



### Machine Learning Fundamentals: From Theory to Practice

### **Heart Disease Classification Using Machine Learning Algorithms**

By Rabab Asr, Enjy Fargally, Zeyad Radwan, Aliaksei Hurnovich and Esraa Hussien.

### **Supervised By:**

Dr. Mohammed Ibrahim Prof. Andrey Nechaevsky

Meshcheryakov laboratory of information technologies, JINR



### 1.Introduction

### 2.Dataset overview

3.Methodology

4.Results

5. Other Real-world projects

### **1. Introduction**

#### Background:

- Medical diagnoses often rely on expert opinions, but in heart disease cases, consensus is difficult due to varying patient symptoms. To improve early detection and treatment outcomes, researchers are developing new methods to identify heart disease in its early stages.
- Machine Learning (ML) is a field of AI that enables computers to learn from data and make predictions without being explicitly programmed.

#### **Project Objective:**

- Develop a predictive model using clinical and demographic data to identify patients at risk for heart disease.
- This project uses supervised learning, specifically classification, to predict heart disease based on patient medical data.

#### ALGORITHMS Automated instructions

#### ARTIFICIAL INTELLIGENCE

Programs with the ability to mimic human behavior

#### MACHINE LEARNING

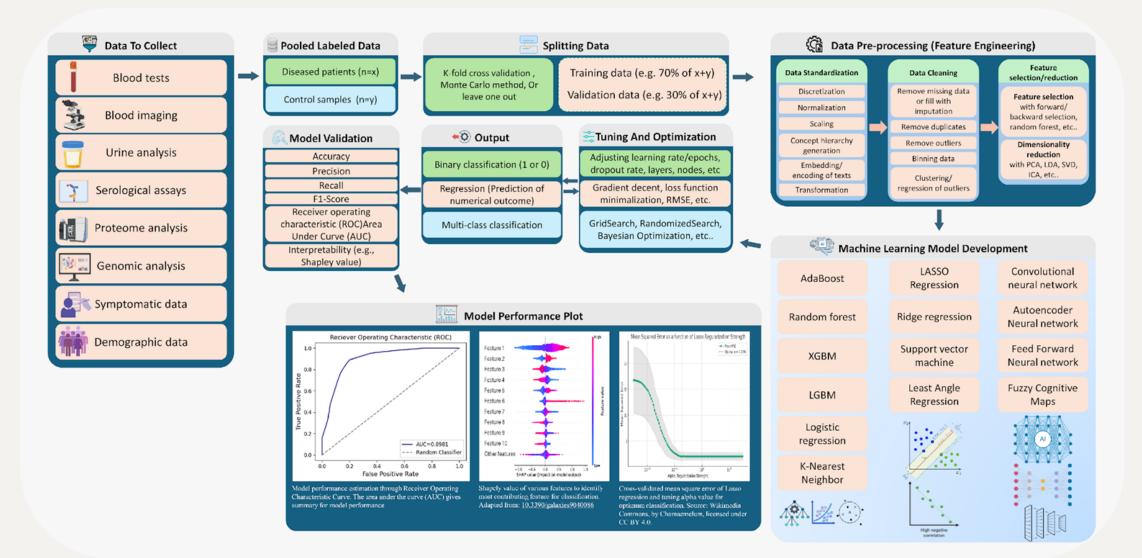
Algorithms with the ability to learn without being explicitly programmed

#### DEEP LEARNING

Subset of machine learning in which artificial neural networks adapt and learn from vast amounts of data

FIG.1: AI and its subcategories

### **1.1.** General Workflow For Classification AI dev.



## 2. Dataset Overview

- Source:
  - Dataset of 303 patients with 13 clinical features and a binary target variable (target: 1 = heart disease, 0 = no heart disease).
  - Obtained from Machine Learning Repository at <u>https://www.openml.org/search?type=data&statu</u> <u>s=active&id=43672</u>.
- Key Features:

Feature	Description & Value Meaning				
sex	1 = male, 0 = female				
chest_pain_type	1 = Typical angina (classic heart-related pain), 2 = Atypical angina, 3 = Non-anginal pain, 4 = Asymptomatic (no pain)				
fasting_blood_sugar	1 = true (>120 mg/dl), 0 = false (≤120 mg/dl)				
target	1 = heart disease present, 0 = no heart disease				

Feature	Description				
Age	Age of participants				
resting blood pressure	blood pressure upon admission				
cholesterol	serum cholestoral conc (mg/dl)				
fasting blood sugar	>120 mg/dl = 1   <120 mg/dl =0				
resting_ecg (electrocardiograph)	<ul> <li>e: normal</li> <li>having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of &gt; 0.05 mV)</li> <li>showing probable or definite left ventricular hypertrophy by Estes' criteria</li> </ul>				
max_heart_rate	maximum heart rate achieved				
exercise_angina	exercise induced angina (1/0)				
oldpeak	ST depression induced by exercise relative to rest				
ST_slope	<pre>slope of the peak exercise ST segment 1: upsloping 2: flat 3: downsloping</pre>				
ECG measurem	Blood measurements				
	Blood Test Results Levels Glycaemic C Fasting 4.4 - 6.1 mm HbA1c < 6.5% Lipids Triglycerides \$ 1.7 mm HDL cholesterol \$ 2.6 m Exercise 150 minut				

