MEDIAPIPE:
Facial Expression Prediction

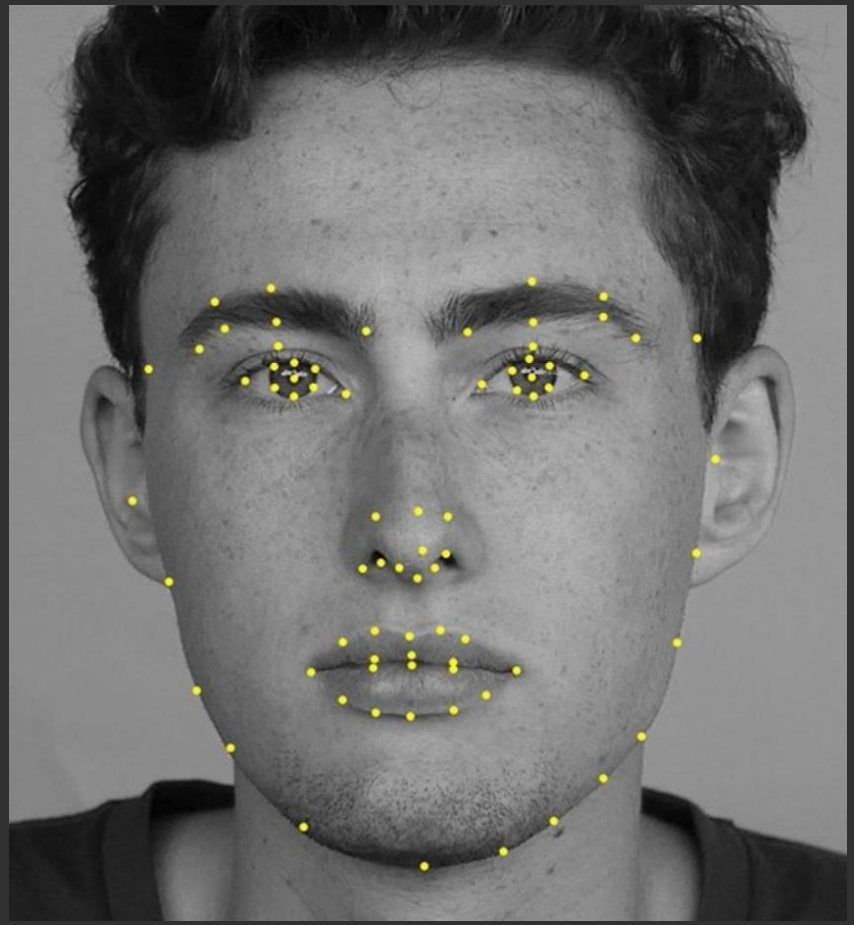
Fondamenti di Visione Artificiale e Biometria

Luigi Vollono - Orazio Cesarano – Davide Alfieri – Vincenzo Sabato

Biometrica utilizzata: IL VOLTO

Face detection:
Rilevamento effettivo del volto in una immagine (tramite landmark).

Face recognition:
Una volta individuato il volto, si procede alle estrazioni delle caratteristiche per poterle elaborare.



Emozioni attraverso espressioni

Un’**emozione** è processo interiore suscitato da un evento/stimolo rilevante per gli interessi dell'individuo.

Le **espressioni facciali** si riferiscono ai movimenti della muscolatura mimetica del volto.

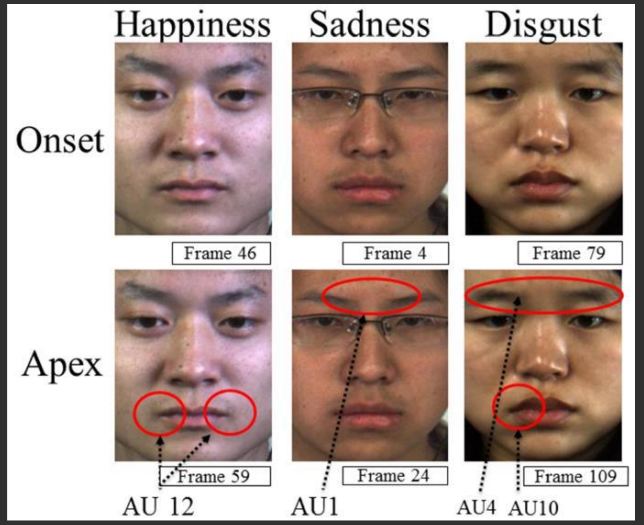
Paul Ekman ha elaborato per primo un modello scientifico per interpretare le emozioni correlate alle espressioni facciali.

Classificare le espressioni

Le espressioni rilevabili allo stato dell'arte sono 8.

Le espressioni facciali si dividono in due categorie:

- **Macro espressioni**, sono espressioni facciali di massima intensità;
- **Micro espressioni,** sono caratterizzate da breve durata e bassa intensità.



Predire le espressioni

L’**obiettivo** di questo progetto è stato quello di fornire uno stato di avanzamento di una espressione sotto forma di percentuale in maniera automatizzata, per le 8 espressioni proposte, a partire da una neutrale.

Le **features** utilizzate, per predire la percentuale di espressione, sono stati i landmark ottenuti mediante mediapipe.

Il **Dataset** utilizzato è noto in letteratura come Cohn Kanade Espression Dataset (CK+).

Predire le espressioni – Passaggi

Preparazione dataset di caratteristiche

1. Estrarre per ogni immagine del dataset i 468 landmark mediante mediapipe;
2. Calcolo delle distanze locali e globali dei 468 landmarks per ogni frame;
3. Etichettature delle distanze con un valore identificativo della classe e della percentuale di espressione;

Predizione dell’espressione

4. Classificare tramite delle classi di percentuale le espressioni. Il problema diventa quindi una classificazione multi-classe, dove le percentuali più basse individuano la micro espressione e le più alte la macro espressione.

FASE 1: Estrazione dei 468 landmarks

Per eseguire questa operazione, abbiamo utilizzato **MediaPipe Face Mesh**.

