

Progetto Machine Learning and Data Processing

Il presente progetto, nell'ambito del corso di Machine Learning and Data Processing, si propone di classificare un dataset composto da 600 immagini denominate "spettrogrammi di target spaziali" mediante l'impiego delle reti neurali AlexNet e VGG16. Il dataset è suddiviso in tre categorie, ognuna delle quali comprende 200 immagini. L'analisi effettuata mira a confrontare le reti neurali in termini di accuratezza e matrice di confusione, nonché a individuare le differenze tra esse applicando la tecnica della data augmentation, che consiste nella modifica leggera delle immagini del dataset. Per l'addestramento delle reti neurali, 80% del dataset di immagini viene utilizzato come dataset di addestramento "train set" e il restante 20% viene suddiviso tra dataset di test "test set" e di convalida "validation set", e viene impiegato l'algoritmo di ottimizzazione SGD (Stochastic Gradient Descent) che aggiorna i pesi della rete neurale muovendoli nella direzione opposta al gradiente negativo rispetto alla funzione obiettivo.

- Il training set è impiegato per addestrare il modello, consentendo alla rete neurale di apprendere dai dati. Durante questa fase, i pesi della rete vengono aggiornati al fine di minimizzare la funzione di perdita (loss function), che rappresenta un indicatore delle previsioni della rete neurale rispetto ai dati reali;
- Il validation set non contribuisce direttamente all'addestramento della rete neurale, ma serve a verificare le performance del modello su un insieme di dati che la rete neurale non ha mai incontrato durante la fase di addestramento;
- Il test set viene utilizzato soltanto dopo che la rete neurale è stata completamente addestrata e convalidata, e serve a fornire una valutazione delle prestazioni del modello in termini di capacità di generalizzazione su nuovi dati non presenti nell'insieme di addestramento.

In seguito, si procede all'analisi delle prestazioni della rete neurale AlexNet, la quale è tipicamente composta da cinque layer convoluzionali seguiti da due layer completamente connessi (fully connected). Un elemento distintivo di AlexNet è l'utilizzo della funzione di attivazione ReLU (Rectified Linear Unit), applicata dopo ciascun livello convoluzionale e completamente connesso, ad eccezione dell'ultimo layer di output. AlexNet introduce inoltre il concetto di dropout, una tecnica di regolarizzazione che disattiva in modo casuale alcuni neuroni durante la fase di addestramento della rete al fine di ridurre il rischio di overfitting. Per l'addestramento della rete, vengono impostati i seguenti iperparametri: un tasso di apprendimento (learning rate) pari a 0.001, che determina l'ampiezza del passo per l'aggiornamento dei pesi; un numero di campioni per l'addestramento (batch size) di 32, in modo tale che la rete venga aggiornata ogni 32 immagini; e un numero di epoche pari a 30, ovvero il numero di iterazioni durante la fase di addestramento. Inoltre, l'architettura della rete neurale AlexNet è stata modificata aggiungendo un ulteriore livello di dropout per mitigare ulteriormente il problema di overfitting.

Poiché le reti neurali come AlexNet e VGG16 necessitano di immagini con dimensioni specifiche in input, è stato effettuato un processo di "resize" per modificare le dimensioni di ciascuna immagine del training set. In particolare, le immagini sono state ridimensionate a 227 x 227 pixel per la rete neurale AlexNet e a 224 x 224 pixel per la rete neurale VGG16

