

# Classification of Alzheimer's by DNA Analysis.

By: Ahmed Samir - Fairuz Soufy - Omar Ehab - Sara El-Bedeawi. Supervised by: Dr. Ashraf Abdelraouf - Eng. Lobna Shaheen In collaboration with Faculty of Pharmacy : Prof. Lamiaa Nabil – Dr. Omar El-Demrdash - Dr. Rawan El-Kady - Dr. Nora El-Smnoudy

# Introduction

### What is Alzheimer's disease?

Alzheimer's is a progressive brain disease in which abnormal protein deposition building up plaques in nerves endings in the brain, causing brain cells to die.

Early diagnosis of **Alzheimer's** may help in slowing down the progression of the disease.<sup>[1]</sup>



#### The analysis of the DNA helps to detect AD.

[1] Molinuevo, José Luis et al. "The Rationale Behind the New Alzheimer's Disease Conceptualization: Lessons Learned During the Last Decades." Journal of Alzheimer's disease : JAD vol. 62,3 (2018): 1067-1077. doi:10.3233/JAD-170698

# **Problem Statement**

The existing classifications of AD is either the patient is healthy, or the patient is in severe case of Alzheimer.

We classify the patient into healthy patient and to a patient that carries AD disease.

But we classify the healthy patients into two stages: *completely healthy patient or a patient that have a risk to have AD.* 

And we are also clustering the diseased patients into three clusters: *Mild, moderate and severe.* 

# Database Diagram



# Sequence Diagram





Researcher

Admin

# **Architecture Diagram**

Login view   Register view   Admin View   Researcher View   Result View     History View   Patient info View   Upload sample   Manipulate patients   Manipulate Users view	Model:
Admin controller     Admin controller     Upload dataset()     TestModel()     Login Controller     Register Controller     Add Sample Controller     Manipulate Users controller     Patient Result Controller	User Admin Researcher Patient SNPS Preprocessing UserType Sample Status Classifier T Fire store Database

# **Class Diagram**



# System Approach

# Activity Diagram (1/2)



# Preprocessing (2/2)

- Filter
- Convert Non Numerical
- Transpose
- Cross Reference
- Labeling and Merging

