



EEG Signal Processing and Machine Learning

Roman Razvan

Scopul proiectului

Scopul proiectului este de a explora diferite tehnici de clasificare, utilizand algoritmi de machine learning, pentru a obține rezultate precise și robuste în clasificarea semnalelor EEG asociate cu imaginile motorii.

In acest proiect, am pus accentual pe urmatoarele tehnici de clasificare: CSP, si LDA.

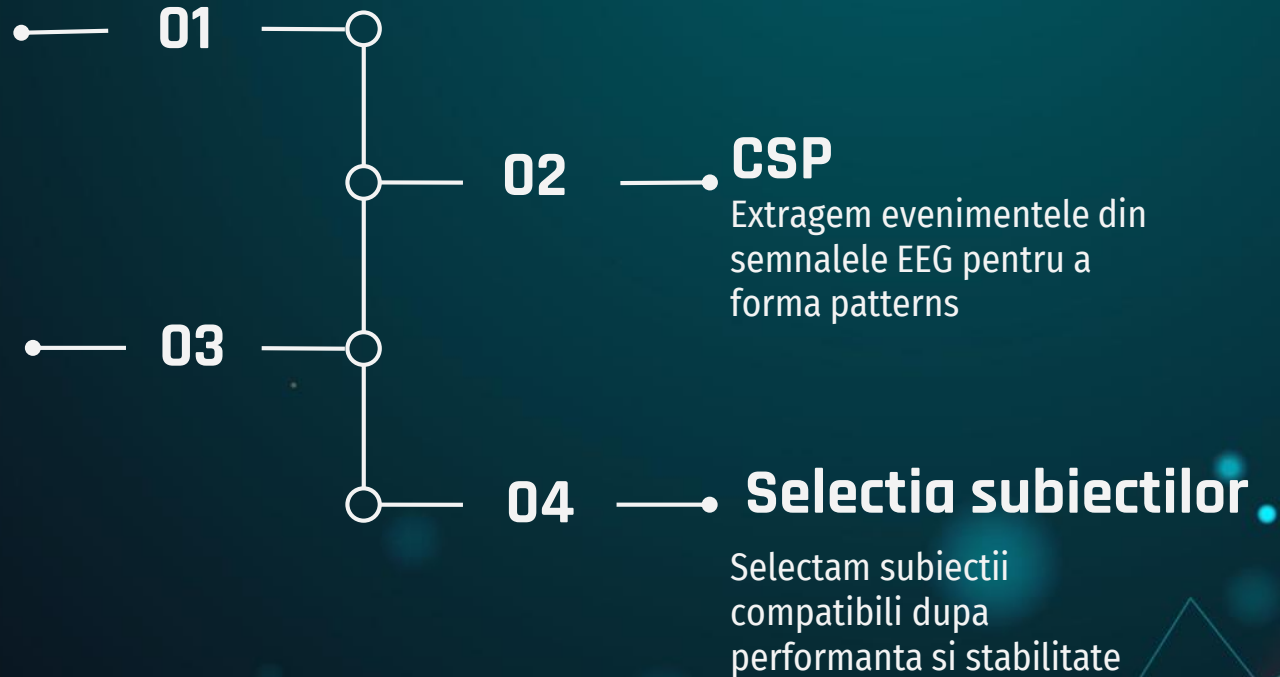
Metodologie

Preprocesare

Folosind un filtru trece banda între 7-30 Hz

LDA

Clasificam pe 2 clase folosind patternurile anterioare



Ce sunt CSP si LDA?



CSP

Common Spatial Patterns

CSP este o tehnică utilizată în prelucrarea semnalelor EEG pentru a extrage caracteristici relevante și distincte asociate cu imaginile motorii. Scopul principal al algoritmului CSP este să identifice combinații lineare de electrozi care maximizează diferența de varianță între două clase de semnale EEG. Aceste combinații lineare sunt numite filtre spațiale.

LDA

Linear Discriminant Analysis

LDA este o tehnică de clasificare utilizată pentru a separa și clasifica datele în două sau mai multe clase. Scopul principal al LDA este de a găsi o proiecție liniară a datelor într-un nou spațiu astfel încât să se maximizeze separabilitatea între clase și să se minimizeze variabilitatea internă a claselor.

Cum functioneaza CSP

Preprocesare

Semnalele EEG sunt prelucrate și segmentate în funcție de evenimentele sau stările mentale de interes.

Transformarea datelor într-un nou spațiu

Semnalele EEG sunt transformate într-un nou spațiu de caracteristici, în care componentele semnalelor sunt ponderate pentru a maximiza variația între clasele de evenimente

Identificarea filtrelor spatiale optime

Identifică filtrele spatiale care permit separarea optimă a semnalelor asociate cu clasele dorite. Aceste filtre sunt utilizate pentru a extrage caracteristici relevante din semnalele EEG.

