

# Risk Factor Hypertension Prediction Model

Vercus Ntirandekura, Jeremie Ndikumagenge

Center of Research in Infrastructure, Environment and Technology (CRIET), University of Burundi, Bujumbura, Burundi

Email: ntirandekuraczi@gmail.com, jeremie.ndikumagenge@ub.edu.bi

**How to cite this paper:** Ntirandekura, V. and Ndikumagenge, J. (2023) Risk Factor Hypertension Prediction Model. *Journal of Applied Mathematics and Physics*, 11, 428-437.

<https://doi.org/10.4236/jamp.2023.112025>

**Received:** December 15, 2022

**Accepted:** February 10, 2023

**Published:** February 13, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

According to the 2020 Ministry of Health reports, the public health sector is facing an acute shortage of logistical resources and qualified competent human resources, as evidenced by the doctor-to-hospital ratio in relation to population [1]. Aside from these structural and cyclical issues, the above ratios are even lower in rural areas with low incomes. Underdevelopment is a major impediment to establishing a normal public health situation, though the Burundian government is working hard to ensure that it is at an acceptable level. Furthermore, some Burundian traditions, customs, and practices are undermining efforts to build an international-standard public health facility. Indeed, the mental state of a people (tradition, culture, and practices) has a significant impact on the fluctuation of risk factors in public health. It is determined by the socioeconomic development and sociocultural behavior of the population. This demonstrates that hypertension is a public health concern in Burundi. Unfortunately, the vast majority of people are completely unaware of the risks that high blood pressure poses to public health. High blood pressure, on the other hand, has always been a key physiological measure in medical examinations, serving as one of the most important biological markers in clinical evaluation. As a result, cardiovascular diseases caused by high blood pressure have a significant impact on mortality worldwide, particularly in Burundi. Predicting high blood pressure based on risk factors can help to reduce complications associated with this disease, which is known as a silent killer. The digital era provides a variety of tools for studying, analyzing, managing, and monitoring the risk factors that contribute to and degenerate high blood pressure. The primary goal of this work is to create a decision-making tool based on the outcomes of high blood pressure epidemic and/or pandemic predictions from sanitarian districts. The current paper work employs a prediction support tool created using linear regression methods from machine learning, one of the fields of artificial intelligence. It is especially useful for optimizing the cost function. The latter allows the predicted values to be determined and defined using the gradient descent algorithm.

## Keywords

Artificial Intelligence, Machine Learning, Prediction, Gradient Descent, Hypertension, Cardiovascular Disease

---

## 1. Introduction

Public health issues have been critical to humanity since the Neolithic era. The current era is no exception to this rule. It even reveals new problems and fragilities that modern societies must face [2].

Periods of public health crisis that the world has experienced since the 1900s marked by epidemics like the Spanish flu, the Ebola virus, the corona virus and the periods marked by recurrent famines since the 1950s, including those experienced by Ethiopia in the 1980s, recall the urgent need to redesign and reframe public health aspects.

Since the invention of vaccines and other useful means in the field of public health, WHO reports on national public health policies and universal health coverage in 2003 show that innovations in this sector have gone through both bottom-up and top-down phases.

Indeed, various methods and techniques have been used to improve, expand, and sharpen health-care services. Campaigns for mass screening, vaccination, and curative treatment are examples of how to improve community health around the world. The integration of information and communication technologies, as well as their effective use in the public health sector, provides valuable insights into the prediction of epidemics and/or pandemics in both developed and developing countries.

In other words, the use of statistical methods and tools for estimating and forecasting, such as statistical and actuarial software, generally provides data and information that does not allow for flexibility in making timely decisions. The gap in these tools results in highly estimated decision-making using either optimistic or pessimistic approaches as appropriate. However, underdeveloped countries and/or certain developing countries lag behind in the intensive use of New Information and Communication Technologies in this sector.

Burundi, like the countries of the African Region, has a health profile marked by a double burden of mortality related to communicable and non-communicable diseases including epidemics and pandemics. The country has experienced a series of epidemics as evidenced by the 2021 WHO report on Burundi. In addition to these health emergencies with high epidemic potential, natural disasters such as floods have worsened the health conditions of the population.

