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**

Agenda

- Introduction
- Related Work
- System Overview
- Problem Statement
- Expected Results
- Demo

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.

[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

Introduction

DNA Structure:

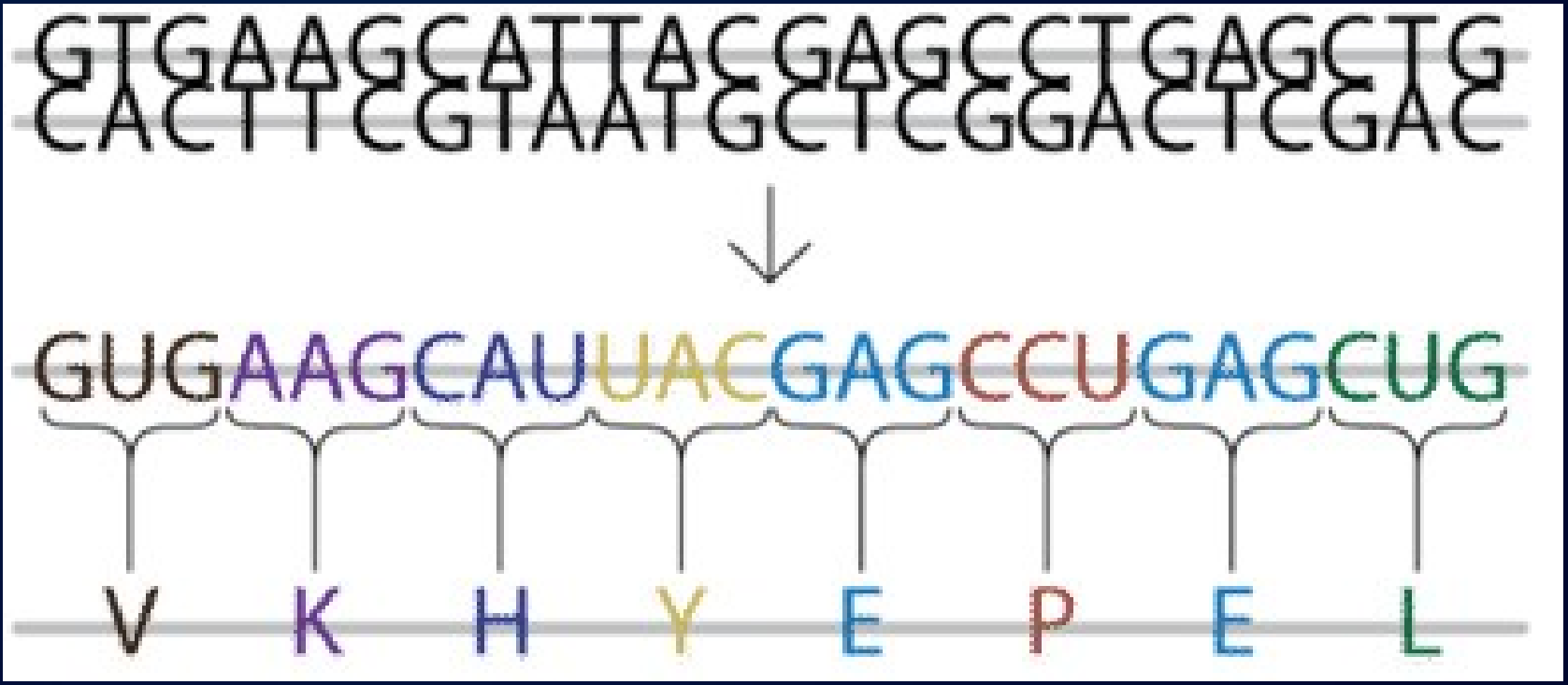
- Our DNA is 99.9% the same [2].
- There are basically four nucleotide bases, which make up the DNA. Adenine (A), Guanine (G), Thymine (T) and Cytosine(C).

[2] Lydia, Ramsey, and Lee Samantha. "Our DNA is 99.9% the same as the person sitting next to us-and were surprisingly similar to a bunch of other living things, Business Insider (2016)."

