



UNIVERSITÀ DEGLI STUDI DI SALERNO

DEEPCODET DETECTION ATTRAVERSO IL MODELLO CLIP

PROFESSORI:
MICHELE NAPPI
LUCIA CASCONE

TUTOR:
MATTEO POLSINELLI

STUDENTI:
CHIARA PUGLIA, 0522501984
LUCA GIULIANO, 0522501931
ANTONIO LANDI, 0522502075



ARGOMENTI TRATTATI



Pre-Processing sul dataset



Scelte progettuali: Approcci utilizzati



Conclusioni e Sviluppi Futuri



ANALISI DEL PROBLEMA



Negli ultimi anni si sono moltiplicati gli approcci di apprendimento automatico per distinguere immagini reali da Deepfake.



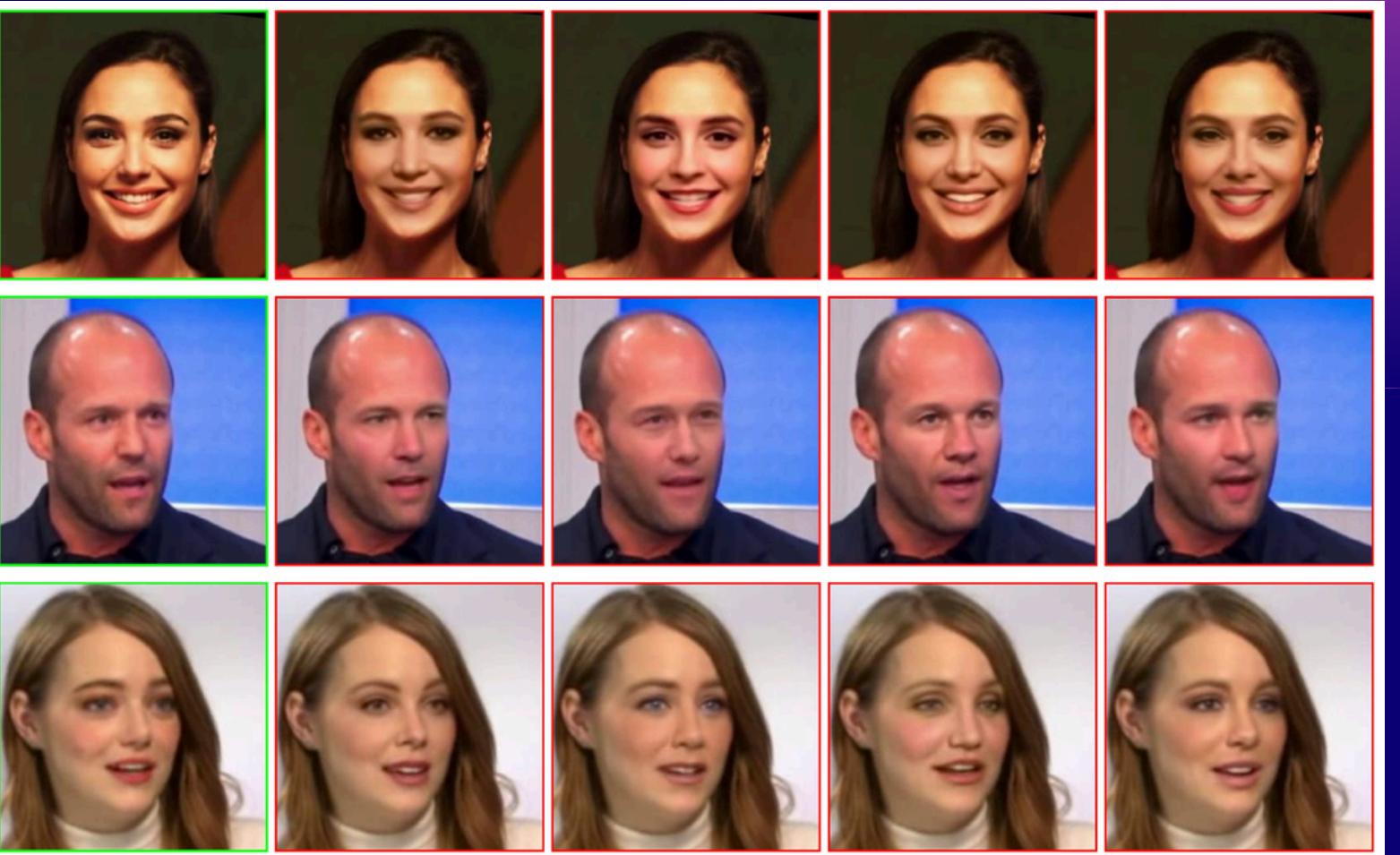
La maggior parte dei lavori risalgono al 2023 e sfruttano reti Siamesi con triplet loss allenata su immagini GAN.



Sono emersi anche metodi multimodali basati su grandi modelli visione-linguaggio.



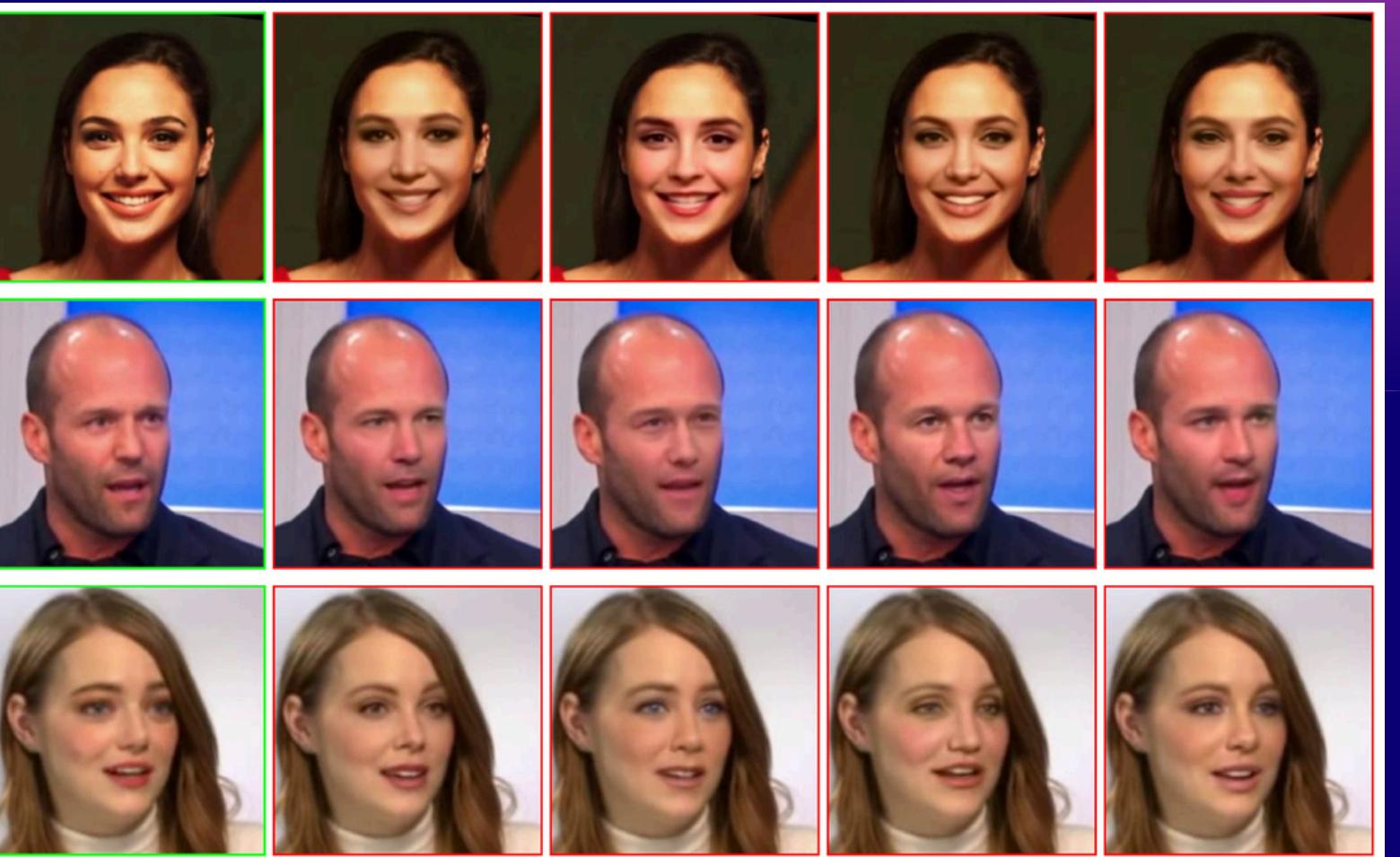
Utilizzo del modello CLIP per ottenere features visive generali.



Ci siamo concentrati sull'utilizzo del modello CLIP, che a sua volta carica un transformer per l'acquisizione delle immagini (e non solo) con l'obiettivo di addestrarlo per farsì che sia in grado di fare distinzione tra immagine reale ed immagine fake.

Sono emersi anche modelli multimediali basati su grandi modelli

Utilizzo del modello CLIP per ottenere features visive generali.



PRE PROCESSING: PRIME ANALISI



Il dataset si compone di 2496738 immagini, di cui 964989 rappresentano le immagini reali, 1531749 rappresentano le immagini fake.



Il dataset contiene le seguenti informazioni: filename, image_path, target e category.



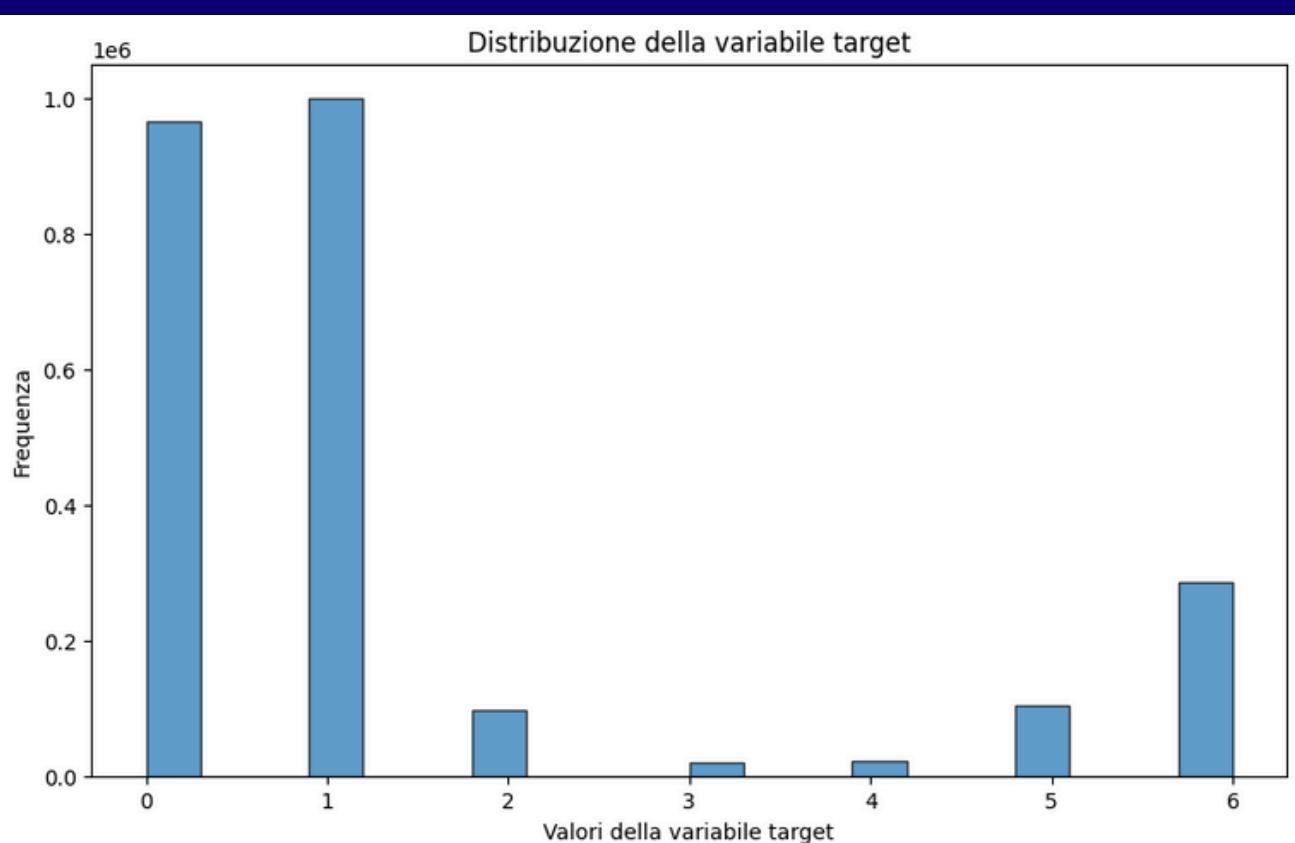
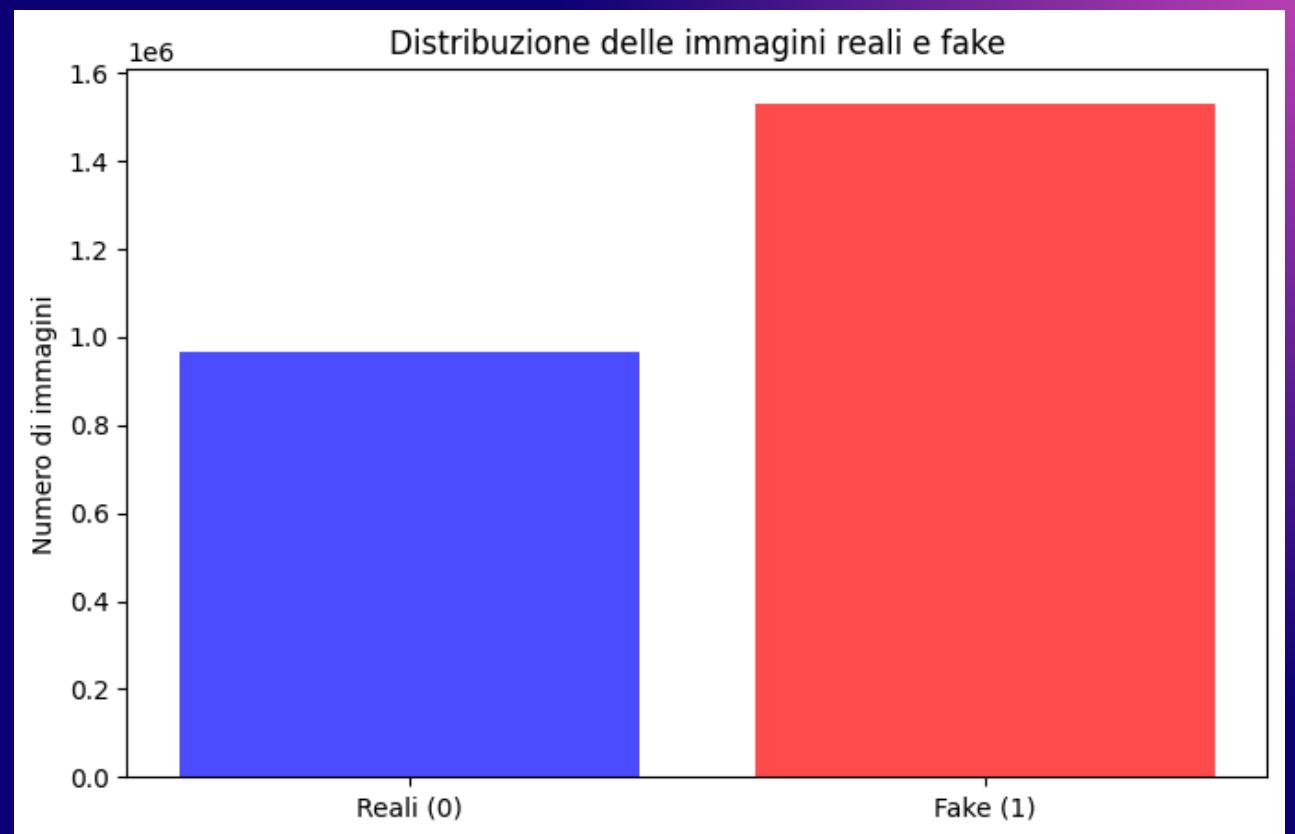
Sbilanciamento dei valori assunti dalla variabile target (valori da 0 a 6)



Tutti i valori maggiori o uguali ad 1 della variabile target sono stati convertiti in 1



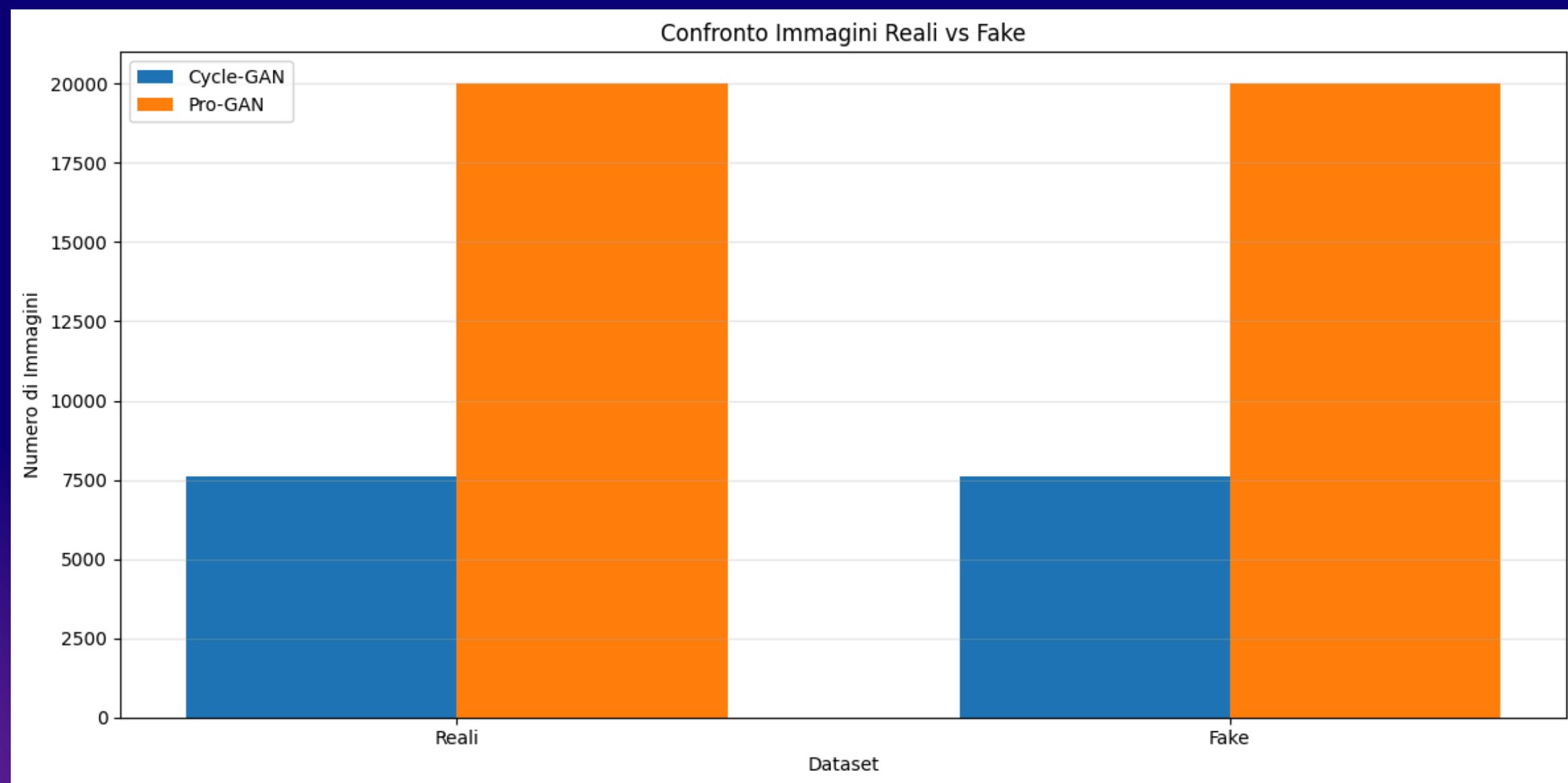
Le immagini risultano ritagliate e ridimensionate a 200x200 pixel e successivamente compresse utilizzando JPEG.



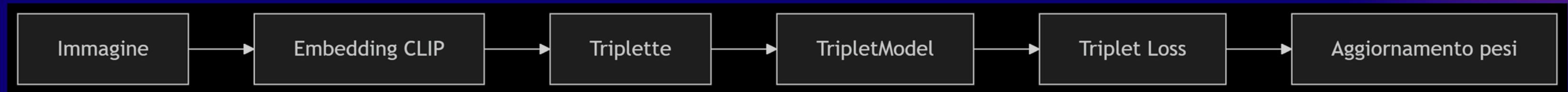
PRE PROCESSING: SCELTA DEI DATASET



Scelta dei dataset Cycle-Gan e Pro-Gan poichè risultavano bilanciati, ovvero costituiti dallo stesso numero di immagini reali e fake.



RICONOSCIMENTO DI IMMAGINI REAL E FAKE TRAMITE L'ESTRAZIONE DELLE INFORMAZIONI PER LE SOLE IMMAGINI



La prima fase consiste nella conversione delle immagini in vettori di embeddings di dimensione 512, che successivamente sono stati ridotti a 128 per proiettare l'embedding dell'immagine in uno spazio di dimensione ridotta.



La seconda fase consiste nell'estrazione delle triplette con l'obiettivo di garantire una maggiore divisione tra immagine reale ed immagine fake. Le triplette identificate sono rispettivamente: anchor, positive e negative.



La terza fase consiste nel passare gli embedding generati nella fase precedente e le rispettive triplette, organizzate in batch di dimensione fissa (ovvero 128), alla rete neurale TripletModel che utilizza la Triplet Loss per regolarizzare i pesi delle triplette sulla rete.

