

## **Cardiovascular Disease Prediction Using Electrocardiogram (ECG) and K-Plus Nearest Neighbors Algorithm: Cases of Chadian Patients**

KHADIDJA Ousman Kossi<sup>1</sup>, ALI Ouchar Cherif<sup>2</sup>, MAHAMAT Charfadine Nimane<sup>1</sup>, ABAKAR Mahamat Ahmat<sup>1</sup>

<sup>1</sup>National School of ICT, Ndjamena-Chad

<sup>2</sup>National Higher Institute of Sciences and Techniques of Abéché, Abéché-Chad

### **Summary**

This article reviews the use of the electrocardiogram (ECG) and the k-nearest neighbor (KNN) algorithm for the prediction of cardiovascular disease. Cardiovascular diseases are a major public health problem, accounting for a significant proportion of global deaths. The ECG offers a non-invasive method to monitor the electrical activity of the heart, detecting abnormalities and predicting risk. The KNN algorithm, a supervised machine learning technique, is used to classify the examples based on the labeled examples. Using pre-processed ECG data, KNN can recognize characteristic patterns of cardiovascular disease, enabling accurate and rapid prediction. This approach has significant medical potential, enabling early detection and informed decision-making. However, cardiovascular disease prediction is complex and evolving, requiring careful selection of attributes and rigorous evaluation of model performance. The article will include a review of prior designs, choice of study design, performance evaluation criteria, results, and an in-depth discussion of those results.

**Keywords:** *ECG, KNN, artificial intelligence, cardiovascular disease, prediction,*

### **Introduction**

Cardiovascular disease (CVD) represents a considerable public health burden worldwide. Early and accurate detection of these conditions is crucial to enable preventative medical intervention and improve patient outcomes. In this article, we examine in detail the use of the electrocardiogram (ECG) and the k-nearest neighbors (KNN) algorithm for the prediction of cardiovascular disease, based on scientific references and solid arguments. .

According to the World Health Organization (WHO), approximately 17.9 million deaths are caused each year by cardiovascular disease, accounting for nearly 31% of all deaths worldwide. [1] These diseases affect people of all age groups and countries, but their prevalence is particularly high in low- and middle-income countries. The main objective of this research is to explore the use of the KNN model in combination with ECG sensors for the prevention and classification of cardiovascular diseases.

The use of ECG sensors offers a non-invasive method to monitor the electrical activity of the heart, thus providing valuable information about its functioning. Analysis of ECG signals can help detect heart abnormalities and predict the risk of developing cardiovascular disease.

The k-nearest neighbor (KNN) algorithm is a supervised machine learning technique that can be used to classify new examples based on the already labeled examples. Using pre-processed ECG data, the KNN algorithm can be trained to recognize characteristic patterns associated with cardiovascular disease, thus enabling accurate and rapid prediction.

This approach has the potential to bring significant advantages in the medical field, by allowing early detection of cardiovascular diseases and facilitating medical decision-making. However, it should be noted that cardiovascular disease prediction is a complex and constantly evolving field, requiring careful selection of attributes and rigorous evaluation of model performance.

In this article, we will first make a state of the art of previous models and methods in order to choose the study model, then we will set the performance evaluation criteria for the validation of the studied model. Will come the results, which will be while evaluating them by the above-mentioned criteria and finally, a discussion of the results after an in-depth analysis.

## **I. Methods and materials**

By definition, ECG and cardiovascular disease is a commonly used technique to assess the electrical activity of the heart. It records the electrical signals emitted by the heart during each beat, providing essential information about heart function. Cardiovascular diseases, such as arrhythmias, myocardial ischemia, and heart disease, often manifest as detectable abnormalities on the ECG. Therefore, ECG analysis plays a crucial role in the detection and prediction of cardiovascular diseases [1].

And, the k-nearest neighbors (KNN) algorithm is a widely used supervised machine learning method for classification. It is based on the principle that similar samples tend to belong to the same class. In the context of cardiovascular disease prediction, features extracted from the ECG, such as interval duration and waveform amplitudes, can be used as parameters to train the KNN model. The algorithm evaluates the k closest samples in the feature space and assigns a class to the new sample based on the majority of neighbor classes.

### **1. State of the art**

With technological advancements in AI, data processing is very fast, so the need for real-time information becomes a major concern. AI is part of these advances and is today at the center of gravity of several fields, such as agriculture, intrusion detection, health, and many others; in this article we are interested in health, in particular cardiovascular diseases. The electrocardiogram (ECG) is a non-invasive medical tool that displays the rhythm and condition of the heart. Therefore, automatic detection of irregular heart rhythms from ECG signals is an important task in the field of cardiology [2]. It finds the distance between an unknown and all training data, selects the specified number of examples (K) closest to the query, and then votes for the very frequent one.

Several methods have been presented in the literature for the classification of cardiovascular diseases using the ECG sensor. First, we summarized the literature [6], Have for goal to detect heart disease, using several supervised machine learning algorithms and find a better accuracy of 95.71% with [8], proposed a heart disease prediction system using the k-nearest neighbors supervised learning algorithm (KNN) and KNN algorithm has been found to have the highest accuracy rate reaching 92.86% .[5],introduce an automatic ECG signal classification system, applied Deep Learning (DL) model to classify four kinds of ECG signals, hence it has 92.5% accuracy. Gutierrez, DMJ Et al., in [4], put a federated learning paradigm on heterogeneous and partitioned sets of high definition electrocardiograms from twelve (12) lead ECG sensor arrays to train AI models a applied, their approach obtained an accuracy close to 97.5%. Thus, we have set up a real-time prediction system for cardiovascular patients in order to intervene and manage them quickly and efficiently. Cardiovascular diseases are a major health problem worldwide, requiring reliable early detection and prediction methods. The use of the electrocardiogram (ECG) combined with machine learning algorithms offers new perspectives for the prediction of

cardiovascular diseases. This article focuses on the use of ECG and the KNN (K-Nearest Neighbors) algorithm for the prediction of these diseases, hence the accuracy is 97.5%.

