SZABADON HOZZÁFÉRHETŐ ARCFELISMERÉSI KÖNYVTÁRAK ADATVÉDELMI SZEMPONTÚ VIZSGÁLATA

CSIKTUSNÁDI-KISS KENÉZ – BSC SZAKDOLGOZAT

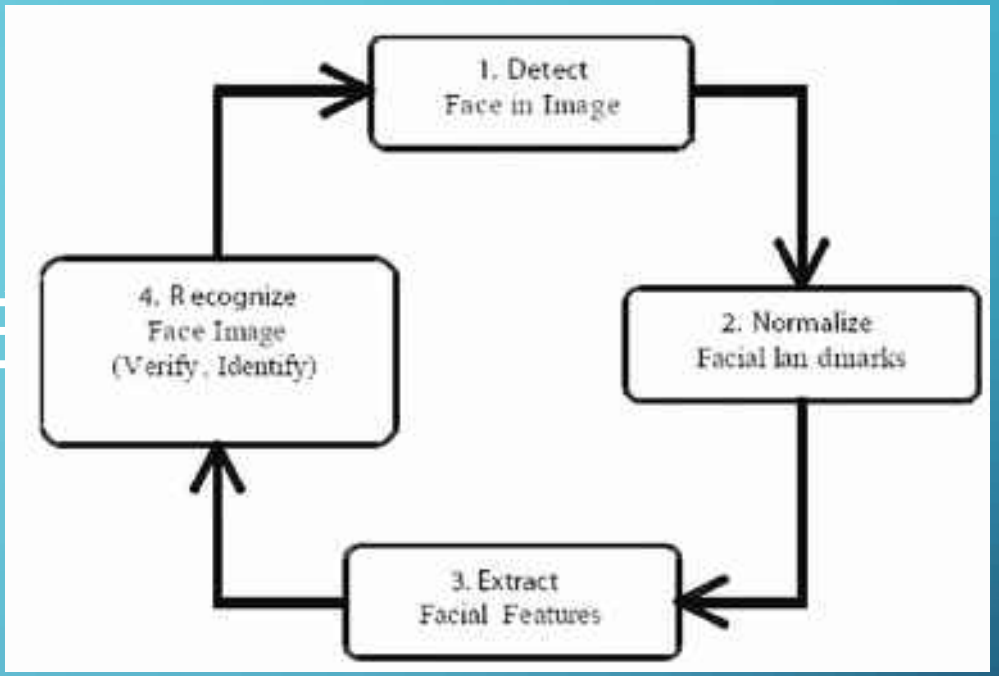
KONZULENSEK: DR. GULYÁS GÁBOR GYÖRGY ÉS FÁBIÁN ISTVÁN

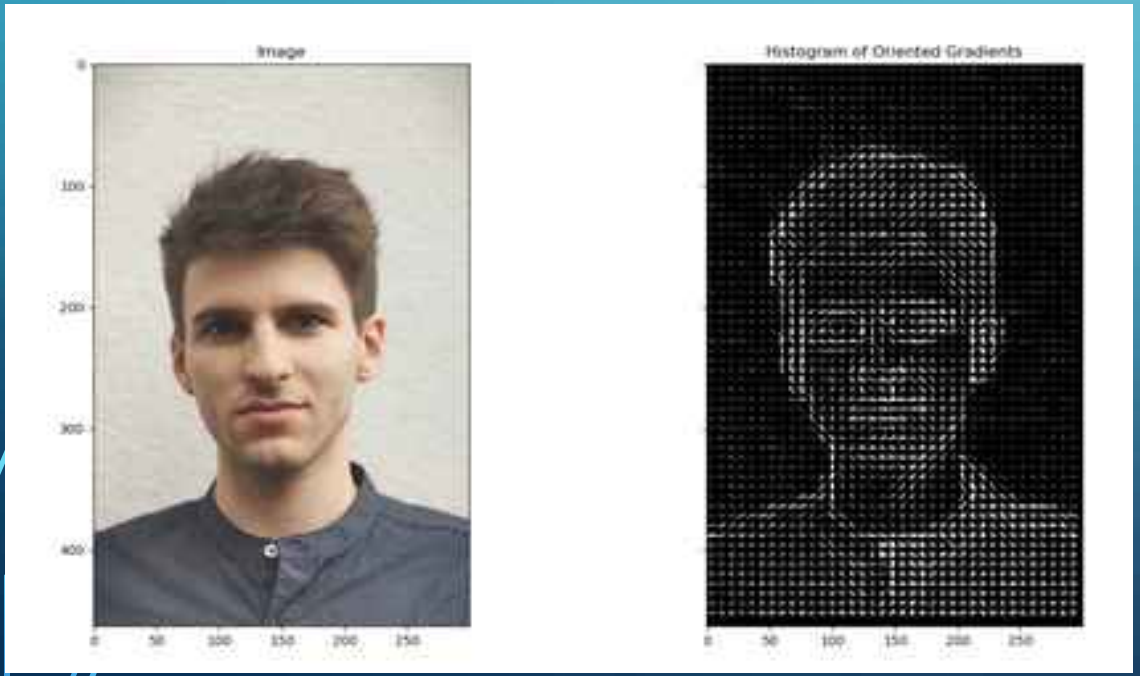
Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Automatizálási és Alkalmazott Informatikai Tanszék

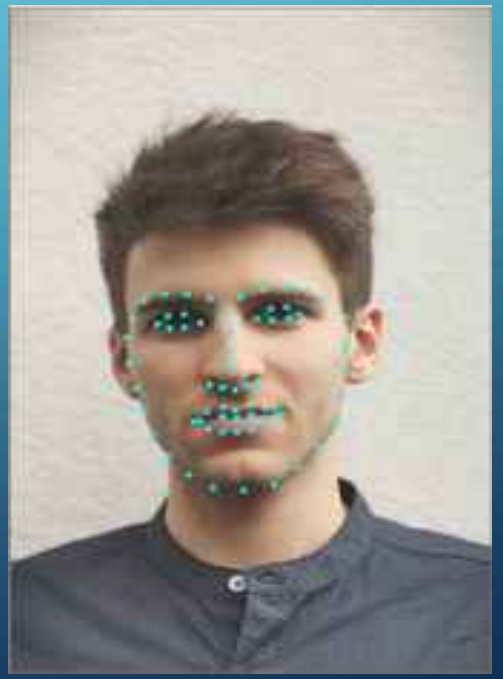