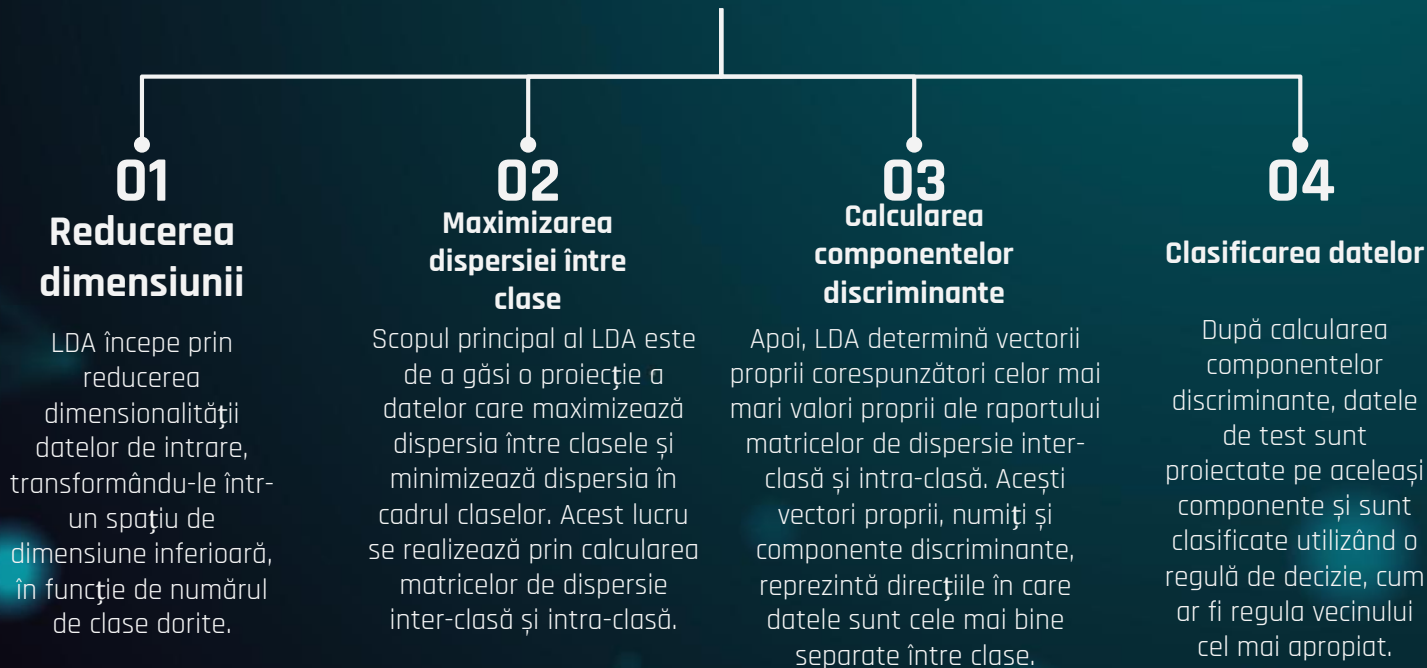
Clasificarea semnalelor

Caracteristicile obținute prin aplicarea filtrelor CSP și selectarea componentelor sunt utilizate pentru a antrena un algoritm de clasificare, cum ar fi LDA

Implementarea matematica CSP

1. Se obțin semnalele EEG brute de la electrozi selectați.
2. Se formează matricea covarianței pentru fiecare clasă de semnale EEG. Aceasta este o măsură a corelației dintre semnalele EEG de la diferiți electrozi.
3. Se calculează media matricelor de covarianță pentru fiecare clasă. Aceasta reprezintă media distribuției semnalelor EEG în cadrul fiecărei clase.
4. Se calculează matricea combinată de covarianță prin adăugarea matricilor de covarianță pentru cele două clase și se calculează inversa acesteia.
5. Se calculează valorile și vectorii proprii ai matricei inverse de covarianță combinată.
6. Se sortează valorile proprii în ordine descrescătoare și se selectează primele N valori proprii. Aceste valori proprii corespund celor mai semnificative caracteristici ale semnalelor EEG.
7. Se calculează filtrele spațiale prin înmulțirea matricei inverse de covarianță combinată cu vectorii proprii selectați.
8. Se aplică filtrele spațiale asupra semnalelor EEG pentru a obține semnalele filtrate CSP.

