Mobile-based Stress Level Detection using Tree-Based Machine Learning Algorithms

Kartina Diah Kesuma Wardani[1], Tony Wijaya[2], Putri Madona[3], Juni Nurma Sari[4], Yuliska[5]

{diah@pcr.ac.id[1], tonywijaya@alumni.pcr.ac.id[2], dhona@pcr.ac.id[3]}

Politeknik Caltex Riau, Jalan Umban Sari Atas No.1 Pekanbaru 28265, Indonesia[1,2,3]

Abstract. Healthcare and physiological specialists have the capability to ascertain if an individual is undergoing a state of stress or not. Automatic detection of stress can minimize the risk of health problems and improve people's well-being because by detecting stress automatically allows for early intervention and prevention of more serious health problems. It's utilize physiological signals and machine learning algorithm to automate the detection of stress levels in individuals. The current research is developing stress level prediction using the XGBoost algorithm in mobile-based stress detection based on four variables: blood pressure, heart rate per minute, body temperature, and Galvanic Skin Resistance (GSR). The mobile application receives signals from these four variables through data published using an MQTT broker from the stress detection device. Using XGBoost algorithm, the data then automatically predicts the stress level and displays the results on the mobile application interface. The model accuracy of XGBoost algorithm used in the aplication is 98% with an f1-score of 97%. Futhermore the results of usability testing of the application carried out on 35 people obtained 91% usability, 93.33% ease of use, 91% user interface, and 91% satisfaction.

Keywords: XGBoost, stress, prediction, mobile.

1 Introduction

Accumulating an excess of stress can trigger the onset of persistent illnesses like hypertension, cardiovascular disease, and even cancer. In extreme scenarios, it may even culminate in fatality. Research employs diverse biological indicators, including electroencephalography, electromyography, oxygen saturation, and pulse waves, to gauge stress levels. Nevertheless, these assessments demand sophisticated, data-intensive systems, come with substantial complexity and cost, and mandate expert analysis of the signals [1]. According to the World Health Organization (WHO), stress management systems are pivotal in identifying stress levels that can perturb one's socioeconomic well-being and way of life. Stress is a mental health problem that affects an individual's life. Stress in humans causes mental and social-fiscal issues, lack of clarity in work, poor work relationships, depression, and ultimately suicidal tendencies[2], [3]. Moreover, Stress is a contributing factor to a wide range of physical and mental health issues, including cardiovascular diseases, depression, and anxiety. Identifying stress early can lead to timely interventions to mitigate its effects. Factors Influencing Stress Levels in Health Workers [4], Stress levels are categorized into multiple degrees, specifically mild, moderate, and severe, and these are shaped by various factors unique to each person. The

ABEC 2023, September 21, Bengkalis, Indonesia

DOI 10.4108/eai.21-9-2023.2342968

research employed a stress factor questionnaire, and stress levels were assessed through the Depression Anxiety and Stress Scale (DASS) questionnaire.

Currently, The ability to ascertain if an individual is experiencing stress or not is exclusive to medical and physiological specialists. One traditional method of stress detection is through the use of questionnaires. Patients are required to respond to a set of inquiries regarding their state, following which the psychologist or psychiatrist will conduct an evaluation of the patient's condition utilizing psychological analysis methods, considering the responses provided. The automation of stress detection reduces the likelihood of health issues and enhances the overall welfare of the community. This underscores the necessity for scientific instruments that rely on physiological signals to autonomously identify stress levels in individuals. [3].

Based on the research conducted by Firman Deza, Putri Madona, and Newal Rahmardy titled "Human Stress Level Detector Based on Body Temperature, Skin Moisture, Blood Pressure, and Heart Rate," a device was created to detect stress levels based on four variables: blood pressure, heart rate per minute, body temperature, and Galvanic Skin Resistance (GSR)[5]. The device can only detect these four variables using sensors and performs manual prediction based on a predefined physical stress level table for young adults to determine an individual's stress level. The goal of creating this device is to display stress levels according to variables obtained from sensors attached to the patient's body.