## 2. Methodology used

Using ECG and KNN Algorithm for CVD Prediction: To predict cardiovascular disease using ECG and KNN algorithm, several steps should be followed:

### 2.1. Data gathering

Data collection is done on the basis of ECG records of patients with or without cardiovascular diseases are collected from the database **Chadian hospitals**. This dataset is from 1988 which was updated on May 28, 2023, contains 76 attributes including the predicted attribute, but all published experiments refer to using a subset of 14 of between them. The "target" field refers to the presence of heart disease in the patient. It has an integer value: 0: no disease and 1: disease.

This step consists of gathering the data necessary for training the model in a consolidated way, so that they are all contained in a single table.

In our study we use the dataset which contains 14 attributes, and 1025 patients. The Chadian hospital database is the only one used in our case.

	Age	sex	TAR	trestbps	chol	glycemia	restecg	FCMA	exang	Depression	slope	NNP	thal	cardio
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	2	1
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0
1022	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0
1023	50	0	0	110	254	0	0	159	0	0.0	2	0	2	1
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	3	0

1025 rows x 14 columns

**Figure 1:** The form of patient data

**Painting1:** shows the clinical characteristics and their description (Ministry of Public Health of Chad)

Numbers	Attributes	Description	Kind
1	Age	Patient's age (29 to 77)	digital
2	Sex	The gender of the person (0: female, 1: male)	categorical
3	ART	The type of chest pain is classified into four types 0: pain caused by typical angina, 1: caused by atypical angina, 2: non-blocking pain 3: asymptomatic	categorical
4	Trestbps	Resting blood pressure (in mm Hg at hospital admission)	digital
5	Hol	Serum cholesterol in mg/dL(126 to 564)	digital
6	blood sugar	Fasting blood glucose >120 mg/dL (1: true, 0: false)	categorical
7	Restecg	Electrocardiographic result at rest (0 to 2)	categorical
8	FCMA	The maximum heart rate of the affected person (71 to 202)	digital
9	Exang	Exercise-induced angina. Is a condition where not enough blood is supplied to the walls of the heart to pump blood. It is caused by exercise or any physical or mental stress (True: 1, False: 0)	categorical

<b>10</b>	<b>depression</b>	exercise-induced versus resting ('ST' refers to positions on the ECG graph.)	Entire
<b>11</b>	<b>slope</b>	the slope of the ST segment peak exercise (0 to 1)	categorical
<b>12</b>	<b>NNP</b>	The number of large vessels (0 to 3)	digital
<b>13</b>	<b>Thal</b>	A blood disorder called thalassemia	categorical
<b>14</b>	<b>cardio</b>	presence or absence of heart disease (1:True or 0:False)	categorical

## 2.2. Data preprocessing

ECG recordings go through a pre-processing process to remove artifacts and extract relevant features. This may include waveform detection, heart rate extraction and other specific measurements.

ECG signals are preprocessed to remove artifacts and normalized to ensure data consistency. Relevant features are extracted from ECG signals to represent the key information of each recording. This may include measurements of QT interval, PR interval, wave amplitude. First, we need to clean the data set that we will use for the training model. We import the necessary libraries, namely: matplotlib, numpy, pandas, seaborn, graphviz and tensorflow. We then import all the data on which we are going to work in the form of a data-frame.

## 2.3. KNN model training

The KNN algorithm is used to build the prediction model, which uses the features of the training samples to predict the class of a new sample based on its nearest neighbors in the feature space. The KNN model is trained using the ECG recordings with their corresponding labels (presence or absence of cardiovascular diseases). The datasets are divided into training and test sets, and the KNN model is adjusted based on the characteristics of the ECG and associated annotations.

## 2.4. The KNN algorithm

It is easy to implement and understand, but has the major disadvantage of slowing down considerably as the size of the data used increases. Its area of application is Comparison of people with similar financial characteristics for granting bank loans, building a profile to offer subscribers suitable films, Classifying a potential voter into the categories “will vote” or “will not vote” for such and such a candidate.

KNN is simple and without any assumptions, but the drawback of the algorithm is that it is slow and can become weak as the number of features increases. It is also difficult to determine the optimal value of K - which is the number of neighbors used.

### 2.4.1. Math function

The K-NN algorithm needs a distance calculation function between two observations. There are several distance calculation functions: Euclidean distance, Manhattan distance, Minkowski distance, Jaccard distance, Hamming distance, etc. The distance function is chosen according to the types of data manipulated. Thus for quantitative data (example: weight, wages, height, amount of electronic shopping cart, etc.) and of the same type, the Euclidean distance is a good candidate. As for the Manhattan distance, it is a good measure to use when the data (input variables) are not of the same type (example: Age, sex, length, weight, etc.) [6]. There is no need to code these distances yourself, generally, Machine Learning libraries like Scikit Learn, perform these calculations internally. You just have to indicate the distance measure you want to use. For the curious, here are the mathematical definitions of the distances we have just mentioned.