Cum functioneaza LDA

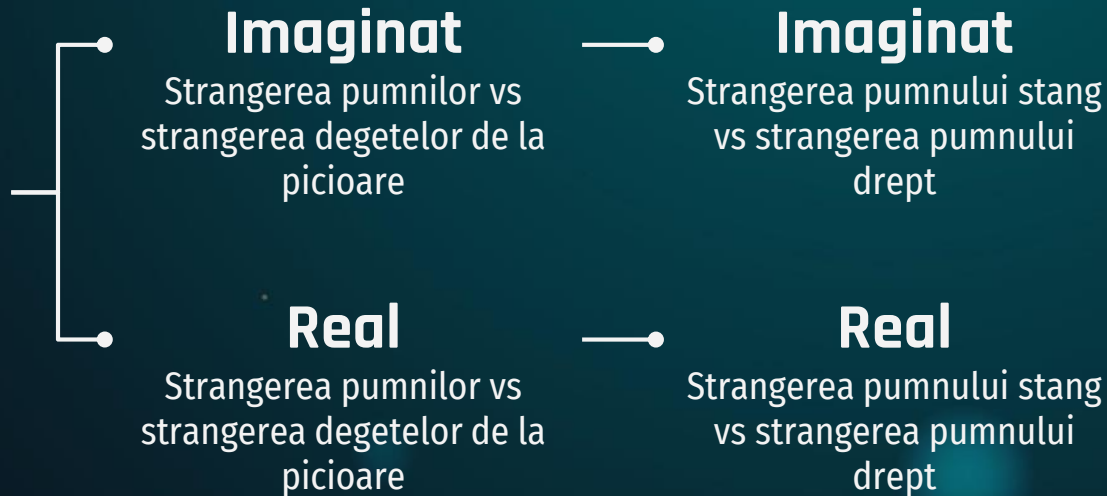


Implementarea matematica LDA

1. Se obțin caracteristicile extrase (de exemplu, semnalele filtrate CSP) și etichetele corespunzătoare ale claselor.
2. Se calculează vectorii medie pentru fiecare clasă. Acești vectori reprezintă media caracteristicilor pentru fiecare clasă.
3. Se calculează matricea de dispersie între clase, care măsoară diferența între medii ale claselor.
4. Se calculează matricea de dispersie internă pentru fiecare clasă, care măsoară variabilitatea caracteristicilor în cadrul fiecărei clase.
5. Se calculează matricea de dispersie totală ca sumă a matricilor de dispersie între clase și matricilor de dispersie interne.
6. Se calculează vectorii proprii și valorile proprii ale matricei inverse a matricei de dispersie totală.
7. Se sortează valorile proprii în ordine descrescătoare și se selectează primele K valori proprii. Acestea reprezintă cele mai semnificative caracteristici pentru clasificare.
8. Se construiește un model liniar folosind vectorii proprii selectați ca bază pentru proiecția datelor într-un nou spațiu.
9. Se clasifică noile instanțe de date proiectându-le în noul spațiu și aplicând un criteriu de decizie (de exemplu, dreapta de decizie).

Setul de date folosit

Pentru realizarea studiului, am folosit un set de date EEG care conținea peste 1500 de înregistrări EEG obținute de la 109 voluntari. Aceste înregistrări au fost efectuate în timp ce subiecții realizau diferite sarcini motorii sau își imaginau aceste sarcini. Setul de date a fost pre-etichetat, ceea ce mi-a permis să asociez fiecare înregistrare cu una dintre cele patru cazuri de clasificare a câte 2 clase

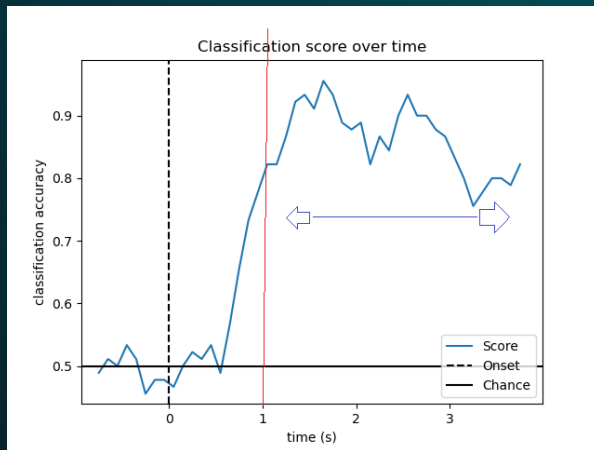


Parametrul de performanta

Scriptul este capabil să detecteze și să ignore subiecții care au o acuratețe în timp instabilă și sub 70%, având în vedere că semnalele EEG pot varia semnificativ de la o persoană la alta.