horizontal

downsloping

upslopin

**Blood Pressure** 

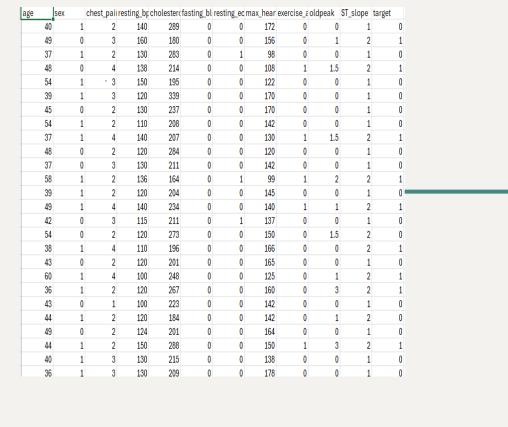
Normal Renal Function

4

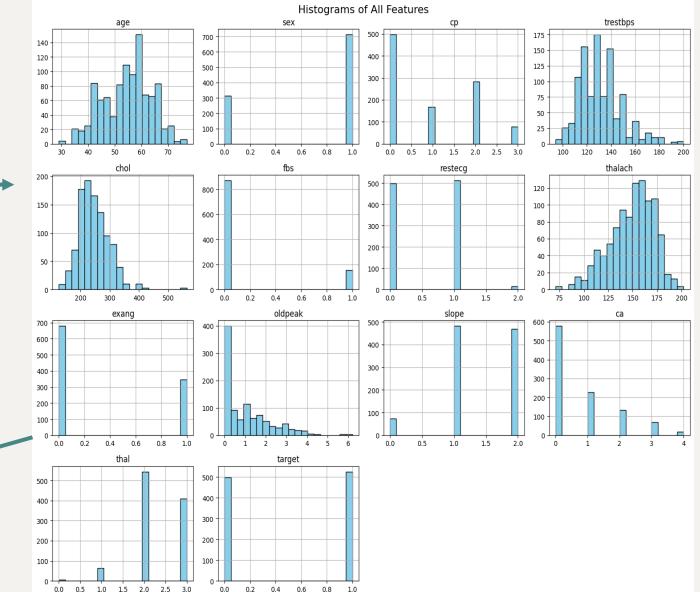
≤ 130/8

< 1251

### **2.1. Exploratory Data Analysis**



- Checking if the data was unbiased, and non-randomly generated.
- Checking distribution, range, outliers, duplicates, etc..



### 3. Methodology

#### Preprocessing:

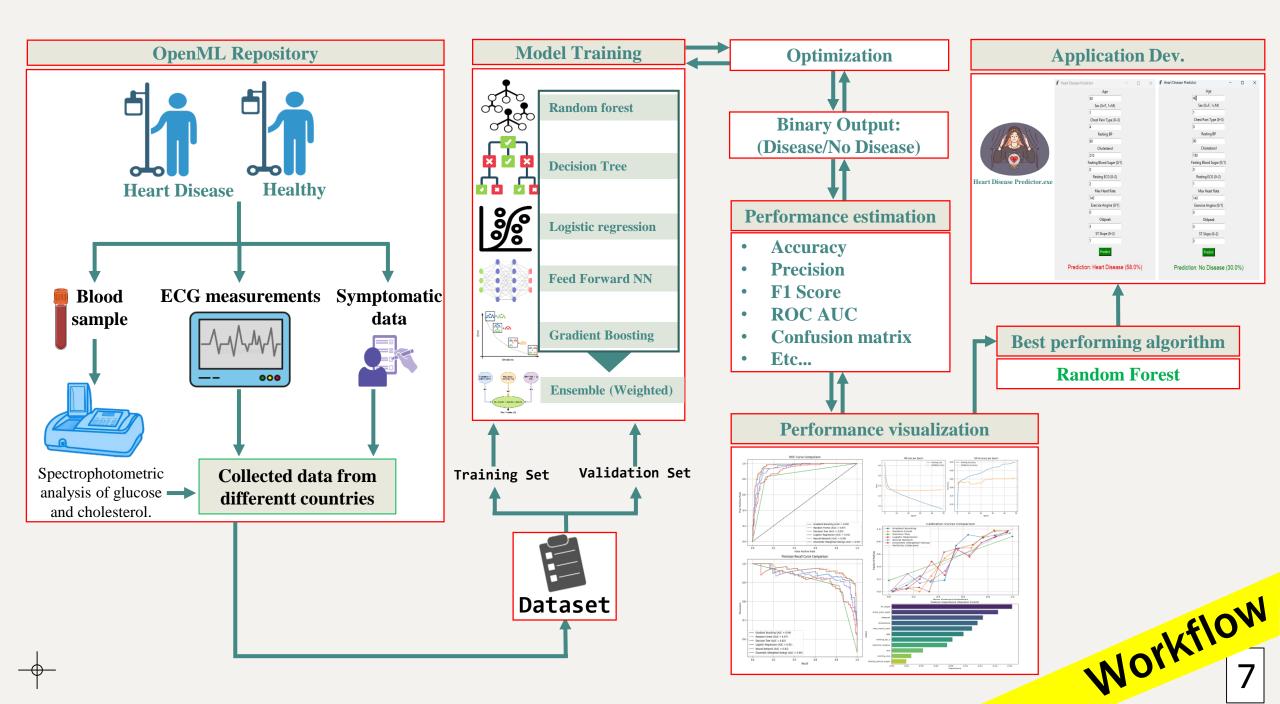
- Checked and cleaned data (no missing values in provided sample)
- Standardized numerical features