- Euclidean distance: Distance that calculates the square root of the sum of the square differences between the coordinates of two points [7]:

$$\text{From}(x, y) = \sqrt{\sum (x_j - y_j)^2} \quad (1)$$

- Manhattan distance: Manhattan distance: calculates the sum of the absolute values of the differences between the coordinates of two points [7]:

$$D_m(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^k |x_i - y_j| \quad (2)$$

- Minkowski distance: The distance between two given points is the maximum difference between their coordinates on one dimension [32]:

$$d(\mathbf{p}, \mathbf{q}) = (\sum_{i=1}^n |p_i - q_i|^c)^{1/c} \quad (3)$$

- Hamming distance: The distance between two given points is the maximum difference between their coordinates on one dimension [7]:

$$D_h(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^k |x_i - y_j| \quad (4)$$

$$\text{with } x = y \Rightarrow D = 0 / x \neq y \Rightarrow D = 1$$

#### 2.4.2. Algorithmic model

The steps of the KNN algorithm are described by the following pseudo code [5]:

The k-nearest neighbors algorithm

*Start Algorithm*

*Input data (or vectors)*

- *a set of data (database).*
- *a function for defining the distance.*
- *An integer K.*

*For a new observation (vector) whose Do class we want to predict:*

*Calculate all the distances of this observation with the other observations of the dataset (training database).*

*Retain observations from the nearest dataset using the distance calculation function.*

*Take the values of the classes of observations selected. Return the most dominant or most voted class.*

*End Algorithm*

```
KNN_model = KNeighborsClassifier(n_neighbors=6)

KNN_model.fit(X_train,y_train)
KNN_predictions = KNN_model.predict(X_test)

print(confusion_matrix(y_test,KNN_predictions))
print('\n')
print(classification_report(y_test,KNN_predictions))
```

**Figure 2: KNN model training code**

### 3. Materials used

For the realization of this last several tools were used: Google Colab or Colaboratory is a cloud service provided by Google (free), based on Jupyter Notebook and intended for training and research in Machine Learning. The platform makes it possible to train machine learning models directly in the cloud [12]. Jupyter Notebook is an open source web application for creating and sharing documents containing code (can be run directly in the document), equations, images, and text. Using this

application, we can do data processing, statistical modeling, data visualization, machine learning, etc... It is available by default in the Anaconda distribution [9].

**Tensor Flow** is an end-to-end open source platform for building machine learning applications. It is a symbolic mathematical library that uses data flow and differentiable programming to perform various tasks focused on deep neural network training and inference [10].

**Keras** is a high-level neural network API, written in Python and able to run on TensorFlow or Theano. It was developed with an emphasis on rapid experimentation. [11].

**matplotlib** is a plotting library available for the Python programming language used to create static, animated, and interactive visualizations [12].

**NumPy (Numerical Python)** is the most popular scientific computing library in Python to perform scientific calculations. Pandas is a BSD-licensed open-source library that provides high-performance, easy-to-use data structures and data analysis tools for the Python programming language [13].

## II. Criterion for evaluating the performance of classification models

Evaluation allows the model to be tested against data that has never been used for training. This lets you see how the model might perform against data it hasn't seen yet. This is meant to be representative of how the model might work in the real world. To evaluate the performance of a classification model, the confusion matrix is used.

### 1. Confusion Matrix

It is a table that summarizes predictions for a particular classification problem. It compares the actual data of the target variable with the data predicted by the model [6]. Correct and false predictions are displayed and divided into 4 categories (see figure 2.7) explained as follows:

- **True Positive (TP)**: the prediction and the actual value are positive. Example: A sick and predicted sick person;
- **True Negative (TN)**: the prediction and the actual value are negative. Example: A healthy and expected healthy person;
- **False Positive (FP)**: the prediction is positive while the actual value is negative. Example: A healthy and predicted ill person;
- **False Negative (FN)**: the prediction is negative while the actual value is positive. Example: A sick and expected healthy person.

The confusion matrix is used to calculate performance metrics: occurrence, recall, F1-score, precision [6].

		<b>Predicted class</b>	
		<b>Positive</b>	<b>Negative</b>
<b>Current class</b>	<b>Positive</b>	<b>True Positive (TP)</b>	<b>False Negative (FN)</b>
	<b>Negative</b>	<b>False Positive (FP)</b>	<b>True Negative (TN)</b>

**Figure 3: confusion matrix**

## 2. Occurrence

The accuracy allows to know the proportion of correct predictions compared to all the predictions. The operation is simply: Number of good predictions /Total number of predictions [6].

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

## 3. Reminder

The recall corresponds to the number of documents correctly attributed to class i compared to the total number of documents belonging to class i (total True positive) [6].

$$\text{Reminder} = \frac{TP}{TP+FN} \quad (6)$$

## 4. Accuracy

Accuracy corresponds to the number of documents correctly attributed to class i compared to the total number of documents predicted as belonging to class i (total predicted positive) [6].

$$\text{Accuracy} = \frac{TP}{TP+FP} \quad (7)$$

## 5. F1score

The F1 – score is a metric to evaluate the performance of classification models with 2 or more classes. It is particularly used for problems using unbalanced data [6]. The F1 – score allows the values of precision and recall to be summarized into a single metric. Mathematically, the F1-score is defined as the harmonic mean of precision and recall, which results in the following equation:

$$\text{F1-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 * TP}{2 * TP + FP + FN} \quad (8)$$

