



Optimization of Malaria Diagnosis by Machine Learning According to the CRISP-DM Model Applied to the University Teaching Hospital Clinics of Lubumbashi (DRC)

Bertin Mazunze¹, Lukusa Muvuala Vicky¹, Kasongo Nyange Franck¹, Moshu Mesu Pierre-Stéphane¹, Kalonji Muya Emery Patrice¹, Kabango Dyuba Desiré², Mutombo Shanga Eddy³

¹Department of Computer Science, Lubumbashi Higher Institute of Commerce, Lubumbashi, Democratic Republic of the Congo

²Department of Computer Science, Higher Pedagogical Institute of Mitwaba, Mitwaba, Democratic Republic of the Congo

³Department of Computer Science, Higher Institute of Applied Techniques of Gombe Matadi, Mbanza Ngungu, Democratic Republic of the Congo

Email: berthmaz1705@unikik.ac.cd

How to cite this paper: Mazunze, B., Vicky, L.M., Franck, K.N., Pierre-Stéphane, M.M., Patrice, K.M.E., Desiré, K.D. and Eddy, M.S. (2025) Optimization of Malaria Diagnosis by Machine Learning According to the CRISP-DM Model Applied to the University Teaching Hospital Clinics of Lubumbashi (DRC). *Open Access Library Journal*, 12: e14143.

<https://doi.org/10.4236/oalib.1114143>

Received: August 20, 2025

Accepted: October 7, 2025

Published: October 10, 2025

Copyright © 2025 by author(s) and Open Access Library 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

Malaria remains a major public health challenge in the Democratic Republic of Congo (DRC), particularly in Lubumbashi, where traditional diagnostic methods are struggling to meet growing demand. The study was conducted at the University Clinics of Lubumbashi (UCL), the teaching hospital affiliated with the University of Lubumbashi. This work proposes an expert system based on artificial intelligence (AI) and the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology to optimize malaria diagnosis in this setting. By leveraging a decision tree classifier trained on local clinical data, the system achieved an accuracy of 90.4%, a recall of 88%, and a specificity of 92%. The results demonstrate a substantial improvement in the speed and reliability of diagnosis, providing a transparent and interpretable decision-support tool suitable for resource-limited healthcare environments.

Subject Areas

Artificial Intelligence/Machine Learning

Keywords

Malaria, Machine Learning, Artificial Intelligence, CRISP-DM, Expert System, Medical Diagnosis, Public Health, DRC, Lubumbashi

1. Introduction

In 2023, malaria caused nearly half a million deaths, 95% of which were in Africa and 76% in children under 5 years of age [1]. In sub-Saharan Africa, malaria diagnostic errors persist due to reliance on rapid diagnostic tests (RDTs) and microscopy, whose sensitivities range from 60% to 85% depending on operational conditions [2]. The Democratic Republic of Congo (DRC) ranks second in the world in terms of the number of malaria cases (12.6%) and deaths (11.3%) [3]. In 2023, it represented 55% of malaria cases in Central Africa, the highest rate in the sub-region [4]. In Lubumbashi, health centers, hospitals and university clinics face major challenges: overload of health professionals, prolonged diagnostic delays and human errors. Traditional methods, such as microscopy and rapid tests, have limitations in terms of sensitivity and specificity. Traditional diagnostic tests, such as microscopy, require expertise and specialized equipment that are not necessary for others [5].

The burden of malaria is particularly high in low-income countries, where health systems are often underdeveloped and resources for prevention, diagnosis, and treatment are insufficient. In the DRC, access to quality tools and the necessary expertise remains a recurring problem, directly threatening patients' lives.

It is in this context that our study takes on its full meaning: it seemed essential to us to propose an innovative solution that could not only help our country, but also benefit other nations facing similar challenges. Thus, artificial intelligence (AI) and machine learning (ML) appear as promising alternatives for automating and optimizing medical diagnoses, particularly in resource-limited contexts. Indeed, several studies have shown the effectiveness of ML in the diagnosis of infectious diseases, including malaria, with accuracy rates sometimes higher than those of conventional methods [6] [7]. According to Kermany *et al.* (2018), AI can even achieve diagnostic accuracy equivalent to that of human experts in certain complex clinical contexts [8]. However, their implementation remains little explored in sub-Saharan Africa, notably due to difficulties in accessing data, user training and infrastructure constraints [9].

This study aims to fill this gap by developing an expert system based on the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, guaranteeing a structured, reproducible approach adapted to field realities.

2. Materials and Methods

The implementation of this project relies on the combined use of technical means, software resources, and a rigorous data processing method. This section describes the tools used and the methodological steps followed to build the intelligent diagnostic system.

2.1. Materials

The development of the system relied on a coherent set of software tools adapted to data science. Python 3.10 was chosen for its wealth of Machine Learning-

oriented libraries, such as *scikit-learn*, *pandas*, and *seaborn*, regularly used in biomedical studies [10]. The Anaconda environment allowed for the efficient management of virtual environments and dependencies, reducing the risk of version conflicts. Finally, the PyCharm IDE was mobilized to structure the project and facilitate the debugging and continuous integration phases. To ensure a stable, isolated, and reproducible working environment, we used Anaconda, a Python distribution dedicated to the management of dependencies and virtual environments, avoiding conflicts between libraries. The source code was developed in the PyCharm integrated development environment, facilitating advanced project management, static code analysis, as well as debugging and integration with version control tools such as Git.

Several specialized Python libraries were used to meet the specific needs of our project. These tools covered all the necessary functionalities, from data manipulation and preparation to modeling, visualization, and user interface management. **Table 1** presents the main libraries used and their role in the development of the expert system.

Table 1. Python libraries used and their roles.

No.	Library	Main role
1	scikit-learn	This library allowed us to train and evaluate the artificial intelligence model (Decision Tree).
2	pandas	It helped us efficiently manipulate and transform clinical data.
3	joblib	Joblib was used to save and reload learning models for later use.
4	streamlit	This library was used to create an interactive and accessible web interface.
5	sqlite3	It allowed us to locally manage the database containing the data and history.
6	datetime	Datetime made it easier to manage the timestamps needed for historical tracking.
7	matplotlib/ seaborn	These libraries allowed us to visualize the model's performance through confusion matrices and ROC curves.
8	numpy	Numpy was used for numerical calculations and efficient manipulation of data arrays.
9	pickle	Pickle was used to save and load Python objects, particularly for model persistence.

2.2. Methods

The methodological approach is based on the **CRISP-DM (Cross-Industry Standard Process for Data Mining) framework**, which is widely adopted for structured data mining projects. This framework consists of six interconnected steps, ranging from business understanding to the deployment phase, including data preparation and modeling. Its effectiveness in medical and epidemiological

contexts is widely documented [11] [12].

In a recent study on the diagnosis of type 2 diabetes, the authors explicitly applied the CRISP-DM methodology to develop a high-performance predictive model. This work demonstrates the relevance and effectiveness of the structured CRISPDM process in a concrete framework of medical data analysis [13]. This methodology is all the more relevant in medical environments with low computerization, as shown by the work of Amato *et al.* (2013), where CRISP-DM made it possible to efficiently structure the analysis and transformation of raw clinical data [14].

This process includes several complementary phases as represented in **Figure 1**.