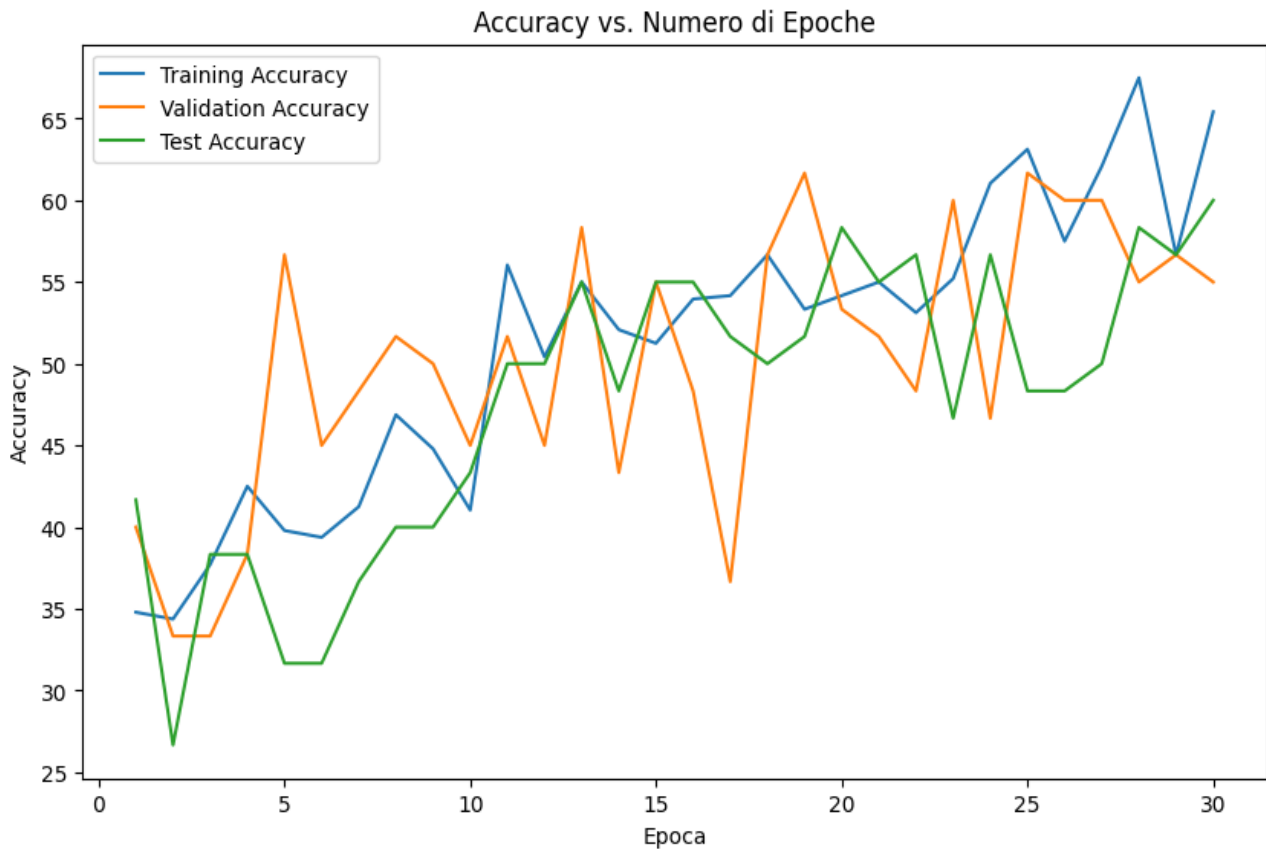


Figura 1 Grafico accuratezza AlexNet con 30 epoche

L'output mostrato in Figura 1 illustra l'accuratezza della rete neurale AlexNet durante la fase di addestramento, considerando un totale di 30 epoche. In dettaglio, sono presenti tre curve distinte che rappresentano l'accuratezza del training set (Training Accuracy), l'accuratezza del validation set (Validation Accuracy) e l'accuratezza del test set (Test Accuracy). L'accuratezza di addestramento riflette il grado di apprendimento della rete neurale basato sul set di dati di addestramento, mentre l'accuratezza di validazione indica il livello di generalizzazione del modello su un set di dati non utilizzato durante l'addestramento. L'accuratezza del test invece si riferisce all'accuratezza calcolata su un set di dati diverso dal validation set e fornisce una valutazione finale delle prestazioni della rete neurale. È possibile osservare che le curve presentano fluttuazioni significative al crescere del numero di epoche.

Confrontando la curva dell'accuratezza del training set con le altre due curve, è possibile determinare se la rete neurale presenta fenomeni di overfitting o underfitting durante l'addestramento. L'accuratezza di addestramento, validazione e test sembra seguire un andamento simile, sebbene con alcune variazioni, suggerendo che la rete neurale potrebbe non essere affetta in modo significativo da overfitting o underfitting. Pertanto, potrebbero essere apportate modifiche agli iperparametri della rete neurale.

Si procederà aumentando il numero di epoche da 30 a 50 per la rete neurale AlexNet, al fine di analizzare se continuerà a non presentare problemi in termini di overfitting o underfitting.

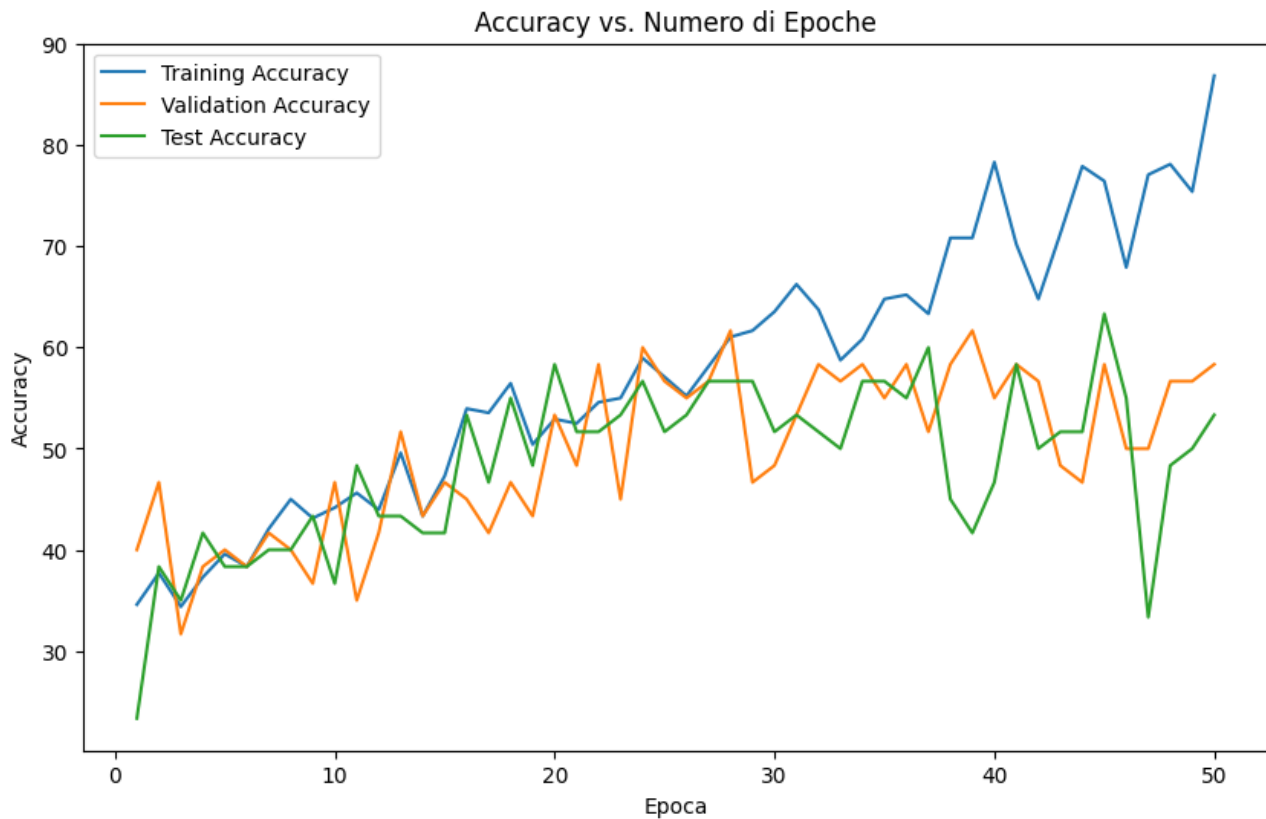


Figura 2 Grafico accuratezza AlexNet con 50 epoche

L'output presentato in Figura 2 mostra l'accuratezza AlexNet durante la fase di addestramento in riferimento a un numero di epoche di 50. A differenza della Figura 1, l'accuratezza del training set, del validation set e del test set non sembra seguire un andamento simile, il che suggerisce che la rete neurale potrebbe essere soggetta a overfitting durante la fase di addestramento, raggiungendo un'accuratezza per il training set vicina al 90%. In confronto all'accuratezza del training set, l'accuratezza del validation set e del test set continua a seguire un trend simile, il che suggerisce di non incrementare ulteriormente il numero di epoche per la rete neurale AlexNet.