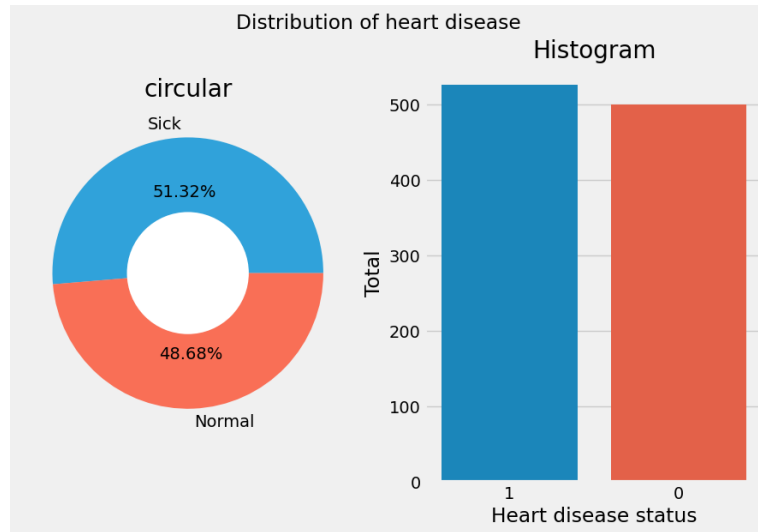
### III. Presentation of the results

#### I. KNN algorithm results

##### 1. Results

##### a. Cardiac Data Visualization

It is important to see the balancing of the data in order to see an accurate result, Figure 1 presents the data visualization graphs. It shows us that the number of healthy people is 474 with a proportion of 48.68% and the number of affected people is 551 with a proportion of 51.32%. However, the percentage of sick and healthy people shows us that the base is almost balanced.

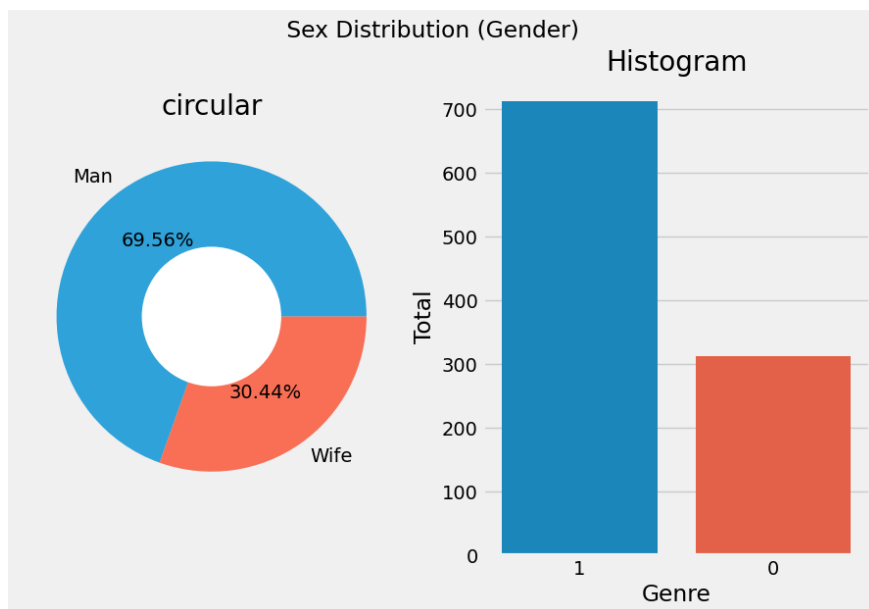


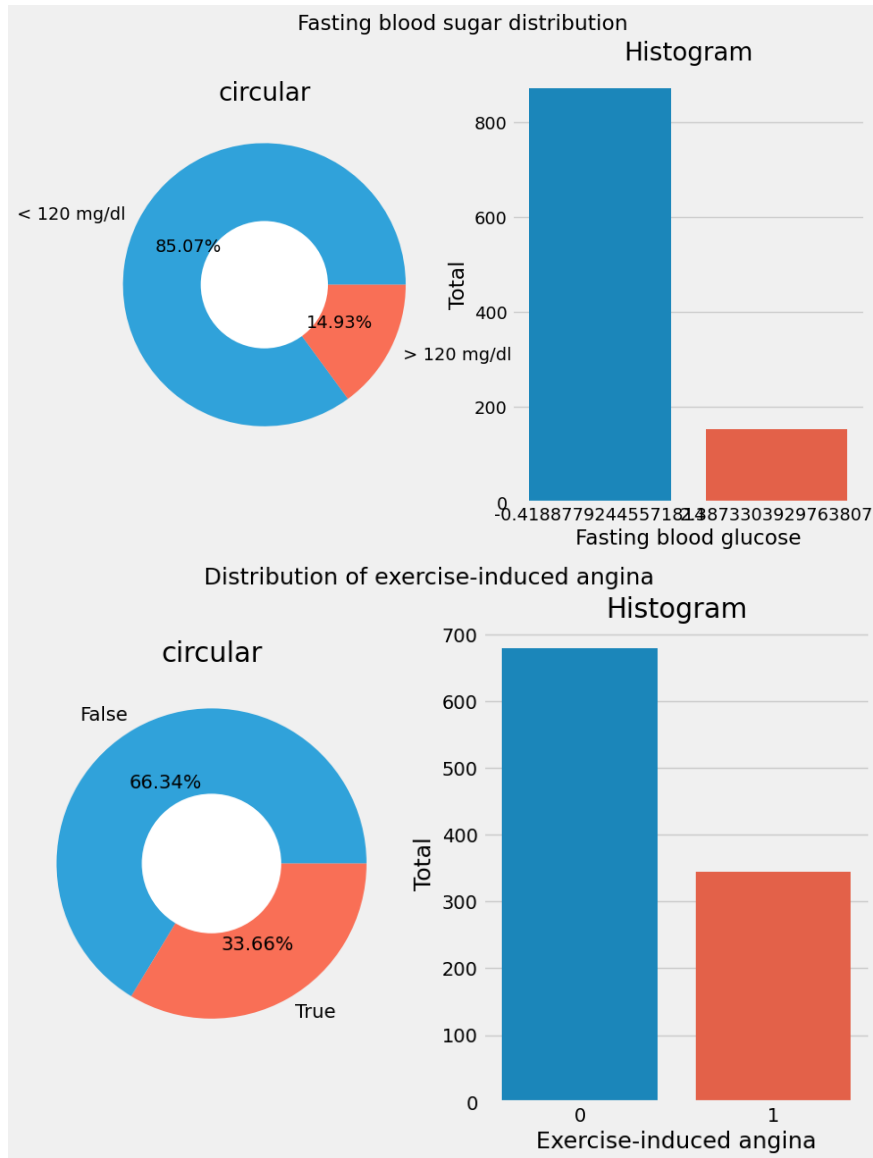
**Figure 4: Visualization of the target class (cardiac).**

b. Visualization of categorical data relative to the target class

The results show the following information:

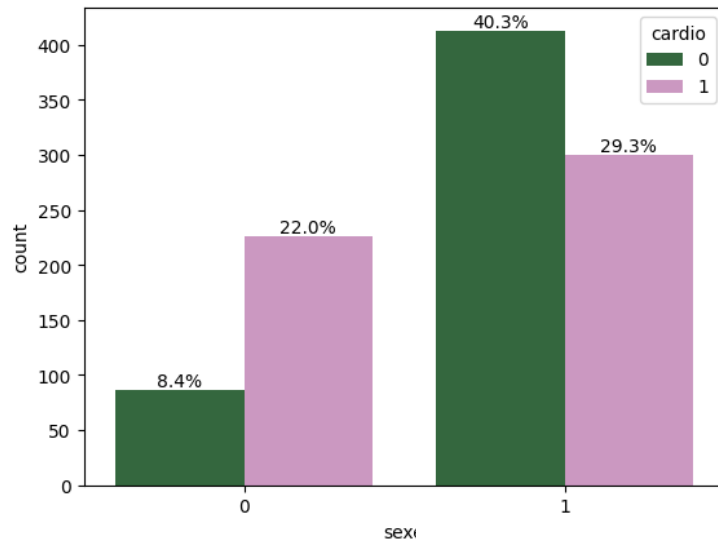
- **Gender:** In the data set, the numbers of women slightly exceed men's alone;
- **Blood sugar:** 872 people have a fasting blood sugar level above 120 mg/dl only 153 have below 120 mg/dl;
- **Exang:** 345 people have exercise-induced angina and 680 do not;





**Figure 5: Visualization of categorical data**

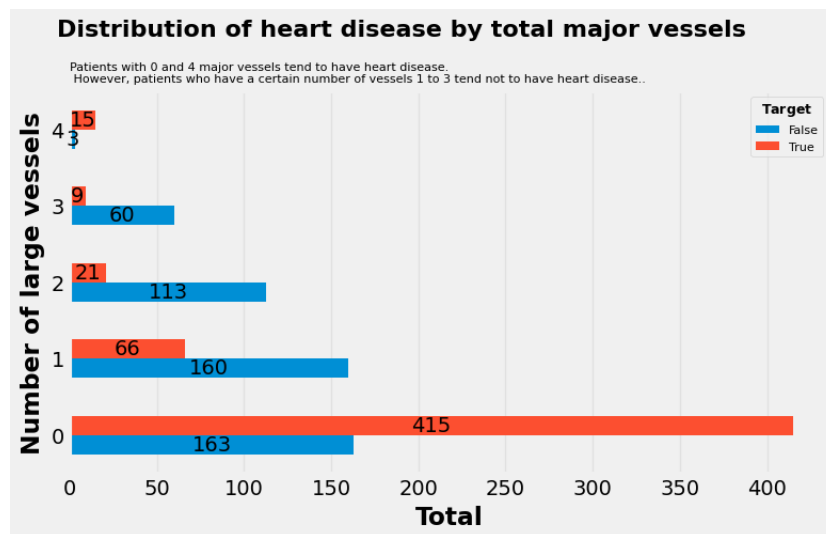
c. Distribution of heart disease by gender  
 Men tend to have heart disease compared to women. In women, the distribution is not unbalanced compared to men who have almost the same distribution.



**Figure 6: Distribution of heart disease by gender**

d. Distribution of heart disease by total major vessels

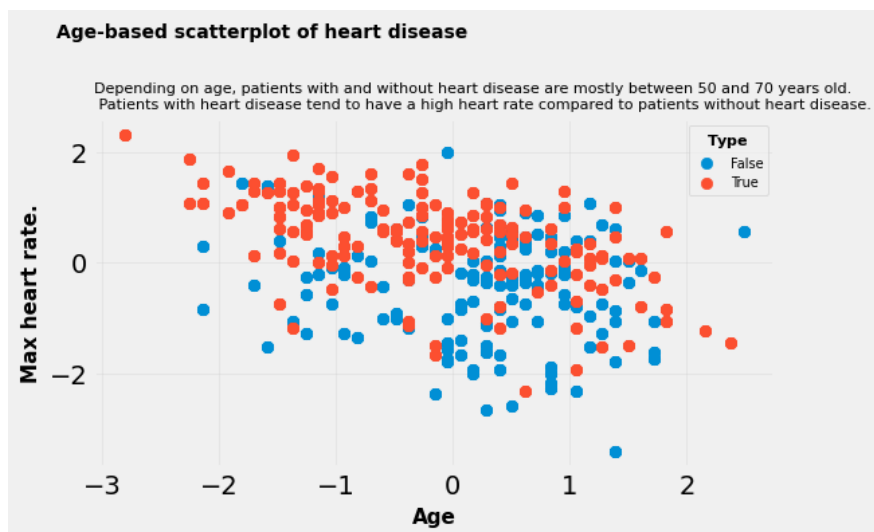
Patients with 0 and 4 major vessels tend to have heart disease. However, patients who have a number of vessels 1 to 3 tend not to have heart disease.



