**Università degli Studi di Salerno**

**Dipartimento di Informatica**

Immagine che contiene regina

Descrizione generata automaticamente

**Corso di Laurea in Informatica**

Fondamenti di Data Science e Machine Learning

Progetto: Driver Drowsiness Detection

|  |  |
| --- | --- |
| Relatori | Studenti |
| Prof. Giuseppe Polese  Dott.re Stefano Cirillo | Domenico Trotta 0522500810  Michele Castellaneta 0522500824 |

**Anno Accademico 2019/2020**

**Sommario**

[Abstract 3](#_Toc47452659)

[1. Introduzione 4](#_Toc47452660)

[2. Lavori presenti in letteratura 5](#_Toc47452661)

[3. Dataset 7](#_Toc47452662)

[3.1 Annotazioni e statistiche del dataset 9](#_Toc47452663)

[4. Modello preliminare 11](#_Toc47452664)

[4.1 Caricamento dataset e Pre-processing 12](#_Toc47452665)

[4.2 Struttura 13](#_Toc47452666)

[4.3 Training modello 16](#_Toc47452667)

[4.4 Testing 18](#_Toc47452668)

[5. Modello proposto 20](#_Toc47452669)

[5.1 Struttura 22](#_Toc47452670)

[5.2 Training 24](#_Toc47452671)

[5.3 Testing 25](#_Toc47452672)

[6. Confronto tra i due modelli 25](#_Toc47452673)

[7. Data Profiling 28](#_Toc47452674)

[7.1 FD 30](#_Toc47452675)

[7.2 RFD 33](#_Toc47452676)

[**7.2.1 Extent** 33](#_Toc47452677)

[**7.2.1 Confronto** 36](#_Toc47452678)

[7.3 Feature selection 39](#_Toc47452679)

[**7.3.1 Selezione Univariata** 40](#_Toc47452680)

[**7.3.2 Feature Importance** 40](#_Toc47452681)

[**7.3.3 Matrice di Correlazione con Heatmap** 42](#_Toc47452682)

[8. Confronto finale 44](#_Toc47452683)

[8.1 Confronto con lavori già svolti 44](#_Toc47452684)

[8.2 Sommario delle sperimentazioni effettuate 45](#_Toc47452685)

[9. Descrizione demo 46](#_Toc47452686)

[9.1 Detection con buone condizioni di luce 46](#_Toc47452687)

[9.2 Allarme visivo/sonoro 48](#_Toc47452688)

[9.3 Acquisizione con camera IR 48](#_Toc47452689)

[9.4 Schermata impostazioni 49](#_Toc47452690)

[10. Conclusioni 50](#_Toc47452691)

[11. Riferimenti 51](#_Toc47452692)

# Abstract

Secondo i rapporti di analisi sugli incidenti stradali degli ultimi anni, è noto che la principale causa di incidenti stradali con conseguenti decessi, lesioni gravi e lievi è dovuta a un guidatore assonnato. Lo stato di sonnolenza può essere causato dalla mancanza di sonno, farmaci, droghe o guida continua per periodi prolungati. Un aumento del tasso di incidenti stradali causati dalla sonnolenza durante la guida indica la necessità di un sistema che rilevi lo stato del guidatore e lo avvisi prima che si verifichino incidenti. Negli ultimi anni, molti ricercatori hanno mostrato interesse per il rilevamento della sonnolenza (Drowsiness Detection). I loro approcci monitorano sostanzialmente le caratteristiche fisiologiche o comportamentali relative al conducente o le misure relative al veicolo utilizzato. Per far fronte a questo problema, proponiamo un algoritmo di monitoraggio degli occhi basato su una rete neurale, per la precisione una rete neurale convoluzionale per determinare lo stato aperto o chiuso dell'occhio e attivare un allarme se il conducente è assonnato. Vengono inoltre presentati risultati sperimentali dettagliati per evidenziare i punti di forza e di debolezza della nostra tecnica. Una precisione del 99% circa è stata registrata per la metodologia proposta. Per il test in condizioni reali è stato prodotto un programma che in real-time grazie all’ausilio di una videocamera riesce a determinare lo stato degli occhi del conducente.

# 1. Introduzione

L’elaborato discusso in questo documento tratta di un progetto svolto durante il corso di *Fondamenti di Data Science e Machine Learning*, del corso di *Laurea Magistrale in Informatica* dell’*Università degli Studi di Salerno*.

La sonnolenza di solito si verifica a causa di sonno insufficiente, acquisizione di farmaci e anche a causa della noia causata dalla guida di veicoli per un lungo periodo di tempo. In uno stato di sonnolenza, il conducente perderà il controllo del proprio veicolo causando un incidente.

Secondo i rapporti statistici, ogni anno più di 1,35 milioni di persone muoiono in incidenti stradali, mentre 50 milioni di persone rimangono ferite [1]. Una delle statistiche più strazianti presenti nel rapporto dell’organizzazione mondiale della sanità è che gli incidenti stradali risultano essere la principale causa di morte tra le persone di età compresa tra 5 e 29 anni.

Per ridurre questo problema ed evitare tali incidenti, lo stato del conducente deve essere osservato costantemente.

Lo scopo di tale progetto prevede la costruzione di un modello che sia in grado di effettuare un’operazione di Drowsiness Detection a partire da un dataset disponibile liberamente in rete. Una volta ottenuto tale modello questo verrà utilizzato per realizzare una demo che effettui la detection del soggetto inquadrato.

Le misurazioni che vengono impiegate per il rilevamento della sonnolenza rientrano in tre categorie di base: misurazioni fisiologiche, comportamentali e basate sul veicolo. L'approccio proposto in questo documento utilizza la misura "comportamentale", ovvero funziona interpretando i segnali visivi del conducente.

Il documento è organizzato come segue:

* la sezione 2 presenta un indagine bibliografica nella quale si discute delle precedenti metodologie di rilevazione della sonnolenza presentate da diversi ricercatori.
* La sezione 3 discute in dettaglio del dataset utilizzato.
* La sezione 4 descrive il primo modello realizzato.
* La sezione 5 descrive i miglioramenti apportanti al modello preliminare.
* La sezione 6 descrive un confronto tra i modelli presentati nelle sezioni precedenti.
* La sezione 7 descrive l’attività di data profiling effettuata sul dataset.
* La sezione 8 descrive i risultati ottenuti confrontandoli con i risultati ottenuti dai lavori già presenti in letteratura.
* La sezione 9 illustra il funzionamento della demo real-time realizzata.
* La sezione 10 descrive le conclusioni del lavoro svolto.

# 2. Lavori presenti in letteratura

In questa sezione si discuterà di varie metodologie che sono state proposte dai ricercatori negli ultimi anni per quanto riguarda l’operazione di Drowsiness detection.

Flores et al [2], nel 2009, hanno presentato un componente per l’Advanced Driver Assistance System (ADAS) in grado di rilevare automaticamente la sonnolenza del conducente. Il modulo utilizza algoritmi di intelligenza artificiale insieme ai dati visivi che vengono acquisiti. Il sistema identifica e monitora il viso e gli occhi e determina la sonnolenza utilizzando Support Vector Machine (SVM). Il sistema è progettato per funzionare in real-time anche in condizioni di luce variabile. Questo sistema oltre ai battiti delle ciglia, prende in considerazione anche altre distrazioni del guidatore, ad esempio lo sbadiglio, l'inclinazione della testa e l'orientamento del viso. Tutti questi fenomeni vengono monitorati per fornire un grado di accuratezza ancora più elevato rispetto al solo rilevamento del battito delle ciglia.

Vitabile et al [3], nel 2011 hanno presentato un rilevatore di sonnolenza in tempo reale da utilizzare nei veicoli. Una fonte di luce a infrarossi da 850 nm è fissata sul cruscotto dell'auto permettendo l’illuminazione della pupilla. Ciò semplifica il rilevamento degli occhi in quanto la retina dell'occhio ha la proprietà di riflettere il 90% della luce proiettata su di essa. Lo stato di sonnolenza viene identificato quando gli occhi sono chiusi per più dell'80% di un certo periodo di tempo. Tecniche di elaborazione delle immagini efficienti sono combinate con una tecnologia hardware consolidata come Field Programmable Gate Array (FPGA). Ciò consente il rilevamento della sonnolenza in tempo reale e consente al sistema di elaborare un intero frame 720x576 in 16,7 microsecondi. La scalabilità e il riutilizzo del codice di FGPA possono aiutare a ridurre i costi di sviluppo. Sono stati osservati dei false allarms con l’utilizzo di occhiali per la vista o di oggetti che riflettono i raggi infrarossi. Per tale motivo, sono necessari ulteriori lavori per superare le limitazioni nel rilevamento degli occhi per i portatori di occhiali, in modo tale da renderlo disponibile per tutti i tipi di driver e consentire al sistema di supportare oggetti che riflettono gli infrarossi.

Singh e Kaur [4] nel 2012, hanno proposto un metodo che rileva la sonnolenza usando l'algoritmo SIFT. Scale-Invariant Feature Transform è un algoritmo utilizzato in computer vision che permette di rilevare e descrivere caratteristiche, o feature locali nelle immagini. L'algoritmo è stato pubblicato da David G. Lowe nel 1999.

Viene effettuato il rilevamento delle palpebre in tempo reale utilizzando una webcam con risoluzione 640 x 480. Gli occhi vengono rilevati da ciascun fotogramma e ogni battito di ciglia viene misurato rispetto a un valore medio. Il sistema confronta l'apertura dell'occhio per ogni blink con un valore medio standard e viene attivato un allarme se l'apertura dell'occhio supera questo valore per un certo numero di fotogrammi consecutivi. Gli autori hanno registrato una precisione del 99%. Inoltre, la risoluzione utilizzata, ovvero 640 x 480 non è altissima, ma nonostante questo sono stati ottenuti ottimi risultati. In questo algoritmo, il sistema deve conservare le informazioni sui fotogrammi passati perché le misurazioni del blink degli occhi su una certa quantità di fotogrammi vengono utilizzate per monitorare la sonnolenza.

Chuang-Wen et al [5], nel 2013 hanno introdotto "CarSafe", la prima applicazione per smartphone Android per il rilevamento della sonnolenza. L'applicazione richiede uno smartphone con doppia fotocamera e funziona effettuando uno switch tra le due fotocamere. Le fotocamera anteriore monitora il blinks rate del conducente e la posa della testa per determinare la sonnolenza. La fotocamera posteriore tiene conto di misure basate sul veicolo. Determina la distanza del veicolo con altri veicoli sulla strada per verificare se il conducente è vicino ad altri veicoli e per verificare se si sta cambiando corsia. Carsafe ha una precision dell'83% e una recall del 75%.