A MUNKÁM CÉLJA

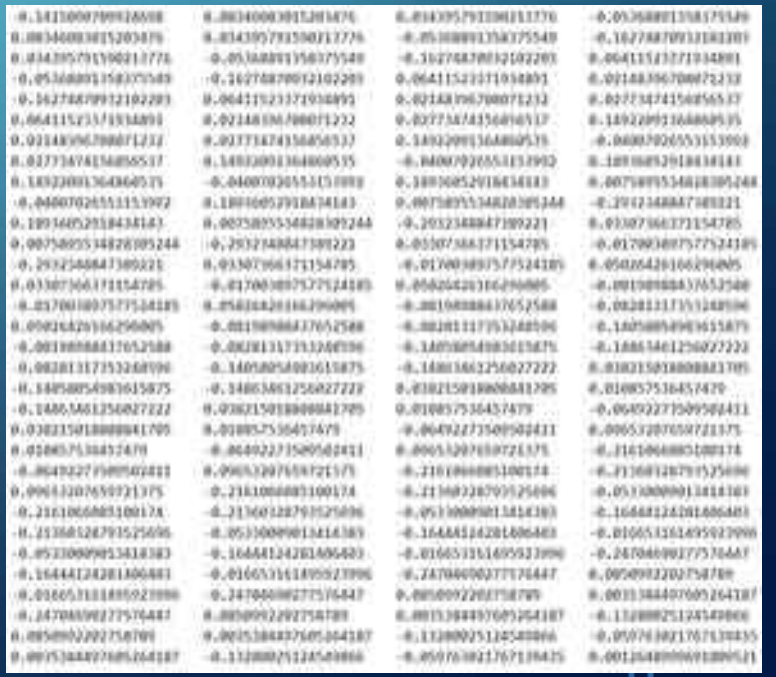
- Elemezni három nyílt forráskódú arcfelismerési eljárást működését
- Összehasonlítani ezen könyvtárak működése során létrejövő ún. embedding-eket, aszerint, hogy milyen mértékben lehet belőlük demográfiai adatokat visszaállítani
- Megvizsgálni, hogy az arcfelismerési folyamat során kinyert biometrikus adatok, milyen adatbiztonsági kockázatokat hordoznak