Această filtrare este posibilă folosind parametrul de performanță regăsit în cod ca și condiția:

```
'if np.mean(mean_performance[1:]) > 0.70
```



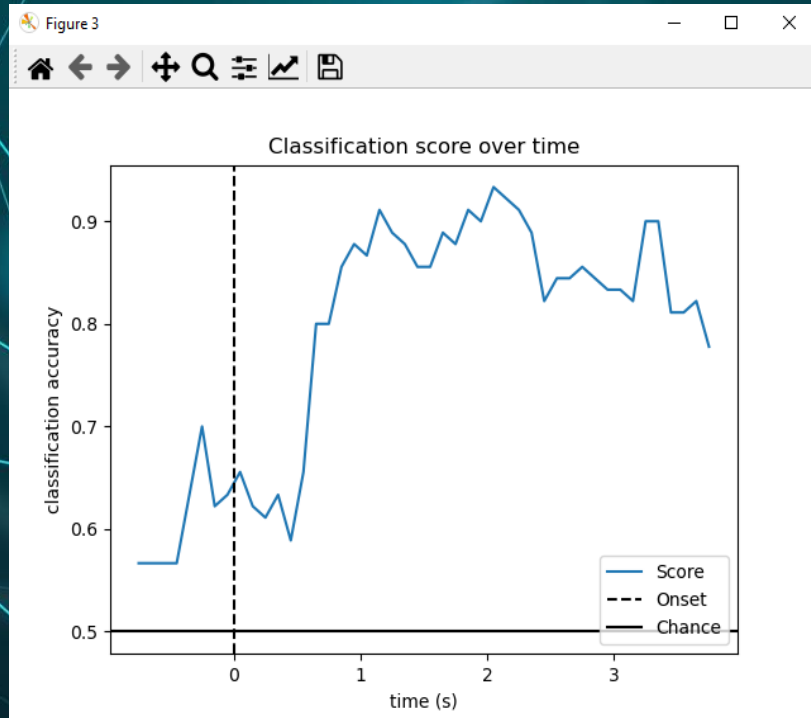
Am rulat pentru 3 valori pe fiecare caz:

0.60

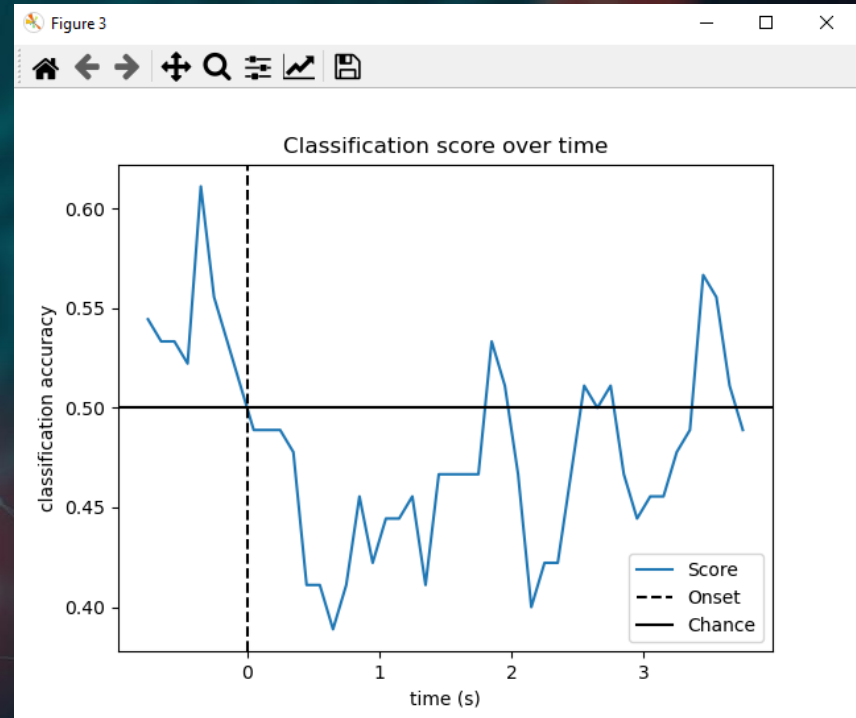
0.65

0.70

Exemplu subiect stabil vs instabil



Subiectul 7



Subiectul 5

Ce s-a folosit:

Conda enviornment

Un mediu izolat în care puteți instala și gestiona versiuni specifice de Python și pachete asociate

