

---

Νευρωνικά Δίκτυα:  
Ενδιάμεση Εργασία

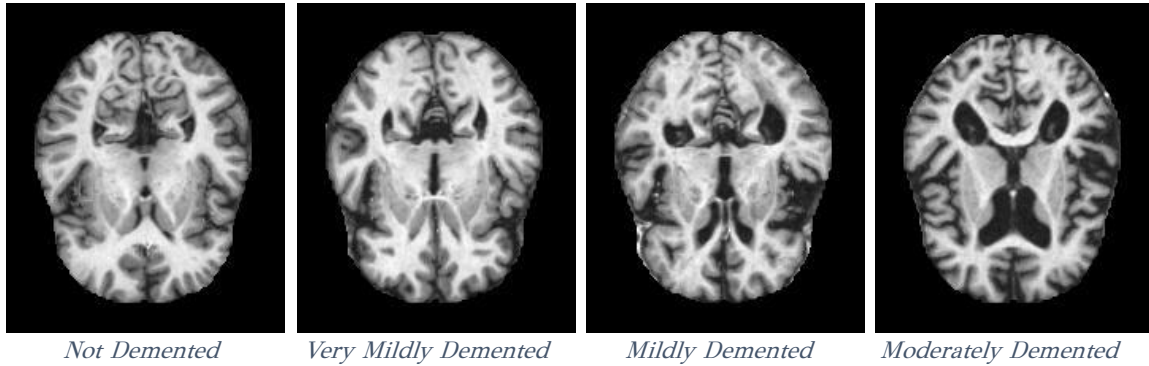
---

Αφροδίτη Λάτσκου

Νοέμβριος 2024

## Σύνολο Δεδομένων

Το σύνολο δεδομένων που επιλέχθηκε για την ερευνητική εργασία είναι το [Augmented Alzheimer MRI Dataset](#) από το Kaggle που περιέχει τόσο τις πραγματικές όσο και τις augmented MRI images ασθενών με και χωρίς Alzheimer σε διάφορα επίπεδα άνοιας.



Το dataset αποτελείται από τα εξής files:

	Original	Augmented	In Total
<i>Not Demented</i>	3200	9600	12800
<i>Very Mildly Demented</i>	2240	8960	11200
<i>Mildly Demented</i>	896	8960	9856
<i>Moderately Demented</i>	64	6464	6528

Επομένως δουλεύουμε με συνολικά 40384 εικόνες

Στο preprocessing βήμα, δημιουργήσαμε τα X και y υποσύνολα που περιέχουν τις εικόνες και τα labels αντίστοιχα και επίσης κάναμε flatten τις εικόνες ώστε να έχουμε μονοδιάστατους πίνακες, εφόσον τα KNN algorithms δουλεύουν με μονοδιάστατα δεδομένα.

Τέλος, πριν την εφαρμογή των αλγορίθμων χωρίσαμε τα δεδομένα μας σε training και testing data, εφόσον δεν υπήρχε ήδη ξεχωριστό testing υποσύνολο (60% to train, 40% to test).

## KNN

Οι KNN αλγόριθμοι ταξινομούν ένα νέο δείγμα με βάση του κοντινότερους σε αυτό “γείτονες” με βάση την απόσταση (εδώ ευκλείδεια). Στην περίπτωση μας έχουμε ένα πρόβλημα ταξινόμησης σε κλάσεις, επομένως το νέο δείγμα παίρνει την κλάση των περισσότερων γειτόνων του.

Χρησιμοποιήσαμε το KNearestClassifier της βιβλιοθήκης sklearn και αρχικοποιήσαμε δύο μοντέλα, ένα με  $k = 1$  και ένα με  $k = 3$ .

KNN with 1 Neighbor:

```
Accuracy: 0.8058685155379472
Classification Report:
              precision    recall  f1-score   support

     0           0.82       0.85       0.84       5116
     1           0.79       0.77       0.78       4488
     2           0.80       0.74       0.77       3899
     3           0.81       0.87       0.84       2651

 accuracy          0.81       0.81       0.81      16154
 macro avg         0.80       0.81       0.81      16154
weighted avg         0.81       0.81       0.81      16154

Confusion Matrix:
[[4359  392  235  130]
 [ 479 3478  355  176]
 [ 354  418 2884  243]
 [ 109  123  122 2297]]
```

KNN with 3 Neighbors:

```
Accuracy: 0.79175436424415
Classification Report:
              precision    recall  f1-score   support

     0           0.72       0.90       0.80       5132
     1           0.79       0.75       0.77       4477
     2           0.88       0.68       0.77       3953
     3           0.86       0.82       0.84       2592

 accuracy          0.79       0.79       0.79      16154
 macro avg         0.81       0.79       0.80      16154
weighted avg         0.80       0.79       0.79      16154

Confusion Matrix:
[[4604  335  118   75]
 [ 840 3344  187  106]
 [ 672  422 2704  155]
 [ 242  157   55 2138]]
```

Όπως παρατηρούμε, είχαν και τα δύο πολύ παρόμοια συμπεριφορά και είχαν ως αποτέλεσμα αρκετά καλές μετρικές. Ειδικά από τα confusion matrices βλέπουμε πως αποδίδουν βέλτιστα στην κλάση των Moderate Demented ασθενών και ταυτόχρονα έχουν πολύ αρκετά recall στην κλάση των Non-Demented (στην ικανότητα δηλαδή να ταξινομούν σωστά ένα δείγμα αυτής της κλάσης), δεδομένα που είναι πολύ σημαντικά όταν έχουμε να κάνουμε με ιατρικά datasets.

## Nearest Centroid

Στον αλγόριθμο nearest centroid, υπολογίζονται τα “κέντρα” κάθε κλάσης (μέσος όρος των δειγμάτων) και κάθε νέο δείγμα ταξινομείται με βάση την απόσταση από αυτό το σημείο.

Χρησιμοποιήσαμε το NearestCentroid της βιβλιοθήκης sklearn.

```
Accuracy: 0.4325244521480748
Classification Report:
              precision    recall  f1-score   support

     0           0.47       0.64       0.54       5132
     1           0.36       0.16       0.22       4477
     2           0.39       0.50       0.44       3953
     3           0.50       0.39       0.44       2592

 accuracy          0.43       0.43       0.43      16154
 macro avg         0.43       0.42       0.41      16154
 weighted avg      0.42       0.43       0.41      16154

Confusion Matrix:
[[3263  591  993  285]
 [1909  730 1445  393]
 [1034  596 1982  341]
 [ 779  101  700 1012]]
```

Βλέπουμε ήδη τεράστια διαφορά από τον προηγούμενο αλγόριθμο, με το NearestCentroid να αποδίδει πολύ χειρότερα και για τις 4 κλάσεις και ειδικά για την VeryMildDemented. Μπορούμε επομένως εύκολα να συμπεράνουμε πως είναι ο πιο αδύναμος εκ των 2.