501 (Fold 7). Although there is some fluctuation between folds, all accuracy values remain above 0.95, and the overall average accuracy is recorded at 0.9683. This indicates that the model performs stably and consistently across various subsets of data. No fold shows an extreme drop in

accuracy, suggesting no signs of overfitting. This validation process reinforces that the Meta MLP + OPTUNA model not only excels on the initial training data but also maintains high performance on previously unseen data.

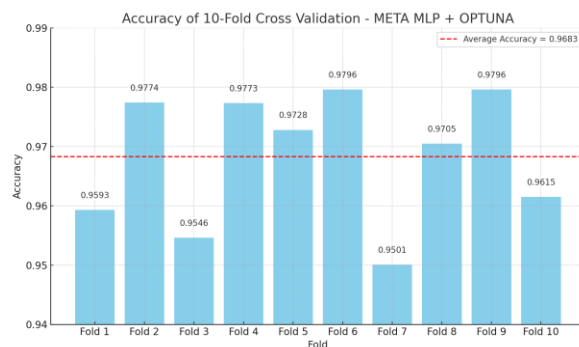


Figure 7. Accuracy of 10-Fold Cross Validation

Although the model consistently achieved high accuracy (average 96.83%) during 10-fold cross-validation, it is important to critically consider the potential for overfitting. To minimize this risk, we applied SMOTE to balance class distribution, introduced dropout regularization in the MLP meta-model, and performed stratified sampling during validation. Furthermore, the stability of performance metrics across folds and the near-perfect AUC scores indicate strong generalization. Nonetheless, we recognize that additional external validation on independent datasets would be a valuable step to further confirm the model's robustness in real-world deployment.

3.6 Model Accuracy Comparison

To determine the relative performance of each algorithm in detecting stress levels based on physiological signals, accuracy comparisons were made between five base models, a stacking model with meta MLP, and a stacking model optimized using Optuna. The accuracy comparison between models is presented in Figure 8.

Figure 8 shows that the Stacking MLP + OPTUNA model achieved the highest accuracy of 0.97, demonstrating a significant advantage over other

models. Meanwhile, the two base models, KNN and Random Forest, achieved the same high accuracy of 0.91, outperforming the Stacking MLP model without tuning, which only reached 0.90. The other three base models (SVM, C4.5, and Decision Tree) had lower accuracy at 0.89. The fact that Stacking MLP has not yet surpassed KNN and RF indicates that the ensemble process without optimization is not yet sufficiently effective. However, after hyperparameter tuning using Optuna, the stacking model showed a significant improvement and outperformed all base models. This underscores the importance of the optimization process to maximize the potential of stacking ensembles in physiological signal-based stress condition classification.

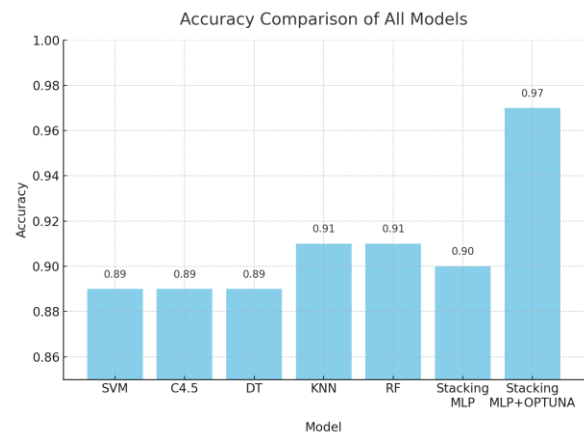


Figure 8. Accuracy Comparison of All Models

Although the base stacking model already showed competitive performance, the use of Optuna allowed fine-grained tuning of key hyperparameters such as learning rate, hidden layer size, and dropout which enhanced the model's ability to generalize and reduce misclassification errors in borderline samples. This explains the notable accuracy gain even from an already strong baseline.

To assess the superiority of the proposed model, a comparison of accuracy was made with several previous studies that used the stacking ensemble approach. Details of the comparison with previous studies are presented in Table 7.