#### Model Selection:

- Random Forest Classifier for its accuracy and interpretability
- Gradient Boosting
- Neural network
- Decision Tree
- Logistic regression
- Ensemble model

#### Training & Evaluation:

- Data split into training and testing sets (e.g., 80/20)
- Performance measured by accuracy, precision, recall, F1-score, and ROC-AUC



#### 1.61 V68.28 92.65 296 252-16 980.01 2.02 569.92 685 8. 10.22 10 .02 So. 2 190 ~ S N \$ 842 20 .8 00 196 .09. 00 915 819.24 10 Sis 4.4 5.65 166.54 40 2861 209.26 V11.68 20 64 12:29 1st 542.48 091 10. 61.82 .9 20 250 5 250-25 000 1530-150

250 cosi or 19:15

.00

100

13.00

123.62

20

200.50

20002

.00

2000

VSLIV

.0

50

619.19

0.

212.91

8.

500

11.08

016:12

P.

62

222

N'

32.029

38.713

1000 CO

20.282

S.

12.101

91.

17.653

2.514

16:55

92.908

201

7.321

0 099

85.08

10.02

10.34

15

N

401.58

.00

1000

100

67-291

51.919

63.81

155

.S.

0

11.57

.60

250

63.032

039

29.131

10

23.323

12.035

51.312

.0

219

60.92

845.61

1st

19.198

16.198

860

2.

6 619 210.00

128.57

598.11

151-69

.0

299.23

es la

.2

120.91

61-28

010.91

199.11

100:10

.5

10.95

~

00:69

1266

19.973

83.155

69.593

73:601

81.201

.ci

.00

100 025.01

.¢,

0,00,000

N.V.

123.00

sil

001

0.

7.10

30.932

96.921

23.121

2005

051

01.859

NA.TA

2

0.109

36.55

not

3.

103

02

N1.1899

S.

53.026

21

125.69

592.24

.8.

81.69

Y

650.05

6.0. 170

.08

012

.2

Not

10.

52.825

30

16.00

1 AS

S.

16.312

6.129

67.672

156.255

00

19:58

26.39

15.11

5

20.112

10h

52.11

28. 69.

16.996

20.148

1:02

01.833

201.00

NOF

63.032

29.131

23.323

12.036

51.315

1.000

5.80

1:00

200

15.026

2

67.217

3º

75.278

216

80'

23.765

51.322

73.16

5

38.161

3 131

01.880

6

1.71

68.971

25.50

2º

A1.158

19

926

3.

S.S.S.

201

20.

19.198

0,998

S

76.198

32.029

38.73

20.282

12.101

91:501

7.321

0.088

2.291.91

21.661

19.36

293

.0.

30

34.609

28.125

\$

21

11.122

17.653

250

10

00000

280

.c.

055

193-68

500

1.12

88

269.02 050

61219

3

81-257

Nº. 1ª 0.5

249.21

50.77

200

141.28

912

.00

5.05

118.05

886.84

162.51

200

510-22

11.09

30.

506.52

89.09

21613

699.991

891.14

.2

00.52

11000

VIVI

10.91

3

698

NOS-24

161.00

026-16

Ves.24

V62.10

0.67

S.

250.16

2051

529.25

809.15

5. 25. C

E.E.S.

0

05.12

N

412.09

4

80.00

54.51 Ses

194

80.

695.95

00.91

95.04

av

82.94

.20

291

61.

2:54

16.565

92.908

85.08

10.102

10.34

15:59

10° 20°

51.919

63.81

69.593

73.601

81.201

23.221

098.16

012

19.973

83.155

7.10

30.932

96.924

0.109

NA.TA

72.098

200

13.65

3

61.859

35.013

28.001

12:518

23.

100

36.551

38.161

19.61

11.11

21.381

91.674

989

53.53

Po-

32.121

59.821

71.173

13.991

9.033

10.110

51.322

~

4

V v

6×1.6×

8

191.95

105

25569

VY. 050 10.02 02 250 692.24

158.15

60.65

121.09

es.

.2

oly

619

86.216

91.886

100.00

11V

VS

8.

81-257

192.68

444

0.

126.96

250.05

.Su

602

42.

2200105

69.62

.08

030.059

105.04

10.001

53.026

23.765

54.322

13.10

10.38

10:00

150.255

S.

5.151

gh-

12:55

2.

88.55

6.

84.191

06.019

95.052

27.328

38.03

30

58.21

0.0

gh.

