

# BIOM - zadanie č.3

Tomáš Jenčík, 111291

Máj 2024

## 1 Verifikácia tvári

Na začiatok zadania sme sa snažili verifikovať tváre. Modely sme použili na datasete normalizovaných tvári z predošlého zadania, kde príklad tvári vyzeral nasledovne:

frame\_45\_face0.jpg frame\_195\_face0.jpg frame\_55\_face0.jpg



frame\_10\_face0.jpg frame\_40\_face0.jpg frame\_120\_face0.jpg



frame\_40\_face0.jpg frame\_95\_face0.jpg frame\_25\_face0.jpg



Obraz č. 1: Normalizované snímky videí z datasetu.

Princípy, ktoré sme použili, boli DeepFace, čo je ľahký nástroj na rozpoznávanie tváre a analýzu atribútov tváre, pričom sme skúsili modely VGG-Face a Facenet512 a skončili s druhým z nich, keďže podľa dostupných informácií je presnejší. Ako druhý princíp sme použili model dlib metódou HOG.

## 2 DeepFace

Deepface je framework na rozpoznávanie tváre a analýzu atribútov tváre (vek, pohlavie, emócie a rasa) pre Python. Je to hybridný framework na rozpoznávanie tvári, ktorý obsahuje najmodernejšie modely: VGG-Face, FaceNet, OpenFace, DeepFace, DeepID, ArcFace, Dlib, SFace a GhostFaceNet.

Pri práci s našimi dátami sme začali pracovať s modelom VGG-Face, no po preskúmaní modelov sme zistili, že práve model Facenet, konkrétna verzia Facenet512. Facenet je 22-vrstvová hlboká neurónová sieť, ktorá priamo trénuje svoj výstup na 128-rozmerné vloženie. Stratová funkcia použitá v poslednej vrstve sa nazýva trojitá strata. FaceNet používa trojicu porovnávaním tváre s tvárou pomocou online metódy ľažby nových trojíc. Samozrejme, tento triplet pozostáva zo súboru obrázkov na zachytenie, kde každý obrázok pozostáva z pozitívnych a negatívnych obrázkov.

Na začiatok sme získali príznaky z normalizovaných framov, ktoré tvorili naše videá. Príznaky sme získavalí pomocou funkcie represent, vďaka ktorej sme získali vektorov. Príklad vektora vyzeral nasledovne:

Column1	Column2	Column3
Folder	Filename	Embedding
normalized_faces_mtcnn_cropped\Karin_Viard\frame_10\face0.jpg	{'embedding': [0.18921302258968353, 0.6513266563415527, 0.2766781449317932, ..., 0.6183100938796997], 'facial_area': {'x': 0, 'y': 0, 'w': 56, 'h': 92, 'left_eye': (50, 29), 'right_eye': (25, 32)}, 'face_confidence': 1.0}	

Obraz č. 2: Príznaky získané modelom Facenet512.

Tieto príznaky sme potom porovnávali na 2 pároch videí: jedna skupina párov, kde 2 videá obsahovali rovnakú osobu a druhá skupina párov, kde sme mali 2 videá, ktoré obsahovali rôzne osoby. Naše obrázky sme reprezentovali ako vektor, aby sme našli mieru podobnosti dvoch vektorov. Ak sú obidva obrázky rovnakej osoby, potom by mala byť miera malá. V opačnom prípade by mala byť miera veľká, ak sú dva obrázky rozdielne osoby. Podobnosť nášho modelu sa merala pomocou vzdialenosťi, pričom sme použili mierku kosínusovej podobnosti, pri ktorej čím menšiu hodnotu sme dostali, tým viac sa na seba podobali 2 obrazy. Zo získaných príznakov sme potom vybrali:

- náhodne vybranú kombináciu príznakov z páru
- priemernú hodnotu pre kombinácie z páru
- maximálnu hodnotu pre kombináciu príznakov z páru
- minimálnu hodnotu pre kombináciu príznakov z páru

Príklad takýchto hodnot na jednom z videí vyzeral nasledovne:

CSV1	CSV2	Similarities
aaa/embeddings/Karin_Viard_1	aaa/embeddings/Karin_Viard_3	[0.57067598 0.55156454 0.5407914, ..., 0.45910066]
		<input type="button" value="IsCorrect"/> <input type="button" value="Random"/> <input type="button" value="Max"/> <input type="button" value="Min"/> <input type="button" value="Average"/>
	TRUE	0.54890614 0.73221928 0.37050052 0.5501783295238095

Obraz č. 3: Nájdené hodnoty pre párs videí.

Taktiež sme si ešte vypočítali distance, podľa ktorej sme mohli určiť, či model správne (alebo nesprávne) určil páru. Skúšali sme viacero threshold hodnôt, no ostali sme na 0.5.