Table 7. Comparison of Accuracy Results with Previous Research

Researcher	Base Model	Meta Model	Accuracy
[46]	KNN, AdaBoost, NB, Decision trees	Logistic Regression	86.00%
[47]	RF, XGBoost, Gradient Boosting, Feed-Forward Neural Networks (FNN)	Feed-Forward Neural Networks (FNN)	92.10%
[48]	KNN, Bayesian Networks, Decision Tree, Gradient Boosting	Logistic Regression	94.12%
[49]	SVM, XGBoost, ANN	Logistic Regression	95.00%
[15]	SVM, LR, MLP, RF	XGBoost	95.00%
[35]	KNN, SVM, C4.5, RF, NB	ADABOOST+OPTUNA	95.50%
This Research			97.00%

Table 7 shows that the accuracy of the model in this study reached 97.00%, higher than all previous studies, which ranged from 86% to 95.50%. The stacking approach with an MLP meta model optimized using

Optuna proved to be capable of providing superior and efficient classification performance in detecting conditions based on physiological data. However, none of these studies incorporated IoT-based real-time

physiological signal acquisition or hyperparameter optimization using Optuna. Study [46] utilized KNN, AdaBoost, NB, and Decision Trees on a road traffic dataset, while [47] applied RF and FNN for student dropout prediction using educational log data. In [48] and [49], stacking models used combinations of Bayesian networks, XGBoost, and ANN on depression and COVID-19 datasets, respectively. However, none of these studies incorporated IoT-based real-time physiological signal acquisition or hyperparameter optimization using Optuna. Our study integrates these components, achieving 97.00% accuracy, which surpasses prior results and demonstrates a more applicable framework for real-time mental health monitoring.

3.7 Real-Time Model Testing with IoT

As a final step, real-time testing of the built model was conducted by integrating it into an Internet of Things (IoT) device. This device consists of several physiological sensors used to read user data directly. The physical appearance of the IoT device is shown in Figure 9.

Figure 9 shows the IoT device consisting of an ESP32 microcontroller, MLX90614 temperature sensor, MAX30102 heart rate and SpO₂ sensor, and a modified digital sphygmomanometer for reading systolic and diastolic blood pressure. All physiological data read by the sensors are automatically sent to a Google Sheet as a real-time database. This data is then analyzed by a Streamlit-based prediction system using a Stacking MLP + OPTUNA model to classify stress levels. Meanwhile, the LCD screen on the device displays sensor values directly to the user. Table 8 shows the sensor data recorded from the IoT device and displayed in Google Sheets.



Figure 9. IoT Device Display

After the IoT system is integrated with the prediction model, a field trial is conducted involving users as participants. The visualization of the testing results and system interface is presented in Figure 10.

Table 8. View Sensor Data from IoT Devices

Temperature (°C)	SpO ₂ (%)	Heart Rate (BPM)	SYS	DIA
36.5	98.2	99	166	100
36.5	98.2	97	164	97
36.5	98.2	94	80	58
36.5	98.2	91	150	80
36.5	98.2	96	138	89
36.5	98.2	69	150	90

Figure 10 shows the results of live system testing, where the sensor reads body temperature, heart rate, SpO₂, and systolic and diastolic blood pressure data, then sends it to Google Sheet and is processed by the predictive model in the backend. The output in the form of stress levels is displayed in real-time on the Streamlit dashboard in the form of classifications such as Relaxed, Anxious, and Tense. In addition, the dashboard also provides a collective view for historical data monitoring