AZ ARCFELISMERÉS MŰKÖDÉ

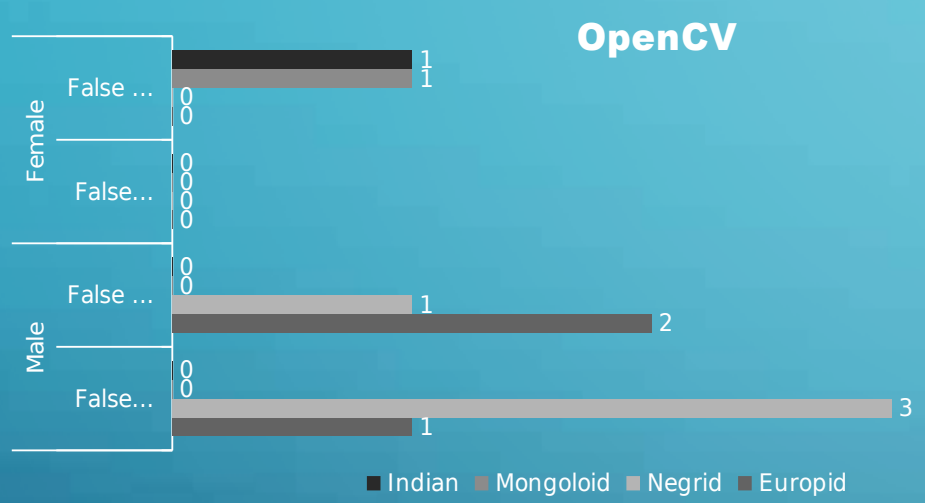


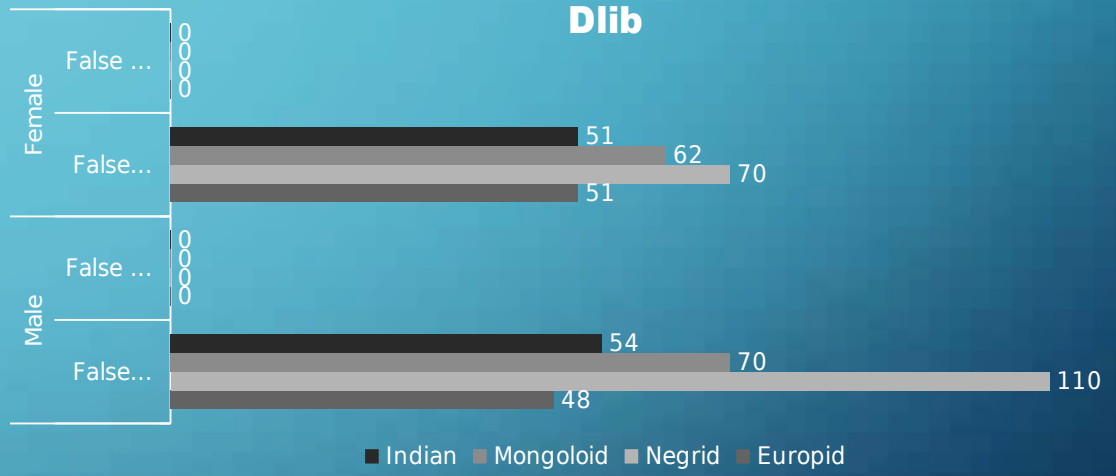


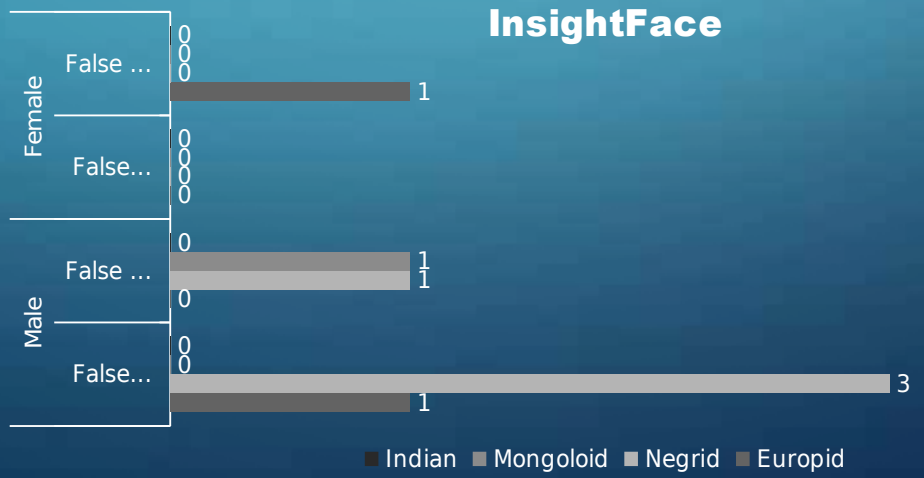




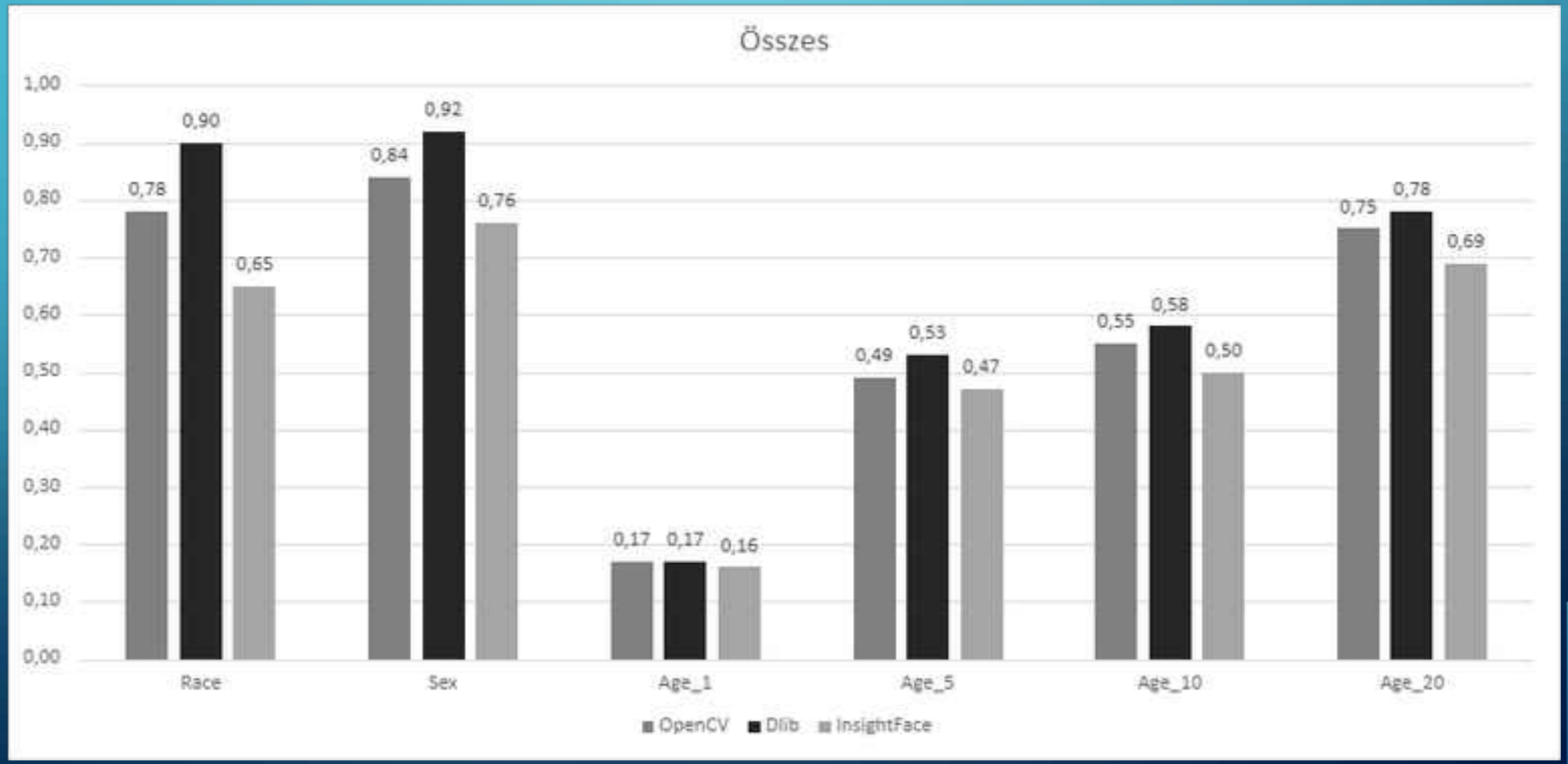
KÖNYVTÁRAK STATISZTIKÁI







EMBEDDINGEK STATISZTIKÁI



A TÁMADÓ MODELL BEMUTATÁSA

Kiinduló adat	• Embedding amit deanonimizálni akarunk
Prediktorok	• A demográfiai adatok kinyerése
Közösségi média	• Azonos demográfiai adatokkal rendelkező profilok keresése
Lehetséges találatok	• Keresett Embeddinggel való összevetés
Megtalált profil	• A kiinduló Embeddinghez tartozó személy

TÁMADÁS ELLENŐRZÖTT KÖRNYEZETBEN