Folder1	Image1	Folder2	Image2	Result
<code>'verified': True</code>				
<code>'isCorrect': TRUE</code>				
				<code>'distance': 0.4756829875547285, 'threshold': 0.5, 'model': 'Facenet', 'detector_backend': 'dlib'</code>

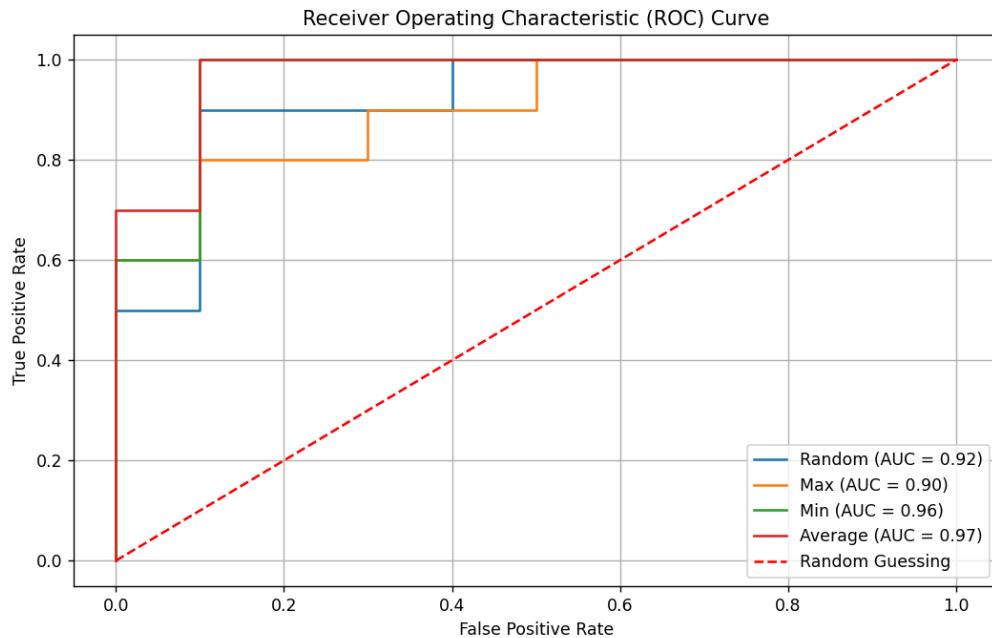
Obraz č. 4: Nájdená zhoda medzi videami.

Ako vidíme na obrázku č. 4, model hoci tesne, ale správne určil, že nájdený páru bol naozaj rovnaká osoba.

Pre dané hodnoty sme ďalej vytvorili ROC krivku. ROC krivka (Receiver Operating Characteristic) je graf, ktorý opisuje kvalitu binárneho klasifikátora v závislosti od nastavenia jeho klasifikačného prahu.

Binárny klasifikátor je zariadenie, ktoré odhaduje klasifikáciu niektorých objektov alebo prípadov do dvoch skupín, všeobecne označovaných ako pozitívne a negatívne. Pritom sa predpokladá, že objekty sú charakterizované nejakou premennou vyjadrenou číslom a že klasifikátor klasifikuje do pozitívnej skupiny práve tie, pre ktoré hodnota tejto premennej presahuje stanovený klasifikačný prah (ktorý je nejakým pevným číslom).

Naša ROC Krivka vyzerá nasledovne:



Obraz č. 5: Facenet512 ROC krivky.

## 2.1 HOG

Ako druhý model sme skúšili HOG. HOG je jednoduchý a výkonný deskriptor príznakov s lineárnym algoritmom strojového učenia SVM na detekciu tváre. Nepoužíva sa len na detekciu tváre, ale je široko využívaný aj na detekciu objektov, ako sú autá, domáce zvieratá a ovocie. HOG je robustný na detekciu objektov, pretože tvar objektu je charakterizovaný pomocou lokálneho rozloženia gradientu intenzity a smeru hrán. HOG architektúra je založená na 3 krokoch:

- Rozdeliť obraz na malé prepojené bunky.
- Vypočítať histogram pre každú bunku.
- Spojiť všetky histogramy dohromady a vytvoriť vektor príznakov, t. j. zo všetkých malých histogramov vytvoriť jeden histogram, ktorý je jedinečný pre každú tváru.

