**Sistemi Cooperativi**

**ERYTHEMA INDEX E LIVELLI DI EMOGLOBINA**

Altamura Nicola, Giglio Valentina Carmela Rita

Università degli studi di Bari Aldo Moro, Piazza Umberto I, 1, Bari 70121, Italy

[n.altamura2@gmail.com](mailto:n.altamura2@gmail.com)

Università degli studi di Bari Aldo Moro, Piazza Umberto I, 1, Bari 70121, Italy

[v.giglio1@studenti.uniba.it](mailto:v.giglio1@studenti.uniba.it)

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**LINK**: <https://github.com/altamuran/HB_prediction?fbclid=IwAR1qdm8uD-AKOvTAsUOh-7_CBg-4zmnopwRLpN5d7_Y6TbXgLfWrsMV93Lo>

**Introduzione**

Lo scopo di questo caso di studio consiste nel verificare l’esistenza di una correlazione tra il valore *erythema index* ottenuto, attraverso opportuni calcoli su frame estratti dall’elaborazione di un piccolo filmato video, ed il livello di emoglobina di soggetti sottoposti al test.

L’ erythema index è un indice empirico che combina alcune immagini monocromatiche a varie lunghezze d'onda sensibili alle variazioni della concentrazione sanguigna.

L’emoglobina è un pigmento dei globuli rossi del sangue, mediante il quale si compie il trasporto dell'ossigeno dai polmoni ai tessuti e dell'anidride carbonica dai tessuti ai polmoni. I valori minimi di riferimento per una persona in salute sono 12.0 g/dL per le donne e 13,0 g/dL per gli uomini. Essendo il polpastrello un tessuto particolarmente vascolarizzato risulta particolarmente interessante analizzare la quantità di colore rosso di un video registrato a diretto contatto col tessuto.

**Tecnologie utilizzate**

Questo progetto è stato sviluppato in Python, un linguaggio di programmazione dinamico orientato agli oggetti utilizzabile per molti tipi di sviluppo software. Offre un forte supporto all'integrazione con altri linguaggi e programmi, è fornito di una estesa libreria standard e può essere imparato in pochi giorni

La scomposizione dei video in frame è stata effettuata mediante l’utilizzo delle funzioni implementate in OpenCV (acronimo in lingua inglese di Open Source Computer Vision Library), una libreria software multipiattaforma nell'ambito della visione artificiale in tempo reale.

È una libreria software libera originariamente sviluppato da Intel, centro di ricerca in Russia di Nižnij Novgorod. Successivamente fu poi mantenuto da Willow Garage e ora da Itseez.

Il linguaggio di programmazione principalmente utilizzato per sviluppare con questa libreria è il C++, ma è possibile interfacciarsi anche attraverso il C, Python e Java.

**Raccolta dati**

Per poter rendere significativo l’esperimento è stato selezionato un campione di persone casuale con analisi valide (cioè effettuate negli ultimi 6 mesi), in modo da poter confrontare il livello di emoglobina registrato con il valore erythema index calcolato mediante l’elaborazione del video.

La registrazione del video è stata realizzata secondo le seguenti modalità: innanzitutto viene lanciata l’applicazione per la registrazione del video, viene impostato il flash in modalità on, si posiziona l’indice destro sulla fotocamera in modo da coprila completamente, infine si fa partire la registrazione per una durata di 30 secondi. Tutti i video sono stati realizzati con lo stesso cellulare: Honor 10 view.

Di seguito, i dettagli relativi ai dati raccolti:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sesso | Età | Scala Fitzpatrick | Emoglobina |
| Persona1 | M | 25 | Tipo II | 14,5 |
| Persona2 | F | 24 | Tipo III | 12,5 |
| Persona3 | M | 24 | Tipo II | 14,9 |
| Persona4 | F | 62 | Tipo II | 12,6 |
| Perona5 | M | 64 | Tipo II | 14,1 |
| Persona6 | F | 61 | Tipo III | 13,6 |
| Persona7 | F | 16 | Tipo III | 12,7 |
| Persona8 | F | 96 | Tipo III | 11,7 |
| Persona9 | F | 47 | Tipo III | 13,0 |
| Persona10 | M | 26 | Tipo II | 15,2 |
| Persona11 | M | 30 | Tipo III | 12,6 |
| Persona12 | F | 24 | Tipo II | 12,4 |
| Persona13 | M | 24 | Tipo III | 16,1 |
| Persona14 | F | 59 | Tipo II | 14,5 |
| Persona15 | M | 22 | Tipo III | 15.7 |

**Elaborazione dati**

Ogni video è stato scomposto in frame, ed è stato ridimensionato ad un massimo di 800 frame in modo da ridurre gli errori commessi durante la raccolta dei dati (anche un decimo di secondo in più avrebbe potuto rendere i valori calcolati lievemente diversi). Successivamente, ogni frame è stato suddiviso nei relativi canali RGB.

