Explainable Artificial Intelligence (XAI) for Identification of Using Obesity Factors Hybrid Artificial Neural Network Approach and SHapley Additive exPlanations

Esti Yogiyanti[1], Yuni Yamasari[1], Ervin Yohannes[1]

[1] Teknik Informatika, Universitas Negeri Surabaya
[1]estiyogiyanti187@gmail.com
[2]yuniyamasari@unesa.ac.id
[3]ervinyohannes@unesa.ac.id

***Abstract*— This study aims to develop and evaluate an obesity classification model using an Artificial Neural Network (ANN) combined with Explainable Artificial Intelligence (XAI) techniques based on SHAP (SHapley Additive exPlanations). The model was trained and tested using two different optimizers, Adaptive Moment Estimation (Adam) and Stochastic Gradient Descent (SGD), across multiple train-test ratios and epoch variations. The experimental results indicate that the Adam optimizer consistently outperformed SGD in terms of accuracy, loss value, and stability of evaluation metrics. The best performance was achieved with a 90:10 train-test ratio at 100 epochs, yielding an accuracy of 94.74%, a loss of 0.1899, precision, recall, and an f1-score of 0.95. To improve interpretability, SHAP was applied to identify the most influential features in the classification process. The analysis revealed that features such as Weight, Height, Gender, and Age significantly contribute to the model's predictions. Based on the SHAP interpretation, feature selection was conducted using the top nine features with the highest SHAP values. Retraining the ANN with these selected features resulted in improved performance, achieving 98.56% accuracy, a loss of 0.0638, and a precision, recall, and F1-score of 0.99 . These findings demonstrate that integrating XAI with ANN not only enhances transparency and interpretability but also boosts classification performance and computational efficiency. This approach shows strong potential for supporting decision-making in healthcare, particularly for early detection and intervention in cases related to obesity.**

***Kata Kunci*— Artificial Neural Network, Machine Learning, SHAP, Explainable AI, Classification of Obesity.**

I. Introduction

The rapid development of technology has driven its widespread use across various sectors, thanks to its significant positive impact on efficiency, cost-effectiveness, operations, and time management, particularly in data processing and analysis, on both small and large scales. One of the most popular and rapidly advancing technologies is Artificial Intelligence (AI), which holds great potential for accurate and effective data classification and prediction processes.

One area that requires special attention is the healthcare sector. In this field, one pressing issue that continues to rise and requires complex intervention is obesity. Obesity is a medical condition that demands serious attention as it can lead to chronic diseases such as diabetes, heart disease, hypertension, and others. According to records from the World Health Organization (WHO) over the past several years, in 2016, approximately 39% of adults aged 18 and over were overweight, with 13% classified as obese. By 2022, an estimated 650 million adults, 340 million adolescents, and 39 million children will be affected by obesity. If the trend continues, by 2025, around 167 million people are projected to be overweight. Based on these statistics and WHO records, researchers predict that by 2030, obesity cases may reach 2.16 billion. The American Association of Clinical Endocrinologists has also stated that being overweight is an early stage of obesity and should be addressed early on [1][2].

With the increasing prevalence of obesity, many studies have leveraged algorithms to build obesity classification models. Some studies have also combined these with optimization methods such as Particle Swarm Optimization (PSO) or Explainable AI (XAI) methods like SHAP. These combinations have successfully produced high-accuracy models, but often lack in-depth exploration of the parameters influencing model performance [3][4][5].

Therefore, this study aims to develop an obesity classification model using the Artificial Neural Network (ANN) algorithm, evaluated and interpreted using the SHAP method from XAI. The model development will include parameter exploration to achieve optimal performance. This research is expected to contribute to the advancement of decision-support systems in the healthcare sector, particularly in identifying risk factors for obesity as part of early intervention and prevention efforts.

II. Literature Review

A. *Artificial Inteligence*

Artificial Intelligence (AI) refers to the development of computer systems capable of performing tasks that mimic or simulate human intelligence. AI is designed to carry out tasks such as learning, reasoning, pattern recognition, problem-solving, decision-making, and more. With basic program training, AI technology can perform complex tasks effectively and efficiently without depending on human instructions. This makes AI a widely used technology in various fields such as healthcare, education, industry, and many others [6]

B. *Machine Learning*

Machine Learning was first developed in the 1950s. It consists of three main subcategories: supervised learning, unsupervised learning, and reinforcement learning [7]. Machine learning is a branch of AI focused on developing algorithms or models to improve performance in problem-solving tasks.