30.

314

21.90

2.

60.

69.932

22.

0.012

2.

31.912

51.981

N.

11

2.

51.734

31.

01.391

21.014

37.196

011

30.534

0

2.02.

10.

63.740

20

050.

60.103

8

29:338

Nº.

85.

923

23.

de

So.

50.

1.

Sh 11 VOS. 24

141-28

012

120000

161

32.

75.278

52.625

V.502

94.933

76.51

1.11

68.971

do.

39.889

11.158

500

12.926

13:291

18.988

1.003

21.599

60.00

118.05

9999

18:51

075021

8.

508-52

550.55

6. S. J. J.

.08

912

5

252.28

15

055-09

5 SVS

1059.2L

462.46

000

Se.

5

00000

516.11

89117

698

005.52

10

026.16

235-15 .cz

S.

009 95.94

.4

361

194

.80

692 0.

3

250.16

.00

12 V

205 629-25

1.

10.91

101

960

.00

199.95

·22.

OLY

.ch

820

.150

099.16

012

230045

c91

.08

721

5.

1.283

51.014

15.991

50.0018

51-330

29.913

72.691

91.255

82.991

12.003

35.31

23.164

13.003

29.608

55 × 100

22.232

23.264

13.088

510

91.821

06.629

0.

11.921

61-555

228

0:

533

83.519

33

cn.

13.051

1.

70.814

84.57

SS

20

66.986

155

51.

66.281

293

1.

16.911

63.129

92.119

31.159

31.12

03.68

63-515

51.981

21-221

51.651

5538

v.

00.136

82.623

81.

65.576

81.15

718

76.513

86.11

89.389

155.982

67.301

14.661

16.016

81.375

16.021

52.098

13.278

03.100

17.018

39.695

08.913

68.683

10.532

23.662

10.139

05-235

85.005

26.354

23.972

12.751

17.121

2.

17.609

69.906

36.01

92.292

8.050

39.774

13.991

14.595

09.683

72.732

82.101

10.106

76.03

8.202

59.12

602 126.96

.0.14

101

NIN VIE

0.

124.52

102

250.05

3.

02000

100

014

3.

050

81.59

2.

109.52

265

.2

619

8 800 595.91

.0:

178-09

3.

64619

162

100 864 861.64

N.

224

131

8.

16.312

6.129

67.672

81.91

55.194

97.891

81.382

31.251

28.19

960

91.883

62.31

d'

92.902

10.086

39.76

7.579

92.23

14.39

73.971

38.12

10

202

77.916

80.11

61.615

A1.754

153-158

75.293

22.662

05.151

66.79

8.519

2.

15.014

N1-809

01.550

18.003

10.316

17.052

1.50

11.19

29.03

92.111

28.00

3

000

g.

83.104

1.1.00

19:518

20.39

15.11

20.13

32.11

28.99

10.000

50.1.48

10.501

78.152

86.566

750

N-958

1

103-032

29.131

.10

200

.2

ELS.

83.150

19:53

73.601

81:201

7.10

30.932

96.921

23.121

M.A.

5. 2. 20

2

0.

0

19.

63.81

201

61.

2.50

10.555

92.908

85.08

10.102

50.00

75.599

122

N

32.029

38.713

20.282

12.101

91:501

7.321

0.088

2.00.101

2.22.062

51.10

71.355

31.121

81.617

18.974

3.

00

21.664

16.293

13.013

22.211

65.631

15.50

1

16.989

19,136

34.609

60 000

82.3

Nº.

3

3

198

0.

23.323

72.036

51-345

14.000

0.00

1.605

15.026

67.217

16.198

17.653

51.919

15109

200

50

.du

11

000

115.64 E

S.

3.

10.00

11918

.0

12.20

Con Con

690 94

20

50.52

43

6.17

3

124.16

355

20.52

So.

50

6.

109.12

880.0

6.

80.98

805-25

120

086.09

00

Ъ

0

6

271

000

5

>

So.

200

4

2.51

200

.00

3

N.

Succession of the second

es.

250

6010

 $\odot$ 

a

3.

20.00

650

02.00

S

.130

Ser.

N.

VIV.VV

20

V26-95

05.73

5000

1.07

691-98

0

120

.2

96.9Y

50

2.00

.00

609.75

.2

200

S.

102

255.05

·co.

.ch

Ner.

204

2

100

I STATE

·2.

515

178-09

64010

152

OS-Y

695 giv

ee

.07

0.00

No

2.

100

020

050 100

11:18

2.

25

20

55

.V

3

296-22

.00

2

.10

12.13 Es la

.2

10,10

101

50.01

ELS.

200

.10

09:98

Si

.5

609

2

285.02

.2

209.11

264

SHE

·6.

10.92

~

0000

6699.71

80.00

629-25

854

Sho

87.00

068.01

68-86

IN IN

050.21

.02

964

3

0.

S.

519

23.25

6.

15V

50

10.01

.2

299.11

.2

01:02

500.84

055-16

678.17

519.58

10-10-1

860 128-16

5-5-V9

125-14

0.

