Explainable AI for Heart Attack Prediction

line 1: up2212278
line 2: *Explainable Artificial Intelligence*
line 3: *University of Portsmouth*
line 4: Portsmouth, United Kingdom
line 5: up2212278@myport.ac.uk

***Abstract* —This paper applies explanatory artificial intelligence (XAI) techniques to interpret a logistic regression model predicting heart attacks. I use Shapley Additive Explanations (SHAP) and Local Interpretable Model-Independent Explanations (LIME) to improve the transparency of model predictions.**

Keywords— "Explainable Artificial Intelligence, XAI, Logistic Regression, SHAP, LIME, Model Interpretation"

I. Introduction

The field of machine learning has seen significant advances in recent years, leading to the development of sophisticated models capable of making highly accurate predictions. However, the complexity of these models often comes at the expense of interpretability, making it difficult for users to understand how decisions are made. This issue is particularly important in healthcare, where understanding the rationale behind predictions can significantly impact clinical decision making and patient outcomes.

Explain Artificial Intelligence (XAI) has become a critical area of research aimed at improving the transparency and interpretability of machine learning models. XAI methods provide insight into how models generate predictions, allowing users to trust those models and use them effectively in practice. By making decision-making more transparent, XAI helps bridge the gap between complex machine learning models and the need for interpretability in mission-critical applications.

This article focuses on the application of XAI methods to a logistic regression model used to predict heart attacks. Logistic regression is a widely used statistical method in healthcare due to its simplicity and interpretability. However, as the number of features and their interactions increases, even logistic regression models can become difficult to interpret comprehensively.

To solve this problem, I use two popular XAI methods: Shapley Additive Explanations (SHAP) and Local Interpretable Model-Independent Explanations (LIME). SHAP values, based on cooperative game theory, provide a single measure of feature importance, offering both global and local explanations for the model's predictions. LIME, on the other hand, generates locally accurate explanations by approximating the model with an interpretable model in the vicinity of each prediction.

The main objectives of this study are:

1. To improve the interpretability of a logistic regression model for heart attack prediction.
2. Identify and understand key features that increase the risk of heart attacks.
3. Demonstrate the practical application and benefits of SHAP and LIME in a healthcare context.

II. Review

A. *Overview of Explainable Artificial Intelligence Methods*

In this section, we'll look at the main Explainable Artificial Intelligence (XAI) techniques that help you understand how machine learning models make decisions. This is important because people need to understand why the model made a certain prediction, especially in areas such as healthcare and finance.

SHAP The SHAP (Shapley Additive Explanations) method helps explain how machine learning models make their predictions using concepts from game theory. Imagine you have a group of people (players) who work together and win a common prize. Shapley values help share this prize fairly among all players based on their contributions.

In the world of machine learning, “players” are the features (factors) that the model uses to make predictions. The SHAP method calculates how much each feature contributes to the model's prediction for each specific case. These calculations are called Shapley values. SHAP takes the data and the model that has already been trained, and for each prediction it determines how important each feature is. This is done in order to understand which features contributed most to a particular prediction.

For example, if a logistic regression model predicts the risk of a heart attack, SHAP may show that age and cholesterol levels were the most important factors in that decision. This helps doctors and patients understand why the model reached its conclusion and makes the model more transparent and trustworthy.

LIME The LIME (Local Interpretable Model-agnostic Explanations) method helps to understand how a model makes decisions by explaining its predictions [2] Think of a complex model as a “black box” that is difficult to look into and understand how it works. The LIME method creates a simple model (linear model) that simulates the behaviour of this complex model, but only close to the specific prediction that will be of interest.

LIME takes the original model and the data it was trained on and creates a bunch of new data by slightly changing the original features. It then trains a simple model on this modified data so that the simple model can predict almost as well as the complex model, but only close to the chosen prediction. This simple model allows us to see which features were most important to the decision in this particular case.

In the context of logistic regression for predicting heart attacks, LIME can show which factors (e.g., age, body mass index) were most influential in predicting that a person would have a heart attack. Thus, LIME makes the operation of a complex model more transparent and understandable.