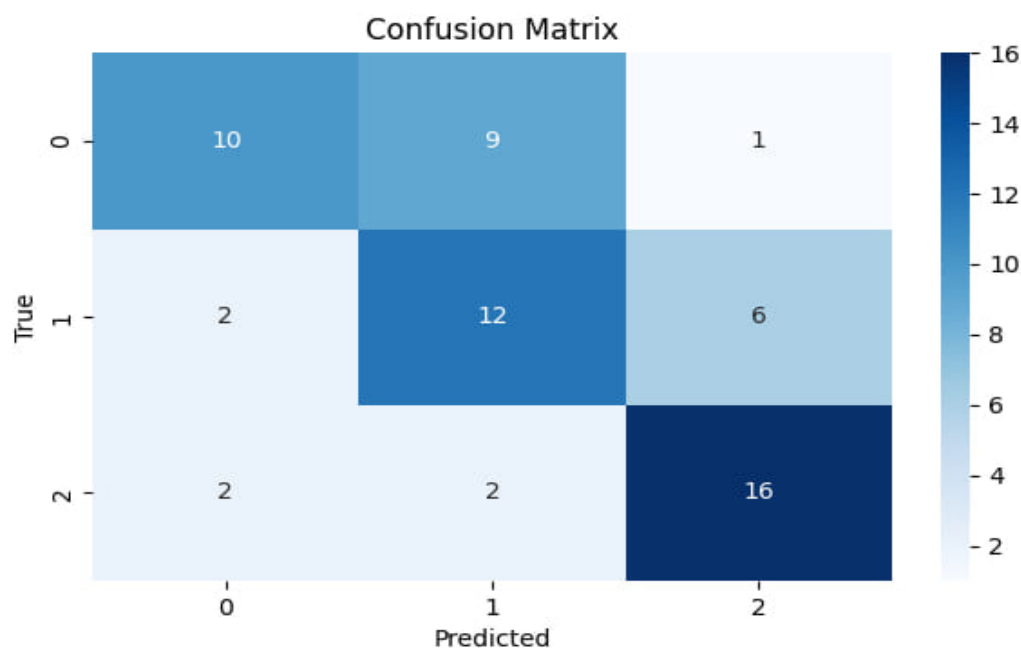


Figura 3 Matrice di confusione AlexNet

L'output mostrato in Figura 3 rappresenta una matrice di confusione, in cui i valori sulla diagonale principale indicano i casi in cui la predizione della rete neurale AlexNet nella classificazione coincide con la classe reale presente nel dataset di test. La classe 1 (label 0) è stata classificata correttamente 10 volte; tuttavia, si sono verificati 9 casi in cui la classe 1 è stata erroneamente classificata come classe 2 (label 1) e un caso in cui è stata confusa con la classe 3 (label 2). La classe 2, invece, è stata classificata correttamente 12 volte, erroneamente 2 volte come classe 1 e 6 volte come classe 3. Infine, la classe 3 è stata classificata correttamente 16 volte, con solamente 2 errori in cui è stata scambiata per la classe 1 e altri 2 per la classe 2.

La rete neurale AlexNet sembra ottenere risultati migliori nel classificare la classe 3, presentando il maggior numero di previsioni corrette e il minor numero di previsioni errate. Tuttavia, come mostrato in Figura 3, la rete neurale incontra difficoltà nel distinguere le caratteristiche che differenziano le immagini della classe 2 e classe 3, classificando erroneamente 6 immagini della classe 2 come classe 3. Si riscontra anche una problematica simile tra la classe 1 e 2, con 9 immagini della classe 1 erroneamente classificate come classe 2.

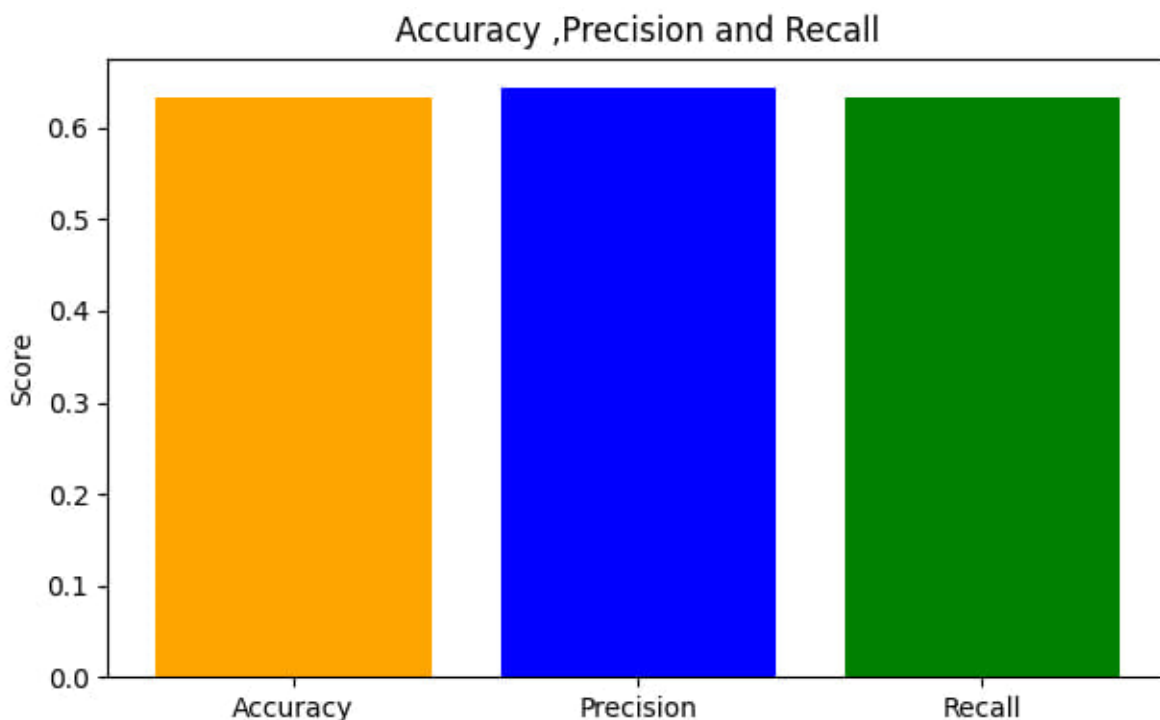


Figura 4 Performance AlexNet in termini di accuratezza, precisione e richiamo

L'output in figura 4 rappresenta graficamente le metriche di valutazione della rete neurale AlexNet con un batch size di 32 e un numero di epoche di 30.

La percentuale del 63% dell'accuratezza (Accuracy) indica il numero di predizioni corrette complessive rispetto al totale delle predizioni fatte ($\frac{10+12+16}{60}$).

La percentuale del 64% della precisione (Precision) indica il numero di predizioni positive che erano effettivamente corrette, in altri termini l'accuratezza per le classi positive.

La percentuale del 63% del richiamo (Recall) indica il numero di casi positivi reali che il modello è stato in grado di identificare correttamente.