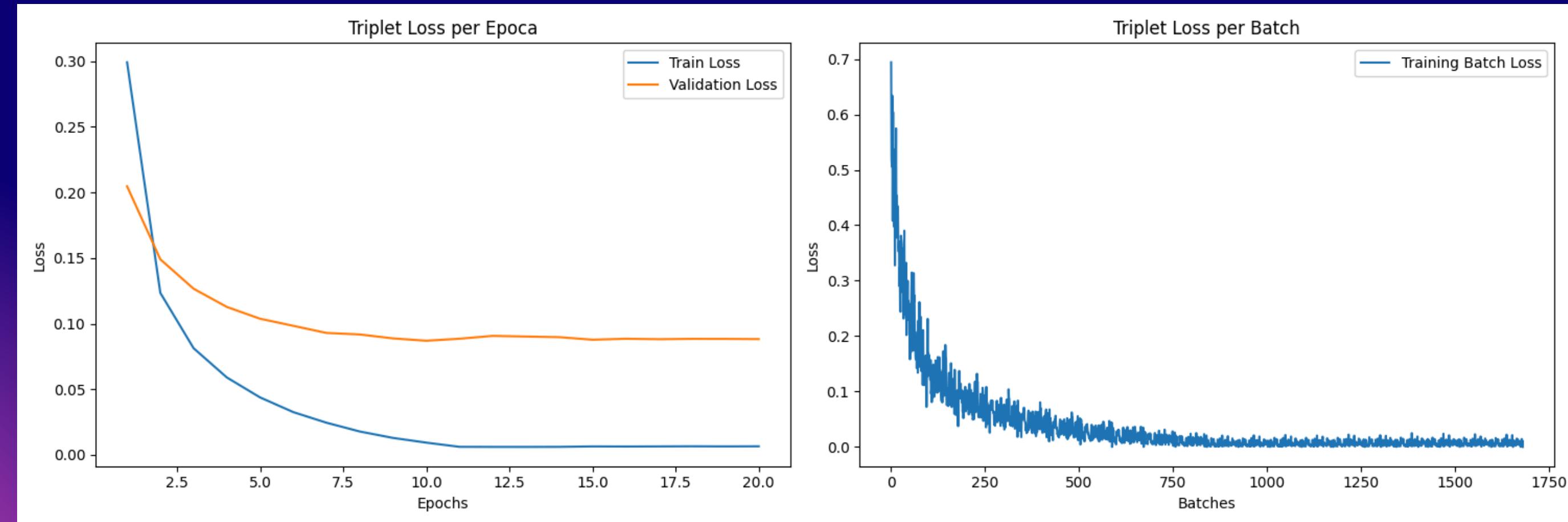
ESTRAZIONE DELLE INFORMAZIONI PER IL DATASET CYCLE-GAN



Il grafico sottostante mostra l'andamento di train e validation-loss su un addestramento di 20 epoche in cui si ferma alla quinta epoca avendo raggiunto il miglior risultato possibile.



La training loss scende rapidamente e si stabilizza a un valore basso, mentre la validation loss inizialmente scende ma, successivamente, risale leggermente e si stabilizza a un valore più alto (circa 0.15) rispetto alla training loss e il gap tra training e validation loss continua ad aumentare fino a stabilizzarsi con una differenza di circa 0.10.



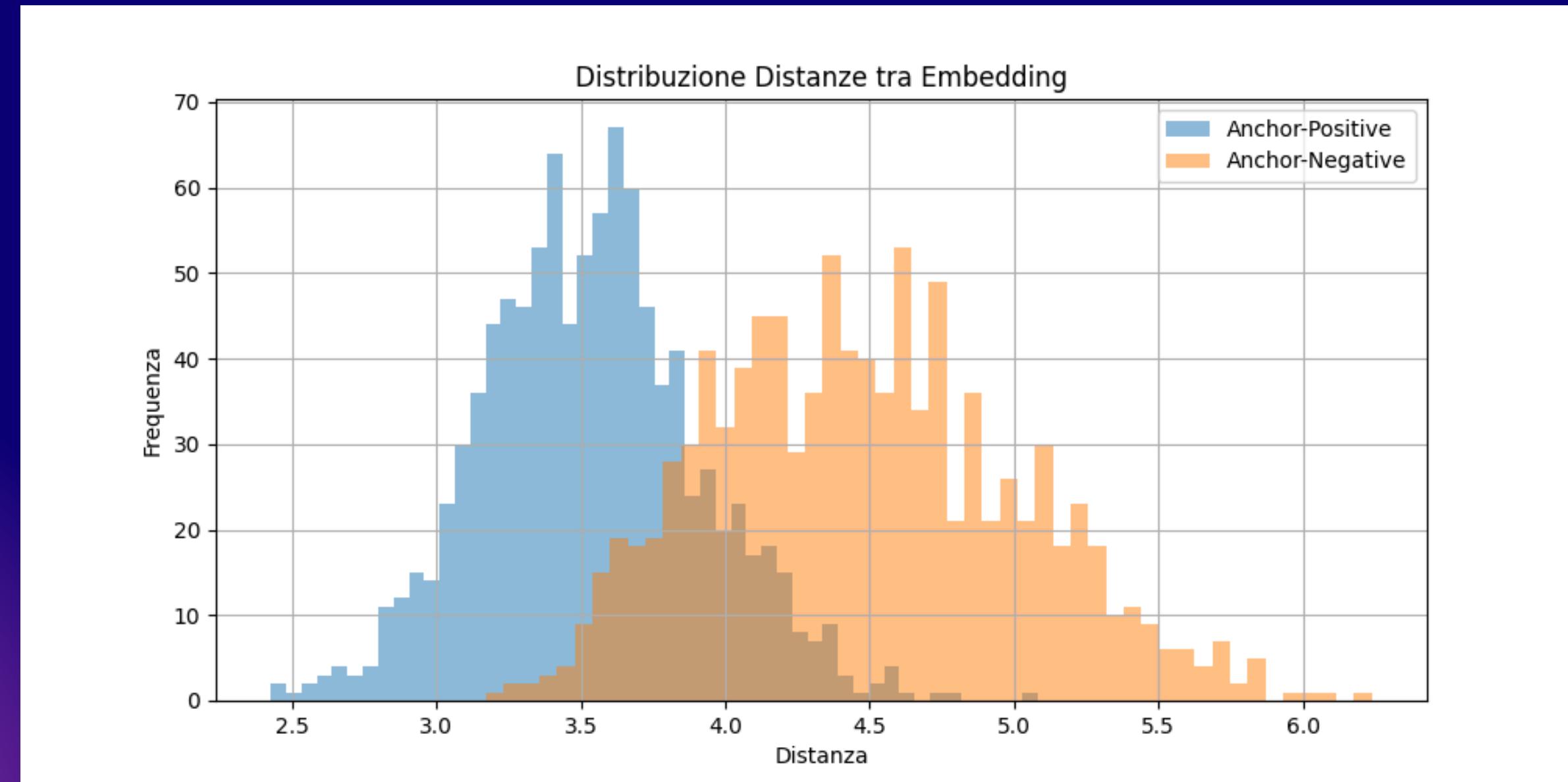
ESTRAZIONE DELLE INFORMAZIONI PER IL DATASET CYCLE-GAN



Il grafico sottostante rappresenta le distanze tra le triplette per i vettori di embeddings, rispettivamente le distanze tra anchor-positive e anchor negative.



Il grafico suggerisce una netta separazione tra anchor-positive e anchor-negative, confermando in questo modo una buona separazione tra immagini real e fake.



RISULTATI RAGGIUNTI PER IL DATASET CYCLE-GAN



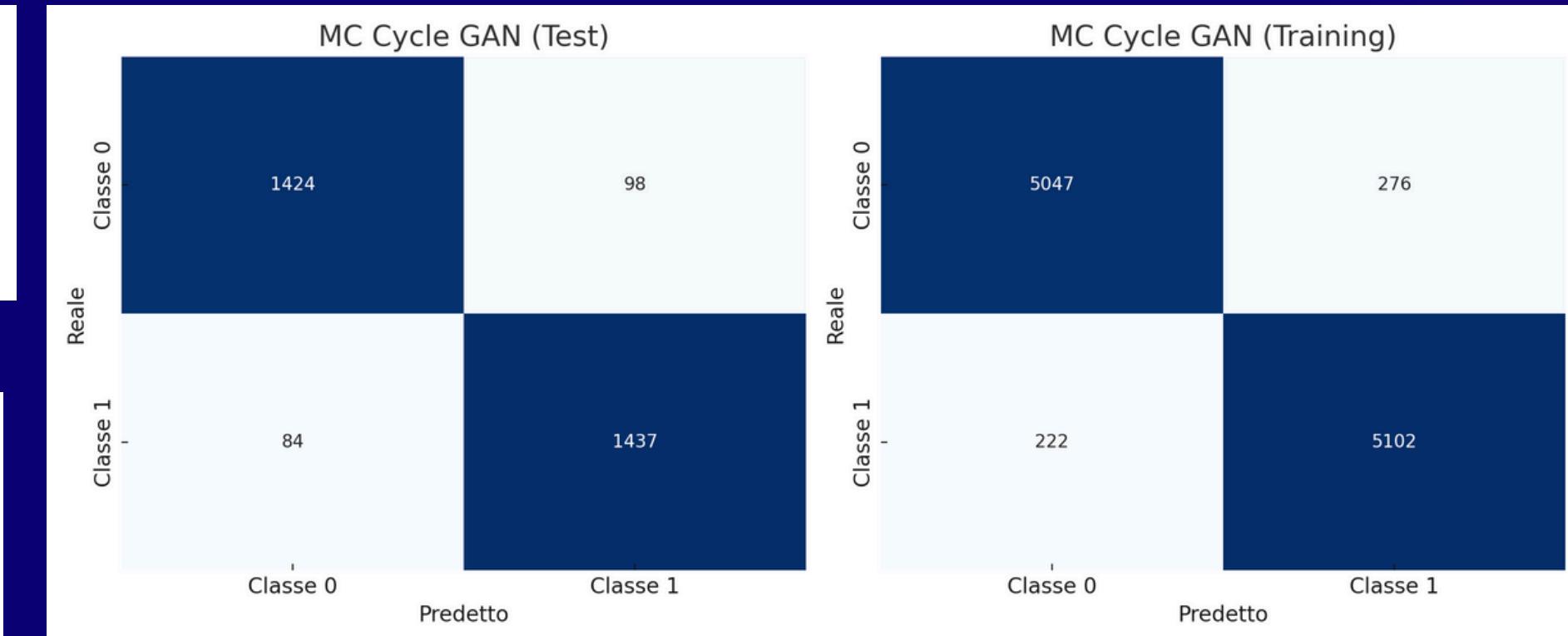
Per l'addestramento è stato utilizzato SVM (Support Vector Machine) i cui dati di addestramento sono stati suddivisi in 70% per il training, 20% per il Test Set e il restante 10% per Validation Set.



L'analisi è stata effettuata prendendo in considerazione 15210 campioni, dei quali 10647 sono stati selezionati per il Training Set, 1520 per la Validation Set ed infine 3043 per il Test Set.

Metriche	Percentuale
Accuracy	95.3%
Precision	95.3% - Media tra Precision Classe 0 = 95.79% e Precision Classe 1 = 94.87%
Recall	95.3% - Media tra Recall Classe 0 = 94.81% e Recall Classe 1 = 95.83%
F1-Score	95.3% - Media tra F1-Score Classe 0 = 95.30% e F1-Score Classe 1 = 95.35%

Metriche	Percentuale
Accuracy	94%
Precision	94% - Media tra Precision Classe 0 = 94.43% e Precision Classe 1 = 93.62%
Recall	94% - Media tra Recall Classe 0 = 93.56% e Recall Classe 1 = 94.48%
F1-Score	94% - Media tra F1-Score Classe 0 = 93.99% e F1-Score Classe 1 = 94.04%



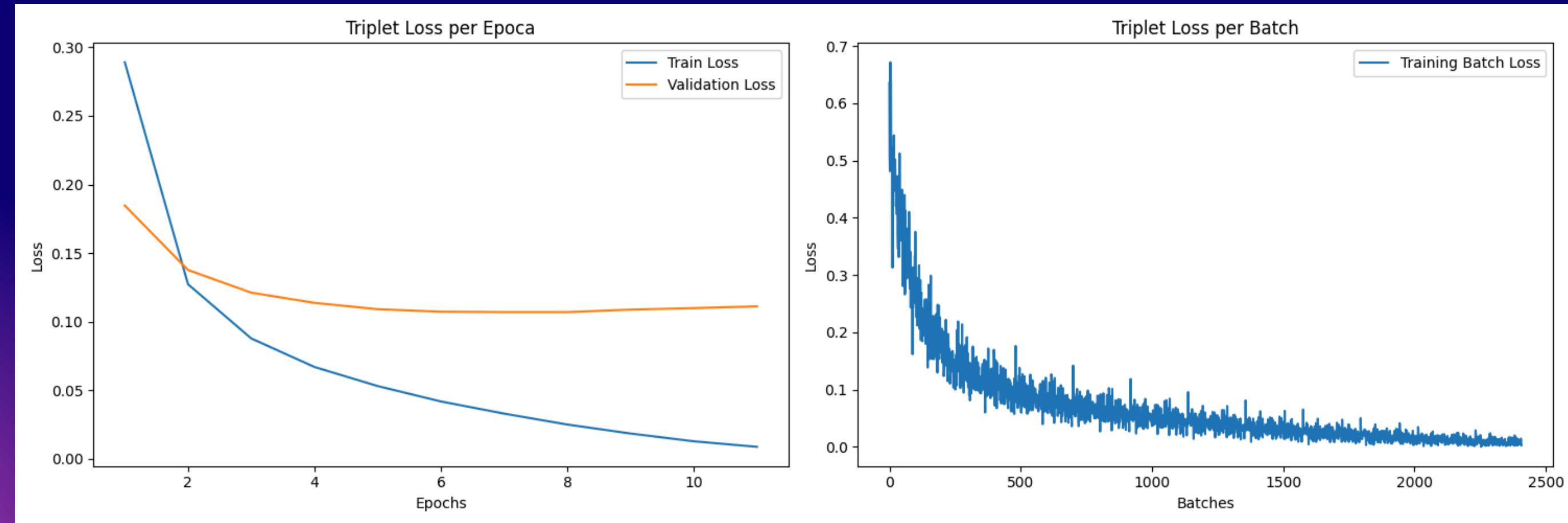
ESTRAZIONE DELLE INFORMAZIONI PER IL DATASET PRO-GAN



Il grafico sottostante mostra l'andamento di train e validation-loss per valutare quanto bene il modello riesce a performare sui dati.



Sia train che validation-loss seguono lo stesso andamento, stabilizzandosi entrambe nelle epoche finali, con un gap minore di 0.03 e suggerendo un leggero overfitting. In conclusione, il grafico riesce a generalizzare bene sui dati a disposizione.



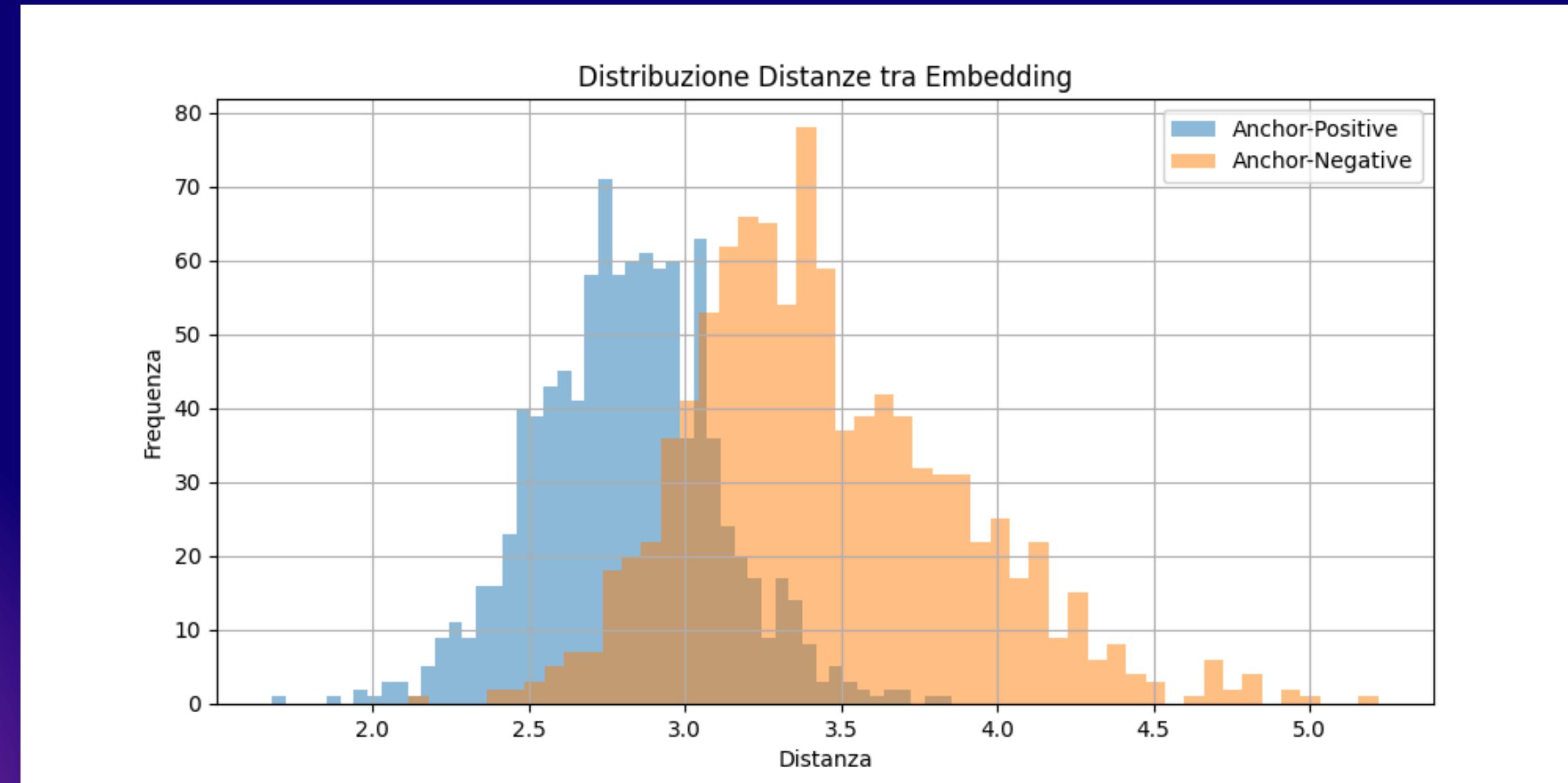
ESTRAZIONE DELLE INFORMAZIONI PER IL DATASET PRO-GAN



Il grafico sottostante rappresenta le distanze tra le triplette per i vettori di embeddings, rispettivamente le distanze tra anchor-positive e anchor negative.



Il grafico suggerisce una netta separazione tra anchor-positive e anchor-negative, confermando in questo modo una buona separazione tra immagini real e fake.



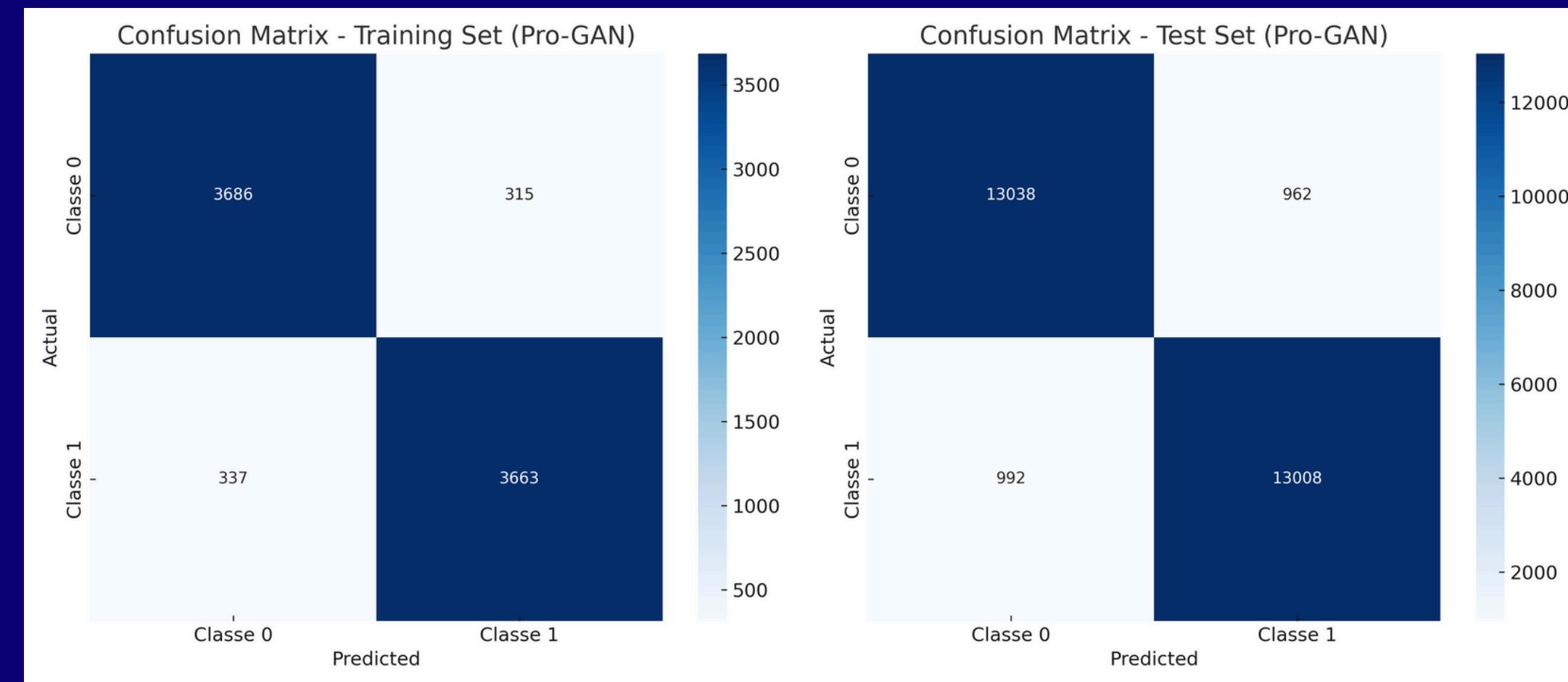
RISULTATI RAGGIUNTI PER IL DATASET PRO-GAN



Anche per questo dataset per l'addestramento è stato utilizzato SVM i cui dati di addestramento sono stati suddivisi in 70% per il training, 20% per il Test Set e il restante 10% per Validation Set.