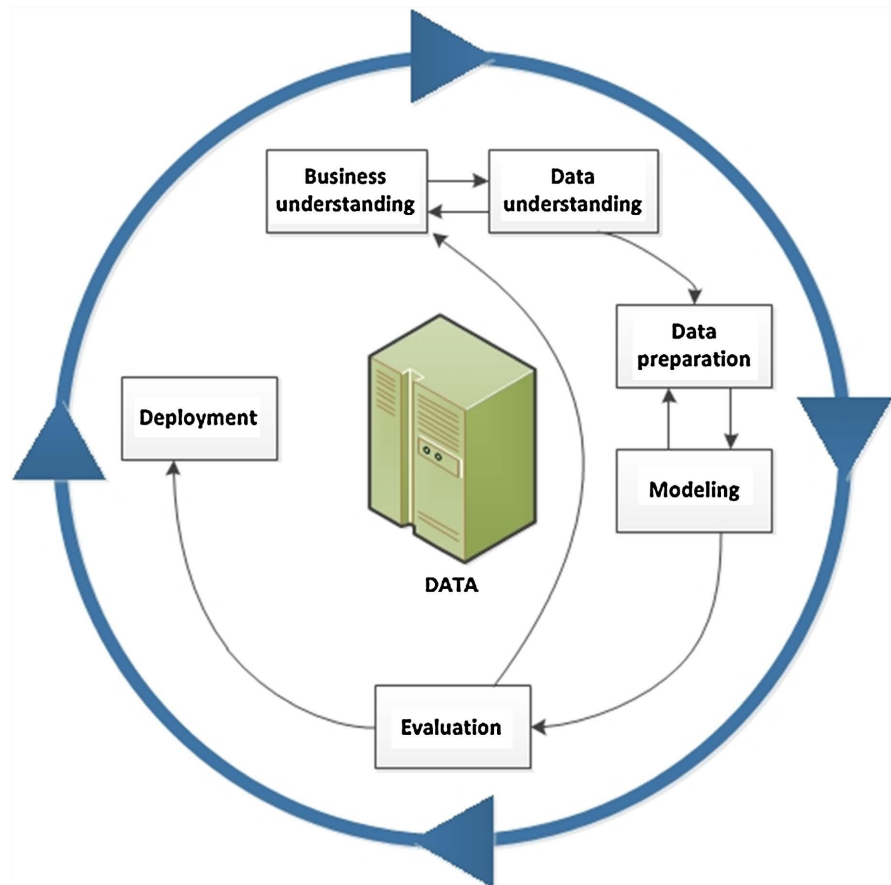


Figure 1. The data mining lifecycle.

To better illustrate how the CRISP-DM methodology was operationalized in this work, **Table 2** provides a concise mapping between each phase of the framework and the specific actions undertaken. This summary strengthens the methodological narrative by linking the theoretical steps of the CRISP-DM process to their concrete implementation in the context of malaria diagnosis at the University Clinics of Lubumbashi.

2.2.1. Understanding the Profession

This step is crucial to align clinical and technical objectives, as highlighted by

Table 2. Mapping of CRISP-DM phases to actions performed in this study.

CRISP-DM phase	Concrete actions performed in this study
Business Understanding	Interviews with physicians; observation of clinical consultations; identification of diagnostic workflow and decision criteria for malaria at UCL.
Data Understanding	Collection of 2500 retrospective patient records (Jan 2023-Mar 2024); exploration of variables (demographic, clinical, diagnostic); assessment of class distribution (224 malaria, 2276 non-malaria).
Data Preparation	Cleaning (handling missing values, duplicates); encoding categorical variables; transformation into numerical formats; class weighting to address imbalance; feature selection based on correlations and importance ranking.
Modeling	Implementation of a Decision Tree classifier; use of 10-fold cross-validation; training with class weights; computation of performance metrics (accuracy, recall, specificity, F1, MCC, ROC-AUC).
Evaluation	Performance analysis with mean \pm SD across folds; confusion matrix inspection; ROC curve analysis; clinical interpretability assessment.
Deployment	Integration of the model into a Streamlit interface; development of a relational SQLite database to store patient records, symptoms, and predictions; user interface for clinicians.

Otero *et al.* (2005), who state that the success of a medical data mining project strongly depends on the quality of the initial business understanding [15].

To gain a concrete understanding of medical practices in the field, we observed several consultations and interviewed healthcare professionals to understand the logic they use to diagnose malaria. This approach made it possible to formalize the essential questions asked of patients, understand the underlying clinical reasoning, and structure this reasoning in a clear and reproducible manner.

The clinical diagnosis of malaria is based on identifying characteristic symptoms and ruling out other possible causes of fever. This process revolves around a focused medical interview, during which the doctor asks a series of key questions. **Table 3** presents these questions, along with the medical objective pursued for each of them:

2.2.2. Understanding Data

To ensure effective deployment of our malaria prediction model, we began with a practical analysis of the collected data, in order to precisely meet the expectations of doctors and demonstrate the relevance of the variables used.

Thorough data analysis from the earliest stages is essential to detect anomalies and better guide future treatments. As Chapman *et al.* (2000) note, the data understanding phase directly influences the quality of predictive models in health [16].

1) Data sources

The dataset used in this study was extracted from the file *DonneesMalaria.xlsx*,

compiled from retrospective and anonymised patient records collected at the University Clinics of Lubumbashi (UCL) between January 2023 and March 2024. This Excel file centralises all the information required for analysis, including clinical symptoms, vital parameters, results of biological tests, and demographic characteristics. Only records with complete and consistent data were retained for the study, and all personal identifiers were removed prior to analysis to ensure patient anonymity (See **Figure 2**).

Table 3. Questions asked by the physician when a suspected diagnosis of malaria is made.

No.	Doctor’s question	Objective of the question
1	What is your gender?	Identify risk factors related to sex, particularly for pregnancy.
2	How old are you?	Appreciate the patient’s vulnerability, especially in children or the elderly.
3	What is your weight?	Assess the patient’s general condition and adjust future management.
4	Are you returning from recent surgery?	Eliminate postoperative fever as an alternative cause.
5	Have you had a fever recently?	Check for the presence of the main symptom of malaria, linked to the lysis of red blood cells.
6	Did you feel chills?	Detect typical infection cycles of the Plasmodium parasite.
7	Do you sweat profusely after a fever?	Confirm the sweating phase following the febrile attack.
8	Do you suffer from headaches?	Identify common headaches in malaria cases.
9	Have you experienced nausea or vomiting?	Detect digestive signs that may indicate a more severe form.
10	Are you pregnant? (if patient concerned)	Detect a high-risk situation requiring appropriate treatment.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	No.	Name	Last Name	First name	Sex	Age	Weight (kg)	Postoperative	Fever	Chills	Sweating	Headaches	Nausea	Pregnancy	Diagnosis
2	1	Ilunga	Ngoy	Kanku	Male	67	63,3	No	No	Yes	Yes	Yes	No	No	Other
3	2	Nzapa	Tshibangu	Benoît	Male	78	66,5	Yes	Yes	Yes	Yes	No	Yes	No	Other
4	3	Malanda	Mwamba	Mbuyi	Female	28	88,4	No	No	Yes	No	Yes	Yes	Yes	Other
5	4	Kabongo	Mubenga	Rosalie	Female	15	33,9	No	No	Yes	No	No	No	Not specified	Other
6	5	Lompala	Kazadi	Mbuyi	Female	71	94,2	Yes	No	No	No	Yes	No	No	Probable malaria
7	6	Ndala	Mubenga	Aline	Female	5	16,6	Yes	No	Yes	Yes	Yes	Yes	Not specified	Probable malaria
8	7	Omari	Mubenga	Moke	Male	14	29,9	Yes	No	Yes	Yes	Yes	No	No	Other
9	8	Ilunga	Ngoy	Nadia	Female	57	54,9	Yes	No	Yes	Yes	No	No	Yes	Other
10	9	Mvemba	Kabela wa Kiluba	Remy-Belli	Male	35	33,1	Yes	Yes	Yes	No	No	No	No	Probable malaria
11	10	Kalala	Mbuyi	Moke	Male	53	57,3	Yes	No	Yes	Yes	No	No	No	Probable malaria
12	11	Mvemba	Mbuyi	Makambo	Male	19	83,5	Yes	No	No	Yes	Yes	No	No	Other
13	12	Mvemba	Kazadi	Rosalie	Female	46	60,5	No	No	Yes	Yes	Yes	Yes	Yes	Other
14	13	Nguz	Mbuyi	Marie-Clai	Female	32	55,7	No	No	Yes	No	Yes	Yes	No	Other
15	14	Bokilo	Mubenga	Aline	Female	22	40,6	Yes	Yes	No	No	No	No	Yes	Other
16	15	Mobutu	Tshibangu	Mboyo	Male	1	43,6	No	No	Yes	No	No	No	No	Other
17	16	Malanda	Tshibangu	Marie-Clai	Female	40	84,2	Yes	Yes	No	No	No	Yes	No	Other
18	17	Ilunga	Mundele	Chantal	Female	52	15,4	No	Yes	Yes	No	No	No	No	Probable malaria
19	18	Luzolo	Mukendi	Emmanuel	Male	29	80,8	Yes	Yes	No	Yes	Yes	No	No	Other
20	19	Kabongo	Kanza	Remy	Male	12	68,7	Yes	No	Yes	Yes	Yes	Yes	No	Other
21	20	Mputu	Lukusa	Bibiche	Female	42	21,4	No	No	Yes	No	No	No	No	Probable malaria
22	21	Gbangbo	Mundele	Aline	Female	57	95,2	Yes	No	No	Yes	Yes	Yes	No	Other
23	22	Malanda	Mundele	Sifa	Female	65	83,7	Yes	No	Yes	Yes	Yes	Yes	No	Other
24	23	Nguz	Mundele	Jean-Paul	Male	43	88,3	Yes	No	No	No	Yes	Yes	No	Other
25	24	Gbangbo	Lusambo	Benoît	Male	35	21,1	No	No	No	Yes	No	No	No	Other
26	25	Ilunga	Mbuyi	Emmanuel	Male	77	79,3	No	Yes	No	No	No	No	No	Other
27	26	Omari	Lusambo	Eliane	Female	80	68,6	Yes	Yes	Yes	Yes	No	No	No	Probable malaria
28	27	Ekokobe	Ngoy	Nadia	Female	47	80,5	No	No	No	No	Yes	No	No	Other
29	28	Mputu	Mbuku	Moke	Male	61	37,8	No	No	Yes	Yes	No	Yes	No	Probable malaria
30	29	Tshibanda	Lusambo	Rosalie	Female	18	97,8	No	No	No	Yes	Yes	No	No	Other

