

Mouhamadou Lamine BA
Université Alioune Diop
Bambey; Senegal

ISMP DATA SCIENCE
SCHOOL
BENIN, 2022

Harnessing Data Science for Diagnosis in Health

Short bio



- **PhD in computer science and networks**
- **Associate professor at Université Alioune Diop, Senegal**
- **Member of the Medical Informatics and ICTL research group**
- **Holding and co-coordinating the AI4CARDIO project funded by AFD**
- **Project Engineer in the Responsible AI Lab hosted by KNUST and funded by IRDC and SIDA**

Healthy vs. Unhealthy



shutterstock.com · 1733873084

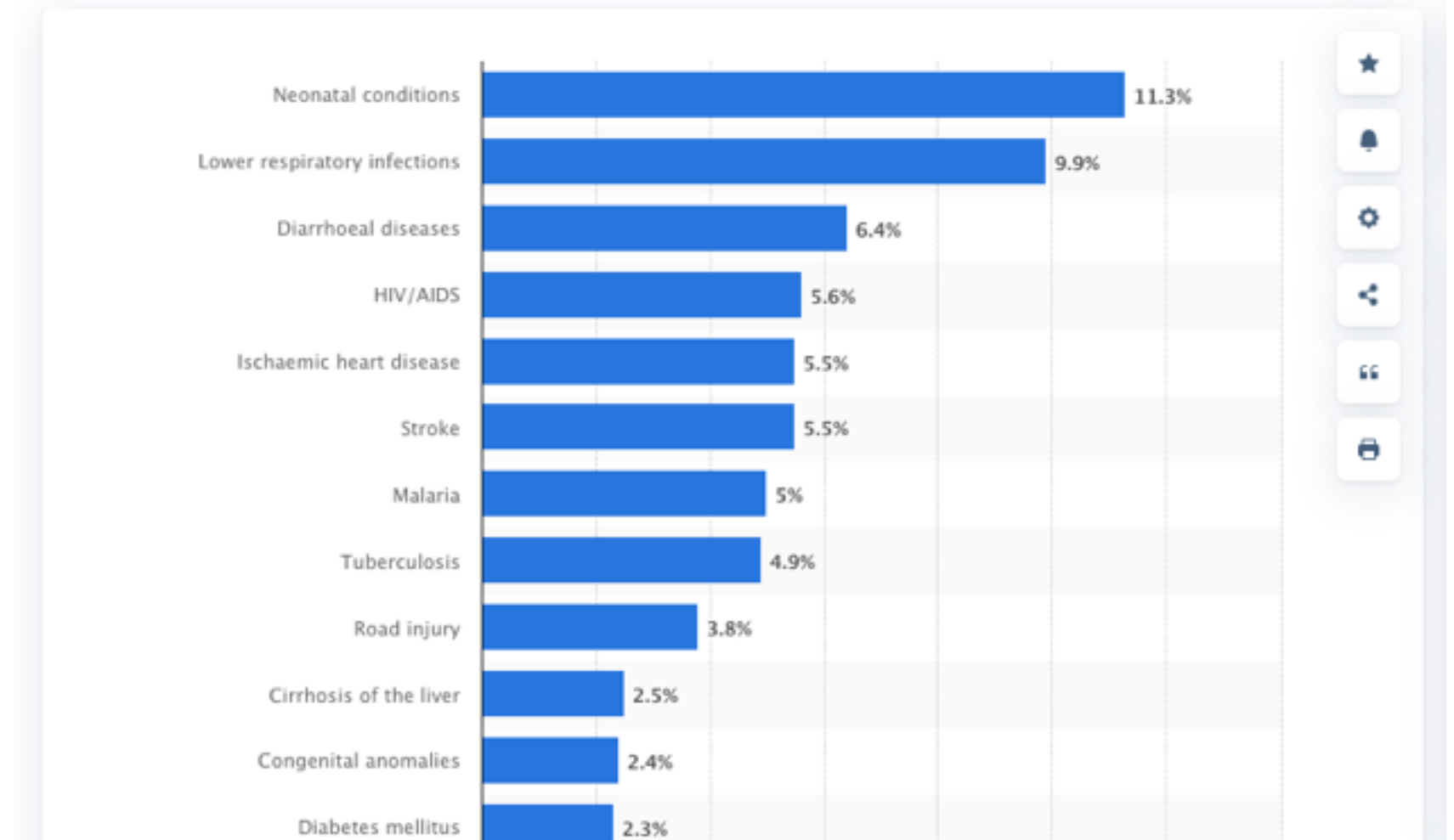


alamy

alamy.com

But diseases
are always
there

Distribution of the leading causes of death in Africa in 2019



Source : <https://www.statista.com/statistics/1029337/top-causes-of-death-africa/>

