# Classification of Alzheimer's by DNA analysis.

Ahmed Samir – Fairuz Soufy – Omar Ehab – Sara Elbedeawi.

Supervised by: Dr Ashraf Abdelraouf – Eng Lobna Shaheen.

In collaboration with Faculty of Pharmacy: Prof. Lamiaa Nabil - Dr. Omar El-Demrdash – Dr. Rawan Elkady – Dr.Noura Elsamanody – Dr. Mohamed Shahin.

### INTRODUCTION I

Alzheimer's disease is a permanent brain disorder that slowly destroys memory, thinking skills and finally the ability to carry out the simplest tasks.

The existing classification of Alzheimer's is either the patient is healthy or the patient is in the late stage of Alzheimer's.

Alzheimer's can be detected by screening the DNA. There are two ways to screen the DNA either using whole genome sequence (WGS) or by using the Single Nucleotide Polymorphism(SNPs).



#### July 2020 🔘

[1] Finckh, Ulrich, et al. "High prevalence of pathogenic mutations in patients with early-onset dementia detected by sequence analyses of four different genes." The American Journal of Human Genetics 66.1 (2000): 110-117.

Udiddalluu

16180861 catageaggt ttgeaacaaa gggggtattt taaagetttt ttetaacate geettageag

16180921 atactgactg atcetttecc aggaactgag agggteceac ceatgeteet gattttagaa

16180981 tcataaacat cagtaggacc aagatgcttt tagagggata aagtgacttt aatttatctt

16181041 gggtggtcaa aacacagcat gtgcaatatt tttacaaact atgattcaga aagcacatag

16181101 ttctagagcg ttttgttttt tgttatgttg gtgtcactgg aaacagttgt getteetgat

16181161 gtettactga gatggeettt ttggteattt tgggeeaett teecaaggat actgcatgtt

16181221 geetgaeetg tatteetgag gaaataagga ggaettagat tgatetgttt tggacagcgt

16181281 acttteteta agaatggega aaggaaaaac ateagaetag gatgetetee agaaataata

16181341 actgagaate teatectgtt gtgtaaacte teettgtaet ttgetetgtg tgtgtttggt

16181401 ggaaatatta tttttaaaca gccatttact ctgaaaaaat atgattttgc attaaattag

16181461 ttttaaagtt aagggatata tataagacct ttttattaac aagttgatga gatgagcatt

16181521 agtataccta taggaagttt ttgaatgagt gccaattctg aatttacaaa tcaacatttc

#### WGS vs SNPs

#### WGS

#### **SNPs**

USLABALLUU

16180861 catageaggt ttgcaacaaa gggggtattt taaagetttt ttetaacate geettageag

16180921 atactgactg atcctttccc aggaactgag agggtcccac ccatgctcct gattttagaa

16180981 tcataaacat cagtaggacc aagatgcttt tagagggata aagtgacttt aatttatctt

16181041 gggtggtcaa aacacagcat gtgcaatatt tttacaaact atgattcaga aagcacatag

16181101 ttctagagcg ttttgttttt tgttatgttg gtgtcactgg aaacagttgt gcttcctgat

16181161 gtettaetga gatggeettt ttggteattt tgggeeaett teeeaaggat actgeatgtt

16181221 geetgaeetg tatteetgag gaaataagga ggaettagat tgatetgttt tggaeagegt

16181281 acttteteta agaatggega aaggaaaaac atcagactag gatgetetee agaaataata

16181341 actgagaate teateetgtt gtgtaaaete teettgtaet ttgetetgtg tgtgtttggt

16181401 ggaaatatta tttttaaaca gccatttact ctgaaaaaat atgattttgc attaaattag

16181461 ttttaaagtt aagggatata tataagacct ttttattaac aagttgatga gatgagcatt

16181521 agtataceta taggaagttt ttgaatgagt gecaattetg aatttacaaa teaacattte

Unianedis	aniple Inc Saniple ID	GC Score	SNP index SNP Name Allele1	-TrAlelez	1-TrAlelet	-RAlele2	-FiAllele1	- N Allele2 - A Chr	Position	GT Score	Chaster Sej SNP	Theta R	2 J			(Raw N	Raw
44191	756 018 5 00	0.5566	44192 m1048069 G	6	5	-	8	8	21 26441982	0.6931	0.7799 jT/Cj	0.979	2.253	0.071	2.182	116	723