Dopo aver analizzato le performance della rete neurale AlexNet con dei buoni risultati in termini di accuratezza, precisione e richiamo possiamo riaddestrare la rete neurale AlexNet con l'aggiunta di data augmentation alle immagini del dataset "spettrogrammi di target spaziali" e aumentando il numero di epoche a 50, rimuovendo il livello di dropout aggiunto precedentemente e lasciando gli altri iperparametri invariati.

Attraverso il procedimento della data augmentation vengono aumentati i dati in input, generando artificialmente nuovi set di immagini a partire da quelli esistenti con cui addestrare la rete neurale durante la fase di train. Questa tecnica viene utilizzata per ridurre l'overfitting durante la fase di addestramento di una rete neurale.

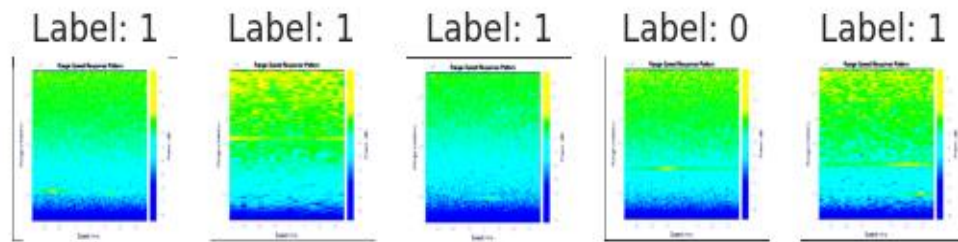


Figura 5 immagini AlexNet con data augmentation

L'output della Figura 5, mostra alcune immagini del set di addestramento della rete neurale AlexNet a cui è stata applicata una rotazione casuale in un intervallo di -0.3 e $+0.3$ gradi, con piccole traslazioni in entrambe le direzioni fino al 1%. Nella figura 5, label 0 e 1 sono alcune immagini di esempio appartenenti alla classe 1 e 2 del dataset "spettrogrammi di target spaziali".

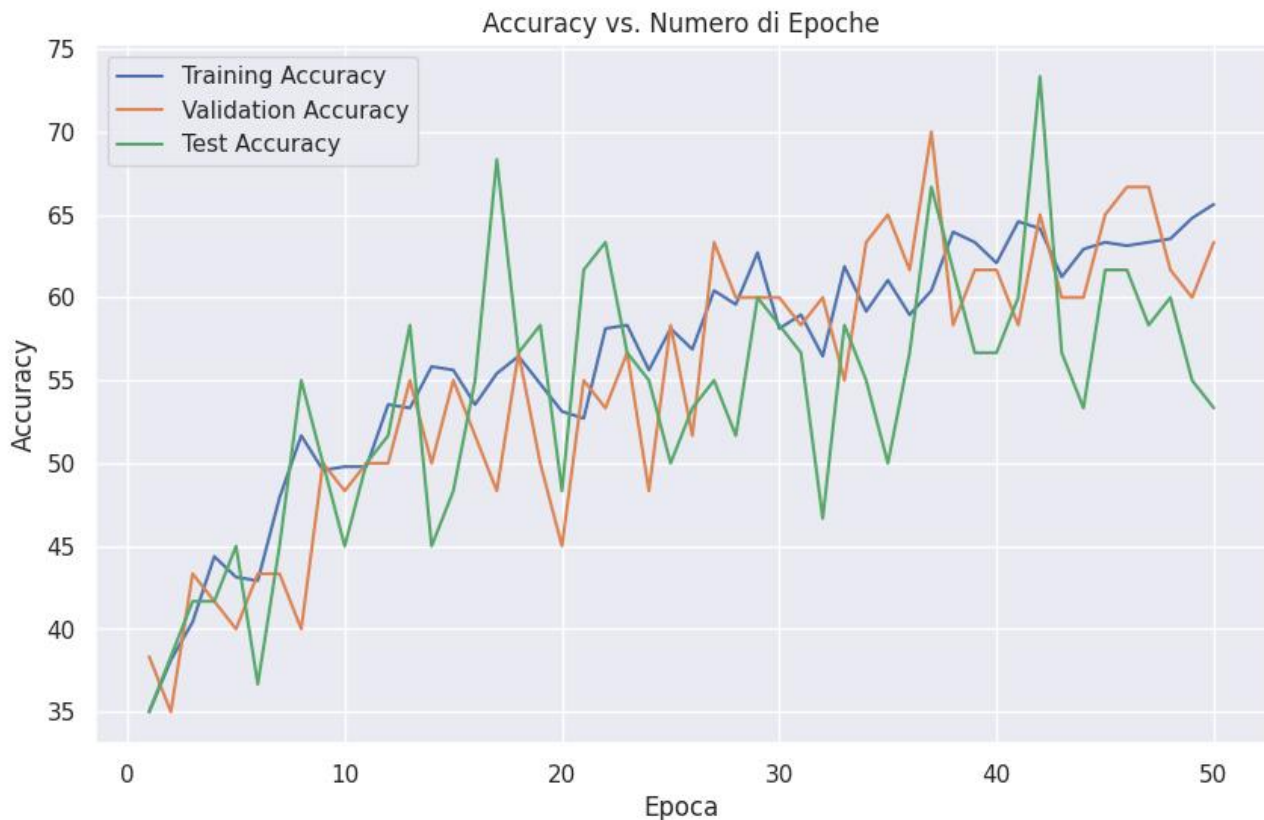


Figura 6 Grafico Accuratezza AlexNet con data augmentation con 50 epoche

L'output mostrato in Figura 6 illustra l'accuratezza della rete neurale AlexNet con data augmentation durante la fase di addestramento, considerando un totale di 50 epoche. Il grafico suggerisce che l'implementazione della data augmentation ha mitigato il problema di overfitting precedentemente riscontrato nell'analisi del grafico della Figura 2, e la curva del training set non mostra fluttuazioni anomale durante l'addestramento. Esaminando il grafico, si possono osservare andamenti simili nelle tre curve al crescere del numero di epoche, con picchi di accuratezza nel test set del 68,33% alla diciassettesima epoca e del 73,33% alla quarantesima epoca.