4	A	В	С	D	orma	F		G	н		1		J	к		L	M	Dis	ea	se	d	Р		Q	R	1	s
741	141_S_1	05 Diseased	1	1	0		0	1		0		0	1		1	2		2		0	0		1	1	2		2
7/12	141 4	B C	DE	F G	H I	J	К	M	N	0	P	015	ample In Sample ID	GC Score SN	IP Inde	SNP Name Allele	1 - T Allele2	- TAllele1	- FAllele2 - F	Allele1 -	Allele2 - # Chr	Position 0	T Score	luster Se SNP	Theta R	x	Y
1.12	254 #CHF	OM POS ID R	REF ALT	QUAL FILTER	INFO FORMAT	T NA06984 NA	06985 NA06	5986 NA069	89 NA0699	4 NA070	00 NA0703	37 1 2	342 002_S_055	0.9299	1	200003 A	A	A	A	A	A	9 1.39E+08	0.8931	1 [T/C]	0.022	1.055	1.019
743	141255	1 2.27E+08 rs1859273 T	г с	100 PASS	AA=T   ;A GT	0 0 0	0 0	0 0	0 0	00	0 0	10 a	342 002 S 055	0.7877	2	200006 G	G	с	с	8	в	9 1.39E+08	0.7824	0.9871 [A/G]	0.989	1.574	0.026
	256	1 2.27E+08 rs5419751 A	A G	100 PASS	AA=A   ;AGT	0 0 0	0 0	0 0	0 0	00	0 0	4	342 002_S_055	0.8612	3	200047 A	A	A	A	A	A	2 2.2E+08	0.8318	0.514 [T/C]	0.056	0.464	0.427
744	141257	1 2.27E+08 rs1900429 A	A G	100 PASS	AA=A   ;AGT	0 0 0	0 0	0 0	0 0	0 0	0 0	0 5	342 002_S_055	0.8331	4	200050 G	G	C	С	в	8	2 2.2E+08	0.8117	0.717 [C/G]	0.996	1.191	0.008
	258	1 2.27E+08 rs7308570 G	G A	100 PASS	AA=G   ;AGT	0 0 0	0 0	0 0	0 0	0 0	0 0	6	342 002_5_055	0.9466	5	200052 T	т	T	T	8	в	2 2.2E+08	0.9126	0.9218 [T/A]	0.988	0.729	0.014
/45	141259	1 2.27E+08 rs5520006 C	СТ	100 PASS	AA=C   ;AGT	0 0 0	0 0 0	0 0	00	00	00	0 7	342 002_S_055	0.7494	6	200053 A	A	T	T	A	A	2 2.2E+08	0.7597	0.7986 [T/C]	0.077	1.045	0.932
	260	1 2.27E+08 rs5639222 A	A C	100 PASS	AA=A  ;AGT	0 0 0	0 0 0	0 0	0 0	00	0 0	8	342 002 S 055	0.9705	7	200070 C	G	G	с	A	8	16 16174099	0.9464	0.828 [G/C]	0.524	0.484	0.233
/46	141261	1 2.27E+08 rs1812813 A	х т	100 PASS	AA=A   ;AGT	0 0 0	0 0 0	0 0	00	00	0 0	0 9	342 002 S 055	0.6531	8	200078 G	G	С	с	в	8	16 16194115	0.7073	0.6003 [C/G]	0.931	1.549	0.152
7.47	262	1 2.27E+08 rs1859366 G	S T	100 PASS	AA=G   ;AGT	0 0 0	0 0 0 0	0 0	0 0	00	0 0	10	342 002_5_055	0.9078	9	200087 A	C	T	G	A	8	16 16153665	0.8708	0.8477 [T/G]	0.593	1.107	0.472
/4/	141263	1 2.27E+08 rs1911814 A	A G	100 PASS	AA=A   ;AGT	010 01	0 0 0 0	010	0 0	00	0 0	10 11	342 002 5 055	0.8205	10	200091 A	C	A	c	A	8	16 16188655	0.8032	1 [A/C]	0.51	2.155	1.06
7.40	264	1 2.27E+08 rs1155403 A	A C	100 PASS	AA=A   ;AGT	0 0 0	0 0 0 0	0 0	010	010	0 0	0 12	342 002 S 055	0.9354	11	200096 A	C	A	с	A	8	16 16195546	0.8992	1 [A/C]	0.456	1.13	0.604
/40	141265	1 2.27E+08 rs1162978 T	C C	100 PASS	AA=T[[];AGT	010 01	0 0 0	010	010	010	010	13	342 002 S 055	0.8902	12	200124 G	G	G	G	8	8	7 1.51E+08	0.855	1 [T/C]	0.972	0.799	0.033
7.40	266	1 2.27E+08 rs1794035 T	C	100 PASS	AA=C[]];AGT	01 0	1 0 0	010	111	010	0 0	0 14	342 002 5 055	0.9599	13	200199 C	G	с	G	A	в	22 40847933	0.9302	0.9179 [G/C]	0.521	0.68	0.329
43	141267	1 2.2/E+08 rs1/94036 G	s A	100 PASS	AA=G[]];#GI	01 01	1 010	010	011	010	010	15	342 002 5 055	0.8783	14	200240 G	G	G	G	8	8	1 20850186	0.8451	0.8297 [C/G]	0.993	0.977	0.01
750	141200	1 2.27E+08 rs5592311 G	5 C	100 PASS	AA=A   ;AGI	010 01	0 010	010	010	010	010	016	342 002 S 055	0.941	15	200251 A	A	т	т	A	A	1 20850335	0.9057	0.8496 [A/C]	0.023	0.95	0.917
100	141269	1 2.27E+08 rs1826027 T		100 PASS	AA=TTTTAGT	010 01	0 010	010	010	010	010	17	342 002 S 055	0.7452	16	200258 A	A	т	т	A	A	1 20869344	0.7573	0.9196 [A/C]	0.043	1.368	1.282
751	9/11271	1 2.27E+08 rs5350319 G		100 PASS	AA=G[[];FGT	010 01	0 010	010	010	010	010	() 18	342 002 S 055	0.7948	17	200260 G	G	c	c	в	8	1 20865406	0.7868	0.5368 [G/C]	1	1.325	0
	271	1 2.27E+08 r55535909 G	5 A	100 PASS	AA=G[[];#GT	010 01	0 010	010	010	010	010	19	342 002 5 055	0.8794	15	200261 A	A	T	т	A	A	1 20841103	0.846	1 [A/G]	0.076	0.991	0.885
752	941272	1 2.27E+08 rs5747507 G	3 I	100 PASS	AA=G[[];FGT	010 01	0 010	010	010	010	010	0 20	342 002 S 055	0.9257	19	200262 A	A	T	T	A	A	1 20866804	0.8885	1 [T/C]	0.017	0.604	0.588
	274	1 2.275+08 153417756 0		100 PASS	AA-ALLIACT	010 01	0 0 0	010	010	010	010	21	342 002 5 055	0.8717	20	200519 A	G	A	G	A	8	6 33274012	0.8399	1 [T/C]	0.507	1.889	0.934
753	941276	1 2.27E+08 IS0118312 A		100 PASS	AA-AIII)POI	010 01	0 010	010	010	010	010	0 22	342 002 5 055	0.9517	21	200521 G	G	G	G	8	8	6 33281976	0.9191	1 [T/C]	0.959	0.459	0.028
	275	1 2.275+08 153735078 C	- T	100 PASS	AA-CIIIAGT	010 01	0 0 0	010	010	010	010	23	342 002 5 055	0.9631	22	200623 G	G	G	G	B	8	4 20150714	0.9349	1 [T/C]	0.988	0.566	0.011
754	941277	1 2.275+08 153034053 C	- I	100 PASS	AA-CIII:AGT	010 01	0 0 0	010	010	010	010	0 24	342 002 5 055	0.9212	23	400530 A	G	A	G	A	8	11 1.19E+08	0.8838	1 [A/G]	0.665	0.65	0.239
_		1 2.27E+08 rs5282547 C	T	100 PASS	AArc111:A GT	010 01	0 010	010	010	010	010	- 25	342 002 5 055	0.951	24	400826 A	G	T	c	A	B	12 20905215	0.9181	0.7394 [T/C]	0.549	0.641	0.296
755	941_5_1	ZUDIseased		U	U	010	U	Ŭ	010	Ű	010	0	~		<b>M</b>	-		-	-	-	· ·		•	¥1	-		-
756	941_5_1	29 Diseased	0	0	0		1	1		0		0	2		0	2		2		0	0		0	0	2		2
757	041 6 1	21 Diseased	0	1	0		1	1		0		0	1		1	3		2		0	1		1	0	2		2
151	941_5_1	51 Diseased	0	1	U		1	-		•			1		1	2		4		•	1		1	0	4		4
758	941_5_1	36 Diseased	0	0	0		1	1		0		0	2		1	2		2		0	1		٥	0	2		2
759	NA06984	4 Normal	0	1	1		0	1		0		0	1		2	0		1		0	0		0	0	2		2
760	NA05985	5 Normal	0	1	0		1	0		0		0	1		2	1		1		0	0		1	0	2		2