Sahayadhas et al [6], nel 2013 hanno usato l’EOG[[1]](#footnote-1) per monitorare i movimenti oculari. I dati ricavati dal EOG vengono quindi utilizzati per rilevare la sonnolenza del conducente. I ricercatori hanno sfruttato le misure basate su segnali fisiologici per rilevare la sonnolenza che risultano essere più accurate e affidabili in quanto utilizzano informazioni sullo stato interno del conducente. Esistono varie altre misure fisiologiche, ad es. EEG (elettroencefalografia), ECG (elettrocardiografia), EMG (elettromiografia) che potrebbero anche essere utilizzati per migliorare ulteriormente l'efficienza del sistema.

A. Rahman et al [7], nel 2015 hanno presentato un nuovo algoritmo per rilevare la sonnolenza in real-time. L’algoritmo acquisisce il volto e tramite l’algoritmo Viola Jones Cascade classifier si effettua il rilevamento del volto e degli occhi. Viola Jones utilizza le feature di Haar per il rilevamento degli occhi. Dopo aver ritagliato la regione dell’occhio individua due angoli degli occhi e un punto sulla palpebra inferiore. Per determinare se l’occhio è chiuso o meno viene calcolato il punto medio dei due angoli superiori degli occhi e successivamente viene calcolata la distanza dal punto individuato sulla palpebra inferiore. Se tale distanza è zero o vicina allo zero, l’occhio viene classificato come chiuso. Inoltre, se l’occhio rimane chiuso costantemente per 2 o più secondi, si presume che il conducente sia assonnato e viene attivato un allarme.

Questa tecnica fornisce risultati estremamente precisi se utilizzata in buone condizioni di illuminazione ed eseguita utilizzando una fotocamera ad alta risoluzione (è stata utilizzata una camera da 16MP).

# 3. Dataset

Dataset che contengono immagini degli occhi sono generalmente registrati in buone condizioni e contengono immagini ad alta risoluzione, che li rendono adatti per il rilevamento della pupilla, il rilevamento dell'iride, il monitoraggio degli occhi o il rilevamento dello sguardo.

Per questo progetto è stato utilizzato un grande dataset di immagini di occhi catturate da diverse fotocamere a infrarossi (NIR) (ad es. Fotocamere Intel Realsense, IDS Imaging).

Il dataset utilizzato si chiama MRL Eye Dataset [8] che risulta essere scaricabile gratuitamente in rete. Questo set di dati contiene immagini a infrarossi in bassa e alta risoluzione, tutte catturate in varie condizioni di luminosità e da diversi dispositivi.

Per ottenere le immagini degli occhi, nella prima fase è stato effettuato un crop manuale di molte migliaia di regioni oculari da diverse immagini NIR; le immagini di input di esempio sono presentate in Figura 1.

Immagine che contiene persona, fotografia, uomo, facciata

Descrizione generata automaticamente

Figura 1 Esempi di immagini catturate da una telecamera a infrarossi per auto.

Nella seconda fase, sono state utilizzate le immagini ritagliate manualmente per addestrare il rivelatore oculare basato sull'istogramma dei gradienti orientati[[2]](#footnote-2) combinato con il classificatore SVM. Questo rilevatore è stato utilizzato per estrarre automaticamente le regioni oculari. L'esempio del rilevamento della regione dell'occhio è mostrato in Figura 2. Dopo la fase di rilevamento, è stata accuratamente controllata ciascuna regione rilevata e sono stati rimossi i falsi positivi. In totale sono state create e controllate 85.000 immagini di occhi di diverse persone (37 persone diverse) catturate in varie condizioni di illuminazione; il set di dati contiene immagini di diversa qualità con proprietà diverse. Successivamente, verranno mostrati gli esempi di occhi che possono essere trovati nel dataset utilizzato.

Immagine che contiene persona, automobile, uomo, guardando

Descrizione generata automaticamente

Figura 2 Esempi di rilevamento degli occhi utilizzando il rilevatore HOG-SVM che è stato creato per il rilevamento automatico della regione degli occhi.

Ad esempio, gli occhi delle persone con gli occhiali sono mostrati nella prima fila della Fig. 3. I problemi che spesso si verificano con gli occhiali sono i riflessi. Nel dataset, vengono fornite molte immagini con il riflesso. Inoltre, per ogni immagine dell'occhio, sono state fornite tre diversi livelli di riflessione basati sulla dimensione delle area di riflesso in ciascuna immagine; nessun riflesso (la prima riga in Fig. 3), piccolo riflesso (la seconda riga in Fig. 3) e grande riflesso (la terza riga in Fig. 3).

Immagine che contiene fotografia, guardando, vista, acqua

Descrizione generata automaticamente

Figura 3 Esempi di occhi con occhiali e riflessi inclusi nel set di dati utilizzato.

In Figura 4, gli esempi di set di dati mostrano che i riflessi possono anche verificarsi senza occhiali (ad esempio nella sclera, nella pupilla e nell’iride).

Immagine che contiene guardando

Descrizione generata automaticamente

Figura 4 Esempi di riflessi oculari senza occhiali inclusi nel set di dati proposto.

In generale, molte donne usano ciglia e sopracciglia sintetiche, che possono causare il problema ai riconoscitori della direzione dello sguardo e dello stato degli occhi. Pertanto, nel set di dati, sono state fornite anche le informazioni sul genere di ogni persona.

## 3.1 Annotazioni e statistiche del dataset

Di seguito sono riportate le caratteristiche del dataset utilizzato:

* **ID soggetto**: nel set di dati, sono presenti i dati di 37 persone diverse (33 uomini e 4 donne);
* **ID immagine**: il set di dati è composto da 84.898 immagini;
* **genere [0 - uomo, 1 - donna]**: il set di dati contiene le informazioni sul genere della persona in ciascuna immagine (uomo, donna);
* **occhiali [0 - no, 1 - sì]**: le informazioni se l'immagine dell'occhio contiene gli occhiali sono fornite per ogni immagine (con e senza gli occhiali); 24.001 immagini con gli occhiali;
* **stato dell'occhio [0 - chiuso, 1 - aperto]**: questa proprietà contiene le informazioni su i due stati oculari (41.945 chiusi e 42.953 aperti);
* **riflesso [0 - nessuno, 1 - piccolo, 2 - grande]**: sono stati annotati tre stati di riflessione in base alla dimensione del riflesso (66.227 nessun riflesso, 5.962 piccolo riflesso e 12.709 grande riflesso);
* **condizioni di illuminazione [0 - pessime, 1 - buono]**: in base alla quantità di luce durante l'acquisizione dei video (53.630 pessime, 31.268 buone);
* **ID sensore [01 - RealSense, 02 - IDS, 03 - Aptina]**: il set di dati contiene le immagini catturate da tre diversi sensori (sensore Intel RealSense RS 300 con risoluzione 640 × 480, sensore IDS Imaging con risoluzione 1280 × 1024 e sensore Aptina con risoluzione 752 x 480).

Di seguito sono riportati alcuni grafici che graficamente mostrano la distribuzione di alcune delle caratteristiche analizzate:

Immagine che contiene screenshot, disegnando

Descrizione generata automaticamente Immagine che contiene screenshot

Descrizione generata automaticamente

Figura 5 Distribuzione reflection a sinistra e distribuzione sensor type a destra.

Per quanto concerne l’attributo eye state che rappresenterà la nostra etichetta, presenta la seguente distribuzione:

Immagine che contiene disegnando

Descrizione generata automaticamente

Figura 6 Distribuzione eye state.

Come si può osservare in Figura 6, il dataset presenta una distribuzione pressoché identica per quanto riguarda lo stato degli occhi.

In sintesi, il set di dati contiene immagini di occhi a bassa risoluzione, immagini con riflessi negli occhi o con riflessi sugli occhiali causati dall'illuminatore IR posto di fronte alla persona. Sono incluse anche alcune immagini oculari in cui la testa non è orientata verso la fotocamera. Tutti questi tipi di immagini in questo set di dati rendono più difficile il rilevamento di parti oculari, ma permettono di addestrare un modello che sia più robusto ad eventuali variazioni di posa e illuminazione.

# 4. Modello preliminare

Nella seguente sezione verrà descritto il lavoro svolto per la creazione di un primo modello basato sulle CNN che abbiamo realizzato basandoci su lavori già svolti [9].

Le Convolutional Neural Networks (**CNN**) sono una famiglia di reti neurali largamente impiegate nell’ambito di computer vision e, più generalmente, con dati che hanno relazioni spaziali.

Se dovessimo rappresentarle graficamente, quello che ne verrebbe fuori è un grafo con una serie di passi dove, partendo da un input complesso, vengono estratte man mano informazioni, chiamate **features**, sempre più semplici e facilmente riconoscibili dal computer.

Le CNN seguono quindi un’architettura a livelli, tipicamente non ciclica. I livelli più importanti sono i**layer di convoluzione**, da cui prende il nome.

La convoluzione è un’operazione matematica che ricorda il funzionamento di uno scanner. Il risultato è funzione della traslazione di una funzione sull’altra, al variare del tempo. Poichè tipicamente le CNN ricevono in input dei dati bidimensionali, come ad esempio le immagini, avremo una o più matrici in input ed una chiamata kernel che verrà traslata al fine di calcolarne il risultato, chiamato feature map. Il passo di traslazione prende il nome di stride.

A seguito di un layer di convoluzione, è comune trovare un layer di sotto campionamento che riduca le dimensioni delle matrici in output. Le operazioni sono più comuni sono:

* **MaxPooling**: ogni cella della matrice risultante contiene il valore massimo della sottomatrice di partenza.
* **AveragePooling**: il valore di ciascuna cella è la media dei valori della sottomatrice di partenza.

Infine, ritroviamo i **layer full connected** che possono essere visti come dei vettori di neuroni. L’output di ciascun neurone è inviato a tutti i neuroni del layer successivo, con diversi valori di peso.

Il nostro obiettivo in questo caso è stato quello di ottenere una base sul quale effettuare delle analisi che verranno descritte nei paragrafi successivi.

Il codice utilizzato per effettuare le operazioni che verranno descritte di seguito è riportato nel file Jupyter Notebook *Drive Drowsiness Detection.ipynb.*

## 4.1 Caricamento dataset e Pre-processing

Il primo task effettuato riguarda il caricamento delle immagini e il loro pre-processing.