Introduction

DNA Structure:

- A DNA sequence looks like this:

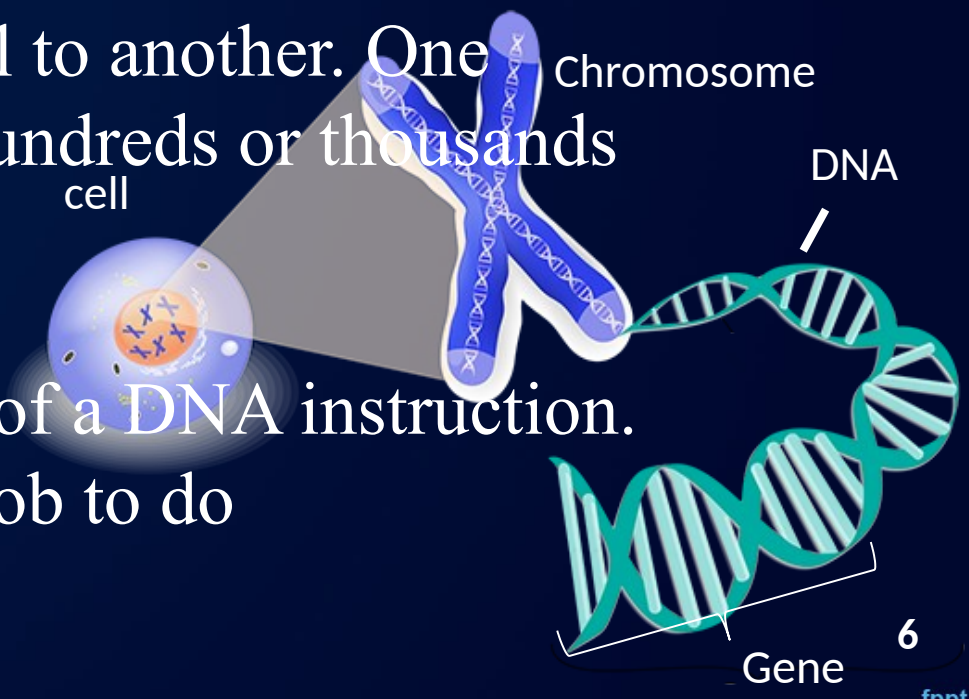


Introduction

What is a chromosome, and what are genes?

- A **chromosome** is how DNA is stored and transferred from one cell to another. One chromosome contains hundreds or thousands of genes.

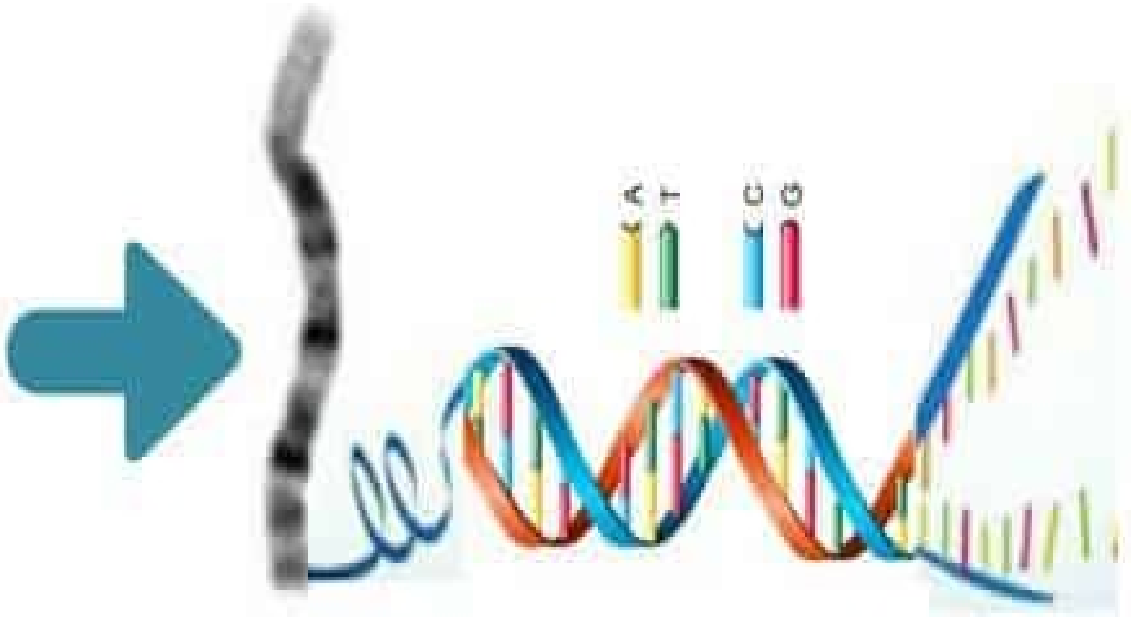
- A **gene** is the basic unit of a DNA instruction. Each gene has specific job to do