B. B. XAI Use Cases

Examples of how XAI methods are used in real life to make models more understandable:

Explainable Artificial Intelligence in Healthcare for Decision-Making XAI helps doctors understand the decisions of AI systems, which is especially useful when the number of beds in a hospital is limited, helping to make decisions about hospitalization of patients with pneumonia and asthma [1].

Explainable Artificial Intelligence in Healthcare for Medicine XAI explains the decisions of AI systems for diagnosis and treatment, which increases doctors' confidence in such systems. For example, XAI helps explain the classification of cardiac pathologies or breast cancer biopsies [1].

Explainable Artificial Intelligence in Healthcare for Diagnosis XAI is used to diagnose various diseases where deep learning methods cannot provide clear explanations. For example, XAI is used to diagnose allergies and analyse chest X-rays in patients with COVID-19 [1].

These examples show how XAI methods can be applied in various fields, making machine learning models more transparent and trustworthy.

III. Methodology

A. Preliminary data preparation and model training

Health and lifestyle data were used to build a model to predict the likelihood of a heart attack [6] The dataset contains 246,022 instances with 27 attribute s including gender, age, general health, presence of diseases (diabetes, angina, stroke), smoking status and others. The target attribute is "HadHeartAttack" It indicates whether the person has had a heart attack or not:

HadHeartAttack

No – 232587 instance s

Yes – 13435 instances

To prepare the data, the following actions were carried out:

- Converting categorical features to numeric values.
- Converting continuous metrics such as body mass index (BMI) and sleep hours (SleepHours) to a numeric format.
- Dividing the data into training and test sets.

A logistic regression model was used to predict heart attack. Model hyperparameters were selected using cross-validation (GridSearchCV).

Using cross-validation, the optimal parameters for the logistic regression model were found using the GridSearchCV method.

Best options:

{'C': 10, 'class_weight': None, 'solver': 'saga'}

The regularization parameter (C) controls the degree of regularization of the model. Regularization helps prevent overfitting by penalizing the model for weights that are too complex or large. The value of C is inversely proportional to the strength of the regularization: a smaller value of C leads to stronger regularization, while a larger value of C weakens the regularization.

In this case, a value of C = 10 indicates relatively weak regularization, allowing the logistic regression model to better fit the data.

The class_weight parameter is used to control the balance of classes in the model. Setting class_weight to None means that the classes in the data are not weighted differently and all classes have equal importance to the model.

In the context of imbalanced data (where one class is represented significantly more often than another), it may be useful to set this parameter to 'balanced' to compensate for the imbalance. However, in this case, the lack of balancing indicates that the model does not take into account class imbalance.

The solver parameter specifies the algorithm used to optimize the model. Saga is a variant of stochastic gradient descent that works well on large data sets and supports L1 (Lasso) and L2 (Ridge) regularization as well as Elastic Net.

The choice of saga as a solver indicates that the model uses an efficient and scalable optimization algorithm suitable for large and sparse data.

Error Matrix:

[[46018 500]

[2044 643]]

The model showed high accuracy but a low F1-score for the "Yes" class, indicating difficulty in predicting rare events (heart attacks).

After training, the model received the following scores:

Classification Report:

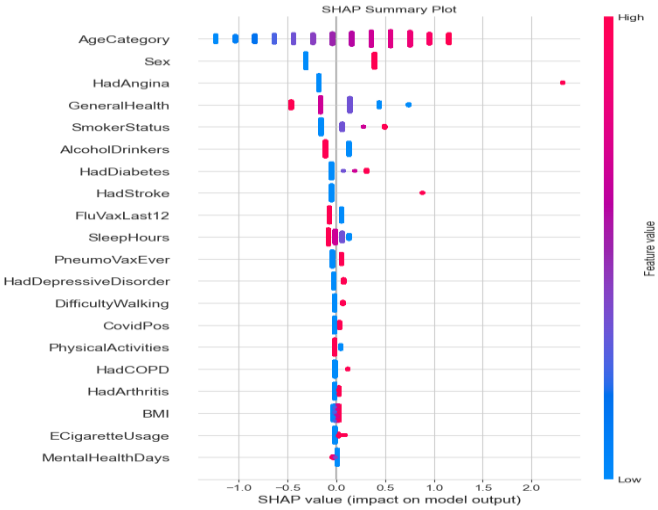
	precision	recall	f1-score	support
No	0.96	0.99	0.97	46518
Yes	0.56	0.24	0.34	2687
accuracy			0.95	49205
macro avg	0.76	0.61	0.65	49205
weighted avg	0.94	0.95	0.94	49205