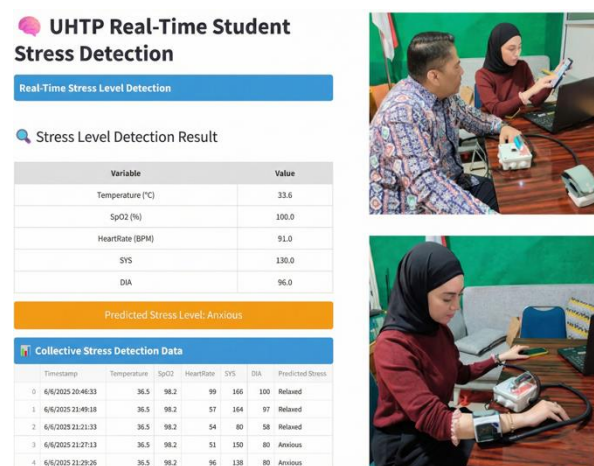


Figure 10. Stress Detection Dashboard View and Test Documentation

4. Conclusions

This study successfully developed and implemented the New Adaptive Stacking Ensemble model for detecting stress levels in students based on physiological data. The model consists of five basic algorithms (SVM, C4.5, DT, KNN, and RF) and one MLP meta-model optimized using Optuna, resulting in the highest accuracy of 97.00%, surpassing all base models and also models without tuning. Evaluation using 10-Fold Cross Validation proves that the model has high predictive stability and shows no indication of overfitting.

Another advantage of this research is the successful integration of the prediction system with IoT devices. Data from body temperature sensors (MLX90614), heart rate and SpO₂ (MAX30102), and digital blood pressure are sent in real-time to Google Sheet, then processed by the model through the Streamlit interface. This system is not only able to classify accurately, but also displays the results to users via the dashboard and LCD screen on the device. This proves that the developed model is not only superior theoretically, but

can also be implemented practically and in real-time in the field.

Further research can develop the integration of this model into a mobile application to make it more portable and easily accessible to users. In addition, it is necessary to explore the addition of psychological variables such as fatigue levels or sleep patterns to improve classification accuracy. The ensemble model can also be extended by utilizing transformer-based meta learners and edge computing to reduce dependence on internet connections and speed up the inference process locally.

Acknowledgements

This research is funded by the Directorate of Research and Community Service, Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology with the Regular Fundamental Research Scheme in 2025, in accordance with Research Contract Number: 010/LL17/DT.05.00/PL/2025.