Moreover, the latter has a glaring lack of information on the problems and threats to community health, of which high blood pressure occupies a substantial share for the Burundian population. Hypertension is one of the invasive diseases of human life worldwide. It caused 9.4 million deaths in 2014 and is the

second most common cause of cardiovascular disease after Diabetes [3].

High blood pressure has always been a key physiological measure in medical examination, being one of the most important biological markers in clinical evaluation. Predicting high blood pressure based on risk factors can help in the management of this deadly, non-communicable disease, particularly due to poor eating habits and lifestyle.

In Burundi, like any other developing country, suffers from a lack of logistical resources as well as qualified and competent human resources in the public health sector. Simultaneously, the digital era provides a variety of tools for the study, analysis, management, and monitoring of risk factors that commonly lead to high blood pressure. Artificial intelligence's achievements and exploits find applications in the medical and health fields, including diagnostic or screening support systems [4].

It is with this in mind that this article uses machine learning in the prediction of hypertension as a function of risk factors through the use of the prediction tool developed and designed in the article entitled "Development of a quantitative prediction support system using the linear regression method". The ultimate goal is to raise awareness among stakeholders about the risks of this public health disease so that they can take appropriate precautions; analyze the prevalence of cardiovascular disease in the specific sanitarian region; and allow public health managers to determine whether the disease is epidemic or pandemic.

## **2. Materials, Tools, Equipment and Methods**

### **2.1. Material**

The material is based on the values of the high blood pressure risk factors as well as the quantitative prediction support tool developed in previous work.

### **2.2. Tools and Equipment**

The Excel spreadsheet aids in summation calculation, whereas python language libraries such as numpy aid in numerical calculation when pandas is in charge of loading the model data. The matplotlib library makes data visualization easier for model analysis.

### **2.3. Methods**

The working method is the minimization of the cost function obtained after applying the linear regression method to the most influential risk factors using gradient descent.

## **3. Obtained Results**

### **3.1. Model Coefficient Values**

The following model parameter or coefficient values were obtained by solving the system of equations formed on the basis of data related to the risk factors of high blood pressure:

$$\begin{aligned}
 a &= 4.4910387 \\
 b &= 5.14996294 \\
 c &= -2.48922975 \\
 d &= 2.06523924 \\
 e &= -4.54950474 \\
 f &= 2.0970228 \\
 k &= -435
 \end{aligned}$$

This changes the model to

$$f(x_i) = 4.49x_1 + 5.14x_2 - 2.48x_3 + 2.06x_4 - 4.54x_5 + 2.09x_6 - 435$$

### 3.2. Correlation Matrix

After studying the model, we arrive at a linear model with two parameters  $X_2$  and  $X_5$ , which represent Body Mass Index and smoking level, respectively, chosen because of their significant influence on pathological elevation of blood pressure, as shown in the matrix of correlations in **Table 1**. From this, we ignore the least influential parameters to keep only those that are most found/given by the matrix of correlations and the model becomes:

$$f(x_i) = 5.14x_2 - 4.54x_5 - 435$$

### 3.3. Optimization of the Model by the Gradient Descent Method

In this part, it comes down to determining the values of the parameters which minimize the cost function giving subsequently a better model but also the predictive values using the descent gradient method.

#### 3.3.1. The Gradient Descent Method

The descent gradient method/algorithm allows to find the minimum of the cost function ( $a, b$ ) with  $a$  and  $b$  randomly selected coordinates.

The descent gradient method consists of three essential steps:

- 1) Calculate the slope of the cost function, *i.e.* the derivative of  $J(a, b)$ .
- 2) Move a certain distance  $\alpha$  in the direction of the steepest slope in order to change the values of parameters  $a$  and  $b$ .