Chromosomes are Books

A chromosome contains a specific group of genes.

Like a book,
a chromosome holds a
group of genes.



Introduction

Causes of Alzheimer's disease?

Mutations with genes even small changes to a gene can cause diseases like Alzheimer's disease.

- Some cases are caused by an inherited change in one of the four genes.
- Certain environmental factors, like exposure to toxic chemicals have long been known to increase the risk of Alzheimer's.[3]

Viatcheslav Wlassoff, PhD ," NEUROSCIENCE & NEUROLOGY 0 Environmental Factors in Development of Alzheimer's Disease " September 15, 2017

Related work 1 :

Predicting cancer type from tumor DNA signatures [4]

Detecting from a person's DNA signatures the type of cancer

Method:

- SVMs with a linear kernel
- logistic regression
- random forests

they also consider precision and recall.

Challenges:

Genomic alterations in tumors are extremely complex, and large differences occur between tumor genotypes of not only different cancer types but also within the same type.

Accuracy : 77.7 %

[4]Soh, Kee Pang, et al. "Predicting cancer type from tumour DNA signatures." *Genome medicine* 9.1 (2017)."

Related work 2:

Detecting Mutations in Patients with Early-OnsetDementia Detected by Sequence Analyses of Four Different Genes [5].

Detecting Mutations in the genes inside DNA sequence PSEN1, PSEN2, APP and PRNP consider with EOD.

Accuracy: In 14 patients, they found a total of 12 different mutations in PSEN1, PSEN2, APP and PRNP Of the 12 mutations, 5 were previously un-described and 7 were known.

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 [3] of Alzheimer's is either the patient is healthy (stage A) or the patient is in the last stage of Alzheimer (stage D). We are Showing two new stages which are:

1. The gene carrier (stage B): A healthy patient at the moment but has inherited the disease from his family and doesn't have any symptoms of the disease.
2. The early symptoms stage (stage C) : A stage where the patient starts to notice some symptoms manifesting .

In order to distinguish between stages A and D we convert the sample DNA sequences into texts and measure the distance between them and a reference sample and if the distance exceeds zero then the patient is in class D. Later on, we are going to use CNN [4](convolutional neural network) for classifications.

[3]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.

[4] Chithrananda, Seyone. "Using CNNs to Understand the Intricacies of Our Genome and Predict Future Diseases." *Medium*, Towards Data Science, 9 Mar. 2019