C. Obesity

Obesity is a medical condition characterized by excessive fat accumulation in the body, which increases the risk of health problems. It is influenced by various factors such as poor diet, consumption of junk food, lack of physical activity, economic factors, and more. Therefore, obesity is considered a multifactorial condition [8].

D. Artificial Neural Network

Artificial Neural Network (ANN) is a machine learning algorithm inspired by the workings of the human brain's neural networks. It is used for tasks such as classification, regression, and pattern recognition [9]. ANN models consist of several processing units called neurons, which correspond to the number of input features used in model training. Each neuron is connected to layers such as the hidden and output layers with specific weights [10]. With its structure and effective mechanism, ANN is capable of processing non-linear data effectively to solve classification problems.

E. Shapley Additive exPlanations

SHAP is one of the methods under the Explainable AI (XAI) approach, designed to provide transparent and easily interpretable model decisions. This method uses Shapley values to calculate the contribution of each feature to the model's output. SHAP is widely used in model training due to its advantages in interpreting feature importance and its impact on the overall model [11].

F. Hybrid Model

Generally, a hybrid model refers to a method that combines two or more different algorithms or techniques within a single research framework to improve system performance. In the context of machine learning and artificial intelligence, it integrates predictive methods, explanation methods, or optimization techniques. By combining various approaches, hybrid models can produce stronger, more adaptive, and suitable solutions for problems that cannot be optimally solved using a single method.

G. Explanation Artificial Intelligence

Around the mid-2000s, the term Explainable AI (XAI) began to be widely introduced to develop models for small-scale systems. XAI is an approach in the development of artificial intelligence systems that aims to produce decisions that are clear, transparent, and easy to understand [12]. As AI technology continues to advance and be implemented across various fields, XAI is increasingly utilized to ensure results are interpretable and comprehensible to users.

H. Epoch

In ANN model training, the term epoch is a crucial concept. An epoch refers to one complete cycle through the entire training dataset during which the model's weights are updated [10]. The greater the number of epochs, the more the model learns from the training data. However, an excessive number of epochs can lead to overfitting, where the model becomes too tailored to the training data and loses its ability to generalize. Therefore, selecting the right number of epochs is essential for maintaining the model's performance and balance.

III. RESEARCH METHOD

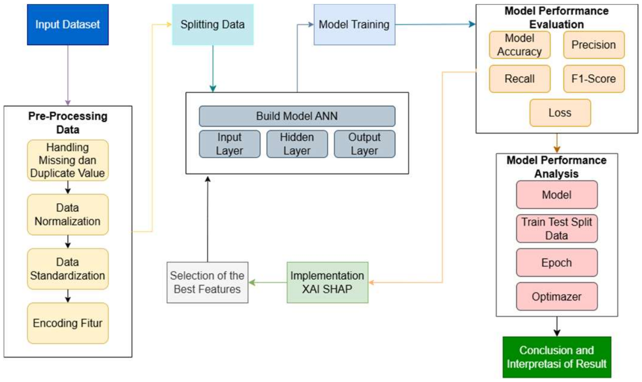

Fig. 1 Research Flow Diagram

A. Input Dataset

The dataset used in this research was obtained from Kaggle under the title "Obesity Levels & Lifestyle". The dataset consists of 2,111 records with 17 features as described in Table 1 covering demographic, anthropometric, dietary, lifestyle data, and obesity level classification labels [13]. For the purpose of this study, 16 independent features were selected for model testing.

TABLE I
DATASET FEATURES

Factor	Questions	Possible Answers
Independent Variable		
Gender	What is your gender?	Female Male
Age	what is your age?	(Tahun)
Height	what is your height?	(Meter)
Weight	what is your weight?	(Kg)
Family History with Overweight	Has a family member suffered or suffers from overweight?	Yes No
FAVC (Frequent consumption of high caloric food)	Do you eat high caloric food frequently?	Yes No