Il modello restituisce le **posizioni dei punti 3D**, nonché la probabilità che un volto sia presente e ragionevolmente allineato nell’input.

FASE 2: Calcolo delle distanze locali e globali

Il successivo passo è stato calcolare le distanze locali e globali sulla base dei landamaks estratti per ogni frame.

Per il calcolo abbiamo applicato la **distanza euclidea**:

$$\sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}$$

FASE 2: Calcolo delle distanze locali e globali

- Per le **distanze locali**, la funzione calcola la distanza tra il primo frame e il secondo frame, poi la distanza tra il secondo e il terzo frame, e così via.

- Per le **distanze globali,** la funzione calcola la distanza tenendo conto sempre del primo frame, quindi, calcola la distanza tra il primo frame e il secondo frame, poi la distanza tra il primo e il terzo frame, e così via.

FASE 3: Etichettature delle distanze

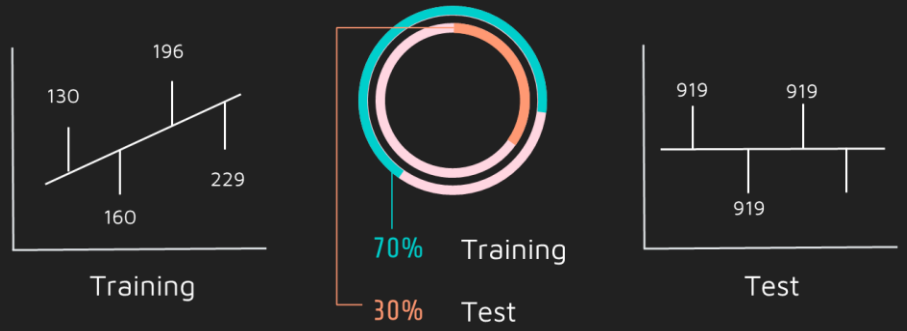
Abbiamo etichettato ogni frame con un valore che è identificativo della classe e della percentuale di espressione.

Abbiamo considerato almeno tre classi per ogni emozione:

- Neutrale;
- Classe 0-33%;
- Classe 33-66%;
- Classe 66-100%.

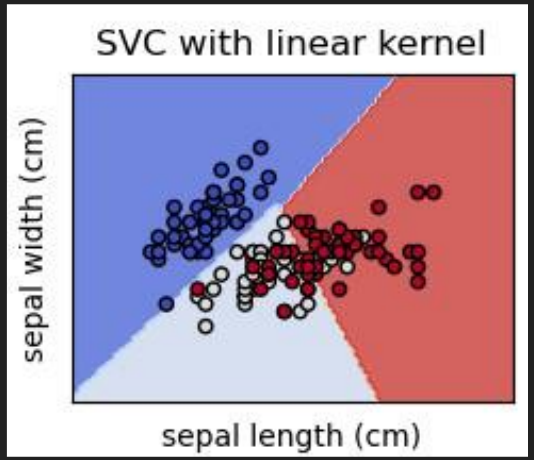
	A	B	C
1	S005_001	0	0
2	S005_001	0	0
3	S005_001	0	0
4	S005_001	3	1
5	S005_001	3	1
6	S005_001	3	1
7	S005_001	3	2
8	S005_001	3	2
9	S005_001	3	2
10	S005_001	3	3
11	S005_001	3	3
12	S010_001	0	0
13	S010_001	0	0

FASE 4: Addestramento e Multi-classificazione



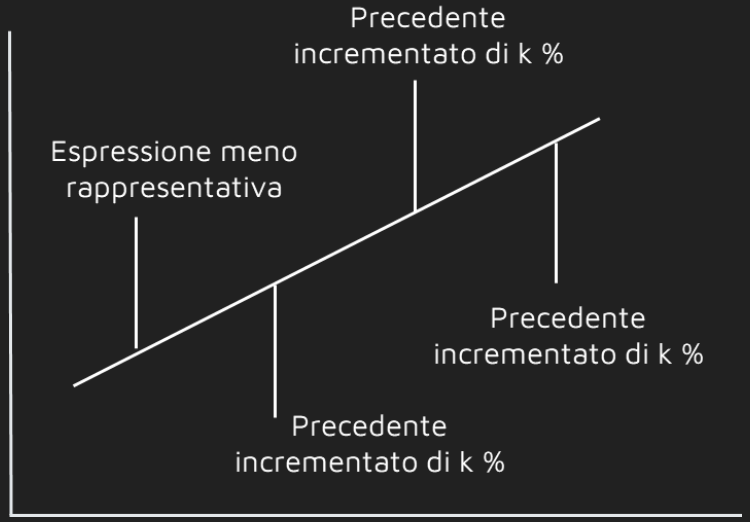
Per il risolvere il problema delle multi-classi abbiamo utilizzato il classificatore **SVC** (**Support Vector Classification**) che è in grado si eseguire la classificazione binaria e multi-classe su un dataset.

Nello specifico abbiamo fatto uso di un **SVC con kernel lineare**.



FASE 4: Sampling

Obiettivo è stato quello di cercare di bilanciare il dataset, cioè ridurre il numero di istanze dell'espressioni più frequenti, scartando alcune di esse, ed aumentare il numero dell'emozioni meno rappresentativi.



RIDUZIONE: Approccio incrementale

L'approccio utilizzato per ridurre le istanze di emozioni più frequenti si basa sullo scegliere quante istanze utilizzare.
Mentre per l'emozione meno rappresentativa utilizziamo tutte le istanze per quelle più frequenti ne utilizziamo una parte.

FASE 4: Sampling

Data Augmentation: SMOTE

Crea nuovi dati a partire da quelli già esistenti nel dataset, per aumentare il numero di espressioni meno frequenti.
Per non avere misure falsate, risulta necessario effettuare l'augmentation solo sui dati di training e non sulla totalità del dataset.

