

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

---

**SCUOLA DI SCIENZE**

*DIPARTIMENTO DI INFORMATICA - SCIENZA E INGEGNERIA (DISI)*

*CORSO DI LAUREA IN INFORMATICA*

**TESI DI LAUREA**

in

Introduzione all'apprendimento automatico

**PRECIPITAZIONI ESTREME: STUDIO SULLE  
PREVISIONI METEOROLOGICHE A BREVE TERMINE  
TRAMITE MODELLI DI DIFFUSIONE**

CANDIDATA:  
Francesca Mazzetti

RELATORE:  
Prof. Andrea Asperti

Anno accademico 2023/2024

Sessione Ottobre



# Sommario

L'aumento degli eventi metereologici estremi è una realtà sempre più evidente, confermata dalle osservazioni degli ultimi anni. La previsione di tali fenomeni pone sfide significative, richiedendo tecniche avanzate per un'analisi accurata, indispensabile per mitigarne l'impatto. I metodi di previsione metereologica, che includono sia approcci numerici sia basati sui dati, incontrano spesso difficoltà nel determinare con precisione la posizione e la tempistica degli eventi estremi. In particolare, i modelli fisici tradizionali risultano meno efficaci nelle previsioni a breve termine, dove la rapidità dei cambiamenti atmosferici richiede approcci più adattivi e reattivi.

Questa tesi si concentra sull'applicazione dell'intelligenza artificiale generativa, con particolare attenzione ai modelli di diffusione, per la previsione di precipitazioni intense. Sono state esplorate diverse strategie per migliorare le prestazioni di un modello reale attraverso esperimenti sui parametri e sulle previsioni; inoltre, sono state esaminate le funzioni di perdita e le metriche di valutazione applicate agli eventi estremi. Vengono discussi i limiti degli approcci attuali e proposte possibili direzioni future per il progresso nel campo della metereologia.

**Parole chiave:** previsioni metereologiche, previsioni a breve termine, precipitazioni estreme, modelli di diffusione, metriche di valutazione



# Indice

<b>Sommario</b>	<b>3</b>
<b>1 Previsione di eventi metereologici estremi</b>	<b>7</b>
1.1 Differenze tra meteo e clima . . . . .	7
1.2 Eventi estremi . . . . .	9
1.2.1 Fenomeno delle precipitazioni . . . . .	9
1.3 Tecniche di previsione metereologica . . . . .	10
1.3.1 Importanza delle previsioni nei settori critici . . . . .	10
1.3.2 Modelli di previsione . . . . .	11
<b>2 Metriche e funzioni di perdita per eventi estremi</b>	<b>13</b>
2.1 Metriche di valutazione . . . . .	13
2.1.1 Premessa . . . . .	14
2.1.2 Accuracy . . . . .	14
2.1.3 Precision e Recall . . . . .	15
2.1.4 $F_{\beta}$ -Score . . . . .	15
2.1.5 Average Precision . . . . .	16
2.1.6 Critical Success Index . . . . .	16
2.1.7 Equitable Threat Score . . . . .	17
2.1.8 Frequency BIAS . . . . .	17
2.1.9 True Skill Statistic . . . . .	18
2.1.10 Heidke Skill Score . . . . .	19
2.1.11 False Alarm Rate e Missed Alarm Rate . . . . .	19
2.1.12 Altre metriche . . . . .	20
2.2 Funzioni di perdita . . . . .	20
2.2.1 Funzioni di perdita pesate . . . . .	20
2.2.1.1 MSE e MAE bilanciate . . . . .	20
2.2.2 Coefficienti di Dice e Jaccard . . . . .	21
2.2.3 Funzione di perdita a penalità logaritmica . . . . .	22
2.2.4 Funzione di perdita composta . . . . .	22
<b>3 Apprendimento automatico nelle previsioni meteorologiche</b>	<b>23</b>
3.1 Introduzione al Machine Learning . . . . .	23
3.1.1 Deep Learning . . . . .	24
3.2 Reti neurali . . . . .	24
3.2.1 Reti Neurali Convoluzionali . . . . .	25
3.2.2 Architettura U-Net . . . . .	26

<b>4</b>	<b>Modelli di diffusione per la previsione delle precipitazioni</b>	<b>29</b>
4.1	Modelli di diffusione . . . . .	29
4.1.1	Applicazioni nella previsione di eventi estremi . . . . .	30
4.2	Introduzione a un modello generativo reale . . . . .	30
4.2.1	Architettura del modello . . . . .	31
4.2.2	Strumenti utilizzati . . . . .	31
4.2.3	Dataset . . . . .	32
4.2.3.1	Informazioni aggiuntive . . . . .	33
4.2.4	Preprocessamento dei dati . . . . .	34
4.2.5	Condizionamento dei dati . . . . .	34
4.2.5.1	Generatore . . . . .	35
4.2.6	Addestramento del modello . . . . .	35
4.2.6.1	Iperparametri . . . . .	35
4.2.6.2	Funzioni di perdita . . . . .	36
4.3	Altri approcci . . . . .	36
<b>5</b>	<b>Esperimenti</b>	<b>39</b>
5.1	Strumenti di analisi . . . . .	39
5.1.1	Funzioni di valutazione . . . . .	39
5.1.2	Funzioni di visualizzazione . . . . .	39
5.2	Lista esperimenti . . . . .	40
5.2.1	Esperimento 1 . . . . .	40
5.2.2	Esperimento 2 . . . . .	43
5.2.3	Esperimento 3 . . . . .	44
5.2.4	Esperimento 4 . . . . .	45
5.2.5	Esperimento 5 . . . . .	46
<b>6</b>	<b>Conclusioni</b>	<b>49</b>
6.1	Lavori futuri . . . . .	49
	<b>Appendices</b>	<b>51</b>
<b>A</b>	<b>Codici</b>	<b>53</b>
<b>B</b>	<b>Risultati</b>	<b>55</b>

# Capitolo 1

## Previsione di eventi metereologici estremi

I processi atmosferici non sono nè costanti nè rigorosamente periodici: se lo fossero, descriverli e fare previsioni meteo sarebbe un compito semplice. L'atmosfera è invece caratterizzata da variazioni e fluttuazioni irregolari, ed è proprio questa imprevedibilità a rendere il suo studio interessante. Per affrontare quantitativamente tale variabilità, è necessario ricorrere alla probabilità, il linguaggio matematico dell'incertezza. In teoria, modelli atmosferici perfetti potrebbero produrre previsioni precise, ma ciò non avviene per due ragioni principali. In primo luogo, i modelli, pur offrendo valide approssimazioni del comportamento atmosferico, non sono rappresentazioni complete della fisica che governa l'atmosfera. Molti processi fisici operano infatti su scale troppo piccole per essere descritti esplicitamente dai modelli; di conseguenza, i loro effetti su scale più ampie devono essere stimati utilizzando solo informazioni di larga scala. In secondo luogo, anche se fosse possibile includere ogni dettaglio fisico nei modelli, l'incertezza persisterebbe per via del fenomeno noto come caos dinamico, scoperto da Edward Lorenz nel 1963.

Il caos dinamico descrive come l'evoluzione di un sistema deterministico non lineare, come quello delle equazioni del moto atmosferico, sia estremamente sensibile alle condizioni iniziali. Questo significa che anche minime variazioni nelle condizioni di partenza possono condurre a risultati radicalmente diversi. Applicando questo concetto alla metereologia, poichè l'atmosfera viene sempre osservata in modo incompleto e impreciso, non è mai possibile avviare un modello matematico, anche se teoricamente perfetto, con lo stato esatto dell'atmosfera. Di conseguenza, non si potrà mai prevedere con precisione l'evoluzione futura della stessa. Sono necessarie strategie probabilistiche per minimizzare questa disparità [1].

Lorenz illustra questa sensibilità attraverso la sua celebre frase: "Il battito d'ali di una farfalla in Brasile può provocare un tornado in Texas", evidenziando come piccole variazioni nelle condizioni iniziali possano avere conseguenze drastiche e imprevedibili nel comportamento di un sistema complesso.

### 1.1 Differenze tra meteo e clima

Il meteo e il clima rappresentano due aspetti distinti del comportamento dell'atmosfera. La metereologia si concentra su fenomeni che si esauriscono in giorni o settimane.

ne, mentre il clima analizza l'andamento atmosferico su periodi di tempo misurati in decenni o secoli. La climatologia, quindi, si occupa dello studio del comportamento atmosferico su scala temporale ampia, analizzando come le condizioni medie si distribuiscono nel tempo e se si registrano variazioni rispetto a tali medie. I valori climatici oscillano tipicamente all'interno di un intervallo considerato normale, che funge da riferimento per il comportamento atteso del sistema atmosferico. Tuttavia, eventi estremi possono infrangere questa normalità.

È importante notare che un singolo evento, come una pioggia intensa o un'estate particolarmente calda, non è di per sé una prova di cambiamento climatico. Come osserva il ricercatore in Scienze Ambientali Luca Angeli: "Solo la frequente ripetizione di tali eventi nel tempo può indicare una modifica specifica nel sistema climatico". Ad esempio, se si verifica una temperatura elevata mai registrata negli ultimi 200 anni, ciò può indicare semplicemente un evento raro con tempi di ritorno lunghi. Tuttavia, se un evento del genere inizia a verificarsi più frequentemente, come ogni 20, poi 10 e infine 5 anni, allora si è davanti a un cambiamento climatico in atto, indicando una tendenza chiara. Questo fenomeno riflette esattamente ciò che stiamo osservando attualmente. Il cambiamento climatico si manifesta gradualmente e può essere difficile da percepire direttamente, ma le misurazioni scientifiche confermano un chiaro aumento delle temperature globali. La rapidità di questi cambiamenti è preoccupante, poiché supera i normali cicli naturali, causando effetti collaterali allarmanti.

Come sottolineato da Filippo Giorgi, direttore della sezione Scienze della Terra presso l'ICTP di Trieste, i principali effetti del cambiamento climatico includono:

- Scioglimento dei ghiacciai: i ghiacciai, principali riserve di acqua dolce, sono in fase di recessione su tutti i continenti, comportando una diminuzione della disponibilità di acqua e un innalzamento del livello del mare;
- Innalzamento del livello del mare: in scenari pessimistici, si prevede un aumento di 1 m entro la fine del secolo, con gravi conseguenze per le aree costiere;
- Aumento degli eventi estremi: l'atmosfera fatica a dissipare il calore in eccesso, conferendo maggiore energia ai fenomeni atmosferici, che si manifestano con maggiore violenza e frequenza. Eventi come alluvioni e siccità sono in aumento, con piogge sempre più intense e periodi di siccità che si allungano;
- Disomogeneità degli effetti: alcune regioni sono più vulnerabili al riscaldamento di altre. Ad esempio, l'Artico si riscalda a un ritmo doppio rispetto alla media globale, mentre il Mediterraneo registra temperature in aumento e precipitazioni sempre meno frequenti.

Studi recenti dimostrano che la frequenza degli eventi estremi è in aumento, soprattutto in Europa, considerato un hotspot per il cambiamento climatico, con una crescita evidente dal 1980<sup>1</sup>. Un'atmosfera più calda è anche più energetica, tendendo a contenere più vapore acqueo e aumentando così la probabilità di eventi estremi. Sebbene non si possa attribuire ogni singolo evento al cambiamento climatico, è evidente che la frequenza e l'intensità di questi eventi sono influenzate da questa dinamica.

---

<sup>1</sup><https://www.ipcc.ch/report/sixth-assessment-report-cycle/>



## 1.2 Eventi estremi

L'IPCC, principale organismo internazionale per la valutazione dei cambiamenti climatici, definisce un evento meteorologico estremo come un fenomeno raro per una specifica località o periodo dell'anno. In questo contesto, "estremo" può essere interpretato come "straordinario". Un evento viene classificato come estremo se supera, per intensità o frequenza, una soglia statistica stabilita in base a un campione di osservazioni raccolte su almeno 30 anni. In genere, questi eventi si collocano al di sotto del 10° percentile o al di sopra del 90° percentile di una distribuzione probabilistica basata su dati osservativi. Le variazioni locali e le differenze nei parametri climatici rendono complesso il confronto degli eventi estremi su scale spaziali e temporali diverse.

Le condizioni meteorologiche estreme possono manifestarsi in molteplici forme, come ondate di calore che pongono rischi per la salute e aumentano la probabilità di incendi boschivi; condizioni oceaniche calde che alimentano cicloni e mareggiate; siccità prolungate che influenzano l'agricoltura e la disponibilità d'acqua; e tempeste di neve o bufere che interrompono i trasporti e danneggiano le infrastrutture. In questa tesi, ci si concentra sulle precipitazioni intense, che possono verificarsi in brevi periodi o attraverso piogge che si protraggono per più giorni consecutivi.

### 1.2.1 Fenomeno delle precipitazioni

In meteorologia, la previsione delle precipitazioni coinvolge l'analisi di numerosi parametri e la comprensione delle loro interazioni per migliorare le tecniche di nowcasting. Le precipitazioni possono presentarsi come pioggia, neve, grandine, nebbia e pioggia gelata, e sono il risultato della condensazione del vapore acqueo in forma liquida o solida, che cade dalle nuvole o che si deposita sul suolo. La disponibilità di vapore acqueo è un elemento cruciale per la formazione di precipitazioni estreme, le quali dipendono dall'interazione di diversi fattori atmosferici:

- **Temperatura:** influisce sui processi di evaporazione e sublimazione. L'aria calda, essendo in grado di contenere più vapore acqueo, può aumentare l'intensità delle precipitazioni;
- **Umidità:** regola la velocità di condensazione del vapore acqueo, determinando la formazione di nuvole e la caduta delle precipitazioni;
- **Vento:** generato dalle differenze di pressione tra regioni diverse, trasporta il vapore acqueo verso l'alto, facilitando la formazione delle precipitazioni.

Prevedere con accuratezza le precipitazioni è un compito complesso, poichè i processi fisici che le governano sono altamente variabili. Basarsi esclusivamente su fattori locali non è sufficiente, in quanto queste rispondono anche a dinamiche globali [2]. Le misurazioni possono essere soggette a errori, poichè ostacoli come alberi ed edifici possono alterare i dati raccolti. Inoltre, la distribuzione delle precipitazioni è specifica per ciascun contesto spazio-temporale. Ad esempio, le zone costiere tendono a ricevere più precipitazioni rispetto alle aree interne, subendo anche variazioni stagionali e d'altitudine [3].

Quando gli elementi atmosferici precedentemente menzionati superano le soglie normali, possono innescare eventi meteorologici estremi con impatti significativi. Nonostante le variazioni stagionali, l'intensità delle precipitazioni estreme rimane relativamente costante durante tutto l'anno, suggerendo che gli ingredienti necessari per generare questi eventi possono manifestarsi in qualsiasi stagione, indipendentemente dai modelli di circolazione e dalle temperature medie.

In sintesi, c'è ancora una mancanza di chiarezza su come i processi atmosferici influenzino la frequenza e l'intensità degli eventi estremi, rendendo incerta la capacità di prevedere le precipitazioni.

### **1.3 Teniche di previsione meteorologica**

Le previsioni meteorologiche sono fondamentali in molte discipline; talvolta una previsione tempestiva e corretta può salvare vite umane. Sebbene le prime previsioni operative siano state avviate negli anni '50 negli Stati Uniti e in Europa occidentale, i tentativi formali di verificarne la qualità sono emersi solo negli anni '80. Negli ultimi anni, l'interesse per le previsioni di eventi estremi è aumentato, in parte a causa degli impatti finanziari e sociali che tali eventi hanno causato, oltre agli sforzi di ricerca legati al cambiamento climatico [4].

Le precipitazioni possono essere classificate in base alla loro durata, distinguendo tra precipitazioni a breve, medio e lungo termine. Tra queste, le precipitazioni a breve termine (0 – 6 ore) hanno un impatto diretto e significativo sui residenti, rendendo la loro previsione cruciale. La previsione accurata delle precipitazioni a breve termine, nota come precipitation nowcasting, è quindi di primaria importanza, nonostante rappresenti una sfida complessa e critica nel campo della meteorologia. Offrendo previsioni accurate, il nowcasting facilita la preparazione, il processo decisionale e la gestione delle emergenze, oltre a contribuire alle strategie di resilienza climatica [5]. Ciò permette a comunità e pianificatori delle infrastrutture di prepararsi e rispondere in modo più efficace agli eventi estremi, contribuendo agli sforzi per affrontare la crisi climatica [6, 7].

Prevedere condizioni meteorologiche estreme non è un compito semplice, poiché per loro natura questi eventi sono rari, il che implica la disponibilità limitata di osservazioni per la ricerca. Inoltre, le osservazioni storiche precedenti al 1980 potrebbero non essere più rilevanti, poiché il clima è cambiato notevolmente nel corso degli anni [8]. Esistono ulteriori limitazioni legate ai modelli di previsione, che complicano la possibilità di effettuare previsioni affidabili.

#### **1.3.1 Importanza delle previsioni nei settori critici**

Le società stanno affrontando un aumento sia nella frequenza che nell'intensità degli eventi meteorologici e climatici estremi. Secondo un'analisi del database NatCatService di Munich Re, il numero di disastri naturali legati al clima è passato da 1.383 tra il 1980 e il 1985 a 4.020 tra il 2013 e il 2018. Questi eventi estremi, uniti alla crescita esponenziale della popolazione e a cambiamenti demografici come l'urbanizzazione e l'aumento delle popolazioni costiere, stanno amplificando l'impatto sociale ed economico di eventi pericolosi [9]. Il contesto di tali eventi è ulteriormente aggravato da tenden-

ze globali non climatiche, tra cui la perdita di biodiversità, il consumo insostenibile di risorse naturali, il degrado degli ecosistemi e le pandemie <sup>2</sup>.