Da quest'analisi, si deduce che la rete neurale riesce a generalizzare efficacemente tra le immagini del validation set e del test set durante la fase di addestramento

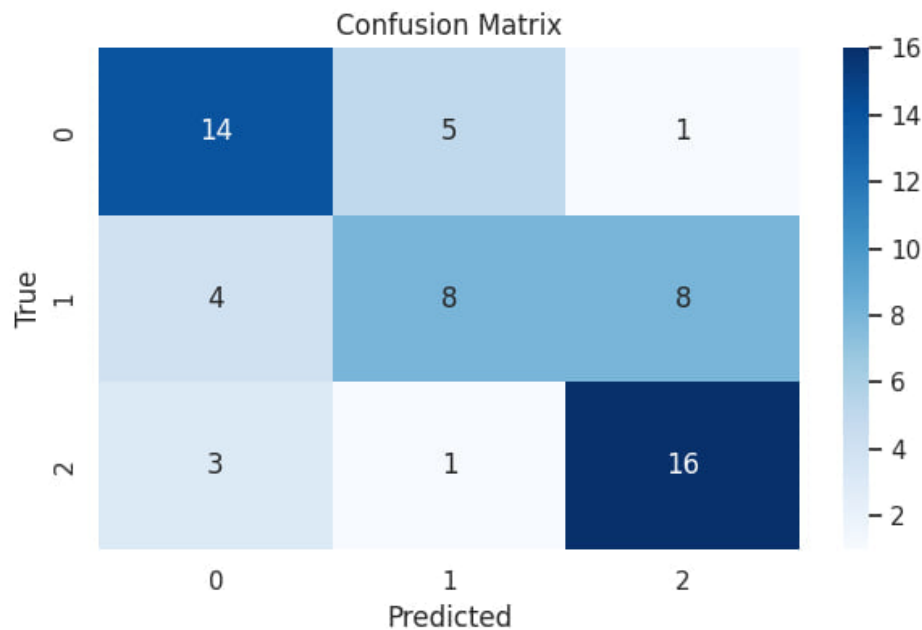


Figura 7 Matrice di confusione AlexNet con data augmentation

Analizzando la matrice di confusione presentata in Figura 7, si osserva che la rete neurale AlexNet, con l'implementazione della data augmentation, è riuscita a classificare correttamente 14 immagini della classe 1 (label 0), 8 immagini della classe 2 (label 1) e 16 immagini della classe 3 (label 2) dopo l'addestramento. In confronto ai risultati ottenuti nella Figura 3, la rete neurale mostra miglioramenti nella classificazione delle immagini della classe 1 (14 rispetto a 10) ma peggiora nelle performance di classificazione delle immagini della classe 2 (8 rispetto a 12), mentre mantiene risultati identici nella classificazione delle immagini della classe 3. Nella figura 7, la rete neurale sembra avere delle difficoltà nel riconoscere le caratteristiche che distinguono le immagini della classe 2 e classe 3, classificando erroneamente 8 immagini della classe 2 come classe 3.

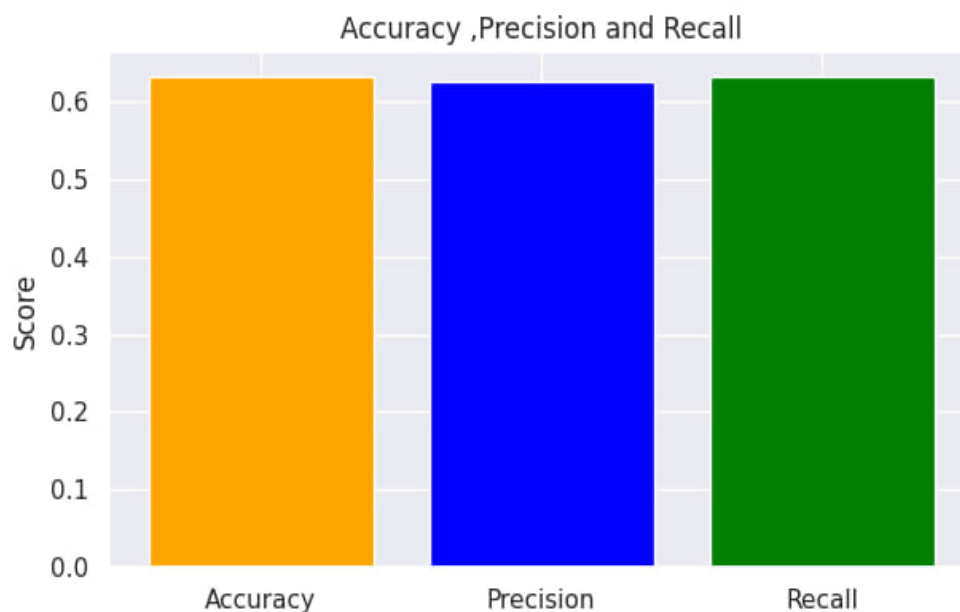


Figura 8 Performance AlexNet con data augmentation in termini di accuratezza, precisione e richiamo

La Figura 8 mostra graficamente le metriche di valutazione della rete neurale AlexNet con l'implementazione della data augmentation, utilizzando un batch size di 32 e un numero di epoche di 50.

In termini di accuratezza, precisione e richiamo, la performance percentuale dell'AlexNet con data augmentation è del 63% per tutte e tre le metriche di valutazione, mentre nella Figura 4, l'AlexNet ha ottenuto un 64% nella precisione.

Dopo aver analizzato e confrontato i risultati e le performance dell'AlexNet e della sua versione con l'aggiunta della data augmentation, si procede con l'addestramento della rete neurale VGG16 e il confronto con la sua versione con data augmentation. Rispetto all'AlexNet, l'architettura della VGG16 è composta da 16 strati pesati: 13 strati convoluzionali e 3 strati completamente connessi, di cui i primi due hanno 4096 canali, un livello di dropout con un tasso del 50% ciascuno e l'ultimo strato con un numero di canali pari al numero di classi da predire. La VGG16 implementa la funzione di attivazione "ReLU" (Rectified Linear Unit) in tutti gli strati nascosti della rete, ad eccezione dell'ultimo che utilizza la funzione "softmax" per la classificazione multiclasse.

Per l'addestramento della VGG16, vengono impostati i seguenti iperparametri: un tasso di apprendimento (learning rate) di 0.001, un batch size di 64, maggiore rispetto all'AlexNet, in modo tale che la rete venga aggiornata ogni 64 campioni di immagini, e un numero di epoche di 30.