95597	756 018 5 004	0,9388	F 95508 == 1150895 A	Α.	T	T	A	A	1 2.256+08	0.9031	1 [A/6]	0.075	0.948	0.941	0.008	811	1
50297B	756 018 \$ 004	0,4658	8 303979 m1166614.G	6	÷¢.	. C.	8	8	19 50105714	0.8676	6.5254 [A/G]	0.991	1.161	0.017	1.145	- 59	4578
205317	756 018_5_00/	0.9533	105118 m1170226 G	G	C	-C	8	8	21 26390611	0.9211	t (A/G)	0.983	1.095	0.028	1,027	57	3408
125365	756 018_5_00	0.7211	125370 es1212692-G	G	G	6	8	B	1 2.25E+08	0.783	0.6402 [T/C]	0.978	1.193	0.04	1.153	74	350
135709	756 008 5 004	0.9078	5 135730 m1240646 A	6	A	10	A	8	1 2.25E+08	0.8705	1 [A/G]	0.622	1.627	0.655	0.972	126	3341
140283	756-018_5_004	0.9163	140384 m1248226.A	A	A	A	Α.	A	21 26399165	1.8809	1 [A/6]	0.018	0.905	0.88	0.025	790	111
340305	756 8t8_5_00/	D.BBS	140306 m1248275 A	A	A	A.	A	A	21 26435708	0.8539	0.8142 (4/5)	0.05	0.83	0,769	0.065	694	236
360844	756-018_5_00	0.8906	0 160845 rt1295640 A	A	Ŧ	T	A.'	A	1 2.25E+08	0.8553	1 [A/6]	0.012	1.516	1,489	0.027	1333	103
163806	356 018 5 004	0.9125	A ESBNDEE 103807 m1304923 A	A	Α.	A.	Α.	A	21 26443288	0.8756	0.7846 [T/C]	0.092	0.456	0.398	0.058	328	17
212072	756 018 \$ 004	0,9475	5 212078 m165933 G	6	c	C.	8		14 72735107	0.9137	1 [4/6]	0.975	1.021	0.038	0.983	65	3261
212073	756 018_S_00/	0.8361	1 212074 m165895 A	A	A	- A	A:	A	14 71756897	0.8069	1 (A/G)	0.035	1.345	1.274	0.07	1144	.240
214396	756 018_3_00/	0,6448	8 234397 rs3684664 C	C .	G	6	8	8	1 2.25E+08	0.7029	1 [T/G]	0.945	2.03	0.161	1.899	192	6421
223216	756 018 5 00	0.854)	223217 rs1700149 A	A	A.	A	Α.	A	21 26200118	0.8341	1 [A/C]	0.043	1.197	151.25	0.076	898	250
243036	756 018_5 004	0.91	243057 m1758861 A	A	T	T	A.	A	21 36439074	0.8757	0.7027 [T/C]	0.021	0.641	0.62	0.021	433	X
249954	756-018_5_004	0.9156	5 249655 m1783016 G	G.	C.	. C	8	8	21 26201909	0.8783	1 [A/6]	80.0	1.552	0.048	1.504	94	5743
249956	756 818_9_00/	0.8674	249957 m1783025 A	G	T	C.	A	8	21 26188634	0.8365	0.8095 (A/G)	0.697	1.02	0.347	0.672	291	2031
250626	756-018_5_00	0.9963	250637 m3787438 C	C	6	6	8	8	21 26195960	0.9001	1 [AVC]	1	0.943	0	0.943	33	3241
271091	756-018_3_004	0.9608	8. 271092 m2014146 A	6	A	6	Α.		21 26256890	0.9315	1 [A%]	6.573	1.334	0.59	0.744	551	2590
27157E	756 018 3 004	D.RDE	6 271577 m2019263 G	a.	6	6	8		21 26174474	0.8356	0.8627 [4/6]	0.99	1.215	0.019	1.195	58	3530
277975	756 018_5_004	0.9375	5 277976 m2070657 A	A	A		A	- A	21 26276888	0.9016	0.8874 [T/C]	0	0.505	0.505	0	462	
278565	756 018_5_00/	0.953	278566 /12073489 A	A	τ	T	Α.	A	1 2.25E+08	0.9194	1 [T/C]	0.02	0.362	0.357	0.011	327	- 40
284228	756-018_5_00+	0.9463	284229 m234260 A	G	Α.	-6	Α.	8	14 72732362	0.9144	1 (AS)	0.617	0.814	0.331	0.483	304	1005
284239	756 018 5 004	0.9504	284340 m214273 G	6	C	. C	8		14 72685877	0.9173	1 [A/6]	0.995	0.695	0.005	0.69	.13	2371
284438	756.818_5_00	0.9019	9 284439 m214488 A	A	T	T	A.	A	21 26209705	0.8653	1 [A/G]	0.043	0.952	0.892	0.05	806	212
285465	756.018_5_00/	0.9400	7 285466 m2116079 G	G	6	6	8	8	21 36470201	0.9053	6.9337 [4/6]	6.995	0.719	0,005	0.713	34	2445
286682	756 018_5_004	0.9594	A 5974651 15236762 A	6	Ť	C	A		21 26258039	0.9295	1 [1/C]	0.507	0.811	0.401	0.45	378	1412
385661	156 dtl 5 00/	0.6467	186603 -116770 4						21.36345525	0.0150	+ 14.91	0.022	1.104	1 02	0.035	021	+ 2