Figure 2. Data presentation.

2) Description of data

The dataset includes 2500 records, each corresponding to a patient described by several clinical and demographic attributes (See [Table 4](#)).

Table 4. Description of data.

No.	Attribute	Description	Kind
1	Sex	Male or female	Categorical
2	Age	In years	Digital
3	Weight (kg)	In kilograms	Digital
4	Postoperative	Yes/No depending on whether the patient is returning from an operation	Binary
5	Fever	Yes/No	Binary
6	Chills	Yes/No	Binary
7	Sweating	Yes/No	Binary
8	Headaches	Yes/No	Binary
9	Nausea	Yes/No	Binary
10	Pregnancy	Yes/No/Not specified (if woman ≥ 18 years old or inapplicable case)	Categorical
11	Diagnosis	Probable malaria/Other	Categorical (target)

3) Data exploration

The dataset contains 2500 rows, each representing a patient, and 11 variables, including demographic information (gender, age, weight), clinical symptoms (fever, chills, headache, etc.) and final diagnosis. The class distribution was as follows: 224 malaria cases (9%) and 2276 non-malaria cases (91%). This imbalance reflects the real-world prevalence observed at the University Clinics of Lubumbashi.

To mitigate the risk of the model being biased towards the majority class, a class weighting strategy was applied during training. This ensured that malaria cases, although underrepresented, had a proportional impact on the learning process.

[Table 5](#) presents all the variables included in the analysis, accompanied by a brief description, their type, as well as the justification for their relevance in the context of our study:

4) Data preparation

The data preparation phase was essential to ensure a clean, consistent, and usable dataset. This step consisted of several sub-phases: presentation, cleaning, transformation, statistical testing, and variable selection (See [Figure 3](#)).

a) Data Presentation

The dataset comes from the file `DonneesMalaria.xlsx` and contains 2500 rows representing patients, each described through 11 clinical, demographic and diagnostic variables. The data is structured in a tabular manner:

- Each line represents a patient;

Table 5. Data exploration.

No.	Variable	Description	Kind	Reason for its presence in the model
1	Sex	Patient’s gender (Male/Female)	Categorical	To detect possible differences in exposure or response to malaria between men and women.
2	Age	Patient’s age in years	Digital	Age can influence vulnerability to malaria (children and the elderly are often at greater risk).
3	Weight (kg)	Patient’s weight in kilograms	Digital	Can indicate the patient’s general condition; useful in post-diagnostic monitoring.
4	Postoperative	If the patient is returning from an operation (Yes/No)	Binary	Helps rule out symptoms related to recent surgery rather than malaria infection.
5	Fever	Presence or absence of fever	Binary	Main symptom of malaria; its detection is crucial.
6	Chills	Presence or absence of chills	Binary	A common symptom of malaria, often associated with fever.
7	Sweating	Presence or absence of excessive sweating	Binary	May accompany fever spikes and help confirm a case of malaria.
8	Headaches	Presence or absence of headaches	Binary	Common symptom that may increase the likelihood of a malaria diagnosis.
9	Nausea	Presence or absence of nausea or vomiting	Binary	May be linked to malaria but also to other conditions; useful for refining prediction.
10	Pregnancy	If the patient is pregnant (Yes/No/Not specified)	Categorical	Detect a high-risk situation requiring specific management.
11	Diagnosis	Clinical observation result: “Probable malaria” or “Other”	Categorical	Model target; serves as a basis for training and validating predictions.

No.	Name	Last Name	First name	Sex	Age	Weight (kg)	Postoperative	Fever	Chills	Sweating	Headaches	Nausea	Pregnancy	Diagnosis
1	Ilunga	Ngoy	Kanku	Male	67	63,3	No	No	Yes	Yes	Yes	No	No	Other
2	Nzapa	Tshibangu	Benoît	Male	78	66,5	Yes	Yes	Yes	Yes	No	Yes	No	Other
3	Malanda	Mwamba	Mbuyi	Female	28	88,4	No	No	Yes	No	Yes	Yes	Yes	Other
4	Kabongo	Mubenga	Rosalie	Female	15	33,9	No	No	Yes	No	No	No	Not specified	Other
5	Lompala	Kazadi	Mbuyi	Female	71	94,2	Yes	No	No	No	Yes	No	No	Probable malaria
6	Ndala	Mubenga	Aline	Female	5	16,6	Yes	No	Yes	Yes	Yes	Yes	Not specified	Probable malaria
7	Omari	Mubenga	Moke	Male	14	29,9	Yes	No	Yes	Yes	Yes	No	No	Other
8	Ilunga	Ngoy	Nadia	Female	57	54,9	Yes	No	Yes	Yes	No	No	Yes	Other
9	Mvemba	Kabela wa Kiluba	Remy-Belli	Male	35	33,1	Yes	Yes	Yes	No	No	No	No	Probable malaria
10	Kalala	Mbuyi	Moke	Male	53	57,3	Yes	No	Yes	Yes	No	No	No	Probable malaria
11	Mvemba	Mbuyi	Makambo	Male	19	83,5	Yes	No	No	Yes	Yes	No	No	Other
12	Mvemba	Kazadi	Rosalie	Female	46	60,5	No	No	Yes	Yes	Yes	Yes	Yes	Other
13	Nguz	Mbuyi	Marie-Clai	Female	32	55,7	No	No	Yes	No	Yes	Yes	No	Other
14	Bokilo	Mubenga	Aline	Female	22	40,6	Yes	Yes	No	No	No	No	Yes	Other
15	Mobutu	Tshibangu	Mboyoy	Male	1	43,6	No	No	Yes	No	No	No	No	Other
16	Malanda	Tshibangu	Marie-Clai	Female	40	84,2	Yes	Yes	No	No	No	Yes	No	Other
17	Ilunga	Mundele	Chantal	Female	52	15,4	No	Yes	Yes	No	No	No	No	Probable malaria
18	Luzolo	Mukendi	Emmanuel	Male	29	80,8	Yes	Yes	No	Yes	Yes	No	No	Other
19	Kabongo	Kanza	Remy	Male	12	68,7	Yes	No	Yes	Yes	Yes	Yes	No	Other
20	Mputu	Lukusa	Bibiche	Female	42	21,4	No	No	Yes	No	No	No	No	Probable malaria
21	Gbangbo	Mundele	Aline	Female	57	95,2	Yes	No	No	Yes	Yes	Yes	No	Other
22	Malanda	Mundele	Sifa	Female	65	83,7	Yes	No	Yes	Yes	Yes	Yes	No	Other
23	Nguz	Mundele	Jean-Paul	Male	43	88,3	Yes	No	No	No	Yes	Yes	No	Other
24	Gbangbo	Lusambo	Benoît	Male	35	21,1	No	No	No	Yes	No	No	No	Other
25	Ilunga	Mbuyi	Emmanuel	Male	77	79,3	No	Yes	No	No	No	No	No	Other
26	Omari	Lusambo	Eliane	Female	80	68,6	Yes	Yes	Yes	Yes	No	No	No	Probable malaria
27	Ekokobe	Ngoy	Nadia	Female	47	80,5	No	No	No	No	Yes	No	No	Other
28	Mputu	Mbuku	Moke	Male	61	37,8	No	No	Yes	Yes	No	Yes	No	Probable malaria
29	Tshibanda	Lusambo	Rosalie	Female	18	97,8	No	No	No	Yes	Yes	No	No	Other

Figure 3. Presentation of the data set.