L'analisi è stata effettuata prendendo in considerazione 40000 campioni, dei quali 28000 sono stati selezionati per il Training Set, 3999 per la Validation Set ed infine 8001 per il Test Set.



Metriche	Percentuale
Accuracy	93.02%
Precision	93.02% - Media tra Precision Classe 0 = 92.93% e Precision Classe 1 = 93.11%
Recall	93.02% - Media tra Recall Classe 0 = 93.13% e Recall Classe 1 = 92.91%
F1-Score	93.02% - Media tra F1-Score Classe 0 = 93.03% e F1-Score Classe 1 = 93.01%

Metriche	Percentuale
Accuracy	91.85%
Precision	91.85% - Media tra Precision Classe 0 = 91.62% e Precision Classe 1 = 92.08%
Recall	91.85% - Media tra Recall Classe 0 = 92.13% e Recall Classe 1 = 91.57%
F1-Score	91.85% - Media tra F1-Score Classe 0 = 91.87% e F1-Score Classe 1 = 91.83%

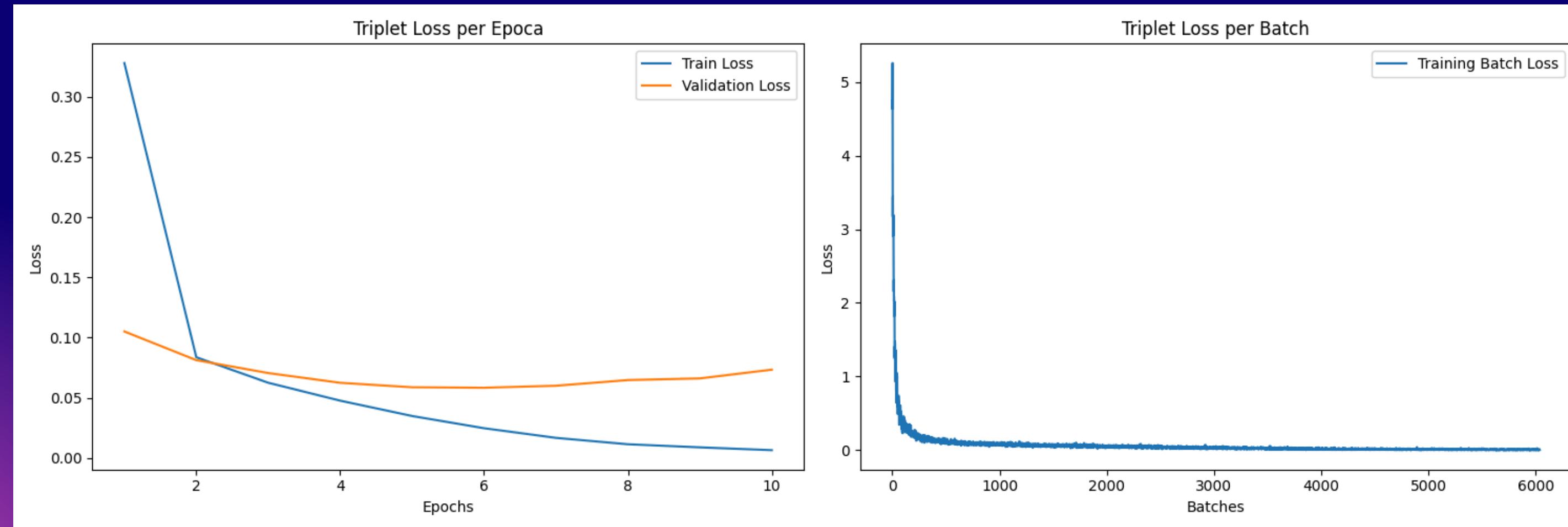
ESTRAZIONE DELLE INFORMAZIONI PER IL DATASET COMBINATO



Il grafico sottostante mostra l'andamento di train e validation-loss per valutare quanto bene il modello riesce a performare sui dati.



La training-loss diminuisce rapidamente nelle prime epoche e in seguito alla sesta epoca, inizia a stabilizzarsi verso un valore basso mentre la validation-loss segue inizialmente l'andamento della training-loss per le prime epoche, ma successivamente si stabilizza dopo la quarta epoca senza migliorare.



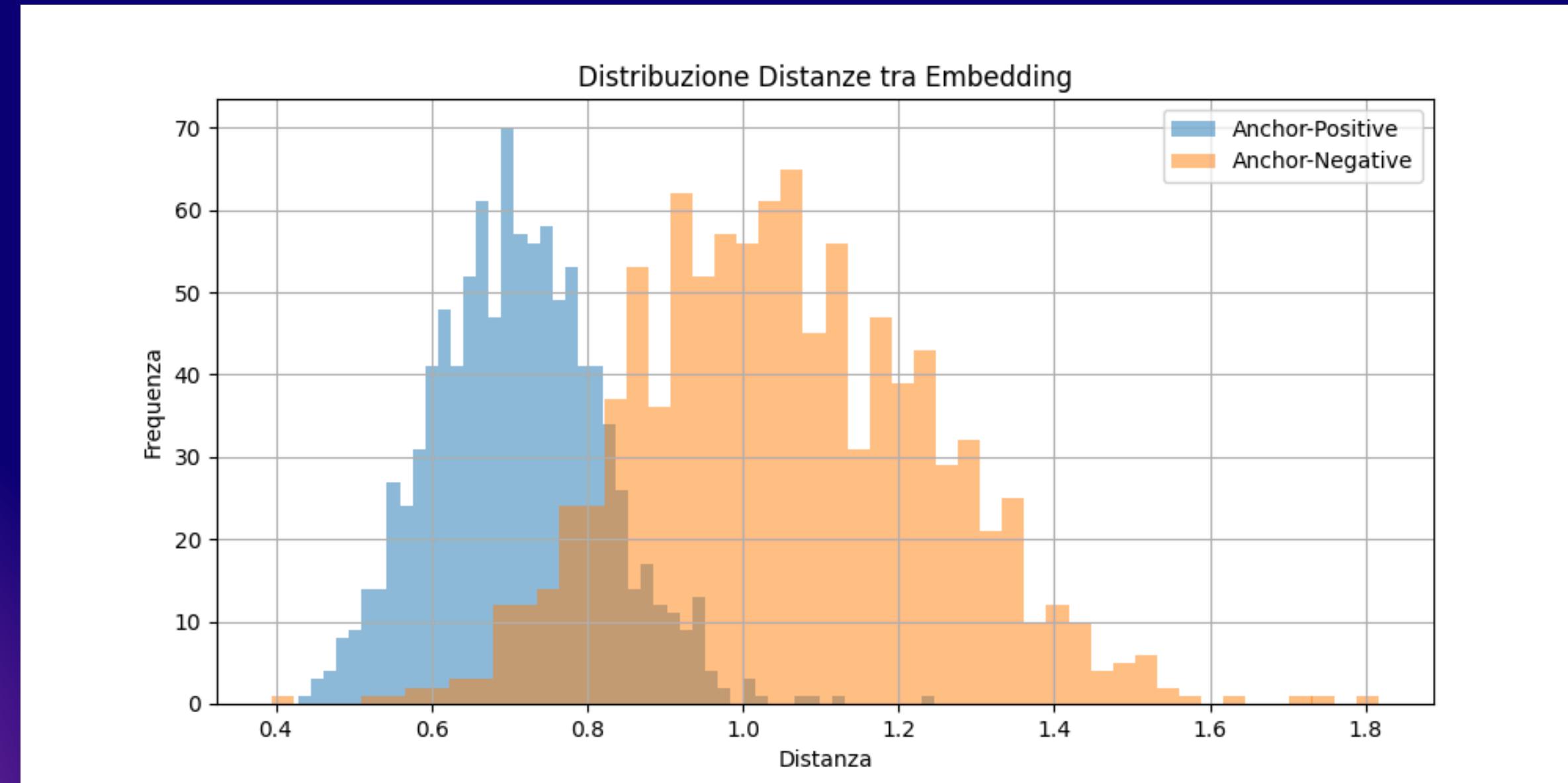
ESTRAZIONE DELLE INFORMAZIONI PER IL DATASET COMBINATO



Il grafico sottostante rappresenta le distanze tra le triplette per i vettori di embeddings, rispettivamente le distanze tra anchor-positive e anchor negative.



Il grafico suggerisce una netta separazione tra anchor-positive e anchor-negative, confermando in questo modo una buona separazione tra immagini real e fake.



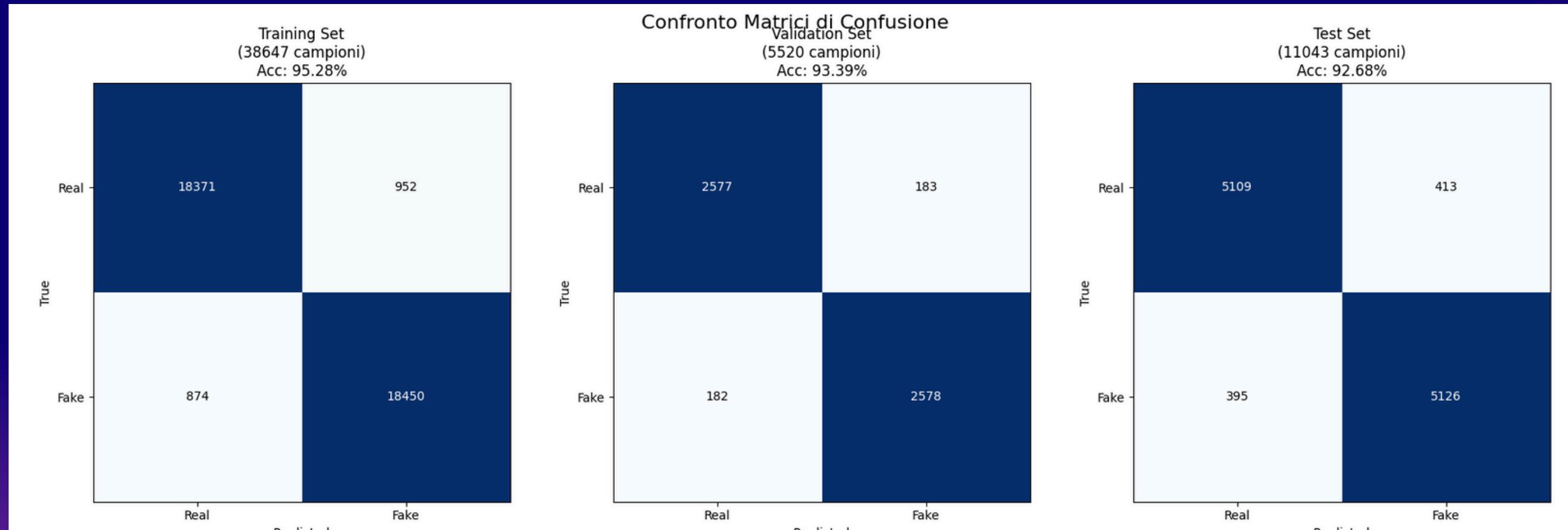
RISULTATI RAGGIUNTI PER IL DATASET COMBINATO



Anche per questo dataset per l'addestramento è stato utilizzato SVM i cui dati di addestramento sono stati suddivisi in 70% per il training, 20% per il Test Set e il restante 10% per Validation Set.



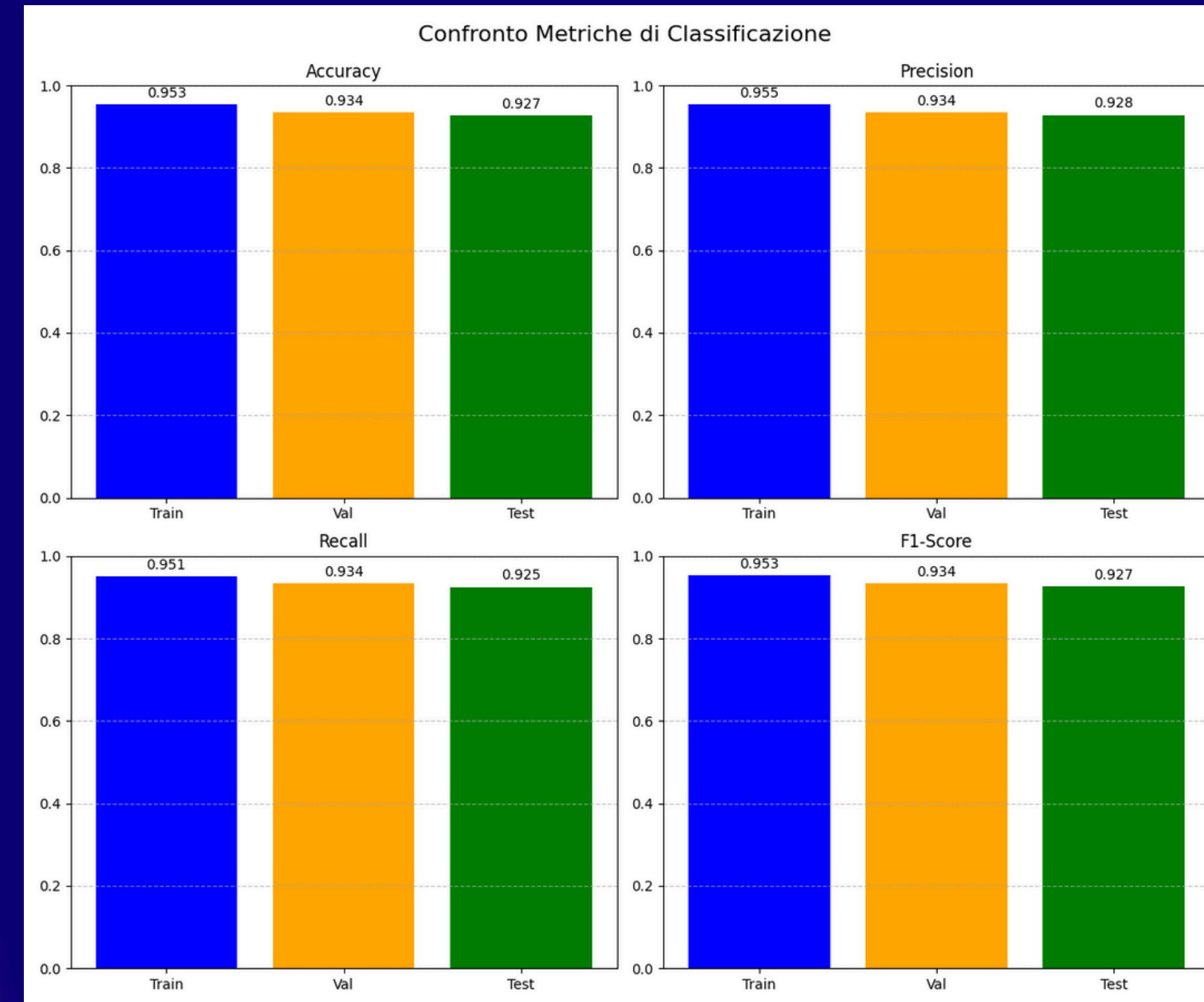
L'analisi è stata effettuata prendendo in considerazione 55210 campioni, dei quali 38647 sono stati selezionati per il Training Set, 5520 per la Validation Set ed infine 11043 per il Test Set. Per ogni set di dati, è stata riportata l'accuracy raggiunta con la corrispondente matrice di confusione ottenuta.



RISULTATI RAGGIUNTI PER IL DATASET COMBINATO



Il seguente grafico mostra le metriche Accuracy, Precision, Recall e F1-Score calcolate per Training, Validation e Test Set con la relativa accuratezza raggiunta.



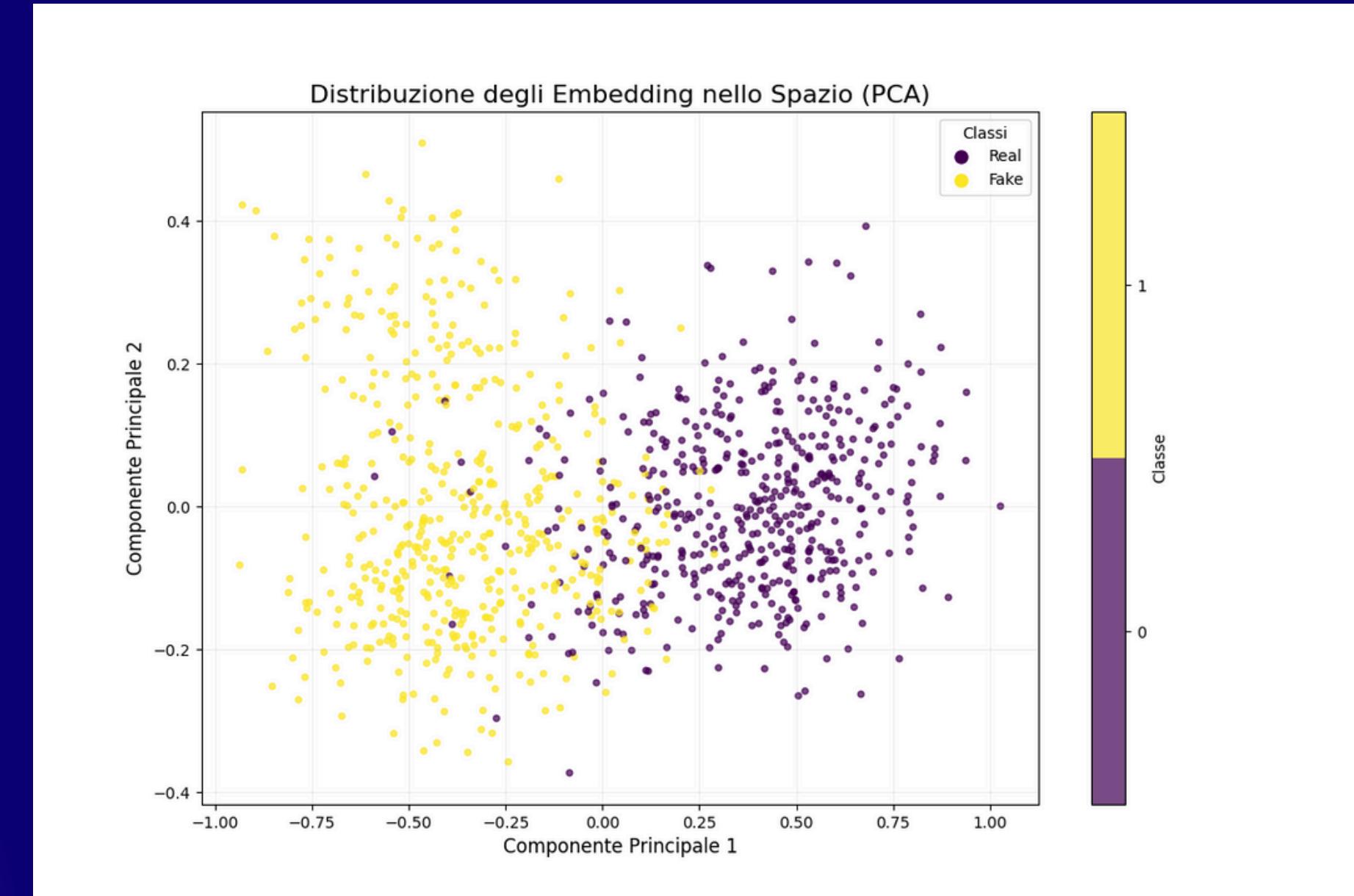
PRINCIPAL COMPONENT ANALYSIS (PCA) PER IL DATASET COMBINATO



Il grafico seguente mostra la distribuzione degli embeddings per le immagini reali e fake utilizzando la PCA con l'obiettivo di visualizzare, quanto bene il modello riesce a differenziare le immagini reali dalle immagini fake.