.2

SE

880.51

192.52

250

¥1651

285-18

168-15

150

51402

164 2

12:11

195.92

20

98. YZ

C00.58

csu 6.

609.65

.0

16.989

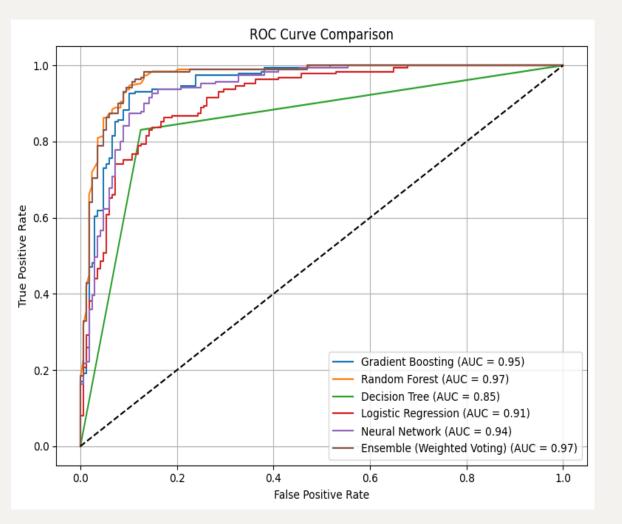


FIG. 2 Compares the performance of different
machine learning models using Receiver
Operating Characteristic (ROC) curves.

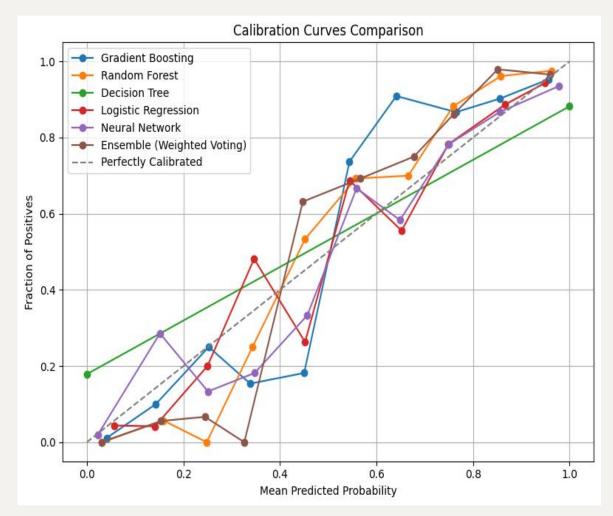


FIG. 3 Assesses how well each model's predicted
probabilities match real-world outcomes, which
is crucial for clinical decision-making.

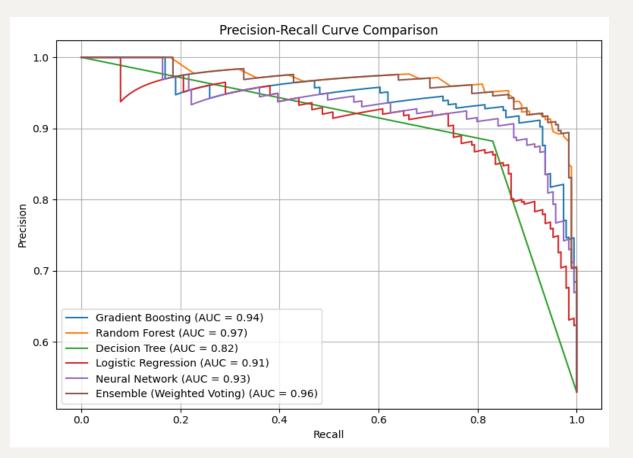


FIG. 4 Evaluates model performance using
precision-recall curves, which is particularly
informative for imbalanced medical datasets.

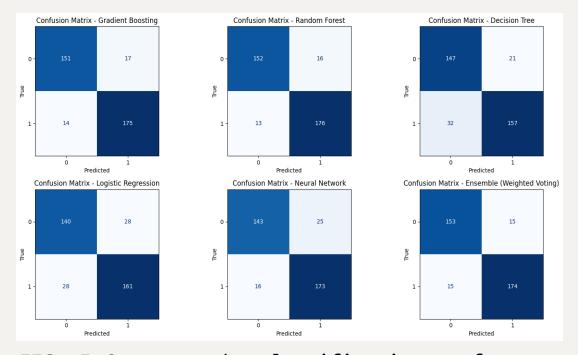


FIG. 5 Compares the classification performance of five different machine learning models through their confusion matrices, showing how each model distinguishes between patients with and without heart disease.