- Each column represents a variable (gender, fever, headache, etc.).
- The variable types are:
- Qualitative: sex, post-operative, diagnosis, symptoms, pregnancy (Yes/No);
- Quantitative: age (in years), weight (in kg).

5) *Cleaning*

Cleaning consisted of managing missing values, correcting inconsistencies, and standardizing formats. Imputation was performed by the mean for numeric variables (age, weight), and by the modal value for categorical variables (sex, pregnancy, symptoms), to limit the loss of information.

This approach, although elementary, is commonly used in medical studies, especially when the rate of missing values is low. It is considered an effective method to preserve initial distributions without introducing major biases. Jakobsen *et al.* (2017) point out that simple imputation, especially by the mean or the most frequent value, is acceptable for exploratory and descriptive analyses when the missing data are random and few in number [15].

Here are the concrete operations carried out:

- Duplicates: identified via a search across all columns, then deleted.
- Input errors: standardization of labels.

Missing values:

- binary variables, the modal (most frequent) value was used for imputation.
- For numeric variables (age, weight), the mean was used to replace missing data.

No individuals were removed in order to preserve the complete sample.

6) *Transformation*

To prepare the data for machine learning analysis, we applied the following transformations:

- Conversion of “Yes”/“No” responses to 1/0 across multiple columns, including the target variable “Diagnosis” (1 for *Probable Malaria*, 0 for *Other*).
- Transformation of the variable “Sex” into two binary columns: “Sex_Female” and “Sex_Male”.
- Transformation of the “Pregnancy” variable into a single variable coded as follows: 0 = No, 1 = Yes and finally 2 = Not specified.
- Grouping the “Age” and “Weight (kg)” columns into numbers, then grouping them into ranges:

Age: 1 = child (<15 years), 2 = adult (15 - 50 years), 3 = senior (>50 years)

Weight: 1 = low (<45 kg), 2 = normal (45 - 75 kg), 3 = high (>75 kg)

These transformations make it easier to integrate data into machine learning models, while maintaining a simple and interpretable structure.

It should be noted that no normalization was performed, because the decision trees used in our study are insensitive to the scale of the variables. This choice is justified by Tan *et al.* (2018), who indicate that tree models are based on cutting thresholds and not on distances or numerical magnitudes [17].

After the transformation, our data takes the form shown in the following **Figure 4**.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	No.	Name	Last Name	First name	Age	Weight (kg)	Postoperative	Fever	Chills	Sweating	Headaches	Nausea	Pregnancy	Diagnosis	Sex_Female	Sex_Male
2	1	Ilunga	Ngoy	Kanku	3	2	0	0	1	1	1	0	0	0	0	1
3	2	Nzapa	Tshibangu	Benoît	3	2	1	1	1	1	0	1	0	0	0	1
4	3	Malanda	Mwamba	Mbuyi	2	3	0	0	1	0	1	1	1	0	1	0
5	4	Kabongo	Mubenga	Rosalie	2	1	0	0	1	0	0	0	2	0	1	0
6	5	Lompala	Kazadi	Mbuyi	3	3	1	0	0	0	1	0	0	1	1	0
7	6	Ndala	Mubenga	Aline	2	1	1	0	1	1	1	1	2	1	1	0
8	7	Omari	Mubenga	Moke	2	1	1	0	1	1	1	0	0	0	0	1
9	8	Ilunga	Ngoy	Nadia	2	2	1	0	1	1	0	0	1	0	1	0
10	9	Mvemba	Kabela wa	Remy-Bell	2	1	1	1	1	0	0	0	0	1	0	1
11	10	Kalala	Mbuyi	Moke	2	2	1	0	1	1	0	0	0	1	0	1
12	11	Mvemba	Mbuyi	Makambo	2	3	1	0	0	1	1	0	0	0	0	1
13	12	Mvemba	Kazadi	Rosalie	2	2	0	0	1	1	1	1	1	0	1	0
14	13	Nguz	Mbuyi	Marie-Clai	2	2	0	0	1	0	1	1	0	0	1	0
15	14	Bokilo	Mubenga	Aline	2	1	1	1	0	0	0	0	1	0	1	0
16	15	Mobutu	Tshibangu	Mboyo	1	1	0	0	1	0	0	0	0	0	0	1
17	16	Malanda	Tshibangu	Marie-Clai	2	3	1	1	0	0	0	1	0	0	1	0
18	17	Ilunga	Mundele	Chantal	3	1	0	1	1	0	0	0	0	1	1	0
19	18	Luzolo	Mukendi	Emmanuel	2	3	1	1	0	1	1	0	0	0	0	1
20	19	Kabongo	Kanza	Remy	1	2	1	0	1	1	1	1	0	0	0	1
21	20	Mputu	Lukusa	Bibiche	2	1	0	0	1	0	0	0	0	1	1	0
22	21	Gbangbo	Mundele	Aline	3	3	1	0	0	1	1	1	0	0	1	0
23	22	Malanda	Mundele	Sifa	3	3	1	0	1	1	1	1	0	0	1	0
24	23	Nguz	Mundele	Jean-Paul	2	3	1	0	0	0	1	1	0	0	1	0
25	24	Gbangbo	Lusambo	Benoît	2	1	0	0	0	1	0	0	0	0	1	0
26	25	Ilunga	Mbuyi	Emmanuel	3	3	0	1	0	0	0	0	0	0	1	0
27	26	Omari	Lusambo	Eliane	3	2	1	1	1	1	0	0	0	1	1	0
28	27	Ekokobe	Ngoy	Nadia	2	3	0	0	0	0	1	0	0	0	1	0
29	28	Mputu	Mbuku	Moke	3	1	0	0	1	1	0	1	0	1	0	1
30	29	Tshibanda	Lusambo	Rosalie	2	3	0	0	0	1	1	0	0	0	1	0

Figure 4. Data transformation.

7) Dimension reduction or attribute selection

To simplify the model and improve its performance, we applied a multi-step selection of key variables. First, the identification columns (number, last name, first name) were removed. Then, the correlation matrix of the numerical variables was used to identify and eliminate redundant variables with a correlation greater than 0.8, thus avoiding multicollinearity. A Random Forest model was then used to estimate the relative importance of the variables, retaining those whose importance exceeded the average.

This method is particularly suitable in medical contexts, as it allows handling heterogeneous datasets while maintaining robust performance. Random Forest-based attribute selection is widely recognized for its ability to identify the most discriminating variables, thus improving the accuracy of predictive models [18]. The final set includes, among others, age, weight, clinical variables, binary sex, as well as the diagnostic target variable.