B. Methods for Explaining Artificial Intelligence (XAI)

SHAP (SHapley Additive exPlanations) method is used to explain the output of machine learning models [1] In the context of a heart attack prediction model, SHAP helps understand the contribution of each feature to the model's predictions. It explains how factors such as age, general health, smoking status and other variables affect the likelihood of a heart attack predicted by a logistic regression model.

The SHAP method calculates the contribution of each feature, taking into account all possible combinations of features. It assigns each feature an importance value for a particular prediction. These values can be aggregated across all predictions to obtain an overall importance score. The SHAP summary plot visualizes the distribution of SHAP values for each trait across the entire data set, showing both the direction and magnitude of the effect.

SHAP Summary Plot



Age category. Age is a critical factor, with higher age categories showing a stronger positive effect on the likelihood of a heart attack. Red dots (high attribute values) are located to the right of zero, indicating an increase in risk.

Floor. Male gender (coded as 1) increases the likelihood of having a heart attack. Red dots (men) are to the right of zero, indicating a positive effect on heart attack prediction.

General health. Poor general health significantly increases the risk of a heart attack. Red dots (poor health) are also located to the right of the zero.

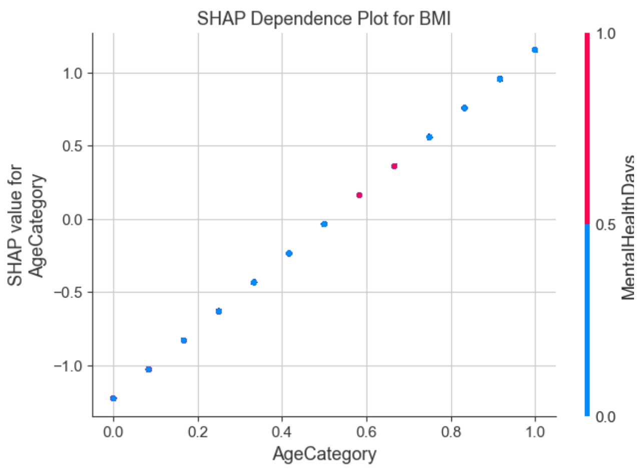
Had angina and stroke. These medical conditions greatly influence the prediction of a heart attack. High trait values (red dots) for these conditions are also located on the right, indicating increased risk.

Smoker status. Current smokers show a higher likelihood of having a heart attack. The red dots (current smokers) to the right of zero confirm this.

Alcohol and diabetes. These factors also have a noticeable impact, although less than the main factors.

The graph shows that high SHAP scores (in red) often correspond to high trait scores for critical factors such as age, general health and existing medical conditions (angina, stroke), indicating that these factors significantly increase the risk of a heart attack.

SHAP Dependence Plot



SHAP Dependence Plot is used to visualize the relationship between a characteristic value and its SHAP value. This method allows us to identify interactions between features and their impact on model predictions.

This graph shows the dependence of SHAP values of the age category (AgeCategory) on body mass index (BMI) values. The colour of the dots reflects the number of days with mental health problems (MentalHealthDays).

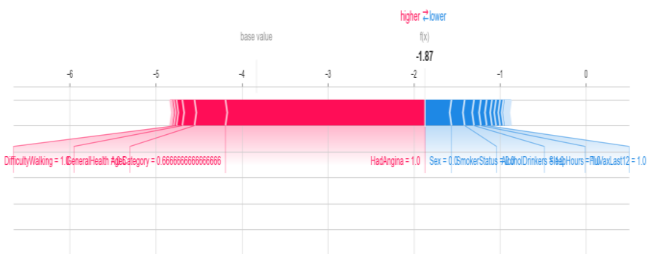
The graph shows that as age (AgeCategory) increases, SHAP values also increase. This indicates that for older adults, age category has a significant impact on model predictions. The older a person is, the higher their SHAP score, indicating a greater risk of heart attack.

