

Valutazione dell'estetica di immagini di cibo

Relatore:

Prof. Paolo Napoletano

Correlatore:

Prof. Gianluigi Ciocca

Candidato:

Sofia Damaso

Matricola:

845189

Anno Accademico 2020-2021

Introduzione al problema

L'Estetica:

- Caratteristica fortemente soggettiva
- Utile per valutazioni automatiche delle immagini
- Aumento dei dataset di immagini per lo studio di essa

L'Estetica e i cibi:

- Attuale soprattutto nell'ambito dei Social Network e del marketing
- Il buono viene legato al bello, analogamente all'effetto Alone¹
- Mangiare inizia con la vista²

Immagini considerate esteticamente belle



Immagini non considerate esteticamente belle



¹ IL PASSE-PARTOUT DEI BELLI. <<6-Psicologia della bellezza>>

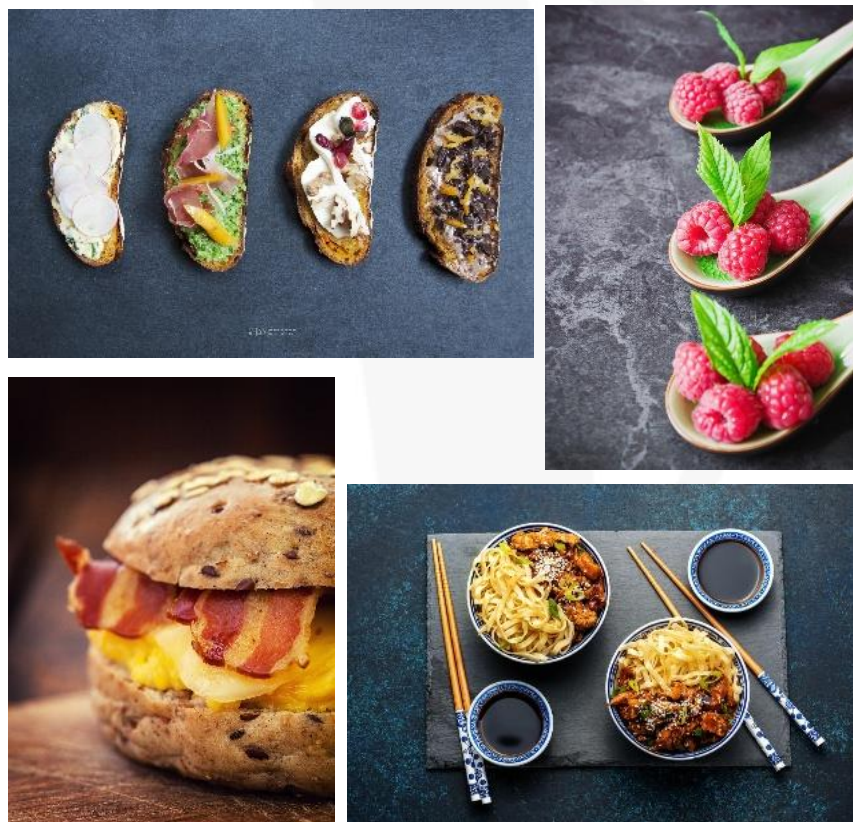
² Raffaele Campo, Giuseppe Loporcaro e Fabrizio Baldassarre. <<The effects of food aesthetics on consumers. Visual stimuli and food marketing>>. In: DIEM: Dubrovnik International Economic Meeting. Vol. 3. 1.

Sveučilište u Dubrovniku. 2017, pp. 553–565

Gourmet Photography Dataset³

24000 immagini di cibo divise in due classi:

Classe positiva



Classe negativa



Terminologia

Classe positiva: immagini esteticamente belle

Classe negativa: immagini esteticamente non belle

³ Kekai Sheng et al. «Learning to assess visual aesthetics of food images». In: Computational Visual Media 7.1 (2021), pp. 139–152

Risultati classificazione su GPD

Sono state utilizzate diverse metodologie:

1. Feature hand-crafted legate al colore e alla texture, eventualmente combinate tra loro, e classificatore SVM



Accuratezza massima raggiunta sul test set con LBP unito alla media dei canali H e S con valore pari al **74,46%**