FCVF (Frequency of consumption of vegetables)	Do you usually eat vegetables in your meals?	1 (Tidak pernah) 2 (Kadang-kadang) 3 (Selalu)
NCP (Number of main meals)	How many main meals do you have daily?	1 (Antara 1 - 2) 3 (Tiga) 4 (Lebih dari 3)
CAEC (Consumption of food between meals)	Do you eat any food between meals?	No Sometimes Frequently Always
Smoke	Do you smoke?	Yes No
CH2O (Consumption of water daily)	How much water do you drink daily?	1 (Kurang dari satu liter) 2 (Antara 1-2 Liter) 3 (Lebih dari 2 Liter)
SCC (Calories consumption monitoring)	Do you monitor the calories you eat daily?	Yes No
FAF (Physical activity frequency)	How often do you have physical activity?	0 (Saya tidak memiliki) 1 (1/2 hari) 2 (2/4 hari) 3 (4/5 hari)
TUE (Time using technology devices)	How much time do you use technological devices such as cell phone, video games, television, computer, and others?	0 (0-2 jam) 1 (3-5 jam) 2 (lebih dari 5 jam)
CALC (Consumption of alcohol)	how often do you drink alcohol?	No Sometimes, Frequently Always
MTRANS (Transportation used)	Which transportation do you usually use?	Automobile Bike Motorbike Public Transportation Walking
Dependent Variable		
NObeyesdad	-	Insufficient Weight Normal Weight Overweight Level I Overweight Level II Obesity Type I Obesity Type II Obesity Type III

B. Pre-Processing Data

Data pre-processing not only improves data quality and leads to better decision outcomes, but also helps reduce testing time [14]. The pre-processing stages performed in this study include handling missing and duplicate values, normalization, standardization, and feature encoding.

Missing values refer to data entries that are not recorded, while duplicate values refer to identical data entries that appear more than once in the dataset. Handling missing and duplicate values is a critical initial step to ensure that the training data is clean, representative, free of noise, and non-redundant. This is essential to prevent errors or biased outcomes during model training [15].

Normalization is the process of transforming feature values to the same scale to prevent certain features from dominating the model during training [15].

Standardization involves scaling input data such that it has a mean of 0 and a standard deviation of 1. This ensures that the features in the dataset are on a balanced scale [14].

Encoding aims to convert categorical features into numerical format so they can be processed by machine learning algorithms. This step is crucial to systematically utilize all categorical feature information without losing its original meaning and to avoid errors during model training [16].

C. Splitting Data

The dataset is divided into two parts: training data and testing data. Data splitting helps the model to generalize better, leading to more accurate results [17]. Various data split ratios are used to evaluate and compare the model's performance, allowing identification of the most optimal split for classification accuracy. Additionally, a validation split of 10% from the training data is applied to prevent overfitting before the final evaluation of the testing data.

D. Build Model ANN

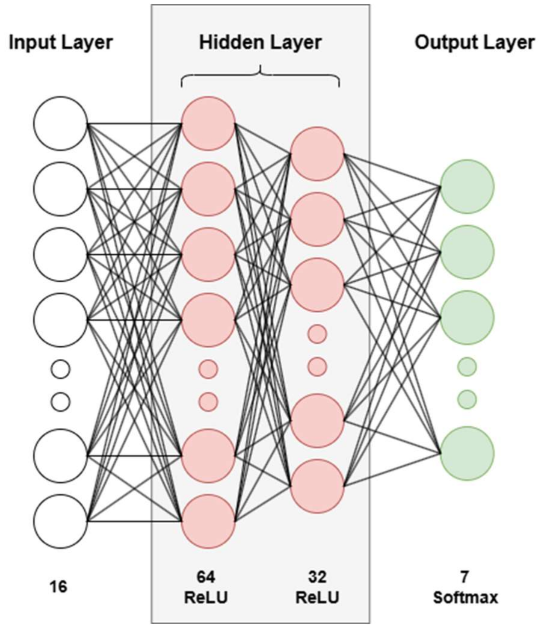

Fig. 2 ANN Model Architecture

The architecture of the model, as illustrated in Figure 2, consists of three layers, which are described as follows:

1) *Input Layer*

This layer is responsible for receiving input from outside the system. The number of neurons in the input layer corresponds to the number of features used as inputs in this study. Each neuron in the input layer represents one feature in the dataset [10]. This study uses 16 features; therefore, the input layer of the ANN model contains 16 neurons.

2) *Hidden Layer*

The hidden layers are intermediate layers positioned between the input and output layers. These layers receive input from the previous layer and transform the data through weights and activation functions [10]. This ANN model includes two hidden layers as follows:

- Hidden Layer 1
  Number of Neurons: 64
  Function: Extracts patterns and complex relationships from the input layer.

- Hidden Layer 2
  Number of Neurons: 32
  Function: Reinforces and enhances the representation of important features learned from the previous layer.