The graph demonstrates that body mass index influences age category SHAP values. Higher BMI values are associated with higher age-specific SHAP values. This indicates that age and BMI interact to enhance each other's influence on model predictions.

The colour of the dots shows how the number of days with mental problems affects the interaction between age and BMI. In this graph you can see that the points that have a higher value on the Y axis (SHAP value of the age category) also have a different colour, indicating the variability of mental problems.

Using the SHAP Dependence Plot, we can see that age and BMI strongly interact to increase the risk of a heart attack. The model reveals that as age increases, the influence of age category on predictions becomes more significant, especially at high BMI values. The colour scale indicated that mental health may also play a role in these interactions, adding another level of detail to the model's interpretation.

SHAP Force Plot for the first prediction:



SHAP Force Plot helps you visualize how each feature contributes to a specific model prediction. This graph shows which features increase or decrease the likelihood of a heart attack for a given observation.

This graph shows the impact of various features on the prediction for the first observation in the test dataset.

Positive impacts (increased risk):

“DifficultyWalking = 1.0” Significantly increases the chance of a heart attack.

“GeneralHealth” Poor general health also increases risk.

“AgeCategory = 0.6666666666666666” Higher age category increases risk.

Negative impacts (risk reduction):

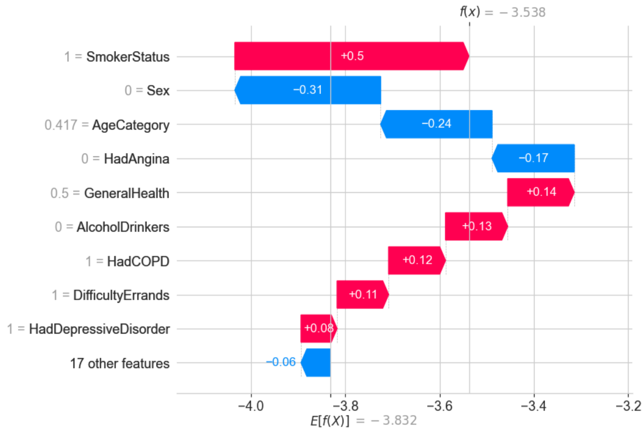
“HadAngina = 1.0” Having angina slightly reduces the risk.

“Sex = 0” Being female reduces the chance of having a heart attack.

“SmokerStatus” “AlcoholDrinkers” “SleepHours” “FluVaxLast12” These signs also reduce the risk in this observation.

SHAP Force Plot allows you to visualize which features contribute most strongly to a prediction, giving you insight into why the model made the decision it did.

SHAP Waterfall Plot for detailed analysis of the contribution of features to specific predictions:



SHAP Waterfall Plot is used to analyse in detail the contribution of each feature to a specific model prediction. This graph helps you understand how each feature contributes to the prediction, starting with a base value (the average prediction) and adding or subtracting the contribution of each feature.

This graph shows how different features affect a specific model prediction.

Positive impacts (increased risk):

“1 = SmokerStatus (+0.5)” Being a smoker significantly increases your chance of having a heart attack.

“0.5 = GeneralHealth (+0.14)” Poor general health also increases risk.

“1 = HadCOPD (+0.12)” Having chronic obstructive pulmonary disease (COPD) increases the risk.

“1 = DifficultyErrands (+0.11)” Difficulty performing daily activities also contributes to increased risk.

“1 = HadDepressiveDisorder (+0.08)” Having a depressive disorder increases the risk.

Negative impacts (risk reduction):

“0 = Sex (-0.31)” Being female reduces the likelihood of having a heart attack.

“0 = HadAngina (-0.24)” The absence of angina also reduces the risk.

“0 = AlcoholDrinkers (-0.17)” Not consuming alcohol reduces the risk.