Continuing previous research, the authors propose further development of their device for automatic prediction through the title "Prediction of Stress Levels Using Stress Detector Tool with Machine Learning and XGBoost Algorithm on Mobile Platform." The mobile app obtains signals from these variables via data published using an MQTT broker from the stress detection device in previous research. Subsequently, it‘s utilize the XGBoost algorithm to automatically predict the stress level based on stress pattern from stress level dataset and presents the outcomes on the mobile app interface so that stress detection can be done using a mobile device automatically. The selection of the XGBoost algorithm for this research was made due to its robustness in handling classification and regression tasks. It is versatile enough to make predictions for both categorical and continuous target variables[6].

The aplication will classify stress levels into four categories: relaxed, anxious, calm, and tense. "Tense" conveys a sense of heightened apprehension or stress, though not to an extreme level. It also represents a state of mild unease or tension, where an individual perceives that nothing significant is unfolding in their life, allowing them to relax to a certain extent. Furthermore, it encompasses a state of tranquility and comfort, enabling individuals to experience happiness and joy[7]. The application periodically presents the patient's health history on several graphs so that the patient can easily understand it.

2 Research Methods

To perform stress level prediction using machine learning, the following steps were carried out using the dataset[8][1], [9] Visible within the results derived from prior research is shows in Figure 1.

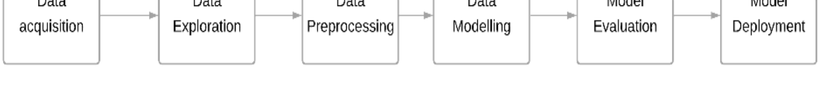

Fig.1. Research Stages

a. Data Acquisition

In this study, data was obtained through a stress detection device and recorded in PDF format. The dataset consists of 546 rows and 6 columns, which are Galvanic Skin Resistance (GSR), Heart Rate (HR), Blood Pressure (BP), Body Temperature, Air Respiration, and the Result, which represents the stress level label[1], [5], [10]. Table 1 provides an explanation of each variable.

Table 1. Variables in classification of stress levels

No	Variabel	Data Type	Value
1	Galvanic Skin Resistance (GSR)	Categorical Data	• 2-4. • 4-6, • <2, • >6.
2	Hearth Rate (HR)	Categorical Data	• 60-70, • 70-90, • 90-100, • >100.
3	Blood Preasure (BP)	Categorical Data	• 100/70-110/75, • 110-75-120/85, • 120/90-130/110, • >130/110.
4	Temperature	Categorical Data	• 33-35, • 35-36, • 36-37, • <33.
5	Air Respiration	Categorical Data	• 16-18, • 19-20, • >20
6	Stress Levels	Categorical Data	• Tenang, • Cemas, • Rileks, • Tegang

b. Data Exploration

The next step involves data exploration to better understand the data. Visualizations were used to characterize each variable. The distribution of stress level is shown in

Figure 2.

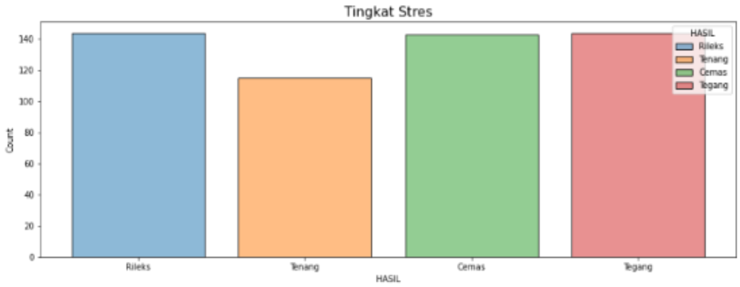

Fig.2. Distribution of stress level data

It was observed that the previous research data was quite balanced, with around 130 data points for each label, except for the "tense" label, which had around 110 data points. In this study, data exploration shows the correlation between variable GSR and Temperature to stress level in figure 3 and figure 4.

