Waste Classification

Andrea Sorgente, 830483 Agostino Tassan Mazzocco, 833933

Gianluca Borchielli, 833003

5 Maggio, 2022

Abstract

Spinti dall'interesse per l'argomento e dall'alta rilevanza e importanza che l'attenzione ambientale ha attualmente, questo studio si pone l'obiettivo di strutturare adeguate reti neurali capaci di classificare correttamente la tipologia di rifiuto: organico oppure riciclabile. Il dataset e le relative informazioni sono state ottenute su kaggle. Per il raggiungimento dell'obiettivo è stato fatto ricorso a svariate tecniche di Deep Learning, tra cui strategie per evitare l'overfitting e modelli di reti neurali più articolate e ad hoc per il dataset analizzato. I risultati sono stati particolarmente incoraggianti, consentendoci di fare importanti considerazioni sul tipo di dataset preso in analisi.

1. Introduzione

Oggi più che mai l'attenzione all'ambiente e al corretto smaltimento dei rifiuti è di alta rilevanza. I rifiuti producono inquinamento: liquami, gas, sostanze tossiche e materiali non biodegradabili che possono inquinare aria, acqua, terra.
I rifiuti costano: rubano spazio e occorrono risorse umane ed economiche per il loro trattamento, ma anche per rimediare ai danni ambientali e sanitari che producono. Affinché l'impatto ecologico di ciascun cittadino sia il minore possibile, tanti sono stati i provvedimenti presi nel tempo dalle autorità competenti: la raccolta differenziata è uno di questi.

Abbiamo pensato di affrontare questa tematica sfruttando il nostro campo professionale di conoscenza, da qui l'elaborato che, sulla base di quanto appreso in materia di Deep Learning, pone il focus sulla classificazione di rifiuti di vario

genere etichettati come organici oppure come riciclabili. Le reti neurali, in particolare quelle convoluzionali, sono le grandi protagoniste del lavoro in oggetto, grazie alle quali si sono ottenuti ottimi risultati in termini di accuracy di classificazione automatizzata e ottimi spunti di riflessione. L'utilizzo di tecniche come la Data Augmentation e l'utilizzo di reti pre-allenate hanno maggiormente contribuito a fornire risultati di alta qualità, infatti è proprio la rete VGG-16 a candidarsi in questo elaborato per essere il modello finale performando sul test set con un'accuracy di poco superiore al 90%.

2. Materiali e metodi

2.1 Materiali

Il dataset analizzato è stato messo a disposizione sul sito kaggle da Sashaank Sekar nel 2019 ma le sue fonti originarie sono Google e ImageNet. Quest'ultima è un'ampia base di dati di immagini realizzata per l'applicazione di tecniche di deep learning finalizzate al riconoscimento di oggetti. Il dataset completo consiste in più di 14 milioni di immagini con modello di colori RGB (Red-Green-Blue) annotate manualmente con l'indicazione degli oggetti in esse rappresentati classificabili in più di 20.000 categorie. Il download del dataset di kaggle prevedeva invece la seguente composizione: 25.077 immagini in formato .jpeg, divise in 22.564 immagini nel training set e 2.513 nel test set, denominate col prefisso dell'etichetta di risposta (dicotomica) O oppure R (Organico o Riciclabile) e con suffisso numerico anche se in fase di pre-processing si è notato che 2 immagini (la O_202 e la O_4430) sono mancanti. La classe dei rifiuti organici è leggermente più presente della classe dei rifiuti riciclabili sia nel training sia nel test set. Inoltre risalta immediatamente all'occhio che le immagini hanno diversa dimensione e che la composizione di training e test set non è randomica, bensì sicuramente a partire da un unico insieme di immagini è stato fatto un taglio netto dividente in due i dataset, a causa del quale si verifica la presenza di immagini molto simili tra loro e raffiguranti lo stesso oggetto solo nel training o solo nel test set.

2.2 Metodi

2.2.1 Data Splitting e Pre-processing

Come emerso nella sezione 'Materiali', la divisione del dataset non è soddisfacente da un punto di vista statistico, ovvero gli oggetti che li compongono non sono stati evidentemente selezionati mediante un sample casuale, inoltre la

mancata considerazione di un validation set ha reso necessaria una modifica nella fase di preparazione dei dati: si è proceduto a riunire tutte le immagini in un unico insieme per poi creare i set di training, validation e test composti da immagini campionate casualmente riuscendo ad ovviare anche al problema, non molto rilevante ed accentuato in verità, del non perfetto bilanciamento delle classi. Infatti il training set si compone di 16.000 immagini, di cui l'esatta metà di rifiuti organici e l'altra metà di rifiuti riciclabili (8.000 immagini per etichetta); validation e test set invece si compongono rispettivamente di 4.000 e 2.000 immagini e anch'essi sono equamente suddivisi rispetto alle due classi.