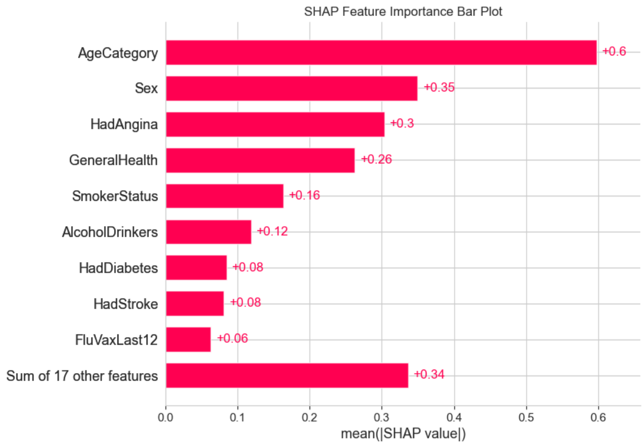
“17 other features (-0.06)” The other 17 features collectively also reduce risk.

Age category contribution (0.417 = AgeCategory) (-0.24):

Age category also plays a significant role, influencing prediction to a lesser extent than some of the most significant features, but still noticeably.

SHAP Waterfall Plot allows you to visualize how each feature affects the prediction for a particular observation, starting with a base value and adjusting it based on feature values.

SHAP Feature Importance Bar Plot



This graph shows the average importance of each feature based on the average absolute SHAP values. It helps determine which features are most important to the model's overall predictions.

AgeCategory: The most important attribute with the highest average SHAP value (+0.6). Age strongly influences the model's predictions.

Sex: Second most important attribute (+0.35). Gender also has a significant influence.

HadAngina: The presence of angina has a significant impact (+0.3).

GeneralHealth: Poor general health has a significant impact on predictions (+0.26).

SmokerStatus: Smoker status is also important for the model (+0.16).

AlcoholDrinkers: Drinking alcohol affects predictions (+0.12).

HadDiabetes, HadStroke, FluVaxLast12: These signs also have an effect, but to a lesser extent.

Using SHAP Force Plot, SHAP Feature Importance Bar Plot, and SHAP Waterfall Plot provided detailed insight into how the model makes decisions and which features have the greatest impact on predictions. These methods together provided a comprehensive understanding of the model's performance, confirming that the main predictors of heart attack include age, gender, health status and the presence of diseases. The logistic regression model relies on these factors to make its predictions, making it more transparent and interpretable.

C. Using LIME (Local Interpretable Model-agnostic Explanations) to explain the model

LIME helps explain the model's predictions by showing how different features affect the outcome. This method uses simple models such as linear ones to locally explain predictions.

Process of using LIME:

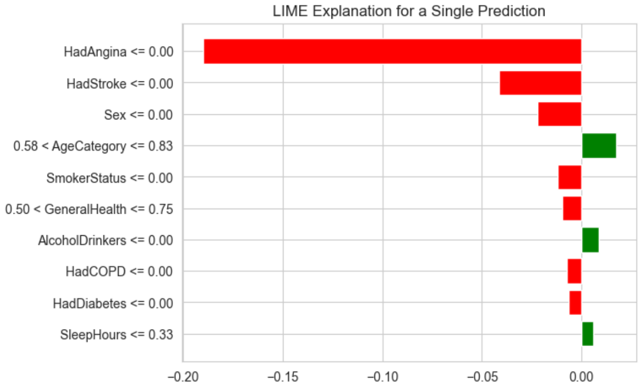
1. Creating an explainer. The LIME explainer was tuned based on the training data. This allows LIME to understand the distribution of data and how it relates to predictions.

2. Selecting an example for analysis. A random sample from the test dataset was selected for analysis. This allows you to test how the model makes specific predictions on new data.

3. Generating an explanation. LIME generated an explanation for the selected example, highlighting 10 important features that influenced the prediction. This shows which features played a key role in the model's solution.

4. Presentation of explanation. The explanation was presented both in text and as a graph, which helps to visually understand the impact of each attribute.