MNE-Python

MNE este o bibliotecă Python pentru analiza semnalelor EEG/MEG



Python 3.10

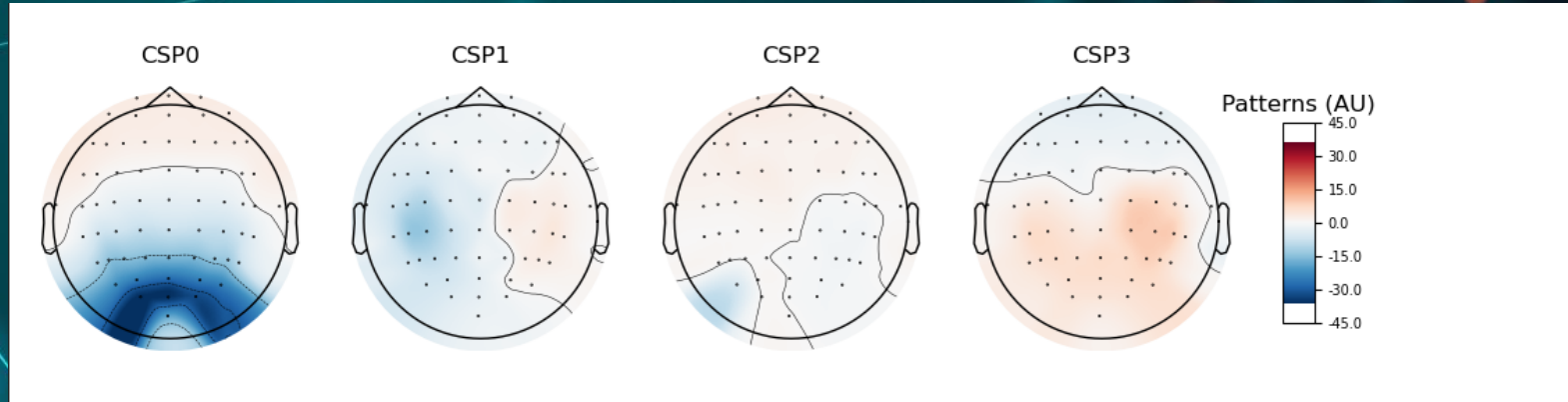
TensorFlow

Este proiectată pentru a permite construirea și antrenarea de modele de învățare automată și rețele neurale profunde

Scikit-learn

A fost creată pentru a oferi un set complet de instrumente pentru diverse sarcini de machine learning, precum clasificare, regresie, clustering și prelucrarea datelor.

Rezultate Modele CSP Patterns

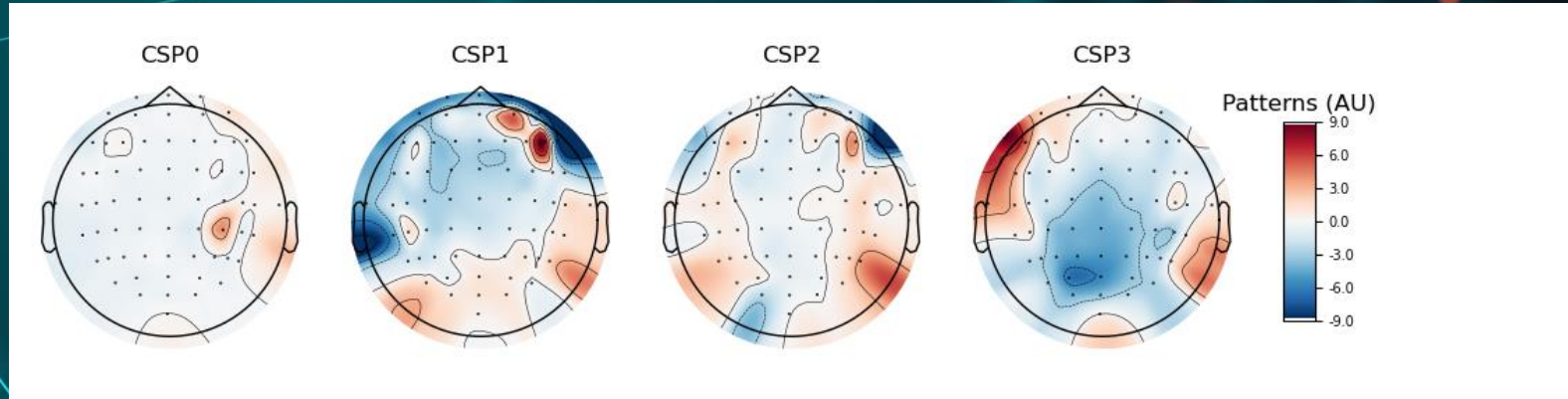


(Subject 7 – Imaginary Hands vs Feet) – Subiectul care a fost selectat în toate performanțele
Fiecare caz de clasificare are un model CSP diferit per subiect.

Se pot observa 4 imagini reprezentând distribuția electrozilor, denumite CSP0, CSP1, CSP2 și CSP3. Aceste imagini reprezintă distribuția componentelor spațiale estimate de algoritmul CSP și reprezintă regiuni diferite ale capului uman.