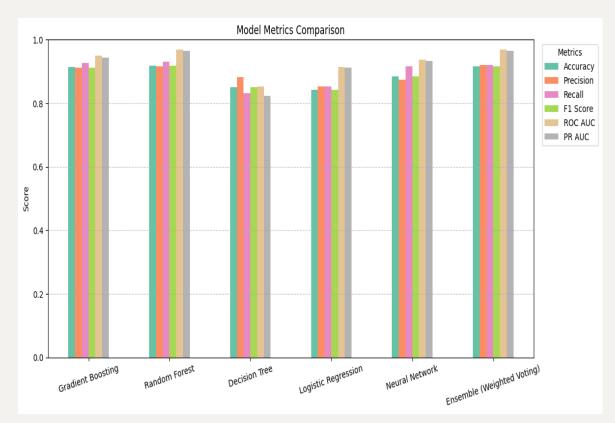


FIG. 6 Presents a side-by-side evaluation of six
 machine learning models across six key
 performance metrics for heart disease detection

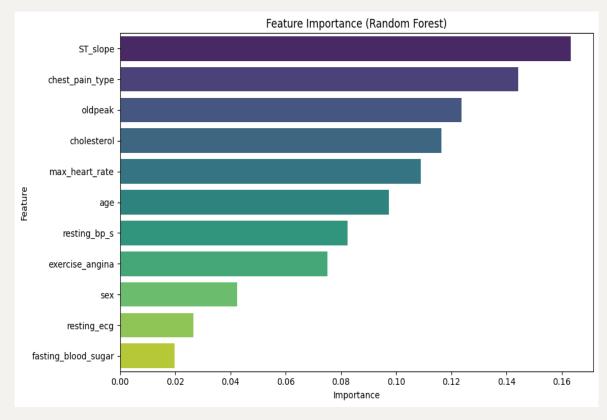


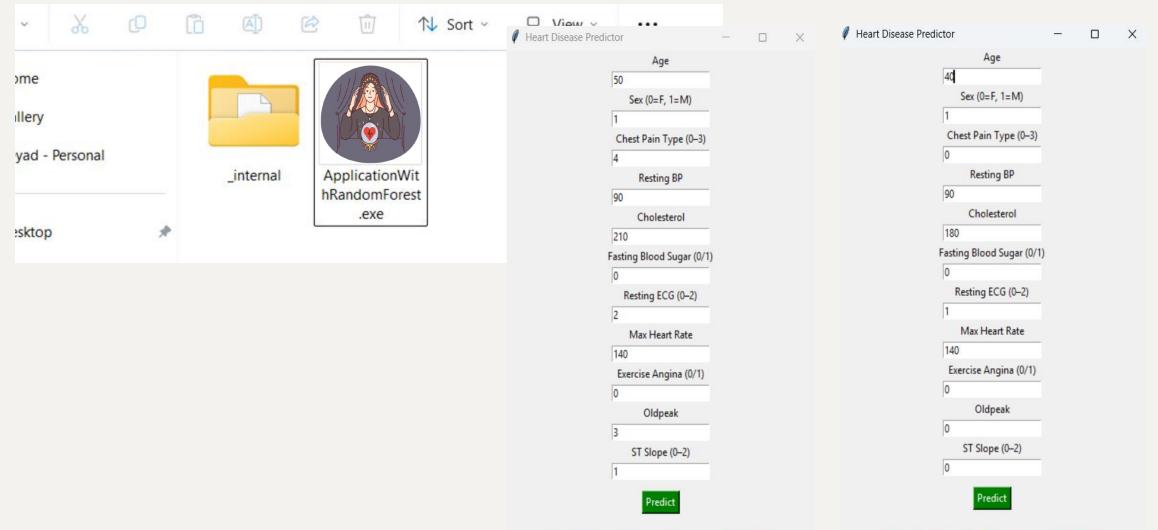
FIG. 7 Reveals the most influential medical factors for predicting heart disease, as determined by a Random Forest algorithm.

### \* Summary Results For Model Performance

	Accuracy	Precision	Recall	F1 Score	ROC AUC	PR AUC
Gradient Boosting	0.913165	0.911458	0.925926	0.912771	0.950145	0.944635
Random Forest	0.918768	0.916667	0.931217	0.918399	0.969309	0.965629
Decision Tree	0.851541	0.882022	0.830688	0.851424	0.852844	0.822321
Logistic Regression	0.843137	0.851852	0.851852	0.842593	0.913045	0.911826
Neural Network	0.885154	0.873737	0.915344	0.884337	0.938146	0.932695
Ensemble (Weighted Voting)	0.915966	0.920635	0.920635	0.915675	0.968191	0.964948