Settori particolarmente vulnerabili agli effetti di questi eventi estremi includono l'energia e l'agricoltura. Molti paesi stanno progressivamente abbandonando le centrali elettriche a combustibili fossili in favore di tecnologie energetiche più sostenibili. Tuttavia, questa transizione comporta nuove sfide, tra cui la stabilità delle reti elettriche. Mentre le centrali convenzionali garantiscono una produzione di energia stabile e flessibile, la produzione di energia rinnovabile dipende fortemente dalle condizioni meteorologiche. Pertanto, diventa fondamentale disporre di previsioni accurate [10, 11]. Inoltre, le interruzioni di corrente causate da eventi estremi hanno spesso un impatto significativo e duraturo, principalmente a causa dei gravi danni subiti dalle strutture di trasmissione e distribuzione, che, se obsolete, risultano particolarmente vulnerabili a queste condizioni [12]. In ambito agricolo, le informazioni meteorologiche e climatiche svolgono un ruolo cruciale nella gestione delle coltivazioni, influenzando decisioni chiave come la preparazione del terreno, la selezione delle varietà di colture, i modelli di coltivazione e i tempi di raccolta. Questo è particolarmente rilevante per i paesi la cui economia dipende fortemente dall'agricoltura [13].

#### 1.3.2 Modelli di previsione

Il campo della previsione metereologica si articola principalmente in due approcci: la previsione meteorologica numerica (Numerical Weather Prediction, o *NWP*) e i metodi di Deep Learning basati sull'analisi radar.

Il paradigma della previsione *NWP* ha dominato per decenni il campo delle previsioni delle precipitazioni. Questi modelli si fondano sui principi di fisica e fluidodinamica per simulare il comportamento dell'atmosfera, risolvendo complesse equazioni matematiche. Nel corso degli anni, i modelli *NWP* hanno visto significativi miglioramenti in termini di risoluzione spaziale e temporale, permettendo di rappresentare con maggiore dettaglio i fenomeni atmosferici. Inoltre, impiegano spesso un approccio ensemble, che consente di eseguire più simulazioni con condizioni iniziali diverse, migliorando così l'affidabilità di risultati. Questi modelli vengono utilizzati quotidianamente per fornire previsioni operative, dalle previsioni locali a quelle su larga scala, come la previsione di uragani o tempeste. Sistemi globali importanti come l'ECMWF (European Centre for Medium Range Weather Forecast) e il GFS (Global Forecast System) impiegano modelli numerici sofisticati e tecnologie avanzate per previsioni a medio e lungo termine. Tuttavia, i progressi nelle prestazioni dei modelli *NWP* sono limitati dalla complessità delle richieste computazionali e dalla gestione di problemi non lineari ad alta dimensione [3]. Inoltre, la rapidità e la complessità dei fenomeni meteorologici su piccola scala ostacolano l'efficacia dei metodi *NWP* nelle previsioni a breve termine, nonostante essi mostrino notevoli successi nelle previsioni a medio e lungo termine [14, 15].

Recentemente, numerosi approcci di Deep Learning hanno dimostrato di superare anche le tecniche tradizionali di estrapolazione dell'eco radar [15]. Questi metodi sfruttano ampi set di dati di osservazioni radar e possono essere addestrati senza vincoli di presupposti fisici predefiniti, migliorando significativamente l'accuratezza delle previsioni [8].

---

<sup>2</sup><https://www.ipcc.ch/report/ar6/wg2/>

A differenza dei modelli tradizionali, i modelli basati sul Deep Learning non si limitano a semplici processi di estrapolazione, ma generano previsioni sfruttando schemi complessi appresi dai dati. Considerano il nowcasting delle precipitazioni estreme come un problema di previsione video, in cui viene analizzata una sequenza temporale di fotogrammi: l'obiettivo è quello di prevedere i fotogrammi futuri basandosi sull'analisi delle osservazione dei fotogrammi passati, catturando l'evoluzione delle condizioni ambientali e meteorologiche nel tempo [7, 16]. Questo permette di modellare dinamiche complesse che i metodi tradizionali faticano a cogliere. Poichè non esistono modelli perfetti, è fondamentale continuare a sviluppare nuove metodologie per migliorare la capacità predittiva. A tal fine, per ottenere previsioni più localizzate su piccola scala, vengono spesso utilizzati modelli ad alta risoluzione che integrano condizioni iniziali e al contorno fornite dai modelli *NWP* [17].

Con l'aumento della disponibilità di dati meteorologici e di previsione, cresce inevitabilmente anche la richiesta di verifiche più approfondite. I meteorologi continueranno a interrogarsi su come migliorare i sistemi di previsione, mentre gli utenti si chiederanno quanto siano utili i prodotti previsti.

# Capitolo 2

## Metriche e funzioni di perdita per eventi estremi

L'addestramento e la valutazione dei modelli per la rilevazione di eventi estremi sono particolarmente complessi a causa della natura fortemente sbilanciata dei dati, che portano gli algoritmi di apprendimento a concentrarsi soprattutto sulle classi di maggioranza, trascurando quelle di minoranza, di maggiore interesse. Gli approcci adottati variano notevolmente, sia per i dataset impiegati, diversi per località, regimi di precipitazione e scale temporali, sia per gli strumenti utilizzati nello studio, su cui non si ha un consenso unanime. Nonostante queste differenze, dalla letteratura emergono elementi comuni. In questo capitolo, vengono presentate le principali metriche di valutazione e alcune funzioni di perdita utilizzate per la previsione di eventi estremi.

### 2.1 Metriche di valutazione

La verifica delle previsioni è il processo attraverso cui si valuta la qualità delle previsioni stesse [1]. Esistono diverse procedure per eseguire questa verifica e le opinioni su cosa costituisca una previsione di successo possono variare. Tuttavia, ogni metodo si basa sul confronto tra le previsioni e le osservazioni effettive. La verifica è quindi un problema statistico [4], poiché richiede l'uso di tecniche statistiche per analizzare e quantificare tali relazioni. Secondo [18], una previsione può essere considerata valida quando è accurata, coerente e utile, specialmente nel supportare il processo decisionale legato alla gestione dei rischi, come nel caso di eventi estremi. Il concetto di utilità nella previsione si può suddividere in tre principali categorie di scopo, ognuna delle quali rappresenta un diverso ambito decisionale:

- Scopi scientifici: per migliorare la comprensione e lo sviluppo dei modelli previsionali;
- Scopi amministrativi: per monitorare e valutare l'efficacia delle previsioni operative nel tempo;
- Scopi economici: per orientare le strategie e gli investimenti futuri in ambito previsionale.

Quindi, una previsione non è valutata solo in termini di precisione, ma anche in base alla sua capacità di essere applicata efficacemente nei vari contesti decisionali. Questo

studio si focalizza principalmente sul contesto scientifico. Verranno discusse diverse metriche <sup>1</sup>, evidenziando i rispettivi punti di forza e debolezza. Sebbene l'attenzione non sia posta sulle diverse architetture dei modelli previsionali, i risultati possono essere facilmente adattati a un'ampia gamma di modelli previsionali diversi.

### 2.1.1 Premessa

Le tradizionali metriche di valutazione per la previsione di eventi rari richiedono una classificazione binaria preliminare degli eventi come positivi (evento si verifica) o negativi (evento non si verifica). Allo stesso modo, le previsioni sono classificate come vere o false, generando quattro categorie principali: veri positivi (*TP*), veri negativi (*TN*), falsi positivi (*FP*) e falsi negativi (*FN*). Questi risultati sono solitamente organizzati in una matrice di confusione (*Confusion Matrix*), uno strumento fondamentale per valutare le prestazioni del modello. La combinazione dei valori presenti nelle sue quattro celle permette di derivare diverse misure di qualità della previsione. Tuttavia, anche con una chiara comprensione degli eventi da classificare, la definizione di queste categorie non è sempre immediata.

I modelli di previsione, infatti, producono spesso probabilità o valori numerici continui, e quindi è necessario stabilire delle soglie per convertire queste previsioni in classi binarie. La selezione delle soglie è un passaggio critico, specialmente in contesti caratterizzati da classi sbilanciate: soglie inadeguate possono portare a sottovalutare la classe rara o, al contrario, a sovrastimare la classe più comune, compromettendo l'accuratezza complessiva. Sebbene nessuna singola metrica possa riassumere tutte le informazioni sulle prestazioni di un modello, una combinazione ben calibrata di diverse misure può offrire una valutazione ragionevole degli errori commessi dal sistema, permettendo una diagnosi più precisa delle sue prestazioni complessive.

### 2.1.2 Accuracy

*Accuracy* (accuratezza, o *ACC*) è una metrica semplice e intuitiva che misura la percentuale di previsioni corrette rispetto al totale delle previsioni effettuate.

Si calcola come:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

rispondendo alla domanda: "Quale frazione delle previsioni è corretta?", con un punteggio da 0 a 1, dove 1 indica una previsione perfetta.

In casi di forte squilibrio tra le classi, la metrica può risultare fuorviante, poiché tende a essere fortemente influenzata dalla classe dominante, solitamente quella del nessun-evento (*TN*). Si può ottenere, quindi, un'accuratezza elevata anche se il modello non è realmente efficace nel riconoscere gli eventi rari, rappresentati dai *TP*. Per questo motivo, è preferibile valutare i modelli in termini di metriche alternative, come *Precision* (2.2), *Recall* (2.3) o una loro combinazione.

---

<sup>1</sup><https://resources.eumetrain.org>, <https://www.cawcr.gov.au/projects/verification/>

### 2.1.3 Precision e Recall

*Precision* (precisione) è definita dal rapporto tra le previsioni effettivamente positive e la quantità totale di positivi:

$$Precision = \frac{TP}{TP + FP} \quad (2.2)$$

rispondendo alla domanda: "Quante delle previsioni positive fatte dal modello si sono effettivamente rivelate corrette?", con un punteggio da 0 a 1, dove 1 indica una previsione perfetta.

*Recall* (richiamo), chiamato anche Probabilità di rilevamento (*POD*) o Sensibilità, indica la relazione tra le previsioni che si sono rivelate positive e il numero totale di eventi positivi:

$$Recall = \frac{TP}{TP + FN} \quad (2.3)$$

rispondendo alla domanda: "Quale frazione degli eventi 'sì' osservati è stata correttamente prevista?", con un punteggio da 0 a 1, dove 1 indica una previsione perfetta.

La precisione rappresenta una misura di esattezza, mentre il richiamo è una misura di completezza. La scelta di dare priorità all'una o all'altra dipende dalle esigenze specifiche e dai rischi associati al problema. Dare priorità al richiamo è preferibile quando è cruciale ridurre i *FN*, come nei casi in cui non individuare un evento può avere conseguenze gravi. Al contrario, la precisione è fondamentale quando ridurre i *FP* è prioritario, ovvero quando ogni previsione positiva deve essere estremamente affidabile. Variare la soglia di classificazione influisce su entrambe le metriche: alzare la soglia riduce il rischio di *FP*, migliorando la precisione, ma aumenta i *FN*. Abbassare la soglia ha l'effetto opposto. Di conseguenza, precisione e richiamo tendono a essere inversamente proporzionali e spesso è necessario trovare un punto di equilibrio tra le due.

### 2.1.4 $F_\beta$ -Score

A causa delle caratteristiche sopra menzionate, precisione e richiamo non sempre sono utili se considerate separatamente. Una possibile soluzione è analizzare la curva *Precision-Recall* che mostra le prestazioni del modello a diverse soglie di classificazione. In alternativa, i valori possono essere combinati in una singola metrica, chiamata *F $\beta$ -Score* (punteggio  $F_\beta$ ). Questa metrica rappresenta la media armonica ponderata di precisione e richiamo, in cui il parametro  $\beta$  regola l'importanza relativa tra le due:

- $\beta > 1$ : dà maggior peso alla precisione, penalizzando di più i *FP* rispetto ai *FN*;
- $\beta < 1$ : favorisce il richiamo, riducendo l'impatto dei *FN*.

Il punteggio  $F_\beta$  varia da 0 a 1, dove 1 indica una previsione perfetta.

La versione più comune del punteggio è *F $_1$ -Score*, che rappresenta un equilibrio equo tra precisione e richiamo:

$$F1-Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN} \quad (2.4)$$

Valore alto di  $F_1$  suggerisce migliori prestazioni del modello, risultando particolarmente utile in contesti di classificazione sbilanciata. Un punteggio  $F_1$  basso, invece, non è del tutto informativo: segnala problemi, ma non specifica se dovuti a bassa precisione o a basso richiamo. Per questo motivo, è necessario analizzare le metriche separatamente per identificare se il problema deriva dai  $FP$  (*Precision*) o dai  $FN$  (*Recall*).

Altri punteggi  $F_\beta$ :

- $F_2$ -Score: dà più peso al richiamo rispetto alla precisione ( $\beta = 2$  rende il richiamo due volte più importante della precisione):

$$F2\text{-Score} = 5 \cdot \frac{Precision \cdot Recall}{4 \cdot Precision + Recall} \quad (2.5)$$

- $F_{0.5}$ -Score: favorisce la precisione rispetto al richiamo ( $\beta = 0.5$  rende la precisione due volte più importante del richiamo):

$$F0.5\text{-Score} = \frac{5}{4} \cdot \frac{Precision \cdot Recall}{0.5^2 \cdot Precision + Recall} \quad (2.6)$$

### 2.1.5 Average Precision

L'*Average Precision* (precisione media, o *AP*) è la metrica di scelta in attività come il recupero di informazioni e rilevamento oggetti, in cui il compromesso tra precisione e richiamo è cruciale. L'*AP* riassume la curva *Precision-Recall* come la media ponderata delle precisioni ottenute a ciascuna soglia, fornendo una misura complessiva delle performance del modello.

Si calcola come:

$$AP = \sum_n (R_n - R_{n-1}) \cdot P_n \quad (2.7)$$

dove  $P_n$  e  $R_n$  sono rispettivamente la precisione e il richiamo alla soglia  $n$ -esima. Il suo punteggio varia da 0 a 1, dove 1 indica una previsione perfetta.

Il calcolo effettivo dell'*Average Precision* non dipende dalle soglie specifiche, ma si basa semplicemente sull'ordinamento delle previsioni in base ai loro valori di confidenza o altre misure non soglia, come l'intensità della pioggia in questo caso. Un vantaggio dell'*AP* è che è meno influenzato dalle tecniche di ensemble rispetto ad altri metodi.

### 2.1.6 Critical Success Index

Il *Critical Success Index* (indice di successo critico, o *CSI*), chiamato anche *Threat Score* (*TS*), è una metrica consolidata in meteorologia, calcolata come:

$$CSI = \frac{TP}{TP + FP + FN} \quad (2.8)$$

rispondendo alla domanda: "Quanto bene gli eventi si previsti corrispondono agli eventi si osservati?", con un punteggio da 0 a 1, dove 1 indica una previsione perfetta.

Il *CSI* penalizza sia i fallimenti ( $FN$ ) che i falsi allarmi ( $FP$ ), senza distinguere però la fonte dell'errore di previsione. Non considerare i non-eventi nel calcolo della metrica può essere visto come un vantaggio pratico, specialmente nelle previsioni di eventi



rari, risolvendo così il problema riguardante l'*Accuracy* (2.1). Tuttavia, proprio questa esclusione dei non-eventi rende il *CSI* dipendente dalla frequenza dell'evento, dimostrando che non è un indicatore imparziale della capacità di previsione. Per chiarire questo punto, consideriamo una riformulazione del *CSI*:

$$CSI = \frac{TP}{T - TN} \quad (2.9)$$

dove  $T$  è composto da tutti i possibili esiti

Quando un evento è comune, il numero di  $TN$  diminuisce, portando a un *CSI* più alto. Ma questo non significa che il modello stia prevedendo meglio; significa solo che c'è meno non-evento da considerare, il che gonfia artificialmente il valore dell'indice, così che anche una previsione imprecisa ha maggiori probabilità di essere corretta.

Ad esempio, in un clima umido, poiché la pioggia è già un evento frequente, un modello potrebbe ottenere un *CSI* più alto semplicemente prevedendo spesso pioggia, anche senza grande precisione. Al contrario, in un clima secco, dove la pioggia è rara, è più difficile ottenere previsioni accurate. Questo è ciò che si intende con il ruolo della casualità: è più facile fare previsioni corrette semplicemente per caso, senza che ciò rifletta necessariamente un'effettiva abilità del modello di previsione. Comunque, la metrica rimane un valido indicatore del valore relativo di diverse tecniche di previsione quando vengono applicate allo stesso ambiente [19].

### 2.1.7 Equitable Threat Score

Per evitare che i successi casuali influiscano troppo sulle misurazioni del *CSI* (2.8), viene utilizzato l'*Equitable Threat Score* (equo  $TS$ , o *ETS*). La metrica misura la frazione di eventi osservati e/o previsti correttamente, aggiustati per i successi associati al caso:

$$ETS = \frac{TP - E(TP)}{TP + FP + FN - E(TP)} \quad (2.10)$$

dove  $E(TP)$  è il numero di veri positivi attesi per caso, calcolato come:

$$E(TP) = \frac{(TP + FP) \cdot (TP + FN)}{TP + FP + TN + FN} \quad (2.11)$$

variando da  $-1/3$  a  $1$ , dove  $1$  indica una previsione perfetta.