**Table 1.** Representation of correlation coefficient calculation for the correlation matrix.

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$Y$
$x_1$	1	0.9573247	0.9725446	0.9972502	0.8473689	0.7987773	0.4312134
$x_2$	0.9573247	1	0.9214967	0.9470634	0.8889943	0.8257268	0.4748895
$x_3$	0.9725446	0.9214967	1	0.9690395	0.7218066	0.6623607	0.4269552
$x_4$	0.9972502	0.9470634	0.9690395	1	0.855431	0.8158909	0.4395764
$x_5$	0.8473689	0.8889943	0.7218066	0.855431	1	0.9871046	0.4685499
$x_6$	0.7987773	0.8257268	0.6623607	0.8158909	0.9871046	1	0.3878939
$Y$	0.4312134	0.4748895	0.4269552	0.4395764	0.4685499	0.3878939	1

3) Repeat steps 1 and 2 until you reach the minimum of  $J(a, b)$  [5].

### 3.3.2. Steps in the Mathematical Descent Gradient Method

1) First we have the linear model expressed in matrix form like this:

$$F(X) = X \cdot \theta \quad \text{where } \theta = \begin{pmatrix} a \\ b \end{pmatrix} \text{ to be given to the machine.}$$

2) Create Cost function

$$J(\theta) = \frac{1}{2m} \sum ((X) -)^2$$

3) The gradient is calculated using the formula:

$$\text{Gradient: } \frac{\partial J(\theta)}{\partial \theta} = \frac{1}{m} X^T \cdot (F(X) - Y)$$

4) We apply the gradient descent, that is, repeat in loop:

$$\theta = \theta - \alpha \frac{\partial J(\theta)}{\partial \theta}$$

### 3.3.3. Descent Gradient Implementation

The implementation of the gradient descent method can result from the following three functions:

1) Function for calculating the cost function

```
def fonction_cout(X,theta,Y):
    m=len(Y)
    return 1/(2*m)*np.sum((Model(X,theta)-Y)**2)
```

2) Function for gradient calculation

```
def grad(X,Y,theta):
    m=len(Y)
    return 1/m*X.T.dot(Model(X,theta)-Y)
```

3) Gradient descent application function

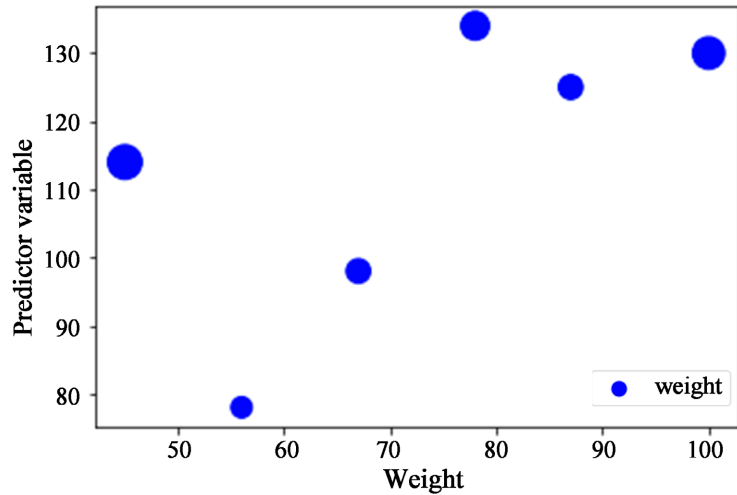
```
def descente_gradient(X,Y,theta,learning_rate,n_iterations):
    cost_history=np.zeros(n_iterations)
    for i in range(0, n_iterations):
        theta=theta-learning_rate*grad(X,Y,theta)
        cost_history[i]=fonction_cout(X,theta,Y)
    return theta, cost_history
```

