Nasrin Moradi
Arezou Habibirad
Hanieh Panahi

# EXPLAINING ADDITIVE WEIBULL MODEL PARAMETER ESTIMATION WITH XAI: A SHAP ANALYSIS



**Abstract** *Conventional machine-learning models face limitations in performing time-to-event analyses due to censoring issues. This study introduces a Deep Additive Weibull (DAW) model that utilizes deep learning techniques for the survival analysis of right-censored COVID-19 patient data. Also, we explore Shapley additive explanations (SHAP) as a method for "opening the black box" of the DAW model to enhance model trustworthiness. The DAW model leverages neural networks for survival analysis, specifically to estimate survival probabilities for each patient using an autoencoder-based network. The DAW model achieved a concordance index of 0.9699 for training and 0.89015 for testing. Our findings show that the DAW model effectively captures nonlinearities and complex interactions. We also assessed the impact of specific features on the model's prediction, providing valuable insights. Based on SHAP, the important features are pneumonia, diabetes, age, and immunosuppression. Moreover, we demonstrated that explainable machine learning (ML) can elucidate how models make predictions, which is crucial to increase trust and adopt innovative ML techniques in healthcare.*

**Keywords** Additive Weibull Model, autoencoder, black box, SHAP, survival analysis

**Citation** Computer Science 27(1) 2026: 5–24

## 1. Introduction

The Weibull distribution is widely used for modeling time-to-event data and has been adapted into various forms because of its broad applicability. The additive Weibull (AW) distribution, introduced by Xie and Lai [18], is based on the generalized Weibull distribution. The probability density function of the AW distribution is given by:

$$f(t) = (abt^{b-1} + cdt^{d-1}) \exp\big(-at^b - ct^d\big);\ t > 0, \tag{1}$$

$a > 0$, $c > 0$, and $b > d > 0$ (or $d > b > 0$) are the scale and shape parameters, respectively. Also, the cumulative distribution and hazard rate function can obtain as follows:

$$F(t) = 1 - \exp\big(-at^b - ct^d\big);\ t > 0, \tag{2}$$

$$r(t) = abt^{b-1} + cdt^{d-1});\ t > 0, \tag{3}$$

Furthermore, in life testing experiments, there may be instances in which some units are lost or not followed, or situations in which certain units are removed before they fail. This approach helps save time and reduce the costs associated with experimentation, as well as free up testing facilities. The most commonly used censoring schemes are right censoring and left censoring. Medical data is often complex and difficult to interpret. While machine learning models can achieve accurate predictions, they are frequently considered "black boxes" due to the opacity of their decision-making processes.

Interpretability is essential in medical applications, as it allows clinicians and researchers to understand how models generate predictions and identify any potential biases or errors. Interpretability techniques enable clinicians to identify high-risk patients and gain insights into factors that affect survival. Several methods have been developed to interpret machine learning models.

Another important technique is SHAP, developed by Lundberg and Lee [11]. Their unified approach to model interpretation demonstrates that many existing methods, such as LIME and DeepLIFT, are specific instances of SHAP values. They also proposed efficient algorithms for computing SHAP values for tree-based models [13]. Additionally, they introduced several interpretation methods based on SHAP values, such as feature importance plots, dependence plots, and summary plots [12]. Recent work has expanded the application of Weibull-based distributions in survival analysis. Sindhu et al. [20] explored the entropy transformed Weibull distribution's properties and demonstrated its applications across various scientific fields. Lone et al. [10] evaluated different estimator approaches using COVID-19 data alongside simulation studies, providing insights into optimal estimation methods for pandemic-related survival data.

Additionally, Sindhu et al. [17] introduced a decreasing failure-rate model that incorporated a novel approach to enhance artificial neural network structures, successfully applying this methodology to both engineering problems and disease data

analysis. These works collectively demonstrate the ongoing development of Weibull-based models integrated with machine learning techniques for survival analysis.

Machine learning techniques ( [1]- [5]) are increasingly being utilized for a wide range of medical applications [16]. Also, various methods have been employed to interpret machine learning models. For instance, Moncada-Torres et al. [14] compared the performance and interpretability of Cox regression and machine learning models to predict breast cancer patient survival. Zhang et al. [19] examined different methods for interpreting neural network models in clinical applications, discussing the strengths and limitations of each approach. Duckworth et al. [3] used the SHAP method to identify characteristics associated with a higher or lower risk of hospitalization due to COVID-19, including age, comorbidities, and symptoms. For further insights [8].

We propose a model that employs neural networks with an autoencoder model to estimate the parameters of the additive Weibull (AW) distribution and survival probabilities while accounting for right-censored data. We refer to this network as the deep additive Weibull (DAW) model. However, this model is often considered a black box, which is undesirable because it makes it difficult to understand how the predictions are derived. Therefore, it is essential to provide an explanation of how the DAW model reaches its output in understandable terms. To our knowledge, no study has been conducted on the additive Weibull distribution using these techniques. In this paper we employ the SHAP method to explain the model's behavior and identify the most important features for prediction. First, we calculate the contribution and impact of each feature using the SHAP algorithm.

The paper can be summarized as follows: Section 2 describes the methods and system components. Section 3 discusses the interpretability methods. The results and discussion are presented in Section 4. The final section concludes the paper.

## 2. Model description

In this section we provide a comprehensive explanation of the model. This includes an in-depth description of the autoencoder neural network.

### 2.1. Autoencoder neural network

Autoencoders are neural networks used for tasks like data compression and feature extraction. They consist of three main components: an encoder, which compresses the input data; a bottleneck, which holds the compressed representation; and a decoder, which reconstructs the data from its compressed form. After training, the compressed features in the bottleneck can be used as inputs for prediction models. In our study, once the autoencoder is trained the encoded layer (bottleneck) can be used to extract features. These features are compressed representations that capture essential characteristics of the input data [9–15]. The autoencoder's structure includes an input layer with 35 nodes, an encoder with 35 nodes, a bottleneck layer with 31 nodes, and

a decoder layer with 35 nodes to reconstruct the data. We work with 35 features, converting non-binary categorical features (such as Medical Unit and Classification Final) into numerical data using the "get dummies" function from the pandas' library. This function generates binary columns for each category, making the transformed data suitable for machine learning models. The details of the autoencoder model are in Table 1.

**Table 1**
The process of building an autoencoder

| | Input | Encoder | Bottleneck | Decoder | Activation |
|---|---|---|---|---|---|
| Layers | 1 | 1 | 1 | 1 | ReLU |
| Nodes per layer | 35 | 35 | 31 | 35 | – |

The input data are compressed through the input layer, processed with a ReLU activation function in the bottleneck layer, and then reconstructed in the output layer using another ReLU activation function.