Il modello RGB è un modello additivo e crea il colore avvalendosi di una sorgente luminosa. È basato sui tre colori Rosso (Red), Verde (Green) e Blu (Blue). Tali colori base, miscelati tra loro, coprono quasi tutto lo spettro dei colori visibili. In questo modello i colori secondari sono creati mediante somma dei colori Rosso, Verde e Blu che sono detti colori primari. Il colore bianco, invece, è generato dalla somma dei tre colori primari nella loro massima intensità. Infine, le tonalità dei grigi vengono create con la somma di valori identici di Rosso, Verde e Blu con intensità decrescenti fino allo zero.

Attraverso la scomposizione nei colori base del modello, è stato possibile calcolare il valore medio del rosso e del verde, ciò ha reso possibile calcolare il valore dell’erythema index utilizzando la formula seguente formula:

.

Questo valore è stato calcolato mediante l’elaborazione di ogni singolo frame del video, ed è stato relazionato con il relativo livello di emoglobina del soggetto prestatosi all’esperimento.

Si è cercato di dar senso alla relazione tra il valore erythema index, così calcolato, ed il livello di emoglobina, è stato quindi creato un modello di regressione lineare.

Per poter stabilire il tipo di correlazione abbiamo scelto di utilizzare l‘indice di Pearson, dove la variabile indipendente è il valore da noi calcolato e la variabile dipendete è il valore di emoglobina.

In statistica, tale coefficiente tra due variabili statistiche esprime un'eventuale relazione di linearità tra esse.

Date due variabili statistiche X e Y, l'indice di correlazione di Pearson è definito come la loro covarianza divisa per il prodotto delle deviazioni standard delle due variabili, cioè:

Dove è la covarianza e , sono le due deviazioni standard.

Il coefficiente assume sempre valori compresi tra -1 e 1:

Nella pratica si distinguono vari "tipi" di correlazione:

* Se >0, le variabili X e Y si dicono direttamente correlate, oppure correlate positivamente;
* se =0, le variabili X e Y si dicono non correlate;
* se <0, le variabili X e Y si dicono inversamente correlate, oppure correlate negativamente.

Inoltre, per la correlazione diretta (e analogamente per quella inversa) si distingue:

* se <0,3 si ha correlazione debole;
* se <0,7 si ha correlazione moderata;
* se >0,7 si ha correlazione forte.

È stato, poi, calcolato il valore R2 (coefficiente di determinazione) per capire, invece, quanto la variabile indipendente influisca sulla variabile dipendente, ed è stato, quindi, tracciato un grafico che rappresenta la distribuzione dei valori (asse x e asse y) e tracciato la retta di regressione.

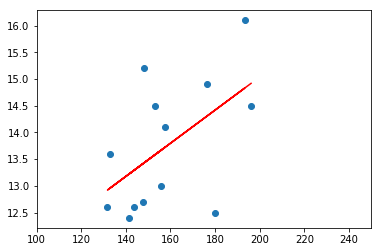
In statistica, il coefficiente di determinazione, è una proporzione tra la variabilità dei dati e la correttezza del modello statistico utilizzato. Esso misura la frazione della varianza della variabile dipendente espressa dalla regressione. Nelle regressioni lineari semplici, quindi nel nostro caso, esso è semplicemente il quadrato del coefficiente di correlazione.

La regressione, invece, formalizza e risolve il problema di una relazione funzionale tra variabili misurate sulla base di dati campionari estratti da un'ipotetica popolazione infinita. Più formalmente, in statistica la regressione lineare rappresenta un metodo di stima del valore atteso condizionato di una variabile dipendente, Y, dati i valori di altre variabili indipendenti, X.

Sono stati, quindi, recuperati i valori di intercetta e pendenza della retta di regressione, ed è stato utilizzato questo modello per cercare di stimare un valore di emoglobina partendo da un nuovo video.

**Risultati**

L’elaborazione dei dati ci ha restituito risultati di interesse rilevante, soprattutto considerando la scarsità di dati.

È stato tracciato un grafico che mostra la relazione dei valori erythema index con i relativi livelli di emoglobina. Tuttavia, a causa della carenza dei dati si è ritenuto utile eliminare il valore out-line della Persona8 a causa dell’età avanzata. 