În ceea ce privește distincția între imaginarea mișcării mâinilor și picioarelor, subiectul 7 prezintă pattern-uri CSP caracteristice și semnificative, evidențiate prin un număr maxim de 35 de patterns (AU).

Rezultate Modele CSP Patterns

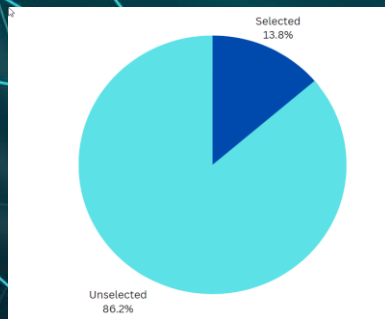


(Subject 5 – Imaginary Hands vs Feet) – Subiectul care NU a fost selectat în nicio performanțe
Fiecare caz de clasificare are un model CSP diferit per subiect.

În schimb, subiectul 5 nu prezintă un număr semnificativ de patterns (AU), având un număr maxim de 9. Acest lucru sugerează că subiectul 12 nu manifestă un model distinctiv al activărilor în contextul distincției între imaginarea mișcării mâinilor și picioarelor. Ploturile CSP asociate acestui subiect prezintă o suprapunere mai mare între cele două stări, făcând mai dificilă distingerea lor.

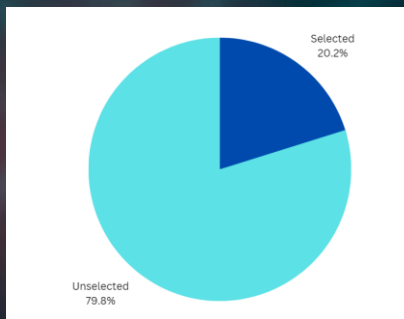
Rezultate Imaginary hands vs feet

Numărul de subiecți
neselectați: 94,
Numărul de subiecți
selecți: 15
Acuratețe: 0.924



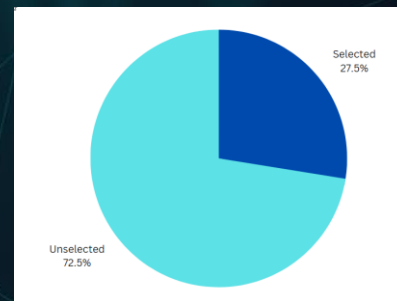
Performanta > 0.7

Numărul de subiecți
neselectați: 87,
Numărul de subiecți
selecți: 22
Acuratețe: 0.891



Performanta > 0.65

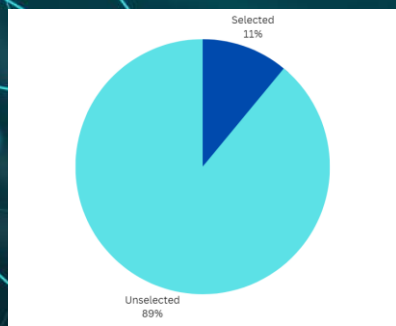
Numărul de subiecți
neselectați: 79,
Numărul de subiecți
selecți: 30
Acuratețe: 0.849



Performanta > 0.60

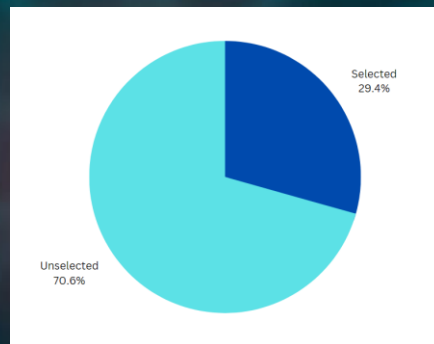
Rezultate Actual Movement hands vs feet (best)

Numărul de subiecți
neselectați: 97,
Numărul de subiecți
selecțați: 12,
Acuratețe: 0.976



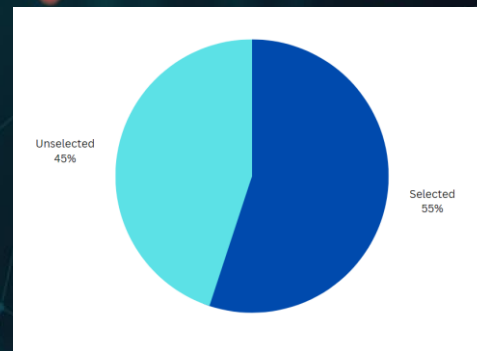
Performanta > 0.8

Numărul de subiecți
neselectați: 77,
Numărul de subiecți
selecțați: 32
Acuratețe: 0.924