The ReLU (Rectified Linear Unit) activation function is used to determine whether the neuron output is linear or non-linear. ReLU sets all negative input values to zero [10][18]. The ReLU function is mathematically defined as follows:

$$ReLU\ (x) = \max(0, x)$$

atau

$$ReLU\ (x) = \begin{cases} 0\ untuk\ x \leq 0 \\ x\ untuk\ x > 0 \end{cases}$$

3) *Output Layer*

This is the final layer of the ANN model that produces the output [10]. The number of neurons in this layer equals the number of target classes in the dataset. Since this study classifies obesity into 7 categories, the output layer consists of 7 neurons. The softmax activation function is applied to convert the output into a probability distribution across the classes. The softmax function is defined as follows:

$$Softmax\ (net_k)_k = \frac{e^{net_k}}{\sum_j e^{net_j}}$$

E. *Training Model*

The model is trained using various parameters, including epoch values, data splitting ratios, and optimizers.

1) *Epochs*

The training process uses 50, 60, 70, 80, and 100 epochs. Varying the number of epochs helps the model gradually learn and improve its understanding of the data.

2) *Data Splitting*

Several data splitting ratios are applied in this study 70:30 (70% training, 30% testing), 80:20 (80% training, 20% testing), 90:10 (90% training, 10% testing). These different ratios are used to evaluate and compare model performance under various training conditions.

3) *Optimizers*

Optimizers are used to update model weights during training in order to minimize loss and improve performance. This study employs four optimizers: Adam and SGD (Stochastic Gradient Descent).

F. *Implementation of XAI: SHAP*

Shapley Additive exPlanations (SHAP) is an Explainable Artificial Intelligence (XAI) method used to measure the contribution of each feature to the model's prediction. The application of XAI is highly beneficial for understanding the results of the study, as each feature is assigned a contribution value (Shapley value). The SHAP visualization used in this study is the feature importance plot, which helps researchers identify the most influential features in the model's decision-making process.

G. *Feature Selection*

After conducting multiple tests and obtaining feature contribution values from SHAP, feature analysis is carried out to select the most significant features influencing the model's decisions. Further testing is performed using the top 5–10 selected features. The purposes of feature selection are as follows:

- To reduce model complexity.
- To minimize noise from irrelevant features.
- To reduce training time.

H. *Model Performance Evaluation*

Evaluating model performance is a crucial step in building a machine learning model. It aims to assess how well the model performs. The performance is evaluated using several metrics, including:

1) *Accuracy*

Accuracy is a basic metric that measures the proportion of correct predictions out of the total predictions. It provides a general view of the model's performance. The higher the accuracy value, the better the model's performance [19].

$$Accuracy = \frac{Number\ of\ Correct\ Classification}{Total\ Number\ of\ Classification}$$

2) Precision

Precision is the ratio of true positive predictions to the total number of positive predictions made by the model. A high precision score indicates that the model makes accurate positive predictions [19].

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

3) Recall

Recall measures the model’s ability to correctly identify positive instances. A high recall value shows that the model successfully detects most obesity cases [19].

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negative}$$

4) F1-Score

The F1-score is the harmonic mean of precision and recall. It is useful when a balance between these two metrics is required. F1-score is ideal for evaluating the effectiveness of multi-class classification in obesity levels. The closer the F1-score is to 1, the better the balance between precision and recall [19].

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

5) Loss Value

Loss value measures how far the model’s predictions deviate from the actual values. A lower loss value indicates a better learning process. Loss is calculated for each epoch to detect overfitting during testing.

I. Model Performance Analysis

This step involves analyzing and comparing the results of model training conducted across various experiments. The analysis includes the following aspects:

- Best model performance based on training and testing data ratios.
- Best optimizer.
- Optimal number of epochs.
- Impact of feature selection on model performance.

J. Conclusion and Result Interpretation

The final step is to draw conclusions and interpret the outcomes of the entire research process, including input data, preprocessing, model development, performance evaluation, model analysis, and XAI-based interpretation. The conclusions of this study include:

- Effectiveness of the Artificial Neural Network (ANN) model.
- Influence of features on model performance.
- Best training parameters.
- Research implications in the healthcare domain.
- Result interpretation using SHAP.

IV. Result and Discussion

A. SGD Optimizer Experimental Result

Model training using the SGD (Stochastic Gradient Descent) optimizer demonstrated competitive performance, particularly when applied with a 90:10 training-to-testing data ratio. In this scenario, the training results of the Artificial Neural Network (ANN) model, as presented in Table 2, show the highest accuracy of 89.95% at epoch 100, along with the lowest loss value of 0.2292. The model's performance is further supported by other evaluation metrics, namely precision, recall, and F1-score, all of which reached 0.90, indicating balanced classification capability. Notably, this experiment revealed a consistent downward trend in loss values from the beginning to the end of the epochs, suggesting that the learning process was stable and effective throughout the training phase.