I valori erythema index calcolati sono i seguenti:

|  |  |  |
| --- | --- | --- |
|  | Erythema Index | Emoglobina |
| Persona1 | 196.10117389322875 | 14.5 |
| Persona2 | 180.02372418619768 | 12.5 |
| Persona3 | 176.52849576822894 | 14.9 |
| Persona4 | 131.76943532986118 | 12.6 |
| Persona5 | 157.72034331597223 | 14.1 |
| Persona6 | 132.84356701388893 | 13.6 |
| Persona7 | 147.5868474361795 | 12.7 |
| Persona9 | 155.80778801716556 | 13.0 |
| Persona10 | 148.1133696614583 | 15.2 |
| Persona11 | 143.8177319824218 | 12.6 |
| Persona12 | 141.4577162873642 | 12.4 |
| Persona13 | 193.26621455078129 | 16.1 |
| Persona14 | 153.16638746178648 | 14.5 |

Da questi dati è stato riscontrato un coefficiente di correlazione uguale a 0.546, che vuol dire che è stata riscontrata una correlazione moderata.

È stata calcolata la retta di regressione. Ricordiamo la formula dove è l’intercetta e rappresenta la pendenza. Nel caso preso in esame la pendenza è uguale alla variazione dell’emoglobina causata dalla variazione unitaria del valore erythema index da noi calcolato. Infine, riportiamo il risultato della pendenza calcolata: 0.031.

Si è ritenuto utile, quindi, costruire una funzione di *“predizione”* che ci permette di ricavare una stima del valore di emoglobina inserendo il valore indice ricavato dall’elaborazione dei video. Per esempio, inserendo un valore *EI* di 130 è stimato un valore di emoglobina 12.87.

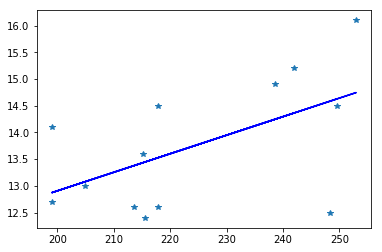
Infine, il coefficiente di discriminazione, che rappresenta la bontà di adattamento ci fornisce un valore di 0.29, ciò significa che il nostro modello riesce a spiegare il trenta percento circa della relazione tra le variabili prese in esame. Questo valore risulta positivo soprattutto vista la quantità limitata dei dati da analizzare.

**Approfondimenti**

1. **Modello unidimensionale RED**

Sono stati utilizzati gli stessi dati e lo stesso tipo di elaborazione dei video per poter correlare il canale R del modello RGB con il valore di emoglobina dei soggetti sottoposti al test. Si è voluto realizzare uno studio di approfondimento su questa relazione poiché nella tipologia dei video realizzati appare subito evidente la predominanza del colore rosso.

In questo caso il valore del coefficiente di correlazione ci mostra un lieve miglioramento: 0.555. Questo miglioramento appare evidente dal grafico in cui viene tracciata la retta di regressione come fatto precedentemente:



Anche per questo tipo di relazione è stata creata una funzione di “*predizione*” che con un valore R=210 ci fornisce un valore di emoglobina stimato di 13.25.

Di rilevante interesse è anche il risultato ottenuto dal coefficiente di discriminazione che ci mostra un piccolo miglioramento rispetto al tipo di elaborazione precedente, infatti il valore risultante è di 0.31.

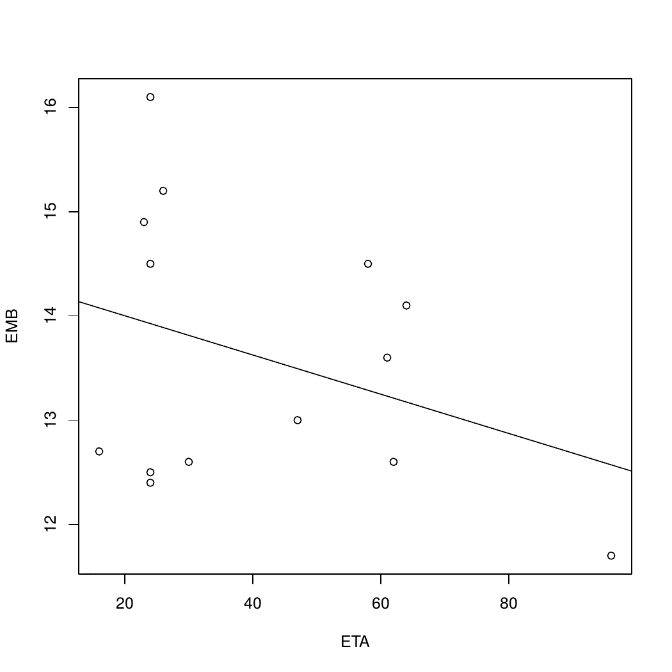
1. **Altri modelli unidimensionali e multidimensionali**

Sono stati sviluppati modelli multidimensionali in codice R per poter mostrare eventuali correlazioni tra gli altri dati rilevati.

Si è voluto quindi prendere in considerazione il colore della pelle, il valore di emoglobina, il sesso, l’età e il valore Erythema Index (calcolato come precedentemente descritto).