**Figure 7: Distribution of heart diseases according to the total of the main vessels**

e. K-nearest-neighbor plot of heart disease based on age

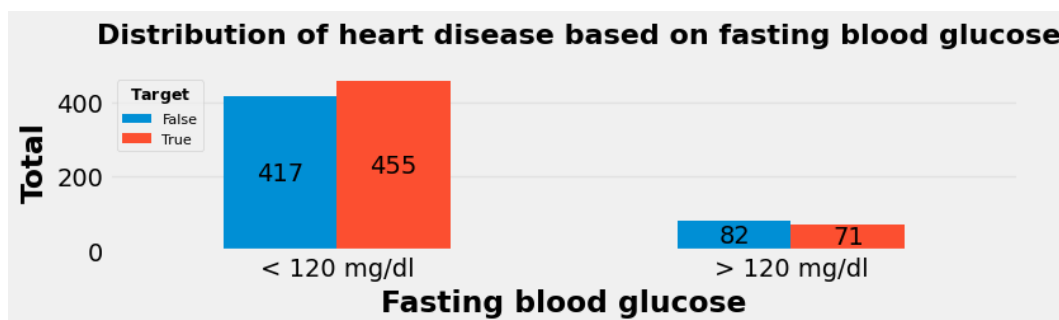
Depending on age, patients with and without heart disease are mostly between 50 and 70 years old. Patients with heart disease tend to have an elevated heart rate compared to patients without heart disease.



**Figure 8: Diagram of heart disease based on age**

f. Distribution of Heart Disease Based on Fasting Glucose

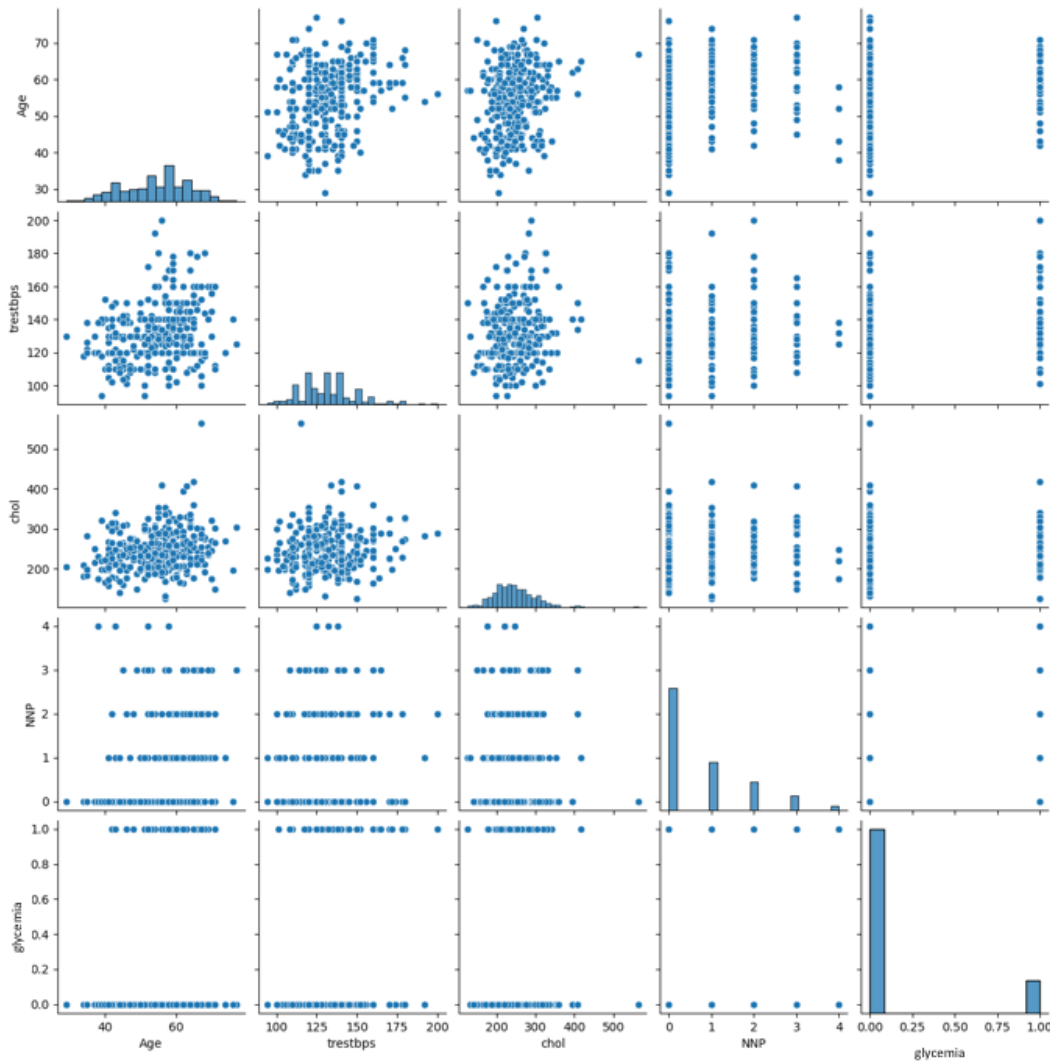
The number of patients with low fasting blood glucose is higher than that of patients with high fasting blood glucose. In the case of low fasting blood sugar, patients tend to have heart disease. In addition, the distribution of heart disease patients with high blood sugar is also distributed.



**Figure 9: Distribution of Heart Disease Based on Fasting Blood Glucose**

g. Visualization of relationship between variables and target class

Figure 7 shows us that the people susceptible to have a heart disease are the middle-aged people or those who have a low maximum heart rate or those who have a high depression. Most of the other variables seem not to influence whether a person is sick or not.