Successivamente, per sopperire al problema delle immagini con lunghezza e larghezza molto variabili, si è dovuto ridimensionarle a misura fissa di 224x224 (dimensione idonea per applicare la rete pre-allenata VGG-16). Dopodiché le immagini sono state opportunamente trasformate dal formato jpeg a RGB (griglia di pixels) per poi convertire il risultato in un tensore di dimensione (16.000, 224, 224, 3) per il training set; per validation e test set i tensori sono rispettivamente (4.000, 224, 224, 3) e (2.000, 224, 224, 3) poiché l'ampiezza campionaria è diversa. Infine i pixels, i cui valori variavano nel range [0,255], sono stati normalizzati (il nuovo range risulta [0,1]).

2.2.2 Feed Forward Neural Network

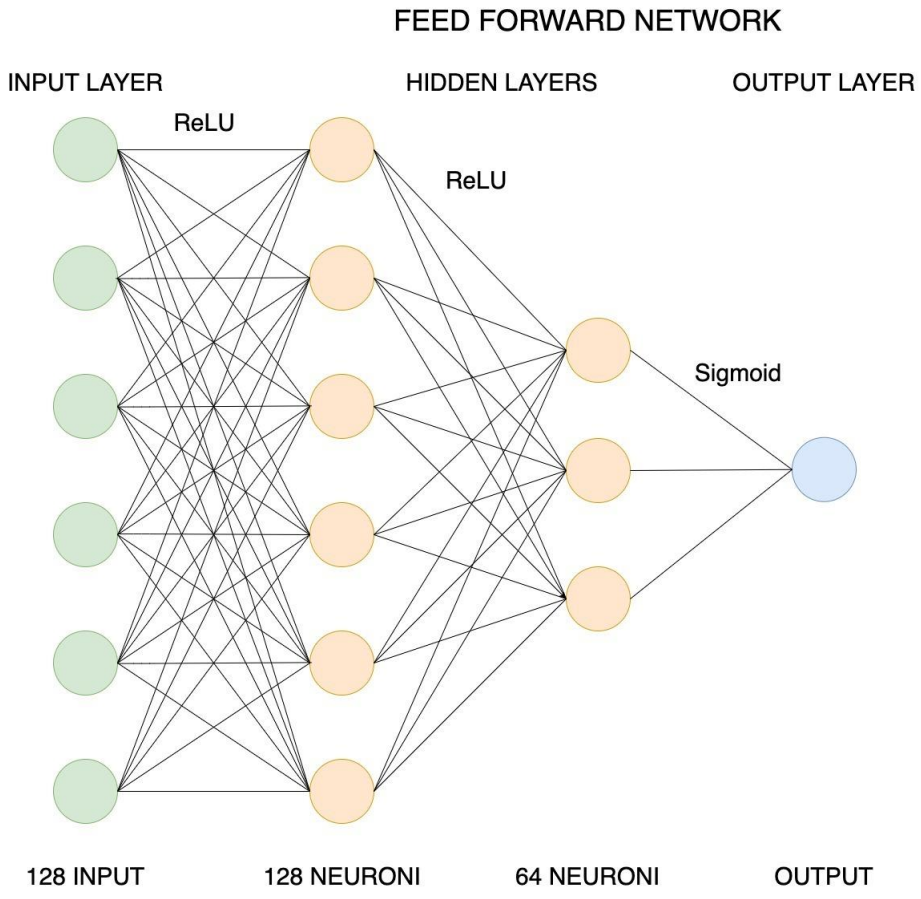
Le reti neurali FFN addestrate con l'algoritmo di apprendimento della back-propagation sono le reti neurali più semplici. Tale tipologia di rete è costituita da neuroni ordinati all'interno dei layers, il primo dei quali è denominato input-layer, l'ultimo output-layer e gli strati intermedi si dicono hidden.

La Feed Forward è caratterizzata dalle seguenti proprietà:

- L'informazione proveniente dall'input layer si propaga in un'unica direzione, in avanti, attraverso gli hidden layers fino ai neuroni di output;
- La connessione tra un layer e i neuroni del layer successivo consiste nella combinazione lineare degli output dei neuroni del primo strato pesati con l'aggiunta del coefficiente di bias a cui viene applicata la funzione di attivazione;

- Il coefficiente di peso di un neurone rispetto ad un altro dello strato successivo riflette il grado di importanza della loro connessione all'interno della rete neurale.
- Esistono differenti funzioni di attivazione tra cui la funzione sigmoidale, la Tanh, step function e la Relu.