Collecting DNA Samples

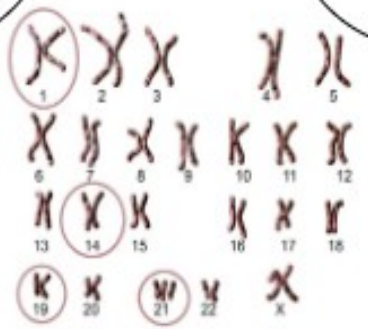
Open DNA on

geneious prime

use it to

select 4 chromosomes

cut the genes from the chromosomes



4 Alzheimer's Genes Identified:

APP | PS-1 | PS-2 | APOE4

CGT → Protein

TAC → Protein

GAT → Protein

extract wanted features

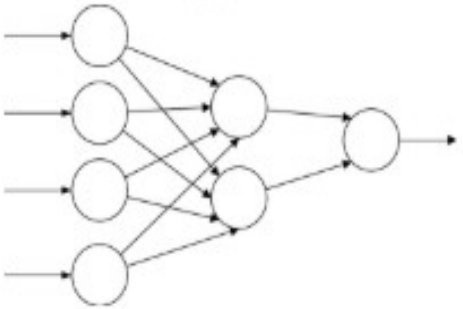
hamming

one hot coding all data inside file

Choose Model

classifying

Train model on Training Set



Dictionary D = A → [00] = 0
T → [01] = 1
G → [10] = 2
C → [11] = 3

Decimal-binary vector

Sequence B = ATAGC

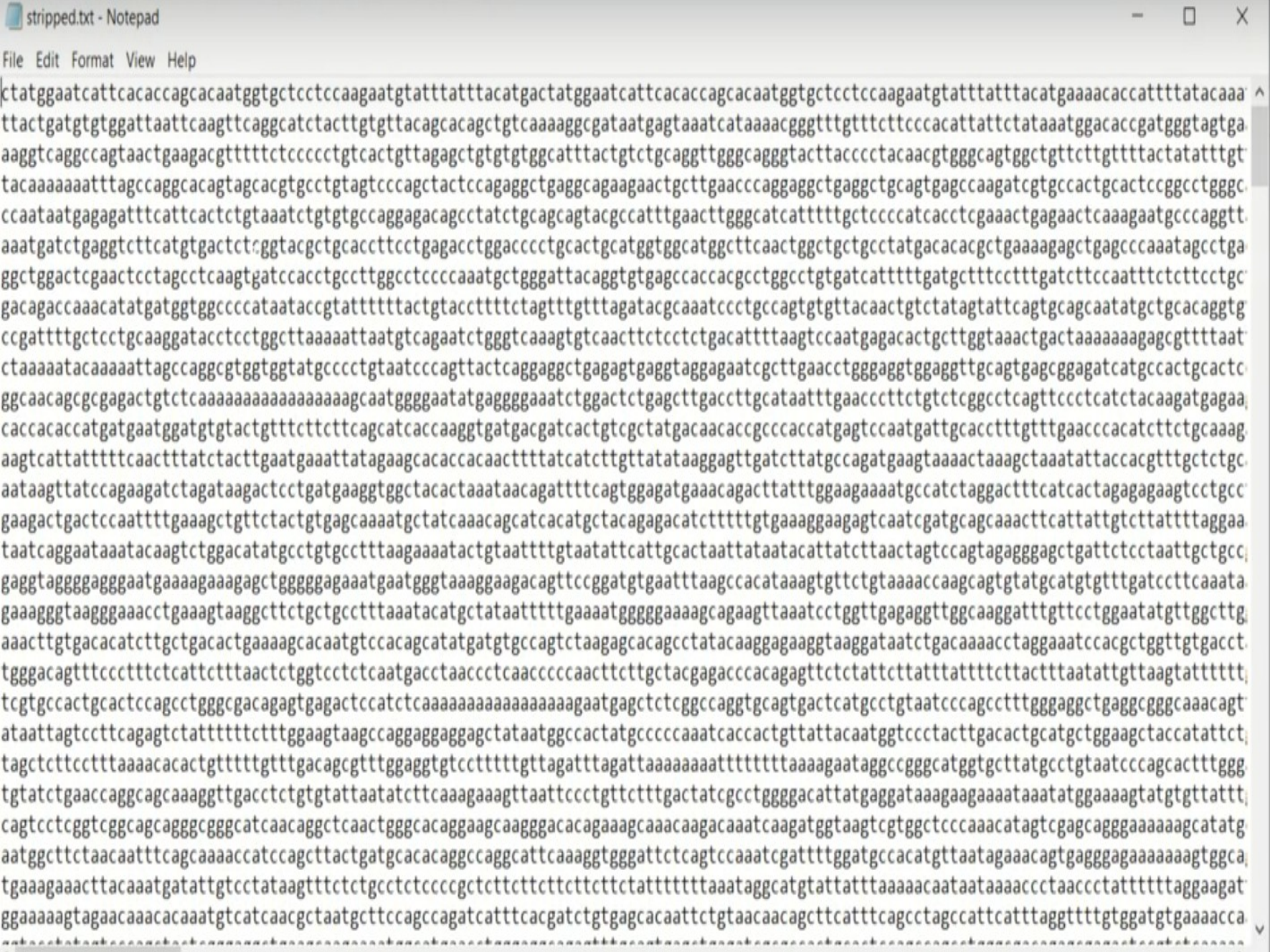
[0 0 0 1 0 0 1 0 1 1]