### 3.4. Data Visualization

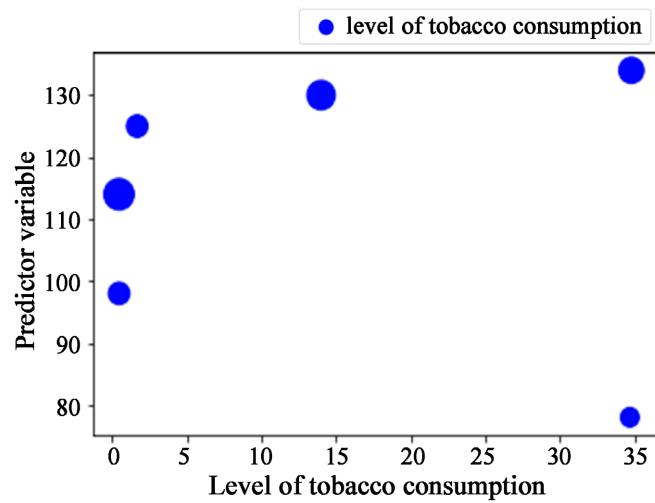
**Figure 1** and **Figure 2** show data visualization by drawing the weight factor and level of tobacco consumption factor representations in relation to the predictor variable, systolic pressure.

### 3.5. Representation of the Initial Model in Relation to the Dataset

This is the representation of the model with the initial parameter values which are:



**Figure 1.** Weight factor representation in relation to the predictor variable.



**Figure 2.** Level of tobacco consumption factor representation in relation to the predictor variable.

5.14,  
 -4.54,  
 -435.

**Figure 3** and **Figure 4** depict the representation of the final factors, weight and level tobacco, in relation to the initial model.

### 3.6. Final Parameter Values (Minimizing the Cost Function)

After dealing with the gradient descent method on the taken dataset, we found that the cost function will be minimal when the values of the following parameters are obtained:

6.53751825,  
 1.36684008,  
 -406.17480706

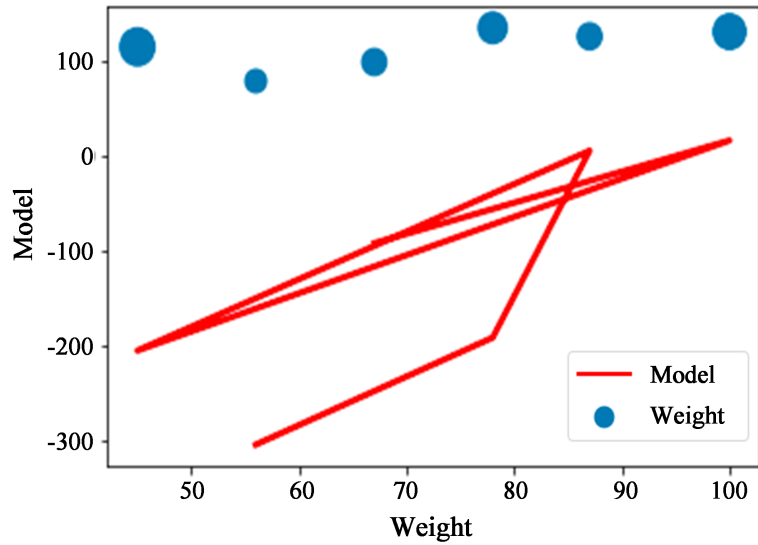


Figure 3. Initial model representation in relation to the weight factor.

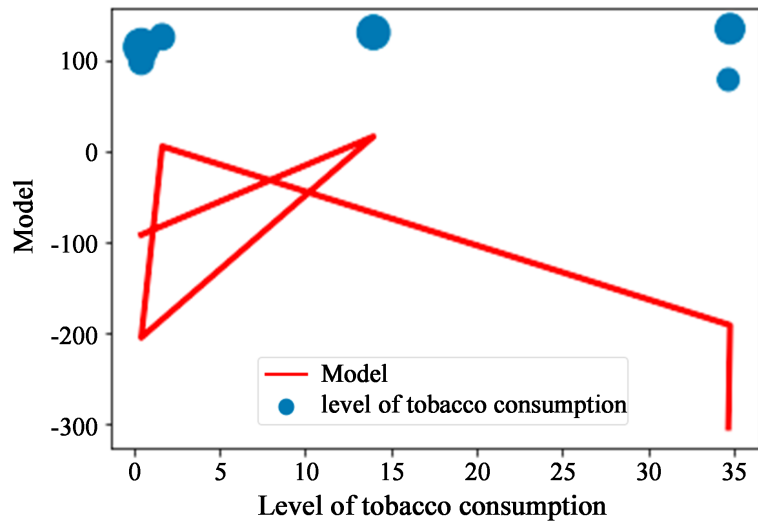


Figure 4. Initial model representation in relation of level tobacco consumption.