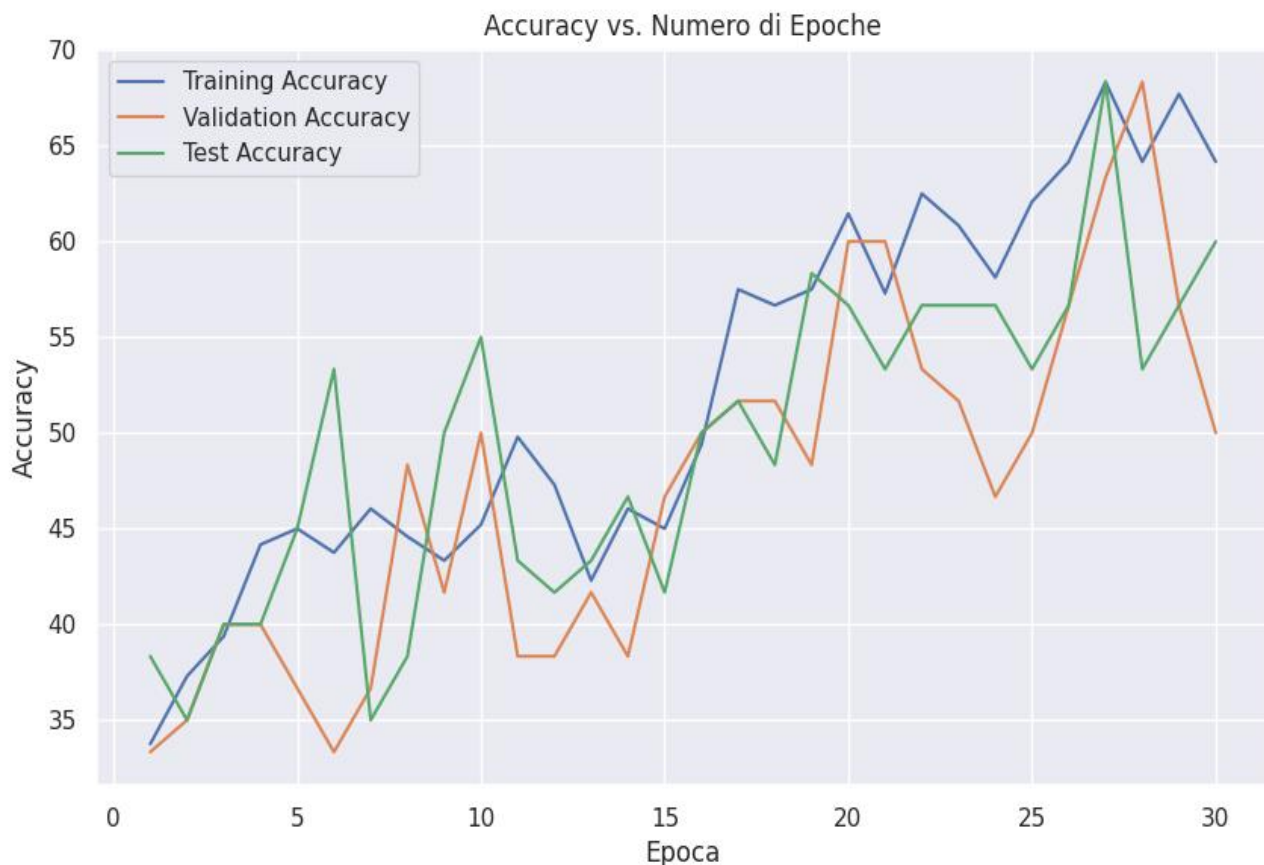


Figura 9 Grafico Accuratezza VGG16 con 30 epoche

L'output presentato in Figura 9 mostra l'accuratezza della rete neurale VGG16 durante la fase di addestramento. Nella figura, si osserva che le curve sono caratterizzate da fluttuazioni significative al crescere del numero di epoche, con un aumento delle tre curve tra la quindicesima e la ventesima epoca. L'accuratezza di addestramento, validazione e test sembra seguire un andamento simile, sebbene con alcune variazioni, in particolare tra la venticinquesima e la trentesima epoca. Questo suggerisce che la rete neurale potrebbe non essere affetta in modo significativo da overfitting o underfitting durante la fase di addestramento.

Inoltre, si può notare che la rete VGG16 dimostra un'efficace capacità di generalizzazione, come evidenziato dalle curve di validazione e di test. Tale capacità di generalizzazione suggerisce che il modello è in grado di adattarsi bene ai dati non visti durante l'addestramento, offrendo una buona base per effettuare previsioni accurate su nuovi campioni di dati.

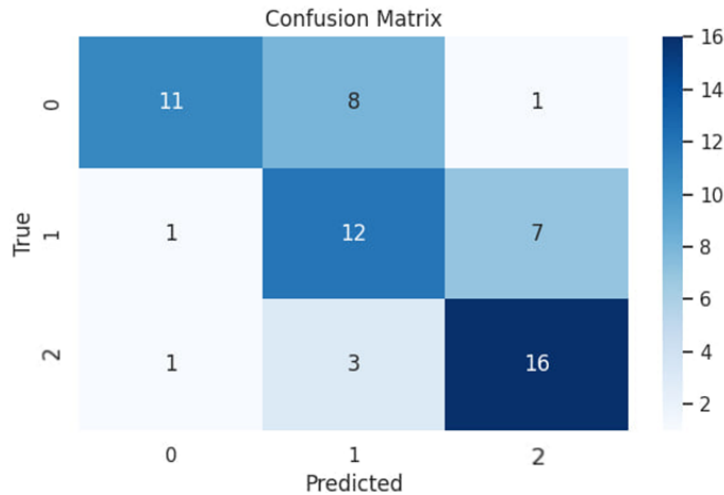


Figura 10 Matrice di confusione VGG16

Analizzando la matrice di confusione presentata in Figura 10, si può osservare che la rete neurale VGG16, dopo l'addestramento, è riuscita a classificare correttamente 11 immagini della classe 1 (label 0), 12 immagini della classe 2 (label 1) e 16 immagini della classe 3 (label 2). I risultati ottenuti nella classificazione corretta della classe 2 per la rete neurale VGG16 sono migliori rispetto a quelli riscontrati nelle matrici analizzate per la rete neurale AlexNet e AlexNet con data augmentation. Tuttavia, la rete neurale VGG16 sembra incontrare gli stessi problemi di classificazione osservati nelle matrici di confusione dell'AlexNet, sia con che senza data augmentation. In particolare, la rete neurale ha difficoltà nel riconoscere le caratteristiche che distinguono le immagini della classe 2 e classe 3, classificando erroneamente 7 immagini della classe 2 come classe 3. Si riscontra anche una problematica simile tra la classe 1 e 2, con 8 immagini della classe 1 erroneamente classificate come classe 2.