# Support Machine Regression

# SVR

Where the red lines represent the boundaries and the red line represent the hyperplane.



## **Clustering Results**

We clustered our data into 3 groups:

- 1. Group A : Normal.
- 2. Group B : Diseased.
- **3.** Group C : Potential.

## **Clustering Results (1/3)**

### Normal Cluster Results:

Normal	Diseased	Total
302	1	303
99.7%	0.3%	100%

## **Clustering Results (2/3)**

### **Diseased Cluster Results:**

Normal	Diseased	Total
27	365	392
7%	93%	100%

## **Clustering Results (3/3)**

### **Potential Cluster Results:**

Normal	Diseased	Total
82	89	171
48%	52%	100%



😰 venv [E:\Users\user\PycharmProject	ts\untitled1\venv]\home.py [ver	nv] - PyCharm							$\times$
<u>F</u> ile <u>E</u> dit <u>V</u> iew <u>N</u> avigate <u>C</u> ode	<u>R</u> efactor Run <u>T</u> ools VC <u>S</u> <u>W</u>	indow <u>H</u> elp							
🖿 venv 👌 💑 home.py							nome (1) 🔻 (	â 🛎 🔳	Q
ម្ 🔲 Projez 😯 🛨 🗢 —	樻 process.py 🛛 👩 home.py	🗶 🛃 SAmir.py 🗵							
venv E:\Users\user\PycharmProje	Q-	(↓ ↑ [↓	9 + K T Match Case W	ords 🔲 Rege <u>x</u> ?					×
Hinclude		tost size - pp mat	th coil(lop(df) * 0.2)						
Lib		cest_size - np.ma	$\operatorname{cn.cell(len(dl) ^ 0.5)}$						Ē
tcl									
tcl8.5		X_train = X[: Ø Alzhe	eimer's Analysis System —						
🕨 🖿 tk8.5									=
🖆 file-1.txt		y_train = y[:	Alzheimer's Analysis System						
🐔 SAmir.py			Alzheimer's Analysis System						
SVR 93 joblib		X  test = X[-t]							
alzheimer-4a38b-firebase-adn	669								
🖧 csvtesting.py		toot[ +							
🚔 data_path.pickle		y_test - yl-t							
design.qss		· · · · · · · · · · · · · · · · · · ·	Logi	in					
home pv		clf.fit(X_tra	First Tir	me?					
Login.py		accuracy=int(	Regist	ter					
🗑 pip-selfcheck.json	674	self.train_re							
🖕 process.py			$\mathbf{X}$						
Register.py									
SVR 93.ioblib	677 🕨 🗄 if na	me == " mai							
🛃 TEST.py									
🐈 venv.rar									
External Libraries									
문 Scratches and Consoles									
7									
				Pecerding Teelbar			× -		
orites				Statistics	Duration Audio		*		
2: Fav				Mode · Video	Duration Audio				
*	Ui MainWindow	v ≥ train0		Frames : 0 (0 fps)	00.00.00	0 0			
▶, <u>4</u> : Run 🗎 6: TODO 🛂 Te	erminal de Python Console			Dropped : 0	00.00.00			1 Event Lo	g
DE and Plugin Updates: PyCharm	n is ready to update. (3/4/2020 4:38	3 PM)		biopped. U			4	spaces 🗧 🔒	曼
⊕ Type here to searce □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □	ch	o 🗄 室 🔒 I	🖬 N 🛇 メ 💽	<u> </u>	89%	🜗 🛡 \land 🍥 📼 🕼	ロッション ENG 12 3,	2:09 AM /7/2020	2