Performanta > 0.70

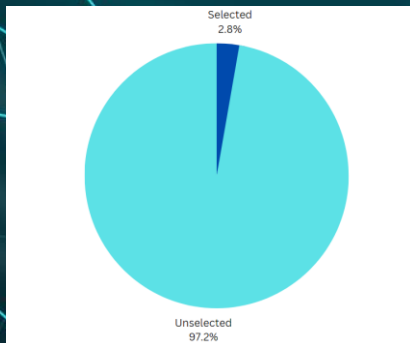
Numărul de subiecți
neselectați: 49,
Numărul de subiecți
selecțați: 60
Acuratețe: 0.862



Performanta > 0.60

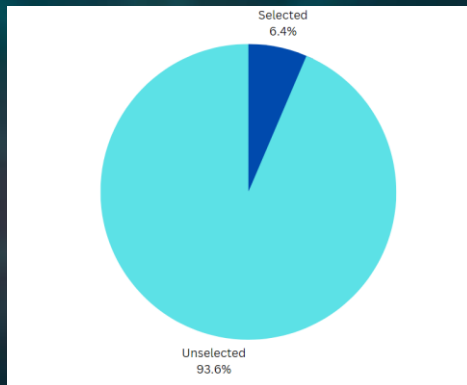
Rezultate Actual Imaginary right vs left hand

Numărul de subiecți neselectați: 106,
Numărul de subiecți selectați: 3, Acuratețe: 0.911



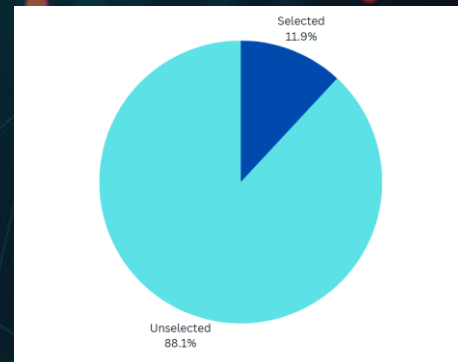
Performanta > 0.7

Numărul de subiecți neselectați: 102,
Numărul de subiecți selectați: 7, Acuratețe: 0.802



Performanta > 0.65

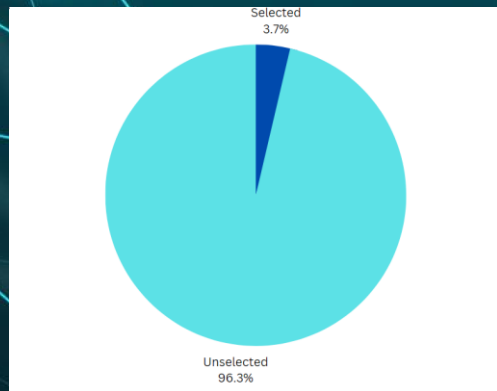
Numărul de subiecți neselectați: 96,
Numărul de subiecți selectați: 13, Acuratețe: 0.776



Performanta > 0.60

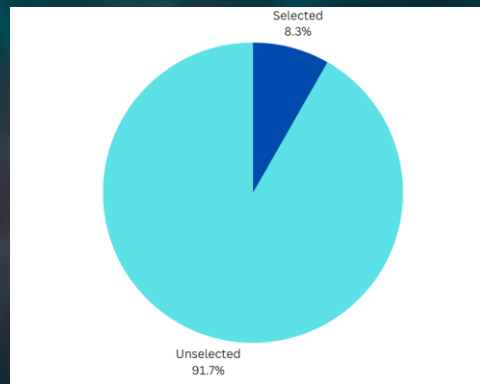
Rezultate Actual Actual movement right vs left

Numărul de subiecți neselectați: 105,
Numărul de subiecți selectați: 4, Acuratețe: 0.911



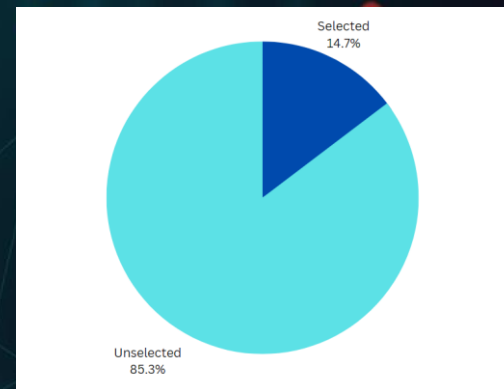
Performanta > 0.7

Numărul de subiecți neselectați: 100,
Numărul de subiecți selectați: 9, Acuratețe: 0.828