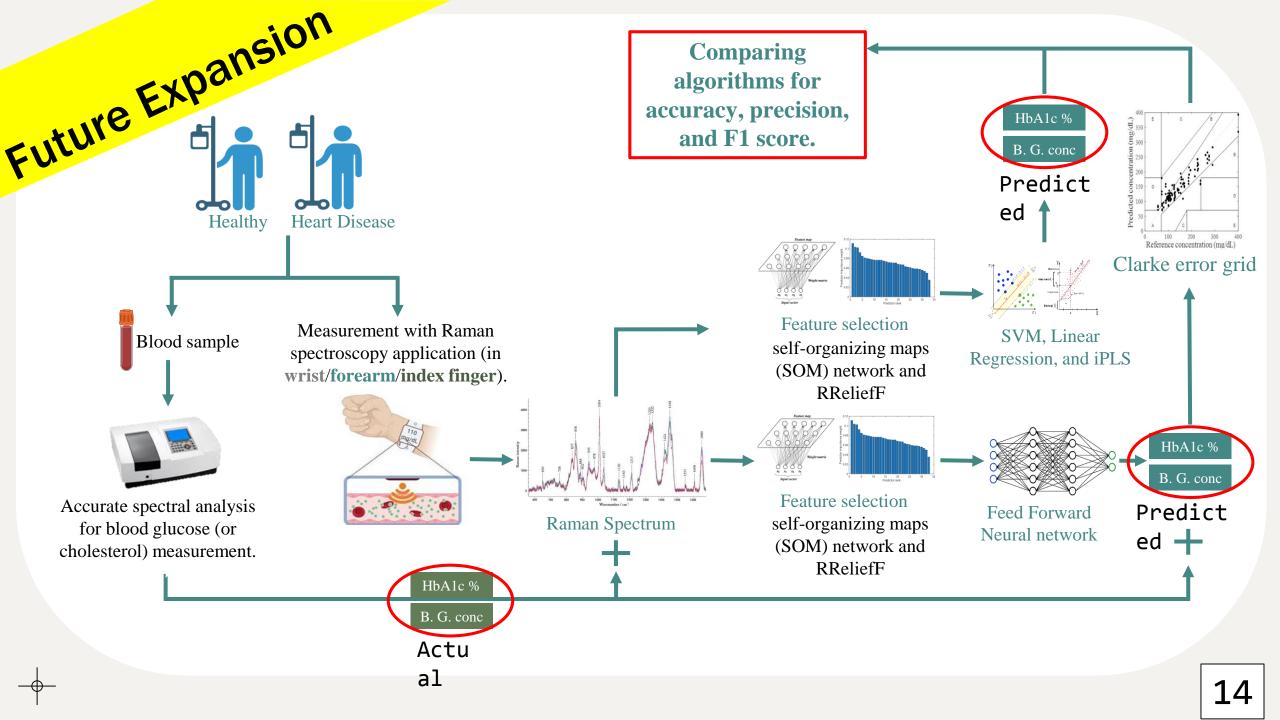
Comprehensive quantitative comparison of six machine learning models across seven key evaluation metrics.

### **Working Desktop Application Draft**



Prediction: Heart Disease (58.0%)

Prediction: No Disease (30.0%)