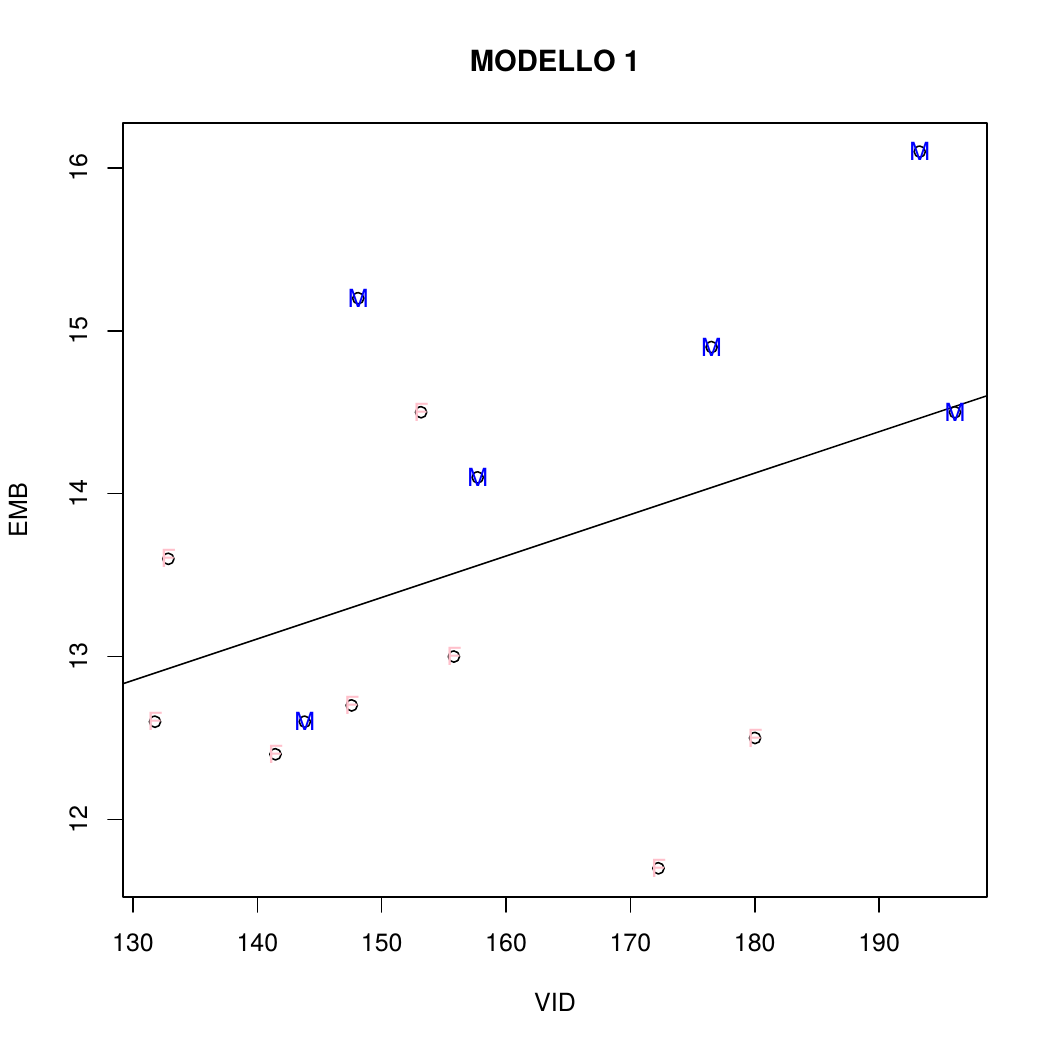
1. Età-emoglobina

È stato creato un primo modello che ha calcolato la correlazione tra i valori di emoglobina e l’età, ma questo modello non è stato ritenuto significativo a seguito dei valori del coefficiente di correlazione e di discriminazione.



1. EI-Emoglobina-Sesso

È estato sviluppato un secondo modello multidimensionale con lo scopo di calcolare la correlazione tra i valori di emoglobina e i valori Erythema Index calcolati distinguendo il sesso dei soggetti sottoposti all’esperimento, ciò è stato possibile poiché la variabile sesso è una variabile dicotomica.



In questo caso esiste una relazione positiva e con una buona capacità di predizione.

1. Emoglobina-Pelle

È stato creato un modello con lo scopo di relazione il colore della pelle con il valore di emoglobina, tuttavia nei dati raccolti, si è dimostrato una correlazione non significativa.

1. Emoglobina-sesso

È stato prodotto un modello con lo scopo di relazionare il valore di emoglobina ed il sesso, ovviamente i valori sono risultati molto correlati poiché anche gli standard di riferimento sono differenti.