Figure 5 shows our data after attribute reduction or selection:

8) Verification of the links between predictor variables and target variables

We analyzed the relationships between predictor variables and the target variable “Diagnosis” by examining correlations for numeric and binary variables (including Sex_Female and Sex_Male), as well as by statistically comparing their distributions across diagnostic classes. Correlation analysis and statistical tests (such as the Mann-Whitney test or the Chi-square test) are essential to ensure the validity of the relationships between attributes and the target variable in a predictive model. These methods not only allow us to detect linear associations, but also

	A	B	C	D	E	F	G	H	I	J	K	L
1	Age	Weight (kg)	Postoperative	Fever	Chills	Sweating	Headaches	Nausea	Sex_Female	Sex_Male	Pregnancy	Diagnosis
2	3	2	0	0	1	1	1	0	0	1	0	0
3	3	2	1	1	1	1	0	1	0	1	0	0
4	2	3	0	0	1	0	1	1	1	0	1	0
5	2	1	0	0	1	0	0	0	1	0	2	0
6	3	3	1	0	0	0	1	0	1	0	0	1
7	2	1	1	0	1	1	1	1	1	0	2	1
8	2	1	1	0	1	1	1	0	0	1	0	0
9	2	2	1	0	1	1	0	0	1	0	1	0
10	2	1	1	1	1	0	0	0	0	1	0	1
11	2	2	1	0	1	1	0	0	0	1	0	1
12	2	3	1	0	0	1	1	0	0	1	0	0
13	2	2	0	0	1	1	1	1	1	0	1	0
14	2	2	0	0	1	0	1	1	1	0	0	0
15	2	1	1	1	0	0	0	0	1	0	1	0
16	1	1	0	0	1	0	0	0	0	1	0	0
17	2	3	1	1	0	0	0	1	1	0	0	0
18	3	1	0	1	1	0	0	0	1	0	0	1
19	2	3	1	1	0	1	1	0	0	1	0	0
20	1	2	1	0	1	1	1	1	0	1	0	0
21	2	1	0	0	1	0	0	0	1	0	0	1
22	3	3	1	0	0	1	1	1	1	0	0	0
23	3	3	1	0	1	1	1	1	1	0	0	0
24	2	3	1	0	0	0	1	1	1	0	0	0
25	2	1	0	0	0	1	0	0	1	0	0	0
26	3	3	0	1	0	0	0	0	1	0	0	0
27	3	2	1	1	1	1	0	0	1	0	0	1
28	2	3	0	0	0	0	1	0	1	0	0	0
29	3	1	0	0	1	1	0	1	0	1	0	1
30	2	3	0	0	0	1	1	0	1	0	0	0

Figure 5. Dimension reduction or attribute selection.

to assess whether the distributions differ significantly across diagnostic groups [19].

The image below shows the correlation matrix between the different variables studied, illustrating the positive or negative relationships between them. This visualization allows us to better understand potential interactions and identify highly correlated variables that may influence the modeling (See **Figure 6**).

2.2.3. Modeling

In accordance with the modeling step of the CRISP-DM process, we implemented a machine learning algorithm to automatically predict malaria diagnosis from the collected clinical data. Our objective is to evaluate whether patient characteristics allow a machine learning model to effectively identify the presence or absence of the disease.

1) *Decision tree*

The choice of the decision tree is based both on its good performance with medium-sized datasets and on its transparency, which is crucial in a medical context. Indeed, this type of model is called “white-box”: it allows health professionals

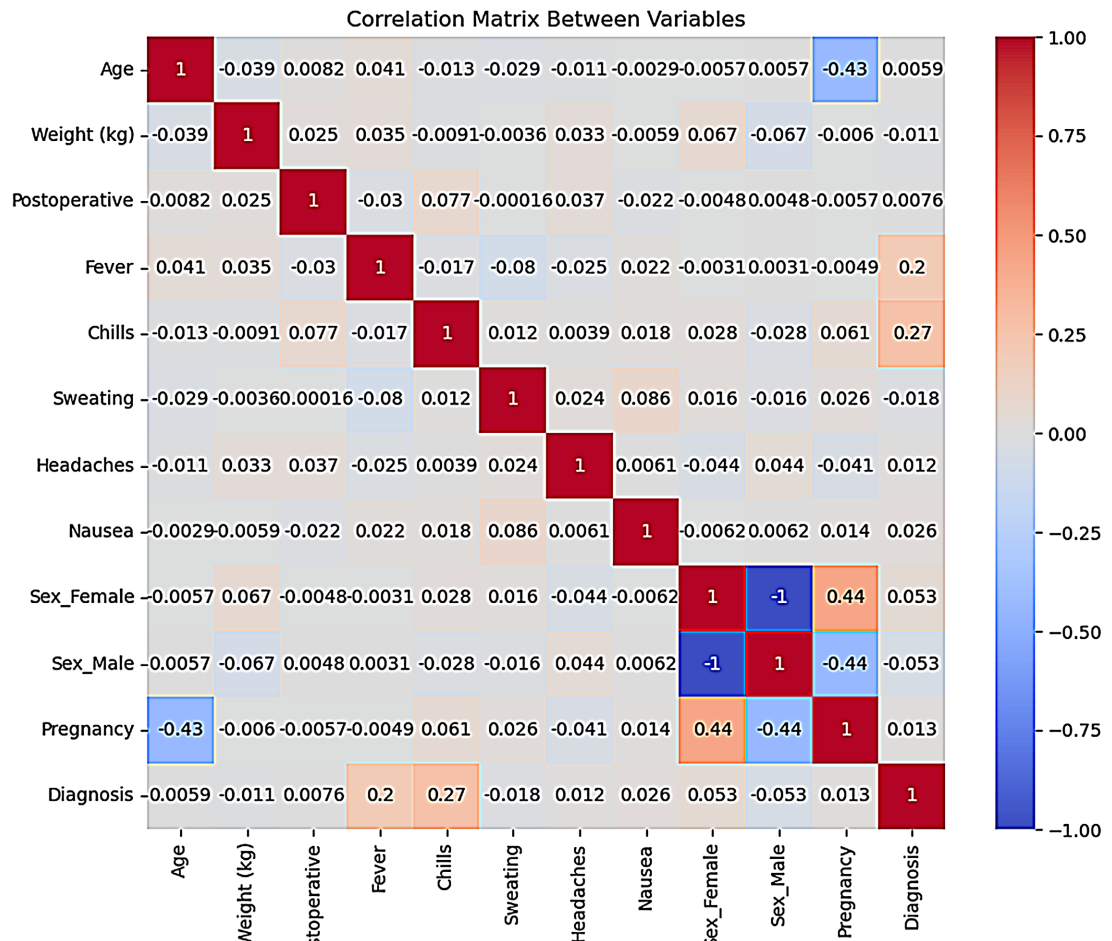


Figure 6. Correlation matrix between variables.

to understand the reasoning leading to a prediction, thus promoting confidence and clinical acceptance [20].

As Shortliffe and Sepúlveda point out, clinicians’ trust in diagnostic support tools strongly depends on the clarity of the reasoning proposed by the system [21]. This explainability is a decisive factor for integrating artificial intelligence into medical practices. In this perspective, Gambetti *et al.* also insist that interpretability is an ethical and functional requirement of clinical decision support systems (CDSS) [22].

To obtain reliable performance estimates, we applied a 10-fold cross-validation strategy instead of a single 80/20 split. At each fold, 90% of the dataset (2,250 records) were used for training and 10% (250 records) for testing. This approach reduces variance in the evaluation and ensures that all cases contribute to both training and testing. The final reported metrics correspond to the mean ± standard deviation across the 10 folds.

In addition, to compensate for the imbalance between malaria and non-malaria cases, the decision tree was trained with class weights, which helped maintain sensitivity for minority-class detection.

Figure 7 illustrates the structure of our decision tree model applied to clinical data.

An AUC ≈ 0.5 corresponds to a random model [24].

Hanley and McNeil demonstrated that AUC is a reliable measure for comparing diagnostic models [25].

Thus, the ROC curve and AUC are reference tools in medicine to evaluate and compare the performance of classification algorithms in diagnostic studies [26].

Figure 8 shows the ROC curve obtained during our study.

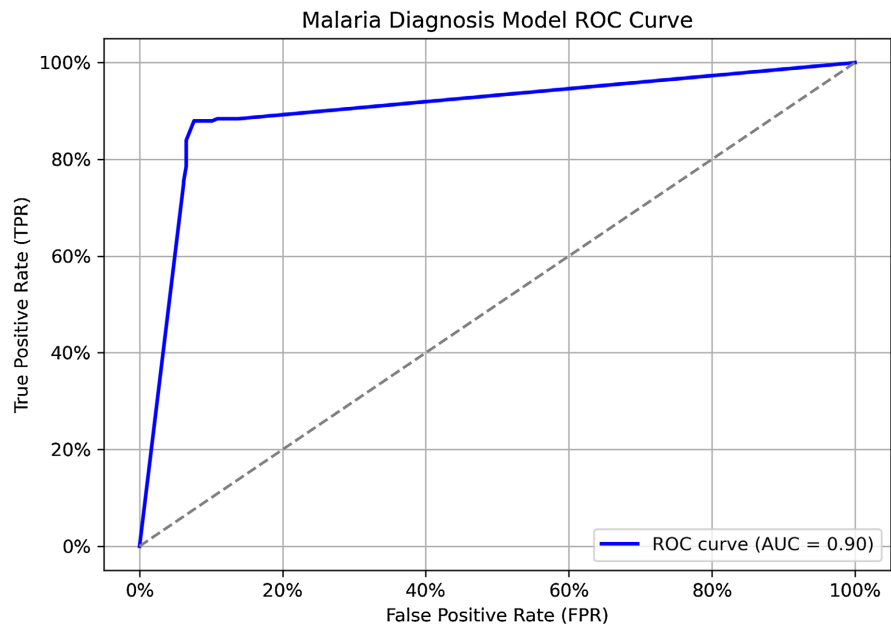


Figure 8. ROC curve.