La PCA è stata effettuata sui dati di testing per una rappresentazione più chiara dei dati e suggerisce una buona separazione tra i due cluster, anche se sono presenti, in alcuni punti, elementi che si mischiano tra loro



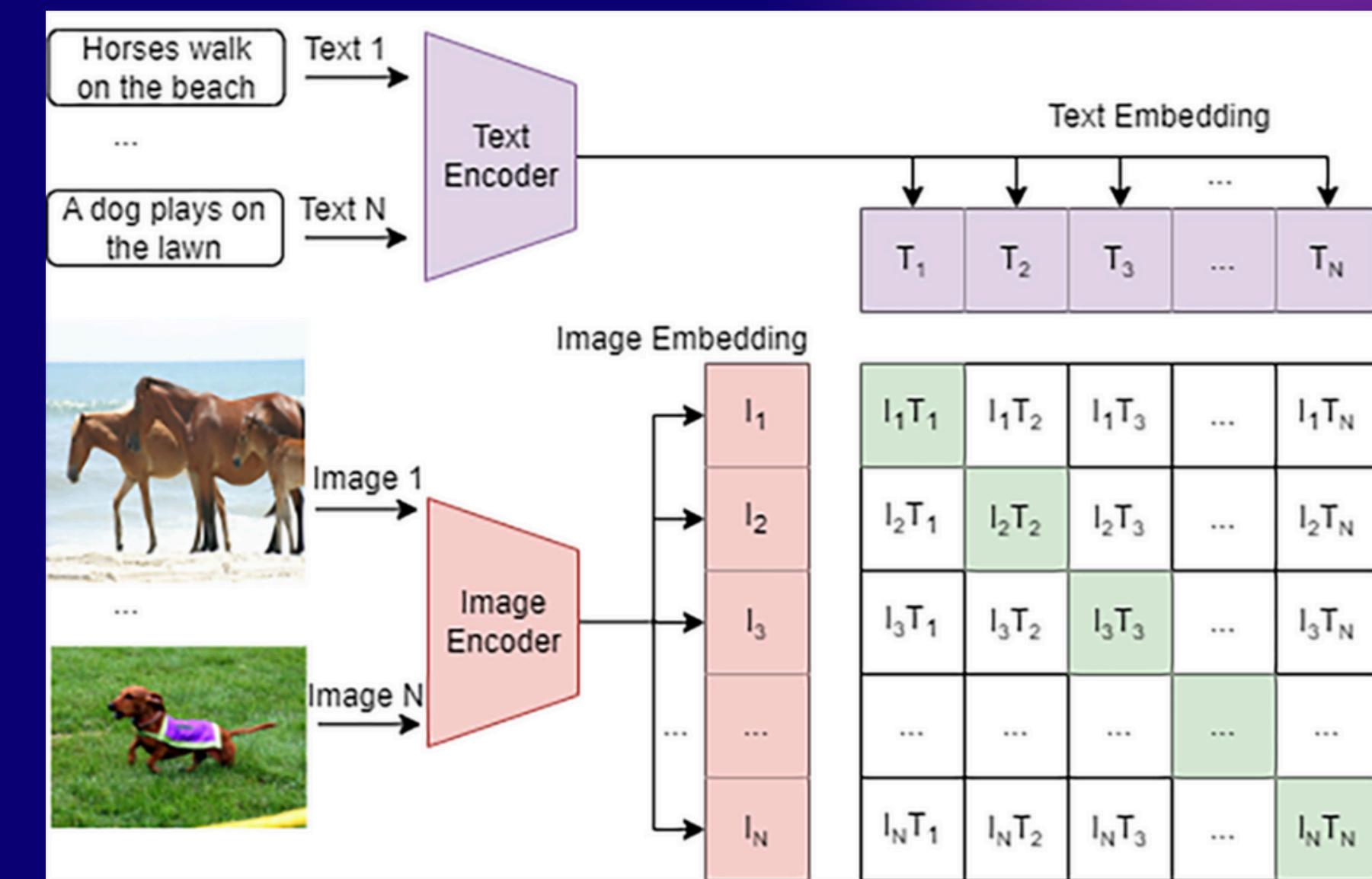
RICONOSCIMENTO DI IMMAGINI REAL E FAKE TRAMITE ESTRAZIONE DELLE INFORMAZIONI DA IMMAGINE E TESTO

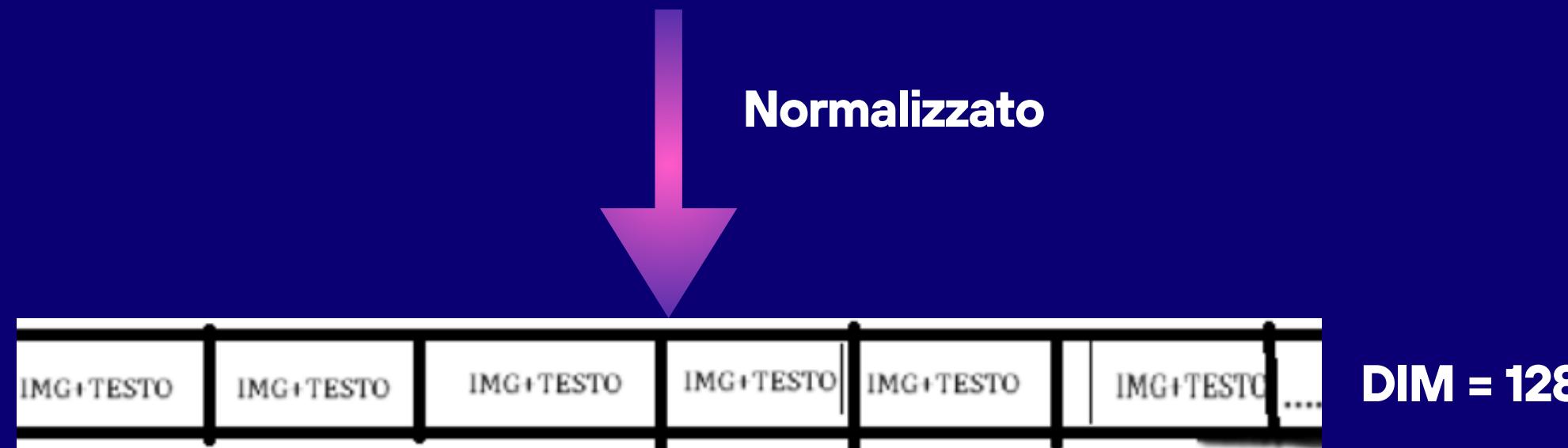
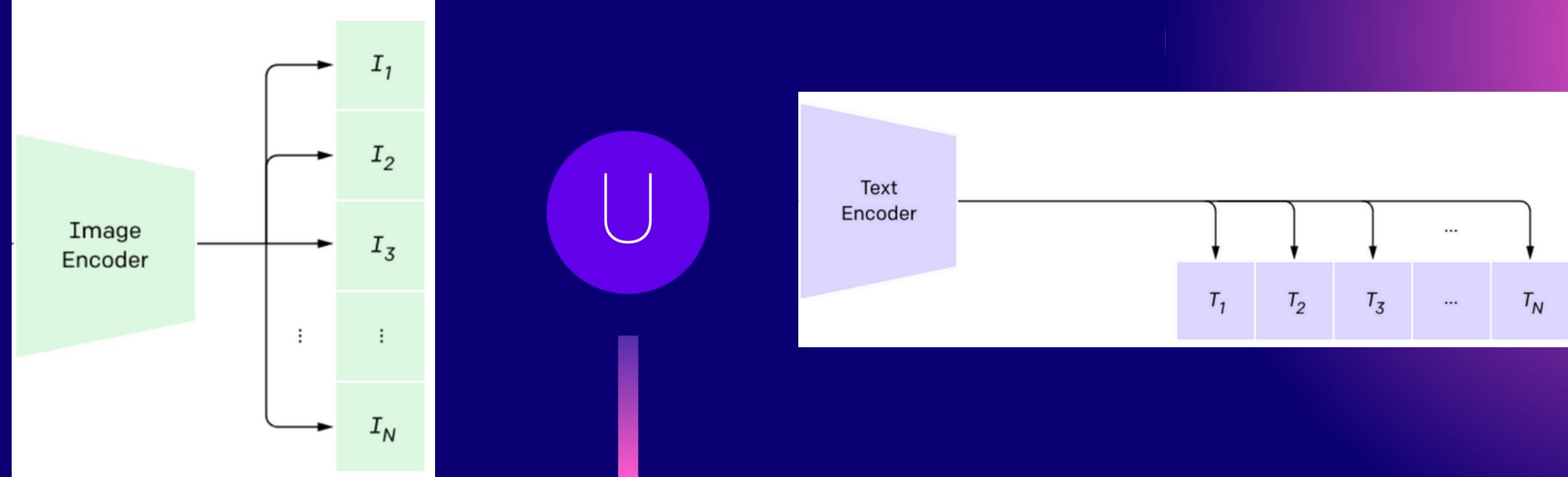


In questa sezione verranno visualizzati i risultati ottenuti effettuando la somma tra le singole componenti contenute negli embeddings.



Sono stati estratti i vettori di embedding dell'immagine, del testo e la combinazione tra i due.





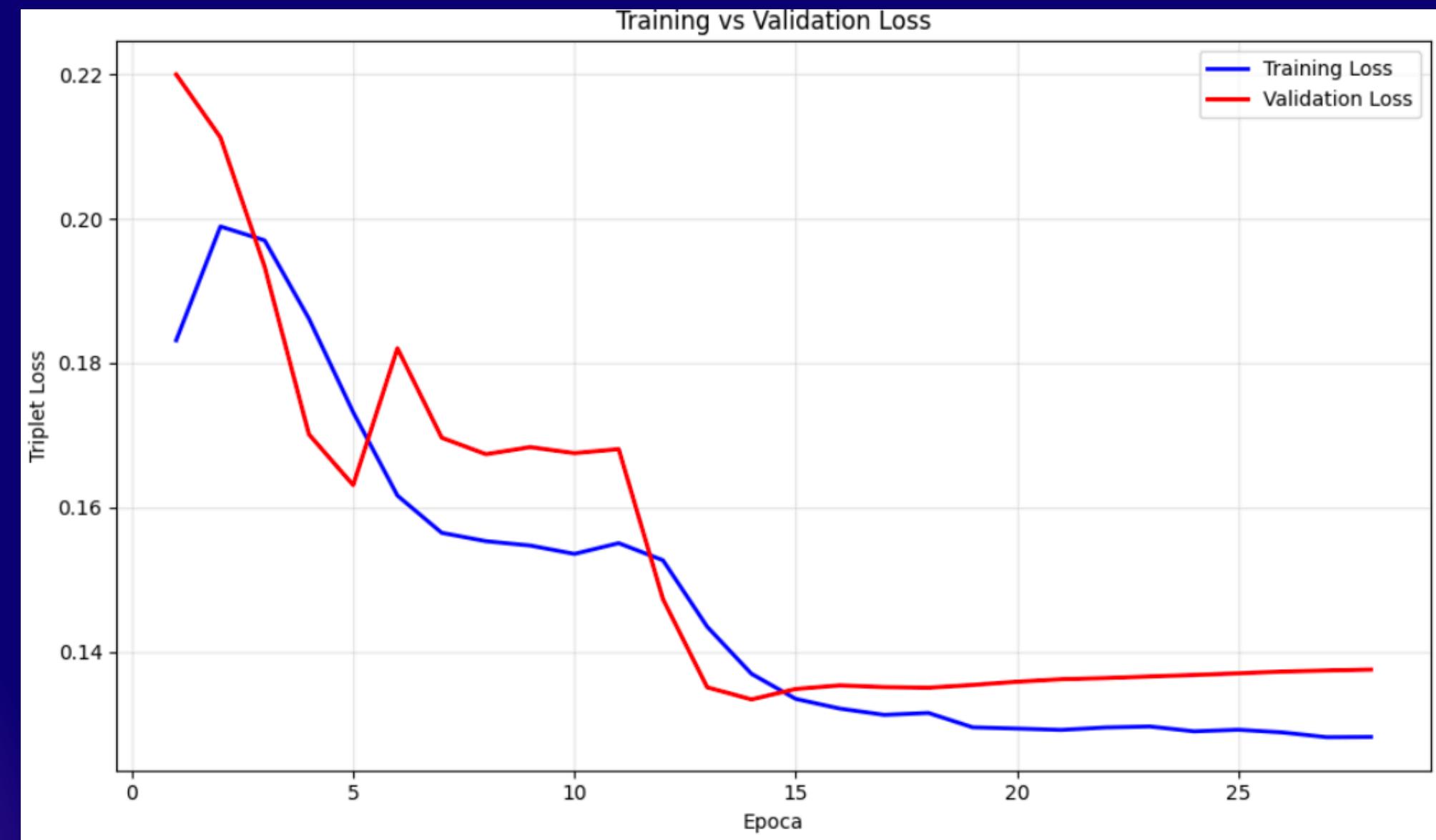
ESTRAZIONE DELLE INFORMAZIONI PER IL DATASET CYCLE-GAN



Il grafico sottostante mostra l'andamento di train e validation-loss per valutare quanto bene il modello riesce a performare sui dati.



si può notare che sia la training che la validation-loss presentano lo stesso andamento per tutte le epocha, in quanto entrambe si stabilizzano a 0.14, indicando che non vi `e presenza di overfitting. Di conseguenza il modello generalizza bene sui dati.



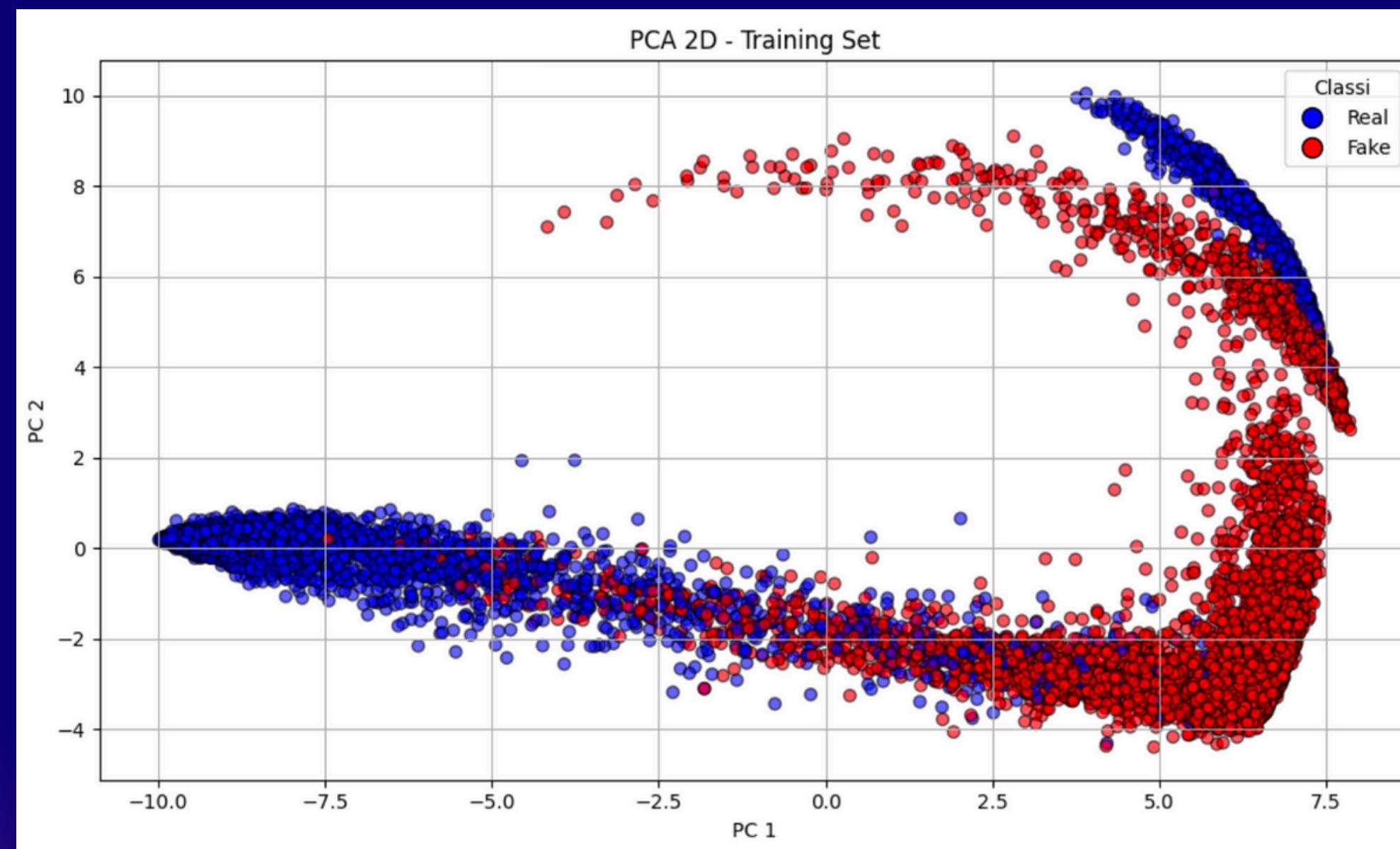
PRINCIPAL COMPONENT ANALYSIS (PCA) PER IL DATASET CYCLE-GAN



Il seguente grafico, invece mostra la distribuzione degli embeddings per le immagini reali e fake utilizzando la PCA.



La PCA è stata effettuata sui dati di training e suggerisce una buona separazione tra immagini reali ed immagini fake, anche se sono presenti alcuni elementi che continuano a mischiarci tra di loro



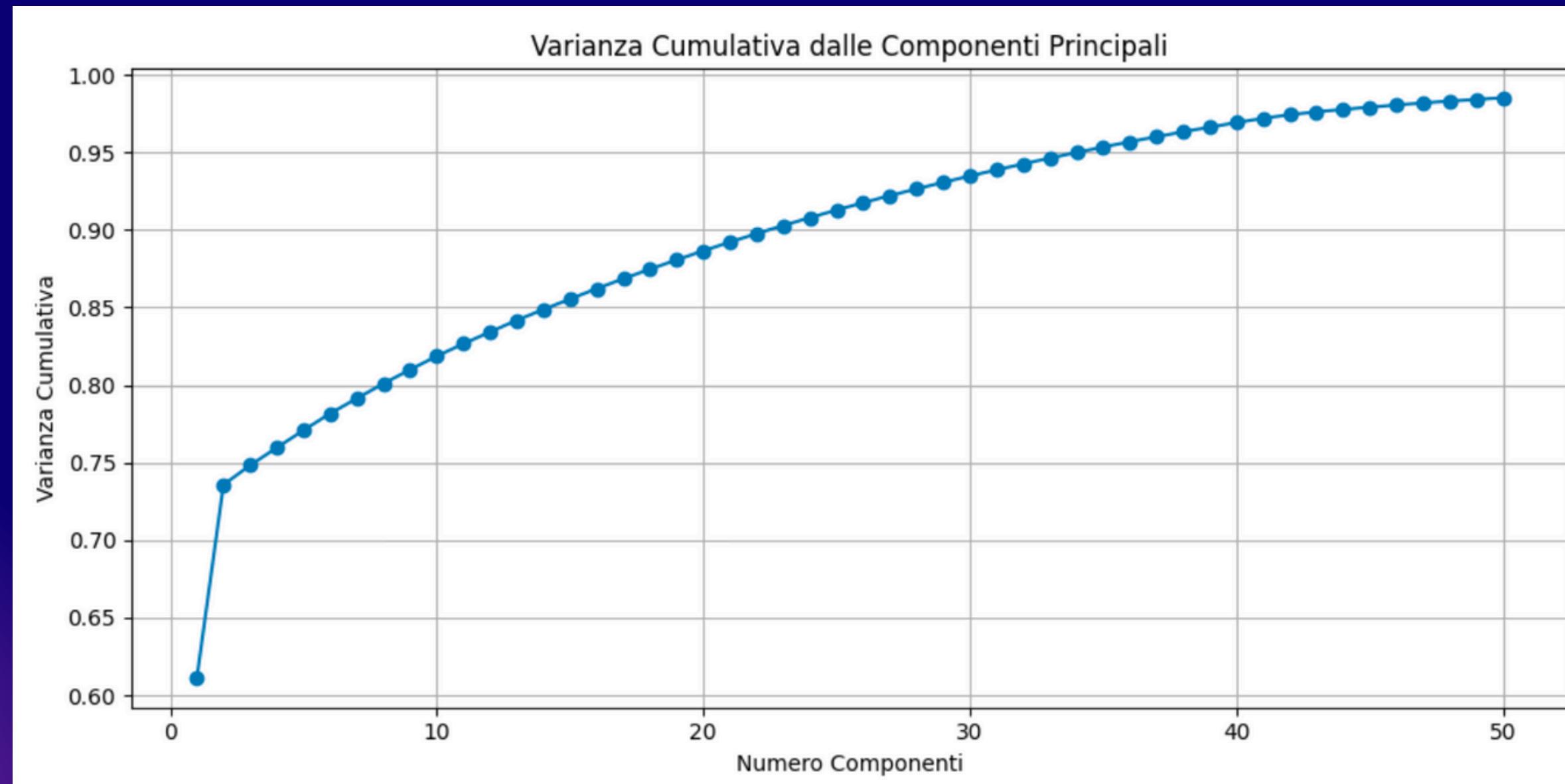
VARIANZA COMULATIVA PER IL DATASET CYCLE-GAN



Dalla PCA è stato effettuato il grafico della varianza cumulativa delle componenti principali.



La prima componente spiega circa il 62% della varianza, con un incremento graduale e uniforme lungo tutte le componenti fino a raggiungere quasi il 100% della varianza cumulativa verso la componente 50.



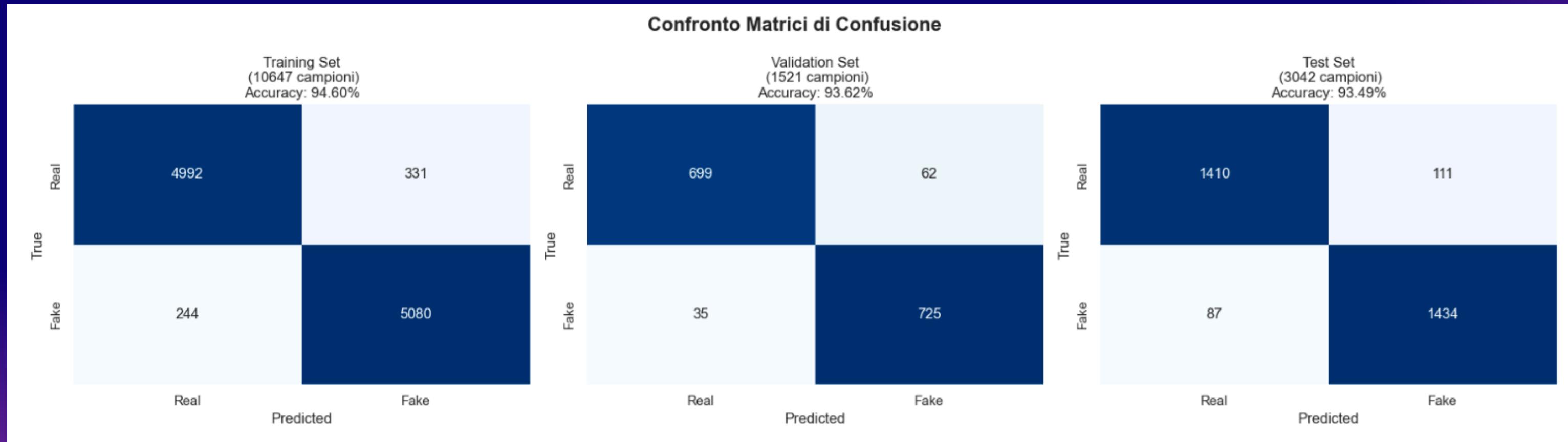
RISULTATI RAGGIUNTI PER IL DATASET CYCLE-GAN



Anche per questo dataset per l'addestramento è stato utilizzato SVM i cui dati di addestramento sono stati suddivisi in 70% per il training, 20% per il Test Set e il restante 10% per Validation Set.



L'analisi è stata effettuata prendendo in considerazione 15210 campioni, dei quali 10647 sono stati selezionati per il Training Set, 1521 per la Validation Set ed infine 3042 per il Test Set.