### 3.7. Representation of the Weight Factor in Relation to the Predicted Values

In the same vein as the previous results, we also found the following predictive values of systolic pressure that could lead to the onset of high blood pressure:

- 32.46219454,
- 266.71277871,
- 111.37140792,
- 164.87190333,
- 151.31765094,
- 7.35556551.

Figure 5 depicts the prediction representation in relation to the weight factor.

### 3.8. Representation of the Level of Tobacco Consumption Factor in Relation to the Predicted Values

Figure 6 depicts the prediction representation in relation to the level tobacco consumption factor.

### 3.9. Viewing the Model Learning Phase

Figure 7 shows the representation of the learning model obtained after 1000 iterations of the gradient descent algorithm with a learning rate of 0.0001. Depending on how this graphic is interpreted, the line representation will either take a regular direction or remain unchanged.

## 4. Discussions

- 1) The matrix of correlations shows that Body Mass Index and the level of

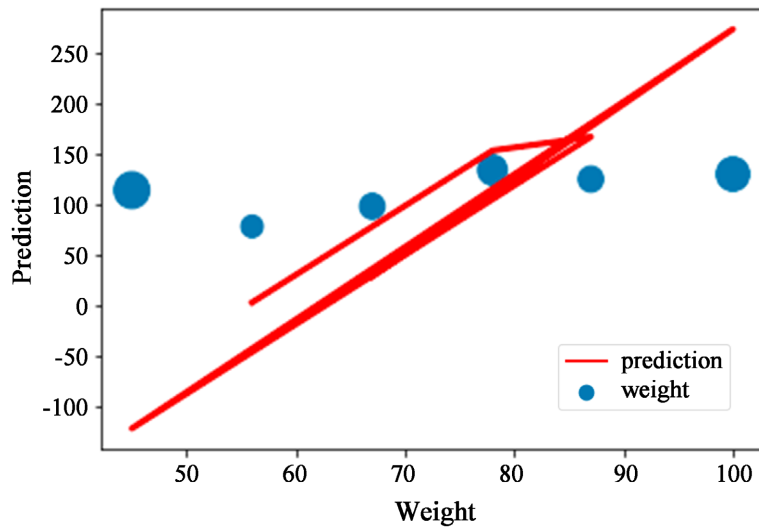


Figure 5. Weight factor representation in the relation to the predicted values.