Table III
SGD Experimental Result

Epoch	Model Accuracy	Loss Accuracy	Precision	Recall	F1-Score
Train (70) – Test (30)					
50	0.8054	0.5162	0.81	0.81	0.80
60	0.8421	0.4494	0.84	0.84	0.84
70	0.8581	0.4031	0.86	0.86	0.86
80	0.8469	0.3897	0.85	0.85	0.85
90	0.9011	0.3050	0.91	0.90	0.90
100	0.8947	0.3138	0.89	0.89	0.89
Train (80) – Test (20)					
50	0.8493	0.4443	0.86	0.85	0.85
60	0.8612	0.4220	0.86	0.86	0.86
70	0.8565	0.3859	0.86	0.86	0.85
80	0.8780	0.3316	0.88	0.88	0.88
90	0.8900	0.3196	0.89	0.89	0.89
100	0.8971	0.2764	0.90	0.90	0.90
Train (90) – Test (10)					
50	0.8517	0.3764	0.86	0.85	0.85
60	0.8660	0.3918	0.87	0.87	0.87
70	0.8947	0.2971	0.89	0.89	0.89
80	0.8804	0.3028	0.88	0.88	0.88
90	0.8947	0.2519	0.90	0.89	0.89
100	0.8995	0.2292	0.90	0.90	0.90

B. Adam Optimizer Experimental Result

The Artificial Neural Network (ANN) model was evaluated using the Adam optimizer (Adaptive Moment Estimation) across three training-to-testing data ratio scenarios. Based on the training results presented in Table 3, the 90:10 train-test ratio yielded the best performance compared to the other configurations. The highest accuracy of 94.74% was achieved at epoch 90, with a corresponding loss value of 0.1899.

Additionally, the evaluation metrics precision, recall, and F1-score all reached a stable value of 0.95, indicating excellent classification performance and strong generalization to the test data.

The use of the Adam optimizer resulted in consistent and efficient training performance, as evidenced by the decreasing loss values and stable accuracy across scenarios. While accuracy is an important metric for model evaluation, test loss serves as a more critical indicator, as it reflects the model's ability to generalize to unseen data. The low and stable loss values suggest that the model did not experience overfitting and was able to perform well on the test set. Based on these results, it can be concluded that the ANN model trained with the Adam optimizer using 90:10 data split at epoch 90 represents the optimal configuration in this study.

TABLE IIIII
ADAM EXPERIMENTAL RESULT

Epoch	Model Accuracy	Loss Accuracy	Precision	Recall	F1-Score
Train (70) – Test (30)					
50	0.9171	0.2452	0.92	0.92	0.92
60	0.9282	0.2325	0.93	0.93	0.93
70	0.9250	0.2698	0.93	0.93	0.93
80	0.9219	0.2666	0.92	0.92	0.92
90	0.9266	0.2981	0.93	0.93	0.93
100	0.9298	0.2331	0.93	0.93	0.93
Train (80) – Test (20)					
50	0.9211	0.2288	0.92	0.92	0.92
60	0.9234	0.1899	0.92	0.92	0.92
70	0.9139	0.2349	0.91	0.91	0.91
80	0.9234	0.2936	0.92	0.92	0.92
90	0.9330	0.2440	0.93	0.93	0.93
100	0.9258	0.2356	0.93	0.93	0.93
Train (90) – Test (10)					
50	0.9139	0.2079	0.91	0.91	0.91
60	0.9282	0.1965	0.93	0.93	0.93
70	0.9187	0.1927	0.93	0.92	0.92
80	0.9426	0.1588	0.94	0.94	0.94
90	0.9474	0.1899	0.95	0.95	0.95
100	0.9426	0.1520	0.94	0.94	0.94

C. Model Performance Evaluation

Based on the experimental results, a clear difference can be observed in the performance of the Artificial Neural Network (ANN) model when using two types of optimizers Adam and SGD. In general, the Adam optimizer demonstrated superior, more consistent, and more accurate performance compared to SGD across various training-testing data ratios and epoch settings. Adam consistently produced lower loss values, indicating its effectiveness in minimizing errors during training. This advantage is due to Adam's combination of momentum and an adaptive learning rate, which enables faster and more stable convergence compared to SGD's fixed gradient approach.

