
Disease Prediction and Self-Diagnosis: i-Diagnose

Ildem Sanli

APRIL, 22th | PARIS

DAFT FEB22

Why Self-Diagnosis?

- Limited availability of medical personnel
- Transportation barriers
- High costs

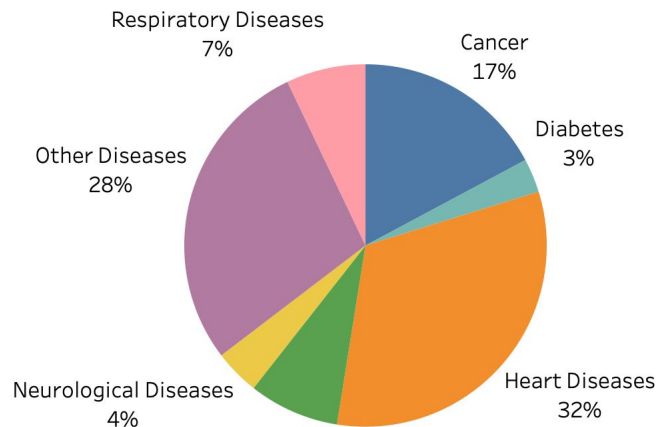


Late diagnosis

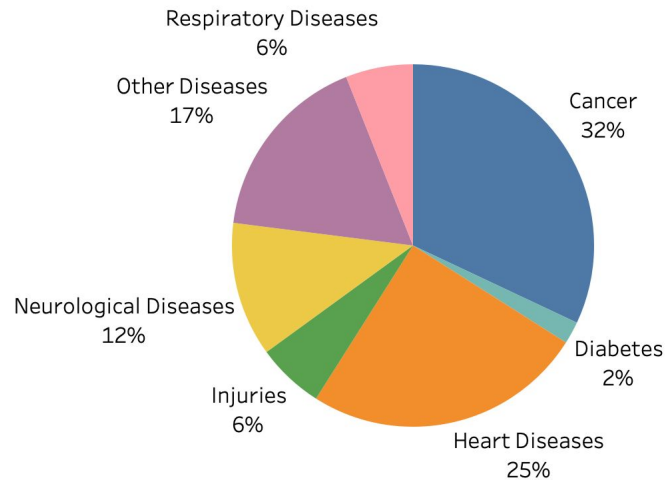
Self-diagnosis could be an initial step to encourage people to seek medical advice