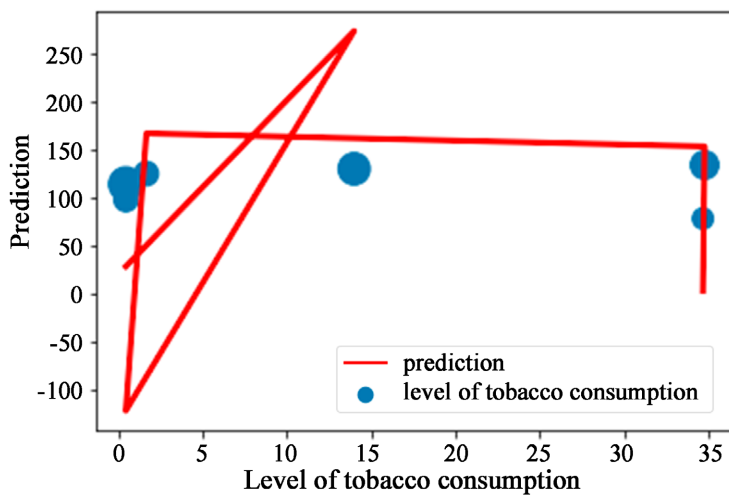
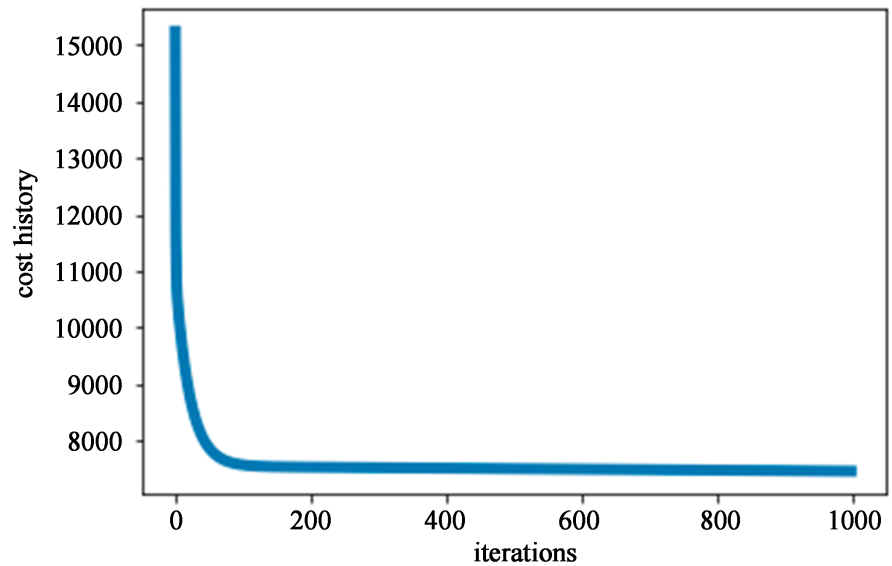


Figure 6. Level of tobacco consumption factor representation in the relation to the predicted values.





**Figure 7.** Learning Model representation.

smoking have a great influence on the pathological elevation of blood pressure.

The model identifies the most influential risk factors likely to cause high blood pressure based on data collected from health districts.

2) The assessment of the model results in a non-performing result. The performance of the machine learning program varies depending on the quantity or number of data learned and the quality.

3) Learning the model is based on the hyper-parameter also called Learning rate. The results show that the smaller the step, the better the model. As a result, the greater the pitch, the less efficient the model is. Ultimately, the model's performance depends on the number of iterations.

4) Compared to the graphs representing the factor values for the sample under consideration based on the obtained predictive values, we find that the model has a large bias.

## 5. Conclusions

Hypertension, a non-communicable disease but considered a silent killer threatens human lives and is a public health problem in Burundi.

The lack of public knowledge about the danger of this type of disease increases the pressure on the public health system, which unfortunately suffers from the lack of logistical resources and qualified and competent human resources.

Fortunately, the quantitative prediction support put in place in previous work applied to estimating high blood pressure based on risk factors is a tool to help public health actors make timely and relevant decisions.

The use of this tool in the field of public health would undoubtedly allow to control the occurrence of cases or phenomena related to high blood pressure due to the associated risk factors.

As a result, the model suffers from a more common problem in machine

learning called under fitting. To eliminate these errors, we are considering using the regularization method in our future work.

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] OMS (2021) Rapport annuel, Bureau de la Représentation de l’OMS au Burundi, bujumbura, 2021.
- [2] Ghemawat, P. (2011) Global Prosperity and How to Achieve It. Havard Business Review Press, Boston.
- [3] Zhao, D., Liu, J., *et al.* (2018) Epidemiology of Cardiovascular Disease in China: Current Features and Implications. *Nature Reviews Cardiology*, **16**, 203-212.  
<https://doi.org/10.1038/s41569-018-0119-4>
- [4] Yu, K.-H., Beam, A.L. and Kohane, I.S. (2018) Artificial Intelligence in Healthcare. *Nature Biomedical Engineering*, **2**, 719-731.  
<https://doi.org/10.1038/s41551-018-0305-z>
- [5] Ray, S. (2019) A Quick Review of Machine Learning Algorithms. *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 14-16 February 2019, Faridabad, 35-39.  
<https://doi.org/10.1109/COMITCon.2019.8862451>