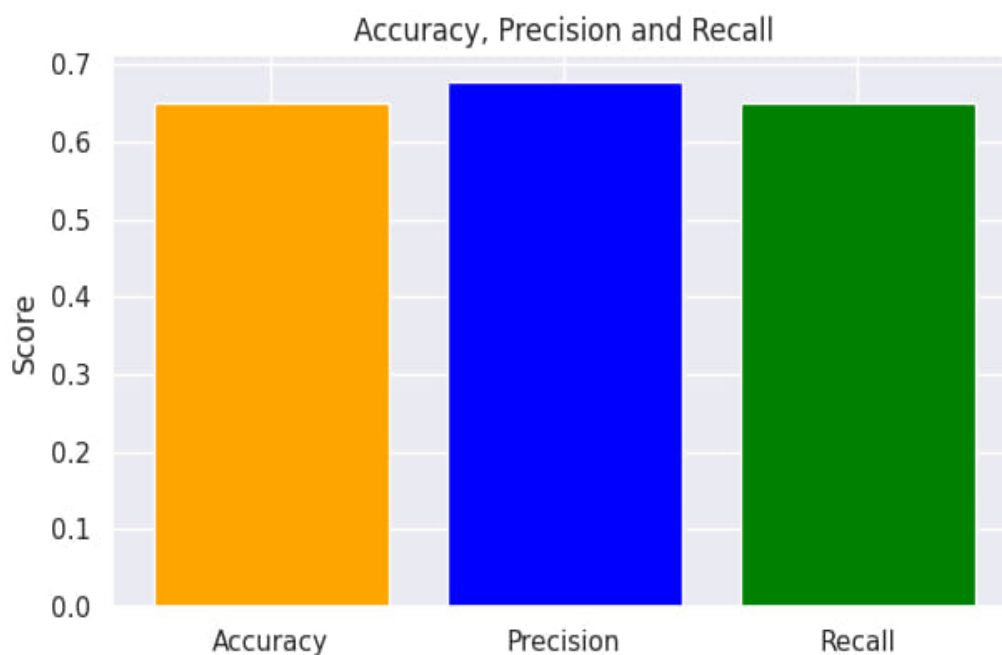


Figura 11 Performance VGG16 in termini di accuratezza, precisione e richiamo

L'output presentato in Figura 11 mostra graficamente le metriche di valutazione della rete neurale VGG16 con un batch size di 64 e un numero di epoche di 30. Esaminando le metriche di valutazione in termini percentuali, si osservano un'accuratezza del 65%, una precisione del 68% e un richiamo del 65%.

Da quest'analisi, si può concludere che la rete neurale VGG16 ha mostrato le migliori performance in termini di metriche di valutazione rispetto alla rete neurale AlexNet e alla versione di AlexNet con data augmentation.

Dopo l'analisi della rete neurale VGG16, si può addestrare nuovamente la rete, aggiungendo la data augmentation, come è stato applicato in precedenza con la rete neurale AlexNet, aumentando il numero di epoche a 50, lasciando gli altri iperparametri invariati.

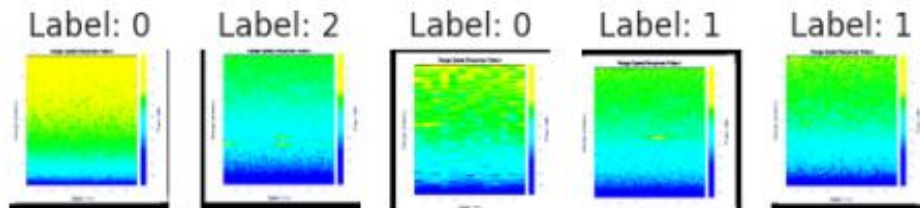


Figura 12 immagini VGG16 con data augmentation

L'output presentato nella Figura 12 mostra alcune immagini del set di addestramento utilizzate per la rete neurale VGG16, alle quali è stata applicata una data augmentation simile a quella utilizzata per AlexNet. Questa data augmentation include rotazioni e piccole traslazioni in entrambe le direzioni fino al 5%.

Le immagini etichettate come label 0, 1 e 2 sono esempi appartenenti rispettivamente alle classi 1, 2 e 3 del dataset "spettrogrammi di target spaziali".

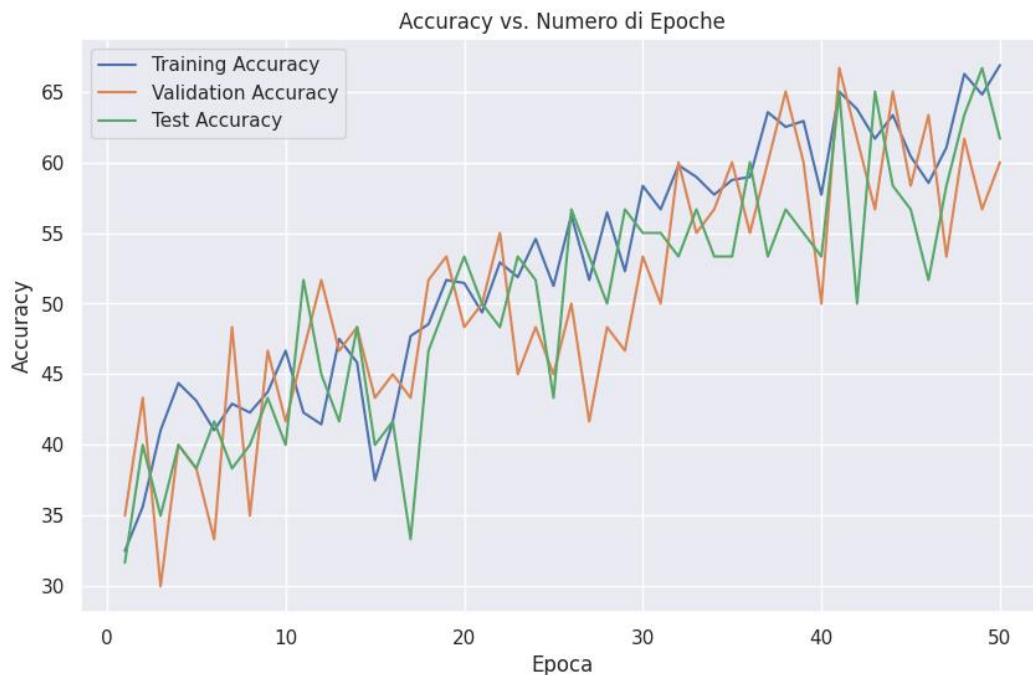


Figura 13 Grafico Accuratezza VGG16 con data augmentation con 50 epoche

L'output mostrato nella Figura 13 illustra l'accuratezza della rete neurale VGG16 con data augmentation durante la fase di addestramento per un totale di 50 epoche. Le curve di accuratezza di test e validation sembrano seguire un andamento simile con dei picchi tra la quarantesima e cinquantesima epoca arrivando oltre il 65%. Tra la trentesima epoca fino alla cinquantesima l'accuratezza del train sembra seguire il medesimo pattern delle curve di validation e test rispetto alle epoche precedenti e non sembrano esserci problemi di overfitting o underfitting.