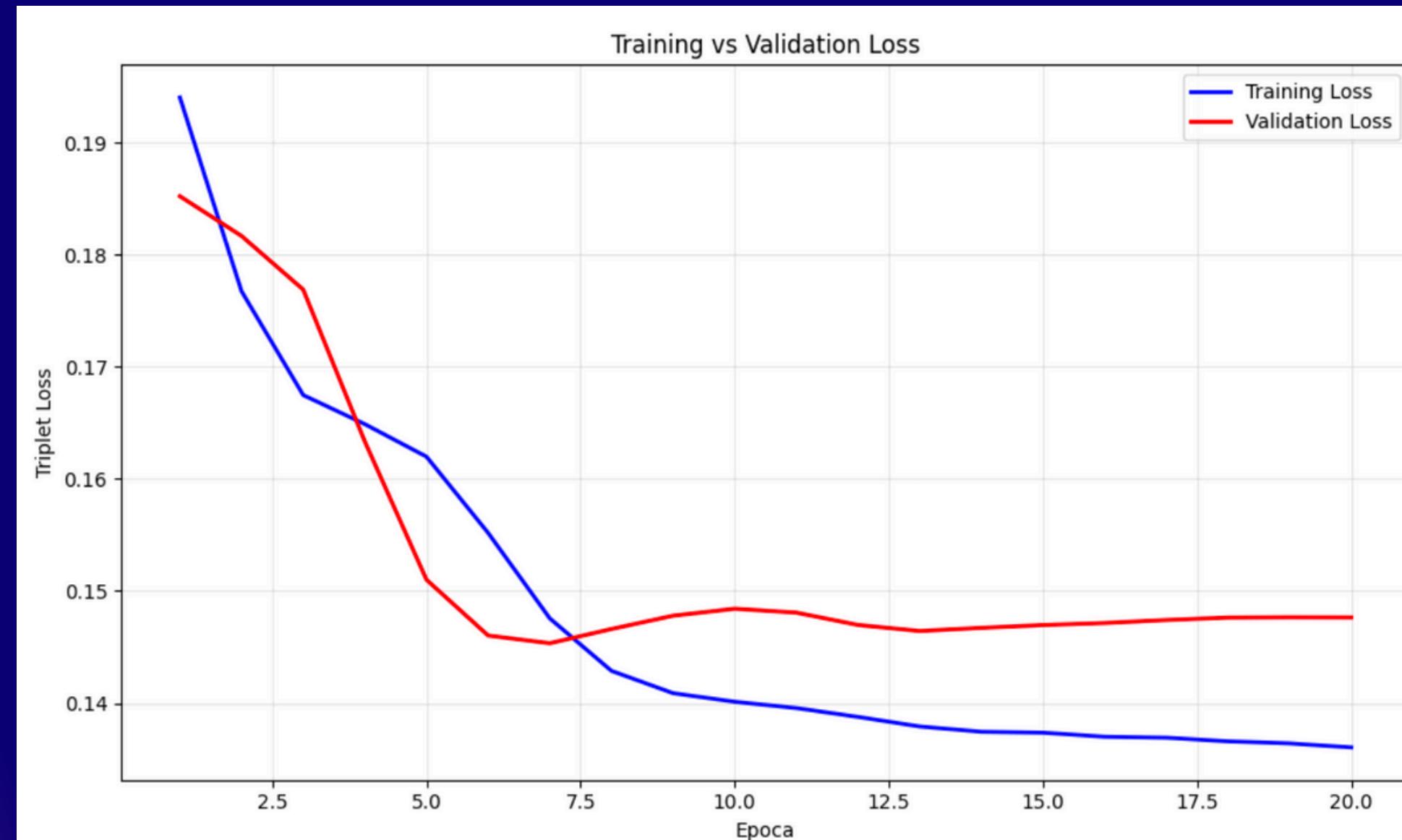


Metriche	Percentuale Training	Percentuale Validation	Percentuale Test Set
Precision	94.61%	93.68%	93.50%
Recall	94.60%	93.62%	93.49%
F1-Score	94.60%	93.62%	93.49%

ESTRAZIONE DELLE INFORMAZIONI PER IL DATASET PRO-GAN

 Il grafico sottostante mostra l'andamento di train e validation-loss per valutare quanto bene il modello riesce a performare sui dati.

 Dal grafico si evince vi è assenza di overfitting dato che, sia train che validation decrescono insieme contemporaneamente, difatti all'epoca 20, entrambe convergono allo stesso valore, ovvero 0.14.



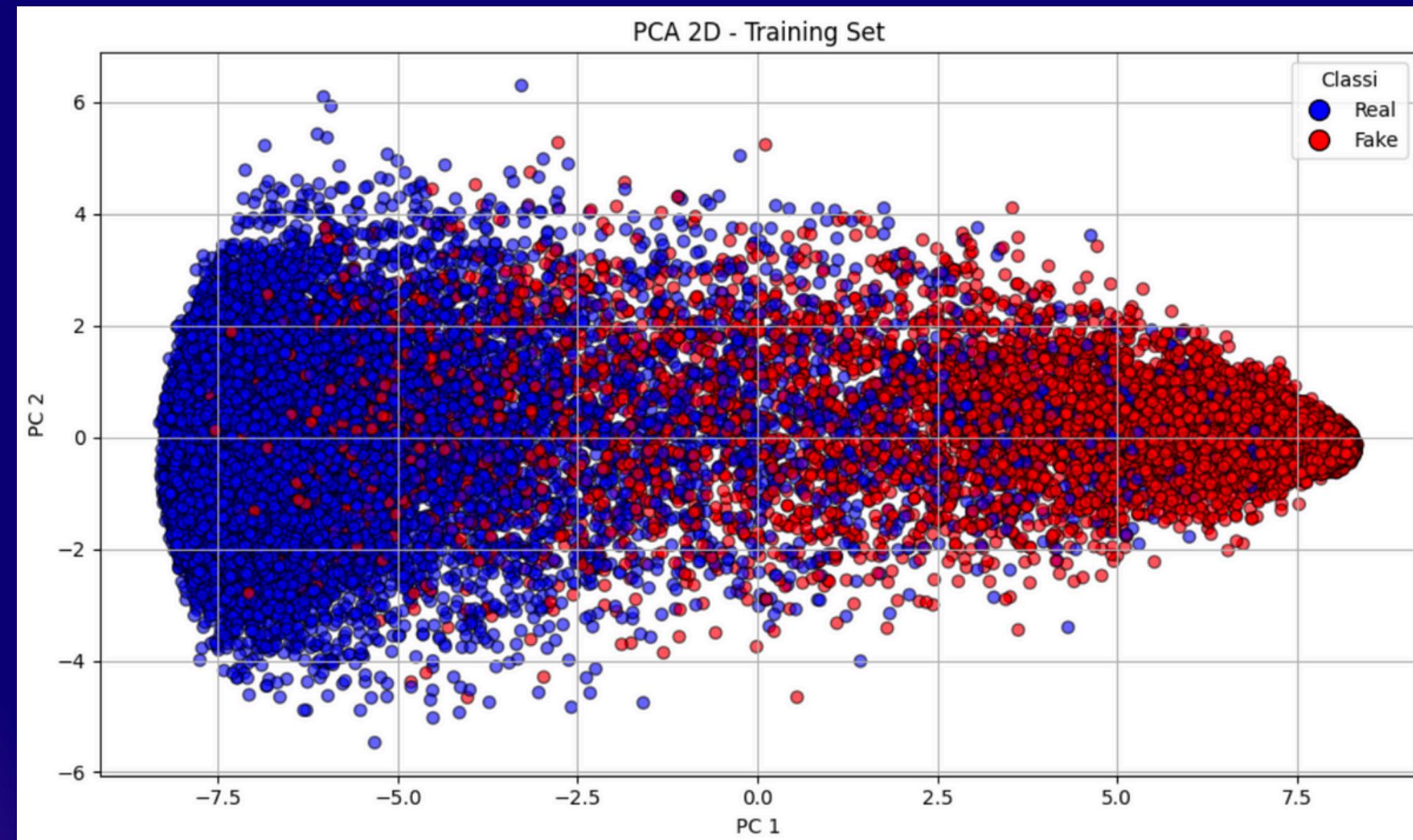
PRINCIPAL COMPONENT ANALYSIS (PCA) PER IL DATASET PRO-GAN



Il seguente grafico, invece mostra la distribuzione degli embeddings per le immagini reali e fake utilizzando la PCA.



La PCA è stata effettuata sui dati di training e suggerisce una buona separazione tra immagini reali ed immagini fake, anche se sono presenti alcuni elementi che continuano a mischiarsi tra di loro.



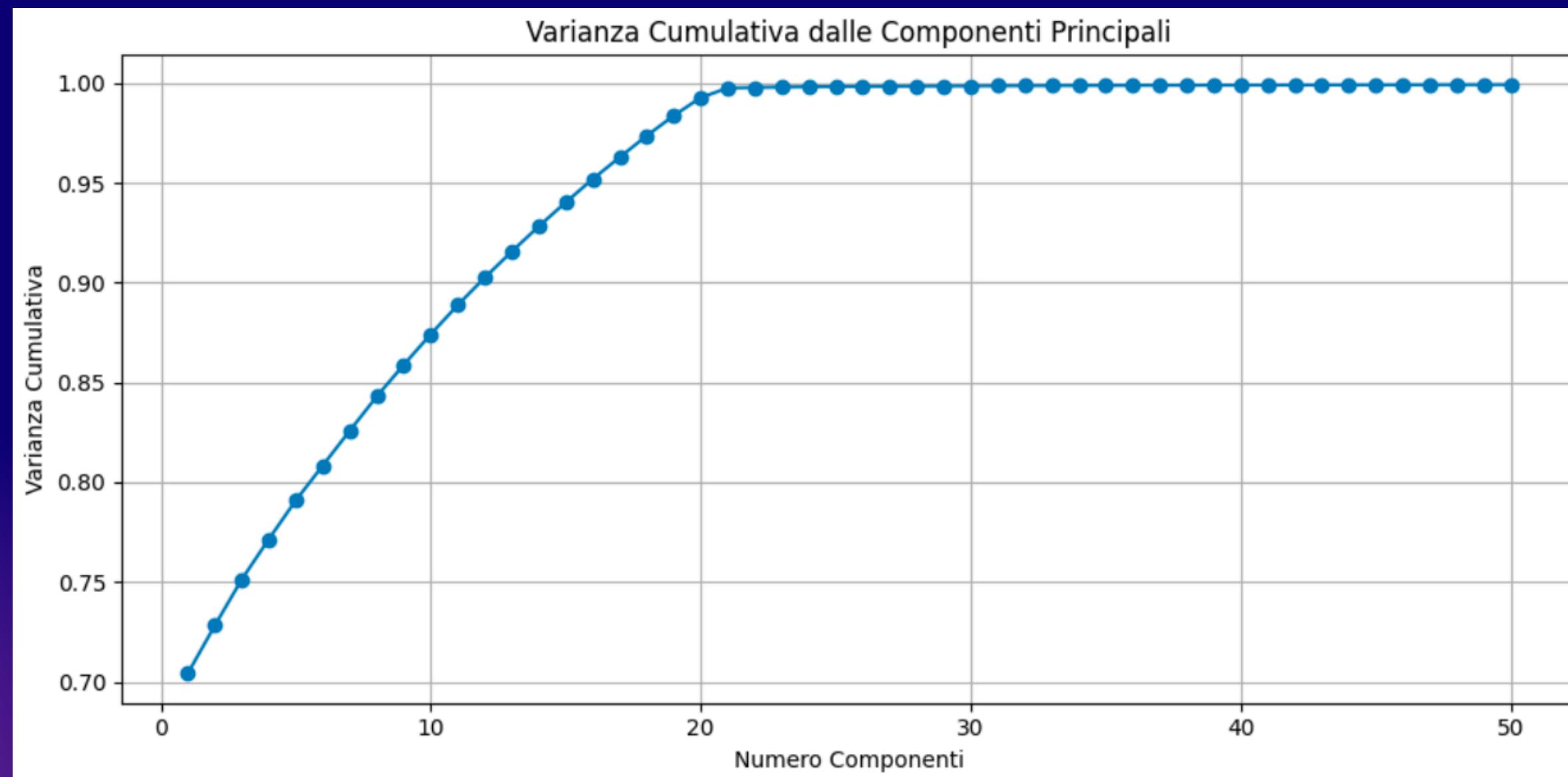
VARIANZA COMULATIVA PER IL DATASET PRO-GAN



Dalla PCA è stato effettuato il grafico della varianza cumulativa delle componenti principali.



La prima componente raggiunge il 70.5% della varianza cumulativa, che di per sé, rappresenta un valore molto alto, successivamente, raggiunge il 100% della varianza già intorno alla componente 20 .



RISULTATI RAGGIUNTI PER IL DATASET PRO-GAN



Anche per questo dataset per l'addestramento è stato utilizzato SVM i cui dati di addestramento sono stati suddivisi in 70% per il training, 20% per il Test Set e il restante 10% per Validation Set.



L'analisi `e stata effettuata prendendo in considerazione 40000 campioni, dei quali 28000 sono stati selezionati per il Training Set, 4000 per la Validation Set ed infine 8000 per il Test Set.



Metriche	Percentuale Training	Percentuale Validation	Percentuale Test Set
Precision	92.99%	92.25%	92.62%
Recall	92.98%	92.25%	92.61%
F1-Score	92.98%	92.25%	92.61%

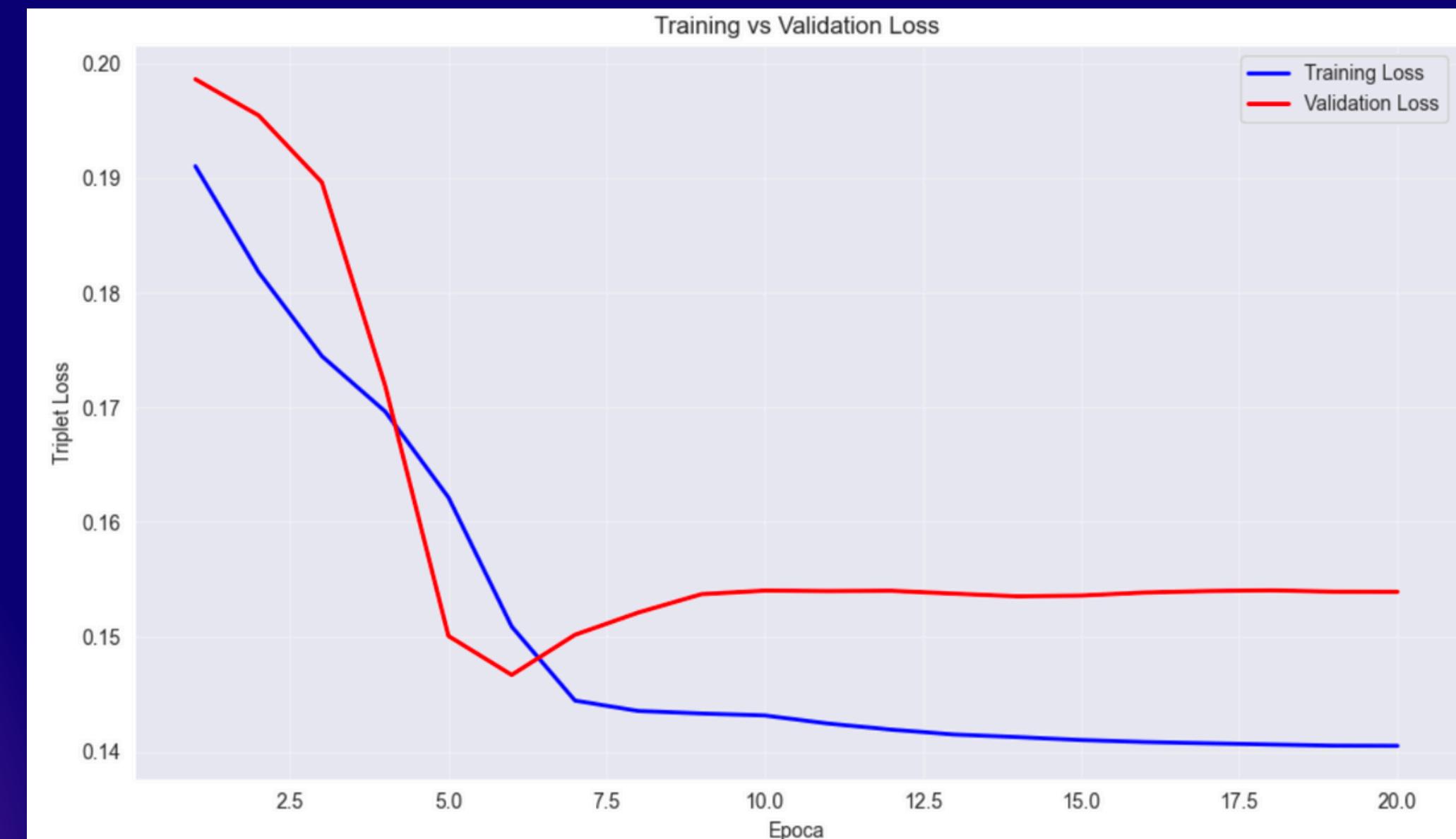
ESTRAZIONE DELLE INFORMAZIONI PER IL DATASET COMBINATO



Il grafico sottostante mostra l'andamento di train e validation-loss per valutare quanto bene il modello riesce a performare sui dati.



la training-loss scende sempre più velocemente, indicando che il modello sta apprendendo efficacemente dai dati di addestramento. Anche la validation-loss segue un andamento simile, fino alla sesta epoca, successivamente inizia ad aumentare, indicando che il modello ha iniziato a sovradattarsi ai dati di training. Questo suggerisce una buona generalizzazione sui dati.



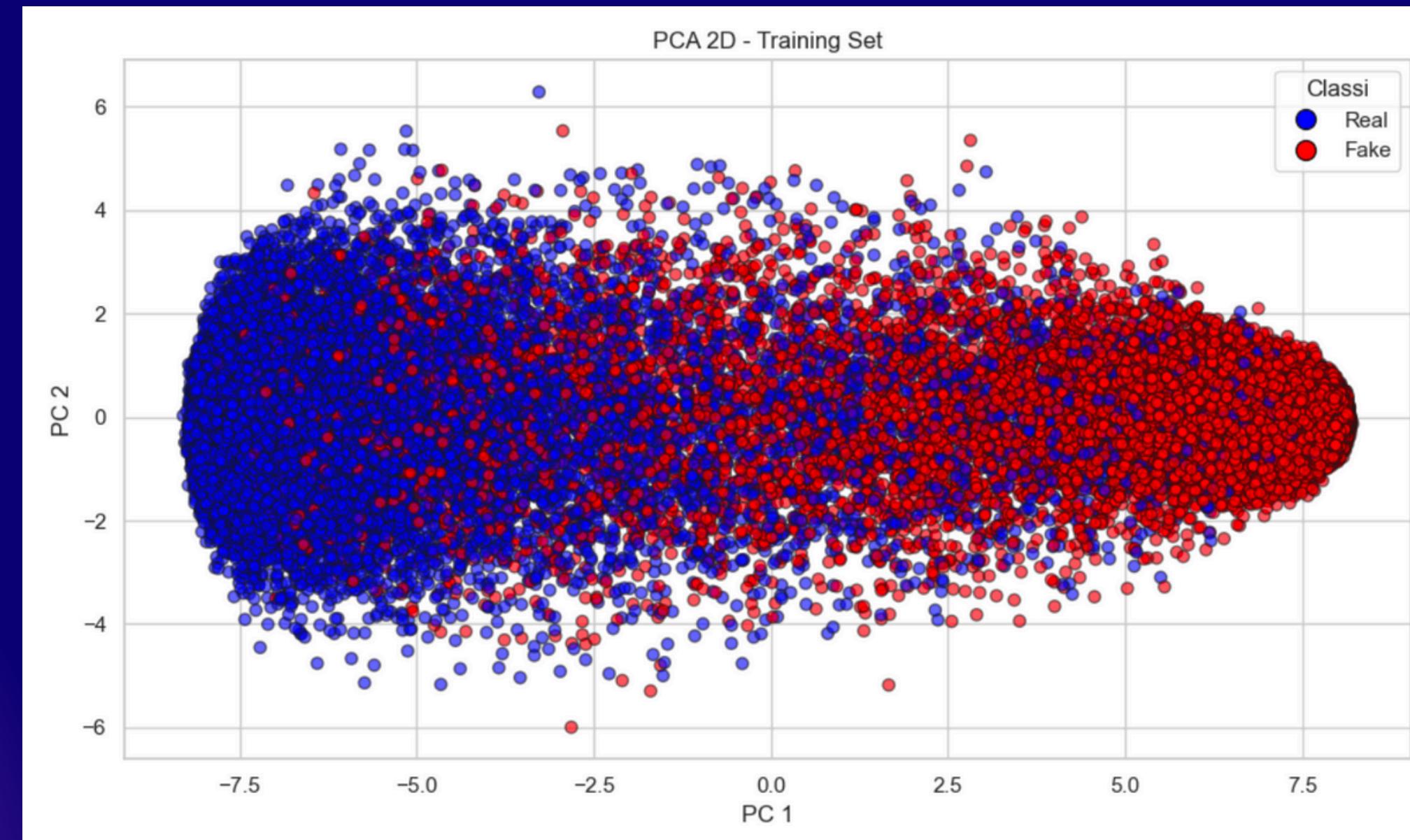
PRINCIPAL COMPONENT ANALYSIS (PCA) PER IL DATASET COMBINATO



Il seguente grafico, invece mostra la distribuzione degli embeddings per le immagini reali e fake utilizzando la PCA.



La PCA è stata effettuata sui dati di training e suggerisce una buona separazione tra immagini reali ed immagini fake, anche se sono presenti alcuni elementi che continuano a mischiarsi tra di loro.