From these tests, it can be concluded that Adam is more effective for obesity data classification using ANN, in terms of accuracy, loss reduction, and the stability of evaluation metrics. However, SGD can still serve as a reliable alternative when using the proper configuration, a sufficient number of epochs, and a larger training data proportion. As a next step, further testing will focus on the best-performing configuration, which is the Adam optimizer with a 90:10 training-testing ratio at epoch 90, as this setup yielded the highest classification performance in this study.

D. SHAP Value Interpretation

The SHAP (SHapley Additive exPlanations) method was used as an Explainable Artificial Intelligence (XAI) approach to interpret the contribution of each feature in the obesity classification predictions made by the Artificial Neural Network (ANN) model trained using the Adam optimizer. SHAP is based on Shapley value theory, where each feature is assigned a contribution value to the model's output. This approach helps identify which features have the most significant influence on the classification results. By applying SHAP, the model's decision-making process becomes more transparent and easier to understand, making it useful for both medical applications and health policy planning.

Table 4 presents the SHAP interpretation results, including the shap values and feature rankings. Each testing parameter combination produced different feature orders and contribution values. Based on the evaluation results, the best configuration was found at epoch 90. Therefore, further testing will use the selected features according to the SHAP interpretation results from this configuration.

TABEL IVV
INTERPRETATION SHAP

Fitur	Epoch 50	Epoch 60	Epoch 70	Epoch 80	Epoch 90	Epoch 100
Gender	3	3	3	3	3	3
	0.0423	0.0408	0.0468	0.0475	0.0431	0.0431
Age	4	4	4	4	4	4
	0.0231	0.0253	0.0240	0.0210	0.0246	0.0201
Height	2	2	2	2	2	2
	0.0539	0.0585	0.0591	0.0549	0.0562	0.0576
Weight	1	1	1	1	1	1
	0.1701	0.1707	0.1712	0.1731	0.1702	0.1748
Family History	8	11	14	13	14	12
	0.0095	0.0084	0.0060	0.0069	0.0049	0.0085
FAVC	12	10	11	10	9	7
	0.0088	0.0086	0.0082	0.0081	0.0089	0.0098
FCVC	5	5	6	8	6	10
	0.0175	0.0138	0.0113	0.0083	0.0115	0.0091
NCP	9	9	13	9	13	11
	0.0091	0.0094	0.0068	0.0083	0.0066	0.0088
CAEC	13	13	12	14	10	14

	0.0083	0.0075	0.0073	0.0069	0.0089	0.0070
Smoke	15	16	16	16	16	16
	0.0027	0.0022	0.0014	0.0021	0.0026	0.0025
CH2O	11	14	10	6	7	9
	0.0089	0.0075	0.0088	0.0106	0.0104	0.0091
SCC	16	15	15	15	15	15
	0.0024	0.0028	0.0032	0.0027	0.0032	0.0028
FAF	14	12	9	11	8	13
	0.0068	0.0082	0.0089	0.0078	0.0095	0.0074
TUE	10	8	7	12	12	8
	0.0091	0.0095	0.0095	0.0069	0.0084	0.0095
CALC	6	6	5	5	5	5
	0.0136	0.0136	0.0117	0.0121	0.0124	0.0122
MTRA NS	7	7	8	7	11	6
	0.0107	0.0104	0.0093	0.0097	0.0086	0.0115

To enhance clarity, a feature importance visualization is also included in Figure 3.

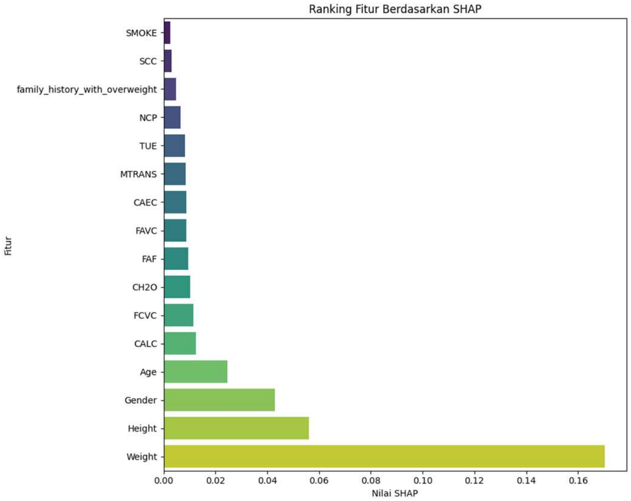

Fig. 3 SHAP Feature Importance

E. Retraining the Feature Selection Model

After interpreting the model using the SHAP approach, the model was re-evaluated using the top five to ten features with the highest contribution values to assess the performance of the Artificial Neural Network (ANN) more efficiently. The results presented in Table 5 show that using the top 9 features yielded the best performance, with an accuracy of 0.9856, the lowest loss value of 0.0638, and precision, recall, and F1-score all at 0.99. This indicates that the model can classify obesity cases very effectively, even without utilizing all available features.