## 2.2. Deep Additive Weibull model

The DAW model builds on the AW model by incorporating deep learning techniques. First, we use an autoencoder model in order to extract the features in a lower dimension. These features represent a compressed version of the input data in which each feature signifies a learned characteristic of the input. We use these features as input for our prediction models. With transformed data we built a DAW model comprising a three-layer neural network and assessed its effectiveness in predicting patient survival rates using test data. The model uses a feedforward neural network to estimate the parameters of the AW distributions and survival probabilities from the right-censored data. To improve generalization we applied a dropout rate of 0.2. The model estimates the likelihood and learns the parameters $a$, $b$, $c$, and $d$ using a neural network with softplus activation functions:

$$a = softplus(k_1);\ b = softplus(k_2);\ c = softplus(k_3);\ d = softplus(k_4)$$

Softplus is defined as: $\log\left(1 + e^{k_i}\right);\ i = 1, 2, 3, 4$. The parameters of a model are selected based on a loss function that accounts for right-censored data and $l_2$-regularization. Figure 1 shows the model architecture. The output layer contains four neurons corresponding to the four parameters of the additive Weibull distribution. Each neuron uses a softplus activation function to ensure that all parameters are positive, which is a mathematical requirement for the additive Weibull distribution. The process of learning the unknown parameters of the AW distribution using the proposed censoring scheme is as follows: Suppose $T_1, T_2, \cdots, T_n$ are random i.i.d. variables given by $f(t)$ in Equation (1).

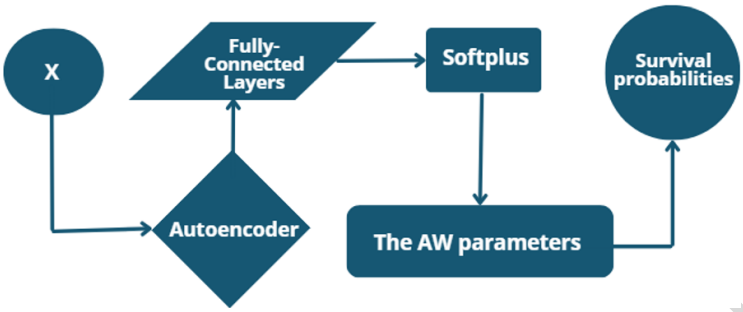



**Figure 1.** Architecture of the DAW network

Letting $\delta_i$ be a death indicator, we get the following likelihood function based on the right-censored samples:

$$\begin{aligned} L(a,b,c,d) &\propto \prod_{i=1}^{n} [f(t_i)]^{\delta_i} [S(t_i)]^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left[ \left( abt_i^{b-1} + cdt_i^{d-1} \right) \exp\left( -at_i^b - ct_i^d \right) \right]^{\delta_i} \\ &\quad \times \left[ \exp\left( -at_i^b - ct_i^d \right) \right]^{1-\delta_i} \end{aligned} \tag{4}$$

The following expression presents the negative log-likelihood (NLL):

$$-\log L(a,b,c,d) \propto -\sum_{i=1}^{n} \left\{ \delta_i \log \left( abt_i^{b-1} + cdt_i^{d-1} \right) - at_i^b - ct_i^d \right\} \tag{5}$$

To boost model performance and prevent overfitting, we include $l_2$ regularization in the loss function as a penalty term. This regularization aids in managing complexity and avoids overfitting by discouraging the model from learning noise in the training data, thereby enhancing the generalization of new data. Thus, we will use this modified loss function to train our model:

$$\ell \propto -\log L(a,b,c,d) + \frac{\lambda_{reg}}{2}(a^2 + b^2 + c^2 + d^2) \tag{6}$$

Here $\lambda_{reg}$ represents the $l_2$-regularization and controls the strength of the regularization. $\lambda_{reg}$ is the hyperparameter in which different values are tested and the one leading to the best model performance is selected.

## 2.3. Estimating the parameters of the AW model using Adagrad

In neural network training for estimating additive Weibull parameters, data is processed through the network to generate an output, which is evaluated using the negative log-likelihood (NLL) loss function to assess prediction accuracy. The gradients

calculated from this loss function are back-propagated to adjust the weights effectively using the Adagrad optimizer. Adagrad updates each weight based on its gradient's magnitude and direction, optimizing the model's performance with each iteration. The weight update process in neural network training can be summarized as follows:

1. Start with accumulators set to zero as $G_\theta = 0$ and compute gradients with $l_2$-regularization:
   To minimize $\ell$, the gradient $\ell$ must be calculated according to the network weight $\theta$.

$$\frac{\partial \ell}{\partial \theta} = \frac{\partial(-\log L)}{\partial \theta} + \frac{\partial(\frac{\lambda_{reg}}{2}(a^2+b^2+c^2+d^2))}{\partial \theta}$$

   It should be mentioned that we use the derivative chain rule to calculate $\frac{\partial(-\log L)}{\partial \theta}$:

$$\frac{\partial(-\log L)}{\partial \theta} = \frac{\partial(-\log L)}{\partial a}\frac{\partial a}{\partial k_1}\frac{\partial k_1}{\partial \theta} + \frac{\partial(-\log L)}{\partial b}\frac{\partial b}{\partial k_2}\frac{\partial k_2}{\partial \theta}$$

$$+\frac{\partial(-\log L)}{\partial c}\frac{\partial c}{\partial k_3}\frac{\partial k_3}{\partial \theta} + \frac{\partial(-\log L)}{\partial d}\frac{\partial d}{\partial k_4}\frac{\partial k_4}{\partial \theta}$$

2. Update sccumulators: Adjust accumulators with the square of the gradients as $G_\theta^{new} = G_\theta^{old} + \left(\frac{\partial \ell}{\partial \theta}\right)^2$.
3. Parameter updates: Parameters are updated using the Adagrad method, in which the learning rate $\eta$ is modified by the accumulators for better convergence. The learning rate is adjusted by the accumulators

$$\theta_{new} = \theta_{old} - \frac{\eta}{\sqrt{G_\theta + \epsilon}} \cdot \frac{\partial \ell}{\partial \theta}$$

   Here $\eta$ is the initial learning rate and is a small constant added for numerical stability. The training process involves several iterations over the data set (epochs) until the loss minimizes sufficiently, indicating optimal parameter estimation based on the network architecture and training data.

## 3. Methods and results

This section covers data set description, preprocessing, interpretability of the model, building, training, and evaluating the model.

### 3.1. Real data

It is more likely that Covid-19, an infectious disease, is caused by a novel type of Coronavirus. Typically, the majority of Covid-19 cases result in mild or moderate respiratory symptoms, and individuals can recover without requiring any specific medical intervention. It is more probable for serious illnesses to develop in older adults

and people with underlying medical problems. An analysis of a clinical data set for Covid-19 patients is conducted in this study[1].