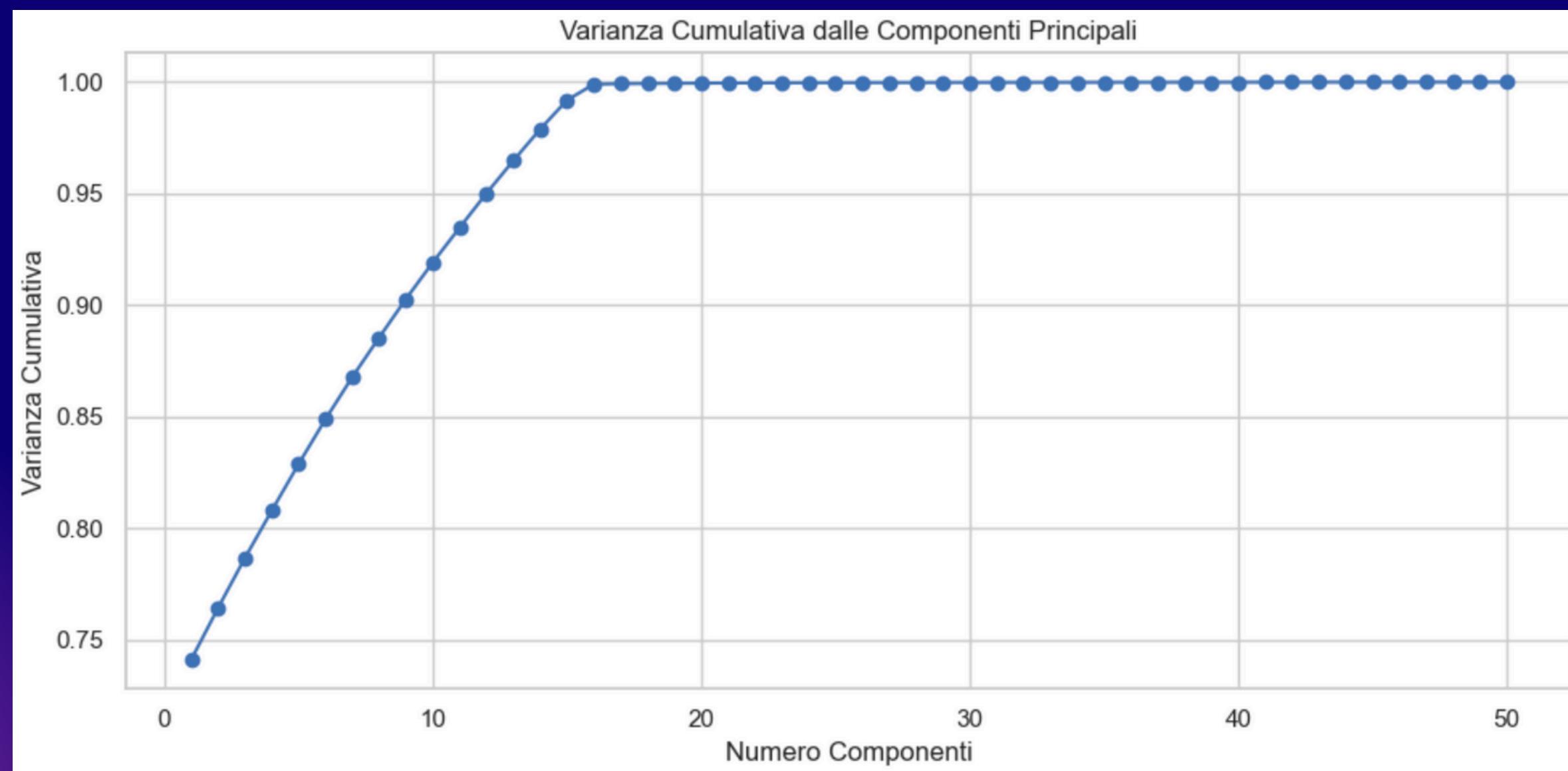
VARIANZA COMULATIVA PER IL DATASET COMBINATO



Dalla PCA è stato effettuato il grafico della varianza cumulativa delle componenti principali.



La prima componente spiega circa il 74% della varianza, aggiungendo le prime 10 componenti, la varianza cumulativa spiegata supera il 90%. Dopo circa 17 componenti, la curva diventa quasi piatta e raggiunge il 100%.



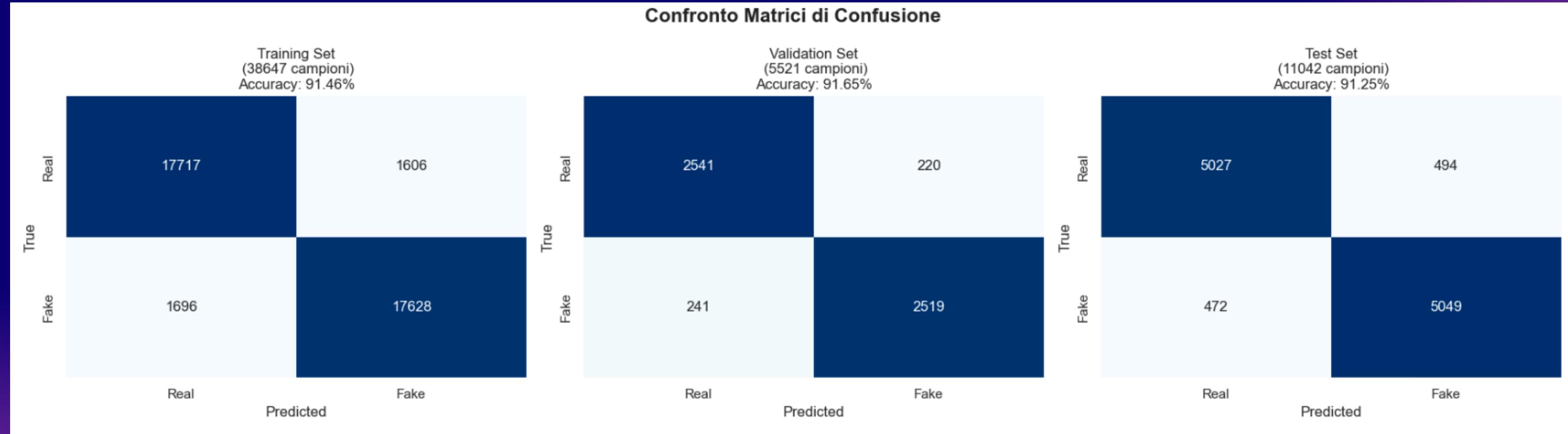
RISULTATI RAGGIUNTI PER IL DATASET COMBINATO



Anche per questo dataset per l'addestramento è stato utilizzato SVM i cui dati di addestramento sono stati suddivisi in 70% per il training, 20% per il Test Set e il restante 10% per Validation Set.



L'analisi è stata effettuata prendendo in considerazione 55210 campioni, dei quali 38647 sono stati selezionati per il Training Set, 5521 per la Validation Set ed infine 11042 per il Test Set.



Metriche	Percentuale Training	Percentuale Validation	Percentuale Test Set
Precision	91.46%	91.65%	91.25%
Recall	91.46%	91.65%	91.25%
F1-Score	91.46%	91.65%	91.25%

RICONOSCIMENTO DI IMMAGINI REAL E FAKE TRAMITE TESTO E IMMAGINE

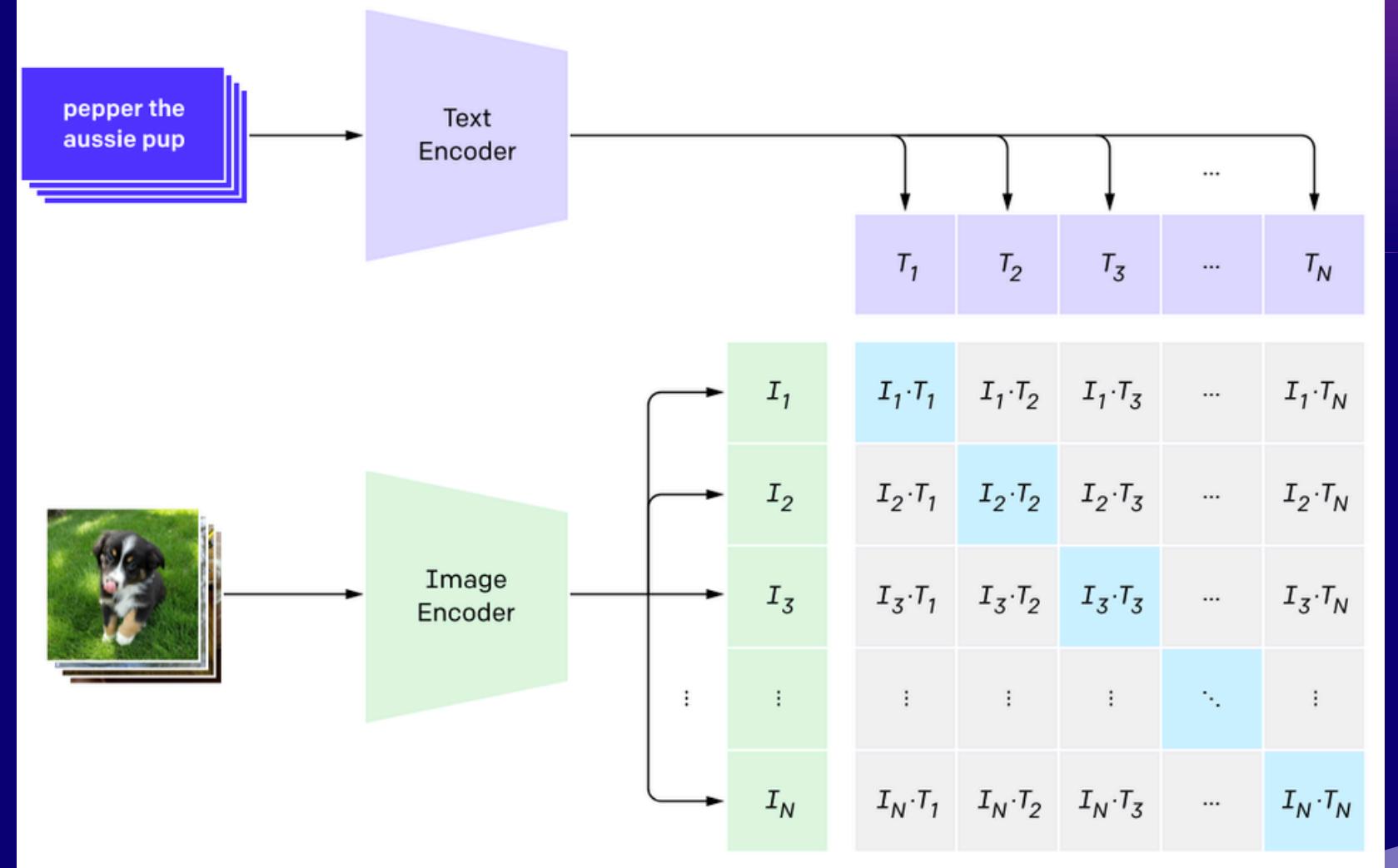


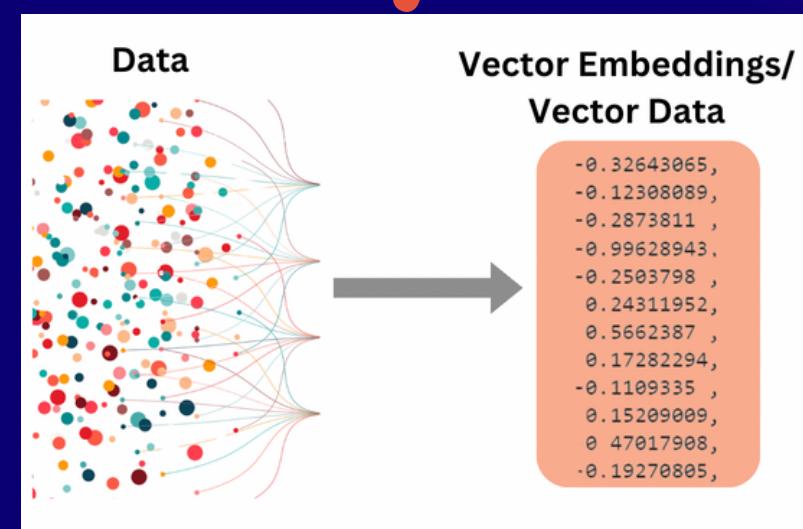
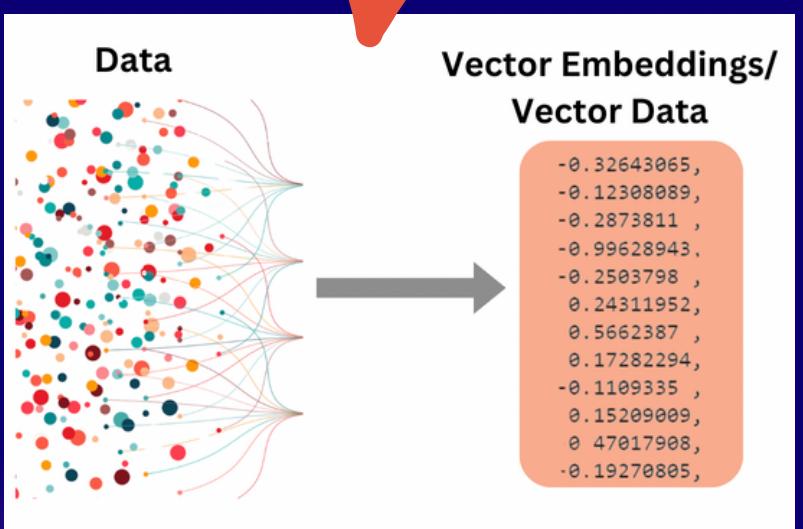
In questa sezione verranno visualizzati i risultati ottenuti effettuando il prodotto tra le singole componenti contenute negli embeddings.



Sono stati estratti i vettori di embedding dell'immagine, del testo e la combinazione tra i due.

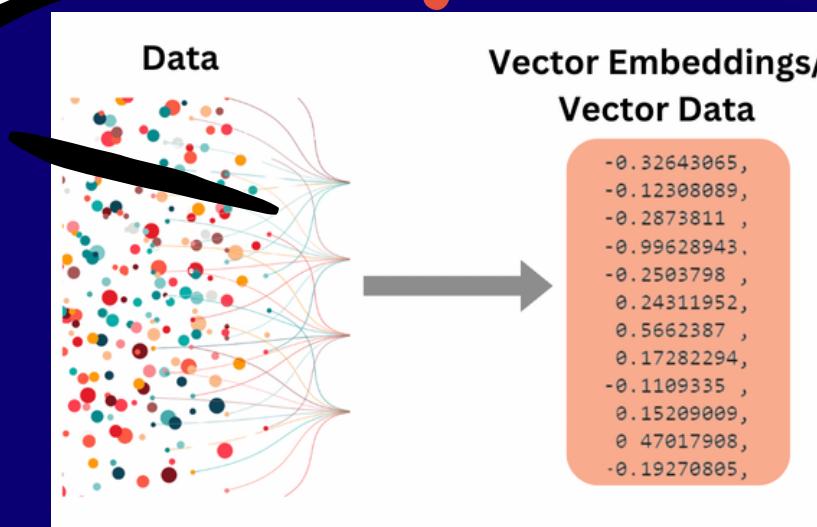
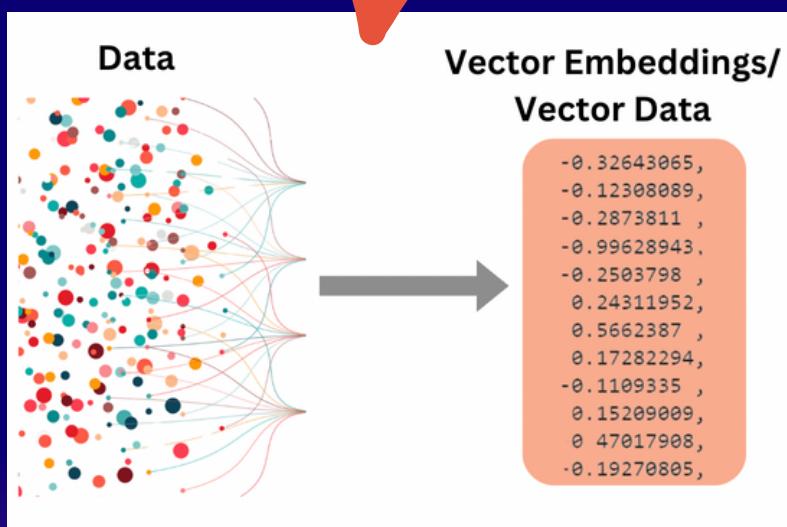
1. Contrastive pre-training







Entrambi gli embeddings vengono
normalizzati in vettori da 128D.



IMG 1

IMG 2

.....

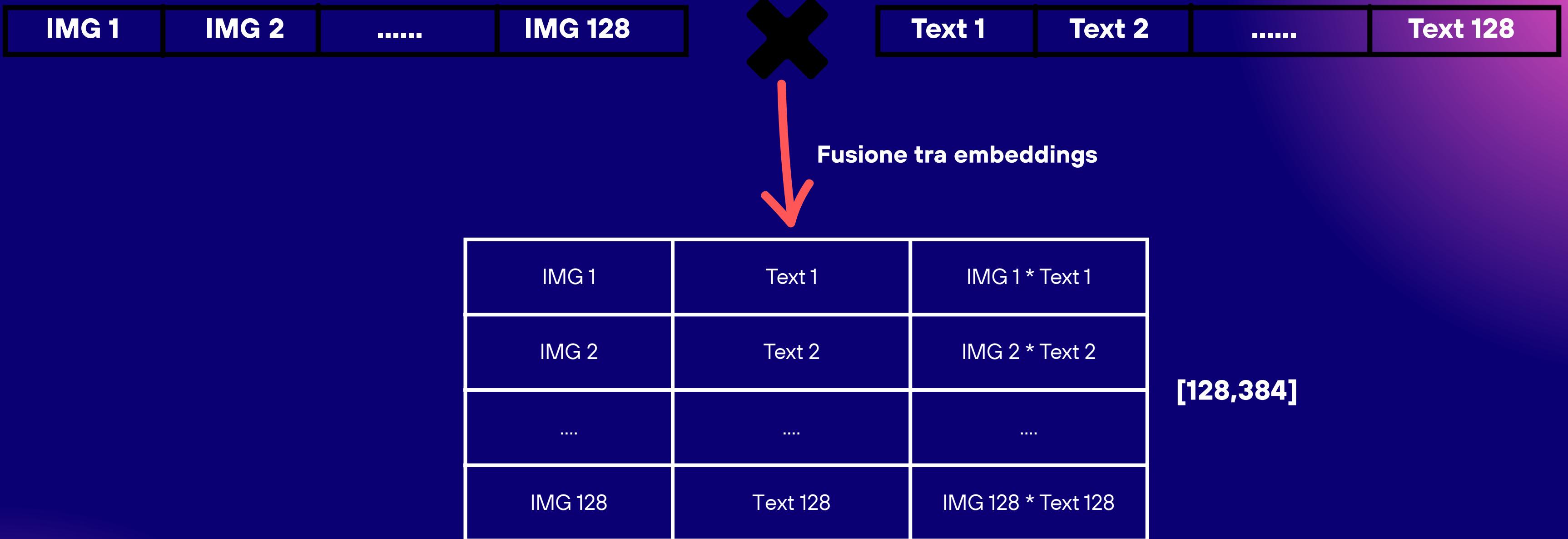
IMG 512

Text 1

Text 2

.....

Text 512



L'embedding ottenuto viene nuovamente normalizzato a 128D.

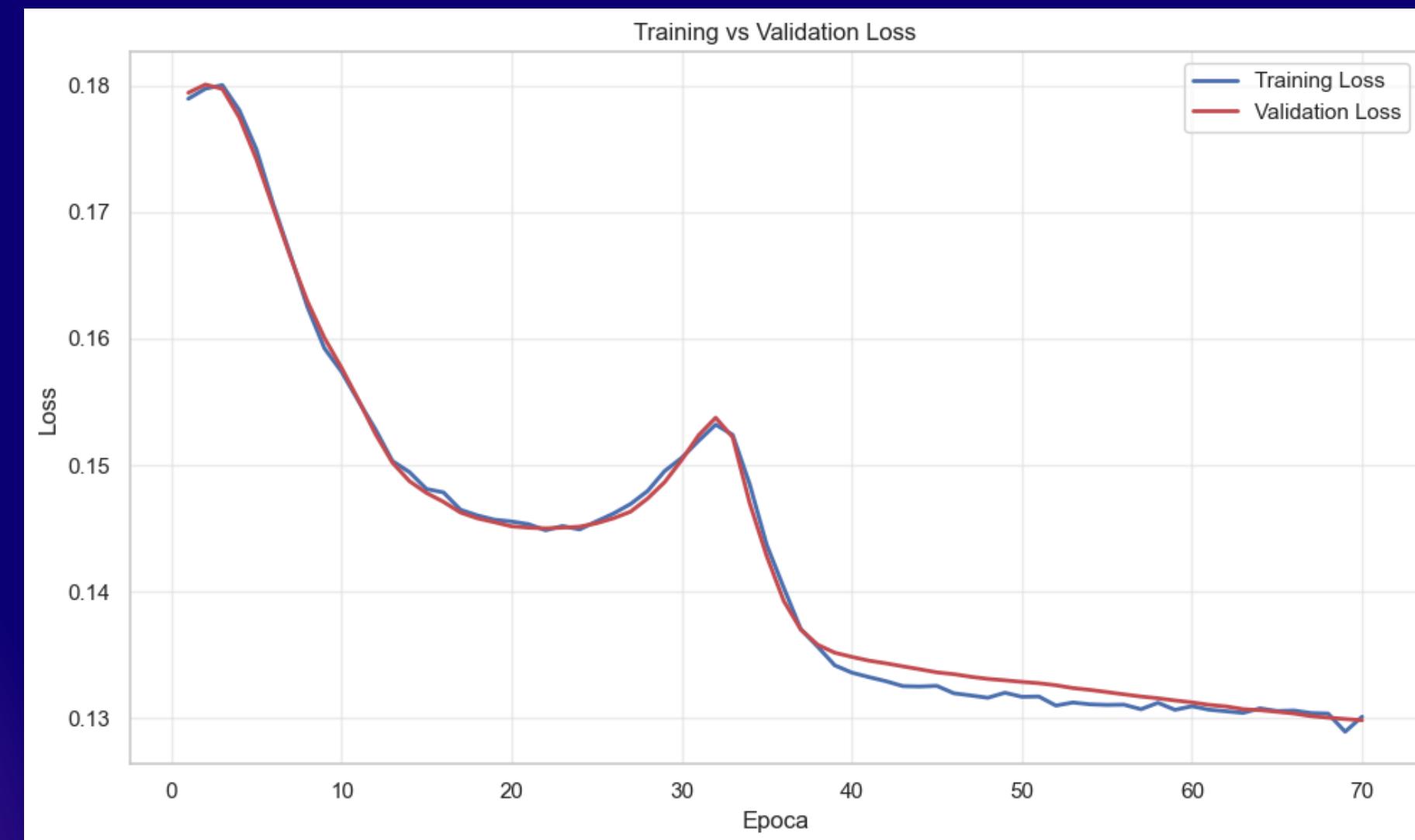
ESTRAZIONE DELLE INFORMAZIONI PER IL DATASET CYCLE-GAN



Il grafico sottostante mostra l'andamento di train e validation-loss per valutare quanto bene il modello riesce a performare sui dati.