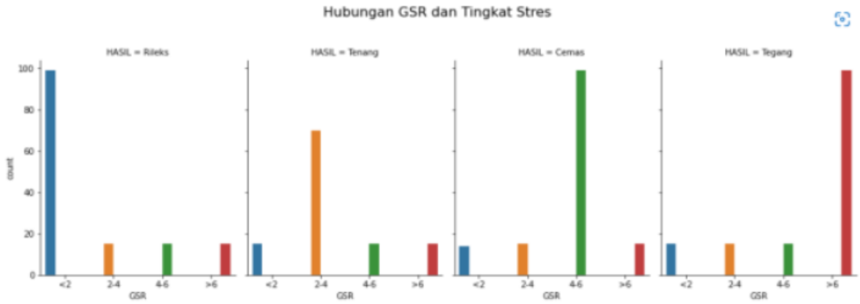

Fig. 1. Correlation between GSR variables and stress levels

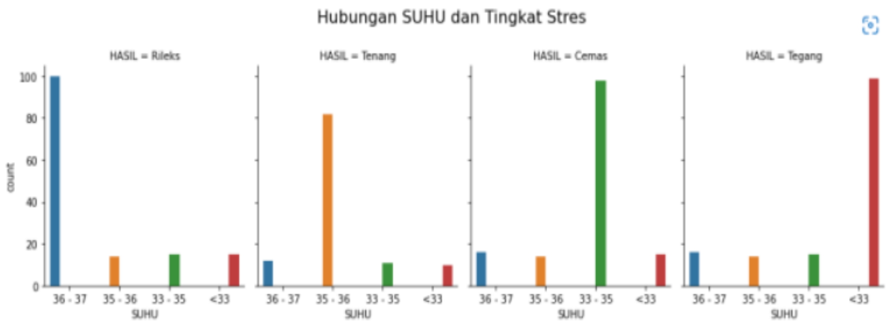

Fig. 2. Correlation between TEMPERATURES and stress levels

c. Data Preprocessing

To achieve the goal of data transformation, each variable was converted into object data types because the xgboost algorithm requires the target variable to be in numerical form to perform calculations and train the model correctly. Label encoding was used as one of the data transformation techniques, where labels were converted into numerical form. By using label encoding on the target variable, it means replacing the categorical values in the target variable with integers. This allows XGBoost to understand and process those target variables in a machine learning context.

The transformation process in this study is shown in table 2.

Table 2. Transformation process with label encoder

No	Variabel	Label	Label Encoder
1	Galvanic Skin Resistance (GSR)	2-4	0
		4-6	1
		<2	2
		>6	3
2	Hearth Rate (HR)	60-70	0
		70-90	1
		90-100	2
		>100	3
3	Blood Preasure (BP)	100/70 – 110/75	0
		110/75 -120/85	1
		120/90 -130/110	2
		>130/110	3
4	Temperature	33-35	0
		35-36	1
		36-37	2
		<33	3
5	Air Respiration	16-18	0
		19-20	1
		>20	2

d. Data Modelling

The next step involves data modeling. To help solve the prediction problem, modeling will be performed using the existing data. In this study data modelling using machine learning algorithms extreme gradient boosting (XGBoost). The XGBoost technique is an evolution of the GBDT (Gradient Boosting Decision Tree) algorithm, originally introduced by Friedman. XGBoost is a Supervised Learning library utilized for making predictions and conducting classifications.[11]–[13]. The decision to use the XGBoost algorithm in this study was based on its strong performance in managing classification and regression tasks. XGboost algorithm is equipped with various powerful optimization techniques, such as tree pruning, loss reduction, and feature clustering, which makes it very efficient in producing good models. Furthermore, this algorithm also could overcome overfitting and underfitting problems and is able to handle large data well.

e. Data Evaluation

To determine the data training and testing proportions, four common ratios were compared: 80:20, 75:25, 70:30, and 67:33. Utilizing a balanced ratio can mitigate the risk of the resulting model becoming excessively intricate and prone to overfitting the training data, which could lead to suboptimal performance when applied to unseen data. Evaluation model in this study using confusion matrix. The Confusion Matrix provides a more intricate breakdown of accurate and inaccurate classifications for each class. In this matrix, the rows correspond to the actual (ground truth) labels, while the columns depict the predicted labels. [14].

f. Model Deployment

The next step involved model deployment. To use the machine learning model in production, modifications were made to enable prediction with other tools, such as mobile devices. For the XGBoost model used in this research, it was transformed into JSON format. Architectural design used in this study is shown in figure 5.