La lettura delle immagini è stata effettuata tramite il metodo **imread** della libreria OpenCV. Successivamente, si è effettuato il resize (**cv2.resize**) e il reshape delle immagini in modo tale che quest’ultime avessero tutte la stessa dimensione. In particolare, la dimensione scelta è stata (*24, 24, 1)* dove i primi due numeri indicano l’altezza e la larghezza dell’immagine, mentre il terzo indica il numero di canali della stessa. Poiché le immagini sono state lette in scala di grigi, il terzo numero risulta essere uno.

Si è poi provveduto all’estrazione delle feature "genere", "occhiali", "riflesso", "condizione di illuminazione" e "tipo di sensore" oltre che delle etichette "stato dell’occhio" dal nome dell’immagini, creando un vettore che contenesse le singole feature per ogni immagine.

Di seguito è riportato il confronto tra un’immagine originale del dataset e la relativa immagine ottenuta dopo averne effettuato il caricamento e il pre-processing.

Immagine che contiene fotografia, guardando, sedendo, indossando

Descrizione generata automaticamenteImmagine che contiene interni, piastrellato, sedendo, fotografia

Descrizione generata automaticamente

Figura 7 Sulla sinistra l’immagine originale e sulla destra l’immagine (24, 24, 1) in scala di grigi caricata.

Infine, come ultima operazione è stata effettuata una normalizzazione delle immagini dividendo i pixel di ogni immagine per 255, in modo tale che questi risultassero avere valori tra 0 e 1. Con tale operazione i valori dei pixel risultano essere più piccoli (si noti che questi valori piccoli rappresentano ancora l'immagine originale), con il beneficio che il tempo richiesto dal modello per convergere si riduce significativamente.

Inoltre, come sarà mostrato nei successivi paragrafi, il lavoro non si è limitato alla creazione di modelli che considerassero soltanto le immagini, ma verranno considerate anche tutte le altre feature di cui dispone il dataset.

Precedente alla creazione del modello, è stata l'operazione di splitting dei dati in training set e test set, ottenuto tramite la funzione **train\_test\_split** di Scikit Learn. Si è deciso di utilizzare l’80% (67918 immagini) dei dati come training set e il restante 20% (16980 immagini) come test set.

Inoltre, dopo aver effettuato lo splitting, le immagini dei due set sono state salvate nelle rispettive cartelle “Training” e “Test”, in modo tale da prevederne un rapido caricamento nelle successive esecuzioni.

## 4.2 Struttura

Il modello di training utilizzato è rappresentato da una CNN, basata sulla libreria Keras. La scelta è ricaduta sulla tipologia di rete CNN in quanto essa viene spesso applicata alla analisi di immagini, su cui il progetto di Drive Drowsiness Detection si basa. La struttura mostrata di seguito riguarda il primo modello che abbiamo creato.

Il modello creato è costituito da:

|  |  |
| --- | --- |
| Input Layer | Input: immagini 24 x 24 in scala di grigi (1 canale) |
| Convolutional layer (CL 1) | Output filters: 32  Kernel size: 3 x 3  Activation: ReLU |
| Pooling layer (PL 1) | Pool size: 1 x 1 |
| Convolutional layer (CL 2) | Output filters: 32  Kernel size: 3 x 3  Activation: ReLU |
| Pooling layer (PL 2) | Pool size: 1 x 1 |
| Convolutional layer (CL 3) | Output filters: 64  Kernel size: 3 x 3  Activation: ReLU |
| Pooling layer (PL 2) | Pool size: 1 x 1 |
| Dropout Layer | Percentage: 25 % |
| Flatten Layer |  |
| Fully connected layer (FCL 1) | Output: 128 nodi  Activation: ReLU |
| Dropout Layer | Percentage: 50 % |
| Fully connected layer (FCL 4) | Output: 2 nodi  Activation: Softmax |

Tabella 1 Struttura del modello con le sole immagini

Immagine che contiene screenshot

Descrizione generata automaticamente

Figura 8 Struttura del modello creato

Tale modello considera come input le sole immagini. Volendo considerare anche altre features di cui il dataset dispone, si rende necessaria una modifica alla rete. Dopo il livello di Flatten, infatti, viene aggiunto un nuovo livello, che permette di concatenare le immagini con le features, appositamente inserite in un nuovo livello di Input separato. Tale operazione verrà descritta nella sezione 5.

Una volta definita la struttura del modello, questo si può creare tramite la funzione **Model** di TensorFlow, che permette successivamente di compilarlo. Per i parametri della compilazione, si è deciso di optare per l’ottimizzatore Adam (computazionalmente molto efficiente), mentre si è scelta la *sparse\_categorical\_crossentropy* come loss function. La cross entropy è una loss function, utilizzata per misurare la dissomiglianza tra la distribuzione delle etichette di classe osservate e le probabilità previste dell'appartenenza alla classe. L'uso della cross entropy categorica sparsa con 2 classi ci dà un risultato finale tra 0 e 1, che è esattamente ciò di cui abbiamo bisogno.

## 4.3 Training modello

Per quanto riguarda il validation set si è utilizzato una percentuale del training set come validation set. Quest’ultimo cambia in modo randomico ad ogni epoca, permettendo così di considerare diversi dati su cui valutare il modello ed ottenere quindi una misura più precisa della bontà dello stesso. Pertanto, abbiamo utilizzato il parametro *validation\_split* del metodo fit assegnandogli il valore 0.1. Per evitare l’overfitting del modello si è utilizzata la funzione EarlyStopping di Keras mostrata di seguito:



Figura 9 Funzione EarlyStopping offerta da Keras.

* **monitor**: indica il valore da monitorare;
* **patience**: indica il numero di epoche senza alcun miglioramento dopo il quale l'allenamento verrà interrotto.
* **min\_delta**: indica la variazione minima della quantità monitorata per poterla qualificare come un miglioramento, ovvero una variazione assoluta inferiore a min\_delta, non verrà considerata come miglioramento.
* **verbose**: livello di verbosità;
* **baseline**: valore di base per la quantità monitorata. Il training si interromperà se il modello non mostra miglioramenti rispetto al valore di base.
* **restore\_best\_weights:** indica se ripristinare i pesi del modello dell'epoca con il miglior valore della quantità monitorata. Se settato a False, vengono utilizzati i pesi del modello ottenuti nell'ultimo step del training.

I risultati da questo ottenuti, come l’accuracy e la loss sul training e validation set sono stati salvati in un file testuale specifico per il modello. Inoltre, sono stati creati dei grafici raffiguranti l’andamento della loss e dell’accuracy all’aumentare del numero delle epoche.

Il modello è stato stoppato dopo solo 14 epoche per evitare l’overfitting.

La *Tabella 2* mostra i risultati registrati sul modello addestrato al termine della fase di training.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Nome | Dim. Immagini | Epoche | Train loss | Train accuracy | Val accuracy | Val loss |
| img\_14\_0.25\_3CL\_0.2 | 24 x 24 | 14 | 0.064 | 0.976 | 0.982 | 0.048 |

Tabella 2 Risultati del training del modello

I risultati nella *Tabella 2* indicano che il modello presenta un’accuracy sul validation set (6792 immagini) del 98% circa. Inoltre, il tempo medio impiegato per ogni epoca del training è di circa 10 secondi.

In Figura 10 sono riportati i grafici che descrivono l’andamento della loss e dell’accuracy all’aumentare del numero di epoche.

Immagine che contiene testo, mappa

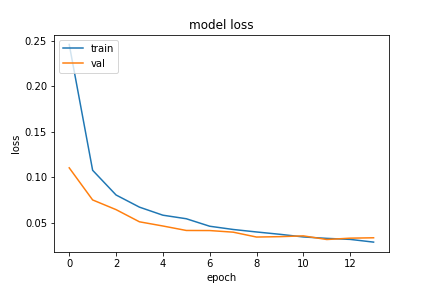
Descrizione generata automaticamente

Figura 10 A sinistra vi è il grafico che descrive l'andamento dell’accuracy, mentre a destra abbiamo la loss.

Dai grafici si può osservare come l’accuracy risulta essere superiore al 98% e la loss inferiore a 0.05. Dai grafici, si può anche notare come l’addestramento sia stato interrotto non appena le due curve hanno iniziato a separarsi e questo per evitare l’overfitting.

Inoltre, dopo aver effettuato il training del modello si è provveduto al suo salvataggio per poterlo testare nella demo illustrata nella sezione 9. Il formato utilizzato per il salvataggio del modello è il formato HDF5[[3]](#footnote-3).

## 4.4 Testing

Una volta effettuato il training del modello, questo è stato testato sul test set tramite la funzione evaluate di Keras. I risultati ottenuti dal modello sul test set (test loss e test accuracy) sono stati aggiunti al file testuale descritto precedentemente.

Nella *Tabella 3* sono riportati i risultati ottenuti dal modello dopo aver effettuato la fase di testing.

|  |  |  |  |
| --- | --- | --- | --- |
| Nome | Epoche | Test loss | Test accuracy |
| img\_14\_0.25\_3CL\_0.2 | 14 | 0.040 | 0.986 |

Tabella 3 Risultati ottenuti sul test set.

Come si può notare dai dati in *Tabella 3*, la loss del modello sul test set è simile alla loss registrata sul validation set, quindi in linea con i risultati ipotizzati.

Vengono ora considerate le previsioni effettuate dal modello su alcune immagini del test set.

|  |  |
| --- | --- |
|  |  |
|  |  |

Figura 11 Esempi di predictions.

Come si può notare dai risultati, il modello risulta essere piuttosto preciso nonostante la risoluzione delle immagini sia di solo 24 x 24 pixel.

Infine, viene mostrato il risultato ottenuto con una immagine non facente parte del test set. Tale immagine è stata acquisita tramite la webcam del laptop. La risoluzione della webcam utilizzata è pari a 1280 x 720 px.