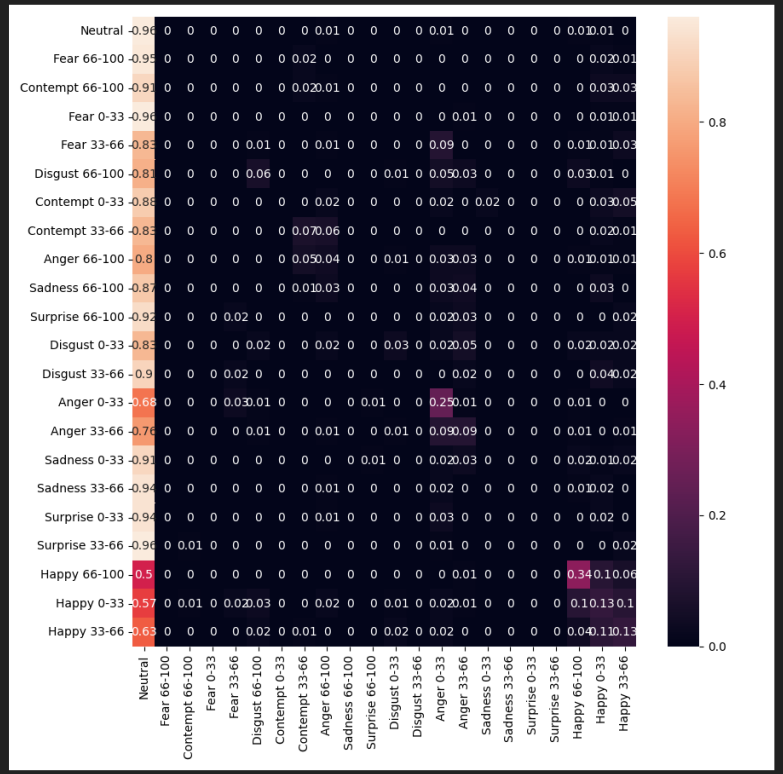
Per l'addestramento del dataset e multi-classificazione sono state applicate diverse tecniche:

1. Addestramento dataset locale e svc;
2. Addestramento dataset globale e svc;
3. Addestramento dataset locale con data augmentation;
4. Addestramento dataset globale con data augmentation;
5. Addestramento dataset locale con pesatura senza data augmentation;
6. Addestramento dataset globale con pesatura senza data augmentation;
7. Addestramento dataset globale con pesatura e data augmentation.

RISULTATI – 1° addestramento (locale)

1. La rete non si comporta bene con il dataset delle distanze locali, poiché i valori delle espressioni di due frame consecutivi o vicini non sono molto diversi.

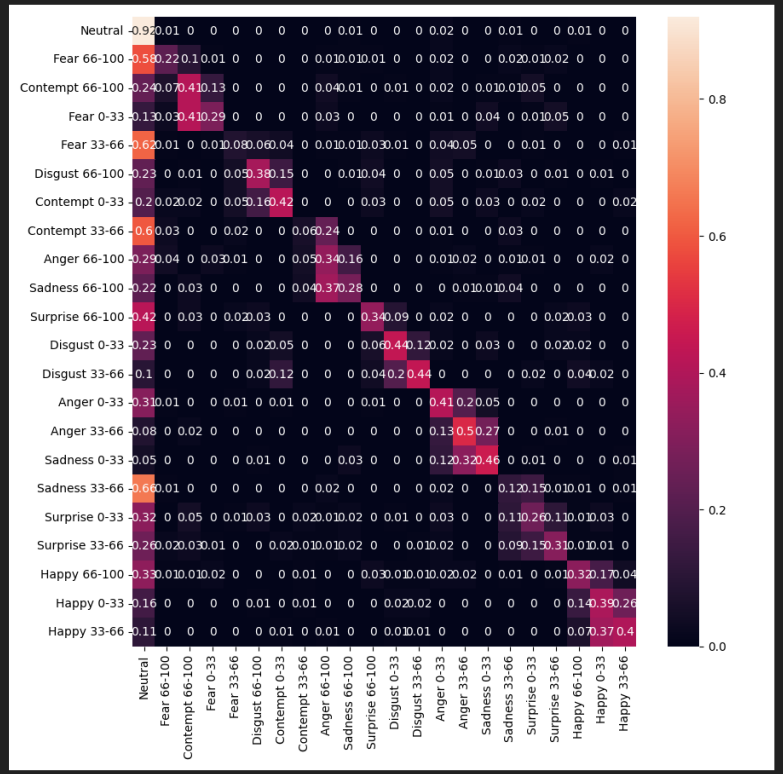
Notiamo che le percentuali sono molto basse e la diagonale della classificazione è poco definita.



RISULTATI – 2° addestramento (globale)

2. La rete si comporta molto meglio dato che le distanze sono calcolate sempre confrontando un dato frame con il primo della sequenza.

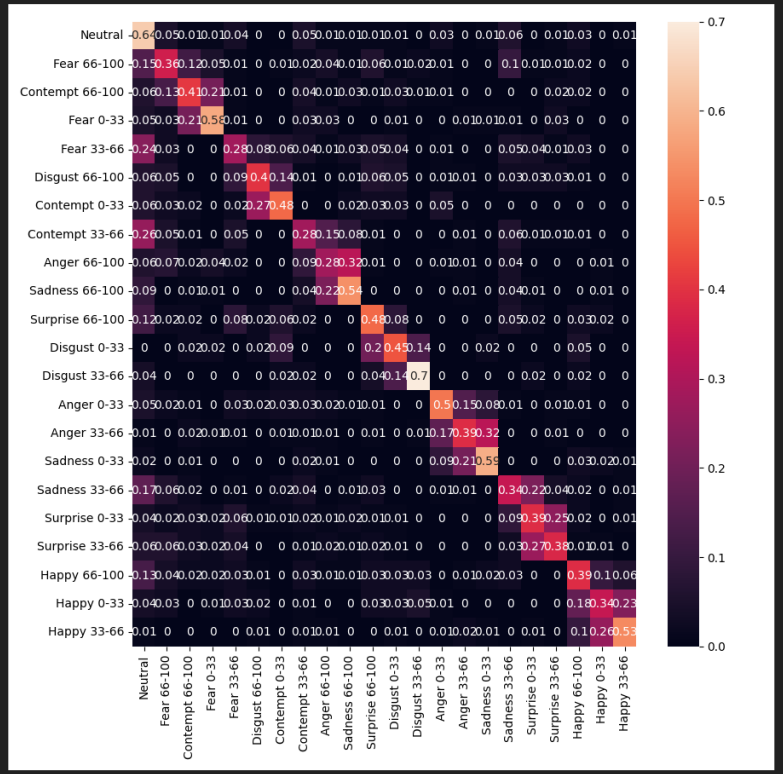
I cambiamenti delle emozioni sono molto più evidenti a livello numerico, con una conseguente migliore classificazione e una diagonale ben definita.