Nel nostro lavoro si è considerata una feed forward composta da un hidden layer da 128 neuroni con funzione di attivazione "ReLU", un secondo layer da 64 neuroni con funzione di attivazione "ReLU" e l'output layer caratterizzato da una funzione di attivazione "sigmoid", in quanto si tratta di un problema di classificazione binario. Il modello viene compilato impostando l'ottimizzatore "rmsprop" con parametro di learning rate pari a 0.0001, la loss function "binary crossentropy" e come metrica di valutazione l'accuratezza della classificazione ed è consultabile graficamente nell'immagine seguente.



Inoltre grazie all'aggiunta del criterio di early stopping l'allenamento della rete si interromperà quando per 5 epoche consecutive l'accuracy del validation set non aumenterà di almeno 0.005. L'allenamento della rete durerà comunque al massimo 30 epoche, con 125 step per ogni epoca sui set di training e validation.

2.2.3 Convolutional Neural Network

La rete neurale convoluzionale (CNN) è una delle architetture di deep learning più popolari, capace di rilevare automaticamente le caratteristiche salienti delle immagini senza alcuna supervisione umana risultando estremamente efficiente dal punto di vista computazionale. È composta da tre livelli di base: layer di convoluzione, layer di sotto campionamento e layer denso. I layer di convoluzione e di sotto campionamento vengono ripetuti a seconda dell'applicazione e infine vengono collegati al layer denso. Il layer di convoluzione esegue l'operazione da cui esso prende il nome, la convoluzione appunto, sulle immagini del dataset mediante un set di filtri. Una volta eseguiti tutti gli strati di convoluzione e sotto campionamento, seguono i layer densi in cui viene utilizzata una funzione di attivazione non lineare chiamata ReLU, la quale converte tutti i valori negativi in zero. Il layer di sotto campionamento si pone l'obiettivo di rendere la rete più robusta e invariante e di ridurre la dimensione dell'immagine mediante diverse tecniche, la cui più comune è quella di maxpooling. Tale tecnica utilizza il valore massimo di un intorno di pixels per rappresentare l'intorno stesso. Lo strato di convoluzione finale viene appiattito attraverso il layer flatten ed è collegato allo strato successivo fino ad arrivare allo strato ultimo di classificazione nel quale viene usata la funzione di attivazione sigmoidale.

La rete convoluzionale considerata nell'elaborato è composta da 3 layers convoluzionali composti rispettivamente da 32, 64 e 128 filtri di dimensione 3x3, alternati con altrettanti layers di max pooling di dimensione 2x2, infine sono stati

aggiunti 2 layers densi con rispettivamente 128 e 64 neuroni e un output layer. Le funzioni di attivazione utilizzate sono la “ReLU” e la “sigmoid” per l’output layer. Anche per questa rete in fase di allenamento è stata scelta una loss function di tipo binary crossentropy, un ottimizzatore rmsprop con parametro di learning rate pari a 0.0001 ed è stato applicato il criterio di early stopping che interromperà l’apprendimento qualora l’accuracy della classificazione sul validation set non aumenti di almeno 0.005 per 5 epoche consecutive; l’apprendimento non supererà comunque le 30 epoche.

2.2.4 Convolutional Neural Network con Data Augmentation

La Data Augmentation è un metodo efficace per ridurre l’overfitting, che generalmente si riscontra in una CNN profonda quando i campioni di addestramento sono ridotti, approssimando lo spazio di probabilità dei dati mediante la manipolazione dei campioni di input sfruttando capovolgimenti orizzontali, ritagli casuali e trasformazioni di scala. In generale, aumentando la quantità, la qualità e la diversità dei dati presenti nel dataset, aumenta conseguentemente anche l’efficacia del modello utilizzato. Un particolare tipo di Data Augmentation è l’abbinamento dei campioni, il quale è semplice e sorprendentemente efficace per classificare le immagini: si tratta di creare una nuova immagine a partire da una originale alla quale va sovrapposta un’altra immagine prelevata casualmente dal training set. Il Dropout è un altro metodo di data augmentation che aggiunge un disturbo all’input space in maniera randomica.

Sebbene i metodi descritti sono in grado di migliorare la generalizzazione del modello, non è ancora stato analizzato rigorosamente l’impatto che il disturbo ha sul decision boundary.

La rete neurale costruita in questo elaborato ha la stessa struttura della rete convoluzionale precedente, con la sola aggiunta del dropout per tentare di rimediare al possibile fenomeno di overfitting.