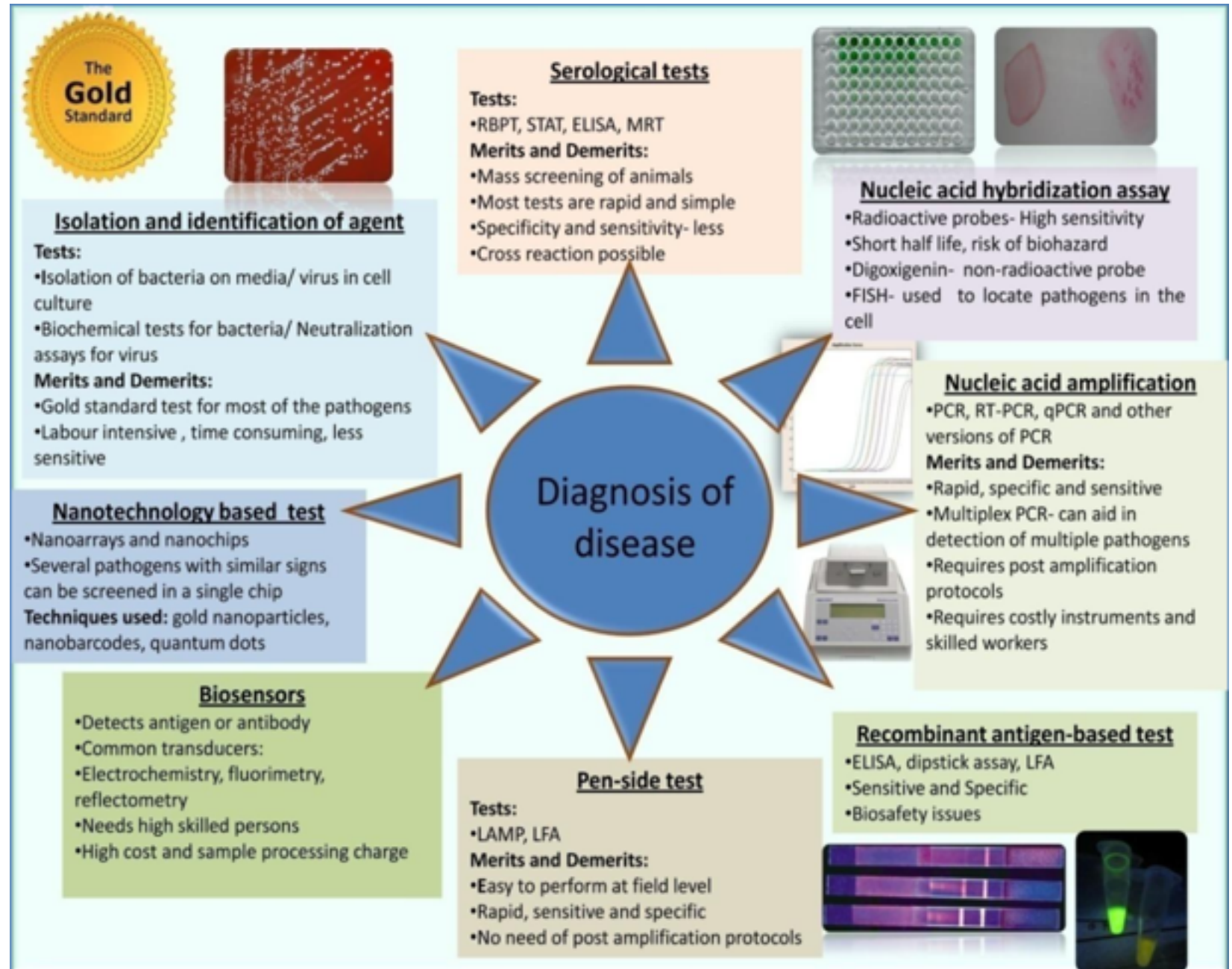
Diagnosis is a key in disease identification and treatment