In questo contesto, "equità" significa che il punteggio dovrebbe essere indipendente dalla frequenza dell'evento; ovvero, dovrebbe riflettere l'abilità del modello in modo uniforme, sia che l'evento sia raro o frequente. Per le sue proprietà, l'*ETS* è spesso utilizzato nei modelli numerici meteorologici (*NWP* 1.3.2) per la verifica delle previsioni di precipitazione e il confronto tra punteggi per diversi regimi climatici. Tuttavia, uno studio in particolare [20] mette in guardia sul fatto che l'*ETS* non sia completamente indipendente dalla prevalenza dell'evento.

### 2.1.8 Frequency BIAS

*Frequency Bias Index* (indice di distorsione della frequenza, o *FBI*) misura la tendenza del sistema di previsione a sovrastimare o sottostimare la frequenza degli even-

ti. È calcolato come il rapporto tra il numero di eventi previsti e il numero di eventi effettivamente osservati:

$$FBI = \frac{TP + FP}{TP + FN} \quad (2.12)$$

In altre parole, l'*FBI* indica se il modello tende a prevedere troppo spesso o troppo raramente un evento, ma non fornisce informazioni sulla qualità delle previsioni, come accuratezza o precisione. L'indice può assumere valori a partire da 0, tendendo verso  $\infty$ , dove:

- *FBI* < 1: indica che il modello sottostima la frequenza degli eventi, adottando un approccio più conservativo, con un alto numero di *FN*. Questo può essere rischioso, soprattutto nella previsione di eventi estremi;
- *FBI* = 1: indica che il modello prevede l'evento esattamente con la stessa frequenza con cui si verifica, indicando assenza di distorsione;
- *FBI* > 1: indica che il modello sovrastima la frequenza degli eventi, mostrando un comportamento più aggressivo e generando un numero elevato di *FP*.

### 2.1.9 True Skill Statistic

*True Skill Statistic* (*TSS*), noto anche come Punteggio di abilità di Peirce o Discriminante di Hanssen e Kuipers, è una metrica usata frequentemente in meteorologia e in altri campi che riguardano la previsione di eventi rari:

$$TSS = Sensitivity + Specificity - 1 \quad (2.13)$$

dove *Sensitivity* (sensibilità) è uguale a *Recall* (2.3). *Specificity* (specificità) misura la proporzione di effettivi negativi identificati correttamente, calcolata come:

$$Specificity = \frac{TN}{TN + FP} \quad (2.14)$$

Effettuando alcune manipolazioni, il *TSS* può essere espresso nel seguente modo:

$$TSS = \frac{TP}{TP + FN} - \frac{FP}{TN + FP} = \frac{TP \cdot TN - FP \cdot FN}{(TP + FN) \cdot (TN + FP)} \quad (2.15)$$

rispondendo alla domanda: "Quanto bene il modello distingue tra gli eventi positivi e negativi?", con un punteggio da -1 a 1, dove 1 indica una previsione perfetta.

La metrica tiene conto sia della capacità di un modello di prevedere correttamente un evento, sia di identificare correttamente un non-evento, oltre ad essere indipendente dalla prevalenza delle classi. Queste caratteristiche rendono il *TSS* una metrica equilibrata, particolarmente adatta quando si devono valutare modelli con classi sbilanciate.

### 2.1.10 Heidke Skill Score

L'*Heidke Skill Score* (*HSS*) misura il miglioramento della previsione rispetto a una previsione casuale. La formula è la seguente:

$$HSS = 2 \cdot \frac{TP \cdot TN - FP \cdot FN}{(TP + FN) \cdot (FN + TN) + (TP + FP) \cdot (FP + TN)} \quad (2.16)$$

rispondendo alla domanda: "*Qual è stata l'accuratezza della previsione rispetto a quella del caso?*", con un punteggio che varia da valori tendenti a  $-\infty$  fino a un massimo di 1, dove 1 indica una previsione perfetta.

L'*HSS* è una metrica ampiamente utilizzata, in parte per la sua semplicità di calcolo e perché consente di superare facilmente le previsioni di base, come quelle basate sulla probabilità. Altri punteggi di riferimento, come la persistenza o la climatologia, sono spesso preferibili nel contesto meteorologico.

### 2.1.11 False Alarm Rate e Missed Alarm Rate

Il *False Alarm Rate*, tasso di falsi allarmi o anche tasso di falsi positivi (*FAR*), misura il rapporto tra il numero di falsi positivi e il numero totale di eventi negativi:

$$FAR = \frac{FP}{TN + FP} \quad (2.17)$$

rispondendo alla domanda: "*Quale frazione degli eventi 'sì' previsti in realtà non si è verificata?*", con un punteggio da 0 a 1, dove 0 indica una previsione perfetta.

Si può ottenere un miglioramento apparente del *FAR* aumentando il numero di previsioni negative ( $TN + FP$ ) e alzando la soglia di classificazione, riducendo così il numero percentuale di falsi allarmi ( $FP$ ). Tuttavia, questo non indica necessariamente un miglioramento reale della qualità del modello, poiché aumenta il rischio di mancati allarmi ( $FN$ ). Inoltre, il *FAR* non considera i  $FN$ , quindi un valore basso potrebbe mascherare l'incapacità del modello di rilevare eventi effettivamente verificatisi.

Per ottenere una visione più completa delle prestazioni di un modello, il *FAR* dovrebbe essere interpretato insieme alla metrica *POD* (2.8), bilanciando la capacità di ridurre i  $FP$  con l'efficacia nel rilevare correttamente gli eventi reali.

Il *Missed Alarm Rate*, tasso di allarmi persi o tasso di falsi negativi (*MAR*), misura la proporzione di falsi negativi sul numero totale di eventi positivi:

$$MAR = \frac{FN}{TP + FN} \quad (2.18)$$

rispondendo alla domanda: "*Quante volte il modello ha mancato di prevedere un evento che si è effettivamente verificato?*", con un punteggio da 0 a 1, dove 0 indica una previsione perfetta.

La metrica è utile per valutare la capacità del modello di rilevare gli eventi reali, ed è cruciale in contesti dove è essenziale non perdere eventi. Complementarmente al *FAR* (2.17), il *MAR* considera solo i  $FN$  e ignora i  $FP$ . Per ottenere una valutazione completa delle prestazioni del modello, è consigliabile usarlo in combinazione con altre metriche, come il *POD* (2.8), talvolta espresso come  $1 - MAR$ .

### 2.1.12 Altre metriche

In questa sezione sono presentate ulteriori metriche di valutazione indicate in [3, 6, 8, 15, 16, 21, 22], che non derivano direttamente dai valori della matrice di confusione. Queste metriche, pur essendo legate a processi di correlazione e accuratezza dei dati, si focalizzano principalmente sulla qualità percettiva e spaziale delle previsioni. A differenza delle metriche tradizionali che analizzano le performance del modello in termini di eventi binari o discreti, si concentrano su come le previsioni si allineano con la realtà osservata in un contesto più ampio:

- *Structural Similarity Index (SSIM)*: misura la somiglianza tra due immagini, valutando i cambiamenti strutturali percepiti, come il degrado della qualità dovuto a operazioni di compressione o rumore;
- *Pearson Correlation Coefficient (PCC)*: valuta la correlazione lineare tra due variabili quantitative. È una statistica descrittiva, riassumendo le caratteristiche di un set di dati, e inferenziale, testando ipotesi statistiche;
- *Fractions Skill Score (FSS)*: metodo di verifica dell'abilità delle previsioni deterministiche di eventi binari. Si concentra principalmente su quanto bene la distribuzione spaziale delle previsioni si sovrappone a quella delle osservazioni, valutando la percentuale di coincidenza tra le due.

Le metriche elencate non considerano il problema dello sbilanciamento delle classi, il che le rende meno adatte per la previsione di eventi rari o estremi. In tali casi, le metriche basate sulla *Confusion Matrix* risultano spesso più appropriate per una valutazione accurata delle previsioni del modello.

## 2.2 Funzioni di perdita

### 2.2.1 Funzioni di perdita pesate

Per enfatizzare l'importanza degli eventi estremi, è possibile attribuire un peso maggiore agli eventi in base all'intensità delle precipitazioni in una specifica località. Questo si realizza solitamente attraverso l'uso di un fattore moltiplicativo, che viene applicato a misure di errore come *MAE*, *MSE* o a una loro combinazione. I fattori di peso per una precipitazione osservata  $y$  sono solitamente limitati a una soglia specifica (threshold), a seconda dell'area e della scala spazio-temporale:

$$w(y) = \min[\max(1, y), th] \quad (2.19)$$

L'assegnazione della soglia evita che i valori estremi sovrastino completamente la valutazione complessiva. Nel lavoro di [23], che propone un modello basato su reti generative avversarie (4.3), la funzione di perdita include anche l'indice *SSIM* (2.1.12) insieme a operazioni di ponderazione.

#### 2.2.1.1 MSE e MAE bilanciate

*BMSE* e *BMAE*, introdotte in [7], sono versioni bilanciate delle metriche *MSE* e *MAE*. Tengono conto delle caratteristiche di distribuzione dei dati sulle precipitazioni, il che è particolarmente utile per prevedere eventi estremi:

$$BMSE = \frac{1}{N} \cdot \sum_{n=1}^N \sum_{i=1}^{128} \sum_{j=1}^{128} w_{n,i,j} (x_{n,i,j} - \hat{x}_{n,i,j})^2 \quad (2.20)$$

$$BMAE = \frac{1}{N} \cdot \sum_{n=1}^N \sum_{i=1}^{128} \sum_{j=1}^{128} w_{n,i,j} |x_{n,i,j} - \hat{x}_{n,i,j}| \quad (2.21)$$

$\hat{x}_{n,i,j}$  e  $x_{n,i,j}$  sono rispettivamente il valore di output derivato dal modello e il corrispondente valore di osservazione. Le somme sono calcolate su tutti gli step temporali  $n$  e le griglie spaziali  $(i,j)$ . I fattori di peso  $w_{n,i,j}$  variano a seconda dell'intensità della precipitazione:

$$w(x) = \begin{cases} 1 & \text{se } 0 \leq x < 1 \\ 2 & \text{se } 1 \leq x < 5 \\ 5 & \text{se } 5 \leq x < 10 \\ 10 & \text{se } 10 \leq x < 20 \\ 20 & \text{se } 20 \leq x < 50 \\ 50 & \text{se } x \geq 50 \end{cases}$$

Combinando queste metriche di errore bilanciate è possibile migliorare ulteriormente le prestazioni del modello. La funzione di perdita *CB Loss* derivata è calcolata come segue:

$$CB\ Loss = \frac{BMSE + \alpha BMAE}{2} \quad (2.22)$$

dove  $\alpha$  è un coefficiente che regola *BMSE* e *BMAE*. In base agli esperimenti condotti,  $\alpha$  è fissato al valore di 0.1.

## 2.2.2 Coefficienti di Dice e Jaccard

Il coefficiente di Dice e l'indice di Jaccard misurano la similarità tra le classi predette e quelle reali. Queste metriche, usate principalmente per scopi di segmentazione, si rivelano utili anche in contesti di classi sbilanciate, come per le previsioni di eventi estremi, dove si può analizzare la concordanza tra:

- Classi predette: eventi previsti dal modello  $PP = TP + FP$
- Classi reali: eventi positivi reali  $P = TP + FN$

Quindi, un modello viene valutato in base alla similarità tra  $P$  e  $PP$ . In generale, la distanza di Dice tra due insiemi  $A$  e  $B$  è data da:

$$Dice(A, B) = \frac{2 \cdot |A \cap B|}{|A| + |B|} \quad (2.23)$$

Nel caso di  $P$  e  $PP$ , si ottiene:

$$Dice(P, PP) = \frac{2 \cdot |P \cap PP|}{|P| + |PP|} = \frac{2TP}{2TP + FP + FN} \quad (2.24)$$

dove l'ultimo termine è equivalente al punteggio  $F_1$  (2.4).

L'indice di Jaccard, anche noto come *Intersection Over Union (IoU)*, differisce dal coefficiente di Dice al massimo di un fattore 2:

$$IoU(P, PP) = \frac{|P \cap PP|}{|P \cup PP|} = \frac{TP}{TP + FP + FN} \quad (2.25)$$

### 2.2.3 Funzione di perdita a penalità logaritmica

Un altro approccio consiste nell'utilizzare una funzione di perdita ponderata che incorpora penalità logaritmiche per previsioni errate, assegnando pesi diversi ai  $FP$  e  $FN$  [8]:

$$EVL = -0.05(1 - \hat{x}) \cdot x \log(\hat{x}) - 0.95\hat{x}(1 - x) \cdot \log(1 - \hat{x}) \quad (2.26)$$

dove  $\hat{x}$  indica le previsioni e  $x$  rappresenta le etichette di verità di base. I coefficienti  $\beta_0 = 0,95$  e  $\beta_1 = 0,05$  regolano la ponderazione dei  $FP$  e dei  $FN$ .

Questa formulazione penalizza efficacemente il modello per eventi estremi in cui le previsioni si discostano in modo significativo dalla verità di base. Ad esempio, quando  $\hat{x}$  si avvicina a 0 e  $x$  si avvicina a 1, la penalità evidenzia errori critici nella previsione di eventi rari con elevata confidenza. Al contrario, anche la previsione di  $\hat{x}$  vicino a 1 quando  $x$  è vicino a 0 comporta una penalità severa per non aver riconosciuto l'assenza di un evento.

### 2.2.4 Funzione di perdita composta

La funzione di perdita  $L1 + L2$  [15, 21] è fondamentale per prevedere precipitazioni estreme. Combina in modo complementare l'errore assoluto  $L1$  ( $MAE$ ) per gestire i valori anomali nei dati e l'errore quadratico  $L2$  ( $MSE$ ) per ridurre deviazioni significative:

$$Composite Loss = \frac{1}{n} \cdot \sum_{t=1}^n (|x_t - \hat{x}_t| + |x_t - \hat{x}_t|^2) \quad (2.27)$$

dove  $n$  rappresenta il numero totale di campioni nel set di dati.

In [21], oltre alla funzione di perdita, lo schema di campionamento inverso aiuta ad accelerare la convergenza del modello, riducendo il numero di iterazioni richieste durante l'addestramento.

## Capitolo 3

# Apprendimento automatico nelle previsioni meteorologiche

All'inizio del 21° secolo, fattori chiave come l'aumento della potenza di calcolo, l'emergere di nuove tecnologie, l'accesso a software specializzati e il miglioramento delle capacità di rianalisi hanno favorito l'uso dell'apprendimento automatico in numerosi settori scientifici. In particolare, il Machine Learning presenta applicazioni di grande significato in meteorologia e climatologia, dove molte questioni relative alla circolazione atmosferica mancano di criteri quantitativi ben definiti [11]. Questa incertezza complica le analisi oggettive. Nei paragrafi successivi, verrà mostrato in dettaglio come il Machine Learning e il Deep Learning possano essere applicati per affrontare le sfide delle previsioni atmosferiche, mettendo in evidenza il loro potenziale nell'elaborazione di modelli complessi <sup>1</sup>.

### 3.1 Introduzione al Machine Learning

Il Machine Learning (apprendimento automatico, o *ML*), è una sottocategoria nell'ambito dell'intelligenza artificiale (*IA*). A differenza di quest'ultima, che è un concetto più ampio che include vari tipi di algoritmi, il Machine Learning si concentra sulla capacità dei sistemi di analizzare ed elaborare dati per migliorare le proprie prestazioni nel tempo. I modelli di *ML* imparano dai dati di addestramento, generalizzando da questi per effettuare previsioni o prendere decisioni su dati nuovi e mai visti. Essi riconoscono pattern e relazioni nei dati, estrapolando automaticamente le caratteristiche più significative. Questa capacità di apprendere dall'esperienza è essenziale per gestire attività complesse, dove le regole e i pattern non sono sempre esplicitamente definiti.

L'invenzione dell'algoritmo di backpropagation negli anni '70 ha rappresentato una svolta nel campo del Machine Learning, permettendo lo sviluppo di reti neurali più complesse e facilitando l'addestramento di relazioni non lineari nei dati. A partire dal 2010, tre importanti sviluppi hanno favorito una rinascita dell'IA, portando all'ampia diffusione che osserviamo oggi [24]:

- L'aumento delle capacità di calcolo grazie all'elaborazione parallela offerta nelle unità di elaborazione grafica (GPU), che ha reso possibile l'addestramento di modelli su larga scala;

---

<sup>1</sup><https://it.eitca.org/categoria/intelligenza-artificiale/>

- L'introduzione delle reti neurali convoluzionali (*CNN* 3.2.1), che ha ottimizzato l'analisi di grandi dataset, in particolare nel campo delle immagini;
- La disponibilità di grandi dataset pubblici su Internet, che ha facilitato l'addestramento dei modelli d'intelligenza artificiale.

Un aspetto fondamentale nel *ML* è la tecnica di addestramento, che può essere supervisionato o non supervisionato. L'apprendimento supervisionato utilizza dati etichettati che permettono al modello la classificazione in base a caratteristiche note. Al contrario, l'apprendimento non supervisionato non richiede etichette e individua autonomamente pattern nei dati, utile quando etichettare i dati è difficile o costoso. In meteorologia, la combinazione di questi approcci permette soluzioni ibride che forniscono precisione e flessibilità, data la disponibilità di dati non etichettati da radar, satelliti e stazioni meteo, e dati storici già etichettati.