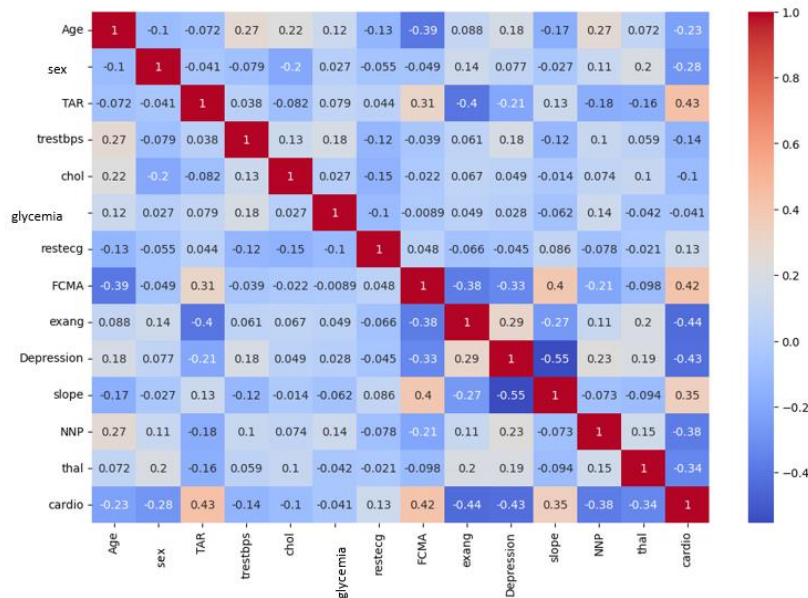
**Figure 10: Visualization of relationship between variables and target class**

a. Correlation Matrix

Correlation is a statistical measure that expresses the notion of a linear connection between two variables (which means that they evolve together at a constant speed). It is a common tool for describing simple relationships without worrying about cause and effect. One of the important processes to improve data is to determine the correlation between variables, where it is used to represent the statistical measure of the linear relationship between two variables. To calculate the correlation ratio between two different variables, we used the Pearson correlation coefficient defined by the formula below. Correlations are described using a unitless measure called the correlation coefficient between  $-1$  and  $+1$  and denoted  $p$  [7]:

- The closer  $p$  is to zero, the weaker the linear relationship;
- Positive values of  $p$  indicate a positive correlation when the values of the two variables tend to increase together;
- Negative values of  $p$  indicate a negative correlation when the values of one variable tend to increase and the values of the other variable decrease.

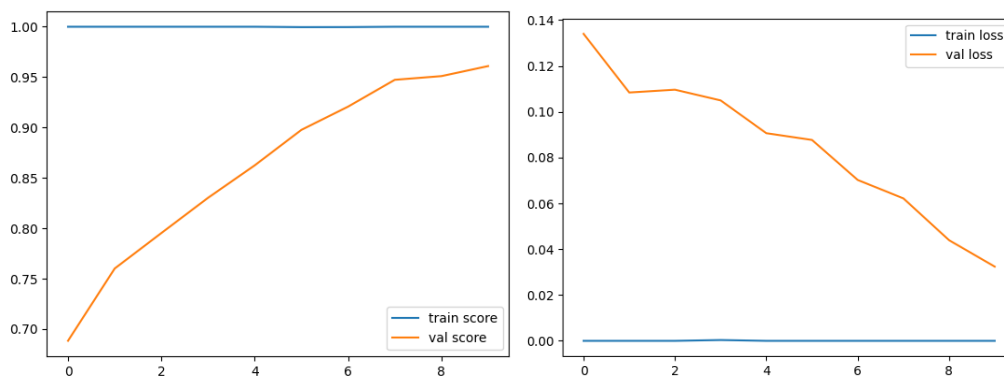
$$\frac{\sum(xi - x) * (yi - y)}{\sqrt{\sum(xi - x)^2 * \sum(yi - y)^2}} \quad (9)$$



**Figure 11: Correlation matrix**

**2. Analysis of the strengths and weaknesses of the proposed models**

By observing figure 9, we notice that the training and validation accuracies increase in parallel and rapidly when the k is closer which means that the model is learning the characteristics present in the data for the first time, then at a certain point, they stop increasing and stabilize, because at this stage the model is looking for new characteristics to learn (characteristics that it has not learned during its training). The validation accuracy reaches the maximum value of 98.7% at k=2. In Figure 10, we see the reverse. The training and validation losses decrease in parallel which indicates that there is an improvement in loss, then from a certain point, they stop decreasing and stabilize.



**Figure 12: Accuracy evolution curves and that of loss evolutions.**

In the KNN algorithm, the data is divided into two parts the number of the data for learning is 718 of which 70% and 307 data as data for the test with a percentage of 30% on the database. The KNN correctly classified 298 data, therefore an accuracy rate of 96.75% and 10 data were misclassified, therefore an error rate of 3.24%.The results obtained show that the use of the ECG combined with the KNN algorithm makes it possible to obtain promising prediction performances for cardiovascular

diseases. The model can identify specific patterns and characteristics associated with different heart conditions, which facilitates accurate prediction.

### 1. Algorithm performance evaluation

After training the models with the KNN, LR and ANN algorithms, we predict the classes using the test dataset to select the best model for use in the prediction phase. The best model is chosen based on its performance. In general, the performance of ANN algorithm is more evaluated based on parameters such as correctness, precision, recall, specificity, etc. In our case we chose accuracy as a performance criterion to evaluate the training model.