This graph, based on a test sample of 500 patients taken from an initial set of 2500, shows excellent discriminatory ability. The curve approaches the upper left corner, indicating a high rate of correct detection of positive cases and a low rate of false alarms.

Performance is quantified by the area under the curve (AUC): a value close to 1 reflects an excellent model, while a value close to 0.5 corresponds to a random model.

Here are the details of our ROC curve results (See **Figure 9**).

2.2.4. Assessment

After training our model on 80% of the data, we evaluated its performance on the remaining 20% unused, in order to measure its ability to generalize to new observations [27]. This step is crucial to assess the model's ability to generalize its predictions to new cases, not used during training.

1) Performance metrics

To quantify the quality of predictions, several standard indicators in supervised learning were calculated:

- **Accuracy (Overall Precision):** proportion of correct predictions (true positives + true negatives) across all observations. Useful but misleading in case of **class imbalance** [28].

◆ ROC curve data with TP, TN, FP, FN rates at each threshold:

Threshold=inf | TP=0 (0.00%) | TN=276 (100.00%) | FP=0 (0.00%) | FN=224 (100.00%)

Threshold=1.000 | TP=168 (75.00%) | TN=259 (93.84%) | FP=17 (6.16%) | FN=56 (25.00%)

Threshold=0.900 | TP=169 (75.45%) | TN=259 (93.84%) | FP=17 (6.16%) | FN=55 (24.55%)

Threshold=0.889 | TP=176 (78.57%) | TN=258 (93.48%) | FP=18 (6.52%) | FN=48 (21.43%)

Threshold=0.800 | TP=188 (83.93%) | TN=258 (93.48%) | FP=18 (6.52%) | FN=36 (16.07%)

Threshold=0.750 | TP=194 (86.61%) | TN=256 (92.75%) | FP=20 (7.25%) | FN=30 (13.39%)

Threshold=0.667 | TP=197 (87.95%) | TN=255 (92.39%) | FP=21 (7.61%) | FN=27 (12.05%)

Threshold=0.500 | TP=197 (87.95%) | TN=248 (89.86%) | FP=28 (10.14%) | FN=27 (12.05%)

Threshold=0.400 | TP=198 (88.39%) | TN=246 (89.13%) | FP=30 (10.87%) | FN=26 (11.61%)

Threshold=0.333 | TP=198 (88.39%) | TN=242 (87.68%) | FP=34 (12.32%) | FN=26 (11.61%)

Threshold=0.250 | TP=198 (88.39%) | TN=241 (87.32%) | FP=35 (12.68%) | FN=26 (11.61%)

Threshold=0.200 | TP=198 (88.39%) | TN=239 (86.59%) | FP=37 (13.41%) | FN=26 (11.61%)

Threshold=0.143 | TP=198 (88.39%) | TN=238 (86.23%) | FP=38 (13.77%) | FN=26 (11.61%)

Threshold=0.000 | TP=224 (100.00%) | TN=0 (0.00%) | FP=276 (100.00%) | FN=0 (0.00%)

Figure 9. ROC curve data.

- **Recall (Sensitivity):** Also known as the true positive rate, this measures the model's ability to correctly detect malaria patients. This metric is crucial because low recall means that many true cases are missed (false negatives), which can have serious clinical consequences.
- **Precision:** proportion of positive predictions (malaria) that are actually correct. It assesses the reliability of positive alerts, limiting the number of false positives [29].
- **F1-score:** harmonic mean between precision and recall, offering a balanced compromise between these two metrics, particularly relevant in the presence of unbalanced classes [30].
- **Specificity:** True negative rate, i.e., the model's ability to correctly identify patients without malaria. Good specificity helps limit false positives, thus avoiding misdiagnoses in healthy patients.

These metrics were calculated by comparing the model's predictions to actual diagnoses provided by physicians on the test set, thus allowing the model's relevance in a clinical context to be accurately assessed.

2) Confusion matrix

The confusion matrix is an essential tool for analyzing model performance in detail. It is a fundamental tool for analyzing detailed model performance, it allows identifying critical errors, especially false negatives, which are essential in a medical context [31] (See Table 6).

Table 6. Confusion matrix.

	Predicted Malaria	Predicted Other
Real malaria	True positives	False negatives
Other real	False positives	True negatives

This representation makes it possible to identify critical errors (false negatives,

in particular) which impact the quality of the diagnosis.

This matrix is presented graphically in **Figure 10**.

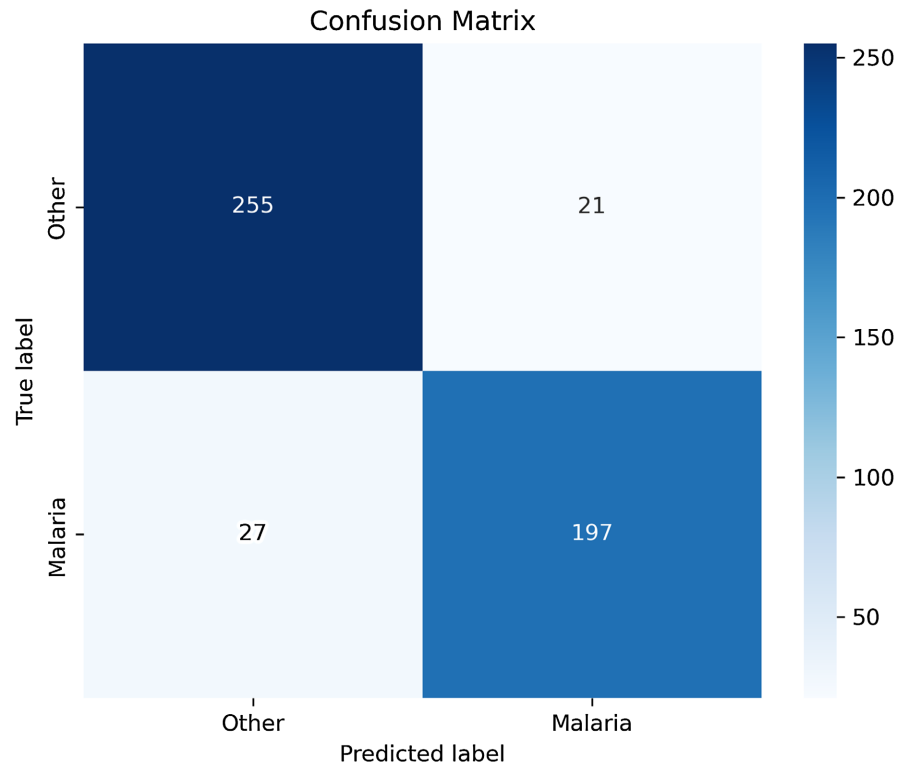


Figure 10. Confusion matrix (comparison between model predictions and actual diagnoses).

In this matrix we found:

- **True positives (TP = 197):** Patients with malaria correctly identified as such.
- **False negatives (FN = 27):** Patients with malaria but not detected by the model a critical error.
- **False positives (FP = 21):** Healthy patients falsely diagnosed as having malaria.
- **True negatives (TN = 255):** Healthy patients correctly identified.

This representation allows us to clearly visualize the most sensitive errors, in particular false negatives, which here represent 12% of real malaria cases (27 out of 224).

2.2.5. Deployment

The deployment of our expert system for malaria diagnosis relies on the seamless integration of the classification model, the user interface, and a relational medical database. The latter constitutes a central element, ensuring the structured and sustainable storage of diagnosed cases.

1) *Medical database architecture*

As part of the deployment, a relational database was developed to meet clinical, technical, and ethical requirements. It allows for the storage, tracking, and consulta-

tion of patient data, thus facilitating the future use of results for research or medical monitoring purposes.