Príznaky pomocou HOG modelu sme vygenerovali v rovnakom formáte, ako Facenet512. Príklad pre vygenerované príznaky vyzeral nasledovne:

Column1	Column2	Column3
Folder	Filename	Embedding
normalized_faces_mtcnn_cropped/Karin_Viard_1	frame_0_face0.jpg	{'embedding': [0.3074967: 2651290894, 0.14757408201694489, 0.08268699049949646, ..., 0.006064385175704956]}

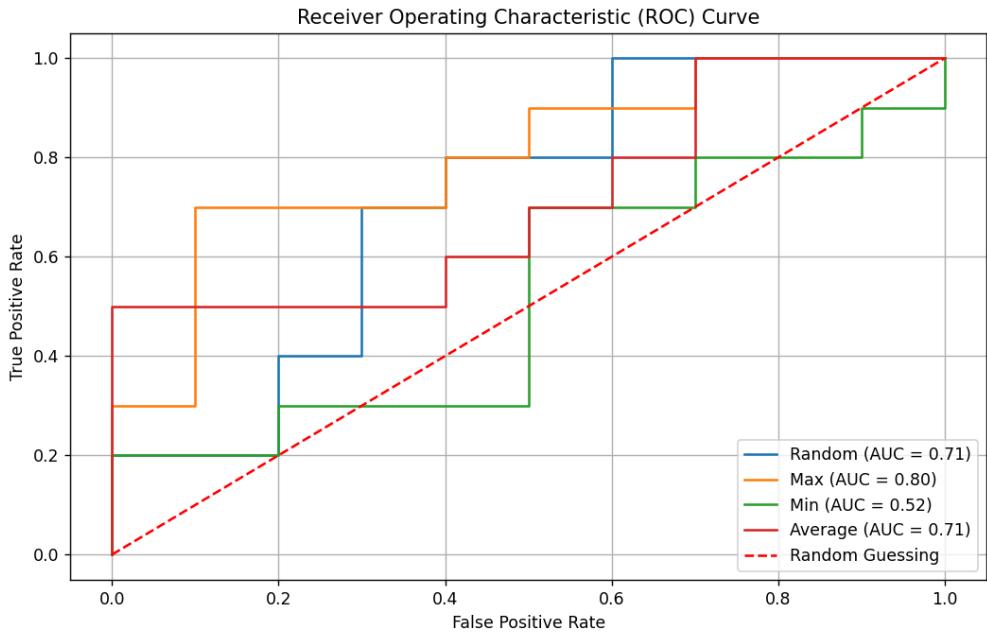
Obraz č. 6: Príznaky získané modelom HOG.

Taktiež sme si vygenerovali 4 hodnoty, ako v predošлом modeli, z ktorých sme potom robili ROC krivky. Príklad vygenerovaných hodnôt na jednom z videí vyzeral nasledovne:

CSV1	CSV2	Similarities
aaa/embeddings-HOG/Karin_Viard_1.csv	aaa/embeddings-HOG/Karin_Viard_3.csv	[0.61873955 0.59978266 0.611474 29, ..., 0.64616773]
IsCorrect	Random	Max Min Average Average Distance

Obraz č. 7: Nájdené hodnoty pre pári videí pomocou modelu HOG.

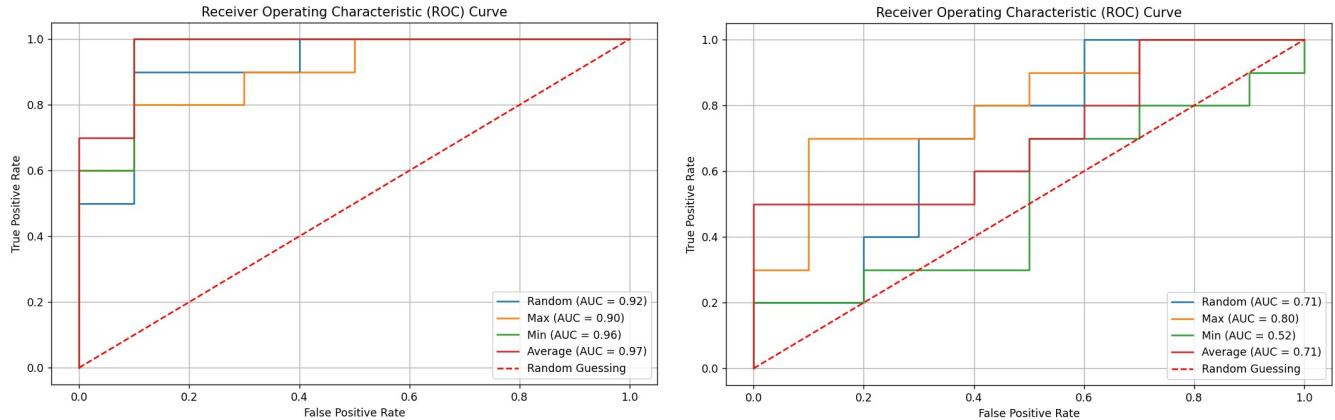
Pre všetky páry, ktoré sme porovnávali, sme taktiež na základe týchto hodnôt vygenerovali ROC krivky.