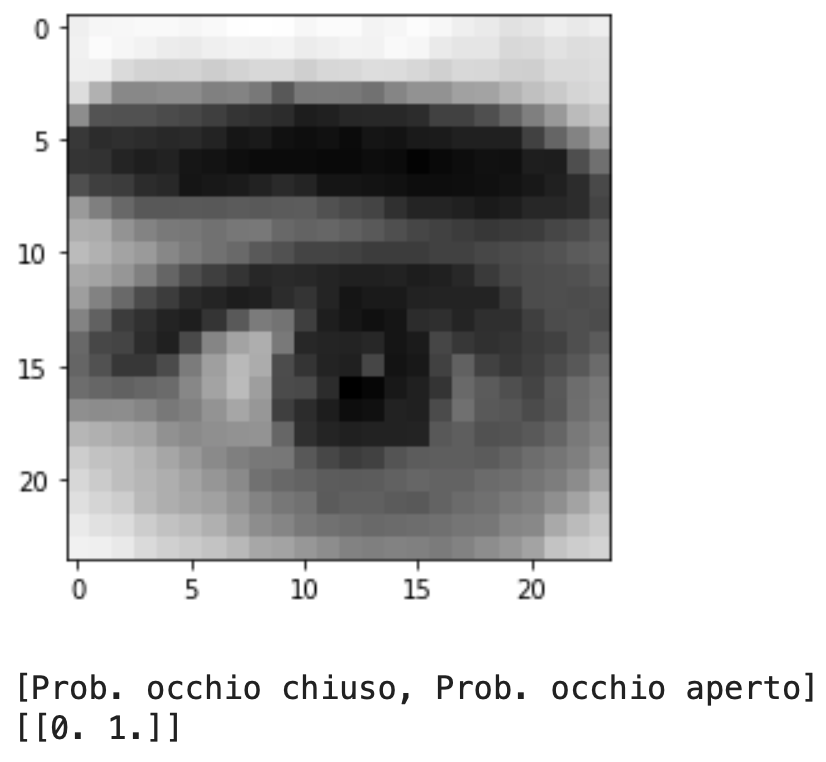


Figura 12 Prediction su un immagine acquisita tramite webcam.

In questo caso il modello con probabilità certa ha classificato l’occhio come aperto.

# 5. Modello proposto

In questa sezione verranno illustrate le modifiche apportate al modello riportato nella sezione 4. Come prima operazione abbiamo verificato se un aumento delle dimensioni delle immagini passate in input al modello portasse ad un miglioramento delle prestazioni e successivamente si è provveduto ad addestrare un modello che considerasse anche tutte le altre feature presenti nel dataset.

Si è deciso di aumentare le dimensioni delle immagini per osservare il comportamento del modello descritto precedentemente all’aumentare della risoluzione. In Figura 13 sono state riportate due immagini per mettere in evidenza la differenza di risoluzione.

Immagine che contiene sedendo, piastrellato, largo, aria

Descrizione generata automaticamente Immagine che contiene ciambella

Descrizione generata automaticamente

Figura 13 Sulla sinistra l'immagine 24 x 24, mentre sulla destra l'immagine 52 x 52.

Come si può osservare dalla Figura 13, l’immagine a destra presenta il doppio della risoluzione rispetto all’immagine sulla sinistra e questo chiaramente permette di avere più dettagli.

Per constatare come questi dettagli aggiuntivi vanno ad impattare sul modello si è resa necessaria la modifica del modello illustrato precedentemente. Quest’ultimo prendeva in input immagini 24 x 24, mentre adesso riceverà in input immagini 52 x 52.

La prima modifica effettuata riguarda l’operazione di resize delle immagini caricate, che ha comportato anche una modifica della dimensione dell’input della rete neurale.

Pertanto, il nuovo modello presenta una struttura simile a quella mostrata in Figura 8, l’unica differenza riguarda le dimensioni dell’input layer.

Effettuando l’addestramento del modello con immagini 52 x 52 si è riscontrato un aumento dei tempi richiesti per ogni epoca, dovuto alle maggiori dimensioni delle immagini. Infatti, il tempo medio richiesto per ogni epoca è passato da circa 10 secondi richiesti per un’immagine 24 x 24 a circa 52 secondi per un’immagine 52 x 52.

Per quanto concerne i risultati ottenuti alla fine della fase di training sono riportati nella tabella seguente, nella quale sono riportati anche i risultati ottenuti sul modello che considera le immagini con dimensione 24 x 24.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Nome | Dim. Immagini | Epoche | Train loss | Val loss | Train accuracy | Val accuracy |
| img\_14\_0.25\_3CL\_0.2 | 24 x 24 | 14 | 0.064 | 0.048 | 0.976 | 0.982 |
| img\_15\_0.25\_3CL\_0.2 | 52 x 52 | 15 | 0.068 | 0.056 | 0.974 | 0.980 |

Tabella 4 Confronto modelli

Dai dati riportati in Tabella 4 si può notare come i risultati ottenuti con il nuovo modello sia quasi identici a quelli registrati sul modello precedente. Pertanto, l’aver aumentato le dimensioni delle immagini a portato ad un aumento dei tempi richiesti per il training non giustificato dai risultati ottenuti.

Dopo aver constatato che la dimensione dell’immagine non ha un grosso impatto sui risultati ottenuti si è deciso di provare a considerare tutte le feature presenti nel dataset.

Da sottolineare che i lavori precendenti illustrati nella sezione 2, non hanno effettuato un’operazione di questo tipo in quanto si sono limitati all’utilizzo delle sole immagini acquisite per l’individuazione dello stato dell’occhio. In questo modo la rete durante l’addestramento considererà delle informazioni aggiuntive di cui non disponeva precedentemente e sarà interessante osservare come tali informazioni influenzeranno il risultato.

Di seguito quindi, verranno illustrate le modifiche che si sono rese necessarie per far si che il modello durante l’addestramento considerasse anche altre feature oltre alle immagini.

Infine, verranno discussi i risultati ottenuti su tale modello.

## 5.1 Struttura

Per considerare anche le seguenti feature:

* **genere**
* **occhiali**
* **riflesso**
* **condizioni di illuminazione**
* **ID sensore**

è necessario modificare la struttura del modello in modo tale da inserire un ulteriore layer chiamato *concatenate layer*. Tale layer si occupa di unire l’output dell’operazione di flatten con un nuovo input layer del modello che è rappresentato delle cinque feature elencate sopra.

Di seguito è mostrata la struttura del modello dopo aver introdotto il concatenate layer e un nuovo input layer.

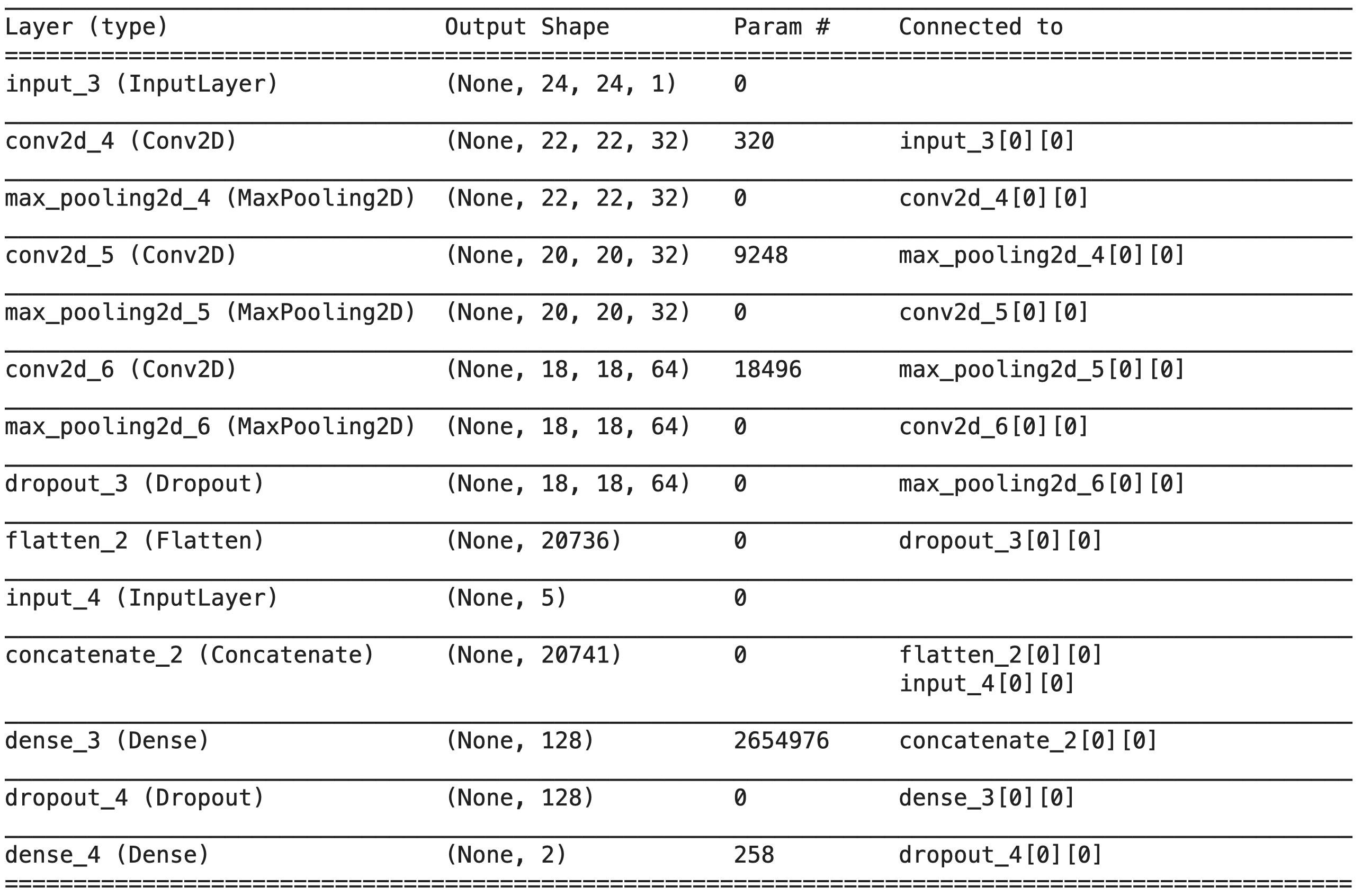


Tabella 5 Struttura del modello in forma tabellare

Dalla Tabella 5 si può notare che dopo l’operazione di flatten sono stati aggiunti due nuovi layer, ovvero un input layer composto da cinque elementi e il concatenate layer che si occupa di concatenare le nuove feature all’output del flatten.

Per una visione grafica del nuovo modello creato si rimanda alla figura illustrata alla pagina seguente.

Immagine che contiene screenshot

Descrizione generata automaticamente

Figura 14 Struttura del modello che considera anche le altre feature del dataset

## 5.2 Training

Per quanto riguarda il training nella tabella di seguito sono riportati i risultati registrati dopo aver effettuato l’addestramento del modello illustrato in Figura 14.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Nome | Dim. Immagini | Epoche | Train loss | Val loss | Train accuracy | Val accuracy |
| all\_14\_0.25\_3CL\_0.2 | 24 x 24 | 14 | 0.057 | 0.044 | 0.987 | 0.984 |

Tabella 6 Risultati del training.