TABLE V
MODEL FEATURE SELECTION MODELS

Number of Features	Model Accuracy	Loss Accuracy	Precision	Recall	F1-Score
5 Features	0.9761	0.1169	0.98	0.98	0.98
6 Features	0.9809	0.0697	0.98	0.98	0.98
7 Features	0.9713	0.0905	0.97	0.97	0.97
8 Features	0.9809	0.0876	0.98	0.98	0.98
9 Features	0.9856	0.0638	0.99	0.99	0.99
10 Features	0.9713	0.1198	0.97	0.97	0.97

F. Model Performance Analysis

The performance analysis of the model before and after the feature selection process in Table 6 shows a significant improvement across various evaluation metrics. When using all 16 features, the Artificial Neural Network (ANN) model achieved an accuracy of 94.74%, a loss value of 0.1899, and precision, recall, and F1-score of 0.95. Although this performance is already considered good, the re-evaluation after applying feature selection using the SHAP method demonstrated that the model's performance could be further enhanced. By using only the top 9 features based on SHAP contributions, the model's accuracy increased to 98.56%, while the loss drastically decreased to 0.0638. Precision, recall, and F1-score also improved to 0.99.

This improvement indicates that the selected features sufficiently represent the most important information in the dataset and help reduce the risk of overfitting that might be caused by less relevant features. Additionally, using fewer features contributes to computational efficiency, making the model lighter and faster during both training and inference. Therefore, the SHAP-based feature selection approach not only enhances the model's interpretability but also optimizes its overall performance.

TABLE VI
MODEL COMPARISON RESULT

Information	Model Accuracy	Loss Accuracy	Precision	Recall	F1-Score
All Features (16)	0.9474	0.1899	0.95	0.95	0.95
After Feature Selection (9)	0.9856	0.0638	0.99	0.99	0.99

V. CONCLUSION

This study aims to develop and analyze an obesity classification model using the Artificial Neural Network (ANN) approach, combined with Explainable Artificial Intelligence (XAI) through SHAP (SHapley Additive exPlanations) for feature interpretation. The training process involved comparing two optimizers Adam and SGD and evaluating the model across various train-test ratios and numbers of epochs.

The results show that the Adam optimizer outperformed SGD, particularly in the training scenario with a 90:10 train-test ratio and 100 epochs. In this configuration, the model achieved an accuracy of 94.74%, a loss value of 0.1899, and precision, recall, and F1-score of 0.95. SHAP-based model interpretation revealed that features such as Weight, Height, Gender, and Age were dominant factors in obesity classification. Based on this analysis, feature selection was performed by retaining the top 9 contributing features. The re-evaluation of the model showed a significant improvement in performance, with an accuracy of 98.56%, a reduced loss of 0.0638, and precision, recall, and F1-score all reaching 0.99. These findings demonstrate that the application of XAI not only enhances model transparency but also contributes to the overall efficiency and accuracy of the obesity prediction system.

In conclusion, the Adam optimizer proved to be more effective than SGD for training ANN in obesity classification tasks. Moreover, interpreting the model using SHAP effectively identified the most influential features, offering a more transparent understanding of model decisions. Feature selection based on SHAP contribution values not only maintained but significantly improved model accuracy and computational efficiency. This research makes an important contribution to the implementation of Explainable Artificial Intelligence (XAI) in health classification systems, particularly in obesity detection. Future work may involve testing the model on more diverse datasets and considering external factors to further enhance its generalization and robustness in real-world conditions.

References

[1] T. Khater, H. Tawfik, and B. Singh, “Explainable artificial intelligence for investigating the effect of lifestyle factors on obesity,” *Intelligent Systems with Applications*, vol. 23, Sep. 2024, doi: 10.1016/j.iswa.2024.200427.

[2] W. Lin, S. Shi, H. Huang, J. Wen, and G. Chen, “Predicting risk of obesity in overweight adults using interpretable machine learning algorithms,” *Front Endocrinol (Lausanne)*, vol. 14, 2023, doi: 10.3389/fendo.2023.1292167.

[3] Z. Helforoush and H. Sayyad, “Prediction and classification of obesity risk based on a hybrid metaheuristic machine learning approach,” *Front Big Data*, vol. 7, 2024, doi: 10.3389/fdata.2024.1469981.