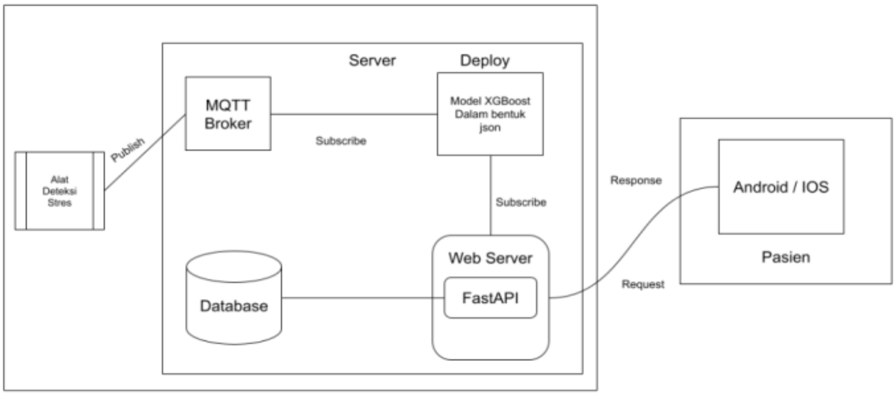

Fig. 5. Android-based stress detection application architecture design

The study began with published data obtained using an MQTT broker and a stress detection tool. The data underwent stress level prediction using the XGBoost algorithm based on the published data. After successful prediction, the data was automatically subscribed and sent to a database through a web server. The research also focused on developing a mobile application that serves as a bridge to display patient data. The REST API was utilized to facilitate communication between the mobile application and the data transmitted through the tool to the database.

3 Result and Discussion

3.1 Interface of the Aplication

The application was built using the Dart programming language with the Flutter framework.

Fig.3. Stress level detector application

The users in this application are patients who can view their recorded data in the form of a dashboard, history, and profile. Figure 8 is user interface in mobile application.

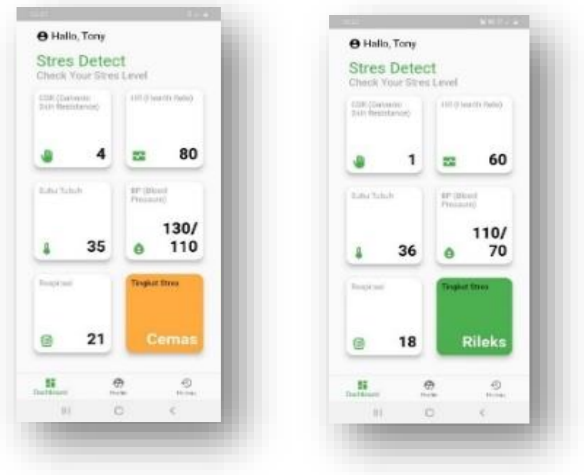

Fig.8. Stress level detector application dashboard

The dashboard displays the latest information from the patient's checkups in card format, making it easier for the patient to understand the data. Visualization of user’s information test shown in figure 9.
The profile page shows patient information, a logout button to log out of the application, a reset password button leading to the password reset page, and a tab bar displaying graphs of the patient's checkups by day, month, and year. User interface for history pages shown in Figure 10.