The objectives of this base are multiple:

- **Structured storage of generated diagnoses:** Each prediction made by the system is recorded, creating a usable history of the cases treated.
- **Medical reference:** Previous cases can be compared with similar new cases, enriching decision-making.
- **Traceability of decisions:** Each diagnosis is linked to clinical data, date, and model output, ensuring total transparency in the decision-making chain.
- **Support for medical research:** Stored data can be reused to refine future models or for epidemiological studies.

In order to meet the above objectives, we have adopted a modular architecture divided into three main tables:

1. Table: Patients

Table 7 gathers patients' personal and biometric information. It is essential for uniquely identifying each individual supported by the system. By centralizing data such as name, age, gender, and weight, it allows for tracking each patient's medical history, ensuring diagnostic traceability, and facilitating inter-patient comparisons during clinical studies.

Table 7. Patient table.

No.	Field	Kind	Description
1	patient_id	INTEGER (PK)	Unique patient identifier
2	name	TEXT	Patient's name
3	first name	TEXT	Patient's first name
4	sex	TEXT	Gender (Male/Female)
5	age	INTEGER	Patient's age
6	weight	REAL	Weight (in kg)
7	pregnancy	Boolean	Pregnancy Yes or No

2. Table: Symptoms

Table 8 contains the various clinical signs noted for each patient at the time of the consultation. It is linked to the Patients table by a foreign key. Its importance lies in the fact that it directly feeds the prediction model with the explanatory variables necessary for the analysis (fever, chills, headaches, etc.). It also allows us to observe the symptomatic evolution and to conduct statistical analyses on the frequency or correlation between symptoms.

3. Table: Diagnostics

Table 9 records the results produced by the artificial intelligence model for each patient. It contains the system's decision (malaria or not), the probability associated with this decision, as well as the prediction date. Its importance is crucial because it constitutes the memory of the predictions made. It allows the model's

performance to be verified on real cases, medical audits to be conducted, and medical decisions made at a given time to be documented.

Figure 11 shows the relational architecture or data schema of our expert system.

Table 8. Symptoms table.

No.	Field	Kind	Description
1	id_symptom	INTEGER (PK)	Symptom Set Identifier
2	patient_id	INTEGER (FK)	Foreign key linked to the patient
3	fever	BOOLEAN	Presence of fever
4	chills	BOOLEAN	Presence of chills
5	headaches	BOOLEAN	Headaches declared
6	sweating	BOOLEAN	Abnormal sweating
7	nausea	BOOLEAN	Nausea symptom
8	postoperative	BOOLEAN	Post-operative status

Table 9. Diagnostic table.

No.	Field	Kind	Description
1	id_diagnostic	INTEGER (PK)	Unique diagnostic identifier
2	patient_id	INTEGER (FK)	Reference to the patient concerned
3	result_prediction	BOOLEAN	1 = Probable malaria; 0 = Other
4	probability	FLOAT	Model confidence rate
5	date_prediction	DATE	Date of diagnosis

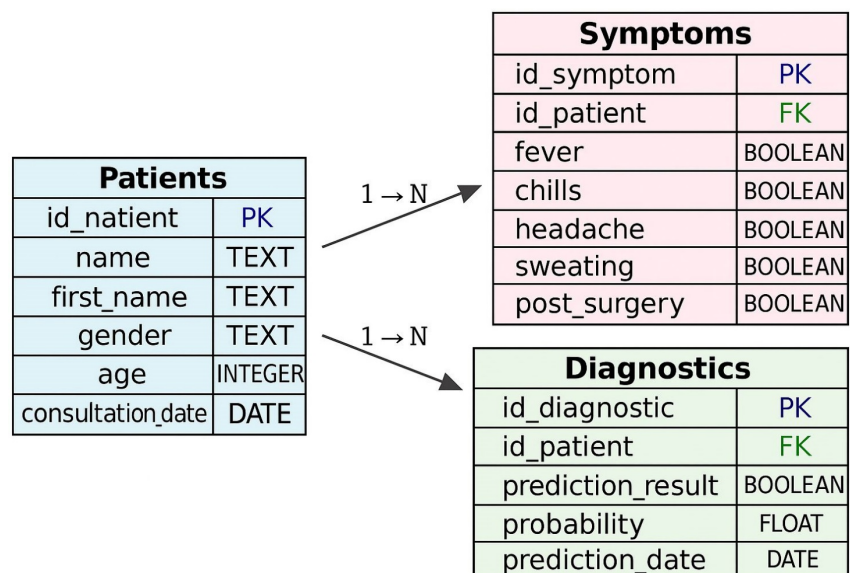


Figure 11. Database architecture.

3. Results and Discussions

The implementation of our malaria diagnostic expert system at the University Clinics of Lubumbashi proved particularly instructive. It allowed us to assess the relevance of the developed model in a real-world environment, taking into account local practices, technical constraints, and the specific needs of medical staff.

3.1. Analysis of the Results Obtained

Tests performed on a sample of 500 patients from a global dataset of 2500 showed that the decision tree-based machine learning model offers robust and reliable performance.

The overall accuracy of 90.4% indicates that the model correctly classifies the majority of cases. The recall (sensitivity) of 88% shows that the system effectively detects most malaria patients, which is crucial to limit false negatives with serious clinical consequences. The precision (90%) ensures that positive alerts are reliable, while the F1-score (0.89) reflects a good balance between recall and precision.

The high specificity (92%) confirms the model's ability to limit false positives, thus avoiding unnecessary overdiagnosis. In addition, the MCC coefficient (0.81) highlights a very good overall quality of the classification, even in the event of imbalance between classes.

These results demonstrate that the model can be an effective diagnostic aid tool, capable of supporting medical decisions while accelerating treatment.

It is important to note that other machine learning models, such as SVMs (Support Vector Machines) or neural networks, could have been considered. However, despite their sometimes-superior performance, these models are often considered "black boxes" due to their lack of transparency. In clinical settings, this can be a major barrier to their adoption, as practitioners need to be able to justify medical decisions. Ribeiro *et al.* (2016) emphasize the importance of explainability in strengthening end-user trust in intelligent systems [32].

3.1.1. Main Clinical Data Entry Interface

As part of the system deployment, a user-friendly graphical interface was developed to facilitate the entry of clinical data by healthcare professionals. This interface allows for the structured entry of essential patient information (name, age, gender, weight), as well as observed symptoms such as fever, chills, headaches, sweating, or nausea. Once the data is entered, the system triggers the prediction process and displays the diagnostic result along with a confidence rate. This interface plays a crucial role in the operationalization of the system, ensuring intuitive handling, rapid data entry, and clear restitution of the medical verdict (See **Figure 12**).

3.1.2. System Limits

Despite these satisfactory performances, several limitations must be highlighted:

- **Data quality:** The system relies heavily on the quality and accuracy of the data entered. Any errors or omissions in symptom collection can impair the model's performance. Proper user training is therefore essential.

ID	LAST NAME	MIDDLE NAME	FIRST NAME	AGE	WEIGHT	GENDER	POST_OP	FEVER	CHILLS	SWEATING	HEADACHE
4	DYUBA	KABANGO	DESIRE	36.0	75.0	Female	1	1	1	0	1
3	TUBONGYE	SHABANI	ARLENE	30.0	75.0	Female	1	1	1	0	1
2	LUKUSA	MUVUALA	VICKY	57.0	95.0	Male	1	1	1	1	1
1	MAZUNZE	MAZUNZE	BERTIN	27.0	65.0	Male	1	1	1	1	0

Figure 12. Input form.

- **Contextual factors not integrated:** The model does not yet integrate epidemiological or environmental variables (seasonality, geographic location, history of epidemics), which could improve the relevance of the predictions.
- **Single-user deployment:** The system currently operates locally without the ability to centralize or share data at the institutional or regional level, limiting coordination and overall epidemiological surveillance.