Which shows us that ANN can make us more reliable results compared to other models, as our dataset is balanced we take the occurrence as our study's prediction rate. Table 4.2 shows the results of the performance evaluation metrics (Occurrence, Precise, Recall, F1-score) extracted from the confusion matrices for each algorithm before attribute selection:

**Painting2:** *Model performance evaluation results*

<b>Models</b>	<b>Occurrence</b>	<b>specify</b>	<b>Reminder</b>	<b>F1-score</b>
<b>KNN</b>	0.97%	0.95%	0.99%	0.97%

## IV. Discussion

Based on the experiments we have done on the KNN model, we find that it gives satisfactory results. We chose to test this model since it proved its efficiency on data classification tasks. The model was able to generalize and therefore to differentiate between sick and healthy people with manifestation of cardiovascular disease and normal ones. KNN performed best with an accuracy of 97.5% and an occurrence of 98.7%. It was able to correctly classify 298 out of 308 data. After training the models with the KNN algorithm, we predict the classes using the test dataset to select the best model to use in the training phase. prediction. The best model is chosen based on its performance. In general, KNN algorithm performance is further evaluated based on parameters such as occurrence, precision, sensitivity and F1-score, in case the k near neighbor equals 2. In our case we chose occurrence as the performance criterion to evaluate the training model. Which shows us that KNN for k=2 can make us more reliable results compared to other models.

By comparing our results with previous works we find that [7], which used the same algorithm to detect heart disease with data selected from UCI, but our result is the best compared to him which is 95.71% . [6], using the same algorithm with the Cleveland database to predict heart disease, the overall accuracy of the KNN classification is 75%, which suggests that the model allows good prediction of the disease. The researchers [26], at their level, compared five supervised learning algorithms with the same database as ours. Comparison Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), The use of ECG and the KNN algorithm offers a non-invasive, data-driven approach to cardiovascular disease prediction. However, it is important to note that other risk factors and medical data may also improve predictive performance. Further studies are needed to validate these results and explore other machine learning techniques for more accurate prediction.

## V. Conclusion

In this article, an effective method of prediction and classification of cardiovascular disease using KNN algorithm and ECG sensor with input data. 1025 database datahospitalsAndwhich contains 14 attributes has been used.The KNN model is designed considering occurrence, precision, F1-score and recall. As a result, our proposed model achieved 98.7% occurrence, 97.5% precision, 98.7% F1-score, 100% sensitivity. Our cardiovascular disease prediction and classification result indicates that disease detection with the ECG sensor and a KNN model can be an effective approach to help experts diagnose cardiovascular diseases that can be seen from ECG signals. For future work, we are building a disease prediction web platform, sending it alert to medicine and patient. The ECG combined with the KNN

algorithm shows promising potential for the prediction of cardiovascular diseases. This approach can contribute to early detection and rapid intervention, which is essential for improving clinical outcomes and reducing cardiovascular disease-related morbidity and mortality. However, further research is needed to confirm and improve the predictive performance of these models.

In this perspective, the use of the KNN algorithm combined with ECG sensors represents a promising approach to predict cardiovascular diseases. This paves the way for new opportunities for research and technology development aimed at improving the early detection of cardiovascular disease and improving clinical outcomes for patients with these conditions.

## **References**

### Webographies

- [1] World Health Organization (2017). Cardiovascular diseases (CVD).[https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))  
Accessed July 17, 2023
- [2] <https://123dok.net/article/conclusion-mod%C3%A9lisation-d%C3%A9bruitage-extraction-charact%C3%A9ristiques-classification.yevpm6ez> Accessed July 18, 2023
- [3] <https://colab.research.google.com/drive/18CV7ePgcJC1zPpzCMtHlAmeGx6u95gKr#scrollTo=O6ttizeGGOLp>, 2023.

### Bibliography

- [4] Md Rahman, Mohammad Abul Kashem, Al-Akhir Nayan, Most Akter, Fazly Rabbi, Marzia Ahmed, Mohammad Asaduzzaman, et al. Internet of things (iot) based ecg system for rural health care. arXiv preprint arXiv:2208.02226, 2022.
- [5] Wahyu Caesarendra, Taufiq Aiman Hishamuddin, Daphne Teck Ching Lai, Asmah Husaini, Lisa Nurhasanah, Adam Glowacz, and Gusti Ahmad Fanshuri Alfarisy. An embedded system using convolutional neural network model for online and real-time ecg signal classification and prediction. *Diagnoses*, 12(4):795, 2022.
- [6] Farida Ghebouli and Amel Lakroum. Application for Classification and Migration between Classes of Heart Disease. Doctoral thesis, University Mohamed el-Bachir el-Ibrahimi Bordj Bou Arréridj Faculty of . ., 2021.
- [7] Lina Aida BEHLOULI and Hichem HAMDJ. A hybrid approach for the prediction of heart disease using the AG-KNN. Doctoral thesis, akli mohand oulhadj-Bouira university, 2022.
- [8] Chad Mourning, David Juedes, Allyson Hallman-Thrasher, Harsha Chenji, Savas Kaya, and Avinash Karanth. Reflections of cybersecurity workshop for k-12 teachers and highschool students. In *Proceedings of the 53rd ACM Technical Symposium on ComputerScience Education V. 2*, pages 1127–1127, 2022.
- [9] Aymen MHAMMEDI, Imane YAKOUB, Abdelwhab OUAHAB, et al. The detection of Covid-19 by deep learning (Deep Learning). PhD thesis, AH-MED DRAIA-ADRAR UNIVERSITY, 2021.
- [10] Abderrahmane BENDJAAFER, Toufik MEDDAH, et al. Image classification using a Totally convolutional network. Doctoral thesis, University of M'sila, 2021
- [11] Milena Carolina dos Santos Mangueira, Renata Passos Machado Vieira, Francisco Regis Vieira Alves, and Paula Maria Machado Cruz Catarino. Leonardo's bivariate and complex polynomials. *Notes on Number Theory and Discrete Mathematics*, 28(1):115–123,2022.
- [12] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.