SNPs is easier to deal with, much faster to work on, and to analyze.

July 2020 🤇

#### **Tools Used**



## Dataset

It is a multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of Alzheimer's disease (AD).

The initial five-year study (ADNI-1) was extended by two years in 2009 by a Grand Opportunities grant (ADNI-GO), and in 2011 and 2016 by further competitive renewals of the ADNI-1 grant (ADNI-2, and ADNI-3, respectively).



### WGS System Overview



July 2020 C



#### **Context Diagram**



July 2020

A	В		ć II	D	E	F	G	н	L.I	1	K	. E	м	N	0	Ρ	Q	R	S	T	U	V	W	Х	¥	Z	AA	AB	AC	l
1 id	rs115	0895 rs207	73489 rs12	295640 rs	2802268 r	s1684664	rs1212692	rs1240546	6 rs231300	7 rs214273	rs8006497	rs362350	rs214260	rs165933	rs362377	rs165935	rs362393	rs362400	rs362346	rs8106922	rs405509	rs769451	rs439401	rs2019263	rs3787620	rs373521	rs2829973	rs178302	5 rs380417	1
2 002 5 025	15	0	1	2	1	0	0	1	1 3	1 2	0	0		2	2 0	1 3	0	0	0	1		2 0	2	2	0	7	0	1	2 7	1
3 002 5 04	3	0	1	2	1	0	1	1	1 1	2 2	0	0		2	2 0	i	0	0	0	1	1.1	2 1	1	2	0	1	1		2 1	L
4 002 5 05	9	1	2	0	0	0	1	(	0 ;	2 2	0	0		2	2 0	1	0	0	0	0		2 0	0	2	0	0	) 2		2 7	2
5 002 5 06	.9	1	2	0	0	1	0	(	0	1 2	0	0		2	2 2	6 3	2	0	2	0		2 0	2	2	0	0	) 1	. 3	1 1	ε
6 002_5_068	15	1	2	0	0	0	0	(	D i	2 2	0	0		2	2 2	i ia	2	0	2	0		2 0	0	2	0	0	2		2 7	2
7 002 5 07	19	0	1	1	0	1	1	(	0	1 2	0	0		2	2 0	é ()	0	0	0	1		2 0	1	2	0	0	1		1 1	£
8 002 5 075	12	1	1	1	0	0	0	(	0 3	1 2	0	0		2	2 0	1	0	0	0	1	- 24	2 0	2	2	0	1	1	4	2 7	1
9 002 5 083	6	1	2	1	0	0	0	1	1 1	1 2	0	0		2	2 0	i - 1	0	0	0	0		2 0	2	2	0	1	1		2 7	2
10 002 \$ 093	18	0	1	1	0	1	1	(	0 3	2 2	0	0		2	2 0	6	0	0	0	2		2 0	2	2	0	0	1		1 1	£
11 002_5_095	4	0	0	2	0	0	1	(	p :	2 2	0	0		2	2 0	i i	0	0	0	1	- 2	2 0	2	2	1	1	0	1	1 1	Ł
12 002 S 095	5	0	2	1	0	1	0	1	1 1	2 2	0	0		2	2 0	i (1	0	0	0	0	1	2 0	1	2	0	1	1	1 3	2 1	2
13 002_5_10	.8	0	2	1	1	0	0	(	0 :	2 2	0	0		2	2 0	1	0	0	0	1		2 0	1	2	0	1	1	. )	2 1	Ĺ
14 002 S_107	70	1	1	1	1	0	0	1	1 :	1 2	0	0		2	2 0	1	0	0	0	1	1	2 0	2	2	0	1	1	. (	2 3	2
15 002_5_113	5	2	2	0	0	0	0	(	0 :	1 2	0	0		2	2 0		0	0	0	1		2 0	1	2	0	- 4	1	- 3	2 7	1