2. Feature estratte da una rete neurale Resnet-18 e classificatore SVM



Accuratezza massima raggiunta sul test set estraendo le feature dal layer res3b_relu con valore pari al **83,91%** e sul validation set pari al **84,78%**

3. Utilizzo di una rete neurale ResNet-18 per l'intero task, la quale è stata sottoposta a Fine Tuning con Early Stopping



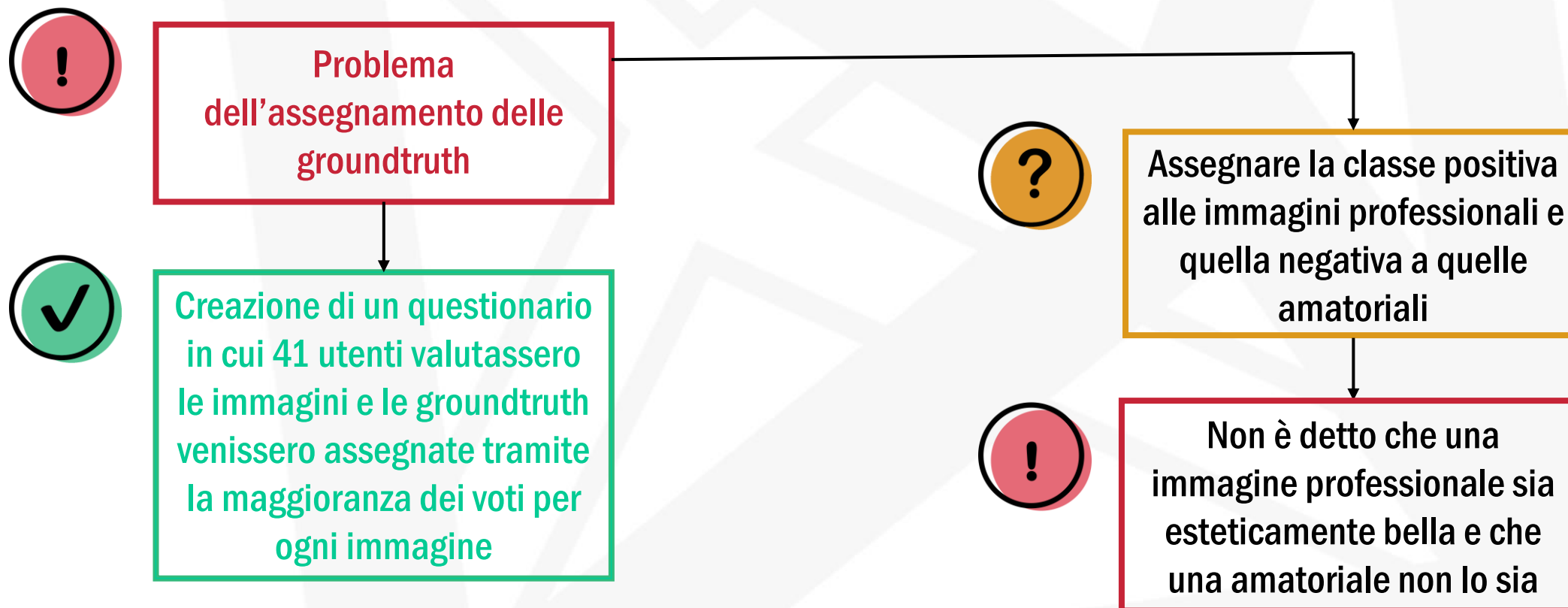
Accuratezza massima raggiunta sul test set pari al **90,51%** e sul validation set pari al **90,43%**