Considerando anche altre informazioni oltre alle sole immagini il tempo medio richiesto per eseguire un epoca è di circa 11 secondi. Pertanto, il modello appena descritto risulta impiegare mediamente un secondo in più per ogni epoca rispetto al modello descritto nella sezione 4.

Di seguito sono riportati l’andamento della loss e dell’accuracy all’aumentare delle epoche.

Immagine che contiene testo, mappa

Descrizione generata automaticamenteImmagine che contiene screenshot

Descrizione generata automaticamente

Figura 15 A sinistra vi è il grafico che descrive l'andamento dell’accuracy, mentre a destra abbiamo la loss.

Osservando i grafici mostrati in Figura 15, si può notare la somiglianza che vi è con quelli mostrati in Figura 10. Anche in questo caso, l’accuracy risulta essere circa del 99%, mentre la loss è circa 0.04. Inoltre, si può notare che anche in questo caso l’addestramento sia stato interrotto non appena le due curve iniziavano a separarsi.

## 5.3 Testing

Di seguito infine, sono riportati i risultati che il modello ha fatto registrare sul test set.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nome | Dim. Immagini | Epoche | Test loss | Test accuracy |
| all\_14\_0.25\_3CL\_0.2 | 24 x 24 | 14 | 0.041 | 0.985 |

Tabella 7 Risultati ottenuti sul test set.

Come si può notare anche in questo caso l’accuracy ottenuta risulta essere alta e la loss bassa. Inoltre, come ci aspettavamo, la loss e l’accuracy registrate sul test sono simili alla loss e all’accuracy registrate sul validation set.

# 6. Confronto tra i due modelli

In tale sezione si discuterà delle valutazioni fatte sui modelli illustrati nelle sezioni precedenti, ovvero quello che considera solo le immagini e nessun’altra feature e quello che considera anche le altre feature oltre che alle sole immagini. Le valutazioni fatte per esso sono state effettuate anche per tutti gli altri modelli creati. I risultati sono riportati nella sezione 8.

Inoltre, si è provveduto a valutare i risultati ottenuti su due ulteriori modelli, che però sono stati addestrati su dataset differenti da quello da noi utilizzato.

Per poter valutare due o più modelli è necessario aver provveduto al salvataggio del modello dopo averne effettuato il training. La prima operazione effettuata consiste nel caricamento dei modelli sopra citati tramite la funzione **load\_model** che restituisce il modello compilato.

Prima di effettuare una comparazione dei modelli descritti nelle sezioni 4 e 5 si vuole mostrare cosa accade ad uno dei modelli dei creati quanti si aumenta il numero di epoche.

Di seguito, sono riportati i due grafici che descrivono l’andamento della loss e dell’accuracy all’aumentare del numero di epoche. In questo caso sono state considerate 25 epoche.

Immagine che contiene mappa, testo

Descrizione generata automaticamente Immagine che contiene screenshot

Descrizione generata automaticamente

Figura 16 A sinistra vi è il grafico che descrive l'andamento dell’accuracy, mentre a destra abbiamo la loss.

Come ci aspettavamo, in tutti e due i grafici mostrati in Figura 16 le due curve tendono a separarsi sempre più all’aumentare del numero del numero di epoche. Infatti, la curva di colore blu che rappresenta la loss e l’accuracy sul training tende a diminuire e aumentare rispettivamente al crescere del numero di epoche, mentre la loss e l’accuracy registrate sul validation set rimane stabile. Il fenomeno appena descritto ci indica che il modello si sta adattando troppo ai dati di training, pertanto, non ha senso aumentare il numero di epoche perché nemmeno con 500 epoche otterremmo un miglioramento dei risultati sul validation set.

Nella tabella mostrata di seguito sono riportati i dati sul test set ottenuti dai due modelli citati precedentemente e di un nuovo modello, simile a quello descritto nella sezione 5, ma che presenta un numero di epoche maggiore.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nome | Dim. Immagini | Epoche | Test loss | Test accuracy |
| img\_14\_0.25\_3CL\_0.2 | 24 x 24 | 14 | 0.040 | 0.986 |
| all\_14\_0.25\_3CL\_0.2 | 24 x 24 | 14 | 0.041 | 0.985 |
| all\_25\_0.25\_3CL\_0.2 | 24 x 24 | 25 | 0.048 | 0.988 |

Tabella 8 Confronto modelli.

Da una lettura della Tabella 8 si evince che i tre modelli presentano dei risultati simili sia sulla loss che sul training.

Come accennato in precedenza, andiamo a valutare i risultati che otteniamo utilizzando il nostro test set (costituito da circa 7000 immagini) su dei modelli addestrati con dataset differenti. Da sottolineare che i risultati che verranno mostrati sono stati registrati su modelli che presentano la stessa struttura della rete neurale convoluzionale illustrata nei paragrafi precedenti.

Uno dei due modelli è stato reperito sul sito sul quale è illustrata le rete neurale utilizzata per la creazione del modello preliminare [9], mentre l’altro modello è stato addestrato utilizzando un dataset ottenuto tramite l’unione di diversi campioni di dataset individuati in rete. L’unione di diversi insiemi di immagini ci ha permesso di ottenere un nuovo dataset contenente poco più di 5000 immagini.

Di seguito sono riportati i risultati registrati sul test set per entrambi i modelli citati.

|  |  |  |
| --- | --- | --- |
| Nome | Test loss | Test accuracy |
| Modello scaricato [9] | 1.031 | 0.703 |
| Modello con nuovo dataset | 0.853 | 0.649 |

Tabella 9 Risultati modelli addestrati con un diverso training set.

Si può notare come la loss e l’accuracy ottenuta sul test set per entrambi i modelli sia superiore e inferiore rispettivamente ai risultati ottenuti sui modelli proposti precedentemente. Pertanto, dopo un’analisi di questo tipo è possibile affermare che il risultato finale risulta essere funzionale al dataset utilizzato per il training. Questo significa, che se il dataset è composto da immagini che ritraggono l’oggetto in analisi (nel nostro caso l’occhio) in diverse pose, illuminazioni e considerando un utente collaborativo e non collaborativo, allora il dataset risulterà essere particolarmente robusto e il modello potrà effettuare il training considerando un’ampia varietà di immagini differenti.

Informazioni più dettagliate su come deve essere strutturato un dataset che risulti essere piuttosto robusto sono riportate nella sezione 3.

# 7. Data Profiling

Successivamente alla creazione del modello si è provveduto ad effettuare diverse operazioni di data profiling come l'estrazione di dipendenze funzionali, dipendenze funzionali rilassate (riferimento) e features selection. Operazioni di tale genere aiutano nella scoperta, comprensione e organizzazione dei dati. Operazione preliminare di tali processi è stata la creazione di due differenti csv. Questi sono stati utilizzati come input per i vari algoritmi di FD discovery e features selection.

Il primo file csv riportava tutte le features del dataset, cioè:

* Image\_Name
* Glasses
* Genders
* Reflections
* Image\_quality
* Sensor\_type
* Eye\_state

Il secondo file csv comprendeva tutte le features del dataset tranne l'attributo " Eye\_state", che rappresenta la feature target del dataset.

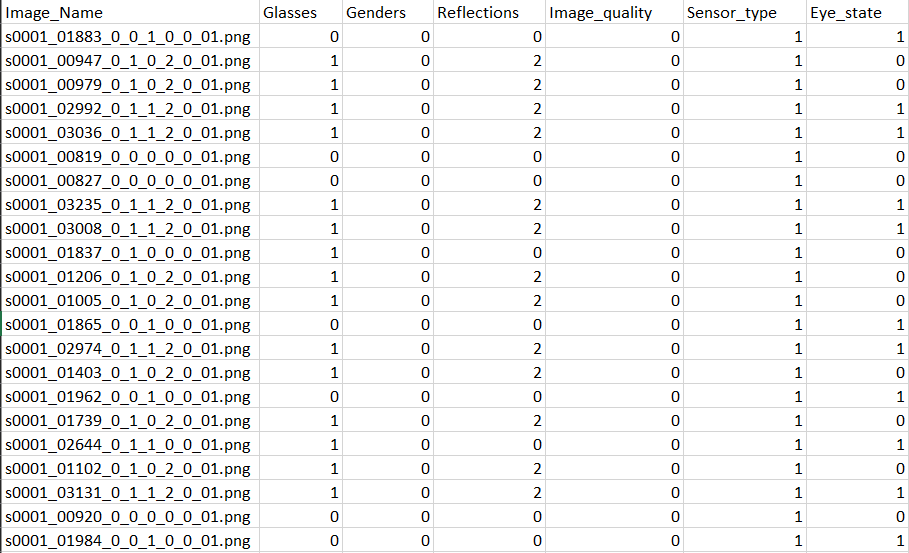


Figura 17 File CSV con etichette.



Figura 18 File CSV senza etichette.

## 7.1 FD

La prima operazione è stata quella di ricercare l'esistenza di Dipendenze Funzionali all'interno del dataset. Tale operazione è stata effettuata tramite il tool COD3. Il file di input utilizzato per l'algoritmo è stato quello che non prevedeva il target attribute, modificato in modo da eliminare l'attributo Image\_Name, non rilevante allo scopo.

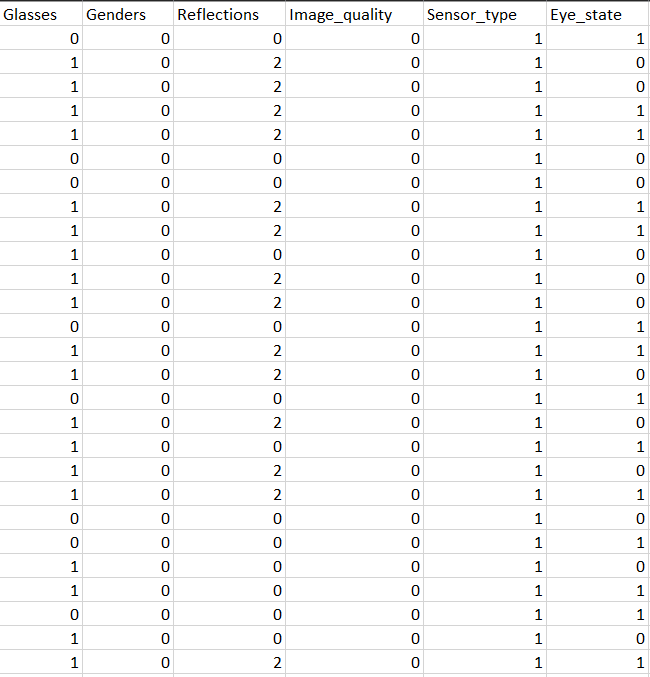


Figura 19 Struttura del file input per l'algoritmo

Applicando l'algoritmo per l'FD discovery su tale file non sono state rilevate dipendenze funzionali. Tuttavia, esaminando i log del tool, si è notato che inizialmente erano state identificate quattro differenti dipendenze, che diventano non più valide oltre un certo numero di tupla.