Obraz č. 8: HOG ROC krivky.

## 2.2 Porovnanie Facenet512 a HOG na ROC krivkách

Pomocou ROC Kriviek si teraz môžeme porovnať, ktorý model dosahoval lepšie výsledky.



Obraz č. 9: Facenet a HOG ROC krivky.

Na krivkách si môžeme všimnúť, že ako sa očakávalo, model Facenet512 dosahoval lepšie výsledky, ako model HOG, keďže Facenet je cnn model. Ďalsia vec čo treba taktiež poznamenať, že Facenet512

model má iný threshold ako HOG. Je to preto, lebo keď sme sa snažili nájsť ideálnu hodnotu threshold pre obe modely rôznym skúšaním, zistili sme, že model Facenet512 mohol mať nižšiu threshold hodnotu a teda presnejšie povedať, či je na páre rovnaká osoba alebo nie. Threshold hodnoty boli:

Tabuľka č. 1: Threshold hodnoty pre Facenet512 a HOG

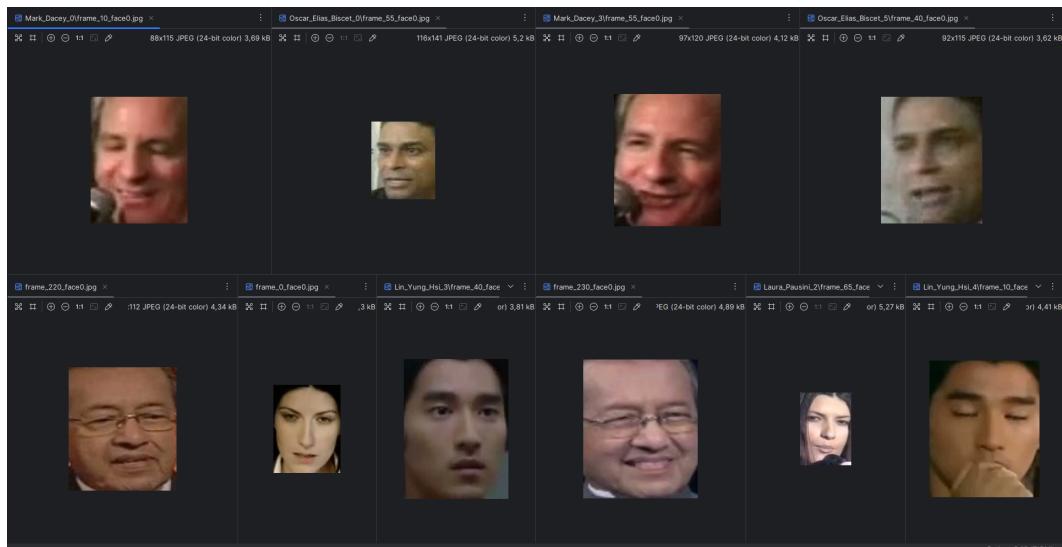
Metóda	Threshold
Facenet512	0.5
HOG	0.68

Taktiež vieme z kriviek povedať, že zatiaľko Facenet512 model určoval najlepšie pravdivosť páru v priemernej hodnote, čo znamená, že často správne určil hodnoty príznakov a taktiež aj v ostatných prípadoch správne určil často pári, tak na druhej strane vidíme, že model HOG pre minimálne hodnoty podobnosti nesprávne určil podobnosť a najlepšie výsledky mal vtedy, keď zobrajal maximálne hodnoty z podobnosti.

## 2.3 Analýza najlepších a najhorších výsledkov pre páry

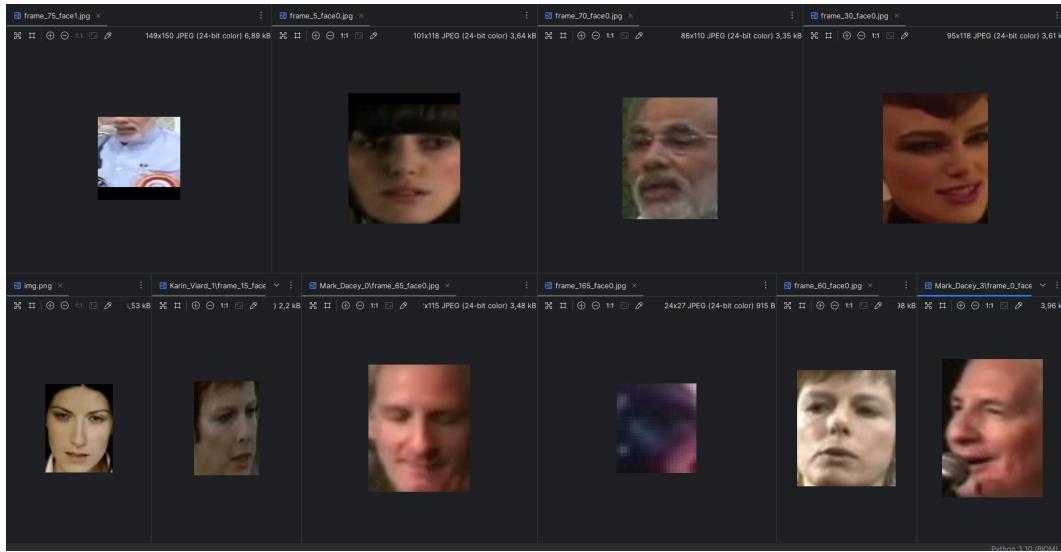
Pozrieme sa na 4 prípady, kde nás model najlepšie a najhoršie určil podobnosť:

- 5 pravých párov s najvyššou podobnosťou
- 5 pravých párov s najnižšou podobnosťou
- 5 odlišných párov s najvyššou podobnosťou
- 5 odlišných párov s najnižšou podobnosťou



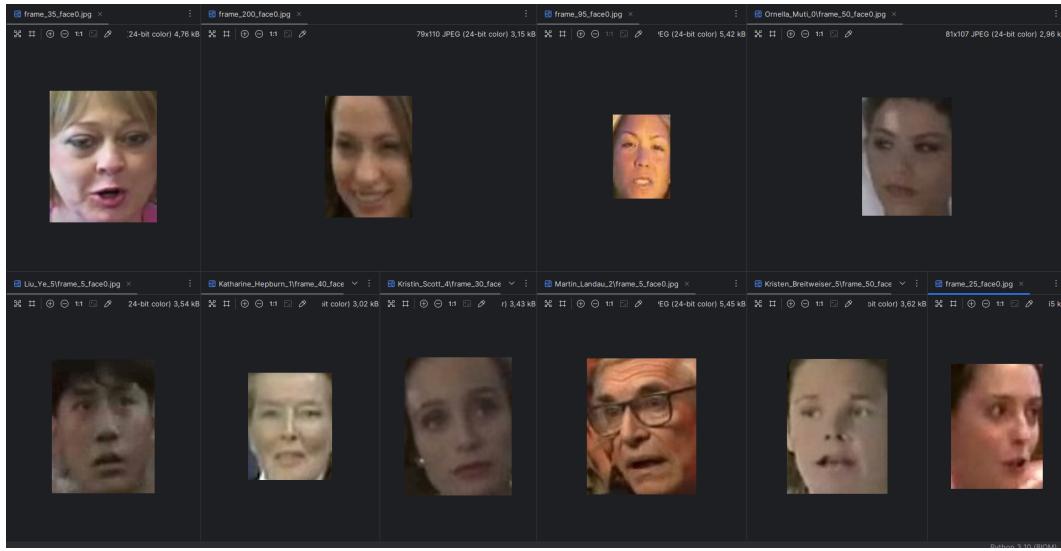
Obraz č. 10: 5 pravých párov s najvyššou podobnosťou.

Ako vidíme, model našiel vysokú podobnosť v týchto videách najmä kvôli tomu, že videá sú takmer identické (osoby sa pozerajú rovnakým smerom, majú objekt pred sebou na oboch videách). Jediný prípad, kde model prekvapivo našiel vysokú pravdepodobnosť bolo na obraze vpravo dole. Je to pravdepodobne z toho dôvodu, že osoba má výraznejšie črty tváre kvôli inej rase a jeho veku.



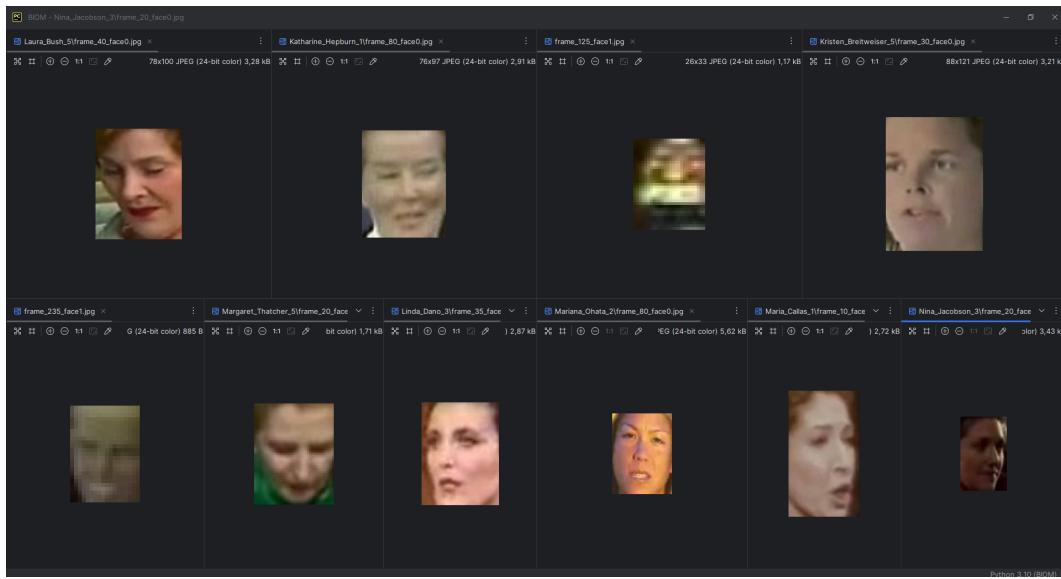
Obraz č. 11: 5 pravých párov s najnižšou podobnosťou.