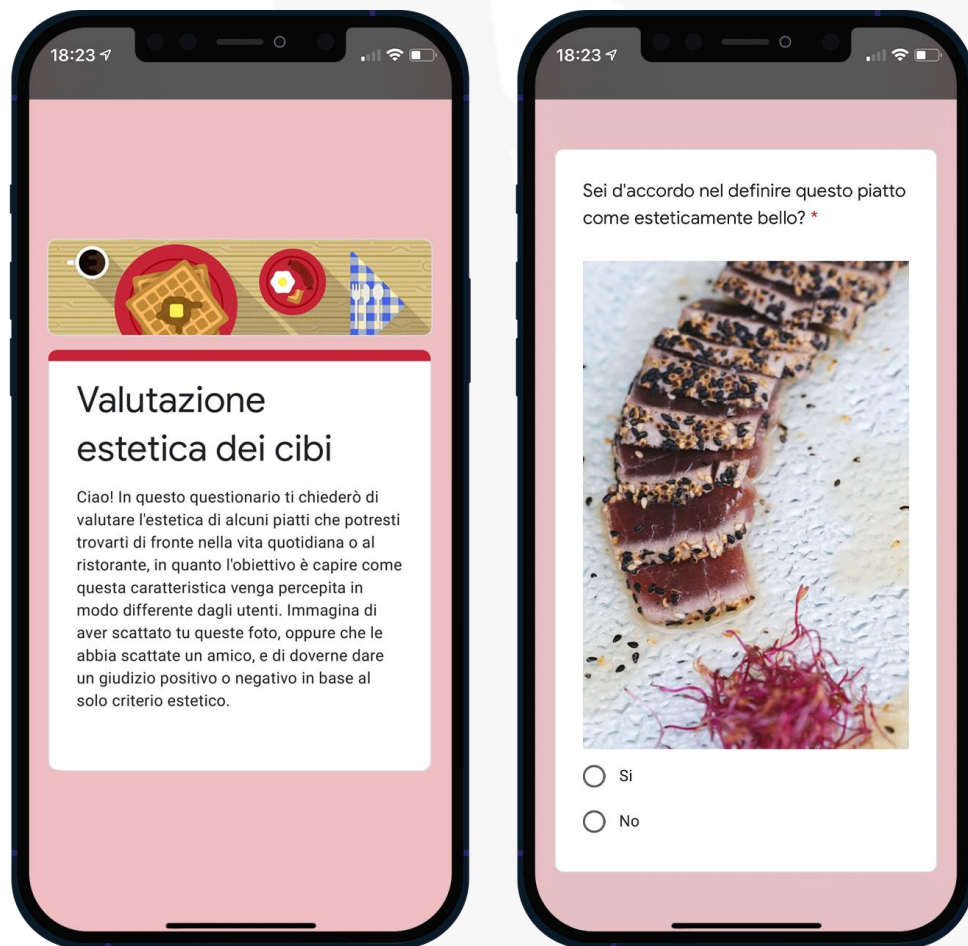
Si è scelto in questi due casi di lavorare con set di immagini bilanciati e di avere anche un validation set in modo tale da avere due set di immagini per valutare le prestazioni

Dataset proposto

- Composto da 50 fotografie amatoriali e 50 professionali
- Si vuole compiere un test su un set di immagini differente dal test set estratto dal GPD



Assegnamento delle groundtruth



Tramite questo questionario 41 utenti hanno valutato le 100 immagini del dataset proposto come segue:

- 73 immagini con groundtruth positiva
- 27 immagini con groundtruth negativa

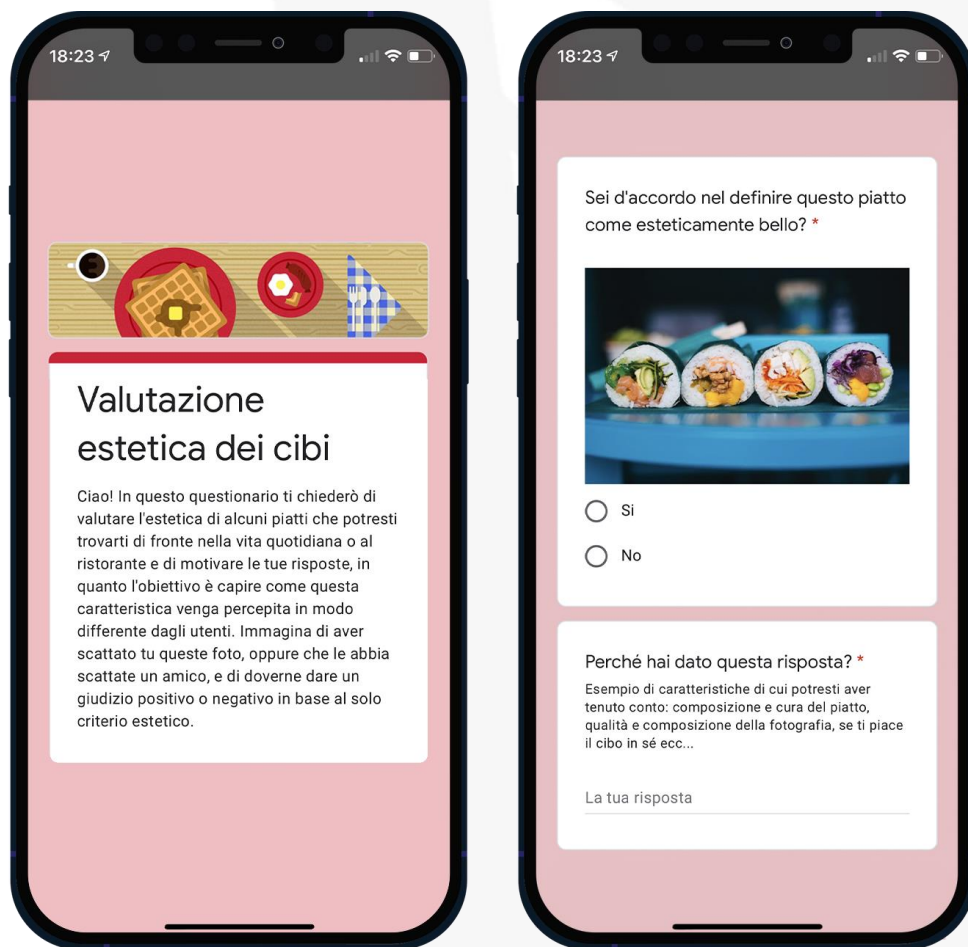
Positive



Negative



Analisi delle motivazioni



Tramite questo questionario un sottoinsieme di 11 utenti ha valutato nuovamente le 100 immagini del dataset motivando le risposte date, in modo tale da poter confrontare le motivazioni degli utenti con le aree evidenziate dalle immagini Grad-CAM⁴ e notare eventuali analogie



11/11 voti positivi poiché c'è un bel contrasto cromatico e i cibi sono composti accuratamente



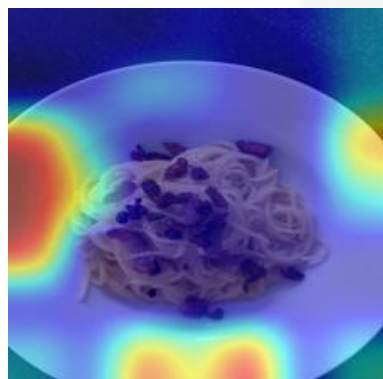
11/11 voti negativi poiché è sfocata e il cibo non ha una composizione accurata

⁴ Ramprasaath R Selvaraju et al. «Grad-cam: Visual explanations from deep networks via gradient-based localization». In: Proceedings of the IEEE international conference on computer vision. 2017, pp 618–626

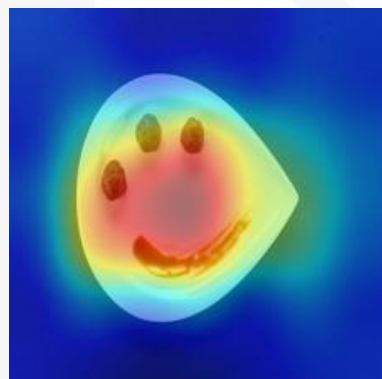
Risultati sul dataset proposto e analisi Grad-CAM

Accuratezza del **72%** con la rete ResNet-18 sottoposta a Fine Tuning con Early stopping

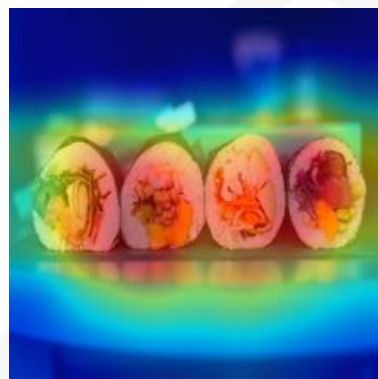
Immagini ottenute con la tecnica Grad-CAM