Inizialmente vi è una rapida diminuzione di entrambe le loss da 0.18 a 0.145 nelle prime 20 epoche, successivamente, seguita da un improvviso aumento di entrambe le loss fino a 0.153 tra le epoche 30-32, seugito da un rapido recupero.



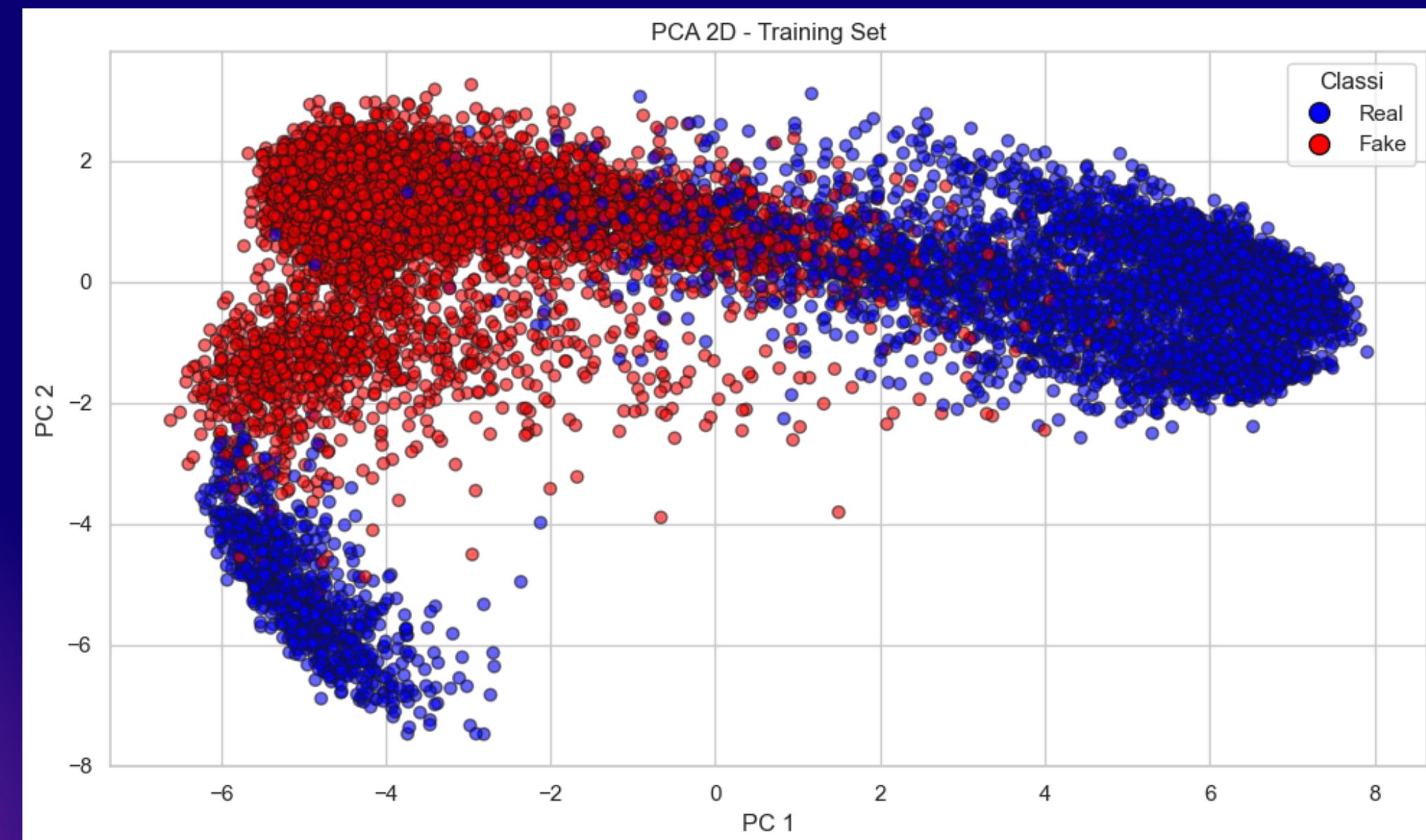
PRINCIPAL COMPONENT ANALYSIS (PCA) PER IL DATASET CYCLE-GAN



Il seguente grafico, invece mostra la distribuzione degli embeddings per le immagini reali e fake utilizzando la PCA.



La PCA è stata effettuata sui dati di training e suggerisce una buona separazione tra immagini reali ed immagini fake, anche se sono presenti alcuni elementi che continuano a mischiarci tra di loro.



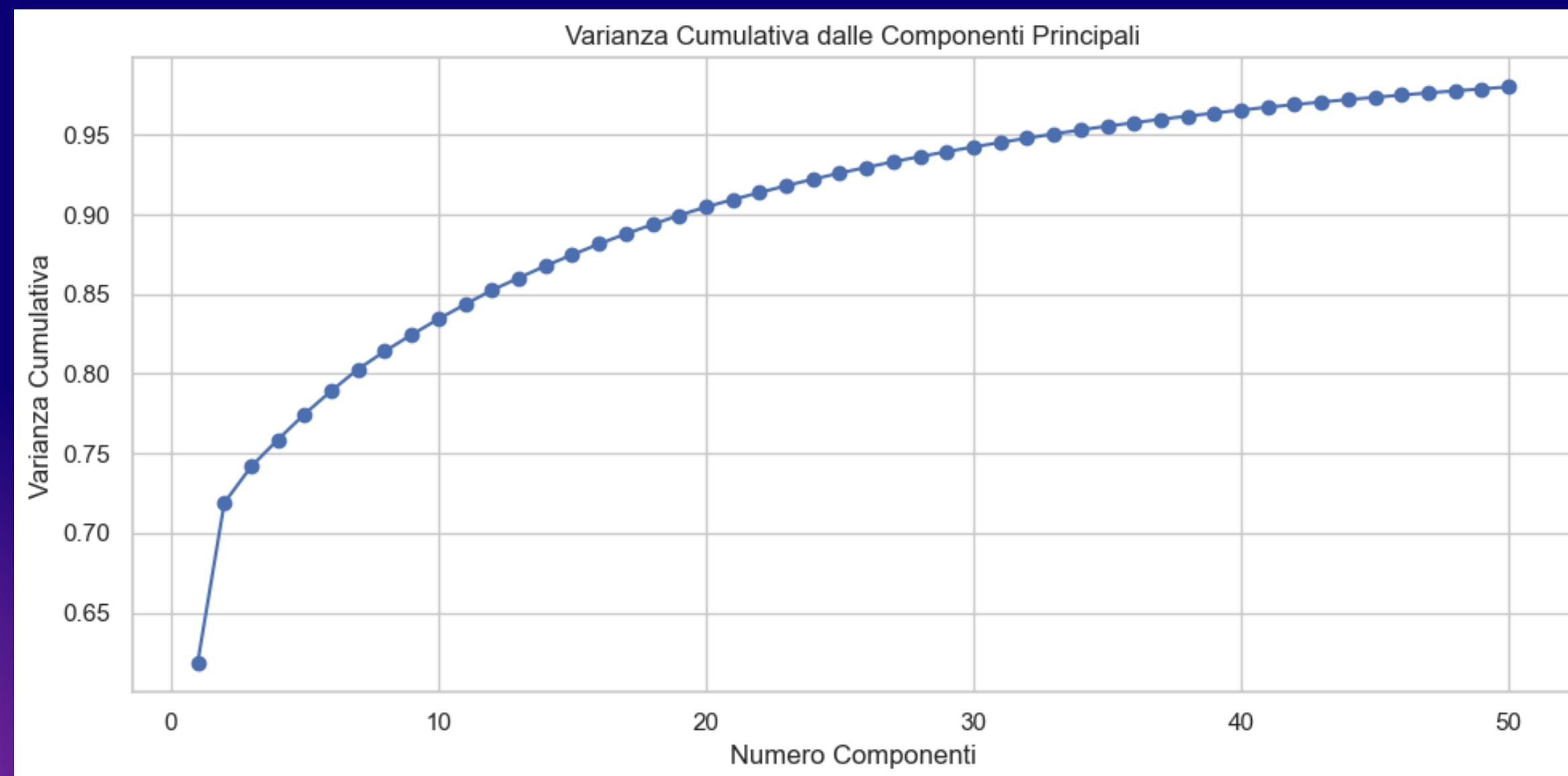
VARIANZA COMULATIVA PER IL DATASET CYCLE-GAN



Dalla PCA è stato effettuato il grafico della varianza cumulativa delle componenti principali.



La prima componente spiega circa il 62% della varianza, con un incremento graduale e uniforme lungo tutte le componenti fino a raggiungere quasi il 100% della varianza cumulativa verso la componente 50.



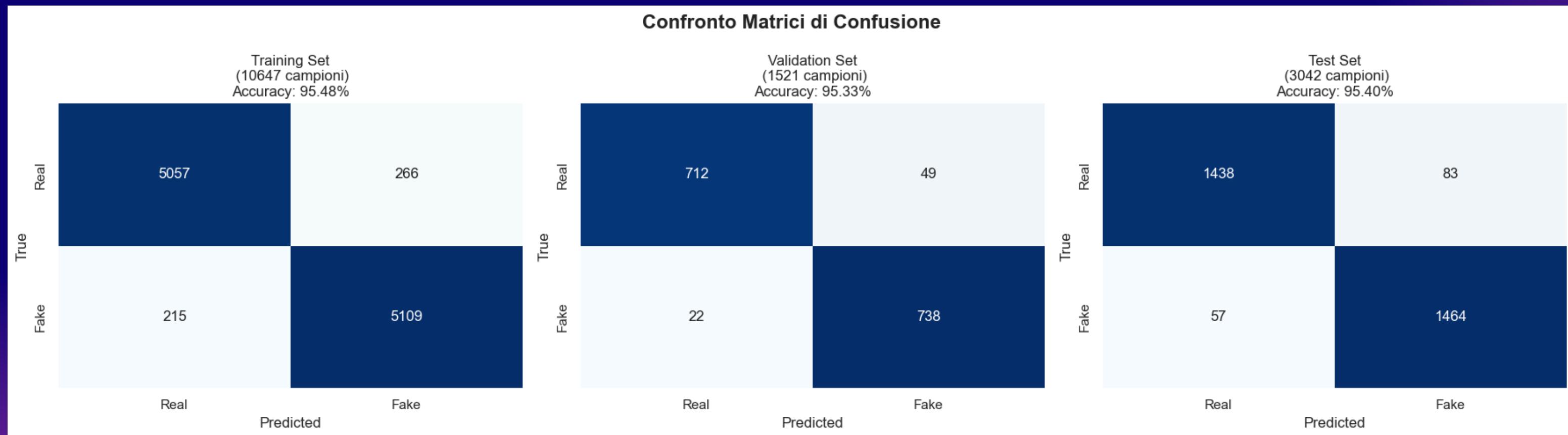
RISULTATI RAGGIUNTI PER IL DATASET CYCLE-GAN



Anche per questo dataset per l'addestramento è stato utilizzato SVM i cui dati di addestramento sono stati suddivisi in 70% per il training, 20% per il Test Set e il restante 10% per Validation Set.



L'analisi è stata effettuata prendendo in considerazione 15210 campioni, dei quali 10647 sono stati selezionati per il Training Set, 1521 per la Validation Set ed infine 3042 per il Test Set.



Metriche	Percentuale Training	Percentuale Validation	Percentuale Test Set
Precision	95.49%	95.39%	95.41%
Recall	95.48%	95.33%	95.40%
F1-Score	95.48%	95.33%	95.40%

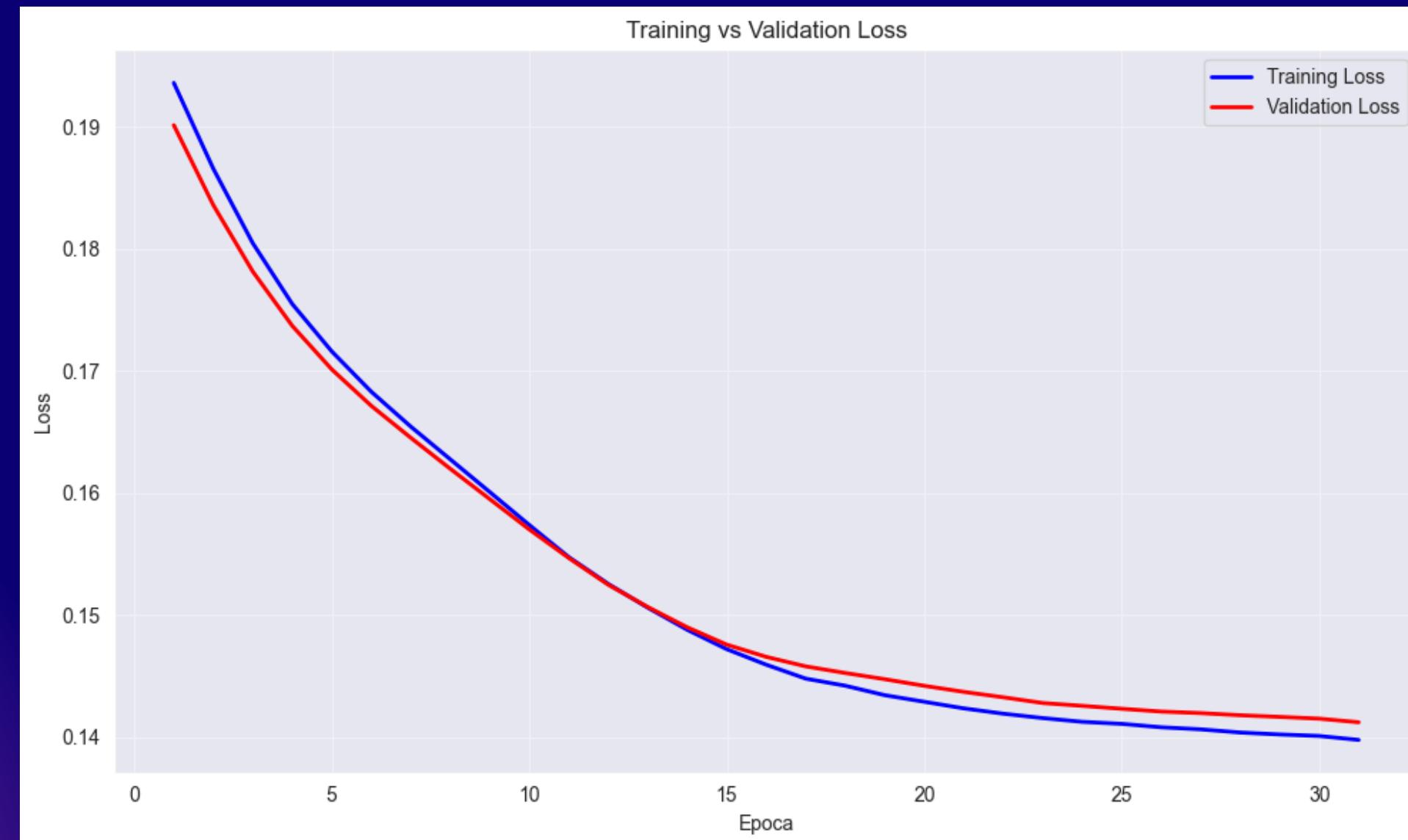
ESTRAZIONE DELLE INFORMAZIONI PER IL DATASET PRO-GAN



Il grafico sottostante mostra l'andamento di train e validation-loss per valutare quanto bene il modello riesce a performare sui dati.



Dal grafico si evince che non vi è presenza di overfitting, poichè entrambe le loss diminuiscono costantemente da 0.195 a 0.14, difatti le curve presentano un andamento parallelo e verso le ultime epocha, le loss si appiattiscono, indicando convergenza.



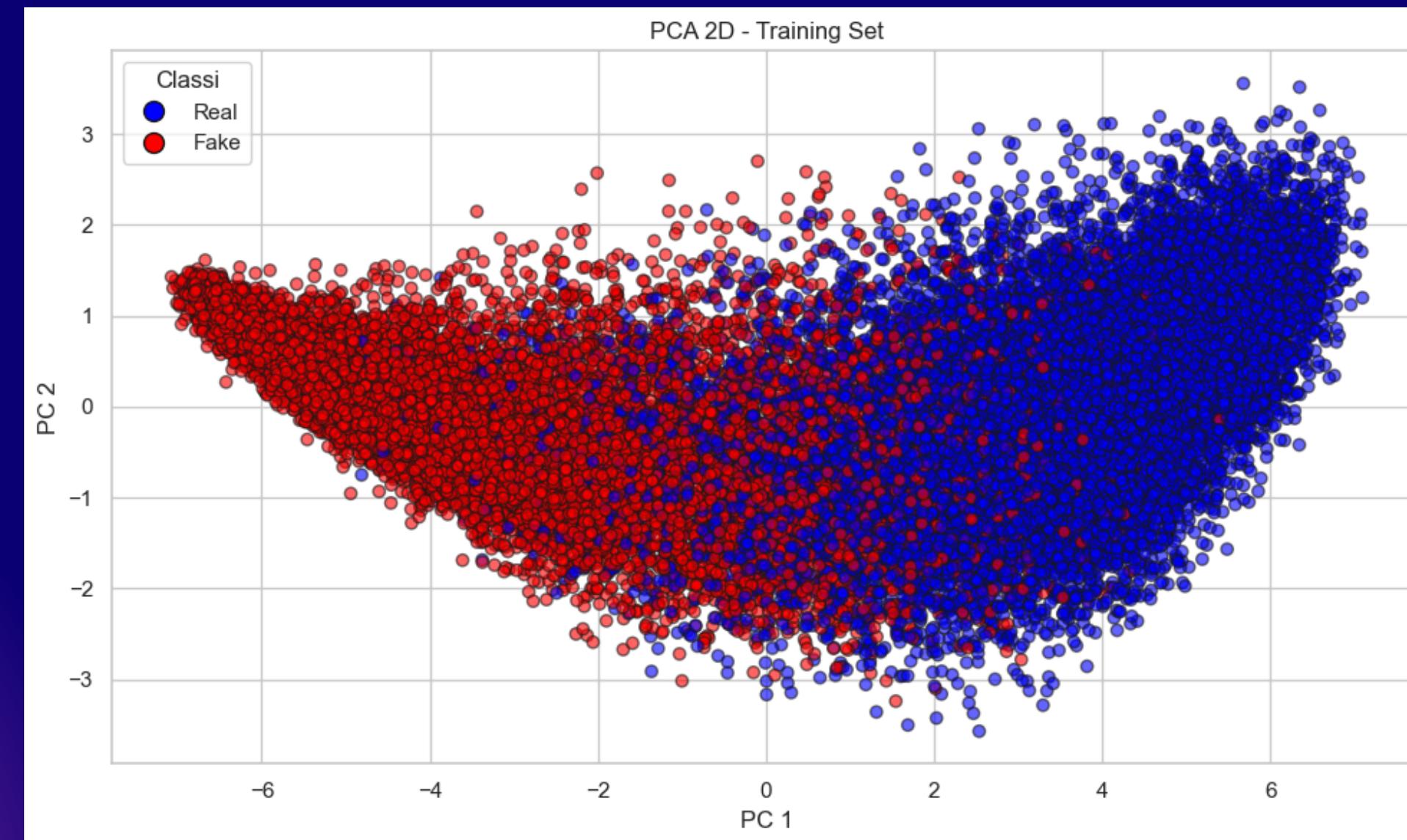
PRINCIPAL COMPONENT ANALYSIS (PCA) PER IL DATASET PRO-GAN



Il seguente grafico, invece mostra la distribuzione degli embeddings per le immagini reali e fake utilizzando la PCA.



La PCA è stata effettuata sui dati di training e suggerisce una buona separazione tra immagini reali ed immagini fake, anche se sono presenti alcuni elementi che continuano a mischiarsi tra di loro.