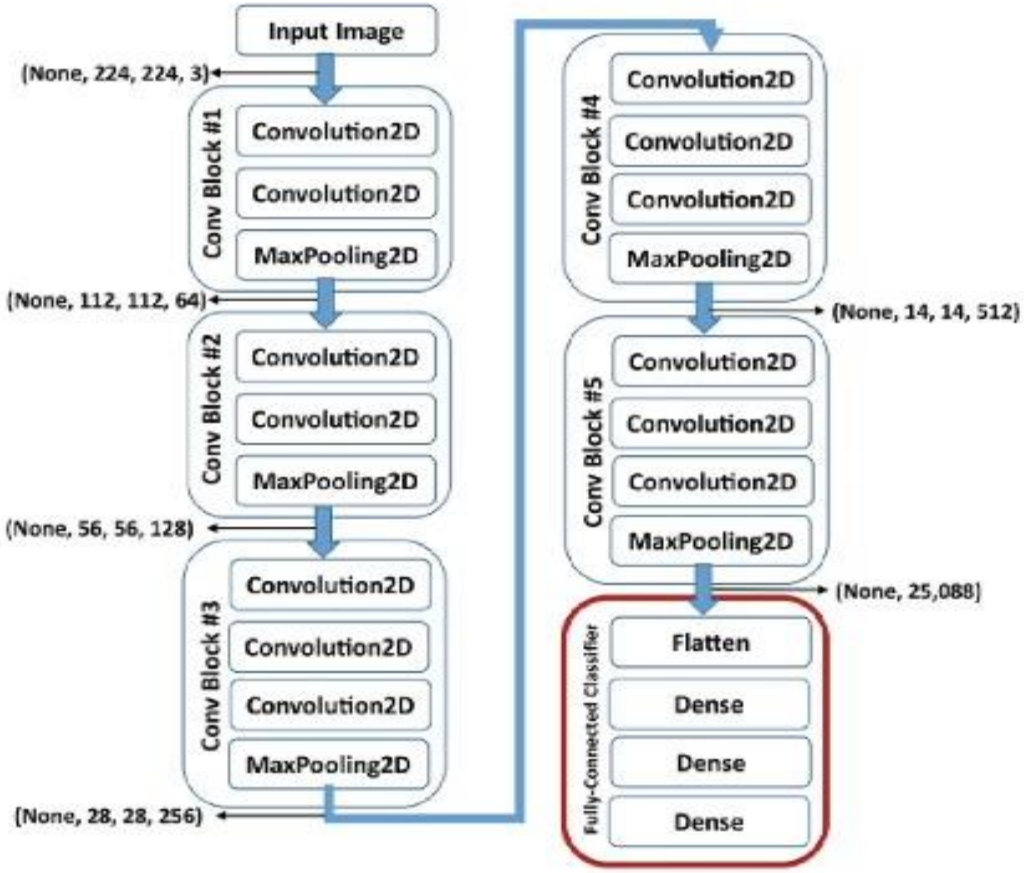
2.2.5 VGG-16 Neural Network

VGG-16 è una rete CNN profonda che è stata pubblicata nel 2015. La rete VGG-16 è stata addestrata sul database ImageNet ed è per questo che ci si aspetta un buon risultato finale. Poiché tale rete è stata sottoposta ad un ampio addestramento fornisce un'eccellente precisione anche quando i dataset delle immagini risultano essere ridotti.

La rete VGG-16 è composta da 16 layer di convoluzione con filtri di dimensione 3×3 e 5 strati di maxpooling di dimensioni 2×2 come viene mostrato nella seguente figura. Infine ci sono 3 layer densi dopo l'ultimo livello di maxpooling.

In generale essa può essere usata per fare features extraction oppure fine tuning. La prima consiste nell’utilizzare le rappresentazioni apprese dalla rete pre-allenata ed estrarre nuove caratteristiche da altri campioni, le quali sono utilizzate in un classificatore che deve apprendere nuovamente dal training set. La seconda consiste nel considerare alcuni strati profondi della rete convoluzionale per costruirne una nuova composta da questi strati e altri fully connected che verranno aggiunti alla fine. La rete così strutturata riceverà in input le immagini del training set al fine di aggiornare i pesi riguardanti la rete pre-allenata.