False negative



False positive



Positiva classificata correttamente



Negativa classificata correttamente



Legenda

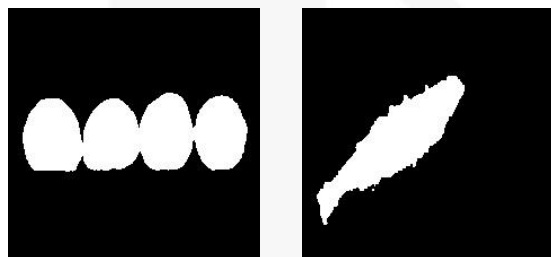
Aree rosse: aree considerate di maggior rilevanza ai fini della predizione da parte della rete
Aree blu: aree considerate di minor rilevanza per la predizione da parte della rete neurale

Indicatore di concentrazione



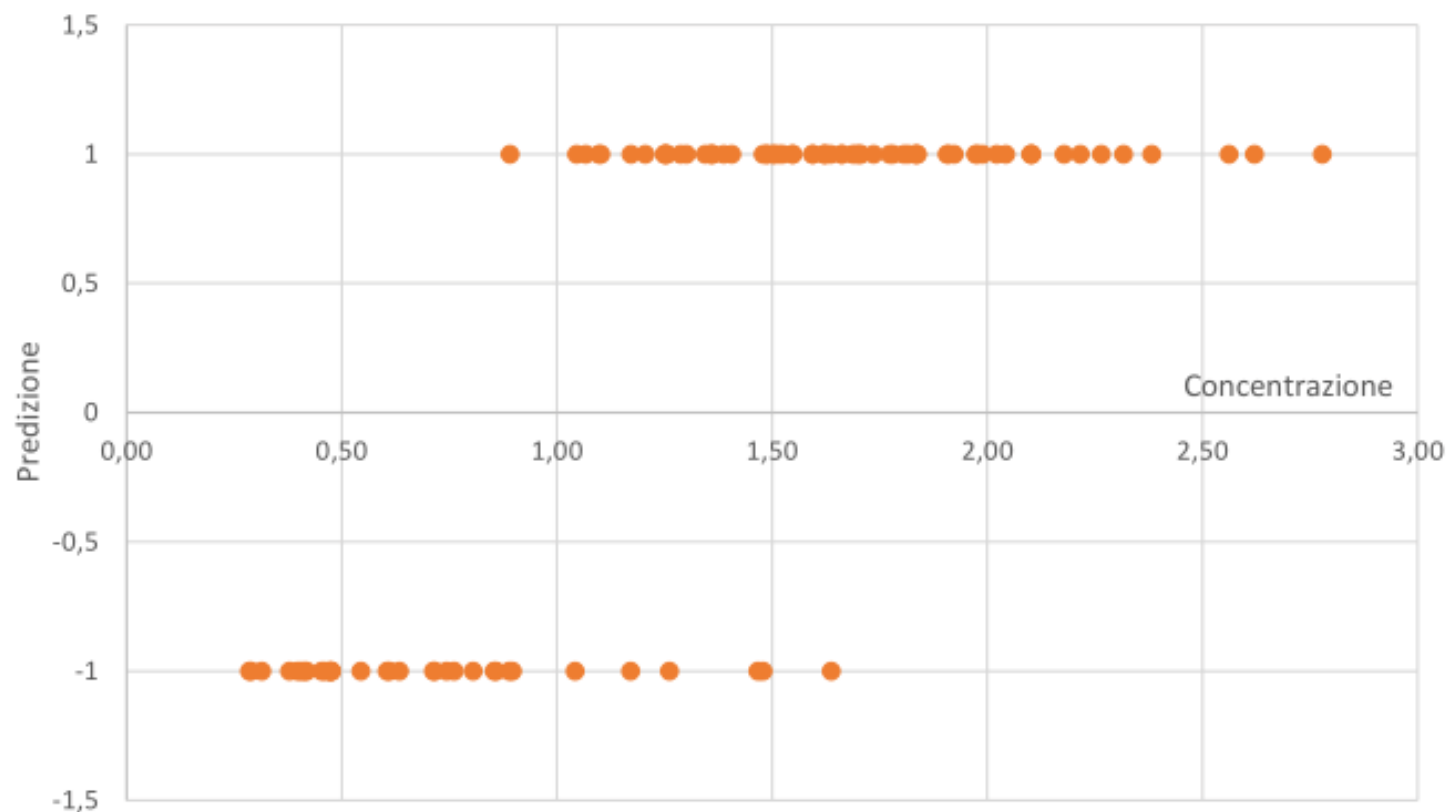
Ipotesi La rete predice la classe positiva quando l'area più significativa della Grad-CAM si concentra sul cibo