16 002_S_126	il.	0	0	2	1	0	1	1	1 3	2 2	0	1		2	2 0	ê () <b>:</b>	. 0	0	0	1		2 0	1	2	1	6	1	()	1 7	Ł
17 002 5 12	18	0	0	2	2	0	0	1	1 :	1 2	0	0		2	2 0	6 3	0	0	0	0		2 0	1	2	1	1	. 0		1 1	L
18 002_5_12	10	0	1	2	1	0	0	(	0 :	1 2	0	1		2	2 0	1 1	0	0	0	0		2 0	1	2	0	0	2	1	2 7	1
19 005_S_022	11	1	1	1	0	0	1	(	0 :	1 2	0	0		2	2 1	1	1	0	1	0		2 0	2	2	1	0	) 1		1	1
20 005_5_022	2	0	0	2	1	0	1	1	1 :	2 2	0	0		2	2 0	1 1	0	0	0	0		2 0	0	2	0	1	1	3	2 7	1
21 005 5 02	13	1	1	1	1	0	0	- 2	2	2 2	0	0		2	2 0	1	. 0	0	0	1	1	2 0	2	2	0	1	1		2 7	1
22 005_5_03	14	2	2	0	0	0	0		2 :	2 2	0	1	-	2	2 0	1 1	0	0	0	0	4	2 0	0	2	0	1	1		1 7	L
23 005_5_04	18	0	0	2	1	0	1		1 1	2 2	0	0		2	2 0	6 3	0	0	0	1		2 0	1	2	0	0	1	1 3	1 1	i
24 005_5_054	16	0	1	1	1	1	0	(	0 2	1 2	0	0		2	2 0	1	0	0	0	1		2 0	1	2	0	- 1	. 1	. 3	1 1	L
25 005_S_05	13	0	0	2	2	0	0	1	1 3	2 2	0	0		2	2 0	<u>ĝ - 0</u>	0	0	0	0	6	2 0	1	2	0		1	1	1 7	L
26 005 5 05	2	1	1	1	1	0	0	(	0 (	0 2	0	0		2	2 1	1	1	0	1	0		2 0	2	2	0		) 1	. 3	2 1	L
27 005_5_060	12	0	0	2	2	0	0		2 3	2 2	1	0		2	2 0	r : 1	0	0	0	1		Z 0	2	2	1	1	Ó	) 8	1 7	L
28 005 \$ 06	.0	0	1	1	0	1	0	(	0	2 2	0	0		2	2 0	1	0	0	0	1	3	2 0	1	2	0	- 2	2 0		2 .7	L
29 005_5_08	4	0	2	0	0	2	0		1 :	2 2	0	0		2	2 0	1 1	0	0	0	0		2 0	1	2	1	6	1 1	3	1 7	L
30 005_5_092	19	0	0	2	2	0	0	1	1 ;	2 2	0	0		2	2 0	6	0	0	0	0	3	2 0	1	2	0	1	1		2 7	L
31 005_5_12	14	0	1	1	1	1	0	1	1 3	2 2	0	0		2	2 0	1 1	0	0	0	1		2 0	1	2	0		) 1		1 1	£
32 005_5_134	I	0	1	2	1	0	0		2 2	2 2	0	2		2	2 0	6 3	0	0	0	1		2 0	1	2	0	1	1	1 3	2 7	Ł
33 006 5 048	14	0	1	2	1	0	0		2 1	2 2	0	0		2	2 0	<u> </u>	0	0	0	1		2 0	0	2	0		2 0	;	1 3	L
34 006_5_049	18	1	2	0	0	1	1	(	0 3	2 2	0	1		2	2 0	<u>6 (</u> 3	0	0	0	0		2 0	2	2	0	1	1	3	2 7	L
35 006 S 054	17	0	1	1	1	1	0	(	0	1 2	0	0	-	2	2 1	1	1	0	1	0		2 0	0	2	1		) 1		1 1	I
36 006_5_06	13	0	1	2	0	0	2	1	1 1	2 2	0	1		2	2 0	1	0	0	0	1		2 0	2	2	0	(	2		1 7	L
37 006 S 067	5	0	1	1	1	1	0	(	0	1 2	0	0		2	2 0	<u> </u>	. 0	0	0	2		2 0	2	2	0	1	0	3	2 7	L
38 006_5_068	11	1	2	1	0	0	1		1 .	2 2	1	0		2	2 0	1	0	0	0	0	- 4	2 0	1	2	1	1	0	<u>i 1</u>	1 1	Ł
39 006 5 073	1	0	1	1	1	1	0	(	0) ;	2 2	0	1		2	2 0		0	0	0	1	3	2 0	1	2	1	(	1	: 3	1 1	L
N	Tota	alScored	÷																											1