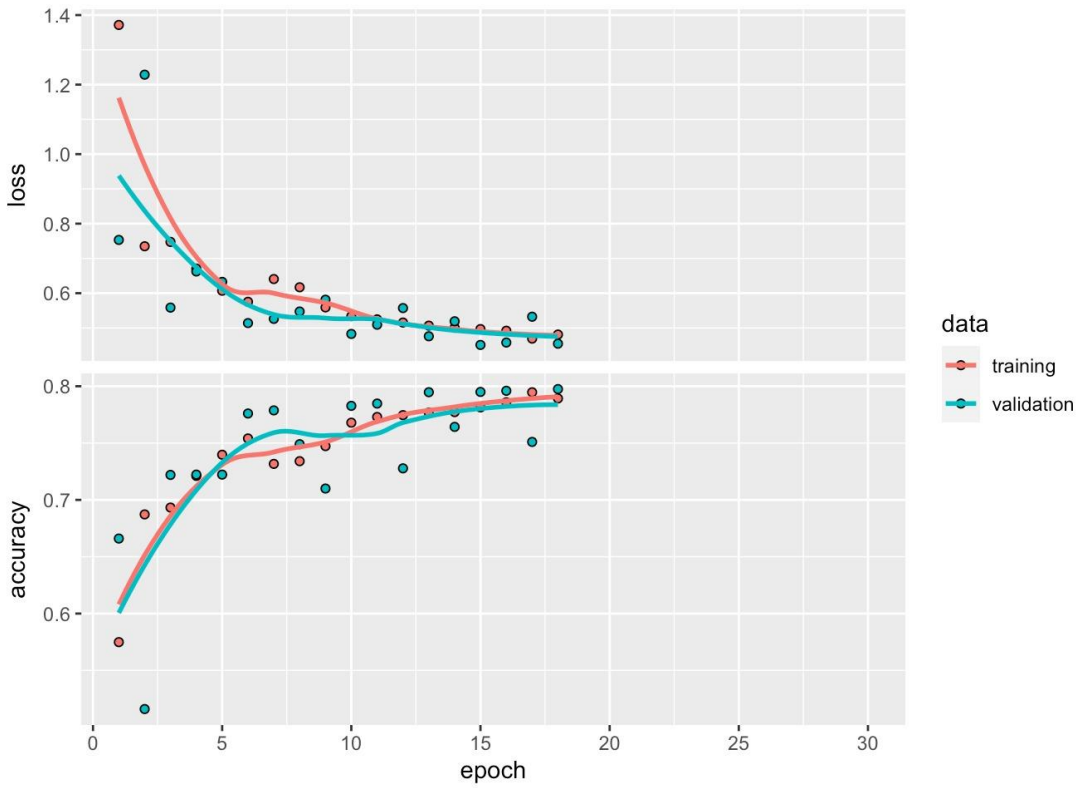
Nel nostro elaborato è stata applicata la features extraction.



3. Risultati

3.1 Feed Forward Neural Network

Applicando la Feed Forward Network sui nostri dati otteniamo il seguente grafico che mette in relazione accuracy e funzione di perdita all'aumentare delle epoche, considerando il training set (curva rossa) e il validation set (curva azzurra).



È possibile notare che l'algoritmo si arresta in corrispondenza della 18-esima epoca a causa del criterio di early stopping impostato in fase di addestramento.

La seguente rete raggiunge quindi i seguenti livelli di accuracy:

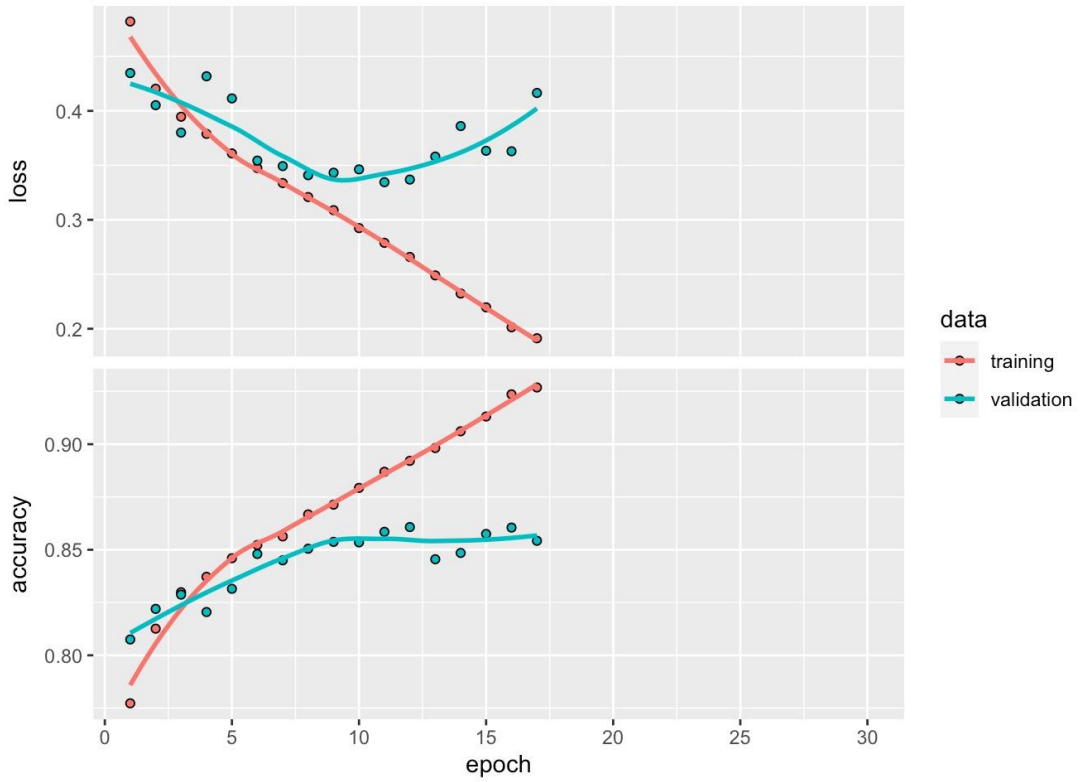
- TRAINING SET: 0.7893
- VALIDATION SET: 0.7975