LIME Explanation graph for one prediction:



The graph shows the influence of various features on a specific model prediction. This allows us to understand why the model made this decision. Red bars represent features that reduce the likelihood of having a heart attack, and blue bars represent features that increase the likelihood. The length of the bar indicates the degree of influence of each feature.

Signs that reduce the risk of a heart attack:

Absence of Angina (HadAngina <= 0.00): Greatly reduces the likelihood of a heart attack (-0.19). This is important because angina is a serious risk factor for heart disease.

No Stroke (HadStroke <= 0.00): Also, significantly reduces risk (-0.04). This is important because a stroke indicates a problem with the cardiovascular system.

Female gender (Sex <= 0.00): Women have a lower risk of heart attack (-0.02) compared to men, which is also reflected in the model.

Non-smoking habit (SmokerStatus <= 0.00): Reduces risk (-0.012) as smoking has a significant impact on heart health.

Good general health (GeneralHealth <= 0.75): Reduces risk (-0.0097) as general health reflects the level of heart risk.

No COPD (HadCOPD <= 0.00): Reduces the likelihood of a heart attack (-0.0074), as COPD negatively affects the cardiovascular system.

Absence of diabetes (HadDiabetes <= 0.00): Also reduces risk (-0.0065), since diabetes is a significant risk factor.

Signs that increase the risk of a heart attack:

Age Category (0.58 < AgeCategory <= 0.83): Slightly increases the risk of heart attack (+0.017).

Getting older increases your risk of heart attack because your risk of heart disease increases with age.

Alcohol consumption (AlcoholDrinkers <= 0.00): Slightly increases risk (+0.0085) as alcohol can have a negative effect on the cardiovascular system.

Lack of sleep (SleepHours <= 0.33): Also, slightly increases risk (+0.006), as lack of sleep affects overall health and the heart.

LIME Explanation text output

“[('HadAngina <= 0.00', -0.19004809499334402),

('HadStroke <= 0.00', -0.04138411467123845),

('Sex <= 0.00', -0.02216019726504595),

('0.58 < AgeCategory <= 0.83', 0.017155160938588987),

('SmokerStatus <= 0.00', -0.012087847862715618),

('0.50 < GeneralHealth <= 0.75', -0.009685797340030514),

('AlcoholDrinkers <= 0.00', 0.008484968626741992),

('HadCOPD <= 0.00', -0.00740294028772874),

('HadDiabetes <= 0.00', -0.006471060574115354),

('SleepHours <= 0.33', 0.006056328073972027)]”

This list shows which signs influenced the prediction and in what direction. Negative values reduce the likelihood of a heart attack, while positive values increase it. For example, the absence of angina reduces the risk by 0.19, while age increases the risk by 0.017.

HadAngina <= 0.00 (-0.19): The absence of angina greatly reduces the likelihood of a heart attack. This makes sense, since angina is an important risk factor for heart disease.

HadStroke <= 0.00 (-0.041): Not having a stroke also significantly reduces the risk of a heart attack, since a stroke indicates a serious problem with the cardiovascular system.

Sex <= 0.00 (-0.022): Being female (value 0) reduces the chance of having a heart attack. This is supported by medical evidence, where women have a lower risk of heart disease than men.

0.58 < AgeCategory <= 0.83 (+0.017): People in this age category have a slightly increased risk of heart attack. This may be due to the fact that the likelihood of developing cardiovascular disease increases with age.

SmokerStatus <= 0.00 (-0.012): Not smoking reduces the risk of heart attack. Smoking is a known risk factor for the heart.

0.50 < GeneralHealth <= 0.75 (-0.009): Good general health also reduces the risk of a heart attack, which makes sense since general health has a big impact on the heart.

AlcoholDrinkers <= 0.00 (+0.008): Drinking alcohol slightly increases the risk of a heart attack, as alcohol can have negative effects on the cardiovascular system.

HadCOPD <= 0.00 (-0.007): Not having COPD reduces the risk of a heart attack because COPD negatively affects the heart.

HadDiabetes <= 0.00 (-0.006): Not having diabetes also reduces your risk, since diabetes is a significant risk factor for the heart.