There are 21 unique features in the data set and 1,048,576 unique patients. It contains patient identifiers such as age, gender, classification, type of patient, pneumonia, pregnancy, diabetes, copd, asthma, inmsupr, hypertension, cardiovascular, renal chronic diseases, other diseases, obesity, tobacco use, usmr, medical unit, intubated, ICU, and date of death. A summary of each patient's death date appears in the summary column, and if the patient is alive it means until the date the data set is downloaded. Thus, it will calculate how long it will take for a person to die or remain alive. An overview of the data set is provided in Table 2, which provides information about the data set.

**Table 2**
Features and its values

| Features | Values | Features | Values | Features | Values |
|---|---|---|---|---|---|
| Sex | 1: Female<br>0: Male | copd | 1: Yes<br>0: No | usmr | Indicates whether the patient was treated in medical units of the first, second or third level. |
| ICU | 1: Yes<br>0: No | asthma | 1: Yes<br>0: No | medical unit | Type of institution of the National Health System that provided the care. |
| Intubed | 1: Yes<br>0: No | inmsupr | 1: Yes<br>0: No | classification | Values 1–3 mean that the patient was diagnosed with Covid in different degrees. Four or higher means that the patient is not a carrier of Covid or that the test is inconclusive. |
| Tobacco | 1: Yes<br>0: No | hypertension | 1: Yes<br>0: No | patient type | 1: The patient returned home<br>0: the patient is hospitalized. |
| Pneumonia | 1: Yes<br>0: No | cardiovascular | 1: Yes<br>0: No | | If the patient died, indicate the date of death, and 9999-99-99 otherwise. |
| Pregnancy | 1: Yes<br>0: No | renal_chronic | 1: Yes<br>0: No | Status | 1: The patient was dead<br>0: The patient was censored |
| Diabetes | 1: Yes<br>0: No | other_disease | 1: Yes<br>0: No | – | – |
| Obesity | 1: Yes<br>0: No | age | Age of the patient | – | – |

[1]https://datos.gob.mx/busca/dataset/informacion-referente-a-casos-covid-19-en-mexico

To indicate whether a patient is dead or alive, we include a death column in our records. If the value entered in this column is "9999-99-99" it shows that the patient is currently alive. On the other hand, any other value entered here signifies that the patient has passed away. Table 3 outlines a summary of multiple statistics, which includes the count of individuals or participants, number of features, rate of censoring, percentage of missing values, as well as the target variables and failure event of interest. Moreover, Table 4 shows the percentage of missing values for each feature.

**Table 3**
Statistics of the gataset

| Size | Features | Train Samples | Test Samples | Target Variables | | Failure Event |
|---|---|---|---|---|---|---|
| | | | | Death/ Survival | Time to Event | |
| 1,025,152 | 21 | 817,581 | 207,571 | Death or survival | Until occurrence of death or survival | COVID-19 caused death |

**Table 4**
Percentage of missing values for each feature

| | | | | | |
|---|---|---|---|---|---|
| Sex | 0.00% | Obesity | 0.29% | Age | 0.03% |
| ICU | 81.64% | Asthma | 0.28% | Classification | 0.00% |
| Intubed | 81.62% | inmsupr | 0.32% | Patient type | 0.00% |
| Tobacco | 0.31% | Hypertension | 0.30% | Medical unit | 0.00% |
| Pneumonia | 1.53% | Cardiovascular | 0.29% | usmr | 0.00% |
| Pregnancy | 50.3% | Renal chronic | 0.29% | Date died | 0.00% |
| COPD | 0.29% | Obesity | 0.29% | | |
| Diabetes | 0.32% | Other disease | 0.48% | | |

## 3.2. Data pre-processing

To prepare a data set suitable for algorithms, data pre-processing is carried out. To begin, conduct data cleaning and pre-processing procedures involve checking for any missing data and removing columns and rows that are irrelevant or contain numerous null values. Additionally, any rows with missing data in specific columns are dropped. Finally, the data is normalized so that it falls within a range of 0 to 1. Firstly, null values were checked and fortunately none were found. The data set consists of one numeric variable, "AGE," two categorical variables, and 17 boolean variables. The original data set had 21 columns, but irrelevant columns and those with numerous null values were removed through cleaning and pre-processing, leaving 18 features. Addi-

tionally, the "Date died" column contained 971,633 values of "9999-99-99," indicating that the individuals are still alive, and missing values were removed from all features except "Intubed," "Pregnant," and "ICU." Due to numerous missing values, these features were dropped. Some features were expected to have only two unique values but had three or four unique values, with 99 representing NaN values. Therefore, only rows with one or two values were retained, reducing the DataFrame length from 1,048,575 to 1,025,152. The data was normalized to a range of [0,1] to rescale and change the distribution of the data. Additionally, resampling was performed using the resample method from sklearn.utils to address the issue of the imbalanced distribution of values. Finally, feature extraction was used in the pre-processing step of data to transform raw data into a representation. The performance of a model heavily depends on the quality of the features used as inputs. That is why, by selecting and transforming the most relevant features, feature extraction helps to reduce noise and irrelevant information, thereby improving the accuracy and efficiency of models.

### 3.3. Interpretability of the model

Among the many explainability techniques available, we chose SHAP and LIME due to their advantageous characteristics for our study. In this study, we used a clinical data set for Covid-19 patients. For the DAW model the concordance index (Antolini, L., et al. [2]) is 0.9699 in the training data and 0.89015 in the test data. Also, the integrated Brier score [7] and integrate binomial log-likelihood value [4–6] are 0.0035 and 0.00989, respectively. Using the SHAP algorithm, we evaluate each feature's contribution to the difference between the final prediction and the average prediction. The steps to calculate the importance of feature $i$ (where $i$ is the feature index) are as follows:

1. Compute all subsets that exclude feature $i$.
2. Determine the marginal contribution of feature $i$ in each subset.
3. Aggregate all marginal contributions to compute the overall contribution of feature $i$.

To find Shapley values with SHAP, we simply input our trained model into 'shap.KernelExplainer'. We can use a SHAP summary plot to visualize the overall impact of these features across multiple instances. For each feature, each point on the plot represents an individual patient. The position of a point along the $x$-axis (i.e., the actual SHAP value) indicates the effect of that feature on the model output for that specific patient. Mathematically, this relates to the probability of survival compared to other patients. For example, a patient with a higher SHAP value has a higher risk of death than one with a lower SHAP value. Features are arranged along the $y$-axis based on their importance, as indicated by their average absolute Shapley values. The higher a feature is on the graph, the more significant it is to the model. Figure 2 presents their summary plots. Each point represents an instance in the data set (e.g., a patient). The position along the $x$-axis (i.e., the actual SHAP value) shows the effect of each feature on the model output for that particular patient. The

wider the spread of SHAP values for a feature, the more impact it has on the model's output. Features are arranged along the $y$-axis based on their importance, as determined by the average of their absolute Shapley values (a higher position indicates greater importance). The color represents the value of the features. Red indicates high feature values, while blue indicates low feature values.