V tomto prípade môžeme vidieť páry, kde hoci boli na videách rovnaké osoby, podobnosť nebola najvyššia. Je to z toho dôvodu, že model pri normalizácii snímok našiel iné objekty alebo iné osoby, a teda porovnával tvár s niečim, čo nebola rovnaká tvár. Ďalšie problémy, ktoré vznikli pri porovnávaní nastali kvôli tomu, že sa osoba pozerala do inej strany ako mala alebo úplne zmenila svoj výzor, ako napríklad môžeme vidieť na obraze vpravo hore.



Obraz č. 12: 5 odlišných párov s najvyššou podobnosťou.

Teraz sa pozrieme na 5 párov, kde model našiel vysokú pravdepodobnosť, hoci nemal, keďže sa jednalo o páry s odlišnými osobami. Ako vidíme, podobnosť bola nájdená buď kvôli rovnakému výrazu tváre, alebo kvôli kvalite fotky (fotka mala podobnú RGB hodnotu ako jej pár).



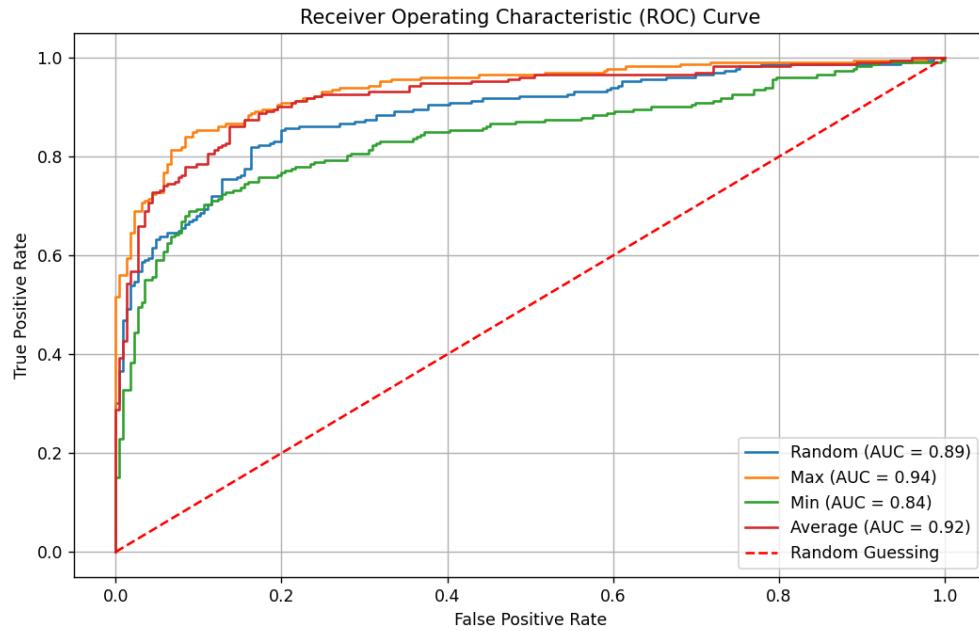
Obraz č. 13: 5 odlišných párov s najnižšou podobnosťou.

Ako pri pravých pároch, aj tu model našiel nízku podobnosť ak sa jednalo o obrazy, kde sa

nenachádzala hľadaná tvár. Taktiež tak ako v predošom prípade, nízka podobnosť bola medzi pármí, kde sa osoby pozerali do rôznych uhlov alebo sa rôzne tváriili.