SleepHours <= 0.33 (+0.006): Lack of sleep slightly increases the risk of a heart attack, as poor sleep can negatively impact your overall health and heart.

These visual representations and textual explanations complement each other by providing both quantitative and visual interpretation of the influence of features.

Data sampling Example of data with quartile sampling:

```
Discretized data (Quartiles):
[[0.         0.5        0.         1.         1.         1.
  0.         0.         0.         0.         0.         0.
  1.         0.         1.         1.         0.         0.
  0.         0.66666667 1.         1.         1.         0.
  0.         0.        ]
 [1.         0.5        0.         0.         1.         0.
  1.         0.         0.         0.         1.         0.
  1.         1.         0.         0.         0.         0.
  0.33333333 0.58333333 1.         0.         0.         0.
  0.         0.        ]
 [0.         1.         0.         1.         1.         0.
  0.         0.         0.         0.         0.         0.
  0.         0.         0.         0.         0.         0.
  0.         0.         0.         0.         0.         0.
```

In this example, each feature was divided into four groups. For example, the value `0.5` means that the attribute value is in the second quartile.

Discretization helps break continuous data into intervals, which allows us to better understand its distribution and impact on model predictions.

Quartile sampling. The data were divided into four parts (quartile ranges). This helped to understand how the data is distributed and what values occur most frequently. Each interval contains 25% of the data, which makes the analysis of feature distribution more visual.

Example:

0.00: Lower quartile (lowest values)

0.50: Median (average values)

0.75: Upper quartile (highest values)

1.00: Upper quartile (maximum values)

Decile sampling. The data were divided into ten parts (decile intervals). This provided a more detailed view of the data distribution, which is important for accurate analysis. Each interval contains 10% of the data.

Example:

0.00: First decile (lowest values)

0.50: Fifth decile (average values)

0.75: Seventh decile (high values)

1.00: Tenth decile (maximum values)

Discretization helps simplify data analysis by making it more understandable. Discretization allows us to better understand the impact of each feature on the model's predictions. Sampling also reduces noise in the data, which can improve the quality of model predictions.

LIME showed how specific features influence individual model predictions. For example, the absence of angina and stroke, female gender, and no smoking habit significantly reduced the likelihood of a heart attack for the selected example.

Data sampling helped to better understand how feature values are distributed and their impact on predictions. Quartile and decile sampling provided a fine-grained view of the data, allowing for better interpretation of the impact of each feature on model predictions.

IV. CONCLUSION

The study used explainable artificial intelligence (XAI) techniques to analyse a logistic regression model predicting the likelihood of a heart attack. The main attention was paid to the SHAP and LIME methods to make the model more transparent and understandable.

The dataset contained health and lifestyle data for 246,022 people with 27 features. The key target feature was "HadHeartAttack" indicating whether the person had a heart attack. Data were carefully prepared, including converting categorical features to numeric values and converting continuous features such as body mass index (BMI) and sleep hours (SleepHours) into numeric format. The data was divided into training and test sets to train and evaluate the model.

A logistic regression model was trained to predict heart attack. Hyperparameter optimization was performed using cross-validation. The model showed high accuracy (95%), but a low F1-measure for the "Yes" class (0.34), which indicates difficulties in predicting rare events (heart attacks). The error matrix confirmed that the model tended to miss heart attack cases despite its high overall accuracy.

During the research, the following methods of explaining artificial intelligence (XAI) were used

SHAP (SHapley Additive exPlanations)

- SHAP Summary Plot
- SHAP Dependence Plot
- SHAP Force Plot
- SHAP Waterfall Plot
- SHAP Feature Importance Bar Plot

LIME (Local Interpretable Model-agnostic Explanations):

- LIME Explanation plot for one prediction
- LIME Explanation Text Representation

LIME Discretize

Key results showed that explanation methods such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) effectively identify and visualize the significance of various factors influencing model predictions. These methods allowed a detailed assessment of the contribution of each characteristic, such as age, gender, general health and the presence of diseases, to the likelihood of a heart attack.