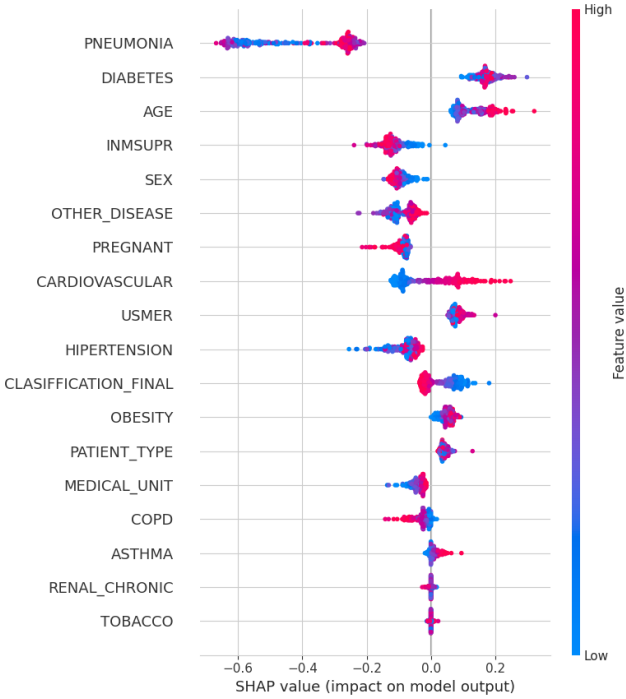



**Figure 2.** Summary plots of SHAP values for each feature

As we can see, for the pneumonia feature, high values (red) generally push the model prediction higher (positive SHAP values). For the diabetes feature, both high and low values have a significant impact, with high values pushing predictions higher and low values lowering predictions. Low feature values (blue) push the prediction lower. Older age (red) tends to increase the prediction, while younger age (blue) tends to decrease it. For the immune suppression (INMSUPR) feature, higher values significantly increase the model's prediction, while lower values decrease it. The impact of gender varies but generally shows that gender differences influence the prediction. The presence of other diseases tends to increase the prediction (red areas). Pregnancy has a notable impact, with high feature values (pregnancy) increasing the prediction. Cardiovascular disease presence pushes the prediction higher. For the USMER feature the importance is shown, but the plot indicates a relatively lesser impact compared to others. High values (presence of hypertension) increase the prediction. For the

final classification there is a significant spread, notably indicating classification impact predictions. Obesity tends to increase the prediction. Various patient types impact the prediction differently, with high feature values pushing predictions up. The type of medical unit has a noticeable impact, although less than some other features. Chronic obstructive pulmonary disease (COPD) increases the prediction. The impact of asthma on prediction is significant but less than other diseases. Chronic renal disease increases prediction, with high values (red) pushing the prediction up. Tobacco use shows a varied impact, with high use generally increasing the prediction.

As we can see in Figure 3, The features that significantly increase the prediction are pneumonia, diabetes, age, and immunosuppression. Sex, other diseases, and pregnancy are the most substantial factors that increase the probability of the predicted outcome. Also, features such as COPD, renal chronic, and tobacco use have little to no effect on the prediction. We calculated the SHAP values for the DAW model based on a given random partition.

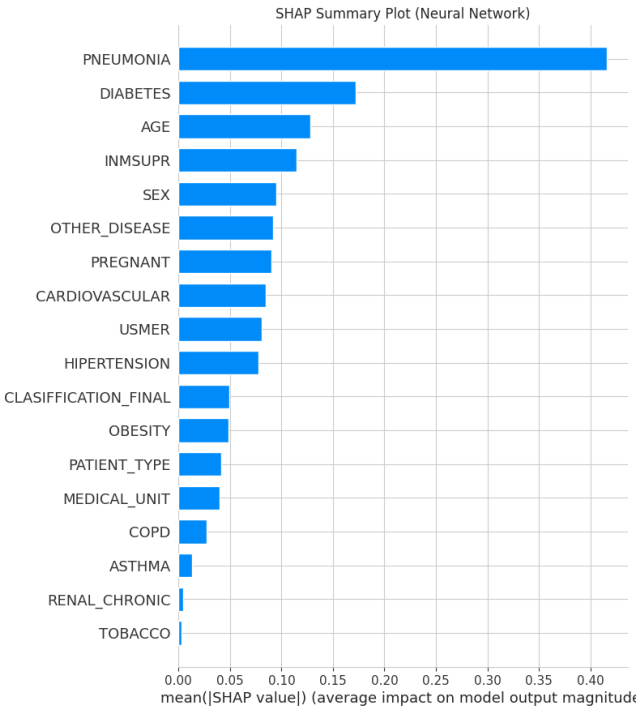
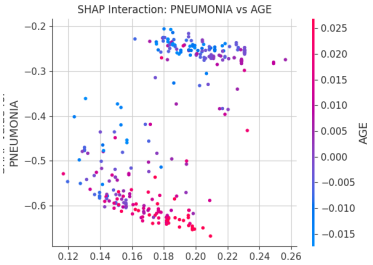
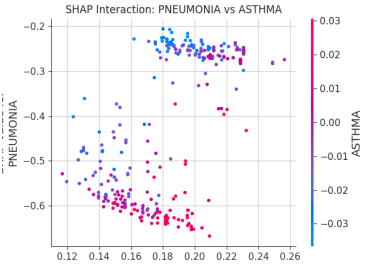
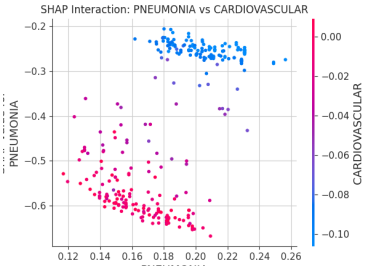


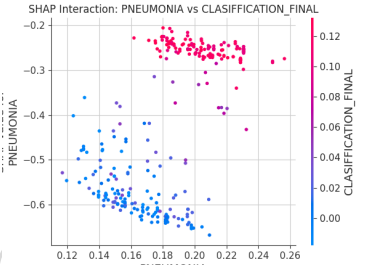

**Figure 3.** SHAP summary plot

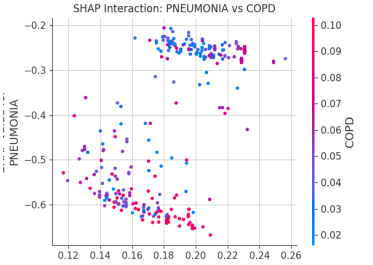
Figure 4, which is related to the different features, shows the impact of each feature on the model prediction and their interaction with the pneumonia feature.