References

- [1] S.-I. Kim *et al.*, "Association of parental depression with adolescent children's psychological well-being and health behaviors," *BMC Public Health*, vol. 24, no. 1, p. 1412, 2024, doi: [10.1186/s12889-024-18337-9](https://doi.org/10.1186/s12889-024-18337-9).
- [2] A. Juliansen, R. S. Heriyanto, M. P. Muljono, C. L. Budiputri, Y. D. S. Sagala, and G. S. Octavius, "Mental health issues and quality of life amongst school-based adolescents in Indonesia," *J. Med. Surgery, Public Heal.*, vol. 2, p. 100062, 2024, doi: <https://doi.org/10.1016/j.glmedi.2024.100062>.
- [3] M. D. Pham *et al.*, "Mental Health Problems Among Indonesian Adolescents: Findings of a Cross-Sectional Study Utilizing Validated Scales and Innovative Sampling Methods," *J. Adolesc. Heal.*, vol. 75, no. 6, pp. 929–938, 2024, doi: <https://doi.org/10.1016/j.jadohealth.2024.07.016>.
- [4] N. Murphy, P. Azzopardi, and M. A. Bohren, "Safeguarding youth sexual and reproductive health and rights in the context of increasing climate-related disasters in the Pacific: A scoping review of policies and responses," *Int. J. Disaster Risk Reduct.*, vol. 86, p. 103561, 2023, doi: <https://doi.org/10.1016/j.ijdrr.2023.103561>.
- [5] C. Conversano *et al.*, "Is Mindfulness-Based Stress Reduction Effective for People with Hypertension? A Systematic Review and Meta-Analysis of 30 Years of Evidence," *Int. J. Environ. Res. Public Health*, vol. 18, no. 6, 2021, doi: [10.3390/ijerph18062882](https://doi.org/10.3390/ijerph18062882).
- [6] R. Kindred and G. W. Bates, "The Influence of the COVID-19 Pandemic on Social Anxiety: A Systematic Review," *Int. J. Environ. Res. Public Health*, vol. 20, no. 3, 2023, doi: [10.3390/ijerph20032362](https://doi.org/10.3390/ijerph20032362).
- [7] Y. Devis, Y. Irawan, Junadhi, F. Zoromi, Herianto, and M. R. Amarta, "Monitoring System of Heart Rate, Temperature and Infusion in Patients Based on Microcontroller (Arduino Uno)," *J. Phys. Conf. Ser.*, vol. 1845, no. 1, pp. 0–6, 2021, doi: [10.1088/1742-6596/1845/1/012069](https://doi.org/10.1088/1742-6596/1845/1/012069).
- [8] Y. Irawan, Y. Fernando, and R. Wahyuni, "Detecting Heart Rate Using Pulse Sensor As Alternative Knowing Heart Condition," *J. Appl. Eng. Technol. Sci.*, vol. 1, no. 1, pp. 30–42, 2019, doi: [10.37385/jaets.v1i1.16](https://doi.org/10.37385/jaets.v1i1.16).
- [9] R. Rosdiana, A. Hasibuan, A. Chairani, M. Daud, and M. Sayuti, "Stress Diagnosis System Using Fuzzy Logic Method," *Motiv. J. Mech. Electr. Ind. Eng.*, vol. 4, no. 3, pp. 355–366, 2022, doi: [10.46574/motivaction.v4i3.162](https://doi.org/10.46574/motivaction.v4i3.162).
- [10] S. Naqvi, A. Z. Shaikh, T. Altaf, and S. Singh, "Fuzzy Logic Enabled Stress Detection Using Physiological Signals BT - Emerging Technologies in Computing," in *Emerging Technologies in Computing*, M. H. Miraz, G. Southall, M. Ali, A. Ware, and S. Soomro, Eds., Cham: Springer International Publishing, 2021, pp. 161–173, doi: <https://doi.org/10.3390/s25237108>.
- [11] M. Amin *et al.*, "ECG-Based Driver's Stress Detection Using Deep Transfer Learning and Fuzzy Logic Approaches," *IEEE Access*, vol. 10, pp. 29788–29809, 2022, doi: [10.1109/ACCESS.2022.3158658](https://doi.org/10.1109/ACCESS.2022.3158658).
- [12] B. Saputra, A. Utami, E. Edriyansyah, and Y. Irawan, "Expert System For Diagnosing Diseases in Toddlers Using The Certainty Factor Method," *J. Appl. Eng. Technol. Sci.*, vol. 4, no. 1, pp. 32–41, 2022, doi: [10.37385/jaets.v4i1.916](https://doi.org/10.37385/jaets.v4i1.916).
- [13] O. Vargas-Lopez, C. A. Perez-Ramirez, M. Valtierra-Rodriguez, J. J. Yanez-Borjas, and J. P. Amezquita-Sanchez, "An Explainable Machine Learning Approach Based on Statistical Indexes and SVM for Stress Detection in Automobile Drivers Using Electromyographic Signals," *Sensors*, vol. 21, no. 9, 2021, doi: [10.3390/s21093155](https://doi.org/10.3390/s21093155).