Early and accurate diagnosis of infectious disease is critically important because:

- **Diagnosis can improve the effectiveness of treatments and avoid long-term complications for the infected patient.**
- **Undiagnosed patients can unknowingly transmit the disease to others. Early diagnosis can help to prevent or stop an outbreak.**
- **Widespread overuse and misuse of antibiotics contribute to antibiotic resistance. Diagnostic tests can determine when antibiotics are an appropriate treatment—and when they are not.**

(Source : MEDICAL TECHNOLOGY life changing innovative)

Types of diagnostic methods

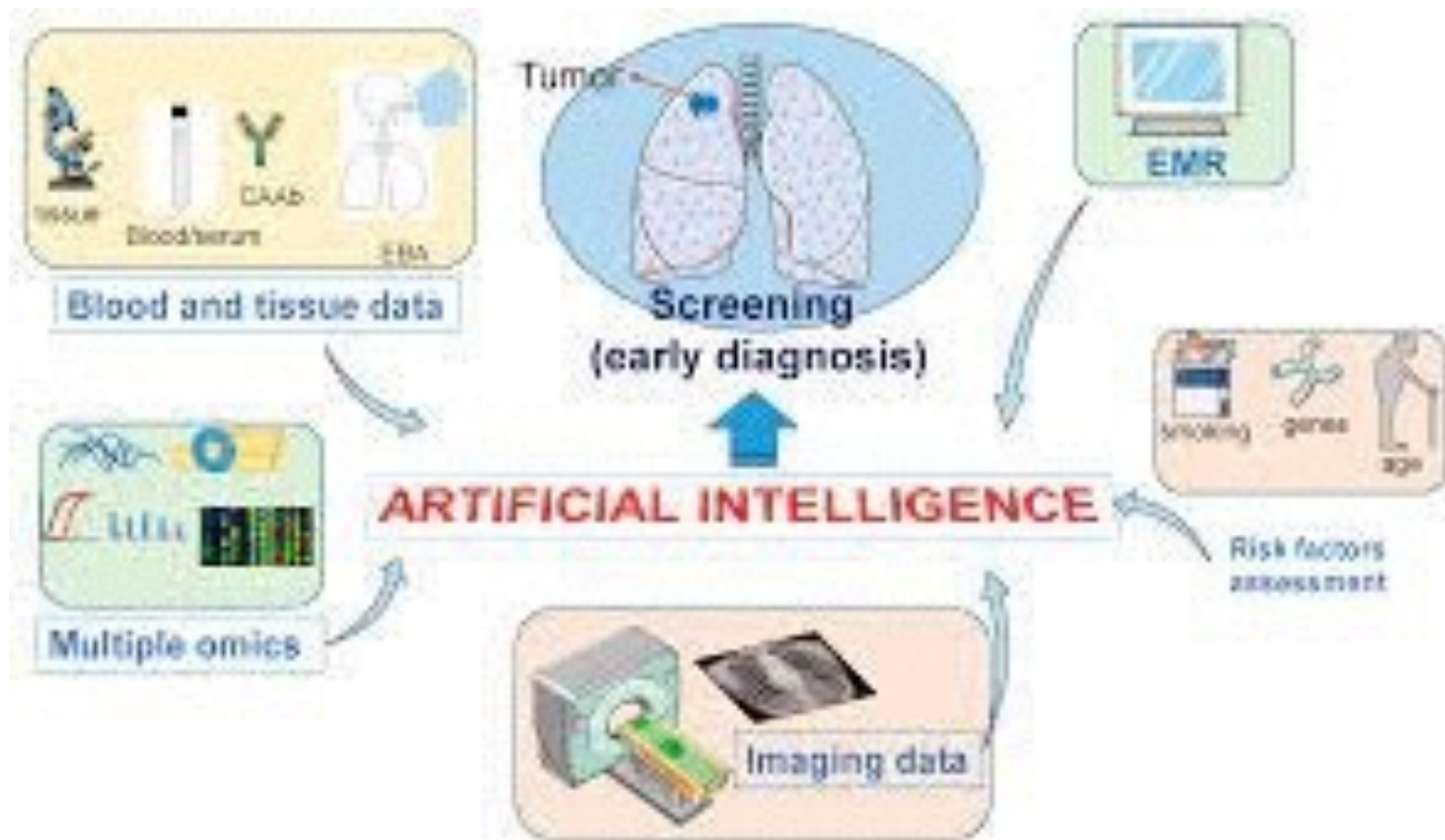


Challenges of the diagnosis of diseases in Africa

- Lack of appropriate diagnosis tools
- Lack of enough or well trained specialists