#### Support Vector Regression (SVR)

Where the red lines represent the boundaries and the blue line represent the hyper-plane.



July 2020 🤇

#### **Classification Results**

•

Class	Precision	Recall	F1-score	Support		
0 (Diseased)	1.00	0.92	0.96	222		
1 (Normal)	0.88	1.00	0.93	129		

#### Confusion Matrix:

220	0
19	112

Accuracy is 94%



12 )

#### **Clustering** (Mini-Batch K-Means)

We used the guidelines on the sklearn website as a guidance to see which clustering algorithm would be best according to our scenario.



July 2020 C

#### **Clustering Results**

•

Class	Precision	Recall	F1-score	Support
1 (Normal)	0.67	0.48	0.56	150
2 (Potential)	0.60	0.87	0.71	219
3 (Diseased)	0.75	0.57	0.65	220

Cont	tusion	$\mathbb{N}$	latr	İX:	

72	65	13
1	190	120
35	60	125

Accuracy is 65%



14 )

## Database Diagram



July 2020

#### Sequence Diagram





Researcher

Admin



-•





July 2020 O-

## Impact

• Early diagnosis helps in slowing down progression of the diseased.

18 )

- Give patients more choice of medications that can improve their symptoms.
- Gives more time for the family and loved ones to understand the disease and plan for their future.

## DEMO

July 2020 C

## THANKYOU!

#### Github: github.com/fairuzsoufy/Alzheimers-GRAD