- [14] T. G. Troyee, M. F. Karim Khondakar, M. Hasan, and M. H. Chowdhury, "A Comparative Analysis of Different Preprocessing Pipelines for EEG-based Mental Stress Detection," in *2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, 2024, pp. 370–375, doi: [10.1109/ICEEICT62016.2024.10534380](https://doi.org/10.1109/ICEEICT62016.2024.10534380).
- [15] H. Fonda, Y. Irawan, R. Melyanti, R. Wahyuni, and A. Muhaimin, "A Comprehensive Stacking Ensemble Approach for Stress Level Classification in Higher Education," *J. Appl. Data Sci.*, vol. 5, no. 4, pp. 1701–1714, 2024, doi: <https://doi.org/10.47738/jads.v5i4.388>.
- [16] Y. Irawan, "Moving load robot using wifi network and android based," *J. Robot. Control*, vol. 2, no. 3, pp. 217–220, 2021, doi: [10.18196/jrc.2382](https://doi.org/10.18196/jrc.2382).
- [17] R. Wahyuni, J. T. Sentana, Muhardi, and Y. Irawan, "Water level control monitoring based on arduino uno R3 ATmega 238p using Lm016l LCD at STMIK Hang Tuah Pekanbaru," *J. Robot. Control*, vol. 2, no. 4, pp. 265–269, 2021, doi: [10.18196/jrc.2489](https://doi.org/10.18196/jrc.2489).
- [18] Y. Irawan, H. Fonda, E. Sabna, and A. Febriani, "Intelligent quality control of shrimp aquaculture based on real-time system and IoT using mobile device," *Int. J. Eng. Trends Technol.*, vol. 69, no. 4, pp. 49–56, 2021, doi: [10.14445/22315381/IJETT-V69I4P208](https://doi.org/10.14445/22315381/IJETT-V69I4P208).
- [19] S. Purwanti, A. Febriani, M. Mardeni, and Y. Irawan, "Temperature Monitoring System for Egg Incubators Using Raspberry Pi3 Based on Internet of Things (IoT)," *J. Robot. Control*, vol. 2, no. 5, 2021, doi: [10.18196/jrc.25105](https://doi.org/10.18196/jrc.25105).
- [20] A. R. Abidin, Y. Irawan, and Y. Devis, "Smart Trash Bin for Management of Garbage Problem in Society," *J. Appl. Eng. Technol. Sci.*, vol. 4, no. 1, pp. 202–208, 2022, doi: [10.37385/jaets.v4i1.1015](https://doi.org/10.37385/jaets.v4i1.1015).
- [21] Y. Irawan, E. Sabna, A. F. Azim, R. Wahyuni, N. Belarbi, and M. M. Josephine, "Automatic Chili Plant Watering Based on Internet of Things (IoT)," *J. Appl. Eng. Technol. Sci.*, vol. 3, no. 2, pp. 77–83, 2022, doi: [10.37385/jaets.v3i2.532](https://doi.org/10.37385/jaets.v3i2.532).
- [22] Y. Irawan, A. W. Novrianto, and H. Sallam, "Cigarette Smoke Detection and Cleaner Based on Internet of Things (IoT) Using Arduino Microcontroller and MQ-2 Sensor," *J. Appl. Eng. Technol. Sci.*, vol. 2, no. 2, pp. 85–93, 2021, doi: [10.37385/jaets.v2i2.218](https://doi.org/10.37385/jaets.v2i2.218).
- [23] R. M. Sari, E. Sabna, R. Wahyuni, and Y. Irawan, "Implementation of open and close a housing gate portal using RFID card," *J. Robot. Control*, vol. 2, no. 5, pp. 363–367, Sep. 2021, doi: [10.18196/jrc.25108](https://doi.org/10.18196/jrc.25108).
- [24] Y. Irawan, R. Wahyuni, and H. Fonda, "Folding Clothes Tool Using Arduino Uno Microcontroller And Gear Servo," *J. Robot. Control*, vol. 2, no. 3, pp. 170–174, 2021, doi: [10.18196/jrc.2373](https://doi.org/10.18196/jrc.2373).
- [25] S. Hadrhi, M. Hadji, and W. Labidi, "A voting ensemble classifier for stress detection," *J. Inf. Telecommun.*, vol. 8, no. 3, pp. 399–416, 2024, doi: [10.1080/24751839.2024.2306786](https://doi.org/10.1080/24751839.2024.2306786).
- [26] R. Richer *et al.*, "Machine learning-based detection of acute psychosocial stress from body posture and movements," *Sci. Rep.*, vol. 14, no. 1, pp. 1–19, 2024, doi: [10.1038/s41598-024-59043-1](https://doi.org/10.1038/s41598-024-59043-1).
- [27] Y. Haque *et al.*, "State-of-the-Art of Stress Prediction from