AI-based disease diagnosis



Outline



Use cases



Data Sources



Analysis
methodology &
Data Science
Tools



Prediction of
Malaria



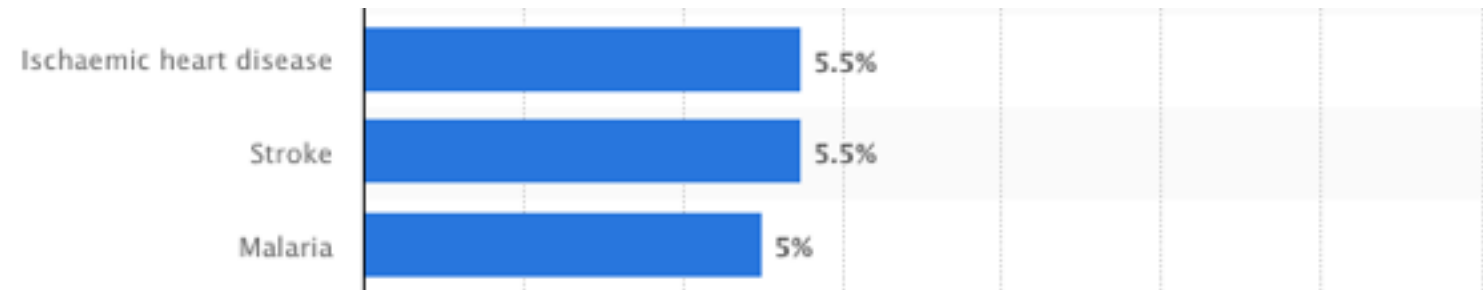
Prediction of
risk factors for
CVDs



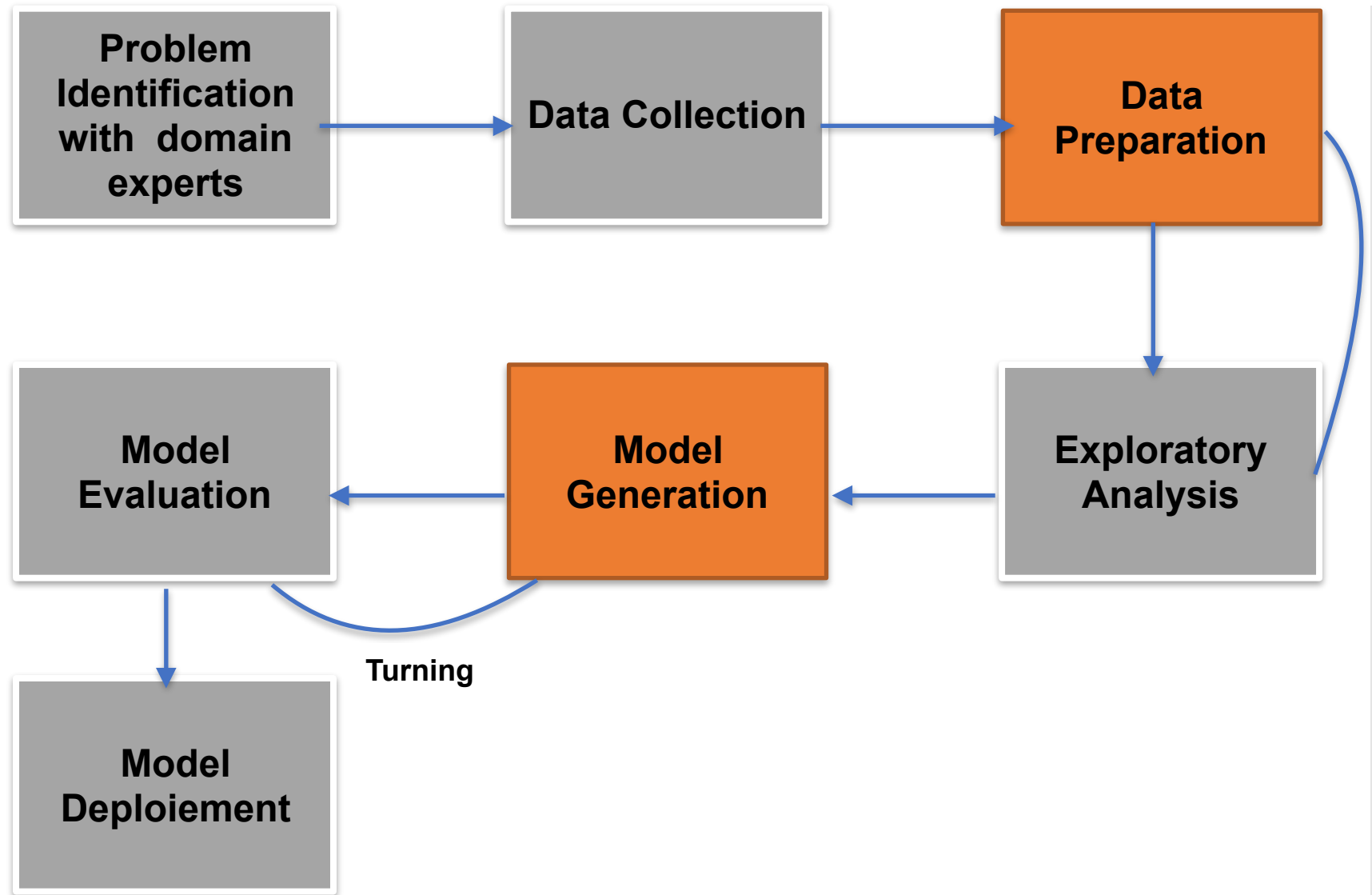
Challenges

Use cases

- Prediction of Malaria using ML algorithms
- Prediction of stroke

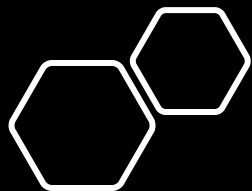


Analysis Methodology



Used Data Science Tools





Data Sources

Two main data sources



Existing patient records on registries

HCAHPS Survey

SURVEY INSTRUCTIONS

- You should only fill out this survey if you saw the patient during the hospital stay named in the cover letter. Do not fill out this survey if you were not the patient.
- Answer **ALL** questions by marking the box to the left of your answer.
- You are sometimes told to skip over some questions in this survey. When this happens you will see an arrow with a note that tells you what question to answer next, like this:
 Yes
 No → If No, Go to Question 1

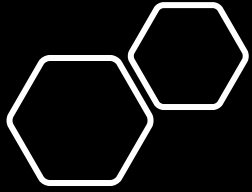
You may notice a number on the survey. This number is used to let us know if you returned your survey so we don't have to send you reminders.
Please note: Questions 1-8 in this survey are part of a national initiative to measure the quality of care in hospitals. QMR #0000000000

Please answer the questions in this survey about your stay at the hospital named on the cover letter. Do not include any other hospital stays in your answers.

YOUR CARE FROM NURSES

<p>1. During this hospital stay, how often did nurses meet you with questions like "How are you?"</p> <p><input type="checkbox"/> Never <input type="checkbox"/> Sometimes <input type="checkbox"/> Usually <input type="checkbox"/> Always</p> <p>2. During this hospital stay, how often did nurses explain things to you?</p> <p><input type="checkbox"/> Never <input type="checkbox"/> Sometimes <input type="checkbox"/> Usually <input type="checkbox"/> Always</p>	<p>3. During this hospital stay, how often did nurses explain things to you in a way you could understand?</p> <p><input type="checkbox"/> Never <input type="checkbox"/> Sometimes <input type="checkbox"/> Usually <input type="checkbox"/> Always</p> <p>4. During this hospital stay, after you pressed the call button, how often did you get help as soon as you wanted it?</p> <p><input type="checkbox"/> Never <input type="checkbox"/> Sometimes <input type="checkbox"/> Usually <input type="checkbox"/> Always <input type="checkbox"/> I never pressed the call button</p>
--	--