### 3.1.1 Deep Learning

Il Deep Learning (apprendimento profondo, o *DL*), è una branca dell'apprendimento automatico focalizzata sull'addestramento di reti neurali artificiali. Questo approccio permette di modellare e comprendere schemi complessi dai dati, eliminando spesso la necessità di una pesante ingegneria manuale delle caratteristiche. Negli ultimi anni, il *DL* è stato utilizzato per migliorare le previsioni delle precipitazioni, offrendo il vantaggio di non basarsi su presupposti fisici specifici [7]. Attualmente, i metodi di nowcasting delle precipitazioni basati sul *DL* trattano il processo come un problema di previsione di sequenze temporali. Numerosi studi dimostrano che questo approccio può superare le tecniche tradizionali nella gestione della natura non lineare delle previsioni meteorologiche. Tuttavia, sebbene i modelli basati su *DL* possano estrarre dettagli rilevanti da grandi quantità di dati, rimane complessa l'implementazione su larga scala. Una delle principali difficoltà è la mancanza di accuratezza nelle previsioni a lungo termine [3, 6].

Il successo di questi modelli dipende spesso anche dall'integrazione di vincoli o informazioni a priori: i ricercatori, spesso inconsapevolmente, impongono vincoli attraverso la selezione dei dati, la scelta dell'architettura del modello o la configurazione di parametri specifici. Negli studi meteorologici, filtrare e preparare i dati in modo adeguato può favorire l'emergere di pattern utili. Per questo motivo, un sistema di *DL* per previsioni meteorologiche risulterà più efficace se integra elementi come vincoli fisici o caratteristiche progettate ad hoc.

## 3.2 Reti neurali

Una rete neurale è organizzata in livelli, o strati, composti da neuroni che elaborano e trasmettono informazioni tramite connessioni ponderate, in cui i pesi determinano la forza delle connessioni tra neuroni. Il primo strato, detto strato di input, riceve i dati iniziali, mentre l'ultimo strato, o strato di output, produce il risultato finale. Tra questi due livelli, ci sono uno o più strati nascosti che trasformano i dati, consentendo alla rete di riconoscere e apprendere schemi complessi. Nel *DL*, la profondità della rete si riferisce al numero di strati nascosti presenti. Reti neurali profonde, con molti strati



nascosti, riescono a catturare rappresentazioni più ricche rispetto alle reti più semplici. Ogni neurone riceve input dal livello precedente, applica una trasformazione matematica e genera un output. Questa trasformazione è tipicamente una somma ponderata degli input, seguita dall'applicazione di una funzione di attivazione che introduce non linearità nel modello. Senza questa non linearità, la rete sarebbe equivalente a un perceptrone a strato singolo e potrebbe risolvere solo problemi lineari.

Al termine del processo, la rete genera un output che viene confrontato con il valore desiderato tramite la funzione di perdita. La funzione di perdita quantifica la differenza tra l'output previsto e quello reale, fornendo un'indicazione sull'errore complessivo della rete. Si cerca di minimizzare questo errore per migliorare le prestazioni del modello tramite algoritmi di ottimizzazione, come la discesa del gradiente. Questo algoritmo in particolare calcola il gradiente della funzione di perdita rispetto ai pesi, determinando la direzione in cui i pesi dovrebbero essere regolati per ridurre l'errore. Poiché il gradiente indica la direzione di massima crescita della funzione di perdita, muovendosi nella direzione opposta si riesce a minimizzarla. L'aggiornamento dei pesi avviene tramite il processo di backpropagation: questo calcola il gradiente dell'errore a ritroso, partendo dallo strato di output fino a quello di input. In pratica, usa la funzione di perdita per determinare quanto ogni peso ha contribuito all'errore totale e aggiorna i pesi di conseguenza, riducendo progressivamente l'errore. Durante l'addestramento, i processi di calcolo dell'errore, ottimizzazione e backpropagation vengono eseguiti in modo iterativo per ciascun esempio del set di addestramento. Questi cicli permettono alla rete di migliorare progressivamente la sua capacità di fare previsioni accurate.

### 3.2.1 Reti Neurali Convoluzionali

Le reti neurali convoluzionali (Convolutional Neural Network, o *CNN*) rappresentano oggi l'approccio standard nella *DL* per il riconoscimento delle immagini. Grazie alla loro architettura specializzata, progettata per gestire efficacemente i dati spaziali, sono diventate fondamentali per numerose applicazioni pratiche. Il successo delle *CNN* è dovuto alla loro capacità di catturare modelli locali, alla robustezza rispetto a piccole traslazioni e alla scalabilità rispetto a grandi set di dati. Inoltre, trovano impiego in compiti generativi. L'architettura tipica di una *CNN* include vari livelli, ognuno con un ruolo distinto nell'elaborazione e nell'analisi dei dati di input (3.1):

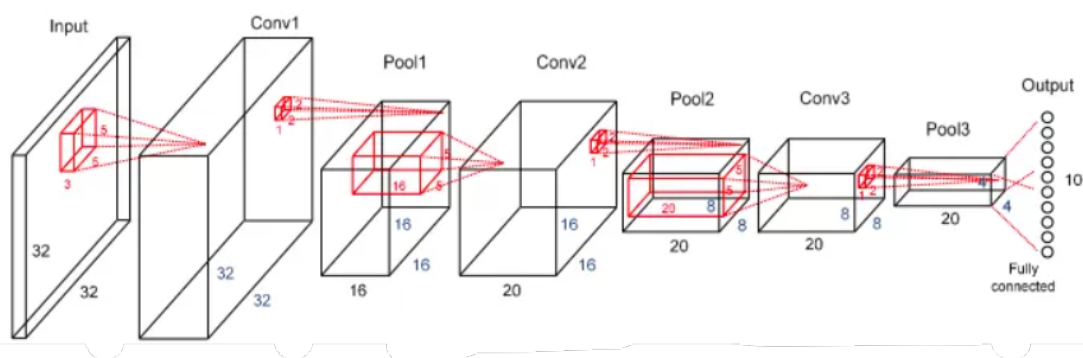


Figura 3.1: Esempio di struttura di una rete CNN

- **Livello di input:** riceve i dati dell'immagine grezza come input, rappresentata come una matrice di valori di pixel. La dimensione del livello di input è determinata dalle dimensioni dell'immagine di input;
- **Livelli convoluzionali:** sono la componente principale di una rete *CNN*. Estraggono le caratteristiche dai dati applicando una serie di filtri, acquisendo modelli locali e relazioni spaziali. Ogni neurone nel livello è collegato a una regione locale dell'immagine e il filtro scorre su questo campo calcolando prodotti scalari tra i pesi del filtro e i pixel. Il risultato è una mappa di caratteristiche (feature map), che identifica la presenza di pattern specifici. Man mano che si procede nella rete, i filtri rilevano caratteristiche sempre più astratte, consentendo l'apprendimento di gerarchie complesse che migliorano la comprensione delle immagini;
- **Livelli di pooling:** utilizzati per ridurre le dimensioni spaziali delle feature map, diminuendo il numero di parametri della rete. La riduzione della dimensionalità contribuisce a prevenire l'overfitting e migliorare l'efficienza computazionale. Il pooling introduce inoltre invarianza di traduzione, ossia la capacità di riconoscere una caratteristica indipendentemente da piccole variazioni nella sua posizione all'interno dell'immagine. Questo consente di concentrarsi sull'estrazione delle caratteristiche più importanti, piuttosto che sulla loro esatta posizione. I livelli di pooling vengono generalmente applicati dopo i livelli convoluzionali e sono responsabili dell'estrazione di rappresentazioni di basso e medio livello dai dati di input;
- **Livelli completamente connessi:** simili a quelli presenti nelle reti neurali tradizionali, in questi ogni neurone è collegato a tutti i neuroni del livello precedente. Nelle *CNN*, i livelli completamente connessi vengono spesso posizionati verso la fine della struttura e combinano le caratteristiche apprese dai livelli convoluzionali e di pooling e per effettuare previsioni finali;
- **Livello di output:** è l'ultimo livello della rete, responsabile della produzione della previsione. Nelle attività di regressione è costituito da un singolo neurone che fornisce previsioni continue.

Agli strati della rete è spesso applicato il dropout, una tecnica di regolarizzazione per prevenire l'overfitting. Durante l'addestramento, una frazione di neuroni viene disattivata casualmente, limitando la co-adattabilità dei neuroni e incoraggiando la rete ad apprendere caratteristiche più robuste e generalizzabili.

### 3.2.2 Architettura U-Net

L'architettura U-Net è una *CNN* avanzata, originariamente sviluppata per la segmentazione di immagini ad alta precisione. Grazie al suo design semplice ed efficiente per l'estrazione delle caratteristiche, ha acquistato grande popolarità in numerose applicazioni di elaborazione e manipolazione delle immagini. La caratteristica distintiva dell'U-Net è la sua struttura a "U" (3.2), composta da due sezioni principali: un percorso di contrazione (codificatore) e un percorso di espansione (decodificatore). Il percorso di contrazione riduce le dimensioni spaziali dell'immagine mediante operazioni di pooling e convoluzione, aumentando al contempo la profondità delle mappe

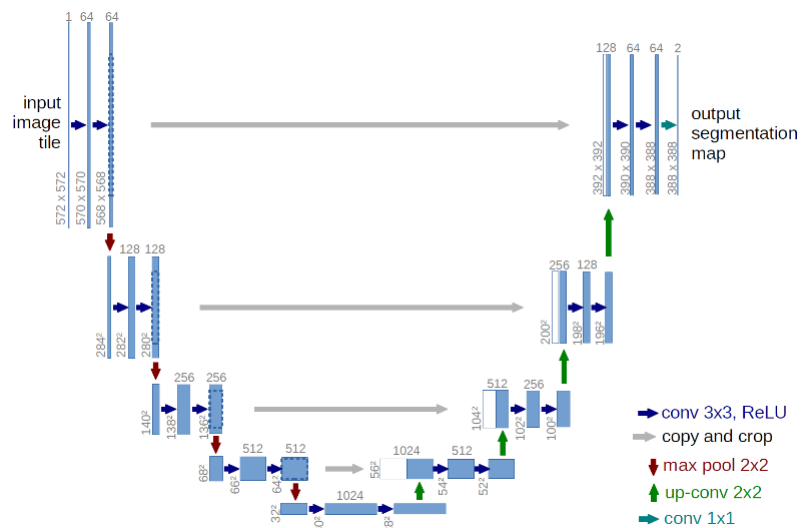


Figura 3.2: Esempio di un'architettura di rete U-Net

delle caratteristiche, il che permette di catturare dettagli di alto livello e schemi complessi. Il percorso di espansione, che rispecchia simmetricamente quello di contrazione, ripristina gradualmente la risoluzione spaziale mediante up-convoluzioni, con l'obiettivo di localizzare con precisione le strutture d'interesse. Un aspetto chiave dell'architettura U-Net è l'uso delle connessioni di salto (skip connections), che collegano direttamente le mappe delle caratteristiche del percorso di contrazione con quelle nel percorso di espansione a pari livello. Queste connessioni aiutano a preservare i dettagli fini dell'immagine che potrebbero andare persi durante il downsampling, migliorando così l'accuratezza della segmentazione.

Teoricamente, le reti U-Net possono essere annidate all'infinito, consentendo di adattare l'architettura alle specifiche esigenze di applicazioni diverse. Questo la rende uno strumento versatile e potente per l'analisi e la previsione basate sulle immagini. Tuttavia, l'aumento esponenziale dei parametri con l'aumentare dei livelli di nidificazione rende l'addestramento più complesso e computazionalmente costoso [6]. La capacità dell'U-Net di elaborare immagini con alto dettaglio la rende particolarmente utile all'interpretazione di dati meteorologici, come quelli provenienti da sensori, immagini satellitari e radar. Grazie alla sua struttura, la U-Net può distinguere le diverse condizioni meteorologiche a livello di pixel. La configurazione della rete si stabilisce scegliendo il numero di blocchi di sottocampionamento e di canali per ciascun blocco, con una fase di sovracampionamento che segue una struttura simmetrica. La risoluzione dell'output dipende dalla risoluzione dell'immagine di input [25].



## Capitolo 4

# Modelli di diffusione per la previsione delle precipitazioni

L'intelligenza artificiale generativa è una sottodisciplina innovativa dell'IA, specializzata nella creazione di dati nuovi e realistici, che siano coerenti con il set di dati in ingresso. Questa branca sta trasformando numerosi settori, inclusi la generazione di testo, la creazione di immagini e la modellazione tridimensionale. I modelli generativi sono in grado di produrre contenuti originali, imitando la struttura e le proprietà dei dati da cui apprendono. La varietà e la rappresentatività dei dati di addestramento influiscono direttamente sulla loro capacità di generalizzare, e, di conseguenza, di generare contenuti utili e pertinenti. Questa capacità di generare dati offre un'opportunità unica per migliorare la precisione delle previsioni meteorologiche, consentendo l'analisi di scenari e l'individuazione di modelli complessi.

### 4.1 Modelli di diffusione

I modelli generativi di diffusione (Denoising Diffusion Models, o *DDM*) rappresentano una classe di modelli generativi particolarmente efficaci nella modellazione di dati complessi e ad alta dimensionalità. Al centro della loro operatività vi è il concetto di diffusione stocastica, un processo matematico che descrive l'evoluzione casuale e continua nel tempo di particelle all'interno di sistemi complessi. Tale diffusione è alla base di fenomeni naturali come il moto browniano e la diffusione di molecole in un fluido, nonché di qualsiasi sistema caratterizzato da incertezze o variabilità intrinseca.

Nel contesto fisico, il comportamento delle particelle si traduce in uno spostamento da regioni ad alta concentrazione a quelle a bassa concentrazione, portando gradualmente ad uno stato di equilibrio. Analogamente, nell'ambito dell'apprendimento automatico, i *DDM* simulano il passaggio da uno stato di elevato rumore o squilibrio a uno stato più omogeneo e controllato, con l'obiettivo di generare nuovi dati significativi. Più in dettaglio, i modelli di diffusione addestrano una rete neurale a trasformare distribuzioni di dati semplici e rumorosi in campioni complessi, attraverso un processo iterativo. L'obiettivo finale è generare immagini realistiche partendo da rumore casuale, il quale può essere controllato in modo parametrico. Questo processo si articola in due fasi principali (4.1):

- Diffusione diretta (forward process): in questa fase, i dati vengono progressivamente corrotti mediante l'aggiunta di rumore. Il processo provoca una perdita

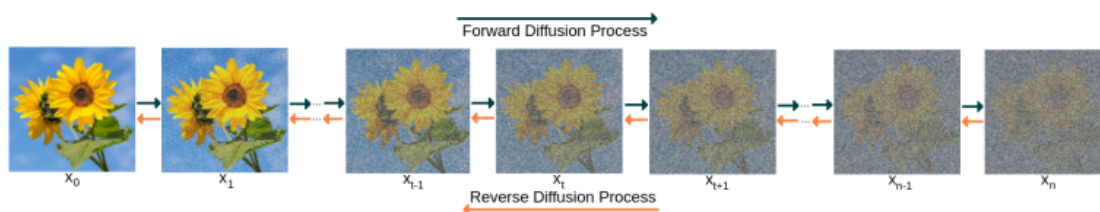


Figura 4.1: Processo di diffusione in avanti e inverso

graduale delle informazioni visive rilevanti, deteriorando la struttura informativa dei dati e rendendoli sempre meno riconoscibili. Questa fase è cruciale per comprendere il modo in cui i dettagli si deteriorano sotto l'influenza del rumore;

- Diffusione inversa (reverse process): il modello apprende come rimuovere gradualmente il rumore, ripristinando la fedeltà dei dati originali. Questo processo di denoising inverte la corruzione introdotta nella fase precedente, migliorando ad ogni passo la qualità dei campioni generati.

L'aggiunta e la rimozione del rumore avvengono senza una formulazione probabilistica complessa. In questo modo, il modello si allena a denoizzare progressivamente i dati, perfezionando continuamente la sua capacità di generare campioni realistici e di alta qualità.

#### 4.1.1 Applicazioni nella previsione di eventi estremi

La generazione dei dati nei modelli *DDM* richiede spesso un meccanismo che orienti il processo verso output specifici e desiderati. Questo approccio è noto come diffusione condizionata o guidata. In altre parole, il modello non si limita a generare dati casuali da una distribuzione generica, ma è diretto attraverso vincoli precisi, come etichette di classe o incorporamenti di immagini e testo. Questa capacità di condizionamento offre un potenziale significativo nella modellazione predittiva.

Nelle previsioni meteorologiche, l'applicazione del condizionamento su variabili rilevanti per gli eventi estremi consente di ottenere una maggiore precisione nella previsione di questi fenomeni complessi, caratterizzati dall'interazione di molteplici variabili ambientali. Inoltre, la natura stocastica dei modelli di diffusione li rende particolarmente adatti all'uso di tecniche di ensemble, ampiamente utilizzate in meteorologia. I modelli di ensemble generano diverse simulazioni a partire dalle stesse condizioni iniziali, variando leggermente i parametri di input. Questo approccio migliora l'accuratezza delle previsioni e consente di valutare le probabilità di scenari rari o catastrofici. L'integrazione delle tecniche di diffusione con metodi di ensemble potrebbe rappresentare una direzione promettente per migliorare ulteriormente la robustezza delle previsioni meteorologiche.

## 4.2 Introduzione a un modello generativo reale

Dopo un'introduzione ai principi della modellazione generativa e dei modelli di diffusione, viene presentato un modello specifico appartenente a questa classe. Il modello descritto in [25] è un modello di diffusione del denoising, addestrato su una sequenza

di dati di pioggia, integrati con altre variabili metereologiche. È progettato per la previsione a breve termine di precipitazioni estreme; prevede un'architettura U-Net per la generazione simultanea della previsione per le tre ore successive.