RISULTATI – 3° e 4° addestramento (augumentation)

3. Il risultato non è molto diverso dal primo addestramento.

4. Applicando l'augumentation, la rete migliora leggermente rispetto alla rete addestrata sul dataset globale.

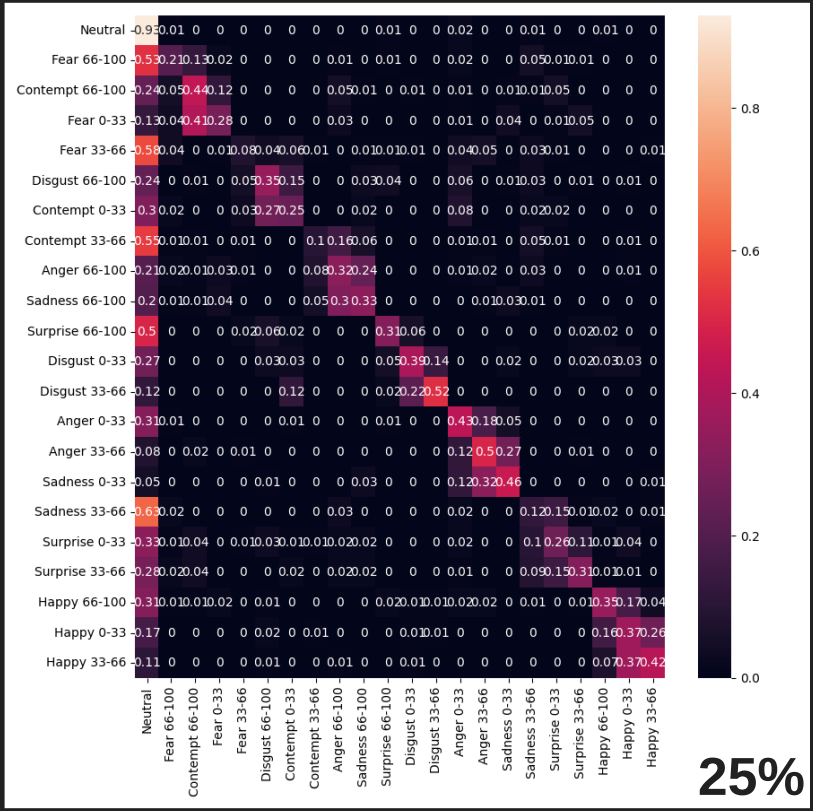


RISULTATI – 5° e 6° addestramento (pesatura)

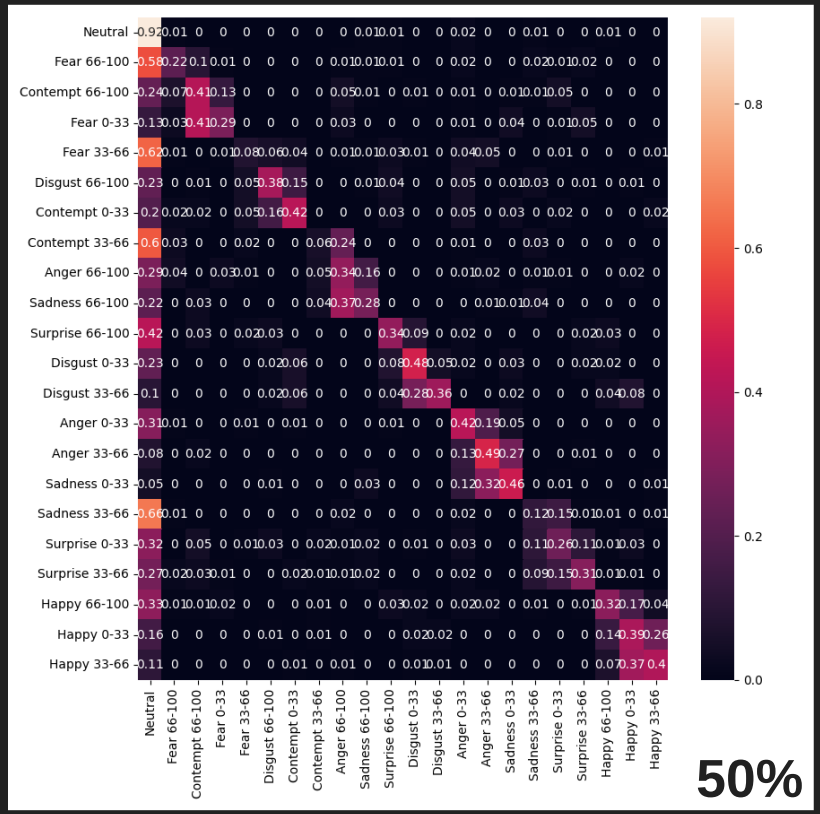
5. Non ha dato nessun effetto positivo, applicando la tecnica di pesatura sul dataset locale.