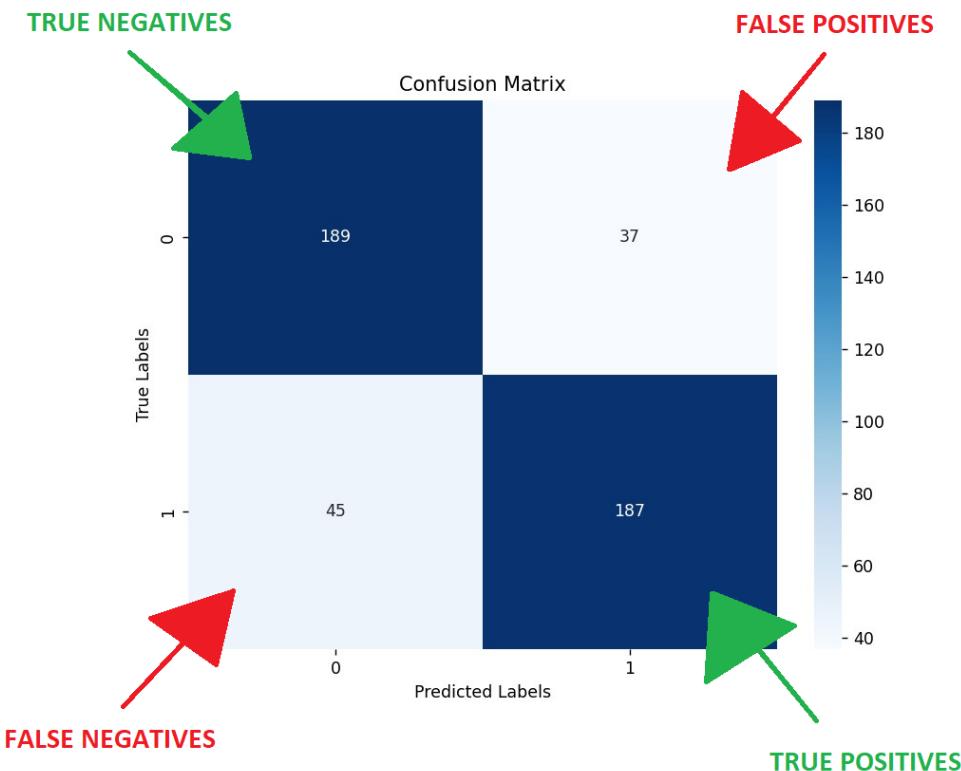
## 2.4 Facenet512 pre všetky páry

Našou ďalšou úlohou vyskúšať model na celom datasete párov. Použili sme model Facenet512, keďže sme s ním dosahovali lepšie výsledky pre určovanie zhody.



Obraz č. 14: ROC krivky pre model Facenet512 na celom datasete.

Ako môžeme vidieť, krvky na väčšom datasete vyzerajú ináč, ale znázorňujú to isté. V tomto prípade najlepšie výsledky zhody sme dostali, ak sme zobraťi maximálne hodnoty zhody, naopak najhoršie (stále dobré) výsledky boli pre zhody porovnávané s minimálnymi hodnotami zhody. Taktiež sme si zobrazili Konfúznu maticu, aby sme vedeli povedať, ako dobre model správne určil, čo je pravdivé a čo nie.



Obraz č. 15: Konfúzna matica pre všetky páry.

Ako môžeme vidieť na obrázku č. 15, náš model vo veľkom množstve správne určil, ktorý pári má byť zhodný a ktorý nie. Tieto výsledky sme dostali pre threshold 0.4. Model pri nižšej hodnote thresholdu mal väčšie hodnoty FP, zatiaľčo pri vyššej hodnote dosahoval väčšie čísla FN.

Výsledky sme ešte vyhodnotili na metrikách Precision, Recall a F1-skóre, a tieto hodnoty pre threshold 0.4 boli:

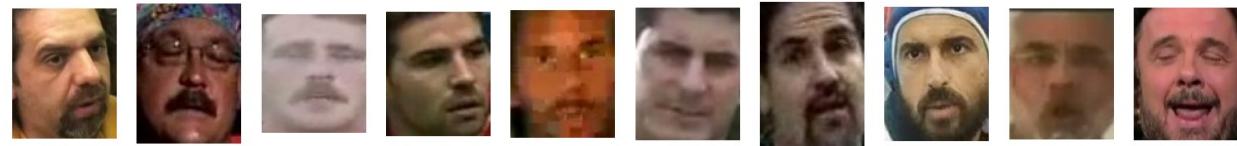
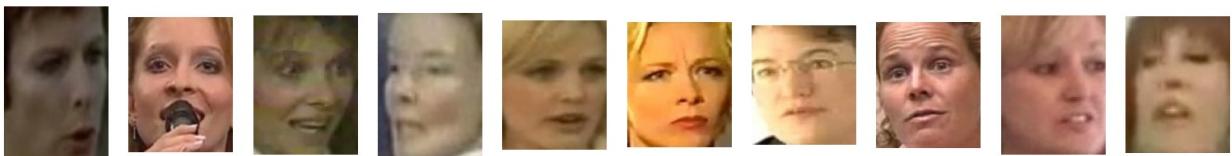
Metrika	Hodnota
Precision	0.8348
Recall	0.8060
F1-score	0.8202

Tabuľka č. 2: Výsledky pre celý dataset.