3.1.3. Career Prospects

In light of the initial tests carried out at the University Clinics of Lubumbashi, several avenues for improvement can be considered:

- ✓ **Larger-scale deployment:** Connect multiple workstations to a central database to facilitate the collection, aggregation and global analysis of medical data.
- ✓ **Model enrichment:** integrate new clinical and biological variables (such as red blood cell count, precise body temperature, etc.) to increase diagnostic accuracy.
- ✓ **Development of a mobile version:** making the system accessible via a light-weight Android application, intended for community health workers or field interventions.
- ✓ **Continuing education:** ensuring lasting ownership of the system by medical staff through training and support sessions.
- ✓ **Longitudinal monitoring:** establish a mechanism for monitoring diagnosed cases in order to measure patient progress and adjust therapeutic recommendations over the long term.

By making the diagnostic tool accessible, transparent and adapted to local realities, this study is part of a logic of equitable and innovative medicine. It also paves the way for other digital health initiatives in low-income countries, where AI can have a direct impact on the care provided to vulnerable populations [8].

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Daneluzzo, L., Daneluzzo, M., Thellier, F., *et al.* (2025) "Severe Imported Malaria in Children in Metropolitan France, 2011-2023." *Medecine et Maladies Infectieuses*, **4**, S24.
- [2] Mbanefo, A. and Kumar, N. (2020) Evaluation of Malaria Diagnostic Methods as a Key for Successful Control and Elimination Programs. *Tropical Medicine and Infectious Disease*, **5**, Article 102. <https://doi.org/10.3390/tropicalmed5020102>
- [3] World Health Organization (2025) World Malaria Report 2024. <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2024>
- [4] World Health Organization (2025) World Malaria Report 2023. <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023>
- [5] Acherar, I.A. (2024) Detection and Identification of *Plasmodium falciparum* on Microscopic Images. PhD Thesis, Sorbonne University. <https://theses.hal.science/tel-04698420/>
- [6] Rajaraman, S., Antani, S.K., Poostchi, M., Silamut, K., Hossain, M.A., Maude, R.J., *et al.* (2018) Pre-Trained Convolutional Neural Networks as Feature Extractors toward Improved Malaria Parasite Detection in Thin Blood Smear Images. *Peer Journal*, **6**, e4568. <https://doi.org/10.7717/peerj.4568>
- [7] Liang, H.Y., Tsui, B.Y., Ni, H., Valentim, S., *et al.* (2019) Evaluation and Accurate Diagnoses of Pediatric Diseases Using Artificial Intelligence. *Nature Medicine*, **25**, 433-438.
- [8] Kermany, D.S., Goldbaum, M., Cai, W.J., *et al.* (2018) Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*, **172**, 1122-1131.
- [9] Nsoesie, E.O., Buckeridge, D.L. and Brownstein, J.S. (2014) Guess Who's Not Coming to Dinner? Evaluating Online Restaurant Reservations for Disease Surveillance. *Journal of Medical Internet Research*, **16**, e22. <https://doi.org/10.2196/jmir.2998>
- [10] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., *et al.* (2011) Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12**, 2825-2830.
- [11] Shearer, C. (2000) The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, **5**, 13-22.
- [12] Marbn, S., Mariscal, G. and Segovi, J. (2009) A Data Mining & Knowledge Discovery Process Model. In: *Data Mining and Knowledge Discovery in Real Life Applications*, I-Tech Education and Publishing. <https://doi.org/10.5772/6438>
- [13] Garcia-Rios, V., Marres-Salhuana, M., Sierra-Liñan, F. and Cabanillas-Carbonell, M. (2023) Predictive Machine Learning Applying Cross Industry Standard Process for Data Mining for the Diagnosis of Diabetes Mellitus Type 2. *IAES International Journal of Artificial Intelligence*, **12**, Article 1713.

- <https://doi.org/10.11591/ijai.v12.i4.pp1713-1726>
- [14] Amato, F., López, A., Peña-Méndez, E.M., Vañhara, P., Hampl, A. and Havel, J. (2013) Artificial Neural Networks in Medical Diagnosis. *Journal of Applied Biomedicine*, **11**, 47-58. <https://doi.org/10.2478/v10136-012-0031-x>
- [15] Chapman, P. (2000) CRISP-DM .0: Step-by-Step Data Mining Guide”. SPSS Inc.
- [16] Jakobsen, J.C., Glud, C., Wetterslev, J. and Winkel, P. (2017) When and How Should Multiple Imputation Be Used for Handling Missing Data in Randomized Clinical Trials—A Practical Guide with Flowcharts. *BMC Medical Research Methodology*, **17**, Article No. 162.
- [17] Tan, P.-N., Steinbach, M. and Kumar, V. (2016) Introduction to Data Mining. Pearson Education India.
- [18] Kursa, M.B. and Rudnicki, W.R. (2010) Feature Selection with the Boruta Package. *Journal of Statistical Software*, **36**, 1-13. <https://doi.org/10.18637/jss.v036.i11>
- [19] Altman, D.G. and Bland, J.M. (1994) Statistics Notes: Diagnostic Tests 1: Sensitivity and Specificity. *British Medical Journal*, **308**, Article 1552. <https://doi.org/10.1136/bmj.308.6943.1552>
- [20] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F. and Pedreschi, D. (2019) A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys*, **51**, 1-42. <https://doi.org/10.1145/3236009>
- [21] Shortliffe, E.H. and Sepúlveda, M.J. (2018) Clinical Decision Support in the Era of Artificial Intelligence. *Journal of the American Medical Association*, **320**, 2199-2200. <https://doi.org/10.1001/jama.2018.17163>
- [22] Gambetti, A., Han, Q., Shen, H. and Soares, C. (2025) A Survey on Human-Centered Evaluation of Explainable AI Methods in Clinical Decision Support Systems.
- [23] Nahm, F.S. (2022) Receiver Operating Characteristic Curve: Overview and Practical Use for Clinicians. *Korean Journal of Anesthesiology*, **75**, 25-36. <https://doi.org/10.4097/kja.21209>
- [24] Lasko, T.A., Bhagwat, J.G., Zou, K.H. and Ohno-Machado, L. (2005) The Use of Receiver Operating Characteristic Curves in Biomedical Informatics. *Journal of Biomedical Informatics*, **38**, 404-415. <https://doi.org/10.1016/j.jbi.2005.02.008>
- [25] Hanley, J.A. and McNeil, B.J. (1982) “The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve.” *Radiology*, **143**, 29-36.
- [26] Çorbacioğlu, Ş.K. and Aksel, G. (2023) Receiver Operating Characteristic Curve Analysis in Diagnostic Accuracy Studies: A Guide to Interpreting the Area under the Curve Value. *Turkish Journal of Emergency Medicine*, **23**, 195-198. https://doi.org/10.4103/tjem.tjem_182_23
- [27] Sathyanarayanan, S. and Tantri, B.R. (2024) Confusion Matrix-Based Performance Evaluation Metrics. *African Journal of Biomedical Research*, **27**, 4023-4031.
- [28] Lasko, T.A., Bhagwat, J.G., Zou, K.H. and Ohno-Machado, L. (2005) The Use of Receiver Operating Characteristic Curves in Biomedical Informatics. *Journal of Biomedical Informatics*, **38**, 404-415. <https://doi.org/10.1016/j.jbi.2005.02.008>
- [29] Flach, P. (2012) Machine Learning: The Art and Science of Algorithms That Make Sense of Data. Cambridge University Press. https://books.google.com/books?hl=fr&lr=&id=Ofp4h_oXsZ4C&oi=fnd&pg=PR15&dq=P.+Flach,+*Machine+Learning:+The+Art+and+Science+of+Algorithms+that+Make+Sense+of+Data*,+Cambridge+University+Press,+2012.&ots=XMtWjqdpSM&sig=ofDk26bPb6y_EDEqA1Zit2euNo
- [30] Davis, J. and Goadrich, M. (2006) The Relationship between Precision-Recall and

ROC Curves. *Proceedings of the 23rd International Conference on Machine Learning*, Pittsburgh, 25-29 June 2006, 233-240. <https://doi.org/10.1145/1143844.1143874>

- [31] Sokolova, M. and Lapalme, G. (2009) A Systematic Analysis of Performance Measures for Classification Tasks. *Information Processing & Management*, **45**, 427-437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- [32] Bent, B., Goldstein, B.A., Kibbe, W.A. and Dunn, J.P. (2020) Investigating Sources of Inaccuracy in Wearable Optical Heart Rate Sensors. *NPJ Digital Medicine*, **3**, Article 18. <https://doi.org/10.1038/s41746-020-0226-6>