I modelli di diffusione per il denoising (Denoising Diffusion Probabilistic Models, o *DDPM*), rappresentano una variante specifica dei *DDM*. I *DDPM* affinano e formalizzano in modo rigoroso il processo di aggiunta e rimozione del rumore, che in questo caso è sempre di tipo gaussiano. Questi modelli si basano su un processo di Markov, in cui ogni aggiornamento dipende unicamente dallo stato precedente, rendendo il meccanismo matematicamente ben definito e controllabile. Grazie a queste innovazioni, i *DDPM* ottengono una maggiore precisione nella stima e nella rimozione del rumore, generando dati di qualità superiore e più stabili rispetto ai modelli di diffusione tradizionali.

### 4.2.1 Architettura del modello

Il principale componente addestrabile in un modello di diffusione è la rete di denoising, che in questo caso è unica e gestisce entrambe le fasi del processo di diffusione. Durante l'addestramento, la rete riceve in input immagini alterate dal rumore, accompagnate da un valore che rappresenta la varianza del rumore in quel momento. Questa varianza riflette l'entità con cui il noise ha alterato l'immagine originale. L'obiettivo è imparare a stimare con precisione la quantità di rumore presente in quell'istante, così da poterlo rimuovere gradualmente nella fase successiva di diffusione inversa. Una volta addestrata, la rete utilizza le sue stime per riconoscere e rimuovere il rumore partendo da uno stato iniziale di rumore puro e procedendo passo dopo passo fino a ottenere un'immagine simile all'originale. Il processo è iterativo perché la rete è allenata per stimare piccole quantità di rumore, suddividendo il compito in parti più gestibili.

La rete di denoising scelta è una U-Net (3.2.2). La sua combinazione con i processi di diffusione ha dimostrato prestazioni notevoli, spesso superiori a quelle di modelli all'avanguardia come le reti generative avversarie (*GAN* 4.3). Grazie alla sua capacità di estrazione delle caratteristiche, la U-Net gestisce efficacemente la complessità del rumore, e la sua modularità ne facilita l'integrazione con tecniche più avanzate.

Negli esperimenti trattati in tesi, la larghezza della rete è impostata a [64, 128, 256]. Il rumore è gestito attraverso una tecnica chiamata sinusoidal embedding, che trasforma la variabile scalare legata al rumore in una rappresentazione periodica, utilizzando funzioni sinusoidali e cosinusoidali. Questo approccio permette alla rete di determinare in quale fase del processo di diffusione si trova, facilitando una rimozione progressiva e controllata del rumore. Invece, la proporzione di segnale, ossia la parte non corrotta dell'immagine originale durante il processo di diffusione, è regolata dalle variabili *min\_signal\_rate* e *max\_signal\_rate*, impostate rispettivamente a 0.015 e 0.95.

### 4.2.2 Strumenti utilizzati

*Leonardo*<sup>1</sup> è un supercomputer di classe pre-exascale, ovvero una macchina ad alte prestazioni che rappresenta un passo passo intermedio tra i supercomputer attuali e quelli di prossima generazione, progettati per raggiungere la soglia dell'exaFLOP, equivalente a un miliardo di miliardi ( $10^{18}$ ) di operazioni in virgola mobile al secondo.

<sup>1</sup><https://leonardo-supercomputer.cineca.eu/>

Inaugurato nel novembre 2022 presso il Tecnopolo di Bologna, un hub specializzato in previsioni meteorologiche e dati climatici, *Leonardo* è ospitato e gestito dal *Cineca*<sup>2</sup>, un consorzio interuniversitario senza scopo di lucro che opera come centro di supercalcolo multidisciplinare. *Cineca* è impegnato nello sviluppo di applicazioni avanzate con un focus particolare sull'IA e le sue applicazioni scientifiche. Secondo la *Top500 List*<sup>3</sup>, *Leonardo* è classificato come il sesto supercomputer più potente al mondo e il secondo in Europa. La macchina, fornita da Atos nell'ambito del progetto europeo *EuroHPC*, è parte di un'iniziativa per lo sviluppo di infrastrutture di supercalcolo in diverse nazioni europee, tra cui Austria, Slovenia, Slovacchia, Grecia e Ungheria. Messo a disposizione della comunità scientifica italiana ed europea, il supercomputer affronta sfide computazionali avanzate in vari settori, come l'IA, la robotica, la farmacologia e la gestione dei rischi legati a eventi naturali estremi. Con la crescente domanda di capacità di calcolo, l'accesso a tecnologie sempre più performanti è cruciale per supportare la comunità scientifica.

L'uso di *Leonardo* è stato fondamentale per l'addestramento e la valutazione del modello descritto, consentendo l'esecuzione di calcoli complessi in tempi ridotti e migliorando la produttività del progetto. L'accesso al supercomputer è stato concesso previa richiesta a *Cineca*, facilitando la collaborazione tra i ricercatori e gli studenti coinvolti.

### 4.2.3 Dataset

L'utilizzo di dati accurati è essenziale per qualsiasi analisi meteorologica. In generale, i dataset per questo campo sono classificati in due categorie principali: dati osservativi e dati di rianalisi. I dati osservativi, considerati la base primaria per le previsioni meteorologiche [17], sono raccolti da enti specializzati tramite misurazioni sul campo. Tuttavia, l'accesso a questi può risultare complesso e frammentato a causa della variabilità nelle metodologie di raccolta, rendendo difficile garantire un'uniformità e una condivisione efficiente delle informazioni. I dati di rianalisi, invece, offrono un'alternativa più omogenea, poichè combinano le osservazioni dirette con modelli numerici che sintetizzano lo stato dell'atmosfera. Partendo da una base adeguata, la rianalisi è in grado di generare dataset dotati di:

- Copertura globale completa;
- Alta rappresentatività delle condizioni meteorologiche;
- Riduzione di disomogeneità spazio-temporali;
- Disponibilità di dati sia storici che in tempo quasi reale (near real-time).

Grazie a queste caratteristiche, i dataset di rianalisi sono particolarmente pratici e utilizzati nelle analisi meteorologiche.

Uno dei dataset più noti e utilizzati nella comunità scientifica è *ERA5* (5th ECMWF ReAnalysis) [26], parte del progetto *C3S* promosso dal centro di previsione *ECMWF*. *ERA5* rappresenta una rianalisi globale affidabile, con aggiornamenti regolari e copre il periodo dal 1940 ad oggi. Offre stime orarie di numerose variabili atmosferiche, terrestri e oceaniche, con una risoluzione spaziale orizzontale di circa 31 km e verticale

---

<sup>2</sup><https://www.cineca.it/it>

<sup>3</sup><https://www.top500.org/lists/top500/list/2023/11/>



di circa 137 livelli. Sono rilevanti le sue applicazioni nell'analisi delle tendenze delle precipitazioni legate ai cambiamenti climatici e agli eventi meteorologici estremi. Il modello trattato utilizza la variabile *Total Precipitation* (4.2) del dataset *ERA5* come parametro principale, misurata in metri (m).

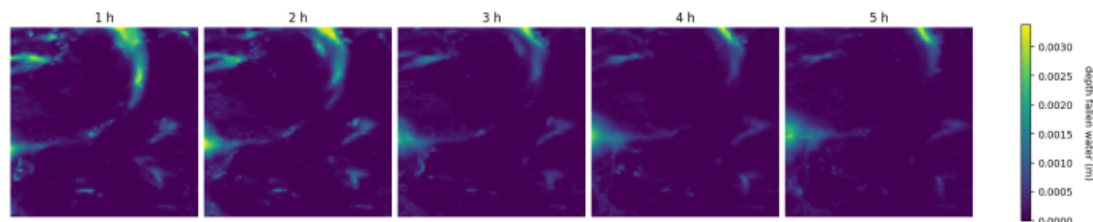


Figura 4.2: Esempio di dati sulle precipitazioni dal set ERA-5

Questo dato cumulativo rappresenta la quantità di acqua, sia liquida che solida, che cade sulla superficie terrestre e fornisce una stima della profondità dell'acqua distribuita uniformemente su ogni cella della griglia di  $31 \text{ km}^2$ . La regione di interesse copre gran parte dell'Europa occidentale, con confini latitudinali che vanno da  $-12^\circ$  a  $12^\circ$  e longitudinali da  $36^\circ$  a  $60^\circ$ . L'analisi considera un intervallo temporale di sei anni, dal 2015 al 2020, con misurazioni orarie. I dati raccolti nel periodo 2016-2020 vengono utilizzati per addestrare il modello, mentre quelli del 2015 vengono impiegati per i test [25].

#### 4.2.3.1 Informazioni aggiuntive

Oltre alla semplice distinzione tra presenza e assenza di precipitazione, vengono presi in considerazione ulteriori parametri derivati dal dataset *ERA5* per rendere le previsioni più accurate e dettagliate. Come discusso nella sezione (1.2.1), variabili come la temperatura, il vento e l'umidità influenzano significativamente la formazione, la distribuzione e l'intensità delle precipitazioni.

I parametri inclusi sono i seguenti (4.3):

- *Maschera terra-mare* (u.m. scalare): indica la percentuale di terra rispetto al mare in ogni punto della griglia, con valori compresi tra 0 (completamente mare) e 1 (completamente terra). Le diverse proprietà fisiche di terra e mare, come la capacità termica e l'albedo (riflettività), influenzano i modelli atmosferici e climatici;
- *Geopotenziale* (u.m.  $m^2 s^{-2}$ ): indica l'energia potenziale gravitazionale per unità di massa, relativa al livello medio del mare. Questo parametro è utile per identificare formazioni atmosferiche come cicloni e anticicloni, poichè influenza direttamente la circolazione dell'aria e la distribuzione della pressione a diverse altitudini;
- *Copertura nuvolosa totale* (u.m. scalare): indica la frazione di cielo coperta da nuvole per ciascun punto della griglia, con valori compresi tra 0 (cielo sereno) e 1 (completamente coperto). Le coperture nuvolose influisce sulla radiazione solare, sul raffreddamento terrestre e sulla formazione delle precipitazioni;
- *Timestamp* (u.m.  $[m, d, h]$ ): indica l'informazione temporale relativa all'inizio della sequenza, codificata in una matrice di dimensioni  $(96 \times 96)$  e comprende

mese, giorno e ora. Questo parametro è utile per contestualizzare i dati all'interno del modello;

- *Velocità del vento a 100 m* (u.m.  $ms^{-1}$ ): indica la velocità del vento a un'altezza di 100 metri dalla superficie terrestre. Calcolata dalle componenti orientale ( $u$ ) e settentrionale ( $v$ ), la velocità totale si ottiene calcolando  $\sqrt{u^2 + v^2}$ . Il vento trasporta umidità da diverse aree geografiche, influenzando la formazione e l'intensità delle precipitazioni.

Una selezione coerente e precisa dei parametri di input, rimuovendo quelli superflui o irrilevanti, può migliorare sensibilmente le prestazioni del modello [27]. Tuttavia, la scelta dei parametri resta in parte sperimentale e richiede valutazioni approfondite per determinare il reale contributo di ciascuna variabile sulla qualità delle previsioni.

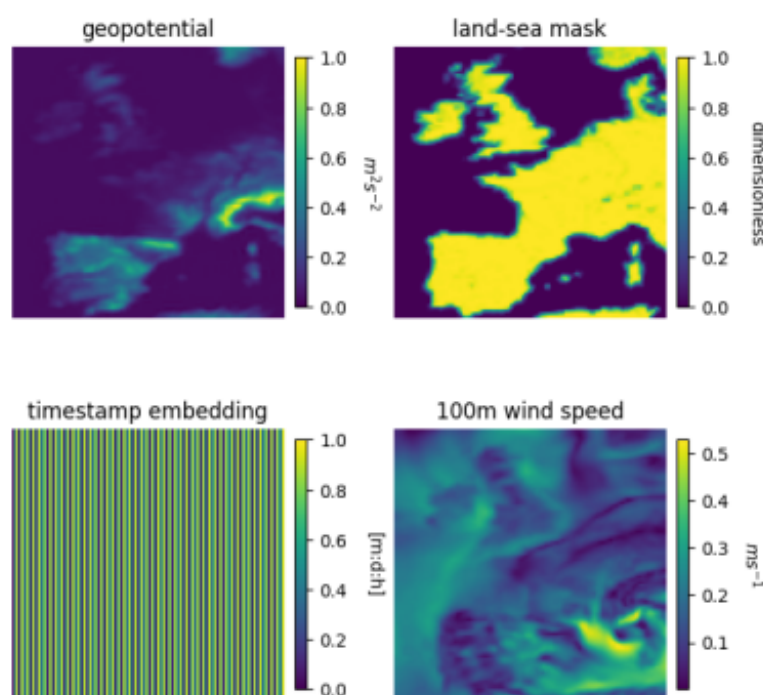


Figura 4.3: Esempio delle funzionalità aggiuntive

#### 4.2.4 Preprocessamento dei dati

Nel campo del *DL*, il preprocessamento dei dati è una fase fondamentale, di cui la normalizzazione rappresenta uno degli approcci più comunemente utilizzati [27]. La normalizzazione dei dati consente di trasformare gli input in un formato compatibile con l'architettura della rete, migliorando l'efficienza dell'addestramento e facilitando la convergenza del modello. Nel caso specifico, i valori del dataset sono normalizzati tra 0 e 1, dividendo ciascun valore per il massimo riscontrato nel dataset, garantendo così una scala coerente per tutti i dati.

#### 4.2.5 Condizionamento dei dati

Il modello di diffusione sfrutta un insieme di informazioni di condizionamento per migliorare la qualità delle previsioni. Questo condizionamento viene implementato

alimentando la rete con una serie di dati concatenati lungo la dimensione del canale, insieme a immagini rumorose che costituiscono l'output della previsione. Ogni porzione temporale dei dati in input viene trattata in modo analogo ai canali di colore di un'immagine RGB, e su ciascuna di esse vengono applicate operazioni di convoluzione in modo indipendente. La struttura dei dati di condizionamento è organizzata come mostrato nella tabella (4.1).

Posizione	Frame
0	Geopotenziale
1	Maschera terra-mare
2-5	Timestamp
6-9	Velocità del vento (u + v)
10	Copertura nuvolosa
11-17	Precipitazioni precedenti
18-20	Precipitazioni future

Tabella 4.1: Organizzazione dei frame nel canale di condizionamento

#### 4.2.5.1 Generatore

Il generatore di dati è un elemento fondamentale nel processo di apprendimento del modello, poichè organizza in modo efficiente le informazioni di condizionamento, ottimizzando la capacità predittiva del modello. Durante l'addestramento, il generatore crea batch di sequenze di dati meteorologici. Con una dimensione batch pari a 32, i dati di input alla rete di denoising hanno una forma di [32, 96, 96, 21], dove l'ultima dimensione include sia le informazioni di condizionamento che le immagini disturbate da rumore. L'output finale ha una forma di [32, 96, 96, 3], rappresentando i 3 frame denoised. Il modello è addestrato per prevedere le precipitazioni future a 1, 2 e 3 ore, in maniera simultanea. La fase di inferenza del modello è descritta in dettaglio nel capitolo (5) dedicato agli esperimenti.

## 4.2.6 Addestramento del modello

Come illustrato nei paragrafi precedenti, l'addestramento del modello è basato su un sottinsieme del dataset *ERA5*, che include dati orari relativi agli anni 2016-2020. Questi dati vengono forniti al modello tramite un generatore che integra le informazioni di condizionamento. Per massimizzare la disponibilità di dati in fase di train, non è stato utilizzato un set di validazione separato oltre al set di test. L'addestramento si basa sulla regolazione di diversi iperparametri e l'ottimizzazione delle funzioni di perdita, al fine di migliorare le prestazioni predittive del modello in particolare per livelli di pioggia di almeno 8 mm in un'ora.

### 4.2.6.1 Iperparametri

Ogni modello richiede un insieme di iperparametri, determinati attraverso un processo noto come regolazione degli iperparametri. A differenza dei parametri, che vengono automaticamente ottimizzati durante il processo di addestramento, gli iperparametri sono variabili di configurazione esterne, impostate manualmente dai data scientist e

sono cruciali per determinare la struttura e il comportamento della rete, con l'obiettivo di massimizzare le prestazioni complessive del modello. Gli iperparametri utilizzati durante l'addestramento, mantenuti costanti per tutti gli esperimenti, sono:

- *optimizer*: *AdamW*, variante dell'algoritmo *Adam* con la regolarizzazione *L2*, per aggiornare i pesi ;
- *batch\_size*: 32, indica il numero di elementi processati in una singola iterazione;
- *n\_epoch*: 40, indica il numero di volte che il modello passa attraverso l'intero set di dati durante l'allenamento;
- *learning\_rate*: 1e-4, rappresenta la velocità con cui il modello si adatta;
- *weight\_decay*: 1e-5, indica il tasso di regolarizzazione applicato ai pesi per prevenire l'overfitting.

#### 4.2.6.2 Funzioni di perdita

Specialmente in modelli complessi come le reti neurali, è comune utilizzare diverse funzioni di perdita per affrontare vari aspetti del compito di apprendimento. Per questo modello, è stato scelto l'errore *MAE* (*Mean Absolute Error*) come principale indicatore della differenza tra le previsioni e i dati reali. Il *MAE* è espresso come segue ed è calcolato a livello di singolo pixel (*pixel-wise MAE*), approccio tipico per i modelli che operano su immagini o mappe spaziali:

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^n |x_i - \hat{x}_i| \quad (4.1)$$

In aggiunta al *MAE*, sono state impiegate per una valutazione più completa:

- *image\_loss*: misura l'efficacia del modello nella generazione delle immagini;
- *noise\_loss*: valuta la capacità del modello di rimuovere il rumore dai dati.

Questo approccio combinato consente di sviluppare un modello più robusto e versatile, capace di affrontare diversi aspetti del problema di previsione delle precipitazioni.