Fig.9. Display menu profile page

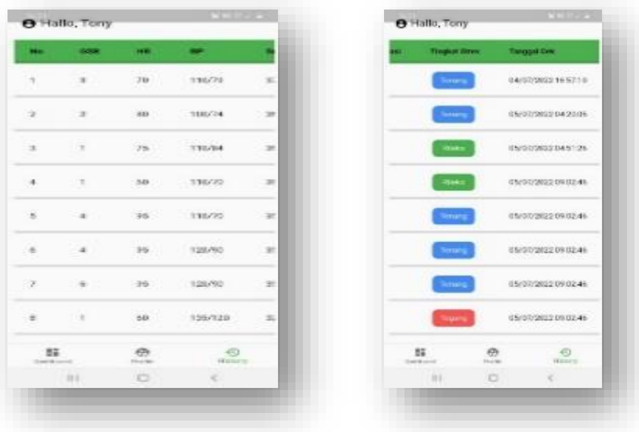

Fig.10. Display of history pages

The history page shows the patient's historical data in the form of a datatable, displaying all the checkup results conducted by the patient while using the tool.

3.2. XGBoost Classification Model

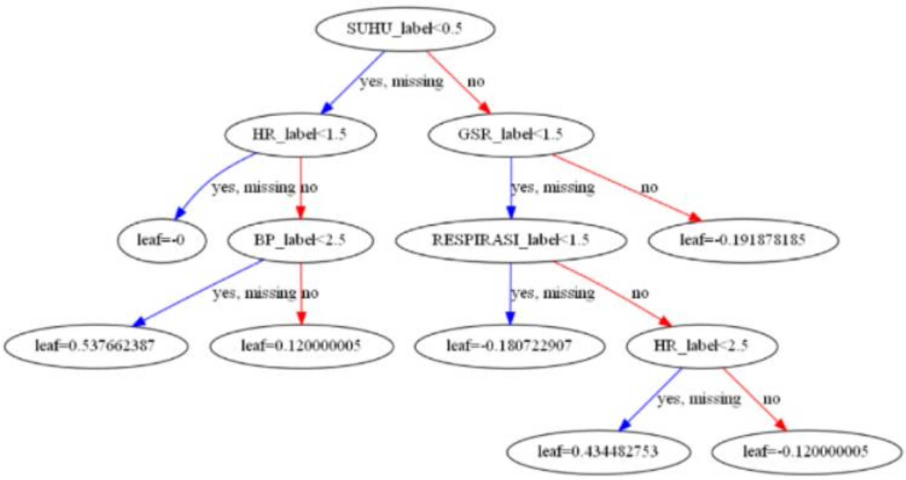

Fig.10. XGBoost Model

Tree model based on Xgboost shown in figure 10. From figure 10 known that temperature attribute is a root. This is due to the fact that temperature is the primary factor that has the most significant influence on stress. Followed by HR and GSR.

3.3. Testing

Testing algorithm accuracy in this study using proportioning comparation between training data and testing data. The highest accuracy of 98.46% was achieved using the 67:33 proportion. Detailed result is shown in table 3.

Table 3. Proportion of training data and testing data to accuracy results

Proportion of training and testing data	Algorithm accuracy
67:33	98.46%
75:25	97.96%
70:30	97.96%
80:20	97.44%

Through testing 65 recorded data results, the Confusion Matrix table was obtained as follows:

		Predicted Class			
		Rileks	Cemas	Tenang	Tegang
Actual Class	Rileks	19	0	0	0
	Cemas	0	17	0	1
	Tenang	0	0	15	0
	Tegang	1	0	0	13

Based on the results of the confusion matrix for stress level detection using the XGBoost algorithm, the precision score reached 98%, and the recall score was also 98%. This indicates a very good agreement between the actual values and the prediction results. The f1-score obtained was 97%, indicating excellent performance in stress level prediction, considering Thera are 4 prediction labels and 5 determinant variables to determine the stress level.

4 Conclusion

In conclusion, the XGBoost algorithm can model data for stress level prediction in a mobile-based stress level detection application. The accuracy of the XGBoost algorithm for stress level modeling in this research was 98%, and the f1-score was 97% using 5 supporting variables namely Galvanic Skin Resistance (GSR), Heart Rate (HR), Blood Pressure (BP), Temperature and Respiration.