6. Applicando la tecnica di pesatura ha prodotto buoni risultati (grafici nelle prossime slide).

RISULTATI – 6° addestramento (pesatura)

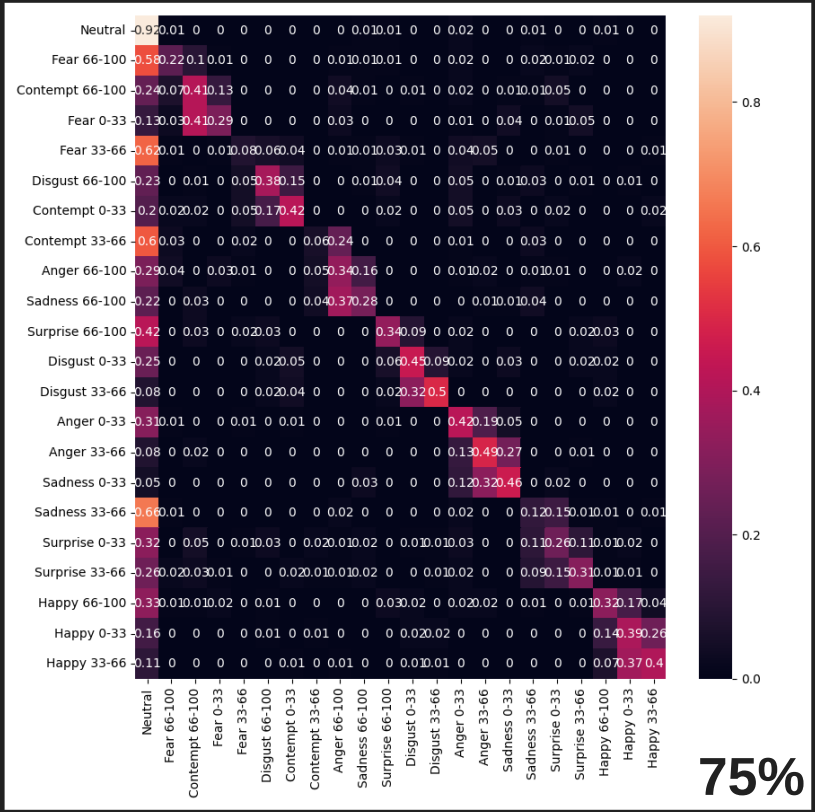


25%

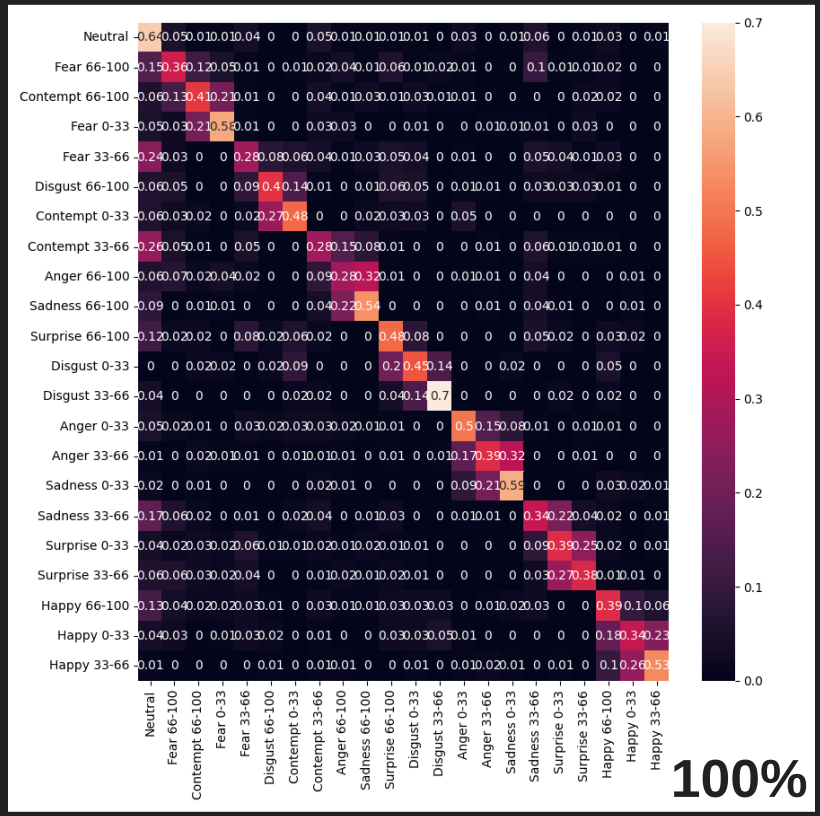


50%

RISULTATI – 6° addestramento (pesatura)