### 4.3 Altri approcci

Attualmente, le tecniche di *DL* per la previsione delle precipitazioni possono essere suddivise in quattro categorie principali [21], tra cui le reti *CNN* (3.2.1) rappresentano una delle opzioni:

- Reti Neurali Ricorrenti (*RNN*): progettate per acquisire le dipendenze a lungo termine nei dati sequenziali. Le *RNN* sono state ampiamente utilizzate nel now-casting delle precipitazioni. Tuttavia, presentano limitazioni significative, come lunghi tempi di addestramento e problemi di gradiente esplosivo o scomparse. Avere il gradiente che svanisce ostacola la capacità del modello di mantenere in memoria le informazioni necessarie per previsioni accurate, specialmente in

relazione a eventi meteorologici estremi che dipendono da condizioni atmosferiche evolutive. Varianti dell'architettura *RNN*, come le *Long Short-Term Memory (LSTM)*, affrontano il problema grazie a celle negli strati nascosti per immagazzinare informazioni cruciali. Alcuni esempi di modelli *LSTM* sono riportati in [15, 16, 21];

- **Metodi multiscala:** questi approcci consentono di considerare simultaneamente diverse scale temporali e spaziali, migliorando la capacità di catturare fenomeni complessi che si manifestano a diversi livelli di granularità;
- **Meccanismi di attenzione:** utilizzati per identificare in modo efficace le correlazioni tra gli elementi nelle sequenze di input. L'architettura *Transformer*, basata su questo approccio, supera la problematica del gradiente che svanisce delle *RNN* e analizza l'intera sequenza dei dati in parallelo, riducendo significativamente i tempi di addestramento. Grazie alla capacità di concentrarsi sugli elementi più rilevanti, questi modelli migliorano la qualità delle previsioni, risultando particolarmente efficaci in contesti complessi come quello della meteorologia. Esempi di architettura *Transformer* includono lo studio [2], il trasformatore autoregressivo esaminato in [8] e il trasformatore gerarchico con attenzione leggera [14], utilizzati con successo nelle nowcasting di precipitazioni estreme.

Esistono altri modelli non rientranti in queste categorie che si rivelano molto efficaci per affrontare scenari meteorologici estremi. Tra questi, i modelli ibridi che combinano punti di forza di diverse architetture [14] e le *Generative Adversarial Networks (GAN)*, ideali per la generazione di nuovi dati [6, 23]. Le *GAN* si basano su una dinamica competitiva tra un generatore, che mira a creare immagini realistiche, e un discriminatore, che tenta di distinguere tra le immagini generate e quelle reali. L'obiettivo è che il generatore prevalga, riuscendo a produrre campioni indistinguibili dai dati reali. Questa varietà di approcci mette in luce il crescente interesse per i modelli generativi e le architetture neurali, evidenziando il loro potenziale nella previsione di eventi estremi.



# Capitolo 5

## Esperimenti

In questo capitolo vengono presentate le valutazioni delle prestazioni del modello descritto in (4.2) per la previsione di precipitazioni estreme. La ricerca della configurazione ottimale del modello rimane ancora un tema aperto [25]. A tal fine, sono stati eseguiti test variando alcuni parametri chiave. I risultati delle analisi mirano a fornire una comprensione più approfondita delle capacità e dei limiti del modello, per meglio adattarlo a situazioni estreme e migliorare la sua precisione in contesti di eventi meteorologici estremi.

### 5.1 Strumenti di analisi

#### 5.1.1 Funzioni di valutazione

- *Single Diffusion*: esegue una singola previsione per batch di dati. È un approccio semplice ed efficiente dal punto di vista computazionale. In input viene fornito un batch di sequenze, che il modello normalizza e utilizza per fare la previsione. Al termine delle iterazioni, le previsioni ottenute vengono confrontate con le osservazioni reali: i valori continui delle previsioni sono convertiti in classificazioni binarie in base a soglie specifiche di pioggia, per l'analisi di metriche derivate dalla matrice di confusione (vedere codice A.1).
- *Generative Ensemble Diffusion*: utilizza un approccio ensemble per migliorare le prestazioni del modello nella previsione di eventi rari e complessi. Esegue 15 previsioni indipendenti per batch di dati e ne calcola la media per produrre il risultato finale. L'idea alla base è che la combinazione di più previsioni aiuta a ridurre l'impatto di rumore o di errori casuali, rendendo la previsione finale più robusta e precisa rispetto a una singola generazione. Il confronto coi dati reali definisce l'errore *MSE* per ogni batch (vedere codice A.2).

#### 5.1.2 Funzioni di visualizzazione

- *plot\_metrics*: permette la visualizzazione dell'andamento di metriche chiave in relazione a diverse ore di previsione (1, 2, 3 ore);
- *plot\_prevision*: progettata per visualizzare come varia una metrica rispetto a diversi livelli d'intensità di pioggia previsti (mm) per tre intervalli temporali di previsione (1, 2, 3 ore).

## 5.2 Lista esperimenti

### 5.2.1 Esperimento 1

L'obiettivo dell'esperimento è analizzare come variano l' $F_1$ -Score (2.4) e l' $F_2$ -Score (2.5) in relazione all'intensità delle precipitazioni, da valori minimi fino a un massimo di 8 mm. Questo permette di comprendere come il modello si comporta in presenza di eventi estremi. Utilizzando la funzione Single Diffusion (A.1), la soglia per i dati reali e previsti è stata variata nell'intervallo tra 1 mm e 8 mm, con incrementi di 1 mm. Per una valutazione coerente, le soglie sono state mantenute corrispondenti.

Analizzando i risultati ottenuti dall'applicazione delle funzioni `plot_metrics` (5.1, 5.2) e `plot_prevision` (5.3, 5.4), si vede che le curve relative ai livelli di precipitazione più elevati mostrano valori inferiori, evidenziando la crescente difficoltà nel fare previsioni accurate in condizioni meteorologiche estreme. Un andamento analogo si osserva nelle curve in funzione delle ore di previsione, le quali mostrano un trend decrescente. Inoltre, per l' $F_1$ -Score si hanno valori leggermente migliori rispetto a quelli per l' $F_2$ -Score: questo potrebbe significare che il modello tende ad avere una buona precisione per intensità di piogge moderate, ma è meno efficace nel catturare tutte le occorrenze di eventi estremi. In altre parole, il modello sta facendo previsioni più sicure, preferendo evitare dei *FP*.

Per monitorare quando e quanto rapidamente il modello perde la sua capacità predittiva, si possono valutare le pendenze delle rette in funzione delle ore di previsione.

Per i valori di precipitazione più estremi (8 mm):

- $F_1$ -Score: pendenza tra 1° e 2° ora (-0.33), pendenza tra 2° e 3° ora (-0.01);
- $F_2$ -Score: pendenza tra 1° e 2° ora (-0.38), pendenza tra 2° e 3° ora (-0.02).

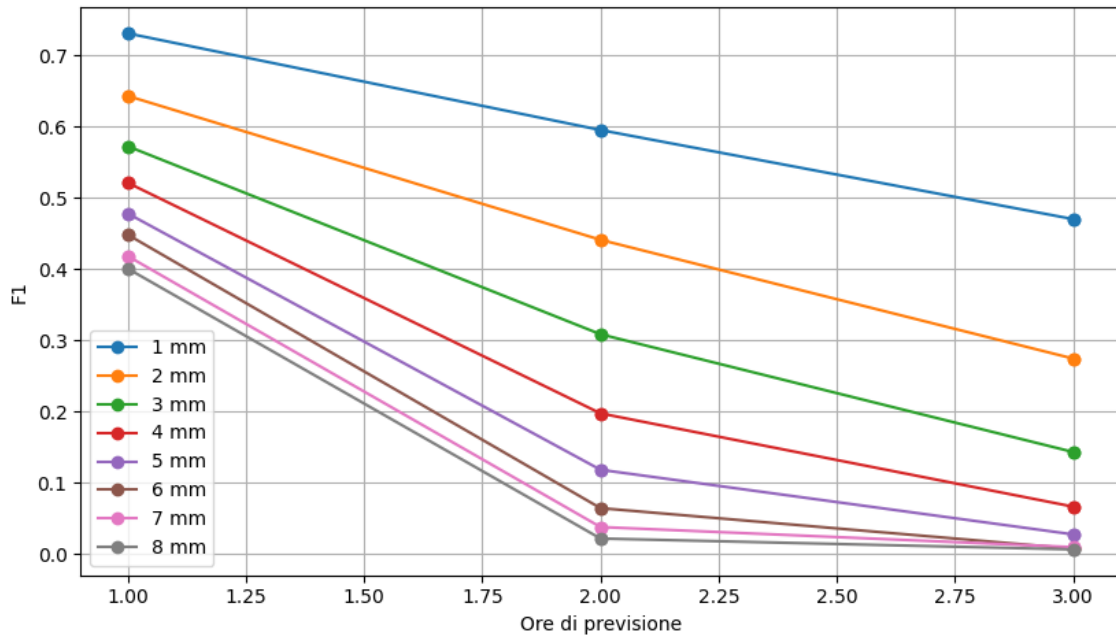
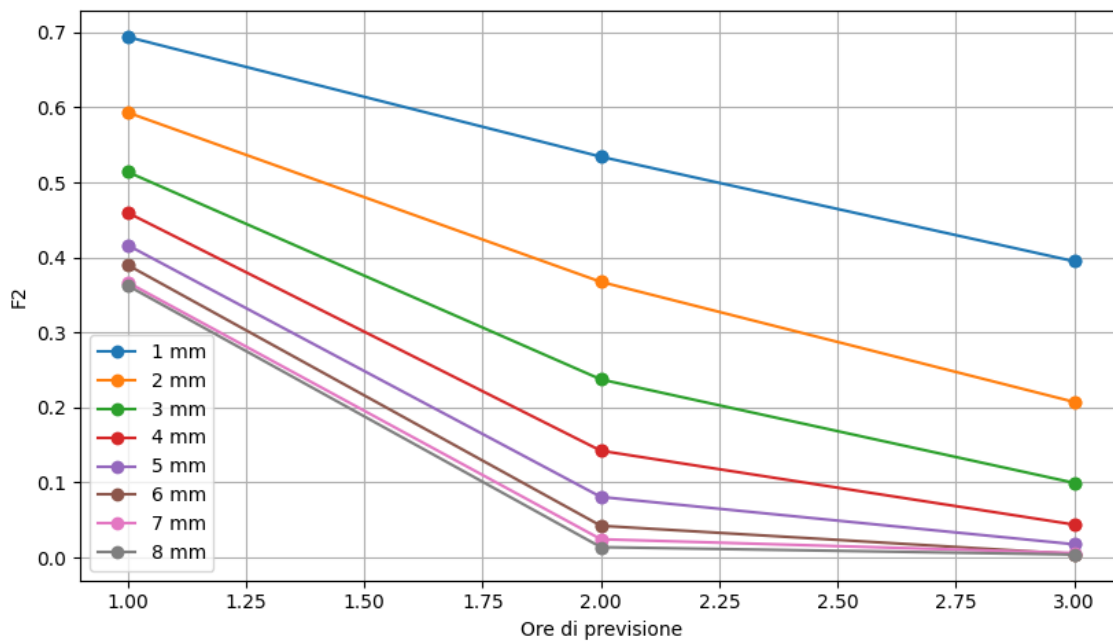
I risultati indicano che il modello subisce una riduzione delle prestazioni più marcata nelle prime ore di previsione, seguita da un rallentamento nelle ore successive.

Questo suggerisce che si conserva una discreta capacità predittiva nel breve termine, ma si incontrano maggiori difficoltà nelle previsioni a lungo termine in condizioni meteorologiche estreme. Inoltre, l' $F_2$ -Score sembra risentire maggiormente delle difficoltà nel fare previsioni accurate già nelle prime ore.

Vista l'importanza delle capacità predittive del modello in condizioni estreme, sono stati condotti ulteriori esperimenti abbassando la soglia di previsione per 8 mm di pioggia. L'obiettivo principale è migliorare le metriche di *Recall* (2.3) e di  $F_2$ -Score: con questa operazione, si aumenta il numero di *TP* identificati dal modello, diminuendo al contempo i *FN*. La riduzione di questi ultimi è particolarmente rilevante in situazioni di emergenza, dove la mancata segnalazione di un evento di pioggia estrema può comportare conseguenze gravi.

Le soglie di previsione considerate variano tra 6.5 mm e 7.5 mm. Dai risultati ottenuti in tabella (B.1), la soglia di 6.9 mm risulta fornire un miglior compromesso tra le diverse metriche analizzate.



Figura 5.1: Andamento  $F_1$ -Score in funzione delle ore di previsioneFigura 5.2: Andamento  $F_2$ -Score in funzione delle ore di previsione

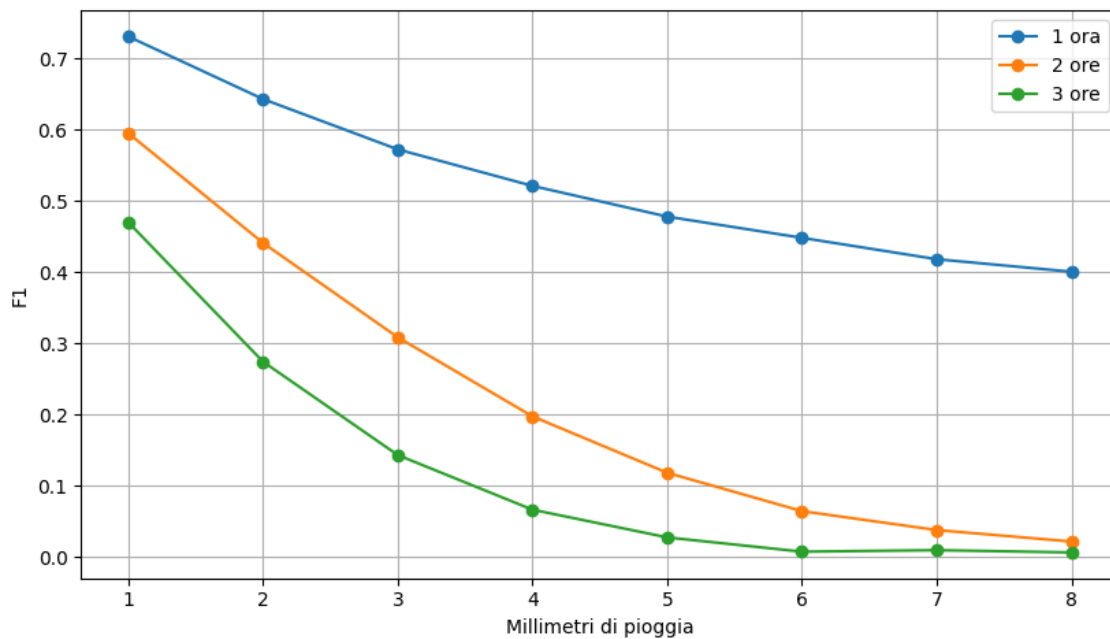


Figura 5.3: Andamento  $F_1$ -Score in funzione dei livelli di pioggia

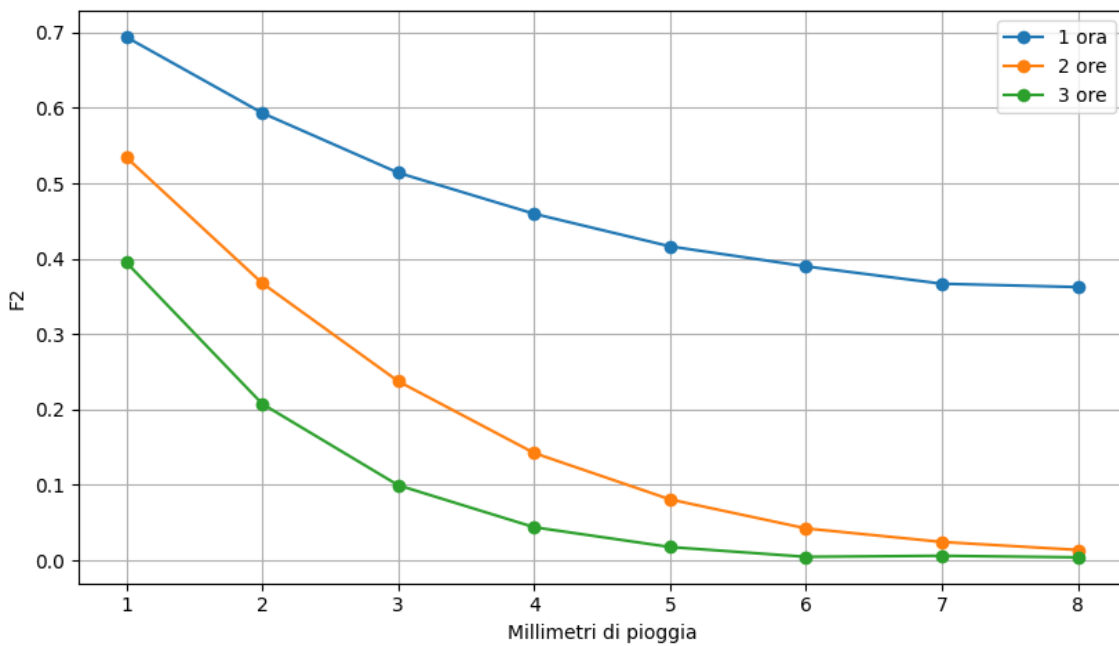


Figura 5.4: Andamento  $F_2$ -Score in funzione dei livelli di pioggia

### 5.2.2 Esperimento 2

Scopo dell'esperimento è valutare la dipendenza delle previsioni dalla scala spaziale, verificando se ridurre la risoluzione porta a un miglioramento dell'*Average Precision* (2.7). Misurare l'effetto dell'esperimento sull'*AP* è particolarmente rilevante perchè fornisce una valutazione complessiva delle prestazioni del modello nel rilevare fenomeni intensi. Altre metriche, come l'*F<sub>1</sub>-Score* e l'*F<sub>2</sub>-Score*, potrebbero non cogliere completamente l'impatto del Max Pooling, specialmente su un intervallo di soglie diverse.