Per quanto concerne i valori della funzione di perdita otteniamo:

- TRAINING SET: 0.4810
- VALIDATION SET: 0.4548

3.2 Convolutional Neural Network

Applicando la Convolutional Neural Network ai nostri dati otteniamo il seguente grafico.



In questo caso l'algoritmo si arresta alla 17-esima epoca per via dell'early stopping.

È possibile notare come il livello di accuracy del training set continui ad aumentare a differenza dell'accuracy relativa al validation set, la quale rimane pressoché costante a partire dalla 9° epoca, ne segue che siamo in presenza di overfitting.

Tale affermazione viene confermata osservando il grafico relativo alla funzione di perdita. In questo caso si nota che, facendo riferimento al training set, i valori diminuiscono in modo inversamente proporzionale all'aumentare delle epoche; diversamente per quanto riguarda il comportamento ottenuto considerando il validation set si nota come a partire dalla 7° epoca il livello della funzione di perdita risulta avere un trend positivo, aumenta quindi il valore di perdita.

Otteniamo quindi i seguenti valori finali di accuracy:

- TRAINING SET: 0.9269
- VALIDATION SET: 0.8543

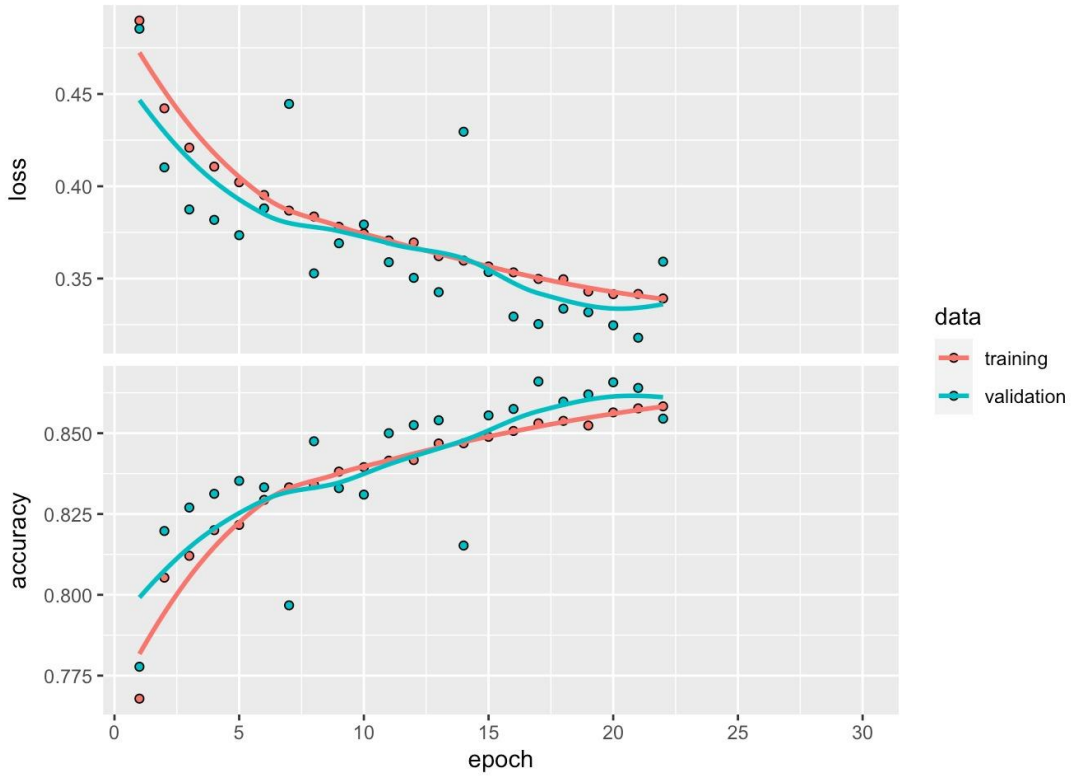
I valori della funzione di perdita risultano essere:

- TRAINING SET: 0.1913

- VALIDATION SET: 0.4161

3.3 Convolutional Neural Network con Data Augmentation

Utilizzando la Convolutional Neural Network con Data Augmentation otteniamo il seguente grafico.



L'algoritmo di apprendimento si ferma alla 22-esima epoca per via dell'early stopping.

Possiamo notare come tale tecnica utilizzata nelle reti convoluzionali risolva il problema di overfitting; infatti osservando il grafico si nota che sia il livello di accuracy che il livello di loss risultano avere un comportamento analogo sia sui dati di training che sui i dati di validation. Inoltre è evidente che aumentando la quantità, la qualità e la diversità dei dati presenti nel dataset, aumenta conseguentemente anche l'efficacia del modello utilizzato. Ciò è riscontrato dal fatto che il livello di accuracy del validation set sia leggermente più alto rispetto a quello relativo alle immagini presenti nel training.

I valori finali di accuracy risultano essere:

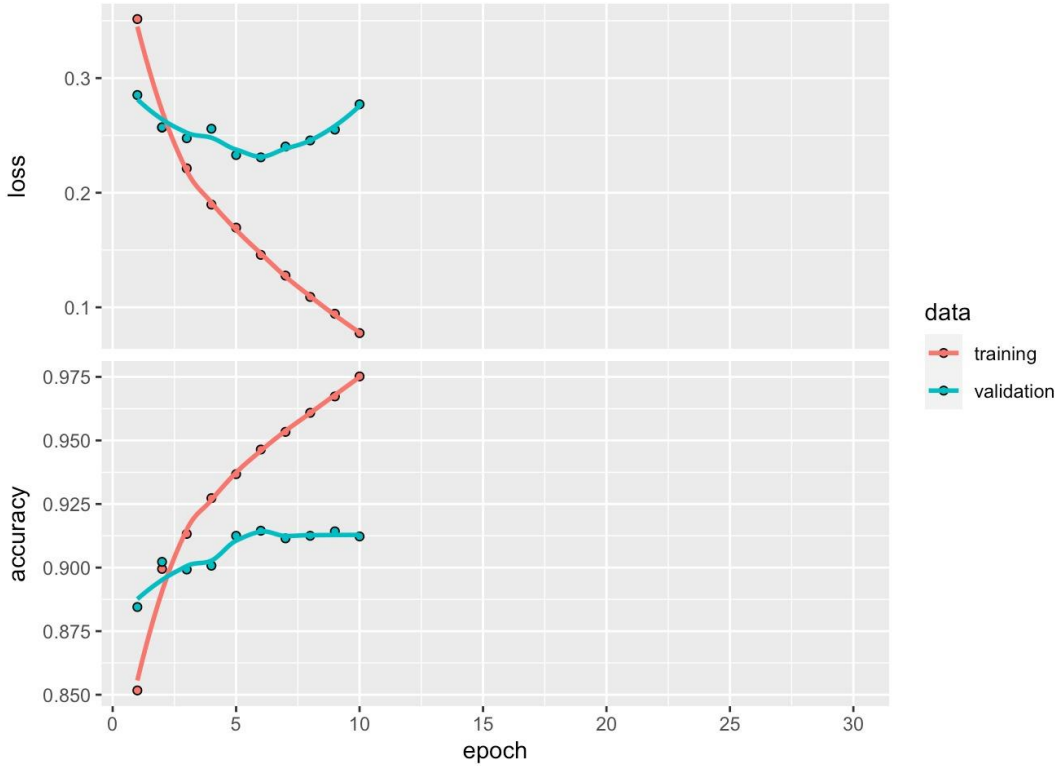
- TRAINING SET: 0.8583
- VALIDATION SET: 0.8640

I valori finali restituiti dalla funzione di loss risultano essere:

- TRAINING SET: 0.3392
- VALIDATION SET: 0.3591

3.4 VGG-16 Neural Network

Eseguendo la rete pre-allenata VGG-16 otteniamo il seguente plot.