75%

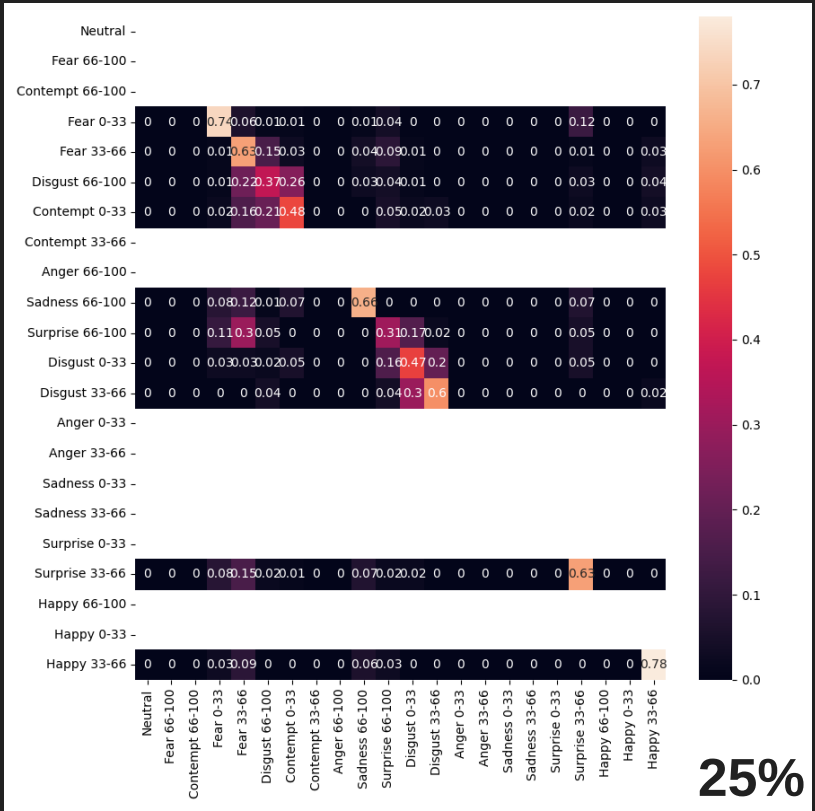


100%

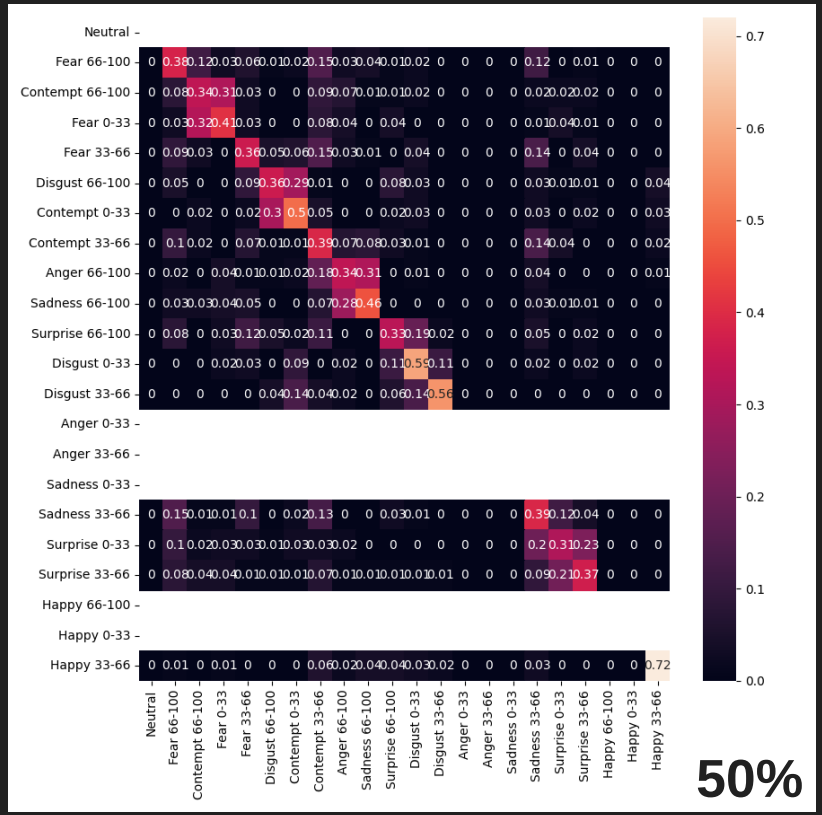
RISULTATI – 7° addestramento (pesatura e augum.)

7. Fatto solo sulle distanze globali, dato che la rete sulle distanze locali si comporta in tutti i casi sempre male, combinando la pesatura e l’augumentation, abbiamo i seguenti risultati:

RISULTATI – 7° addestramento (pesatura e augum.)

