

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is light green. They are positioned diagonally, with the blue one partially covering the green one.

Random forest με εξαγωγή χαρακτηριστικών



Χρήση media pipe για εξαγωγή χαρακτηριστικών

🔍 Εφτιαξα κώδικα όπου αναζητά όλα τα αρχεία βίντεο με κατάληξη .MOV σε συγκεκριμένους φακέλους, και για κάθε βίντεο εξάγει δεδομένα από σημεία-κλειδιά του σώματος (pose landmarks) με τη βοήθεια του MediaPipe BlazePose.

📄 Αποθηκεύει τα δεδομένα πόζας (x, y, z και visibility για επιλεγμένα σημεία του σώματος) σε ένα αρχείο CSV για κάθε βίντεο, περιλαμβάνοντας πληροφορίες όπως το ID του βίντεο, την πάθηση (π.χ. ΚΟΑ ή PD) και την πλευρά (π.χ. 01 ή 02).

Εξαγωγή χαρακτηριστικών και δημιουργία ακολουθιών με στατιστικά στοιχεία των χαρακτηριστικών

✓ 1. Μεταδεδομένα:

- ID
- Disease
- Side
- Level
- Window
- Source_File

✓ 2. Bilateral Similarity Features (Διμερή Ομοιότητα)

Για κάθε άρθρωση: Hip, Knee, Ankle, Heel

- Hip_Similarity_Mean
- Hip_Similarity_Max
- Hip_Similarity_Std
- Knee_Similarity_Mean
- Knee_Similarity_Max
- K.Λ.Π.

✓ 3. Step Length Features (Μήκος Βήματος)

Υπολογίζονται από τις αποστάσεις μεταξύ των αστραγάλων (ankles):

- Step_Length_Mean
- Step_Length_Max
- Step_Length_Std

✓ 4. Features ανά πλευρά (Left / Right)

Για κάθε πλευρά (Left / Right), εξάγονται τα εξής:

> Hip (γωνία)

- Left_Hip_Mean, Left_Hip_Max, Left_Hip_Std
- Right_Hip_Mean, Right_Hip_Max, Right_Hip_Std

> Knee γωνία (γωνία)

- Left_Knee_Mean, Left_Knee_Max, Left_Knee_Std
- Right_Knee_Mean, Right_Knee_Max, Right_Knee_Std

> Ankle (γωνία)

- Left_Ankle_Mean, Left_Ankle_Max, Left_Ankle_Std
- Right_Ankle_Mean, Right_Ankle_Max, Right_Ankle_Std

> Heel (ύψος πτέρνας)

- Left_Heel_Max, Left_Heel_Min
- Right_Heel_Max, Right_Heel_Min



Συνολικός αριθμός χαρακτηριστικών (εκτός μεταδεδομένων):

- Bilateral Similarity: 12 χαρακτηριστικά (4 αρθρώσεις \times 3 metrics)
 - Step Length: 3
 - Left πλευρά: 11
 - Right πλευρά: 11
- Σύνολο: 37 χαρακτηριστικά + 6 μεταδεδομένα = 43 στήλες**



Κανονικοποίηση δεδομένων προτού εισαχθούν στο RF με MinMaxScaler



Εκπαίδευση του Μοντέλου Random Forest

Για την ταξινόμηση των δεδομένων, χρησιμοποιήθηκε ο αλγόριθμος Random Forest με τις παρακάτω ρυθμίσεις (υπερ παραμέτρους):

- `n_estimators = 500` → αριθμός δέντρων στο δάσος
- `max_depth = 20` → μέγιστο βάθος κάθε δέντρου
- `min_samples_split = 10` → ελάχιστος αριθμός δειγμάτων για διαχωρισμό κόμβου
- `min_samples_leaf = 5` → ελάχιστα δείγματα σε φύλλο δέντρου

Χαρακτηριστικά Εκπαίδευσης:

- Χρησιμοποιήθηκε ομαδοποίηση με βάση το Id (Group-based training)
- Εκπαίδευση με δύο τεχνικές:
 - Train Test Split (Group ShuffleSplit τυχαίος χωρισμός έχοντας ως group το id των ατόμων)
 - Leave-One-Group-Out (LOSO) (κάθε φορά αποκλείεται ένα άτομο από την εκπαίδευση για δοκιμή)

Στόχοι ταξινόμησης:

- Μόνο η ασθένεια (Disease)
- Συνδυαστικά ασθένεια και στάδιο (Disease + Stage)