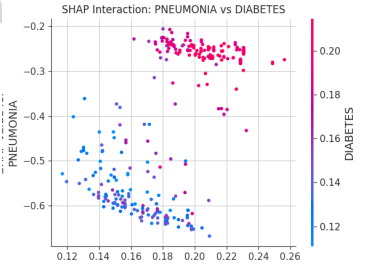

SHAP Interaction: PNEUMONIA vs AGE
PNEUMONIA
AGE

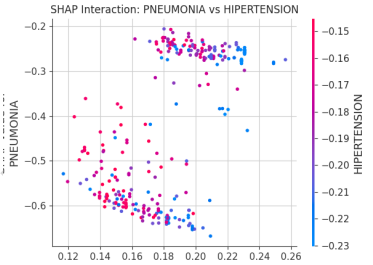

SHAP Interaction: PNEUMONIA vs ASTHMA
PNEUMONIA
ASTHMA

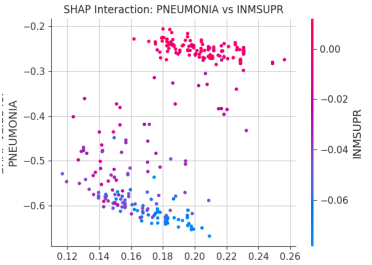

SHAP Interaction: PNEUMONIA vs CARDIOVASCULAR
PNEUMONIA
CARDIOVASCULAR

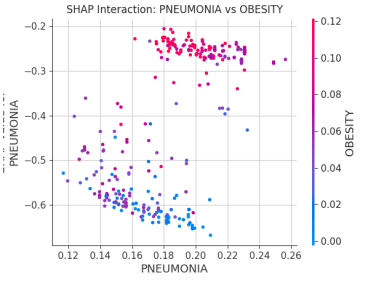

SHAP Interaction: PNEUMONIA vs CLASIFFICATION_FINAL
PNEUMONIA
CLASIFFICATION_FINAL

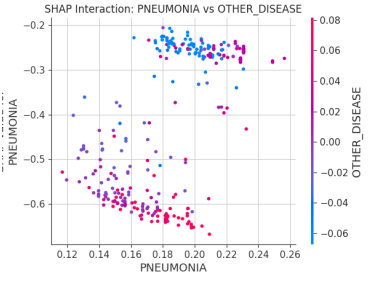

SHAP Interaction: PNEUMONIA vs COPD
PNEUMONIA
COPD


SHAP Interaction: PNEUMONIA vs DIABETES
PNEUMONIA
DIABETES


SHAP Interaction: PNEUMONIA vs HIPERTENSION
PNEUMONIA
HIPERTENSION


SHAP Interaction: PNEUMONIA vs INMSUPR
PNEUMONIA
INMSUPR


SHAP Interaction: PNEUMONIA vs OBESITY
PNEUMONIA
OBESITY
PNEUMONIA


SHAP Interaction: PNEUMONIA vs OTHER_DISEASE
PNEUMONIA
OTHER_DISEASE
PNEUMONIA

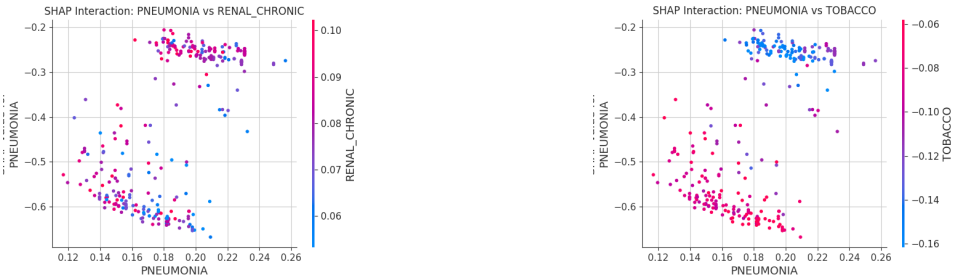



**Figure 4.** The effect of the PNEUMONIA in interaction with other features

The SHAP plots show that with increasing age, the negative impact of pneumonia on the model prediction increases, such that older individuals are at greater risk of developing pneumonia compared to younger individuals. Individuals show more negative SHAP values if they have diabetes and pneumonia, indicating worse health outcomes in these conditions. Chronic obstructive pulmonary disease and asthma also increase the negative impact of pneumonia, with individuals with these conditions experiencing worse outcomes if they develop pneumonia.

INMSUPR is another factor that leads to more negative SHAP values in the presence of pneumonia, indicating a stronger negative impact on health outcomes. Hypertension is also associated with worse outcomes from pneumonia, with people with high blood pressure showing more negative SHAP values when they have pneumonia. The presence of other underlying diseases such as cardiovascular disease and chronic kidney disease also enhances the negative effect of pneumonia. Obesity, especially in people with pneumonia, increases the risk. Tobacco use is also associated with negative SHAP values, indicating that people who smoke or use tobacco experience more severe effects from pneumonia. Higher scores in the final classification are associated with more negative SHAP values, indicating that people with more severe health problems do worse in the presence of pneumonia. Overall, across all SHAP plots, it is clear that pneumonia interacts negatively with a range of health conditions. People with chronic conditions such as diabetes, COPD, kidney disease, asthma, cardiovascular disease, and those with weakened immune systems are at significantly higher risk if they develop pneumonia. Older age, obesity, and tobacco use also exacerbate the negative impact of pneumonia on health outcomes.