- Célja, hogy statisztikai adatokkal szolgáljon arról, hogy a támadási modell, milyen hatékonysággal képes egy embedding alapján egy személyt azonosítani
- A siker ráta: **36.7%**

```
match_candidates = {}
for i in range(0, len(deanonim_candidates_df)):
    target = deanonim_candidates_df.loc[[i], :]
    if (int(target.at[i, 'Sex']) == int(sex_prediction) and
            int(target.at[i, 'Race']) == int(race_prediction) and
            (
            int(target.at[i, 'Age']) - 20 <= int(age_prediction) <= int(target.at[i, 'Age']) + 20
            )):
        target = target.drop(columns=['Sex', 'Race', 'Age'])
        result = face_recognition.compare_faces([target.to_numpy().astype(np.float)[0]],
                                                embedding_to_deanonim.to_numpy().astype(np.float)[0], tolerance=0.5)
        if result[0]:
            match_candidates[i] = target.to_numpy().astype(np.float)[0]
```

TÁMADÁS A KÖZÖSSÉGI MÉDIA BEVONÁSÁVAL

A közösségi médiában a keresési szempontok:

- A személyek halmazát kellő mértékben szűkítsék egy sikeres azonosításhoz
- Az általuk megtalált személyeknek legyen olyan közös kapcsolódási pontjuk, mely miatt elképzelhető, hogy az arcukról készült *embedding*ek egy adatbázisba kerülnek

FEJLESZTÉSI LEHETŐSÉGEK

- A keresési fázist könnyű automatizálni
- Párhuzamosítani lehet a keresést több közösségi média között
- A demográfiai adatok prediktálásának a pontosítása

ÖSSZEFOGLALÁS