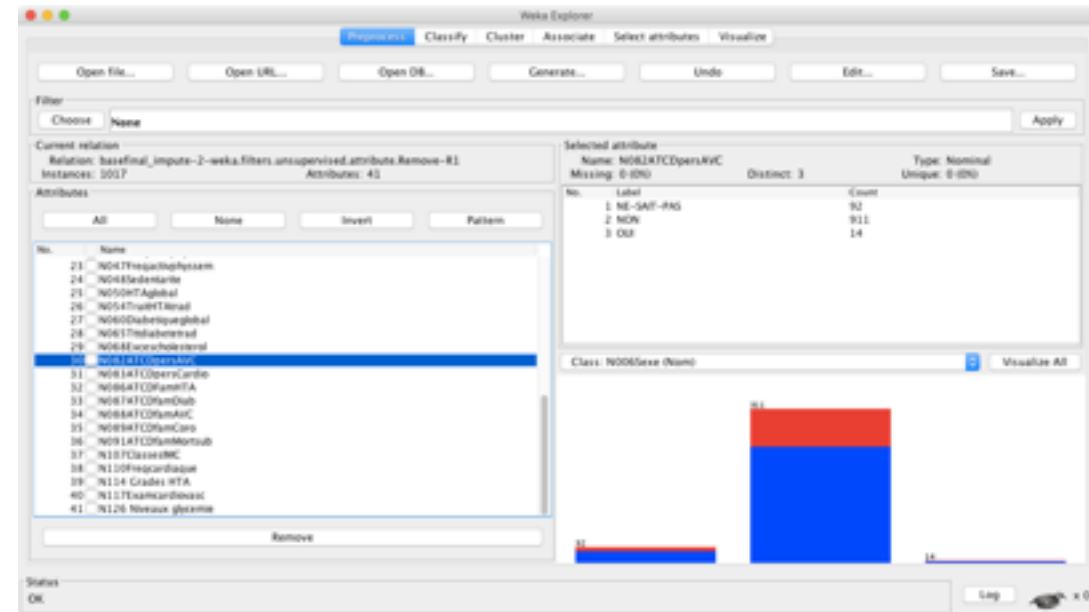
Surveys



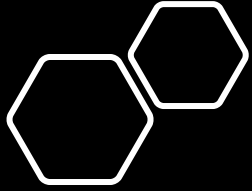
Demo

The screenshot shows the OpenRefine interface with a data table. The table has columns for various attributes and rows of data. The interface includes a sidebar with filters and a main table area.

OpenRefine for data preprocessing
(<https://openrefine.org/>)

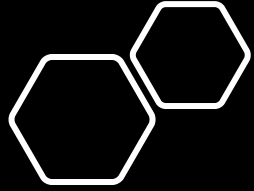


Weka for exploratory analysis & Model Generation
(<https://www.cs.waikato.ac.nz/ml/weka/>)



Malaria prediction

Can we do better with ML models than Rapid Diagnostic Test using sign and symptom data ?

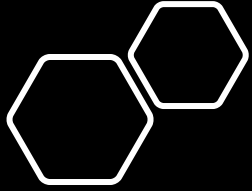


Malaria prediction

Data description

Variables	Observations	Variables types		Classes	
		Numeric	Boolean	Malaria	Not Malaria
16	21083	2	14	614	20469

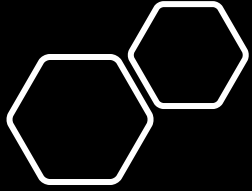
Table 1. Main characteristics of our real-world dataset of patients in Senegal