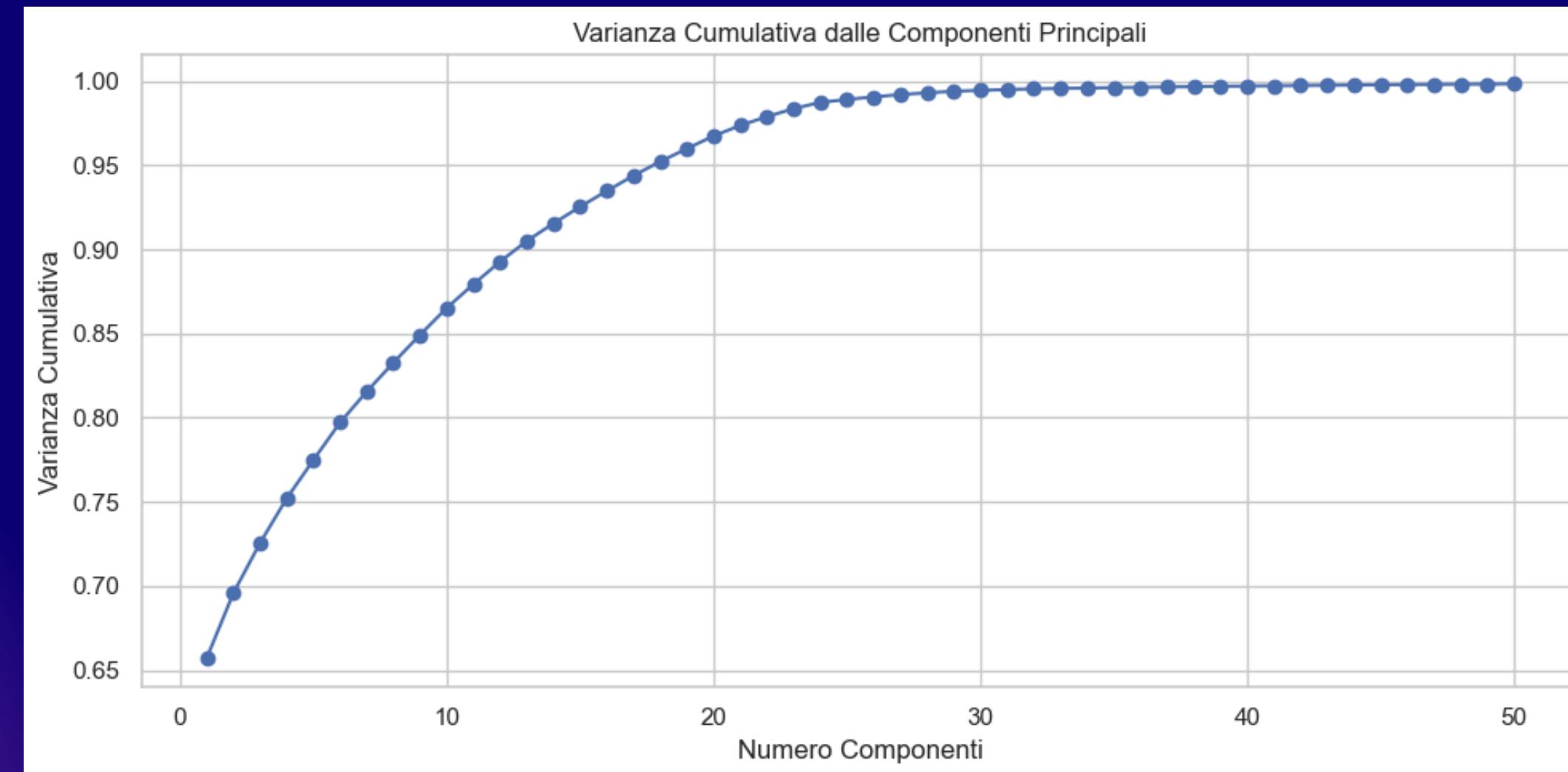
VARIANZA COMULATIVA PER IL DATASET PRO-GAN



Dalla PCA è stato effettuato il grafico della varianza cumulativa delle componenti principali.



La prima componente raggiunge il 65% della varianza cumulativa, come avveniva per il dataset Cycle-GAN ma, successivamente, raggiunge il 100% della varianza già intorno alla componente 25-30 .



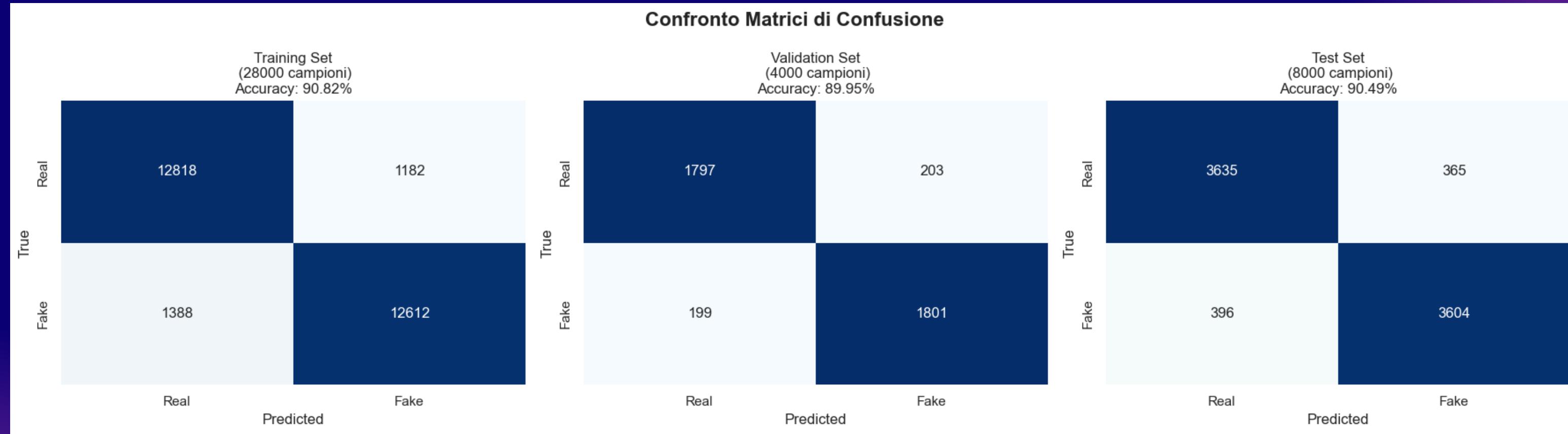
RISULTATI RAGGIUNTI PER IL DATASET PRO-GAN



Anche per questo dataset per l'addestramento è stato utilizzato SVM i cui dati di addestramento sono stati suddivisi in 70% per il training, 20% per il Test Set e il restante 10% per Validation Set.



L'analisi `e stata effettuata prendendo in considerazione 40000 campioni, dei quali 28000 sono stati selezionati per il Training Set, 4000 per la Validation Set ed infine 8000 per il Test Set.



Metriche	Percentuale Training	Percentuale Validation	Percentuale Test Set
Precision	90.83%	89.95%	90.49%
Recall	90.82%	89.95%	90.49%
F1-Score	90.82%	89.95%	90.49%

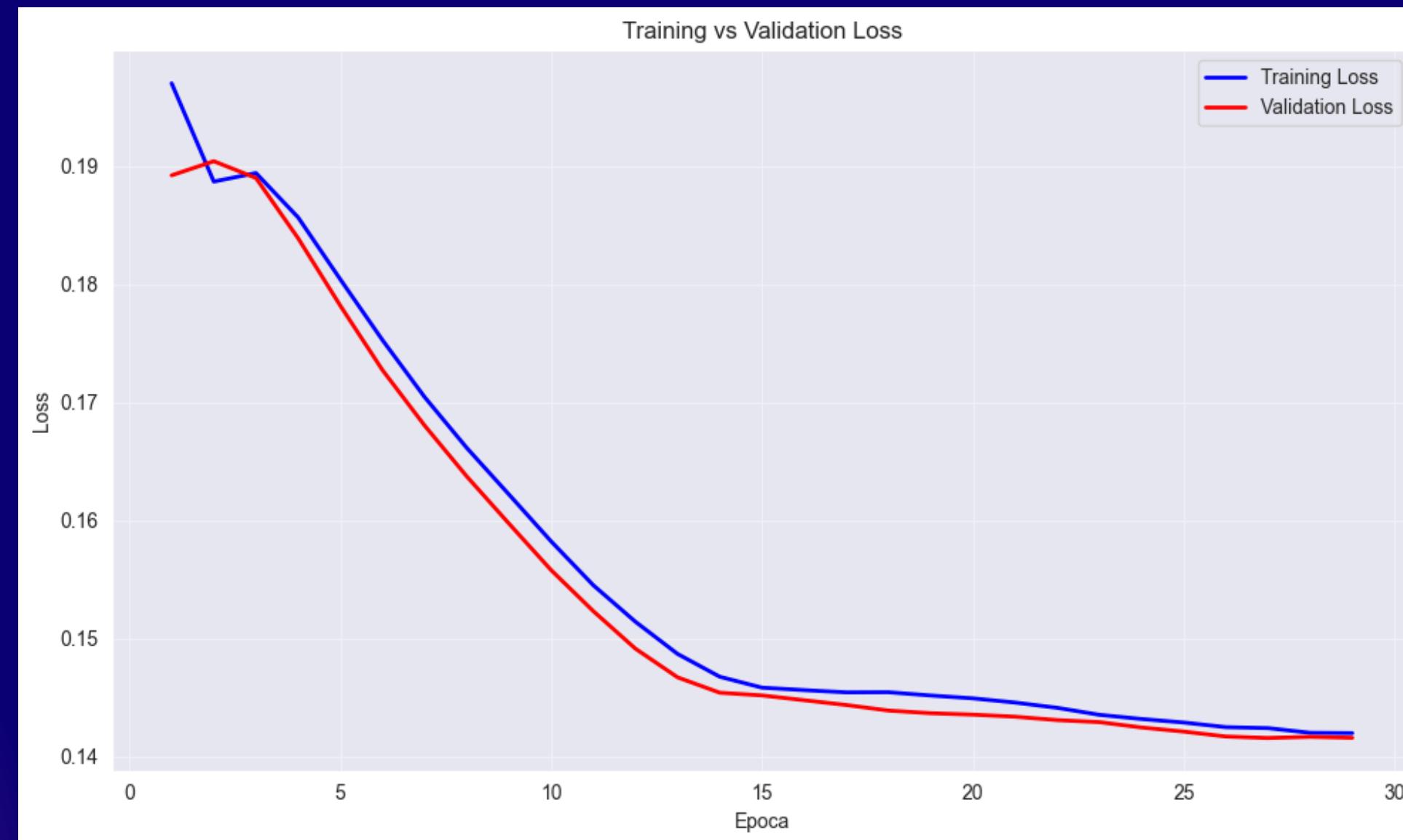
ESTRAZIONE DELLE INFORMAZIONI PER IL DATASET COMBINATO



Il grafico sottostante mostra l'andamento di train e validation-loss per valutare quanto bene il modello riesce a performare sui dati.



Il grafico mostra un andamento simile a Pro-GAN, con qualche differenza con le epoche iniziali dove le loss presentano una leggera oscillazione seguite da una rapida discesa per entrambe come nel grafico per Pro-GAN.



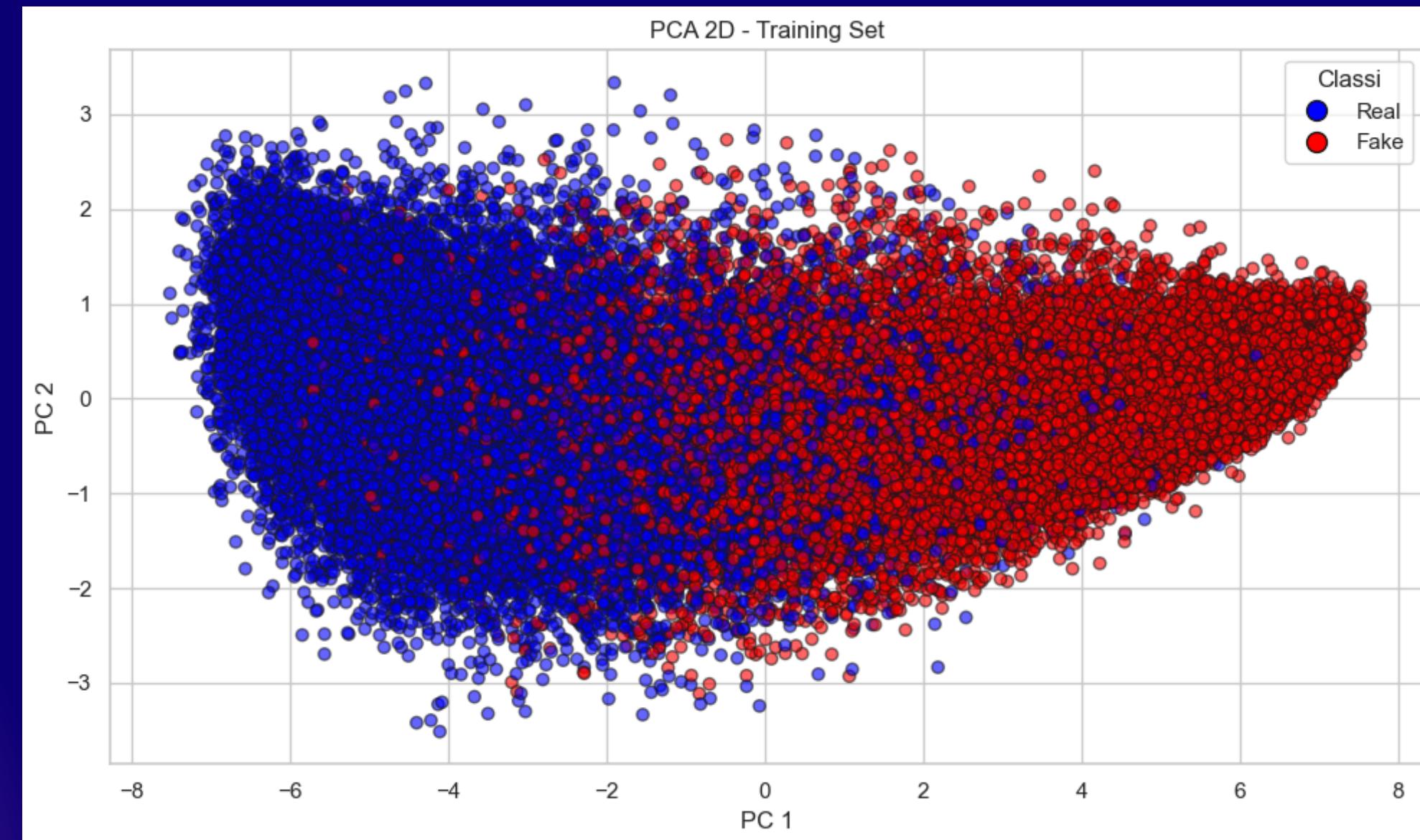
PRINCIPAL COMPONENT ANALYSIS (PCA) PER IL DATASET COMBINATO



Il seguente grafico, invece mostra la distribuzione degli embeddings per le immagini reali e fake utilizzando la PCA.



La PCA è stata effettuata sui dati di training e suggerisce una buona separazione tra immagini reali ed immagini fake, anche se sono presenti alcuni elementi che continuano a mischiarsi tra di loro.



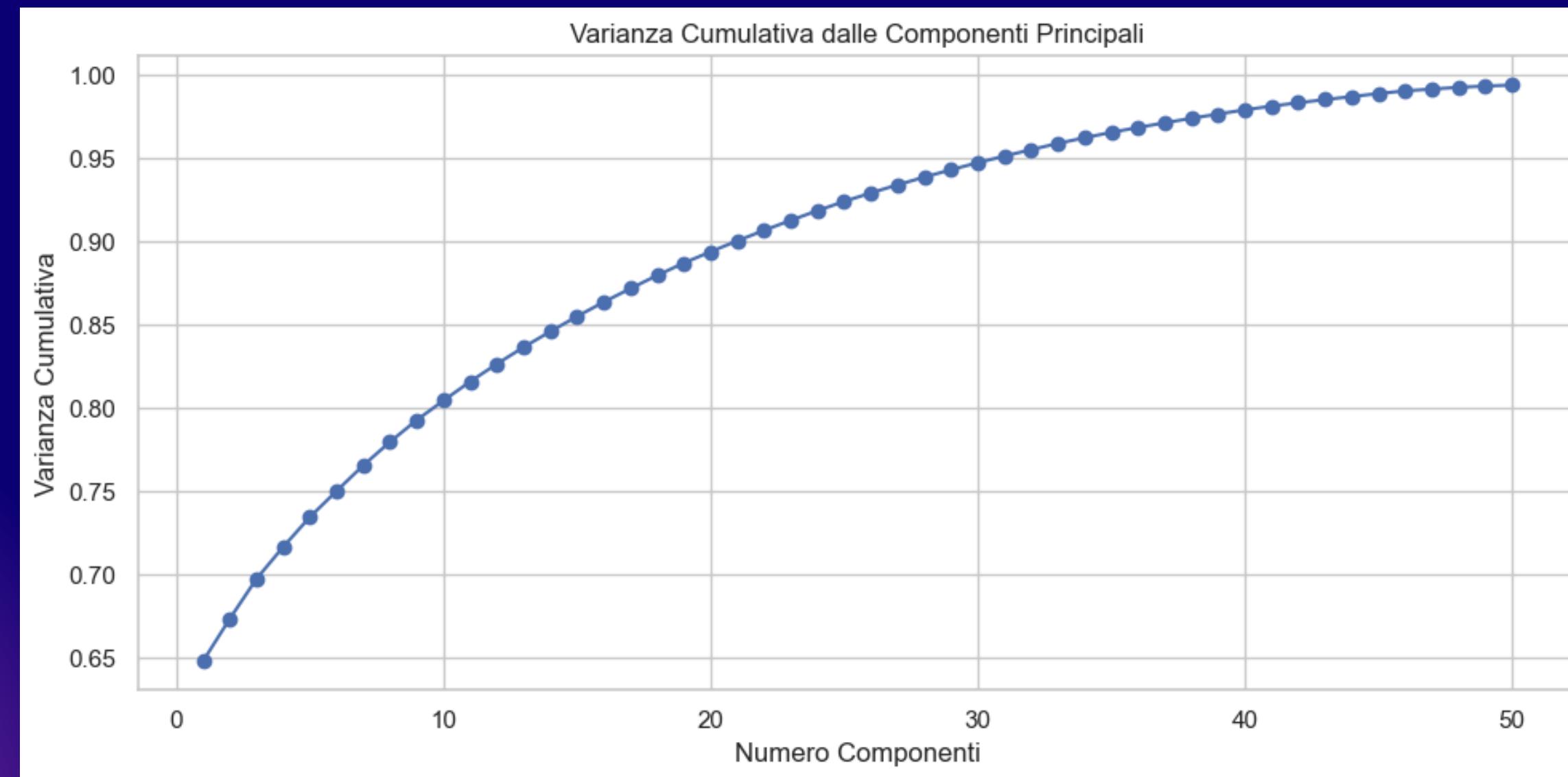
VARIANZA COMULATIVA PER IL DATASET COMBINATO



Dalla PCA è stato effettuato il grafico della varianza cumulativa delle componenti principali.



La prima componente spiega circa il 65% della varianza, come per Cycle-GAN, con un incremento graduale e uniforme lungo tutte le componenti fino a raggiungere quasi il 100% della varianza cumulativa verso la componente 50.



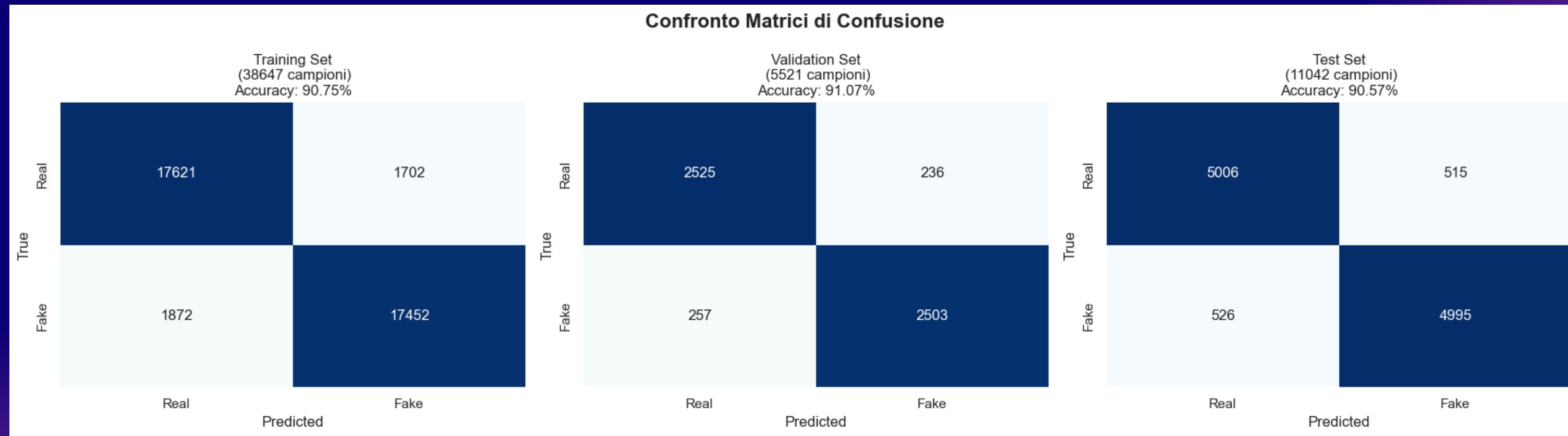
RISULTATI RAGGIUNTI PER IL DATASET COMBINATO



Anche per questo dataset per l'addestramento è stato utilizzato SVM i cui dati di addestramento sono stati suddivisi in 70% per il training, 20% per il Test Set e il restante 10% per Validation Set.



L'analisi è stata effettuata prendendo in considerazione 55210 campioni, dei quali 38647 sono stati selezionati per il Training Set, 5521 per la Validation Set ed infine 11042 per il Test Set.



Metriche	Percentuale Training	Percentuale Validation	Percentuale Test Set
Precision	90.76%	91.07%	90.57%
Recall	90.75%	91.07%	90.57%
F1-Score	90.75%	91.07%	90.57%

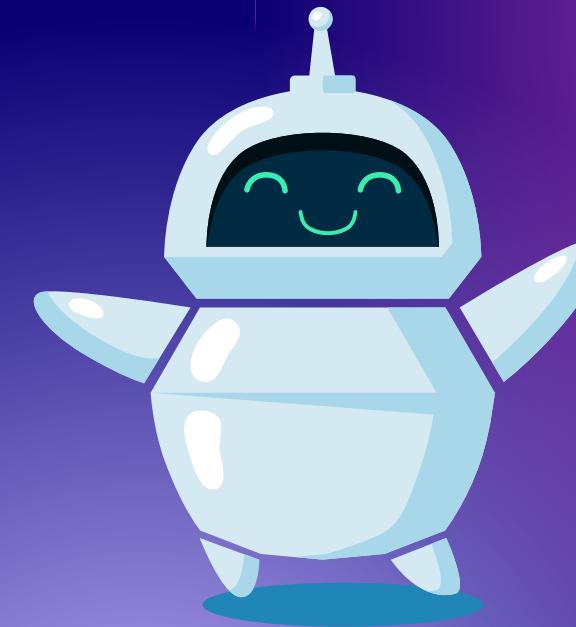
CONCLUSIONE E SVILUPPI FUTURI



Entrambi gli approcci utilizzati hanno portato al raggiungimento di buoni risultati, in quanto il modello riesce a generalizzare bene sui dati.



Sviluppi futuri: è possibile effettuare ulteriore ricerca al fine di migliorare i risultati raggiunti per farsi che il modello sia in grado di effettuare una separazione netta tra immagini reali ed immagini fake.



**GRAZIE PER
L'ATTENZIONE**