Evaluate model on Validation Set

Pick the model that does the best validation

Confirm results onTest Set

Result


stripped.txt - Notepad
File Edit Format View Help
ctatggaatcattcacaccagcacaatggtgctcctccaagaatgtatttatttacatgactatggaatcattcacaccagcacaatggtgctcctccaagaatgtatttatttacatgaaaacaccattttatacaaa
ttactgatgtgtggattaattcaagttcaggcatctacttgtgttacagcacagctgtcaaaaggcgataatgagtaaatcataaaacgggtttgtttcttcccacattattctataaatggacaccgatgggtagtga
aaggtcaggccagtaactgaagacgtttttctccccctgtcactgttagagctgtgtgtggcatttactgtctgcaggttgggcagggtacttacccctacaacgtgggcagtggctgttcttgttttactatatttgt
tacaaaaaaatttagccaggcacagtagcacgtgcctgtagtcccagctactccagaggctgaggcagaagaactgcttgaacccaggaggctgaggctgcagtgagccaagatcgtgccactgcactccggcctgggc
ccaataatgagagatttcattcactctgtaaatctgtgtgccaggagacagcctatctgcagcagtacgccatttgaacttgggcatcatttttgctccccatcacctcgaaactgagaactcaaagaatgcccaggtt
aaatgatctgaggtcttcatgtgactctcggtacgctgcaccttcctgagacctggaccctgcactgcatggtggcatggcttcaactggctgctgcctatgacacacgctgaaaagagctgagcccaaatagcctga
ggctggactcgaactcctagcctcaagtgatccacctgccttggcctccccaaatgctgggattacaggtgtgagccaccacgcctggcctgtgatcatttttgatgctttccttgatcttccaatttctcttcctgc
gacagaccaaacatatgatggtggccccataataccgtatttttactgtacctttctagtttgtttagatacgcaaatccctgccagtgtgttacaactgtctatagtattcagtgcagcaatatgctgcacaggtg
ccgattttgctcctgcaaggatacctcctggcttaaaaattaatgtcagaatctgggtcaaagtgtcaacttctcctctgacattttaagtccaatgagacactgcttggtaaactgactaaaaaaagagcgttttaat
ctaaaaatacaaaaattagccaggcgtggtggtatgcccctgtaatcccagttactcaggaggctgagagtgaggtaggagaatcgcttgaacctgggaggtggaggttgcagtgagcggagatcatgccactgcactc
ggcaacagcgcgagactgtctcaaaaaaaaaaaaaaaaaagcaatggggaatatgaggggaaatctggactctgagcttgaccttgcataatttgaaccccttctgtctcggcctcagttccctcatctacaagatgagaa
caccacaccatgatgaatggatgtgtactgtttcttcttcagcatcaccaaggtgatgacgatcactgtcgctatgacaacaccgcccaccatgagtccaatgattgcacctttgtttgaacccacatcttctgcaaag
aagtcattatttttcaactttatctacttgaatgaaattatagaagcacaccacaacttttatcatcttgttatataaggagttgatcttatgccagatgaagtaaaactaaagctaaatattaccacgtttgctctgc
aataagttatccagaagatctagataagactcctgatgaaggtggctacactaaataacagattttcagtggagatgaaacagacttatttggaagaaaatgccatctaggactttcatcactagagagaagtcctgcc
gaagactgactccaattttgaaagctgttctactgtgagcaaaatgctatcaaacagcatcacatgctacagagacatctttttgtgaaaggaagagtcaatcgatgcagcaaacttcattattgtcttattttaggaa
taatcaggaataaatacaagtctggacatatgcctgtgcctttaagaaaatactgtaattttgtaatattcattgcactaattataatacattatcttaactagtccagtagagggagctgattctcctaattgctgcc