Malaria prediction

Data preparation

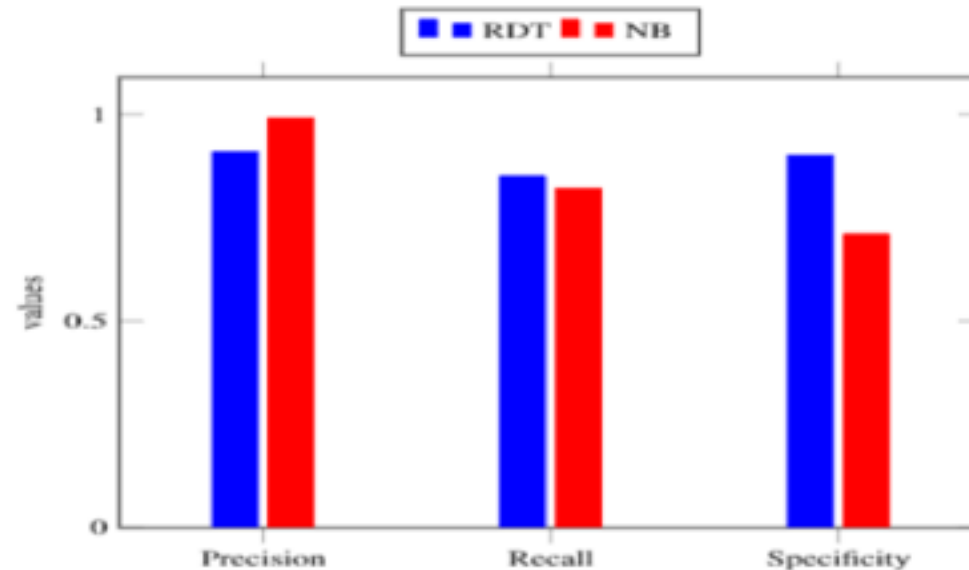
- Data cleaning and normalization
- Feature extraction based on experts of the domain
- Missing data imputation
- Over-sampling of data

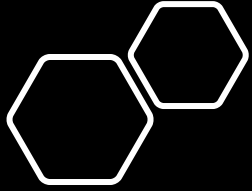


Malaria prediction

Models and performance

ML Algorithms	Precision	Recall	F1-score	AUC	Specificity
Decision Tree	0.99	0.84	0.91	0.76	0.58
Random Forest	0.99	0.84	0.91	0.76	0.60
Logistic Regression	0.90	0.78	0.88	0.84	0.75
Naive Bayesian	0.99	0.82	0.90	0.84	0.71
Support Vector Machine	0.99	0.86	0.92	0.80	0.62
Artificial Neural Networks	0.99	0.84	0.91	0.79	0.65

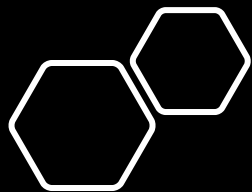




Prediction of Stroke

Can we predict accurately the type of stroke based on risk factors with ML?

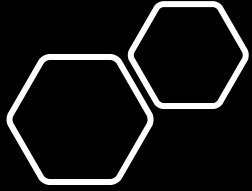




Prediction of Stroke

Data description

Name	Value
Rows	1,018
Columns	43
Discrete columns	32
Continuous columns	11
All missing columns	0
Missing observations	1,210
Complete Rows	543
Total observations	43,774

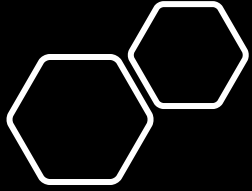


Prediction of Stroke

Data preparation

- Removal of useless attributes
- Data values normalization
- Missing data imputation
- Handling inconsistencies

Models & Performance



Prediction of Stroke

Algorithm	Accuracy (%)
K-Nearest Neighbors	95
Random Forest	93
Decision Tree	86

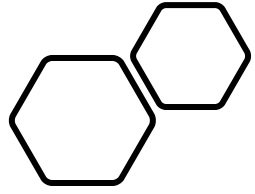
General Challenges

Deployment and use:

- Design the physical equipment that will host the model
- Evaluate the model in a real condition
- Train doctors

Other challenges:

- The lack of in-domain datasets.
- Sharing the data and privacy.
- Interpretability of the models.
- Sensitivity of medical scenarios



Thanks
Any
questions?

mouhamadouamine.bai@uadb.edu.sn
<https://sites.google.com/a/uadb.edu.sn/mlba/>