In contesti meteorologici, è importante comprendere se la riduzione della risoluzione influisca sulle capacità del modello di identificare aree critiche, aumentando la probabilità di rilevare correttamente fenomeni estremi. La riduzione effettuata mediante il Max Pooling aiuta a diminuire il rumore locale e a rafforzare i segnali più significativi, poichè l'operazione preserva i picchi rilevanti, come quelli associati a precipitazioni intense. In presenza di fenomeni estesi su larga scala, i dettagli più fini potrebbero risultare meno rilevanti.

Per eseguire questa valutazione, è stata utilizzata un'operazione di Max Pooling con un kernel di dimensioni (2 x 2), applicata sia alle osservazioni reali che alle previsioni. Questo processo riduce la risoluzione spaziale complessiva di un fattore 2, portando le immagini a una risoluzione finale di (48 x 48). I risultati sulla soglia di osservazione a 8 mm sono riportati nella tabella (5.1).

Ora	Senza maxpooling	Maxpooling	Differenze assolute	Miglioramento (%)
1	0.364333661	0.432077472	0.067743811	18.6%
2	0.070391526	0.104988432	0.034596906	49.1%
3	0.019452393	0.036449750	0.016987357	87.3%

Tabella 5.1: Prestazioni Average Precision con e senza maxpooling con kernel (2 x 2)

L'incremento percentuale nei risultati evidenzia un miglioramento significativo nell'*AP* per tutte e tre le ore esaminate, indicando una maggiore efficienza del modello nell'identificazione del target.

Se si utilizzassero kernel di dimensioni maggiori, la riduzione della risoluzione sarebbe più drastica rispetto a prima. Questo consentirebbe di diminuire ulteriormente il rumore, ma al costo di eliminare dettagli potenzialmente importanti, riducendo maggiormente la possibilità di riconoscimento di fenomeni su scala ridotta.

Prima di procedere con esperimenti che prevedono l'uso di kernel di dimensioni maggiori, è necessario valutare se la scala spaziale dei fenomeni meteorologici in esame possa tollerare una riduzione della risoluzione.

Per il modello analizzato, che lavora con immagini di dimensioni (96 x 96), l'applicazione di un kernel (3 x 3) e (4 x 4) porta ad una risoluzione finale delle immagini pari rispettivamente a (32 x 32) e (24 x 24). Un'altra informazione importante da considerare è lo stride applicato: se uguale alla dimensione del kernel, non si genera sovrapposizione tra i blocchi elaborati durante il Pooling, influenzando ulteriormente il mantenimento dei dettagli.

- *Average Precision* con kernel (3 x 3) e stride 2, 3: (B.2);
- *Average Precision* con kernel (4 x 4) e stride 3, 4: (B.3).

I risultati ottenuti con kernel più grandi potrebbero sembrare sorprendentemente elevati, ma ciò non toglie la loro plausibilità. Il Max Pooling potrebbe aver ridotto notevolmente il rumore, migliorando la capacità di riconoscimento dei pattern associati agli eventi estremi. È probabile che i dati di test includano prevalentemente fenomeni su larga scala, piuttosto che eventi altamente localizzati.

### 5.2.3 Esperimento 3

L'obiettivo dell'esperimento è valutare la dipendenza del modello dal numero di step impiegati durante il processo di generazione.

Il numero ottimale di passaggi non è fisso e può variare a seconda dei modelli.

In questo esperimento, si considera il numero di iterazioni da 2 a 16 step. Si utilizza la funzione di valutazione Ensemble (A.2). Nella tabella indicata sono riportati i risultati per l'errore *MSE* sul dataset (5.2).

Step	Esemble 1h	Esemble 2h	Esemble 3h
2	0.0001737	0.0002989	0.0004205
3	0.0001707	0.0003019	0.0004256
4	0.0001695	0.0003049	0.0004302
5	0.0001691	0.0003077	0.0004342
6	0.0001693	0.0003107	0.0004381
7	0.0001696	0.0003132	0.0004411
8	0.0001700	0.0003153	0.0004442
9	0.0001706	0.0003173	0.0004464
10	0.0001716	0.0003202	0.0004495
11	0.0001724	0.0003219	0.0004522
12	0.0001730	0.0003245	0.0004558
13	0.0001742	0.0003254	0.0004562
14	0.0001741	0.0003266	0.0004578
15	0.0001749	0.0003283	0.0004593
16	0.0001763	0.0003303	0.0004611

Tabella 5.2: MSE con valutazione tramite Ensemble

In base ai risultati riportati, le prestazioni del modello variano leggermente a seconda dell'orizzonte temporale della previsione. Determinare il numero ottimale di step dipende dalle priorità specifiche richieste, ma si può affermare che 5 step rappresenti il miglior compromesso tra le previsioni a breve (1 ora) e medio termine (2-3 ore) nel contesto di questo studio. Oltre i 5 step, i miglioramenti diventano trascurabili o si riscontra un leggero peggioramento delle performance. A questo punto, il modello sembra aver già estratto tutte le informazioni utili dai dati, e ulteriori iterazioni possono introdurre rumore.

L'esperimento è stato eseguito parallelamente sulla funzione Single Diffusion per ulteriori metriche, utilizzando una soglia di osservazione pari a 8 mm, per focalizzarsi sulle prestazioni per eventi estremi. La soglia di previsione è stata impostata sul valore ottimale ricavato nell'esperimento (5.2.1), in relazione a quella soglia di osservazione.

- $F_1$ -Score (B.5);

- $F_2$ -Score (B.6);
- *Average Precision* (B.7).

Coerentemente con le misurazioni ottenute con l'Ensemble, anche nel modello a Single Diffusion si osserva che il numero ottimale di step per la funzione di previsione è intorno a 5, punto in cui il miglioramento delle prestazioni si stabilizza.

Per la metrica *MSE*, risulta più accurata l'approccio Ensemble rispetto alla diffusione singola (B.4): il deterioramento delle prestazioni è più marcato nelle previsioni a 2 e 3 ore, indicando una minore capacità di catturare con precisione i fenomeni meteorologici.

### 5.2.4 Esperimento 4

L'obiettivo dell'esperimento è valutare la dipendenza delle prestazioni del modello in funzione della cardinalità dell'ensemble (A.2), cioè dal numero di modelli combinati per fare una previsione. L'idea è verificare se un numero maggiore di modelli porta a una riduzione dell'errore e a una maggiore robustezza nelle previsioni.

In particolare, analizzando la deviazione standard (*std*) dei risultati, si può valutare la loro variabilità: un ensemble più ampio, teoricamente, riduce la varianza delle previsioni e quindi produrre stime più stabili, meno soggette a fluttuazioni casuali. La deviazione standard misura infatti la dispersione dei valori rispetto alla loro media: una *std* bassa indica che i valori sono molto vicini alla media, mentre una *std* alta segnala una maggiore variabilità.

Nell'esperimento, l'output è calcolato utilizzando un ensemble di cardinalità pari a 32 e poi si valutano le prestazioni su sottoinsiemi 1, 2, 4, 8 e 16. Per ogni cardinalità, si riporta sia il valore dell'errore *MSE*, sia la *std* per garantire una stima robusta delle prestazioni.

Size	MSE 1h	MSE 2h	MSE 3h	Std 1h	Std 2h	Std 3h
1	0.0002202	0.0003823	0.0005166	0.0001779	0.0003228	0.0004367
2	0.0001946	0.0003492	0.0004807	0.0001619	0.0003026	0.0004156
4	0.0001811	0.0003324	0.0004626	0.0001526	0.0002917	0.0004051
8	0.0001747	0.0003242	0.0004537	0.0001484	0.0002865	0.0004000
16	0.0001715	0.0003200	0.0004493	0.0001463	0.0002840	0.0003972
32	0.0001699	0.0003179	0.0004469	0.0001453	0.0002827	0.0003960

Tabella 5.3: MSE e std per diverse cardinalità dell'ensemble

Dai valori riportati in tabella (5.3) si nota che:

- L'*MSE* si riduce all'aumentare della cardinalità dell'ensemble per ogni orizzonte temporale considerato. Utilizzare più modelli combinati migliora l'accuratezza delle previsioni;
- La *std* mostra una diminuzione generale all'aumentare della cardinalità dell'ensemble. Le previsioni diventano più stabili e meno soggette a variazioni casuali quando si usano più modelli combinati;

- Il miglioramento nell'*MSE* è più evidente tra le cardinalità più basse (1 - 2 e 4). Aumentando la cardinalità dell'ensemble, il miglioramento diventa sempre più marginale (16 - 32).

L'operazione con una cardinalità di 32 è risultata costosa, soprattutto in termini di tempo di esecuzione. Pertanto, potrebbe essere opportuno considerare come compromesso tra accuratezza e complessità una dimensione di 16 per l'ensemble.

Inoltre, l'uso di un numero dimezzato di batch si è rivelato vantaggioso per le prestazioni del modello. Nella tabella (B.8) sono riportati i valori ottenuti per l'*MSE* e la *std* con la configurazione a batch dimezzato.

In sintesi, l'esperimento evidenzia che un aumento della cardinalità dell'ensemble migliora l'accuratezza delle previsioni. Si ha una chiara correlazione tra maggiore cardinalità e una diminuzione dell'*MSE*. Tuttavia, incrementi oltre alla cardinalità di 16 sono per lo più marginali.

### 5.2.5 Esperimento 5

L'obiettivo dell'esperimento è calcolare l'*MSE* per la previsione di persistenza.

Questo tipo di previsione implica l'assunzione che il futuro non cambierà rispetto al passato recente, replicando quindi l'ultimo valore noto di pioggia in tutte le previsioni future. La previsione di persistenza è spesso utilizzata come baseline per valutare l'efficacia della previsione di modelli predittivi complessi. Se il modello sviluppato non riesce a superare le prestazioni di questa previsione di base, ciò potrebbe indicare che ha difficoltà nel catturare efficacemente i pattern nei dati.

A livello di codice, le modifiche apportate alla funzione di valutazione (A.1) comportano:

- Sostituire `tmp = model.generate2(np.copy(sample), 10)`, per eliminare l'uso del modello predittivo e basare la previsione esclusivamente sull'ultima osservazione disponibile;
- Inserire `tmp = np.repeat(hist[:, :, -4:-3], 3, axis=-1)`, cioè utilizzare l'ultima ora di pioggia osservata (-4) e ripeterla lungo l'asse temporale per prevedere le ore future (-3, -1, -1), come mostrato nella tabella (4.1).

Per i valori di persistenza si ottengono i seguenti di risultati di *AP* (2.7) e *MSE*:

- *AP*: [0.250278105, 0.047034982, 0.013906078]
- *MSE*: [0.0003617, 0.0007473, 0.00102699]

Ponendo una soglia di osservazione a 8 mm e soglie di previsione tra 6.5 mm e 7.5 mm, si confrontano i valori di *AP* e *MSE* ottenuti con la previsione di persistenza con quelli del modello di previsione condizionato. La tabella 5.4 mostra il miglioramento percentuale ottenuto per ogni coppia di valori.

In tutti i casi, l'*AP* del modello condizionato supera quello della persistenza soprattutto per le previsioni di 1 e 2 ore. I miglioramenti più significativi sono quelli per le soglie comprese tra 6.5 mm e 6.9 mm, ma l'*AP* rimane generalmente più alto rispetto alla persistenza anche per soglie superiori. Si ha una consistente riduzione anche dell'*MSE*

Soglia	AP 1h	AP 2h	AP 3h	MSE 1h	MSE 2h	MSE 3h
0.0065	+52.70%	+55.87%	+39.52%	-39.46%	-49.13%	-49.87%
0.0066	+48.65%	+57.50%	+72.27%	-39.54%	-49.00%	-49.77%
0.0067	+47.37%	+46.91%	+32.31%	-39.52%	-49.00%	-49.75%
0.0068	+42.38%	+49.69%	+39.35%	-39.35%	-49.11%	-49.85%
0.0069	+44.59%	+51.06%	+16.67%	-39.37%	-49.08%	-49.96%
0.0070	+46.08%	+38.23%	+26.91%	-39.39%	-49.10%	-49.85%
0.0071	+41.06%	+25.68%	+26.54%	-39.08%	-48.98%	-49.71%
0.0072	+44.33%	+50.65%	+26.56%	-39.19%	-49.05%	-49.85%
0.0073	+47.25%	+66.41%	+45.47%	-39.19%	-49.06%	-49.84%
0.0074	+41.68%	+67.06%	+94.42%	-39.17%	-49.02%	-49.84%
0.0075	+45.18%	+37.34%	+42.57%	-39.12%	-48.96%	-49.84%

Tabella 5.4: Miglioramento (%) delle previsioni di persistenza rispetto al condizionamento

rispetto alla persistenza per tutte le ore di previsioni. Il miglioramento è più marcato nelle previsioni per le ore successive. Questi valori di *MSE* risultano in generale stabili. In generale, il modello condizionato è superiore alla persistenza per queste metriche. La persistenza non è una buona strategia a lungo termine. Il modello condizionato cattura meglio le dinamiche delle piogge, con previsioni più accurate anche su orizzonti temporali più lunghi. Sembra quindi che le precipitazioni non siano statiche.

Lo stesso comportamento si osserva aggiungendo un'operazione di Max Pooling con kernel (2 x 2), come eseguito nell'esperimento (5.2.2): i miglioramenti in termini di *AP* sono inferiori rispetto ai risultati precedenti. Questo suggerisce che l'uso del Max Pooling non compromette drasticamente le prestazioni del modello, ma potrebbe far perdere alcuni dettagli importanti, riducendo leggermente l'accuratezza delle previsioni:

- *AP*: [0.335245374, 0.083635160, 0.027253892]
- *MSE*: [0.0003617, 0.0007473, 0.00102699]





# Capitolo 6

## Conclusioni

Questa tesi delinea gli sforzi compiuti per applicare tecniche di intelligenza artificiale nel dominio meteorologico, evidenziando sia l'importanza di affrontare lo squilibrio dei dati che la natura dinamica degli eventi estremi.

In particolare, è stato analizzato un modello di previsione delle precipitazioni che utilizza una tecnologia di diffusione del denoising basata su un'architettura U-Net. Il modello è progettato per fare previsioni a breve termine, mirate a stimare le condizioni delle prossime tre ore. L'addestramento ha integrato diversi elementi meteorologici come informazioni di condizionamento per fornire un contesto più completo per la generazione delle previsioni. I dati utilizzati corrispondono ad un sottoinsieme del set di *ERA5* su base oraria, relativi ad una regione d'Europa centrale. Il modello è stato addestrato sui dati degli anni 2016-2020 e testato con dati del 2015. I risultati di una serie di esperimenti specifici dimostrano che il modello studiato ottiene prestazioni accettabili per le previsioni a breve termine. Tuttavia, per soglie particolarmente critiche, le sue performance non sono ottimali, a causa di vari fattori. Nonostante ciò, i miglioramenti osservati in alcuni esperimenti suggeriscono che esista del margine per ulteriori ottimizzazioni, contribuendo a perfezionare significativamente le previsioni. Il modello, quindi, si pone come una base solida per ulteriori sviluppi.

### 6.1 Lavori futuri

In futuro, l'attenzione potrà focalizzarsi sul miglioramento delle capacità di modellare e comprendere le dinamiche complesse degli eventi meteorologici estremi. Un aspetto fondamentale sarà l'individuazione di variabili chiave capaci di migliorare le prestazioni dei modelli, considerando che piccoli aggiustamenti possano determinare cambiamenti significativi nei risultati. Per il modello analizzato, un possibile sviluppo è l'ampliamento della scala spaziale oltre le regioni specifiche dell'Europa, sfruttando le risorse del supercomputer *Leonardo*. Questo permetterà una comprensione più approfondita dei sistemi meteorologici. Inoltre, l'integrazione di variabili di condizionamento più complesse, provenienti anche da campi interdisciplinari, rappresenta un'area di ricerca promettente [4]. Si prevede un'ulteriore cambiamento nella natura delle previsioni e delle osservazioni, con una maggiore prevalenza di approcci ensemble e di dati satellitari, per gestire dati meteorologici legati agli eventi estremi. I modelli basati sull'intelligenza artificiale, sempre più leggeri e avanzati, continueranno a migliorare l'accuratezza e l'affidabilità di queste previsioni grazie a dati multicanale e modelli ottimizzati [17].

In conclusione, questo lavoro contribuisce al progresso delle metodologie di previsione delle precipitazioni e offre una promettente prospettiva per l'impiego di modelli di diffusione, aprendo nuove strade per una comprensione più approfondita dei fenomeni metereologici estremi.