Performanta > 0.65

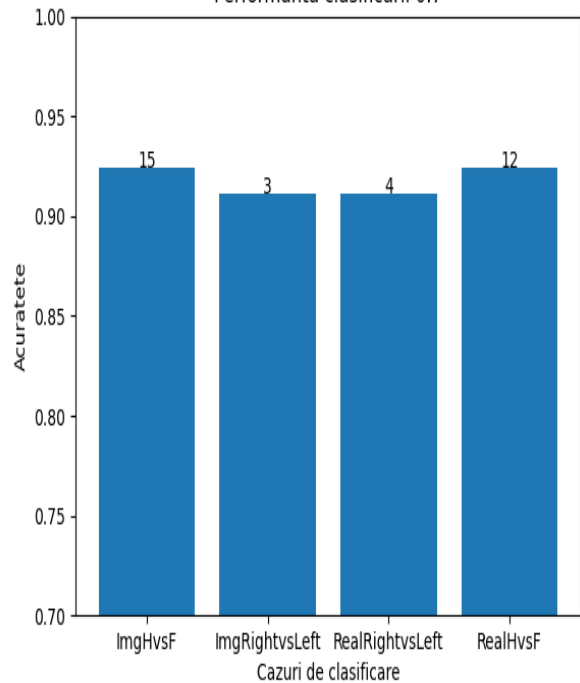
Numărul de subiecți neselectați: 93,
Numărul de subiecți selectați: 16, Acuratețe: 0.769



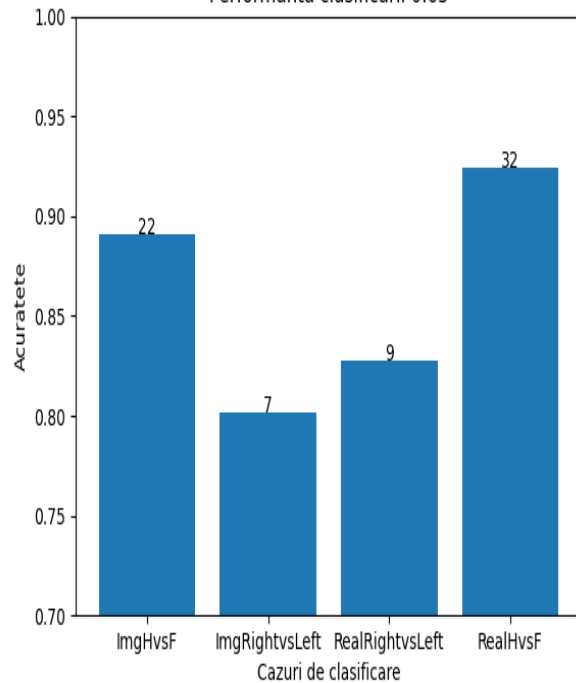
Performanta > 0.60

Concluzie

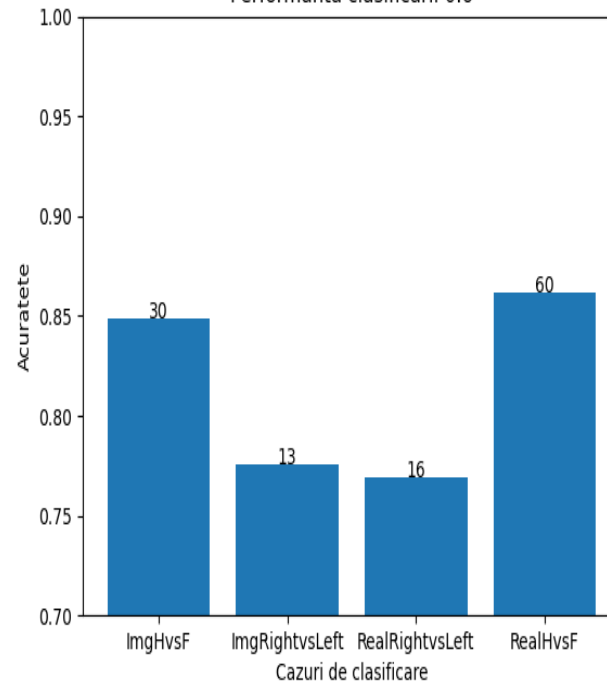
Performanta clasificarii 0.7



Performanta clasificarii 0.65



Performanta clasificarii 0.6



Concluzie

În general, calitatea rezultatelor este bună, cu performanțe excelente în cazul detecției mișcărilor reale ale mâinilor versus picioarelor. Cu toate acestea, se pot observa anumite dificultăți în distingerea imaginilor imaginare ale mâinii drepte față de mâna stângă și mișcărilor reale ale mâinii drepte față de mâna stângă, care pot necesita investigații și îmbunătățiri ulterioare. Totodată, este important de menționat că selecția adecvată a subiecților a fost crucială pentru obținerea rezultatelor precise.