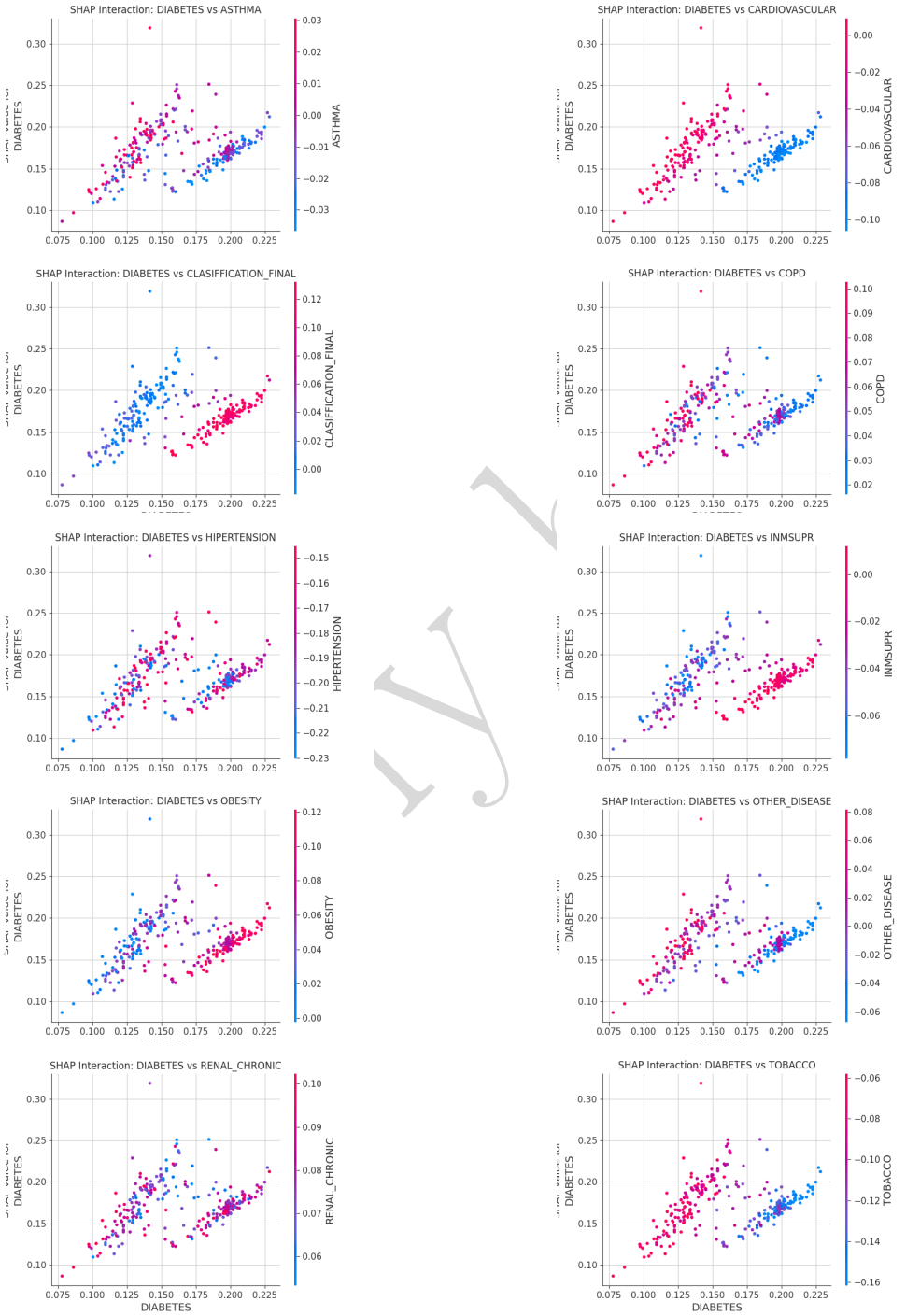
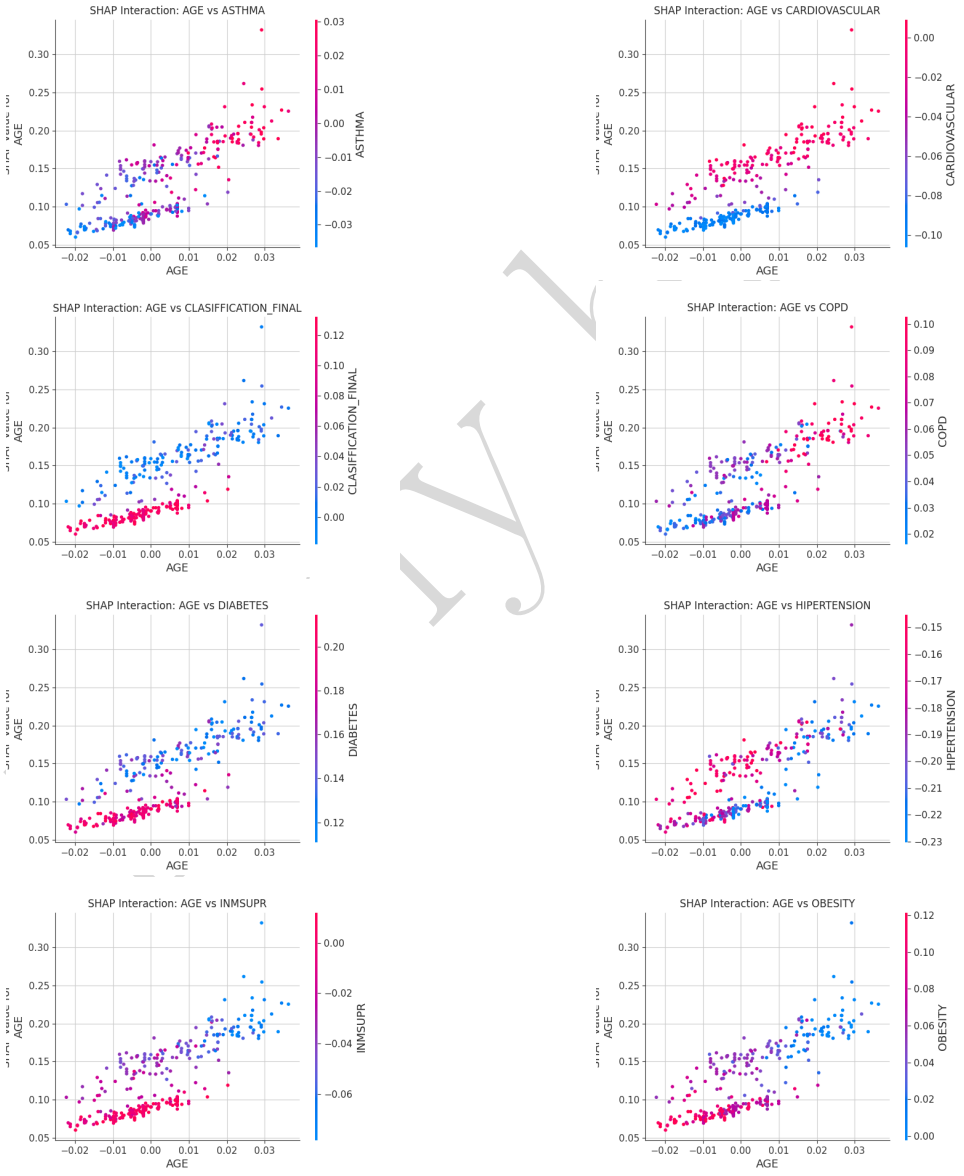



**Figure 5.** The effect of diabetes in interaction with other features

In Figure 6, the effect of age on the model's prediction is examined and how this feature interacts with other features, including diabetes, asthma, etc. In the analysis of SHAP plots, it was found that increasing age acts as a significant factor in increasing the probability of death, especially in people with certain conditions.