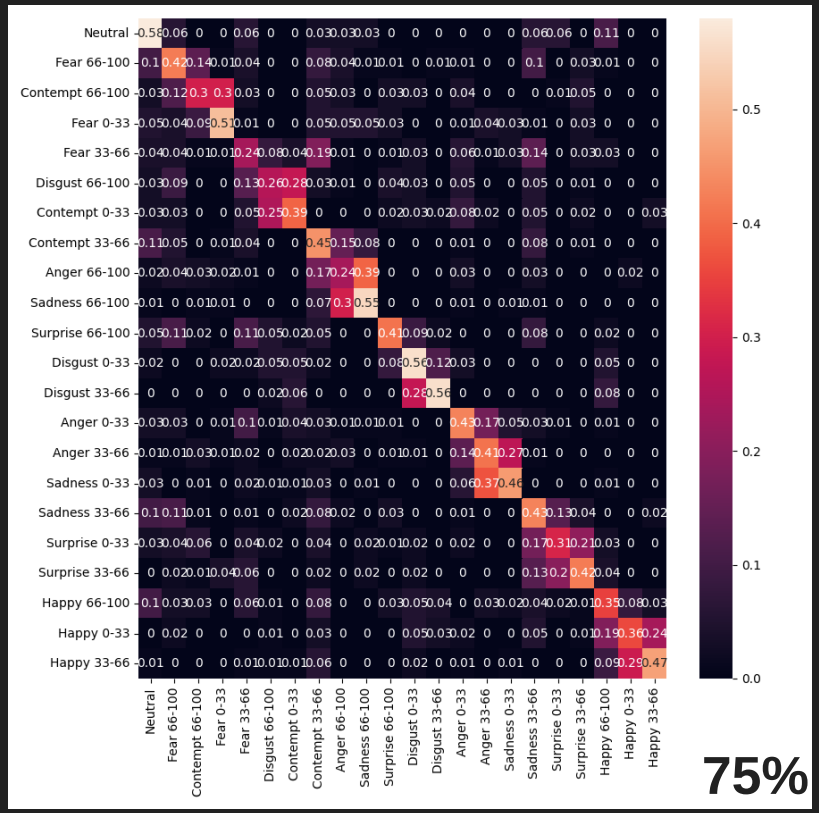


25%

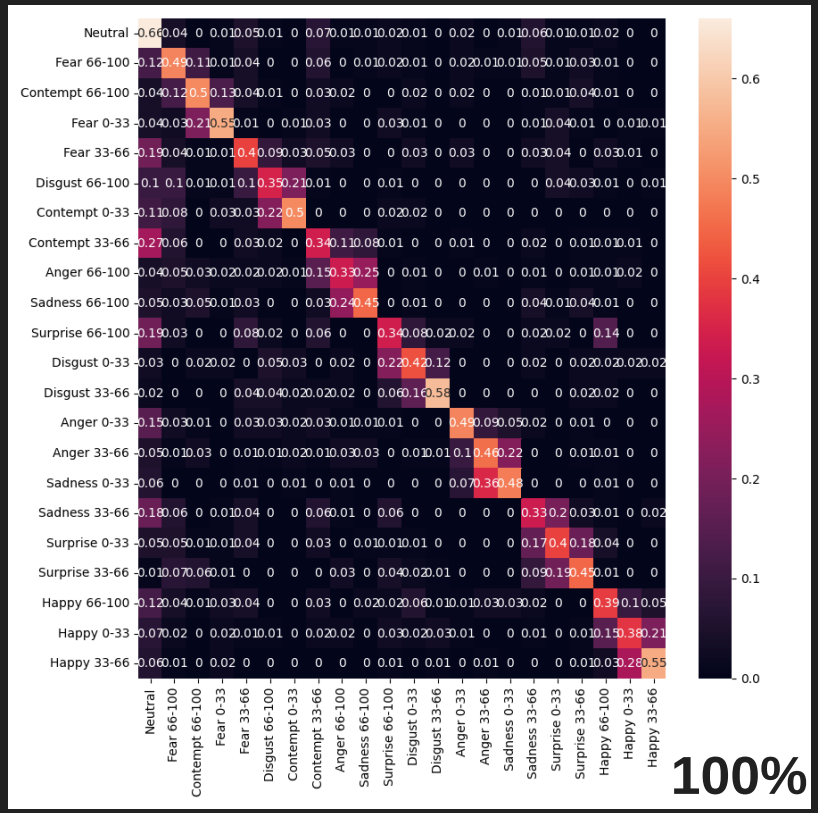


50%

RISULTATI – 7° addestramento (pesatura e augum.)

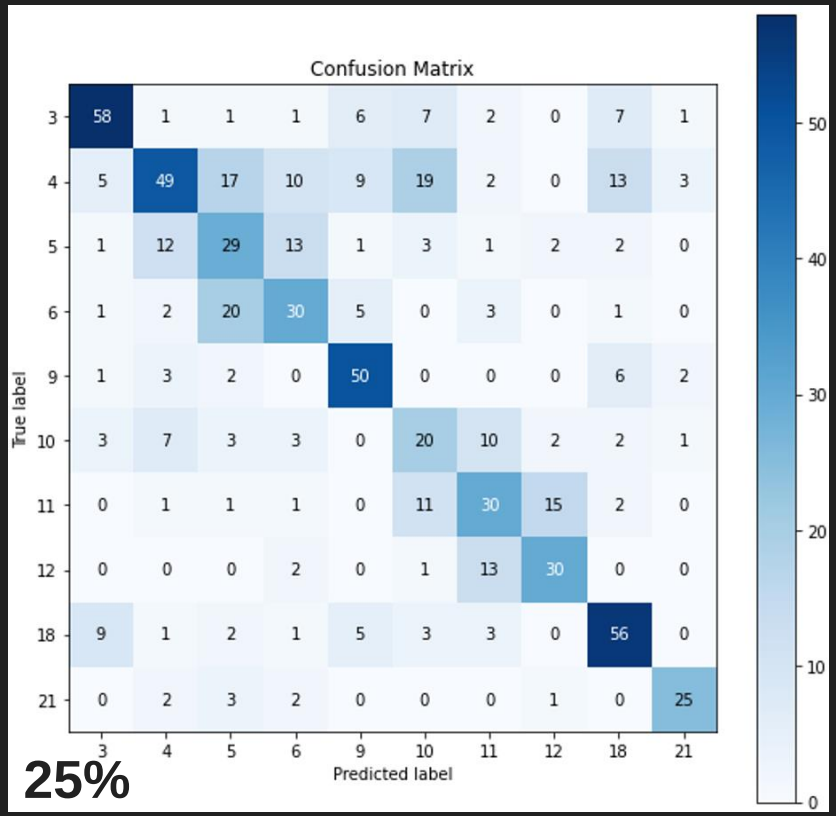


75%

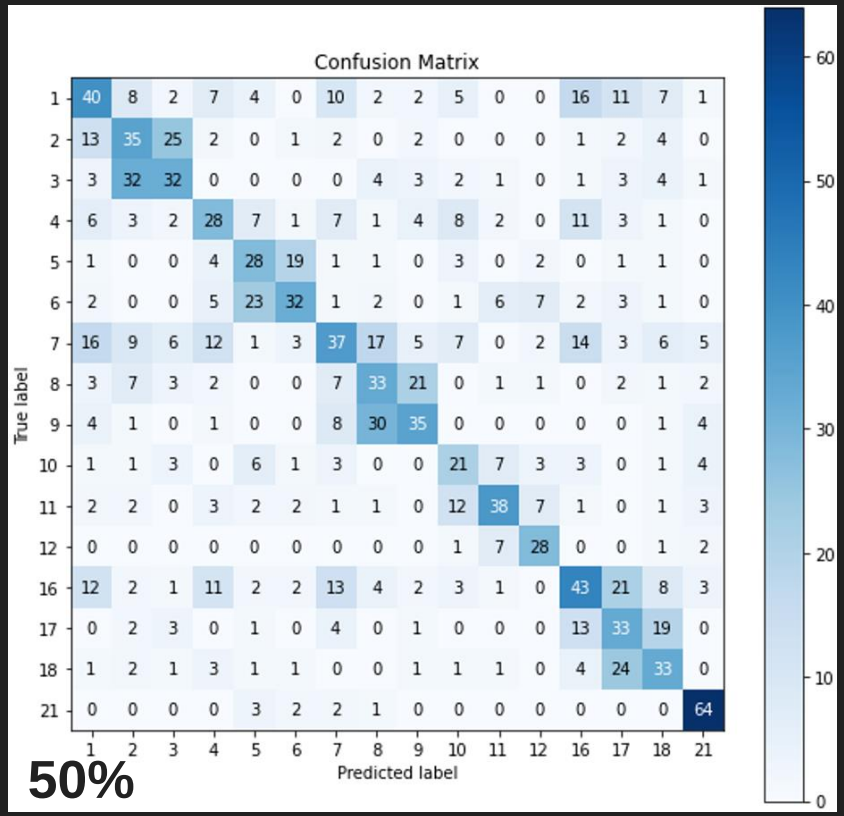


100%

RISULTATI – 7° addestramento (matrice di confusione)