The SHAP method provided insight into the impact of each trait on prediction, showing that age and gender were key predictors of heart attack. Visualizations using SHAP Summary Plot and SHAP Force Plot demonstrated how specific feature values increased or decreased the likelihood of a heart attack. The LIME method, in turn, provided local explanations for individual predictions, revealing which features played a decisive role in each particular case. This confirms that XAI methods can be effective tools for analysing and interpreting complex models, improving their usefulness and user confidence.

One of the main limitations of the study is the use of one dataset and one machine learning model. Although the data set is large, it may not reflect all of the variability in the real population. Logistic regression, although an interpretable model, may not fully capture complex nonlinear relationships between traits. Additional limitations are associated with possible data distortions caused by preprocessing and the choice of model hyperparameters. The limitations of XAI methods in explaining interactions between traits and possible interpretation errors should also be considered.

Application of XAI methods in decision support systems allows medical professionals to better understand what factors influence the risk of a heart attack in a given patient. This may improve diagnostic accuracy and quality of decision making in clinical practice. SHAP and LIME methods can be implemented in medical information systems, increasing their transparency and user trust. In addition, these methods can be used to train medical personnel, helping them better understand how machine learning models work and their predictions. [3]

Future research should aim to apply XAI methods to other machine learning models such as neural networks and ensemble methods to compare their interpretability and prediction accuracy. It is also important to examine the use of these methods on different datasets, including data from different populations and sources, to increase the generalizability and validity of the results. Additionally, the impact of different data preprocessing strategies on model interpretation results should be considered. The development of new XAI methods that can better explain complex dependencies between features is also a promising direction for further research.

The study demonstrated that explainable artificial intelligence techniques such as SHAP and LIME are powerful tools for improving the transparency and interpretability of machine learning models used to predict heart attacks. These methods provide a detailed understanding of how various factors influence model predictions, which is especially important for use in medical decision support systems. The study's findings highlight the need and importance of using interpretable models in medicine and other critical fields, promoting the development of more reliable and trusted artificial intelligence systems.

References

[1] Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. In Advances in Neural Information Processing Systems (pp. 4765-4774).

[2] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144).

[3] A. L. Imoize and J. Hemanth, et al., Explainable Artificial Intelligence in Medical Decision Support Systems. London, UK: The Institution of Engineering and Technology, 2022. ISBN 978-1-83953-620-5 (hardback), ISBN 978-1-83953-621-2.

[4] S. Tulli and D. W. Aha, Explainable Agency in Artificial Intelligence Research and Practice, 1st ed. Boca Raton, FL: CRC Press, 2024.

[5] O. P. Jena, M. Panda, U. Kose, et al., Medical Data Analysis and Processing using Explainable Artificial Intelligence, 1st ed. Boca Raton, FL, USA: CRC Press, 2024.

[6] Pytlak, K. (2022). Indicators of Heart Disease (2022 UPDATE). Retrieved from
https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease/data

[7] Molnar, C. (2023). Interpretable Machine Learning A Guide for Making Black Box Models Explainable. Retrieved from
https://christophm.github.io/interpretable-ml-book/

[8] Trevisan, V. (2022). Using SHAP Values to Explain How Your Machine Learning Model Works. Towards Data Science. Retrieved from
https://towardsdatascience.com/using-shap-values-to-explain-how-your-machine-learning-model-works-732b3f40e137

[9] Cooper, A. (2021). Explaining Machine Learning Models: A Non-Technical Guide to Interpreting SHAP Analyzes. Retrieved from https://www.aidancooper.co.uk/a-non-technical-guide-to-interpreting-shap-analyses/

[10] SHAP Documentation. Retrieved from
https://shap-lrjball.readthedocs.io/en/latest/index.html#welcome-to-the-shap-documentation

[11] LIME Documentation. Retrieved from
https://lime-ml.readthedocs.io/en

[12] Sharma, A. (2018). Decrypting your Machine Learning model using LIME. Towards Data Science. Retrieved from
https://towardsdatascience.com/decrypting-your-machine-learning-model-using-lime-5adc035109b5