Causes of Mortality

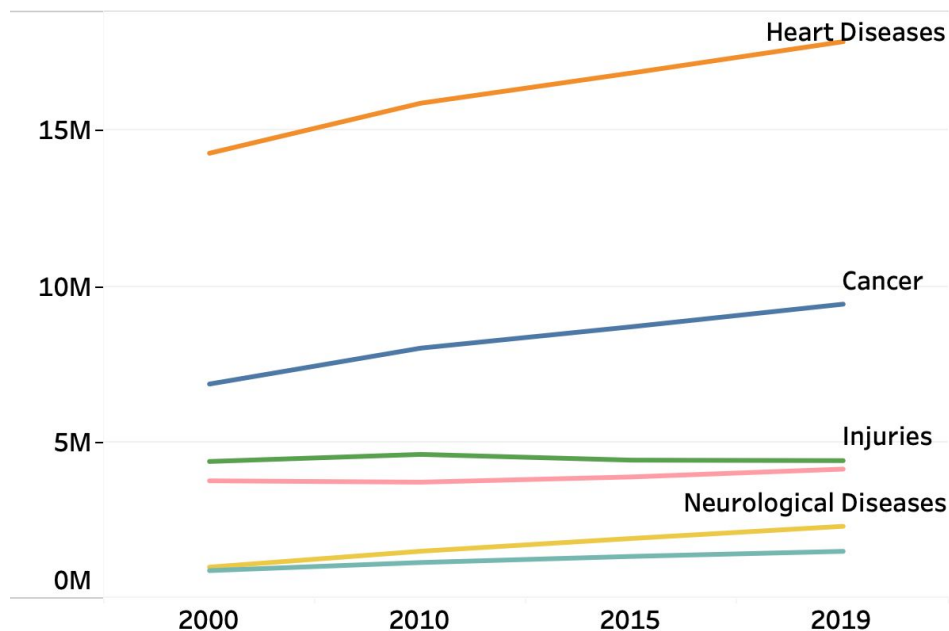
Worldwide



France



Upward trend in the recent years



2000	2010	2015	2019
14 281 200	15 887 600	16 859 300	17 864 100
28%	31%	32%	32%

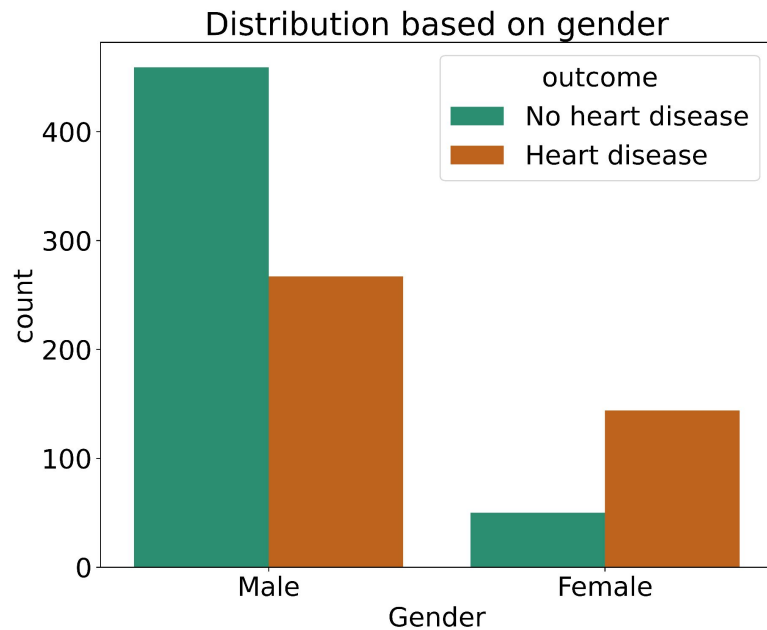
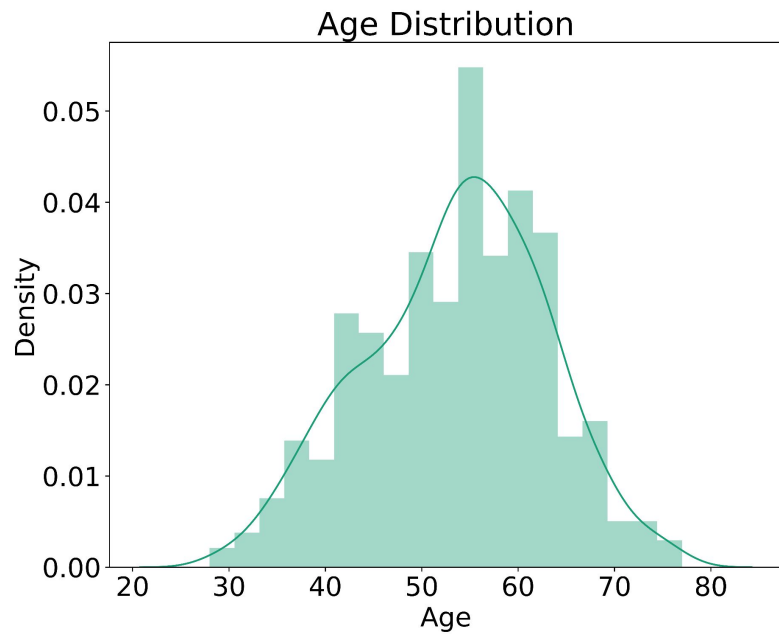
Plan

- Analyze patient data about heart diseases
- Create a machine learning model for disease prediction
- Create a self-diagnosis tool for easy access and cost-efficient diagnosis