- Heart Rate Variability Using Artificial Intelligence,” *Cognit. Comput.*, vol. 16, no. 2, pp. 455–481, 2024, doi: [10.1007/s12559-023-10200-0](https://doi.org/10.1007/s12559-023-10200-0).
- [28] A. K. Ahuja, B. P. Mishra, C. Shankar, and T. P. Dubey, “Stress Detection Using Bio-Signal Processing: An Application of IoT and Machine Learning for Old Age Home Residents,” *Eng. Proc.*, vol. 81, no. 1, 2024, doi: [10.3390/engproc2024081012](https://doi.org/10.3390/engproc2024081012).
- [29] E. A. Sağbaş, S. Korukoglu, and S. Ballı, “Real-time stress detection from smartphone sensor data using genetic algorithm-based feature subset optimization and k-nearest neighbor algorithm,” *Multimed. Tools Appl.*, vol. 83, no. 1, pp. 1–32, 2024, doi: [10.1007/s11042-023-15706-1](https://doi.org/10.1007/s11042-023-15706-1).
- [30] S. Reza, R. Amin, R. Yasmin, W. Kulsum, and S. Ruhi, “Heliyon Improving diabetes disease patients classification using stacking ensemble method with PIMA and local healthcare data,” *Heliyon*, vol. 10, no. 2, p. e24536, 2024, doi: [10.1016/j.heliyon.2024.e24536](https://doi.org/10.1016/j.heliyon.2024.e24536).
- [31] E. Alalwany and I. Mahgoub, “An Effective Ensemble Learning-Based Real-Time Intrusion Detection Scheme for an In-Vehicle Network,” *Electron.*, vol. 13, no. 5, pp. 1–14, 2024, doi: [10.3390/electronics13050919](https://doi.org/10.3390/electronics13050919).
- [32] G. Taskasaplidis, D. A. Fotiadis, and P. D. Bamidis, “Review of Stress Detection Methods Using Wearable Sensors,” *IEEE Access*, vol. 12, no. December 2023, pp. 38219–38246, 2024, doi: [10.1109/ACCESS.2024.3373010](https://doi.org/10.1109/ACCESS.2024.3373010).
- [33] R. Wahyuni, Herianto, Ikhtiyaruddin, and Y. Irawan, “IoT-Based Pulse Oximetry Design as Early Detection of Covid-19 Symptoms,” *Int. J. Interact. Mob. Technol.*, vol. 17, no. 3, pp. 177–187, 2023, doi: [10.3991/ijim.v17i03.35859](https://doi.org/10.3991/ijim.v17i03.35859).
- [34] I. Moraiti and A. Drigas, “Measuring the stress of autistic people with the help of a smartwatch, Internet of Things Technology,” *Brazilian J. Sci.*, vol. 3, no. 2, pp. 45–56, 2024, doi: [10.14295/bjs.v3i2.466](https://doi.org/10.14295/bjs.v3i2.466).
- [35] Herianto, B. Kurniawan, Z. H. Hartomi, Y. Irawan, and M. K. Anam, “Machine Learning Algorithm Optimization using Stacking Technique for Graduation Prediction,” *J. Appl. Data Sci.*, vol. 5, no. 3, pp. 1272–1285, 2024, DOI: <https://doi.org/10.47738/jads.v5i3.316>.
- [36] U. M. Butt et al., “Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications,” vol. 2021, 2021, doi: <https://doi.org/10.1155/2021/9930985>.
- [37] M. A. Rahim, M. A. Hossain, M. N. Hossain, J. Shin, and K. S. Yun, “Stacked Ensemble-Based Type-2 Diabetes Prediction Using Machine Learning Techniques,” *Ann. Emerg. Technol. Comput.*, vol. 7, no. 1, pp. 30–39, 2023, doi: [10.33166/AETIC.2023.01.003](https://doi.org/10.33166/AETIC.2023.01.003).
- [38] N. Singh and P. Singh, “ScienceDirect Stacking-based multi-objective evolutionary ensemble framework for prediction of diabetes mellitus,” *Integr. Med. Res.*, pp. 1–22, 2019, doi: [10.1016/j.bbe.2019.10.001](https://doi.org/10.1016/j.bbe.2019.10.001).