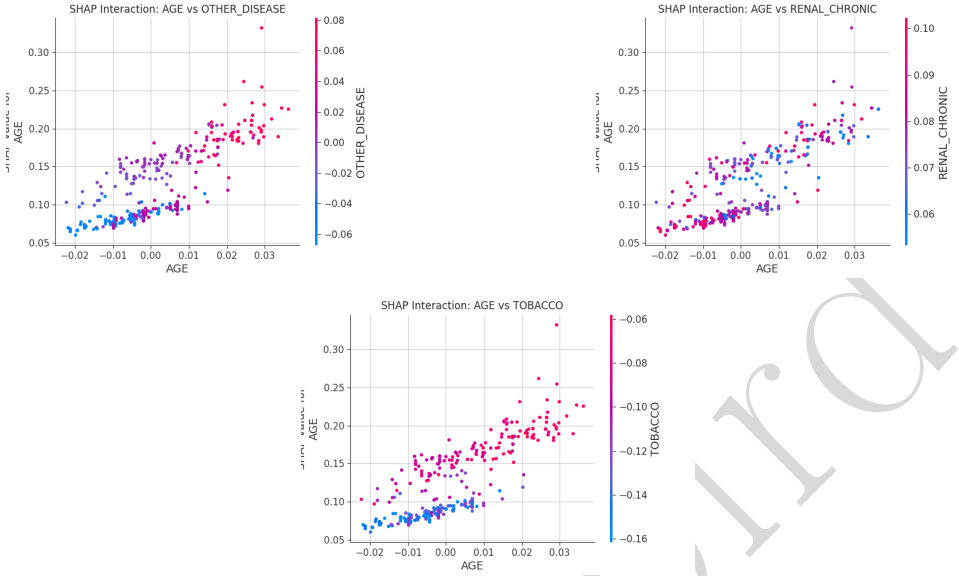



**Figure 6.** The effect of Age in interaction with other features

Renal chronic disease and asthma also exacerbate the negative effect of increasing age on the probability of death, and people with these diseases face a higher risk as they age. A weak immune system and high blood pressure also interact with increasing age to increase the probability of death. People with chronic kidney disease experience a higher probability of death as they age than people without this disease. Tobacco use, along with increasing age, acts as a serious risk factor in predicting death; people who use tobacco are at a higher risk as they age. In general, increasing age, in conjunction with chronic diseases and high-risk behaviors such as tobacco use, significantly increases the likelihood of death. In contrast, for people without these diseases or behaviors the impact of age on death is less. These results suggest that managing chronic diseases and adopting a healthy lifestyle can reduce the negative impact of aging on survival.

Figure 7 examines the effect of the INMSUPR along with other features on the model output. In the analysis of the SHAP plots, it was found that INMSUPR in combination with other diseases and conditions significantly increases the risk of death or worsening of an individual's condition. In the case of cardiovascular disease, blue colors indicate more severe heart disease. The combination of heart disease with immunodeficiency (blue dots) reduces the SHAP value, meaning an increased risk of worsening an individual's condition. The combination of obesity with immunodeficiency can increase the risk of worsening an individual's condition, while the dots to the right and blue indicate less obesity and lower risk.

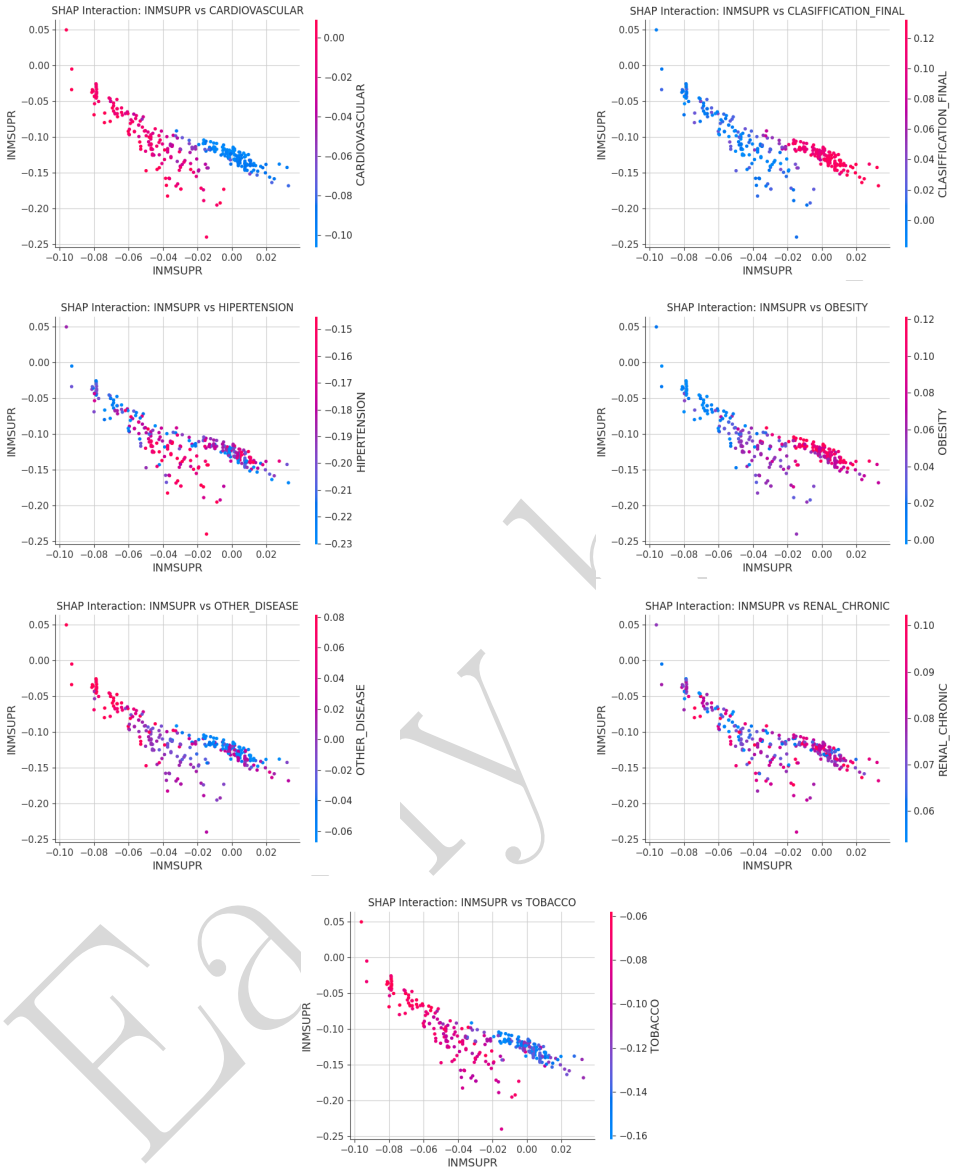



**Figure 7.** The effect of the INMSUPR in interaction with other features

Renal chronic disease, combined with immunodeficiency, increases the risk of death or worsening of an individual's condition. In the context of tobacco use, people with immunodeficiency and tobacco use (blue dots) have lower SHAP values, indicating an increased risk. Finally, in the final classification, people with weakened immune

systems and higher classifications (blue dots) are more likely to have a worsening condition. Since the size of the features is large, only the most important features were considered to examine the interaction of features.