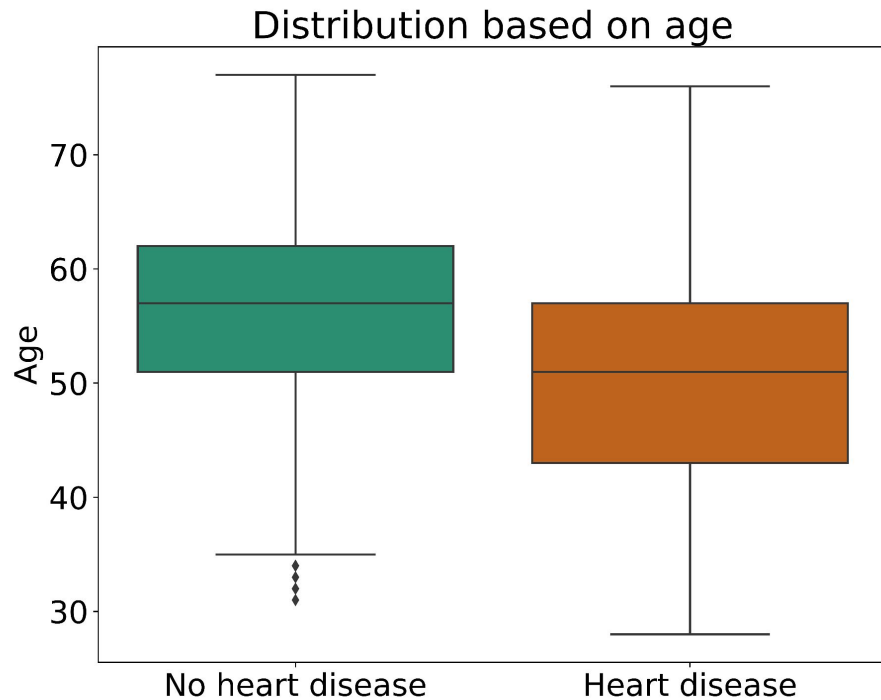
Data

- Historical patient data from 4 medical centers
- 920 rows and 14 columns
- Outcome column showing if the patient had heart disease

EDA



EDA



Data Preparation

- 13 features → age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise induced angina, oldpeak — ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, number of major vessels and Thalassemia.
- Dropped 87 rows with missing values and 172 rows with 0 in cholesterol column
- 661 rows, 5 columns

Supervised Learning on Patient Data

TPOT Classifier to pick the best model

```
StackingEstimator(estimator=BernoulliNB(alpha=0.001, fit_prior=True)),  
    MaxAbsScaler(),  
    Normalizer(norm="l1"),  
    ExtraTreesClassifier(bootstrap=False, criterion="gini", max_features=0.7500000000000001, min_samples_leaf=8,  
min_samples_split=4, n_estimators=100)
```

Bernoulli NB: Naive-Bayes classifier designed for binary/boolean features, suitable for discrete variables

Extra Trees Classifier: uses a number of randomized decision trees and takes the average prediction

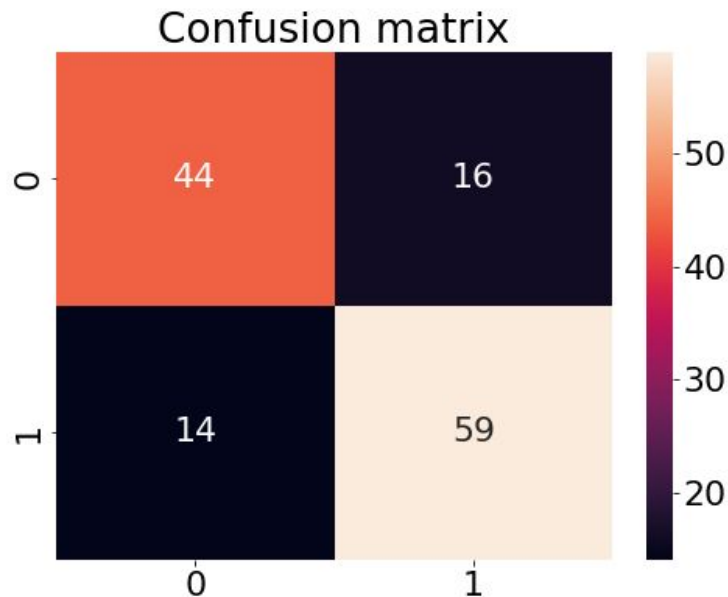
Model Evaluation

Accuracy on train data: 0.81

Accuracy on test data : 0.77

Precision score: 0.77

Recall score: 0.77



Streamlit App for Self-Diagnosis: demo

i-Diagnose

Conclusion

- Created a tool that allows self-diagnosis
- Early diagnosis step to direct patients to appropriate medical care and reduce mortality due to heart diseases