- [39] K. Kannadasan, D. R. Edla, and V. Kuppli, “Type 2 diabetes data classification using stacked autoencoders in deep neural networks,” *Clin. Epidemiol. Glob. Heal.*, vol. 7, no. October 2018, pp. 530–535, 2019, doi: [10.1016/j.cegh.2018.12.004](https://doi.org/10.1016/j.cegh.2018.12.004).
- [40] U. Rahmalisa, A. Febriani, and Y. Irawan, “Detector leakage gas LPG based on telegram notification using wemos D1 and MQ-6 sensor,” *J. Robot. Control*, vol. 2, no. 4, pp. 287–290, 2021, doi: [10.18196/jrc.2493](https://doi.org/10.18196/jrc.2493).
- [41] Khairunisa, Mardeni, and Y. Irawan, “Smart aquarium design using raspberry Pi and android based,” *J. Robot. Control*, vol. 2, no. 5, pp. 368–372, 2021, doi: [10.18196/jrc.25109](https://doi.org/10.18196/jrc.25109).
- [42] M. K. Anam et al., “Sara Detection on Social Media Using Deep Learning Algorithm Development,” *J. Appl. Eng. Technol. Sci.*, vol. 6, no. 1, pp. 225–237, Dec. 2024, doi: [10.37385/jaets.v6i1.5390](https://doi.org/10.37385/jaets.v6i1.5390).
- [43] A. Lubis, Y. Irawan, Junadhi, and S. Defit, “Leveraging K-Nearest Neighbors with SMOTE and Boosting Techniques for Data Imbalance and Accuracy Improvement,” *J. Appl. Data Sci.*, vol. 5, no. 4, pp. 1625–1638, 2024, doi: [10.47738/jads.v5i4.343](https://doi.org/10.47738/jads.v5i4.343).
- [44] M. K. Anam, A. Yunianta, H. J. Alyamani, Erlin, A. Zamsuri, and M. B. Firdaus, “Analysis and Identification of Non-Impact Factors in Smart City Readiness Using Technology Acceptance Analysis: a Case Study in Kampar District, Indonesia,” *J. Appl. Eng. Technol. Sci.*, vol. 5, no. 1, pp. 1–17, 2023, doi: [10.37385/jaets.v5i1.2401](https://doi.org/10.37385/jaets.v5i1.2401).
- [45] A. Febriani, R. Wahyuni, Y. Irawan, and R. Melyanti, “Improved Hybrid Machine and Deep Learning Model for Optimization of Smart Egg Incubator,” *J. Appl. Data Sci.*, vol. 5, no. 3, pp. 1052–1068, 2024, doi: <https://doi.org/10.47738/jads.v5i3.304>.
- [46] J. O. Oyoo, J. S. Wekesa, and K. O. Ogada, “Predicting Road Traffic Collisions Using a Two-Layer Ensemble Machine Learning Algorithm,” 2024, doi: <https://doi.org/10.3390/asi7020025>.
- [47] J. Niyogisubizo, L. Liao, E. Nziyumba, E. Murwanashyaka, and P. Claver, “Computers and Education: Artificial Intelligence Predicting student’s dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization,” *Comput. Educ. Artif. Intell.*, vol. 3, no. November 2021, p. 100066, 2022, doi: [10.1016/j.caeai.2022.100066](https://doi.org/10.1016/j.caeai.2022.100066).
- [48] A. Daza, J. Carlos, H. Miranda, J. Bobadilla, C. Fidel, and S. Ponce, “Predicting the depression in university students using stacking ensemble techniques over oversampling method,” *Informatics Med. Unlocked*, vol. 41, no. June, 2023, doi: [10.1016/j.imu.2023.101295](https://doi.org/10.1016/j.imu.2023.101295).
- [49] M. Hasan, P. A. Bath, C. Marincowitz, L. Sutton, R. Pilbery, and F. Hopfgartner, “Pre-hospital prediction of adverse outcomes in patients with suspected COVID-19: Development, application and comparison of machine learning and deep learning methods,” *Comput. Biol. Med. J.*, vol. 151, no. May, 2022, DOI: [10.1016/j.compbiomed.2022.106024](https://doi.org/10.1016/j.compbiomed.2022.106024).