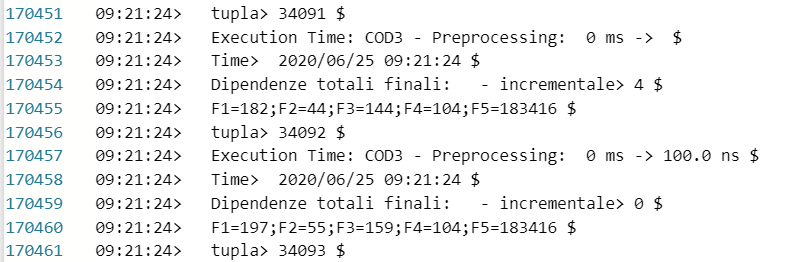


Figura 20 Punto di terminazione di validità delle FD

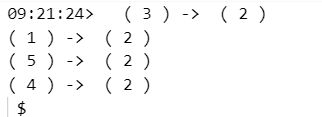


Figura 21 Dipendenze rilevate fino alla tupla 34091

Nello specifico, non avendo considerato l'attributo Image\_Name, la corrispondenza degli attributi era la seguente:

|  |  |
| --- | --- |
| Nome | Numero |
| Glasses | 1 |
| Genders | 2 |
| Reflections | 3 |
| Image\_quality | 4 |
| Sensor\_type | 5 |

Tabella 10 Mapping features

Per cui, le corrispondenze trovate erano:

Glasses →Genders

Reflections → Genders

Image\_quality → Genders

Sensor\_type → Genders

Si è quindi provveduto ad analizzare il file allo scopo di trovare il motivo della terminazione della validità della dipendenza.

Immagine che contiene mappa

Descrizione generata automaticamente

Figura 22 Dettaglio del file csv in cui si interrompe la validità delle dipendenza

Come si può analizzare dalla figura, nella tupla 34092 si ha la prima occorrenza del valore 1 per l'attributo genere. Ciò vuol dire che dalla tupla 1 alla tupla 34091 i valori di genere erano tutti uguali, e pari a 0. Il cambiamento del valore di tale attributo ha provocato la terminazione della validità delle dipendenze. Le dipendenze rilevate sono state considerate banali e non utili ai fini della feature selection, in quanto basate sulla considerazione di un valore univoco sulla parte destra.

## 7.2 RFD

Recentemente si è posto un interesse sulle cosiddette Relaxed Functional Dependencies [9], vale a dire delle dipendenze funzionali inesatte che rilassano su uno o più vincoli delle dipendenze funzionali canoniche. Considerando che non sono state rilevate dipendenze funzionali canoniche sul dataset, si è provveduto a ricercare RFD che rilassano sull'extent e sul confronto tramite l'algoritmo BIRD (extent) e DOMINO (extent e comparison) [10].

### **7.2.1 Extent**

Rilassare sull'extent significa che la dipendenza funzionale vale per "quasi tutte" le tuple, o su un sottoinsieme di esse. Diversi sono i casi per cui una dipendenza potrebbe non valere per tutte le tuple, come la presenza di valori mancanti, differenti tipi di dati o errori presenti negli stessi.

Applicando l'algoritmo sul dataset considerato, è stata rilevata la dipendenza funzionale approssimata (Genders Reflections Sensor\_type) -> (Glasses) con soglia di extent g3-error di 0.9. Il g3-error corrisponde al minimo numero di tuple da rimuovere da una istanza di relazione r affinché una dipendenza funzionale X → Y valga. Nello specifico caso ciò significa che affinchè la dipendenza trovata valga, sarebbe necessario eliminare il 90% delle tuple. Nonostante la soglia della dipendenza sia molto alta, si è effettuata la costruzione di un nuovo modello che considerasse tutte le features tranne la feature Glasses, che rappresenta la parte destra della dipendenza funzionale, quindi la feature che è possibile derivare dalle altre. La struttura del modello, a meno delle features utilizzate per il training, è la medesima del modello descritto nel 6.

In Figura 23 sono mostrati i grafici riportanti l'andamento della loss e dell'accuracy al crescere del numero di epoche:

Immagine che contiene testo, mappa

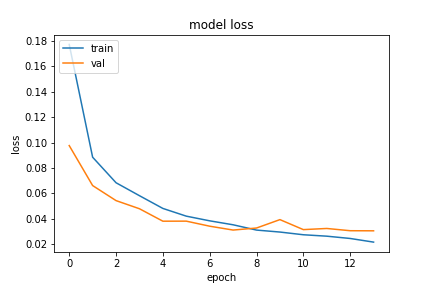
Descrizione generata automaticamente

Figura 23 A sinistra vi è il grafico che descrive l'andamento dell’accuracy, mentre a destra abbiamo la loss.

I risultati del modello sono mostrati nella tabella 11:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Training** | | **Validation** | | **Test** | |
| Nome | **Mean epoch training time** | **Loss** | **Accuracy** | **Loss** | **Accuracy** | **Loss** | **Accuracy** |
| all-glasses\_14\_0.25\_  3CL\_0.2\_  basicModel\_24 | 10.926 | 0.051 | 0.980 | 0.043 | 0.984 | 0.032 | 0.989 |

Tabella 11 Risultati modello che non considera la feature Glasses

Di seguito viene mostrato il confronto tra il modello discusso nel 6. e il modello discusso in questo paragrafo.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Training** | | **Validation** | | **Test** | |
| Nome | **Mean epoch training time** | **Loss** | **Accuracy** | **Loss** | **Accuracy** | **Loss** | **Accuracy** |
| all-glasses\_14\_0.25\_  3CL\_0.2\_  basicModel\_24 | 10.926 | 0.051 | 0.980 | 0.043 | 0.984 | 0.032 | 0.989 |
| all\_14\_0.25\_3CL\_0.2\_basicModel\_24 | 10.772 | 0.057 | 0.978 | 0.044 | 0.984 | 0.041 | 0.985 |
|  | + 0.154 | - 0.06 | + 0.002 | -0.001 | 0 | - 0.009 | + 0.004 |

Tabella 12 Confronto tra i modelli

Come si può notare dalla tabella, i risultati ottenuti non variano di quantità rilevanti.

### **7.2.1 Confronto**

Rilassare le dipendenze funzionali sul confronto significa usare paradigmi di matching approssimato per confrontare i valori dell'attributo sul lato sinistro e quelli del lato destro della dipendenza funzionale. Ciò permette di catturare relazioni semantiche tra gruppi di valori che sembrano essere simili, piuttosto che identici. Generalmente, questo tipo di RFD viene utilizzato per risolvere problemi di data quality, query optimization e knowledge discovery. Queste si basano su una soglia di comparison, che è la differenza massima ammissibile tra i valori di due tuple su un attributo.

Tramite l'algoritmo DOMINO, si è provato ad estrarre RFD basate sul confronto dal dataset utilizzato nel progetto. Nel fare ciò si sono utilizzate differenti soglie di comparison: 1, 2, 4 e 15.

Utilizzando una soglia di comparison pari ad uno non sono stati trovati risultati, vale a dire che non vi sono due attributi X e Y per cui la differenza massima tra i valori di X sia minore ad uno e la differenza massima tra i valori di Y sia minore ad 1.

Di seguito vengono riportati i risultati ottenuti con le diverse soglie:

SOGLIA 2

\*\*\*\*\*\* DISCOVERED RFDs \*\*\*\*\*\*\*

0@2.0->0@2.0 cc:0.9685232103301251

1@2.0->0@2.0 cc:0.8694514821618544

0@2.0->1@2.0 cc:0.5990717281682707

SOGLIA 4

\*\*\*\*\*\* DISCOVERED RFDs \*\*\*\*\*\*\*

0@4.0->0@2.0 cc:0.9685232103301251

1@4.0->0@2.0 cc:0.8694514821618544

0@4.0->1@2.0 cc:0.5990717281682707

SOGLIA 15

0@15.0->0@2.0 cc:0.9685232103301251

1@15.0->0@2.0 cc:0.8694514821618544

0@15.0->1@2.0 cc:0.5990717281682707

Nonostante le tre diverse soglie, le RFD sul confronto e le cc (CHE SONO?) dal tool sono le medesime:

0->0

1->0

0->1

E cioè:

Glasses → Glasses

Genders → Glasses

Glasses → Genders

Poichè la prima dipendenza è banale, ed il modello senza la feature Glasses è stato descritto nell' 8.2.1, si discuterà del modello sviluppato includendo tutte le features tranne Genders. La struttura del modello, a meno delle features utilizzate per il training, è la medesima del modello descritto nel 6.

In Figura 23 sono mostrati i grafici riportanti l'andamento della loss e dell'accuracy al crescere del numero di epoche:

Immagine che contiene mappa, testo

Descrizione generata automaticamenteImmagine che contiene screenshot

Descrizione generata automaticamente

Figura 24 A sinistra vi è il grafico che descrive l'andamento dell’accuracy, mentre a destra abbiamo la loss.

I risultati del modello sono mostrati nella Tabella 13:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Training** | | **Validation** | | **Test** | |
| Nome | **Mean epoch training time** | **Loss** | **Accuracy** | **Loss** | **Accuracy** | **Loss** | **Accuracy** |
| all-genders\_14\_0.25\_3CL\_0.2\_basicModel\_24 | 10.995 | 0.050 | 0.981 | 0.042 | 0.985 | 0.040 | 0.987 |

Tabella 13 Risultati modello che non considera la feature Genders

Di seguito viene mostrato il confronto tra il modello discusso nel 6. e il modello discusso in questo paragrafo.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Training** | | **Validation** | | **Test** | |
| Nome | **Mean epoch training time** | **Loss** | **Accuracy** | **Loss** | **Accuracy** | **Loss** | **Accuracy** |
| all-genders\_14\_0.25\_3CL\_0.2\_basicModel\_24 | 10.995 | 0.050 | 0.981 | 0.042 | 0.985 | 0.040 | 0.987 |
| all\_14\_0.25\_3CL\_0.2\_basicModel\_24 | 10.772 | 0.057 | 0.978 | 0.044 | 0.984 | 0.041 | 0.985 |
|  | + 0.223 | -0.007 | + 0.003 | -0.002 | + 0.001 | -0.001 | + 0.002 |

Tabella 14 Confronto tra i modelli

Come si può notare dalla tabella, i risultati ottenuti non variano di quantità rilevanti.