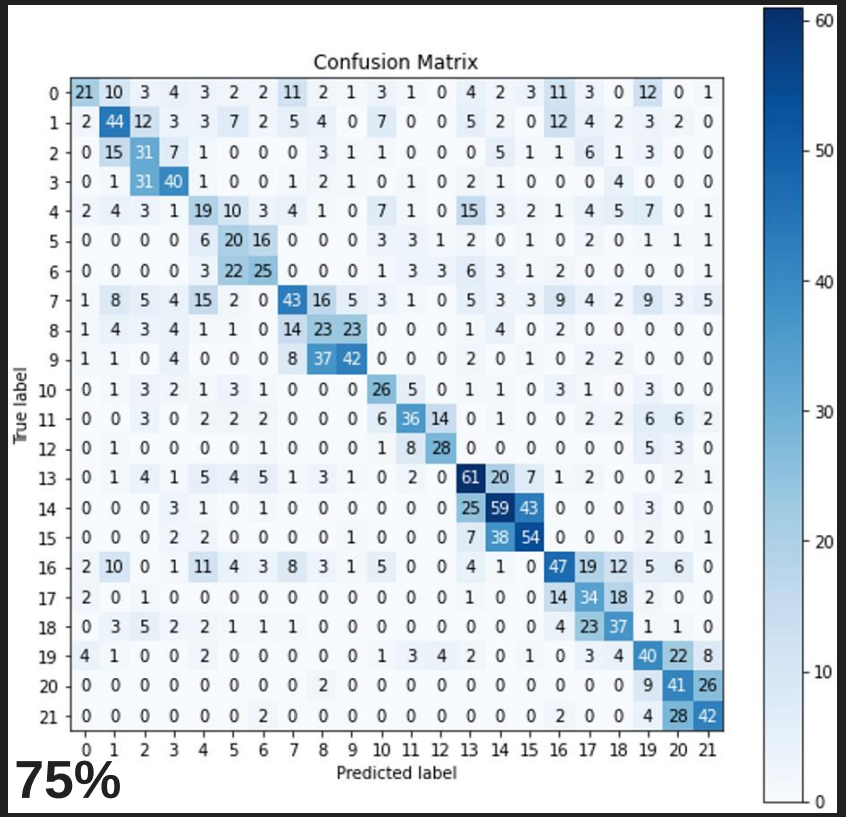


25%

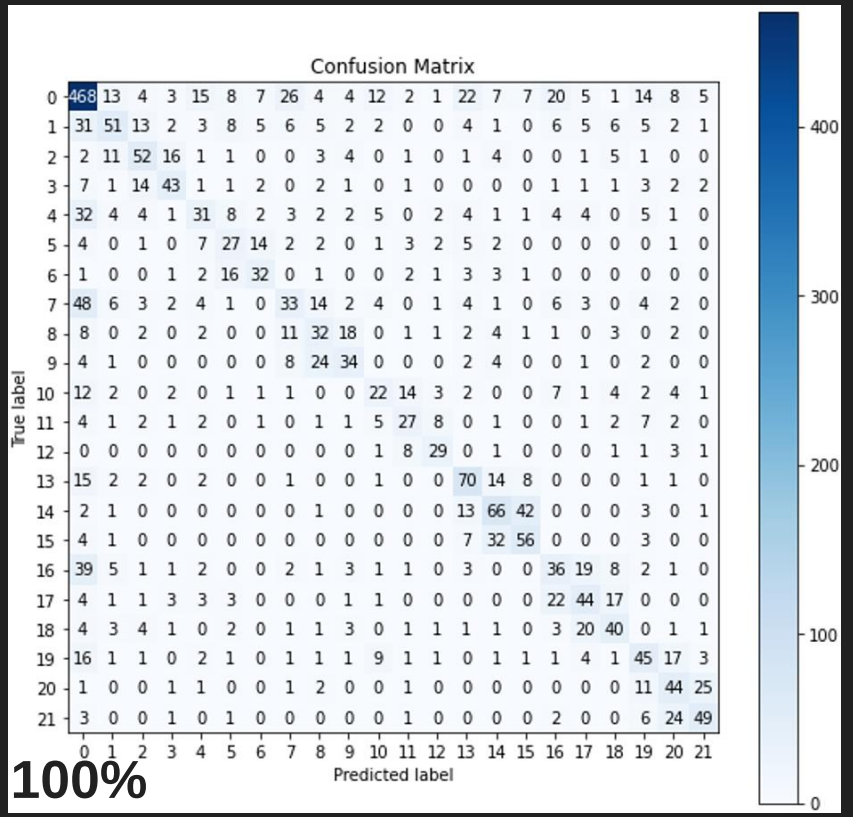


50%

RISULTATI – 7° addestramento (matrice di confusione)



75%



100%

CONCLUSIONI

Sulla base delle accuratezze totali calcolate, possiamo dire che l'addestramento che ha prodotto il miglior risultato è quello fatto sul dataset globale tramite l'augumentation e pesatura 25%, l'accuratezza totale è del 56%.

A seguire, l'addestramento migliore è stato quello della pesatura sul dataset globale con augumentation, fatto sul dataset globale 100%, dove l'accuratezza totale è del 49%.

Addestramento	Score_Accuracy
Addestramento dataset locale con svc	30%
Addestramento dataset globale con svc	48%
Addestramento dataset locale con data augmentation	32%
Addestramento dataset globale con data augmentation	48%
Addestramento dataset locale con pesatura senza data augmentation	-
Addestramento dataset globale con pesatura senza data augmentation	48% (pesatura 25%), 48% (pesatura 50%) 48% (pesatura 75%), 48% (pesatura 100%)
Addestramento dataset globale con pesatura e data augmentation	**56% (pesatura 25%)**, 41% (pesatura 50%) 40% (pesatura 75%), 49% (pesatura 100%)

SVILUPPI FUTURI

Obiettivi futuri:

- Raffinare la fase di addestramento del classificatore, applicando nuove tecniche, raggiungendo nuovi risultati che siano potenzialmente migliori rispetto a quelli ottenuti finora.

- Tecniche di undersampling e decision threshold.

GRAZIE PER L’ATTENZIONE