In questo caso la stima dei parametri attraverso la back propagation si arresta alla 10° epoca. Dai grafici possiamo notare come questo tipo di rete pre-allenata risulta essere molto efficace, infatti si evince che dopo una sola epoca il livello di accuracy passi da 0.8517 a 0.8995.

Anche in questo caso si può notare come vi sia overfitting nel modello, infatti l'aumento costante dell'accuracy nel training set è contrapposto ad un livello di accuracy pressappoco costante a partire dalla 5° epoca sul validation set.

I risultati ottenuti considerando l'accuracy risultano essere:

- TRAINING SET: 0.9752
- VALIDATION SET: 0.9122

I livelli finali della funzione di loss risultano essere:

- TRAINING SET: 0.0775
- VALIDATION SET: 0.2772

3.5 SELECTION MODEL

Per selezionare il modello finale abbiamo deciso di considerare come criterio da massimizzare il valore di accuracy del validation set.

Si riporta la tabella riassuntiva.

	FNN	**CNN**	**CNN AUGMENTATION**	**VGG-16**
ACCURACY VALIDATION SET	0.7975	0.8543	0.8640	0.9122

La scelta fatta quindi risulta essere la rete neurale pre-allenata VGG-16, la quale applicata al Test set ha ottenuto un livello di accuracy pari a 0.9040.

4. Discussioni

Nel presente elaborato si è analizzato un dataset con oltre 20.000 immagini, scaricate dal web, di rifiuti generici, da classificare come organici o riciclabili. In un periodo in cui l'attenzione all'ecosistema e all'ambiente è alta, l'argomento trattato desta particolare interesse. Al fine di portare a termine il task di classificazione delle immagini, si sono utilizzate quattro diverse tipologie di reti neurali: Feed Forward, Convoluzionale, Convoluzionale con Data Augmentation e Convoluzionale pre-allenata VGG-16. I risultati ottenuti mostrano che il modello di rete neurale più adatto al riconoscimento delle immagini del dataset analizzato è quello pre-allenato, la VGG-16, che ha riportato un'accuracy sul validation set pari a 0.9122. Tale modello ha performato sul test set sostanzialmente in maniera analoga, 0.9040 di accuratezza. L'utilizzo di tali tecniche particolarmente sofisticate hanno fornito un esito di alta qualità, rendendo l'elaborato considerevole.

E' evidente che il mondo dell'intelligenza artificiale sia in continuo e rapido sviluppo e pertanto è altresì chiaro che il nostro elaborato è da considerarsi come

un buon punto di partenza nell'ambito del riconoscimento automatico di oggetti di rifiuto che può essere di fondamentale importanza nella vita quotidiana attuale. I possibili sviluppi futuri sono effettivamente tanti: a partire da queste immagini e tenendo inalterate le etichette di classificazione, una rete pre-allenata, come la VGG-16, con l'implementazione di un qualche tipo di Data Augmentation potrebbe fornire risultati ancora migliori di quelli presentati. Infatti è da sottolineare il fatto che le tecniche che tentano di ridurre il fenomeno di overfitting sono molte: i metodi di Neural Augmentation e Smart Augmentation, ad esempio, insegnano alla rete neurale come apprendere autonomamente la generazione di nuovi campioni minimizzando l'errore della rete stessa. Inoltre, esistono altre diverse tecniche di regolarizzazione implementate all'interno dei layer intermedi come DisturbLabel e SoftLabel che aggiungono rumore alle etichette per meglio approssimare lo spazio di probabilità dell'output space. Anche i modelli pre-allenati sono tanti: Xception, Inception V3, ResNet50, VGG19, MobileNet sono per esempio altre reti strutturate sulla base di immagini molto simili, ed in parte uguali, a quelle che compongono il dataset esaminato nel presente elaborato.

Altri spunti futuri possono derivare dalla diversa concezione del problema: considerare tante etichette di risposta quante sono le tipologie di rifiuti principali (plastica, carta, vetro, materiale indifferenziato o non riciclabile ecc) piuttosto che solo due categorie avrebbe di sicuro una maggiore rilevanza pratica.

Infine testare l'efficacia delle reti nel riconoscimento di rifiuti non presenti nel dataset, come foto scattate personalmente e non scaricate dal web per poter implementare la base di dati dal momento che gli oggetti rappresentati nelle immagini presenti nel dataset sono spesso integri e si allontanano dal concetto di rifiuto.