1. Creazione di maschere binarie dei cibi



2. Calcolo di A_m ovvero il rapporto, espresso in percentuale, tra area della maschera corrispondente al cibo e area totale dell'immagine
3. Normalizzazione dei valori di energia delle mappe Grad-CAM rispetto al valore massimo e al valore minimo
4. Calcolo di E_m ovvero il rapporto, espresso in percentuale, tra l'energia che cade all'interno della maschera corrispondente al cibo e l'energia totale
5. Calcolo dell'indicatore di concentrazione C come rapporto tra E_m e A_m

Risultati dell'analisi delle Grad-CAM



Legenda

Il valore -1 della predizione corrisponde alla classe negativa, mentre il valore 1 corrisponde alla classe positiva

Si osserva una correlazione tra predizioni e concentrazione dell'energia in corrispondenza del cibo, in particolare al crescere della concentrazione le predizioni tendono a diventare tutte corrispondenti alla classe positiva

- 67 immagini hanno avuto predizione positiva
- 33 immagini hanno avuto predizione negativa

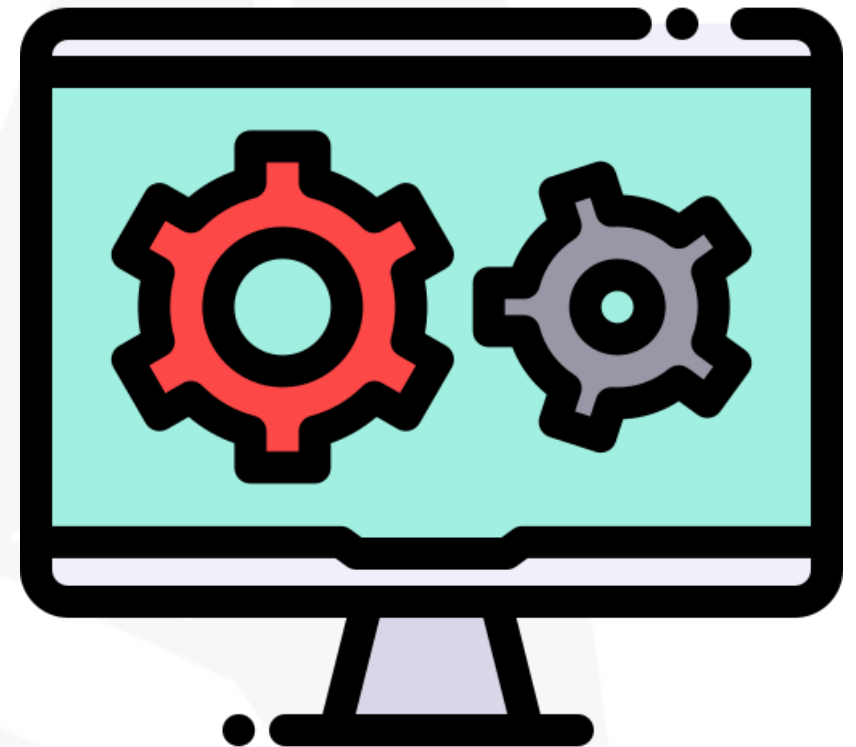
Conclusioni e possibili sviluppi

Conclusioni:

- Gli utenti tendono a valutare le immagini concentrandosi sul cibo
- L'uso delle reti neurali permette di ottenere i migliori risultati di accuratezza

Possibili sviluppi futuri:

- Ampliamento del dataset proposto
- Ampliamento della popolazione che valuta le immagini
- Eventuale integrazione di questo studio all'interno di un'applicazione mobile



Grazie per l'attenzione!



Relatore:

Prof. Paolo Napoletano

Correlatore:

Prof. Gianluigi Ciocca

Candidato:

Sofia Damaso

Matricola:

845189