## 7.3 Feature selection

La feature selection è un importante concetto del machine learning, che può influire in maniera rilevante sulle performance di un modello. Infatti, features irrilevanti possono penalizzare le performance dello stesso, per cui si rende necessaria una attenta fase di scelta delle feature da utilizzare. La feature selection può essere intesa come un processo in cui vengono selezionate (automaticamente o manualmente) le feature più rilevanti per la feature target, cioè la variabile predittiva.

In generale, una operazione di questo tipo consente di:

· Ridurre l'Overfitting: Meno features → Minore probabilità di effettuare decisioni basate su noisy data.

· Migliorare l'Accuracy

In questo progetto sono state utilizzate tre tecniche tra le diverse presenti in letteratura. Esse sono:

1. Selezione Univariata

2. Feature Importance

3. Matrice di Correlazione con Heatmap

Nel seguito verranno mostrate le diverse tecniche applicate sul dataset utilizzato nel progetto, con degli snippet di codice per mostrarne il funzionamento.

Operazione preliminare alle tecniche di feature selection è stata quella di separare la colonna target dalle altre:

Immagine che contiene mano, tenendo, stanza, telefono

Descrizione generata automaticamente

### **7.3.1 Selezione Univariata**

I test statistici sono uno dei modi utilizzati per selezionare le features che hanno la più alta correlazione con la variabile di output.

La classe SelectKBest della libreria di Scikit-learn consente di effettuare un insieme di test statistici per selezionare uno specifico numero di features.

Di seguito viene mostrato il codice per restituire le 5 features del dataset insieme ai relativi punteggi, utilizzando il test statistico del chi-squared (chi²).



Figura 25 Snippet per la selezione univariata.

Il risultato di tali operazioni è mostrato nella Figura 25.

Immagine che contiene uccello

Descrizione generata automaticamente

Figura 26 Risultati selezione univariata.

Analizzando la Figura 25, si può notare che la feature più correlata all'attributo target risulta essere "Reflections".

### **7.3.2 Feature Importance**

Alternativamente alla selezione univariata, si può considerare l'importanza di ciascuna feature del dataset utilizzando una apposita proprietà del modello, cioè quella della "feature importance". Questa restituisce un punteggio per ogni feature del dataset: più è alto il punteggio e più la feature è rilevante rispetto alla feature target.

Tale proprietà è una classe propria dei classificatori ad alberi. Per usarla, quindi, ci sarà la necessità di usare un'istanza di questo tipo di classificatori. Nello snippet di codice seguente viene mostrato come sia possibile estrarre i punteggi per le 5 feature del dataset.



Figura 27 Costruzione del classificatore ad albero

Stampando il contenuto della variabile *model.feature\_importances\_* si ottiene una lista di valori indicante l'importanza di ognuna delle feature del dataset.



Figura 28 Risultati della feature importance

L'ordine dei risultati restituiti si riferisce a quello del salvataggio degli attributi nel dataframe. Per permettere una visualizzazione più chiara è possibile creare un grafico che mostra , oltre al valore delle feature, anche la loro denominazione.

Immagine che contiene stanza, uomo, tenendo

Descrizione generata automaticamente

Figura 29 Snippet per la costruzione dell'istogramma della feature importance

Eseguendo il codice si ottiene il seguente grafico:

Immagine che contiene screenshot

Descrizione generata automaticamente

Figura 30 Istogramma della feature importance

Dal grafico in Figura 29, è possibile osservare che la feature che ha riportato il maggiore punteggio è stata "Sensor\_type", differentemente a quanto accaduto con il metodo della selezione univariata descritto nell' 8.3.1. Le altre features del dataset presentano dei punteggi di importanza molto minori di quello ottenuto da "Sensor\_type", addirittura inferiori alla metà di quest'ultimo.

### **7.3.3 Matrice di Correlazione con Heatmap**

Un ulteriore metodo per la feature selection è rappresentato dalla matrice di correlazione. Essa mostra come le features sono correlate rispetto alle altre o all'attributo target. La correlazione può essere positiva o negativa. Nel primo caso l'incremento del valore della feature provoca l'incremento di quello dell'attributo target, mentre nel secondo l'incremento del valore della feature provoca il decremento del valore dell'attributo target.

La matrice di correlazione può essere associata ad una heatmap. Questa è una tecnica di data visualization che mostra la magnitudine di un fenomeno attraverso un colore.

Tramite la heatmap è facile verificare quali features sono più correlate alla variabile target. Il seguente codice mostra come sia possibile crearla tramite la libreria Seaborn.

Immagine che contiene uccello

Descrizione generata automaticamente

Figura 31 Snippet per la costruzione della matrice di correlazione con heatmap

Eseguendo le operazioni, si ottiene il risultato in Figura 32.

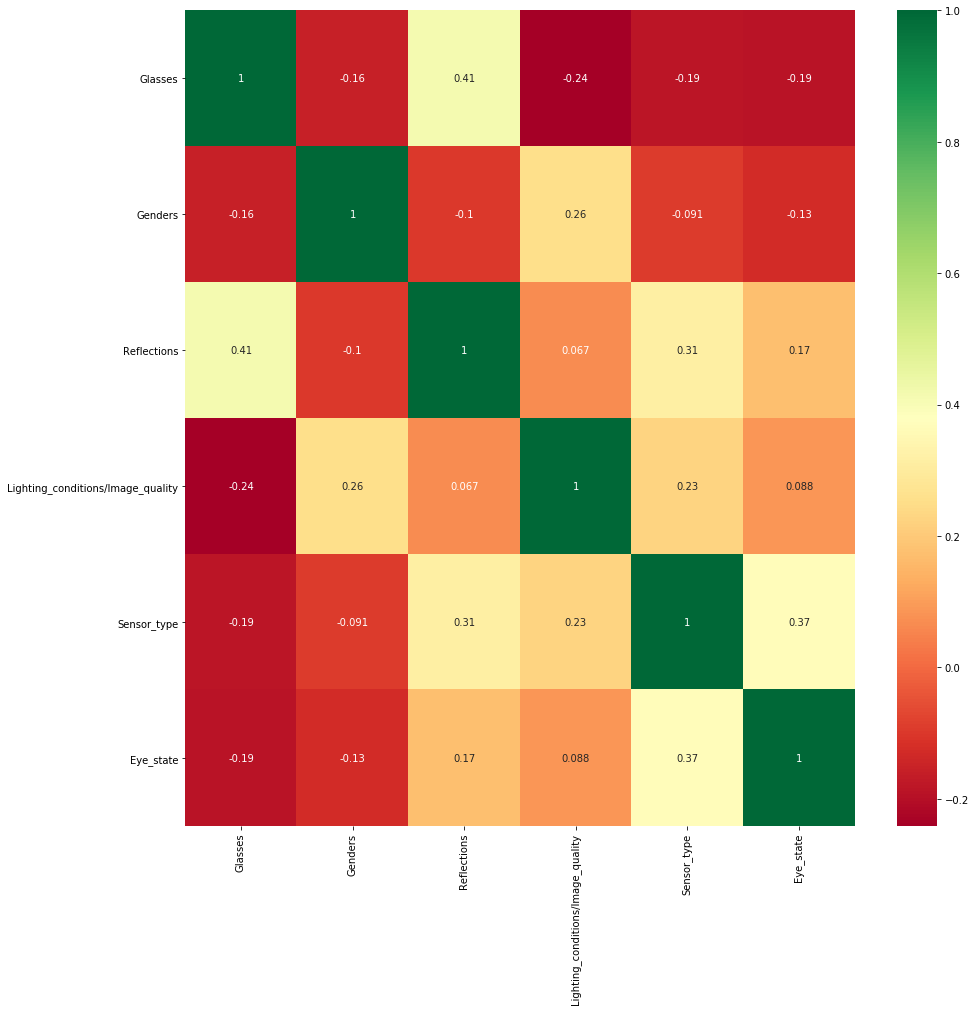


Figura 32 Matrice di correlazione con heatmap

Analizzando l'immagine in Figura 31, si nota che nessuna feature è fortemente correlata con la variabile target "Eye\_state": la correlazione più alta è risultata essere 0.37, corrispondente alla feature "Sensor\_type".

# 8. Confronto finale

In questa sezione verrà prima mostrato un confronto tra i lavori già svolti e la tecnica proposta e successivamente verrà mostrato un sommario delle diverse sperimentazioni effettuate che sono state descritte nelle sezioni precedenti.

## 8.1 Confronto con lavori già svolti

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Autori** | **Tecnica** | **Accuracy** | **Testing Data** | **Intrusivo** | **Scarsa illuminazione** | **Tipo camera** | **Funziona con occhiali**  **da vista/**  **sole.** | **Funziona con**  **faccia inclinata** | **Real time** | **Error rate** |
| Mandeep et al 2012. [3] | Eye Blink Monitoring using Mean Sift Algorithm | 99.4% | Non specificato | No | No | Webcam | Non specificato | No | Si | 1% |
| Salvatore, et al 2011.[2] | FPGA based prototyping | Non specificata | Non specificato | No | Si | IR CCD Camera | Non specificato | Si | Si | Non specificato |
| Flares, et al 2009.[1] | Face and Eye monitoring based on neural networks & visual information | 95.6% | 5 video | No | Si | Monocular Camera | Si | Si | Si | 4.4% |
| Chuang- Wen, et al 2013.[4] | Computer Vision & Machine Learning Algorithm | 83% | 12 video | No | Non specificato | Dual camera smartphone | Non specificato | Non specificato | Si | 17% |
| A.Sahayadhas. et al 2013.[5] | Electroculogr am and vehicle based measures | Non specificata | 15 video | Si | Si | Non specificato | Si | Si | No | Non specificato |
| A. Rahman, et al 2015.[6] | Real Time Drowsiness Detection using Eye Blinks Monitoring | 94% | 16 video | No | No | 16MP webcam | No | Si | Si | 6% |
| Tecnica proposta | Real Time Drowsiness Detection using Convolutional Neural Network (CNN) | 99% | Circa 7000 immagini | No | Si | 7MP  Webcam,  5MP Infrared Camera. | Solo con occhiali da vista | Si | Si | 1% |

Tabella 15 Confronto tecniche

La Tabella 15 illustra il confronto tra la tecnica proposta e le metodologie precedenti. La tecnica proposta ha un'accuratezza stimata del 99% che risulta essere quasi uguale all'algoritmo proposto da Mandeep et al che presenta la massima accuratezza tra le tecniche sopra menzionate. La qualità distintiva di questo approccio è che è stato testato su un test set abbastanza grande (circa 7000 immagini), che include condizioni di illuminazione e risoluzioni della telecamera variabili. L'algoritmo proposto è abbastanza economico e può essere implementato utilizzando solo una webcam e non richiede altri dispositivi hardware per funzionare. Inoltre, non è invadente e funziona bene in real-time. Per quanto riguarda le condizioni di scarsa illuminazione si consiglia l’utilizzo di una camera infrared. Infine, se gli occhi sono coperti con occhiali da sole, l'algoritmo fallisce nel rilevamento degli occhi e pertanto, non si ottengono i risultati sperati.