### **5.Other Real-World Projects**

ID

2 45654403

44731507 3

45774419 Λ

A         B         C         D         E         F         G         H         L         J         K           ID         Price         Levy         Manufacti, Model         Prod. year Category         Leather int Fuel type         Engine vol: Mileage         Cylir           45654403         13328         1399         LEXUS         RX 450         2010         Jeep         Yes         Hybrid         3.5         186005 km           44731507         16621         1018         CHEVROLE Equinox         2001         Jeep         No         Petrol         3.3         20000 km           45774419         8467         HONDA         FIT         2006         Hatchback No         Petrol         1.3         20000 km           45769185         3607         862         FORD         Escape         2014         Hechback Yes         Petrol         2.3         160931 km           45809263         11726         446         HONDA         FIT         2016         Hechback Yes         Petrol         2.4         2160931 km           45805768         1803         761         TOYOTA         Prius         2013         Sedan         Yes         Hybrid         2.5         398069 km <t< th=""><th>4 Variator       Profixe       04-May Left wheel White       4         4 Automatic Front       04-May Left wheel White       4         4 Automatic Front       04-May Left wheel White       4         4 Automatic Front       04-May Left wheel White       12         4 Automatic Front       04-May Left wheel White       12         4 Automatic Front       04-May Left wheel Grey       12         4 Automatic Front       04-May Left wheel Silver       12         4 Automatic Front       04-May Left wheel Silver       12         6 Automatic Av4       04-May Left wheel Silver       12         6 Automatic Rear       02-Mar Left wheel White       12         8 Manual Rear       02-Mar Left wheel Blue       0         8 Manual Front       04-May Left wheel White       12         6 Automatic Front       04-May Left wheel White       12         6 Automatic Front       04-May Left wheel White       12         6 Automatic Front       04-May Left wheel White       12</th><th>Used-cars prices prediction</th></t<>	4 Variator       Profixe       04-May Left wheel White       4         4 Automatic Front       04-May Left wheel White       4         4 Automatic Front       04-May Left wheel White       4         4 Automatic Front       04-May Left wheel White       12         4 Automatic Front       04-May Left wheel White       12         4 Automatic Front       04-May Left wheel Grey       12         4 Automatic Front       04-May Left wheel Silver       12         4 Automatic Front       04-May Left wheel Silver       12         6 Automatic Av4       04-May Left wheel Silver       12         6 Automatic Rear       02-Mar Left wheel White       12         8 Manual Rear       02-Mar Left wheel Blue       0         8 Manual Front       04-May Left wheel White       12         6 Automatic Front       04-May Left wheel White       12         6 Automatic Front       04-May Left wheel White       12         6 Automatic Front       04-May Left wheel White       12	Used-cars prices prediction
12 45621750 FORD Transit 1997 Goods wa(No Cree 3.5 138038 km	4 Automatic Front of May Left wheel White 12	
14 45815568 3000 - 10F5 LEXUS RX 450 2015 Leon Yes Diesel 2 74146 km	4 Automatic From an May Left wheel Blue	M N O P wc rc htn
15 45661288 1019 891 HYUNDAI Santa FE 2010 leep Yes Petrol 1.8 121840 km	an 4 Automatic Front 04-May Left Wheel Wheel	J K sc sod pot nento 44 7800 5.2 yes
16 45/32004 549 1079 TOYOTA China 2016 Sedan Yes Potrol 1.8 54317 km	D E F	ba bg 36 1.2 11 3 38 6000
AE772281 7085 Was Petrol	A D al su	notpresent 18 0.8 9.6 31 7.00 3.9 yes
19 45797221 26362 age HyUNDAI Sonata 21 Apr Yes HyDrid 1.6 112645	48 80 4 0 1011	425 a.o. 111 2.5 ar 7300 4.0 m
20 45772104 515 1850 LEXUS RX 400 2012 Sodan Yes Petrol 18 187000		notpresen 117 ac 14 ap 39 7800 4.4 yes
21 45653306 941 531 HYUNDAI Elantra 2012 Sedan Yes Plug-in Hyt 2.5 397091	a final abiormal abiormal abiormal	notoresen notpresen 106 20 11 142 3.2 12.2 no
22 45801686 1862 586 TOYOTA Prius 2017 Sedan Yes Petrol 230000	3 48 70 reat 2 0 normal normal	notoresen notpresen 74 24 104 4 6900 5 10
23 45168471 20698 1249 TOYOTA Camry 2017 Microbus No Diesel 2.0 test	6 4 51 00 1 015 3 0	notpresen notpresen 100 31 1.1 10.8 33 9600 4 70
24 45751951 17502 FORD Transit 1005 Sedan No Petrol 2.2 Turbo 210000		notpresen notpresen 410 1.9 9.5 29 12100 3.7 900
25 45814106 OPEL Vectra and Sedan Yes Dicset 4.7.0 km	8 6 00 1015 2 4 ioinit short	notpresent notpresent 70 107 7.2 114 9,4 28 28 28
26 45813492 MERCEDE E 220 2008 Jeep Yes Pertor	9 7 24 100 1015 3 0 Holmer	al present notpresent 100 55 4 10.8 32 4500
27 45/82835 - 2455 LEXUS 0X470	10 0 10 102 2 abnorm	al present 1000 380 60 2.1 120 58 9.7 2.0 ves
28 45759112 549 2455 4	11 9 55 50 1.01 2 4 a shormal abnorm	al present notpresent 2009 72 2.1 130 2.4 9.8
	12 10 63 70 1.01 5 1 normal	oresetti hotp
	13 11 68 70 1.015 J	notpresent 157 90 4.2 141 4.9 7.6 24 no
	14 12 abnor	nal present protoresen 76 162 3.2 138 4.1 12.6 yes
	15 13 68 80 1.01 3 0 norma	actoresen notpresen 35 52 139 5.7 5.7 11400 4.3 yes
- •	16 1 40 80 1.015 2 0 norm	actoresen notpresen 114 07 13 135 4.5 10 2 30 5300 5.7 year
nicnase	18 16 47 70 1.015	1 potpresen notpresen 203 21 1.6 7.7 24 9200 5.2 900
I have 117 SEase	10 17 47 80 norm	resent notpresent 100 148 3.9 135 5.2 10.9 32 6200 5.0 yes
Vidnev ULSe	20 18 60 100 1.025 0 abno	180 76 4.3 0.0 9.8 32 0900
K LUIIC)		
Kidney Disease	22 20 61 00	notpresen notpresen 95 111 39 8300 40 Fe
-dict1011	23 21 60 co 1025 4 0 Hormstrong	mal notpresent notpresent 50 1.4 125 5.2 9.9 29 8400 4 years
prediction	24 22 46 1.01 0 0 normal abr	normal notpresent resent 75 1.9 141 at 116 35 10500
	25 23 21 100 1.015 4 0 normal no	mal notpresen notpresent and 45 2.4 140 to 13.5
Γ	26 24 1025	mal notpresen notpresen 156 45 17 170 a
	75 00 A MAR	example entropy
<b>ħ</b>	28 26 70 101	
Ϋ́		

15

V W

appet

good

good

poor

poor

good

good

good

good

good

poor

good

poor

poor

poor

poor

good

good

poor

good

good

poor

good

good

poor

poor

good

poor

acad

U

dm

yes

no

yes

no

no

yes

no

yes

yes

yes

yes

yes

yes

yes

yes

no

no

no

yes

no

yes

yes

no

no

no

yes

yes

1100

cad

no

yes

yes

yes

no

no

no

yes

yes

yes

yes

no

no

no

no

no

4 yes

4.1 1000



# Thank you.

For more information, please contact: <u>zeyad.mansour@aucegypt.edu</u> <u>alexbe323@gmail.com</u> <u>Rabab.asar17@fsc.bu.edu.eg</u> <u>Enjy\_sayed@aun.edu.eg</u> <u>Esraa Hussien@aun.edu.eg</u>