### 3 Zhluky

Ako poslednú časť sme hľadali zhluky medzi obrazmi. Snažili sme sa nájsť 10 obrazov v datasete, na ktorých by sa dali ukázať nejaké podobnosti, na základe ktorých model našiel zhodu. Nájdené

4 zhluky boli:



Obraz č. 16: 4 nájdené zhluky.

Na nasledovných zhlukoch si môžme všimnúť, že model ako prvé našiel chlapov, potom ženy, potom hľadal iné príznaky a našiel ľudí, ktorí nosia okuliare a nakoniec našiel mužov, kde veľké množstvo z nich malí fúzy alebo bradu.

## 4 BONUS

Ako bonus sme zistili, ako sa tvária dané osoby a akej rasy podľa modelu sú. Treba poznamenať, že model hľadal horšie emócie pre normalizované fotky, ako doterajšie výsledky, keď sme hľadali príznaky ako vektoru. Príklad vyzeral nasledovne:



Obraz č. 17: Fotka Liu Ye.

Listing 1: Data

```
{  
    'emotion': {  
        'angry': 55.728250059097995,  
        'disgust': 0.18920805378013208,  
        'fear': 6.427607664155047,  
        'happy': 0.0315535659539919,  
        'sad': 32.864805118464815,  
        'surprise': 0.14992734027029989,  
        'neutral': 4.608636917634378  
    },  
    'dominant_emotion': 'angry',  
    'region': {  
        'x': 73,  
        'y': 67,  
        'w': 85,  
        'h': 106,  
        'left_eye': (137, 112),  
        'right_eye': (99, 111)  
    },  
    'face_confidence': 1.0,  
    'age': 25,  
    'gender': {  
        'Woman': 0.7773519493639469,  
        'Man': 99.22264814376831  
    },
```

```

'dominant_gender': 'Man',
'race': {
    'asian': 79.56084256638974,
    'indian': 5.726312163315754,
    'black': 1.205430408664893,
    'white': 2.3076507305919254,
    'middle-eastern': 0.3863551426181027,
    'latino-hispanic': 10.81339662511389
},
'dominant_race': 'asian'
}

```



Obraz č. 18: Fotka Martin Luther King.

Listing 2: Data

```

{
    'emotion': {
        'angry': 96.97123747892603,
        'disgust': 6.038913302987769e-05,
        'fear': 2.582216230464949,
        'happy': 0.002563147580056866,
        'sad': 0.43137406707549497,
        'surprise': 0.00022245864291021427,
        'neutral': 0.012329090539423557
    },
    'dominant_emotion': 'angry',
    'region': {
        'x': 155,
        'y': 34,
        'w': 55,
        'h': 73,
        'left_eye': (184, 61),
        'right_eye': (161, 60)
    },
    'face_confidence': 1.0,
    'age': 30,
}

```

```

'gender': {
    'Woman': 0.07266912143677473,
    'Man': 99.92733597755432
},
'dominant-gender': 'Man',
'race': {
    'asian': 13.627302646636963,
    'indian': 22.30982333421707,
    'black': 29.935255646705627,
    'white': 8.950115740299225,
    'middle-eastern': 7.012484222650528,
    'latino-hispanic': 18.165017664432526
},
'dominant_race': 'black'
}

```



Obraz č. 19: Fotka Mary Matalin.

Listing 3: Data

```

{
    'emotion': {
        'angry': 9.1596243774667,
        'disgust': 0.05468249763332293,
        'fear': 1.4868971846638432,
        'happy': 0.6327722598774588,
        'sad': 78.90856140306958,
        'surprise': 0.17259464230952515,
        'neutral': 9.584867035440626
    }
}

```

```
},
'dominant_emotion': 'sad',
'region': {
    'x': 29,
    'y': 37,
    'w': 37,
    'h': 48,
    'left_eye': (57, 53),
    'right_eye': (39, 56)
},
'face_confidence': 1.0,
'age': 38,
'gender': {
    'Woman': 79.52503561973572,
    'Man': 20.474962890148163
},
'dominant_gender': 'Woman',
'race': {
    'asian': 1.358839205434275,
    'indian': 0.3173531455086556,
    'black': 0.03143512190957349,
    'white': 84.68729166312764,
    'middle-eastern': 8.257622018910768,
    'latino-hispanic': 5.347460082021956
},
'dominant_race': 'white'
}
```