## 8.2 Sommario delle sperimentazioni effettuate

In questo sezione è riportata una tabella riepilogativa delle diverse sperimentazioni effettuate e dei risultati ottenuti su ciascuno di esse. La seguente tabella è stata estratta dal file *Sommario sperimentazioni.xlsx.*

Immagine che contiene computer

Descrizione generata automaticamente

Figura 33 Tabella riassuntiva delle sperimentazioni effettuate.

Sul file Excel, per agevolare la lettura dei risultati è stata fornita la possibilità di ordinare i modelli in base ad uno dei valori di interesse, come Test loss, Test accuracy, ….

# 9. Descrizione demo

In questa sezione verrà descritta la demo realizzata per testare il modello in casi reali. La demo creata permette in real-time di determinare se gli occhi del soggetto inquadrato sono aperti o chiusi. Nel caso gli occhi del soggetto inquadrato rimangono chiusi per un tempo superiore ad un valore preimpostato viene lanciato un allarme sonoro e visivo.

Per la realizzazione della demo si è fatto uso di OpenCV **(***Open Source Computer Vision Library*), ovvero, una libreria software multipiattaforma utilizzata nell'ambito della visione artificiale in real-time.

La libreria offre un insieme di algoritmi di visione artificiale e machine learning sia classici che all'avanguardia. Tali algoritmi possono essere utilizzati per rilevare e riconoscere volti, identificare oggetti, classificare azioni umane in video, tracciare oggetti in movimento, estrarre modelli 3D di oggetti, trovare immagini simili in un database di immagini, ecc.

Per ulteriori informazioni si rimanda alla documentazione di OpenCV [12].

Nei paragrafi seguenti, tramite delle immagini verrà illustrato il funzionamento della demo, mentre per la visione del codice realizzato si rimanda al file Live\_Drowsiness\_Detection.py.

## 9.1 Detection con buone condizioni di luce

Immagine che contiene persona, uomo, schermo, interni

Descrizione generata automaticamenteImmagine che contiene monitor, automobile, fotografia, autobus

Descrizione generata automaticamente

**4**

**2**

**1**

**3**

Figura 34 Drowsiness detection con buone condizioni di luce.

**1**

Tasto per accedere alle impostazioni della demo descritte nel paragrafo 9.4.

**2**

Viene mostrato lo stato degli occhi, in questo caso “Open”.

**3**

Tasto per chiudere la demo.

**4**

Viene mostrato un timer che ci permette di visualizzare per quanto tempo il soggetto ha tenuto gli occhi chiusi. Inoltre, se si tengono costantemente gli occhi chiusi per un intervallo di tempo predefinito verrà riprodotto un allarme sonoro e visivo come mostrato nel paragrafo successivo.

Nell’immagine a destra è stata utilizzata una camera con una qualità inferiore a quella utilizzata per l’acquisizione dell’immagine sulla sinistra. Tale camera è stata posta all’interno di un’automobile per simulare il comportamento del sistema in un ambiente quanto più realistico possibile. Inoltre, anche la distanza dagli occhi risulta essere maggiore rispetto alla camera utilizzata per la prima acquisizione.

Per quanto concerne il rilevamento degli occhi si sono utilizzati classificatori a cascata basati su feature di Haar che è un metodo di rilevamento di oggetti molto efficace proposto da Paul Viola e Michael Jones nel 2001. Le feature si estraggono applicando all’immagine una serie di maschere binarie. Applicare una maschera in una determinate posizione significa sommare i valori dei pixel nell’area bianca e sottrarre il valore dei pixel nell’area nera. Di seguito sono riportati degli esempi di Feature di Haar.

Immagine che contiene screenshot, disegnando, microonde

Descrizione generata automaticamente

Figura 35 Esempi di Feature di Haar. Nelle aree chiare e nelle aree scure l'area sottesa viene sommata o sottratta rispettivamente.

Per ulteriori informazioni si rimanda alla documentazione di Cascade Classifier [13].

## 9.2 Allarme visivo/sonoro

Immagine che contiene fotografia, schermo, sedendo, uomo

Descrizione generata automaticamenteImmagine che contiene monitor, interni, fotografia, schermo

Descrizione generata automaticamente

Figura 36 Allarme sonoro/visivo.

Come accennato precedentemente, dopo un intervallo temporale predefinito l’utente viene avvisato tramite un allarme sia visivo che sonoro. Tale allarme permette al conducente di svegliarsi nel caso in cui stesse dormendo ed evitare un eventuale incidente.

## 9.3 Acquisizione con camera IR

Di seguito sono riportate le acquisizioni effettuate con una camera Infrared per simulare la guida del veicolo in assenza di illuminazione.

Immagine che contiene persona, fotografia, interni, tenendo

Descrizione generata automaticamenteImmagine che contiene fotografia, tenendo, monitor, schermo

Descrizione generata automaticamente

Figura 37 Acquisizioni effettuate con camera IR.

Da notare, che nonostante la bassa qualità della camera infrared a disposizione, il sistema riesce comunque a rilevare un’eventuale situazione di pericolo avvertendo l’utente con un allarme sonoro. Chiaramente, si è riusciti ad ottenere un tale risultato poiché all’interno del dataset sono presenti anche immagini di scarsa qualità che hanno fatto si che il modello risulti essere piuttosto robusto a variazioni di illuminazione e di qualità dell’immagine acquisita.

## 9.4 Schermata impostazioni

Il menu impostazioni della demo realizzata è mostrato nell’immagine seguente.

**Immagine che contiene screenshot

Descrizione generata automaticamente**

**1**

**2**

Figura 38 Impostazioni.

La schermata mostrata in Figura 38 offre le seguente opzione:

**1**

Possibilità di modificare la soglia del riconoscimento degli occhi. Se gli occhi non vengono riconosciuti vi è la possibilità di abbassare la soglia, mentre nel caso in cui si voglia essere più selettivi ed evitare il più possibile false alarms è possibile alzare tale soglia.

**2**

Per confermare le modifiche effettuate è necessario cliccare la spunta verde.

# 10. Conclusioni

# 11. Riferimenti

**[1]** Global status report on road safety 2018. Geneva: World Health Organization; 2018.

**[2]** M. J. Flores, J. M. Armingol and A. Escalera,” Real-time warning system for driver drowsiness detection Using Visual Information”, Journal of Intelligent & Robotic Systems, Volume 59, Issue 2, pp 103-125, August 2010.

**[3]** S. Vitabile, A. D. Paola and F. Sorbello, “A real-time non- intrusive FPGA-based drowsiness detection system”, Journal of Ambient Intelligence and Humanized Computing, Volume 2, Issue 4, pp 251-262, December 2011.

**[4]** M. Singh, G. Kaur,” Drowsiness detection on eye blink Duration using algorithm”, International Journal of Emerging Technology and Advanced Engineering, Volume 2, Issue 4, April 2012.

**[5]** C. W. You, N. D. Lane, F. Chen, R. Wang, Z. Chen, T. J. Bao, M. Montes-de-Oca, Y. Cheng, M. Lin, L. Torresani and A. T. Campbell, “CarSafe App: Alerting Drowsy and Distracted Drivers using Dual Cameras on Smart phones”, In MobiSys’13, June 25-28, 2013.

**[6]** Sahayadhas, K. Sundaraj and M. Murugappan,” Drowsiness detection during different times of day using multiple feature”, Australasian Physical & Engineering Sciences in Medicine, Volume 36, Issue 2, pp 243-250, June 2013.

**[7]** A. Rahman, M. Sirshar and A. Khan, "Real time drowsiness detection using eye blink monitoring," 2015 National Software Engineering Conference (NSEC), Rawalpindi, 2015, pp. 1-7, doi: 10.1109/NSEC.2015.7396336.

**[8]** Fusek R. (2018) Pupil Localization Using Geodesic Distance. In: Bebis G. et al. (eds) Advances in Visual Computing. ISVC 2018. Lecture Notes in Computer Science, vol 11241.

**[9]** https://data-flair.training/blogs/python-project-driver-drowsiness-detection-system/

**[10]** Caruccio, Loredana & Deufemia, Vincenzo & Polese, Giuseppe. (2016). Relaxed Functional Dependencies - A Survey of Approaches. IEEE Transactions on Knowledge and Data Engineering. 28. 147-165. 10.1109/TKDE.2015.2472010.

**[11]** L. Caruccio, V. Deufemia, F. Naumann and G. Polese, "Discovering Relaxed Functional Dependencies based on Multi-attribute Dominance," in IEEE Transactions on Knowledge and Data Engineering, doi: 10.1109/TKDE.2020.2967722.

**[12]** https://opencv.org

**[13]** <https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html>

1. Segnale elettrooculogramma che viene registrato per individuare i movimenti degli occhi utili nella stadiazione del [sonno](https://it.wikipedia.org/wiki/Sonno). La registrazione si basa sulla differenza di potenziale esistente tra [cornea](https://it.wikipedia.org/wiki/Cornea) (positiva) e [retina](https://it.wikipedia.org/wiki/Retina) (negativa) e viene effettuata mediante [elettrod](https://it.wikipedia.org/wiki/Elettrodi)i. [↑](#footnote-ref-1)
2. È un descrittore di caratteristiche usate in computer vision ed in elaborazione delle immagini per il riconoscimento di oggetti. [↑](#footnote-ref-2)
3. Hierarchical Data Format versione 5 (HDF5), è un formato di file open source che supporta dati di grandi dimensioni, complessi ed eterogenei. [↑](#footnote-ref-3)