## 4. Conclusion

Traditional machine learning models face challenges in conducting time-to-event analyzes due to censoring. This study introduces the DAW model that leverages deep learning techniques for the survival analysis of right-censored Covid-19 patient data. The SHAP value analysis reveals that certain features, such as pneumonia, age, diabetes, and immunosuppression, have a significant and consistent impact on the model's prediction. These factors generally increase the predicted risk or outcome. On the other hand, features such as pregnancy, COPD, asthma, and hypertension often show a negative correlation, suggesting a protective or nuanced role in the model. The interactions between these features further highlight the complexity of the model's decision-making process, indicating that the predictions are influenced by a combination of factors rather than isolated variables. This holistic understanding of feature interactions and their contributions to model prediction can guide further refinements in the model and inform healthcare interventions targeted at high-risk groups.

## References

[1] Aggarwal C.C.: *Neural networks and deep learning*, vol. 10, Springer, Cham, 2018. doi: 10.1007/978-3-031-29642-0.

[2] Antolini L., *et al.*: A time-dependent discrimination index for survival data, *Statistics in Medicine*, vol. 24(24), pp. 3927–3944, 2005. doi: 10.1002/sim.2427.

[3] Duckworth C., Chmiel F.P., Burns D.K., Zlatev Z.D., White N.M., Daniels T.W., *et al.*: Using explainable machine learning to characterise data drift and detect emergent health risks for emergency department admissions during COVID-19, *Scientific Reports*, vol. 11(1), p. 23017, 2021. doi: 10.1038/s41598-021-02481-y.

[4] Gerds T.A., Schumacher M.: Consistent estimation of the expected Brier score in general survival models with right-censored event times, *Biometrical Journal*, vol. 48(6), pp. 1029–1040, 2006. doi: 10.1002/bimj.200610301.

[5] Goodfellow I., Bengio Y., Courville A.: *Deep learning*, MIT Press, 2016.

[6] Graf E., Schmoor C., Sauerbrei W., Schumacher M.: Assessment and comparison of prognostic classification schemes for survival data, *Statistics in Medicine*, vol. 18(17–18), pp. 2529–2545, 1999. doi: 10.1002/(sici)1097-0258(19990915/30)18:17/18<2529::aid-sim274>3.0.co;2-5.

[7] Harrell F.E., Califf R.M., Pryor D.B., Lee K.L., Rosati R.A.: Evaluating the yield of medical tests, *JAMA*, vol. 247(18), pp. 2543–2546, 1982. doi: 10.1001/jama.1982.03320430047030.

[8] Hu C., Li L., Li Y., Wang F., Hu B., Peng Z.: Explainable machine-learning model for prediction of in-hospital mortality in septic patients requiring intensive care unit readmission, *Infectious Diseases and Therapy*, vol. 11(4), pp. 1695–1713, 2022. doi: 10.1007/s40121-022-00671-3.

[9] Kramer M.A.: Nonlinear principal component analysis using autoassociative neural networks, *AIChE Journal*, vol. 37(2), pp. 233–243, 1991. doi: 10.1002/aic.690370209.

[10] Lone S.A., Unal C., Cekim H.O.: Evaluation of different estimator approaches based on COVID-19 data and a simulation study, *Alexandria Engineering Journal*, vol. 107, pp. 174–183, 2024. doi: 10.1016/j.aej.2024.06.098.

[11] Lundberg S.M., Erion G.G., Lee S.I.: Consistent individualized feature attribution for tree ensembles, *arXiv preprint arXiv:180203888*, 2018.

[12] Lundberg S.M., Lee S.I.: A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[13] Lundberg S.M., Nair B., Vavilala M.S., Horibe M., Eisses M.J., Adams T., *et al.*: Explainable machine learning predictions to help anesthesiologists prevent hypoxemia during surgery, *BioRxiv*, 206540, 2017. doi: 10.1101/206540.

[14] Moncada-Torres A., van Maaren M.C., Hendriks M.P., Siesling S., Geleijnse G.: Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival, *Scientific Reports*, vol. 11(1), 6968, 2021. doi: 10.1038/s41598-021-86327-7.

[15] Rokach L., Maimon O., Shmueli E. (eds.): *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*, Springer Nature, 2023. doi: 10.1007/978-3-031-24628-9.

[16] Sidey-Gibbons J.A., Sidey-Gibbons C.J.: Machine learning in medicine: a practical introduction, *BMC Medical Research Methodology*, vol. 19, pp. 1–18, 2019. doi: 10.1186/s12874-019-0681-4.

[17] Sindhu T.N., Çolak A.B., Lone S.A., Shafiq A., Abushal T.A.: A decreasing failure rate model with a novel approach to enhance the artificial neural network's structure for engineering and disease data analysis, *Tribology International*, vol. 192, p. 109231, 2024. doi: 10.1016/j.triboint.2023.109231.

[18] Xie M., Lai C.D.: Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function, *Reliability Engineering and System Safety*, vol. 52(1), pp. 87–93, 1996. doi: 10.1016/0951-8320(95)00149-2.

[19] Zhang Z., Beck M.W., Winkler D.A., Huang B., Sibanda W., Goyal H.: Opening the black box of neural networks: methods for interpreting neural network models in clinical applications, *Annals of Translational Medicine*, vol. 6(11), 2018. doi: 10.21037/atm.2018.05.32.

## Affiliations

**Nasrin Moradi**
Ferdowsi University of Mashhad, Department of Statistics, School of Mathematical Sciences, Moradi.nasrin06@gmail.com

**Arezou Habibirad**
Ferdowsi University of Mashhad, Department of Statistics, School of Mathematical Sciences, ahabibi@um.ac.ir (corresponding author)

**Hanieh Panahi**
Islamic Azad University, Department of Mathematics and Statistics, Lahijan Branch, Lahijan, Iran, h.panahi@yahoo.com and panahi@liau.ac.ir

**Received**: 20.01.2025
**Revised**: 25.02.2025
**Accepted**: 15.08.2025