[4] J. H. Bae, J. W. Seo, X. Li, S. Y. Ahn, Y. Sung, and D. Y. Kim, “Neural network model for prediction of possible sarcopenic obesity using Korean national fitness award data (2010–2023),” *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-64742-w.

[5] C. Ogami *et al.*, “An artificial neural network−pharmacokinetic model and its interpretation using Shapley additive explanations,” *CPT Pharmacometrics Syst Pharmacol*, vol. 10, no. 7, pp. 760–768, Jul. 2021, doi: 10.1002/psp4.12643.

[6] S. S. Singh Rana, J. S. Ghahremani, J. J. Woo, R. A. Navarro, and P. N. Ramkumar, “A Glossary of Terms in Artificial Intelligence for Healthcare,” *Arthroscopy - Journal of Arthroscopic and Related Surgery*, Feb. 2024, doi: 10.1016/j.arthro.2024.08.010.

[7] S. Azmi *et al.*, “Harnessing Artificial Intelligence in Obesity Research and Management: A Comprehensive Review,” *Diagnostics*, vol. 15, no. 3, Feb. 2025, doi: 10.3390/diagnostics15030396.

[8] A. K. Tripathi, N. R. Chauhan, and A. Sharma, “Obesity Classification and Prognosis Using Machine Learning,” *AIP Conf Proc*, vol. 3224, no. 1, Feb. 2025, doi: 10.1063/5.0245938.

[9] N. K. A. P. S. Dewi, A. W. Wijayanto, and J. A. Nursiyono, “Comparison of Machine Learning Algorithms in Classifying Districts/Cities in Indonesia According to the Human Development Index (HDI) in 2021,” *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, vol. 4, no. 1, pp. 26–33, Jan. 2025, doi: 10.20885/snati.v4.i1.4.

[10] F. Arifin, H. Sibyan, and N. Hasanah, “Rancang Bangun Chatbot Pada Sistem EKAPTA Berbasis Natural Language Processing dengan Algoritma Artificial Neural Network,” *Jurnal Ilmiah Informatia dan Komputer*, vol. 4, no. 1, pp. 1–8, Jan. 2025, doi: 10.32699/biner.v4i1.7687.

[11] C. Özkurt, “Examination and Evaluation of Obesity RiskFactors with Explainable Artificial Intelligence,” *Computers and Electronics in Medicine*, Jul. 2024, doi: 10.69882/adba.cem.2024072.

[12] Q. Sun, A. Akman, and B. W. Schuller, “Explainable Artificial Intelligence for Medical Applications: A Review,” *ACM Trans Comput Healthc*, no. 2, Feb. 2025, doi: 10.1145/3709367.

[13] F. M. Palechor and A. de la H. Manotas, “Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico,” *Data Brief*, vol. 25, Aug. 2019, doi: 10.1016/j.dib.2019.104344.

[14] S. Mondal, R. Maity, and A. Nag, “An efficient artificial neural network-based optimization techniques for the early prediction of coronary heart disease: comprehensive analysis,” *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-85765-x.

[15] A. S. Ritonga and I. Muhandhis, “Teknik Data Mining Untuk Mengklasifikasikan Data Ulasan Destinasi Wisata Menggunakan Reduksi Data Principal Component Analysis (PCA),” *Jurnal Ilmiah Edutic*, vol. 7, no. 2, 2021, doi: 10.21107/edutic.v7i2.9247.

[16] S. Layeghay, M. Portmann, M. Gallagher, and L. Manocchio, “An Empirical Evaluation of Preprocessing methods for Machine Learning based Network Intrusion Detection Systems,” *SSRN*, Jan. 2025, [Online]. Available: https://ssrn.com/abstract=5079226

[17] H. Syahidah, N. Irsandi, A. N. Ajizah, and A. Amelia, “Obesity Prediction Using Machine Learning Algorithms,” *IJATIS: Indonesian Journal of Applied*

Technology and Innovation Science, vol. 2, no. 1, pp. 53–62, Mar. 2025, doi: 10.57152/ijatis.v2i1.1869.

[18] A. Bhandari and G. Adhikari, “Artificial Neural Network for Digits Classification,” *TechRxiv*, Feb. 2025, doi: 10.36227/techrxiv.173895043.31733566/v1.

[19] M. Azad, M. F. K. Khan, and S. A. El-Ghany, “XAI-Enhanced Machine Learning for Obesity Risk Classification: A Stacking Approach with LIME Explanations,” *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3530840.