# **Appendici**



# Appendice A

## Codici

Listing A.1: Funzione Single Diffusion

```
def SD(generator, n_iter, model, maxR):
    history = np.zeros((n_iter, 3))
    raw_data = np.zeros((n_iter, batch_size, 96, 96, 6))
    mses = np.zeros((3))

    for i in range(n_iter):
        sample = generator.__getitem__(i)
        hist = np.copy(sample)
        sample = model.normalizer(sample)
        tmp = model.generate2(np.copy(sample), 10)

        hist = hist * maxR
        tmp = tmp * maxR

        raw_data[i, :, :, :, :3] = hist[:, :, :, -3:]
        raw_data[i, :, :, :, 3:] = tmp[:, :, :, -3:]

        mse = np.mean(np.sum((hist[:, :, :, -3:] - tmp[:, :, :, -3:])**2,
            axis=(1,2)), axis=0)
        history[i] = mse
        mses += mse[-3:]

    test_dataset = generator.data * maxR
    true = np.reshape(raw_data[... , :3], (-1, 96, 96, 3))
    pred = np.reshape(raw_data[... , 3:], (-1, 96, 96, 3))

    for h in range(3):
        obs = np.copy(true[... , h]).flatten()
        preds = np.copy(pred[... , h]).flatten()
        assert obs.shape == preds.shape

        true_obs = (obs >= 0.008)
        true_preds = (preds >= 0.006)
```

```
[...]  
  
f1 = fb(true_obs, true_preds, beta=1)  
f05 = fb(true_obs, true_preds, beta=0.5)  
f2 = fb(true_obs, true_preds, beta=2)  
ap = aps(true_obs, preds)  
  
return mses / n_iter, history, raw_data
```

Listing A.2: Funzione Generative Ensemble Diffusion

```
def GED(generator, model, n_iter, ensemble_iter, maxR, tot_frames):  
    mses = np.zeros((n_iter, 3))  
    raw_data = np.zeros((n_iter, batch_size, 96, 96, 6))  
  
    for i in range(n_iter):  
        test = generator.__getitem__(i)  
  
        print(i)  
  
        res = np.zeros((batch_size, ensemble_iter, 96,  
                        96, tot_frames))  
        print(res.shape)  
        for j in range(ensemble_iter):  
  
            sample = np.copy(test)  
  
            sample = model.normalizer(sample)  
            tmp = model.generate2(np.copy(sample), 10)  
  
            tmp = tmp * maxR  
            res[:, j] = tmp  
  
        average = np.mean(res, axis=1)  
        hist = test * maxR  
  
        raw_data[i, :, :, :, :3] = np.copy(hist[:, :, :, -3:])  
        raw_data[i, :, :, :, 3:] = np.copy(average[:, :, :, -3:])  
  
        mse = np.mean(np.sum((hist[:, :, :, -3:] -  
                             average[:, :, :, -3:])**2,  
                             axis=(1,2)), axis=0)  
        mses[i] = mse  
  
return mses, raw_data
```

# Appendice B

## Risultati

Soglia di previsione (mm)	F2-Score 1h	F2-Score 2h	F2-Score 3h
6.5	0.470319054	0.061408061	0.008932559
6.6	0.460458193	0.060811398	0.015638963
6.7	0.459751201	0.049135737	0.004938493
6.8	0.445368569	0.049135737	0.006730683
6.9	0.451232802	0.044846993	0.002696387
7.0	0.445595499	0.040112845	0.006724648
7.1	0.428289659	0.030517470	0.004934948
7.2	0.426362089	0.036402379	0.002252658
7.3	0.426362089	0.036402379	0.002252658
7.4	0.398439778	0.027663752	0.008102998
7.5	0.403902798	0.018339596	0.002254893

Tabella B.1: F<sub>2</sub>-Score a soglia di osservazione 8 mm

Ora	Stride 2	Miglioramento (%)	Stride 3	Miglioramento (%)
1	0.463513763	27.2%	0.480296956	31.8%
2	0.140384787	99.4%	0.138806061	97.2%
3	0.052531155	170.0%	0.054430443	179.8%

Tabella B.2: Prestazioni Average Precision con maxpooling a kernel (3 x 3)

Ora	Stride 3	Miglioramento (%)	Stride 4	Miglioramento (%)
1	0.514188977	41.1%	0.485285441	33.2%
2	0.176110588	150.0%	0.149549709	112.5%
3	0.076365856	292.6%	0.061476458	216.0%

Tabella B.3: Prestazioni Average Precision con maxpooling a kernel (4 x 4)

Step	Single 1h	Single 2h	Single 3h
2	0.0001918	0.0003200	0.0004422
3	0.0001964	0.0003321	0.0004567
4	0.0002009	0.0003421	0.0004681
5	0.0002049	0.0003509	0.0004786
6	0.0002088	0.0003593	0.0004894
7	0.0002116	0.0003653	0.0004970
8	0.0002151	0.0003710	0.0005042
9	0.0002171	0.0003769	0.0005098
10	0.0002193	0.0003805	0.0005139
11	0.0002218	0.0003862	0.0005225
12	0.0002233	0.0003894	0.0005266
13	0.0002261	0.0003926	0.0005296
14	0.0002274	0.0003954	0.0005319
15	0.0002287	0.0003997	0.0005380
16	0.0002302	0.0004017	0.0005397

Tabella B.4: MSE con valutazione tramite Single Diffusion

Step	F1-Score 1h	F1-Score 2h	F1-Score 3h
2	0.365197481	0.058452341	0.005003574
3	0.377435162	0.063576159	0.002866356
4	0.389007942	0.054340623	0.002857143
5	0.395877944	0.067452521	0.014064698
6	0.404497501	0.065883888	0.018284107
7	0.415890570	0.069828722	0.007130125
8	0.411251504	0.065840939	0.009869581
9	0.407260525	0.069934641	0.008429926
10	0.415295408	0.066427874	0.004247788
11	0.410705751	0.058804312	0.004941758
12	0.412193935	0.066710268	0.010456605
13	0.398458445	0.065454545	0.009943182
14	0.407094595	0.050892267	0.007741027
15	0.401353638	0.053333333	0.005576856
16	0.391167192	0.062213491	0.004964539

Tabella B.5: F<sub>1</sub>-Score con valutazione tramite Single Diffusion



Step	F2-Score 1h	F2-Score 2h	F2-Score 3h
2	0.468126321	0.038896747	0.003153437
3	0.473662107	0.042399081	0.001803101
4	0.475237157	0.036222281	0.001801639
5	0.468839156	0.045354469	0.008972633
6	0.469848436	0.044426850	0.011664424
7	0.470902261	0.046749581	0.004502071
8	0.457221219	0.044419034	0.006284791
9	0.449283166	0.047090925	0.005382131
10	0.451232802	0.044846993	0.002696387
11	0.442104541	0.039605703	0.003143524
12	0.439245231	0.044898319	0.006714414
13	0.416637466	0.043704750	0.006296663
14	0.423669222	0.033989582	0.004935834
15	0.417282387	0.036040788	0.003581021
16	0.398372242	0.041831792	0.003147199

Tabella B.6: F<sub>2</sub>-Score con valutazione tramite Single Diffusion

Step	Average Precision 1h	Average Precision 2h	Average Precision 3h
2	0.362439342	0.077668097	0.030399638
3	0.369732443	0.080496614	0.025485855
4	0.372341397	0.073833552	0.022096690
5	0.362674336	0.074371008	0.026923779
6	0.367850928	0.073206546	0.030936365
7	0.377818747	0.079210018	0.025403482
8	0.367379231	0.070118963	0.022334254
9	0.363222345	0.069422856	0.019822503
10	0.361769707	0.071054491	0.016223161
11	0.361066936	0.062385070	0.015568950
12	0.360247206	0.068039359	0.015227220
13	0.345169493	0.067557897	0.019086657
14	0.355538713	0.060491754	0.017094429
15	0.338535812	0.050272688	0.010476403
16	0.337977426	0.067590247	0.017960063

Tabella B.7: Average Precision con valutazione tramite Single Diffusion

Size	MSE 1h	MSE 2h	MSE 3h	Std 1h	Std 2h	Std 3h
1	0.0001606	0.0002632	0.0003529	0.0001161	0.0001845	0.0002459
2	0.0001392	0.0002358	0.0003234	0.0001028	0.0001676	0.0002289
4	0.0001287	0.0002225	0.0003082	9.63550e-05	1.59918e-04	2.19521e-04
8	0.0001233	0.0002155	0.0003004	9.32900e-05	1.55797e-04	2.14965e-04
16	0.0001207	0.0002122	0.0002965	9.14810e-05	1.53856e-04	2.12501e-04
32	0.0001193	0.0002103	0.0002946	9.06031e-05	1.52537e-04	2.11108e-04

Tabella B.8: MSE e std per diverse cardinalità dell'ensemble con batch dimezzati



# Elenco delle figure

3.1	Esempio di struttura di una rete CNN . . . . .	25
3.2	Esempio di un'architettura di rete U-Net . . . . .	27
4.1	Processo di diffusione in avanti e inverso . . . . .	30
4.2	Esempio di dati sulle precipitazioni dal set ERA-5 . . . . .	33
4.3	Esempio delle funzionalità aggiuntive . . . . .	34
5.1	Andamento $F_1$ -Score in funzione delle ore di previsione . . . . .	41
5.2	Andamento $F_2$ -Score in funzione delle ore di previsione . . . . .	41
5.3	Andamento $F_1$ -Score in funzione dei livelli di pioggia . . . . .	42
5.4	Andamento $F_2$ -Score in funzione dei livelli di pioggia . . . . .	42



# Bibliografia

- [1] D. S. Wilks, *Statistical methods in the atmospheric sciences*. Academic press, 2011.
- [2] Q. Jin, X. Zhang, X. Xiao, Y. Wang, S. Xiang, and C. Pan, “Preformer: Simple and efficient design for precipitation nowcasting with transformers,” *IEEE Geosci. Remote. Sens. Lett.*, vol. 21, pp. 1–5, 2024.
- [3] Q. Jin, X. Zhang, X. Xiao, Y. Wang, G. Meng, S. Xiang, and C. Pan, “Spatiotemporal inference network for precipitation nowcasting with multimodal fusion,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.*, vol. 17, pp. 1299–1314, 2024.
- [4] I. T. Jolliffe and D. B. Stephenson, *Forecast verification: a practitioner’s guide in atmospheric science*. John Wiley & Sons, 2012.
- [5] S. Verma, K. Srivastava, A. Tiwari, and S. Verma, “Deep learning techniques in extreme weather events: A review,” *arXiv preprint arXiv:2308.10995*, 2023.
- [6] X. Ling, C. Li, P. Yang, Y. Huang, and F. Qin, “Tu2net-gan: A temporal precipitation nowcasting model with multiple decoding modules,” *Pattern Recognition Letters*, vol. 178, pp. 98–105, 2024.
- [7] W. Kim, C.-H. Jeong, and S. Kim, “Improvements in deep learning-based precipitation nowcasting using major atmospheric factors with radar rain rate,” *Computers & Geosciences*, vol. 184, p. 105529, 2024.
- [8] C. Meo, A. Roy, M. Lica, J. Yin, Z. B. Che, Y. Wang, R. Imhoff, R. Uijlenhoet, and J. Dauwels, “Extreme precipitation nowcasting using transformer-based generative models,” *CoRR*, vol. abs/2403.03929, 2024.
- [9] G. Brunet, D. B. Parsons, D. Ivanov, B. Lee, P. Bauer, N. B. Bernier, V. Bouchet, A. Brown, A. Busalacchi, G. C. Flatter, R. Goffier, P. Davies, B. Ebert, K. Gutbrod, S. Hong, P. K. Kenabatho, H.-J. Koppert, D. Lesolle, A. H. Lynch, J.-F. Mahfouf, L. Ogallo, T. Palmer, K. Petty, D. Schulze, T. G. Shepherd, T. F. Stocker, A. Thorpe, and R. Yu, “Advancing weather and climate forecasting for our changing world,” *Bulletin of the American Meteorological Society*, vol. 104, no. 4, pp. E909 – E927, 2023.
- [10] A. Amin and M. Mourshed, “Weather and climate data for energy applications,” *Renewable and Sustainable Energy Reviews*, vol. 192, p. 114247, 2024.
- [11] B. Bochenek and Z. Ustrnul, “Machine learning in weather prediction and climate analyses—applications and perspectives,” *Atmosphere*, vol. 13, no. 2, p. 180, 2022.

- [12] M. Panteli and P. Mancarella, "Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies," *Electric Power Systems Research*, vol. 127, pp. 259–270, 2015.
- [13] T. P. Agyekum, P. Antwi-Agyei, and A. J. Dougill, "The contribution of weather forecast information to agriculture, water, and energy sectors in east and west africa: A systematic review," *Frontiers in Environmental Science*, vol. 10, p. 935696, 2022.
- [14] W. Li, Y. Zhou, Y. Li, D. Song, Z. Wei, and A. Liu, "Hierarchical transformer with lightweight attention for radar-based precipitation nowcasting," *IEEE Geosci. Remote. Sens. Lett.*, vol. 21, pp. 1–5, 2024.
- [15] C. Zheng, Y. Tao, J. Zhang, L. Xun, T. Li, and Q. Yan, "Tise-lstm: A lstm model for precipitation nowcasting with temporal interactions and spatial extract blocks," *Neurocomputing*, vol. 590, p. 127700, 2024.
- [16] F. Naz, L. She, M. Sinan, and J. Shao, "Enhancing radar echo extrapolation by convlstm2d for precipitation nowcasting," *Sensors*, vol. 24, no. 2, p. 459, 2024.
- [17] N. Yang and X. Li, "Lightweight ai-powered precipitation nowcasting," *The Innovation Geoscience*, vol. 2, no. 2, pp. 100066–1, 2024.
- [18] A. H. Murphy, "What is a good forecast? an essay on the nature of goodness in weather forecasting," *Weather and Forecasting*, vol. 8, no. 2, pp. 281 – 293, 1993.
- [19] J. T. Schaefer, "The critical success index as an indicator of warning skill," *Weather and Forecasting*, vol. 5, no. 4, pp. 570 – 575, 1990.
- [20] R. J. Hogan, C. A. T. Ferro, I. T. Jolliffe, and D. B. Stephenson, "Equitability revisited: Why the "equitable threat score" is not equitable," *Weather and Forecasting*, vol. 25, no. 2, pp. 710 – 726, 2010.
- [21] J. Hu, B. Yin, and C. Guo, "Meteo-dlnet: Quantitative precipitation nowcasting net based on meteorological features and deep learning," *Remote. Sens.*, vol. 16, no. 6, p. 1063, 2024.
- [22] J. Gong, L. Bai, P. Ye, W. Xu, N. Liu, J. Dai, X. Yang, and W. Ouyang, "Cascast: Skillful high-resolution precipitation nowcasting via cascaded modelling," *CoRR*, vol. abs/2402.04290, 2024.
- [23] W. Li, X. Zhou, and X. Mou, "A spatial-temporal dual discriminator precipitation nowcasting method using sngan with inceptiongru," *IEEE Access*, 2024.
- [24] M. G. Schultz, C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L. H. Leufen, A. Mozaffari, and S. Stadler, "Can deep learning beat numerical weather prediction?," *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, p. 20200097, 2021.
- [25] A. Asperti, F. Merizzi, A. Paparella, G. Pedrazzi, M. Angelinelli, and S. Colamona-co, "Precipitation nowcasting with generative diffusion models," *arXiv preprint arXiv:2308.06733*, 2023.

- 
- [26] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut, “The era5 global reanalysis,” *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.
- [27] M.-H. Yen, D.-W. Liu, Y.-C. Hsin, C.-E. Lin, and C.-C. Chen, “Application of the deep learning for the prediction of rainfall in southern taiwan,” *Scientific Reports*, vol. 9, p. 12774, Sep 2019.
- [28] B. Bochenek and Z. Ustrnul, “Machine learning in weather prediction and climate analyses—applications and perspectives,” *Atmosphere*, vol. 13, no. 2, 2022.
- [29] X. Jiang, J. Chen, X. Chen, W. kin Wong, M. Wang, and S. Wang, “Comparative study of cloud evolution for rainfall nowcasting using ai-based deep learning algorithms,” *Journal of Hydrology*, vol. 639, p. 131593, 2024.
- [30] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, “Striving for simplicity: The all convolutional net,” *arXiv preprint arXiv:1412.6806*, 2014.
- [31] B. Pan, K. Hsu, A. AghaKouchak, and S. Sorooshian, “Improving precipitation estimation using convolutional neural network,” *Water Resources Research*, vol. 55, no. 3, pp. 2301–2321, 2019.
- [32] A. Perera, V. Nik, D. Chen, J.-L. Scartezini, and T. Hong, “Quantifying the impacts of climate change and extreme climate events on energy systems,” *Nature Energy*, 02 2020.
- [33] Z. Pan, F. Ji, Y. Zhou, R. Hang, Q. Liu, and X.-T. Yuan, “Deep precipitation nowcasting with dual regions displacement information and global spatiotemporal representations learning,” *IEEE Geoscience and Remote Sensing Letters*, 2024.
- [34] Z. Ma, H. Zhang, and J. Liu, “DB-RNN: an RNN for precipitation nowcasting deblurring,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.*, vol. 17, pp. 5026–5041, 2024.
- [35] Y. Zhou, R. Hang, F. Ji, Z. Pan, Q. Liu, and X. Yuan, “Spatiotemporal enhanced adversarial network for precipitation nowcasting,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.*, vol. 17, pp. 7608–7620, 2024.
- [36] L. Vatamany and S. Mehrkanoon, “GD-CAF: graph dual-stream convolutional attention fusion for precipitation nowcasting,” *CoRR*, vol. abs/2401.07958, 2024.
- [37] E. Reulen and S. Mehrkanoon, “Ga-smaat-gnet: Generative adversarial small attention gnet for extreme precipitation nowcasting,” *CoRR*, vol. abs/2401.09881, 2024.
- [38] S. Yang, Y. Chen, L. Wang, S. Liu, and Y. Chen, “Denoising diffusion step-aware models,” *arXiv preprint arXiv:2310.03337*, 2023.