References

[1] M. Kang, S. Shin, J. Jung, and Y. Kim, "Classification of Mental Stres Using CNN-LSTM Algorithms with Electrocardiogram Signals.," *J. Healthc. Eng.*, 2021.

[2] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on Psychological Stress Detection Using Biosignals," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 440–460, 2022, doi: 10.1109/TAFFC.2019.2927337.

[3] S. Sriramprakash, V. D. Prasanna, and O. V. R. Murthy, "Stress Detection in Working People," *Procedia Comput. Sci.*, vol. 115, pp. 359–366, 2017, doi: 10.1016/j.procs.2017.09.090.

[4] R. S. Tiara Dewi, Muhammad Amir Masruhim, "Faktor-Faktor Yang Mempengaruhi Tingkatan Stres Pada Tenaga Kesehatan Di Rumah Sakit Universitas Tanjungpura Pontianak Tahun 2015," *Lab. Penelit. dan Pengemb. FARMAKA Trop. Fak. Farm. Univ. Mualawarman, Samarinda, Kalimantan Timur*, no. April, pp. 5–24, 2016.

[5] F. Deza, P. Madona, and N. Rahmardy, "Alat Pendeteksi Tingkat Stress Manusia Berdasarkan Suhu Tubuh, Kelembaban Kulit, Tekanan Darah dan Detak Jantung (Human Stress Level Detection Tool Based on Body Temperature, Skin Moisture, Blood Pressure and Heart Rate).," *J. Elektro dan Mesin Terap.*, vol. 3, no. 2, pp. 31–42, 2017.

[6] M. Li, X. Fu, and D. Li, "Diabetes Prediction Based on XGBoost Algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 768, no. 7, 2020, doi: 10.1088/1757-899X/768/7/072093.

[7] E. Suwarto, "Alat Pendeteksi Parameter Stres Manusia Berbasis Mikrokontroler ATMega 16," *Politek. Negeri Semarang.*, 2012.

[8] A. Roihan, P. Sunarya, and A. Rafika, "Pemanfaatan Supervised Learning dalam Berbagai Bidang: Review paper," *Indones. J. Comput. Inf. Technol.*, 2020.

[9] P. R. Sihombing and I. F. Yuliati, "Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 2, pp. 417–426, 2021, doi:

10.30812/matrik.v20i2.1174.
[10] K. Wiradharma, H. Mukhtar, and W. A. Cahyadi, "Sensor Galvanic Skin Response (Gsr) Berbasis Arduino Nano Sebagai Pengukur Perubahan Konduktansi Listrik Kulit Dalam Kondisi Tenang Atau Distress," *J. e-Proceeding Eng.*, vol. 9, no. 2, p. 219, 2022.
[11] K. D. K. Wardhani and M. Akbar, "Diabetes Risk Prediction Using Extreme Gradient Boosting (XGBoost)," *J. Online Inform. 7(2), 244-250.*, vol. Vol 7.No 2, 2022, doi: 10.15575/join.v7i2.970.
[12] H. Mo, H. Sun, J. Liu, and S. Wei, "Developing window behavior models for residential buildings using XGBoost algorithm," *Energy Build.*, vol. 205, pp. 1–23, 2019, doi: 10.1016/j.enbuild.2019.109564.
[13] N. N. Pandika Pinata, I. M. Sukarsa, and N. K. Dwi Rusjayanthi, "Prediksi Kecelakaan Lalu Lintas di Bali dengan XGBoost pada Python," *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, vol. 8, no. 3, p. 188, 2020, doi: 10.24843/jim.2020.v08.i03.p04.
[14] M. Heydarian, T. E. Doyle, and R. Samavi, "MLCM: Multi-Label Confusion Matrix," *IEEE Access*, vol. 10, pp. 19083–19095, 2022, doi: 10.1109/ACCESS.2022.3151048.