gaggtaggggagggaatgaaaagaaagagctgggggagaaatgaatgggtaaaggaagacagttccggatgtgaatttaagccacataaagtgttctgtaaaaccaagcagtgtatgcatgtgtttgatccttcaaata
gaaagggtaagggaaacctgaaagtaaggcttctgctgcctttaaatacatgctataatttttgaaaatgggggaaaagcagaagttaaatcctggttgagaggttggcaaggatttgttcctggaatatgttggcttg
aaacttgtgacacatcttgctgacactgaaaagcacaatgtccacagcatatgatgtgccagtctaagagcacagcctatacaaggagaaggtaaggataatctgacaaaacctaggaaatccacgctggttgtgacct
tgggacagtttccctttctcattctttaactctggtcctctcaatgacctaaccctcaacccccaacttcttgctacgagacccacagagttctctattcttatttatttttcttactttaatattgttaagtatttttt
tcgtgccactgcactccagcctgggcgacagagtgagactccatctcaaaaaaaaaaaaaaaaagaatgagctctcggccaggtgcagtgactcatgcctgtaatcccagcctttgggaggctgaggcgggcaaacagt
ataattagtccttcagagtctatttttttctttggaagtaagccaggaggaggagctataatggccactatgcccccaaatcaccactgttattacaatggtccctacttgacactgcatgctggaagctaccatattct
tagctcttcctttaaaacacactgtttttgtttgacagcgtttggaggtgtccttttttgttagatttagattaaaaaaaattttttttaaaagaataggccgggcatggtgcttatgcctgtaatcccagcactttggg
tgtatctgaaccaggcagcaaaggttgacctctgtgtattaatatcttcaaagaaagttaattccctgttctttgactatcgcctggggacattatgaggataaagaagaaaataaatatggaaaagtatgtgttattt
cagtcctcggtcggcagcagggcgggcatcaacaggctcaactgggcacaggaagcaagggacacagaaagcaaacaagacaaatcaagatggtaagtcgtggctcccaaacatagtcgagcagggaaaaagcatatg
aatggcttctaacaatttcagcaaaaccatccagcttactgatgcacacaggccaggcattcaaaggtgggattctcagtccaaatcgattttggatgccacatgttaatagaaacagtgagggagaaaaaaagtggca
tgaaagaaacttacaaatgatattgtcctataagtttctctgcctctccccgctcttcttcttcttcttctatttttttaaataggcatgtattatttaaaaacaataataaaaccctaaccctatttttaggaagat
ggaaaaagtagaacaaacacaaatgtcatcaacgctaatgcttccagccagatcatttcacgatctgtgagcacaattctgtaacaacagcttcatttcagcctagccattcatttaggttttgtggatgtgaaaacca

Preprocessing:

- Load the DNA sample into Geneious Prime.
- Export the desired chromosomes using Geneious Prime.
- Cut the desired genes from the chromosomes .
- One hot encoding the files.
- Load the files into a CSV file to be later loaded to the classifier.

Expected Results

- Detection of Alzheimer's.
- Differentiate between classes B and C based on the DNA sample so Alzheimer's is classified into four classes :

Class A	Class B	Class C	Class D
Non-defected person	person with a high risk of Alzheimer's(Family history)	Defected person with early symptoms	a severe case of the disease

Benefits:

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

"Detecting Alzheimer's Decades Early." *Alzheimers.net*, 23 July 2019, https://www.alzheimers.net/7-27-15-detecting-alzheimers-decades-early/.

Demo

Any Questions?