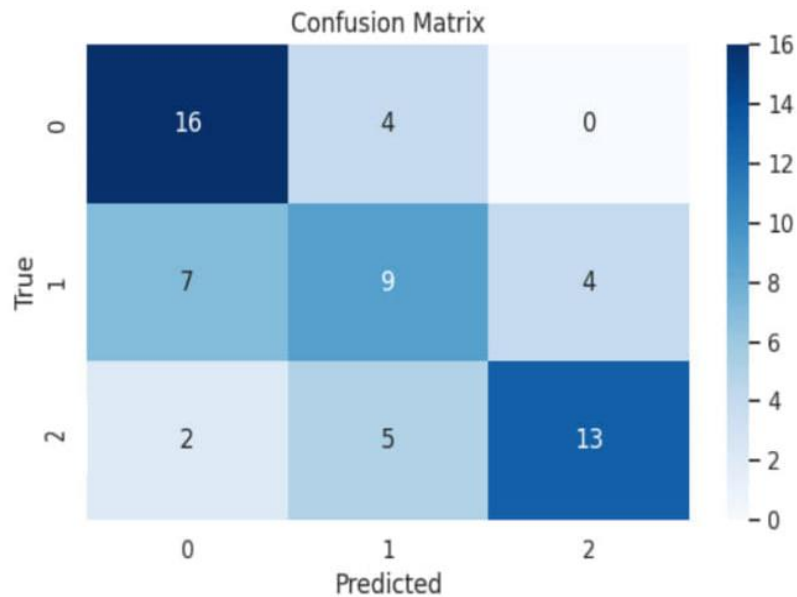


Figura 14 Matrice di confusione VGG16 con data augmentation

Esaminando la matrice di confusione mostrata nella Figura 14, è possibile notare che la VGG16, implementata con data augmentation, ha correttamente classificato 16 immagini appartenenti alla classe 1 (label 0), 9 immagini della classe 2 (label 1) e 13 immagini della classe 3 (label 2). In confronto alle altre reti neurali analizzate in precedenza, la VGG16 con data augmentation dimostra una migliore capacità di classificazione per la classe 1, mentre risulta meno efficace per la classe 3. I risultati ottenuti per la classe 2 sono simili a quelli della AlexNet con data augmentation. Tuttavia, emergono ancora difficoltà nel distinguere le caratteristiche delle immagini appartenenti alle classi 1 e 2, poiché 7 immagini della classe 2 sono state erroneamente classificate come appartenenti alla classe 1.

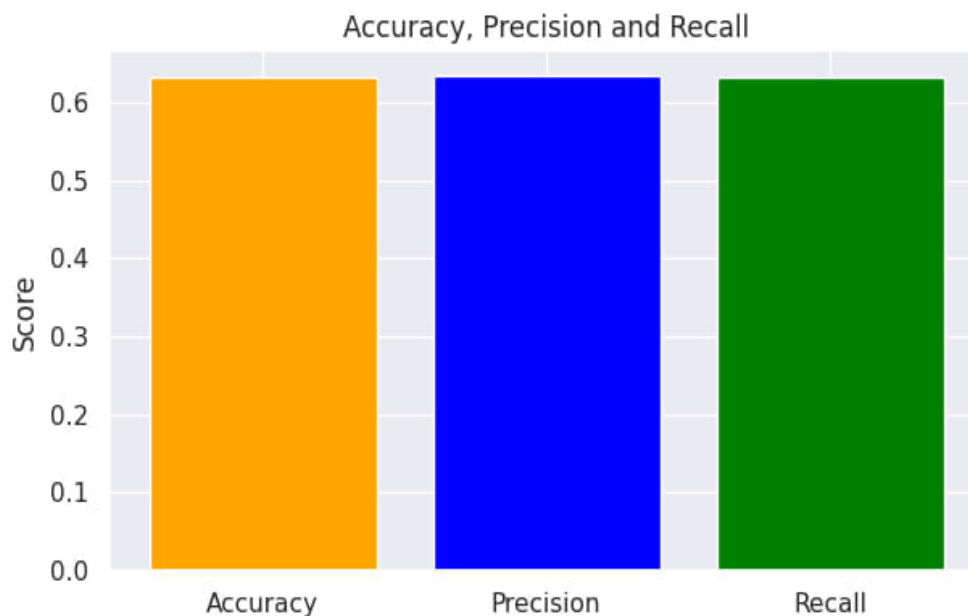


Figura 15 Performance VGG16 con data augmentation in termini di accuratezza, precisione e richiamo

La Figura 15 mostra graficamente le metriche di valutazione per la rete neurale VGG16 con l'implementazione della data augmentation, utilizzando un batch size di 64 e 50 epoche. In termini di accuratezza, precisione e richiamo, i risultati ottenuti sono analoghi a quelli della AlexNet con data augmentation.

In sintesi, le reti neurali addestrate e analizzate hanno dimostrato buone prestazioni nella classificazione delle immagini degli "spettrogrammi di target spaziali", in particolare per le classi 1 e 3. Tuttavia, sono state riscontrate difficoltà nella classificazione corretta delle immagini appartenenti alla classe 2 in tutte le matrici di confusione analizzate. Questo errore di classificazione potrebbe essere attribuito alla dimensione ridotta del dataset, composto da soli 600 immagini, con 200 per ciascuna classe.

Per quanto riguarda la classificazione della classe 1, la rete neurale VGG16 con data augmentation ha ottenuto i risultati migliori, classificando correttamente 16 immagini per la classe 2. Le reti neurali VGG16 e AlexNet hanno conseguito i migliori risultati per la classe 3, classificando correttamente 12 immagini. Ad eccezione della VGG16 con data augmentation, tutte le reti neurali addestrate hanno ottenuto i migliori risultati con 16 immagini classificate correttamente.

Infine, in termini di metriche di valutazione, la rete neurale VGG16 senza l'implementazione della data augmentation ha raggiunto le migliori prestazioni, avvicinandosi al 70% di precisione, richiamo e accuratezza.