

L'uso dell'IoT per la monitoraggio e la predizione della qualità dell'aria

Arlind Pecmarkaj^{1*}, Emanuele Lattanzi²

Sommario

La qualità dell'aria rappresenta una problematica di crescente rilevanza, particolarmente nelle aree urbane e industriali italiane, dove l'esposizione prolungata agli inquinanti atmosferici ha un impatto significativo sulla salute pubblica. Questo progetto, sviluppato nell'ambito dell'Internet of Things (IoT), propone un sistema innovativo per il monitoraggio e la predizione della qualità dell'aria attraverso l'applicazione di tecniche di machine learning. È stata realizzata una stazione meteorologica IoT in grado di rilevare TVOC, particolato (PM), CO₂, umidità relativa, pressione atmosferica e temperatura. I dati raccolti nel periodo dal 9 giugno al 29 giugno 2025 sono stati utilizzati per analizzare le correlazioni esistenti tra condizioni ambientali e livelli di inquinamento, con particolare focus sulla possibilità di stimare la qualità dell'aria utilizzando esclusivamente parametri meteorologici a basso consumo energetico. L'obiettivo è quello di ridurre la complessità e i costi dei sistemi di monitoraggio tradizionali, offrendo una soluzione scalabile e a basso consumo energetico, particolarmente adatta per l'implementazione in reti di sensori urbani distribuiti.

Keywords

Air Quality — Meteorology — Internet of Things — Machine Learning — ESP32

¹ Laurea Magistrale in Informatica Applicata, Università degli Studi di Urbino Carlo Bo, Urbino, Italia

² Docente di Programmazione per l'Internet of Things, Università degli Studi di Urbino Carlo Bo, Urbino, Italia

*Corresponding author: a.pecmarkaj@campus.uniurb.it

Introduzione

L'inquinamento atmosferico costituisce una delle sfide ambientali più pressanti del nostro tempo, con effetti diretti sulla qualità della vita e sulla salute pubblica, particolarmente evidenti in aree critiche come la Pianura Padana. I sistemi tradizionali di monitoraggio della qualità dell'aria, pur garantendo elevata precisione nelle misurazioni, presentano costi significativi e una distribuzione geografica limitata che ne restringe l'utilizzo su larga scala.

L'Internet of Things (IoT) offre un approccio alternativo attraverso l'impiego di sensori compatti ed economici, capaci di creare reti di monitoraggio capillari e distribuite sul territorio. Tuttavia, i sensori dedicati alla rilevazione di inquinanti specifici come particolato (PM), composti organici volatili totali (TVOC) e anidride carbonica presentano consumi energetici elevati che ne limitano l'impiego in dispositivi alimentati a batteria o in applicazioni domestiche a lungo termine.

Il presente progetto esplora una strategia innovativa basata sull'utilizzo del machine learning per la predizione della qualità dell'aria, utilizzando esclusivamente dati meteorologici a basso consumo energetico quali temperatura, umidità relativa e pressione atmosferica. Come verrà illustrato nella sezione seguente, queste variabili meteorologiche influenzano significativamente i meccanismi di dispersione e accumulo

degli inquinanti atmosferici.

Sebbene la correlazione tra parametri meteorologici e qualità dell'aria non sia sempre diretta, un modello predittivo accurato può fornire stime sufficientemente affidabili per numerose applicazioni pratiche. L'obiettivo finale è lo sviluppo di dispositivi economici e a basso consumo energetico da integrare in sistemi domotici intelligenti, in grado di gestire automaticamente purificatori d'aria o sistemi di ventilazione sulla base di predizioni della qualità dell'aria.

1. Relazione tra parametri meteorologici e qualità dell'aria

La qualità dell'aria non dipende esclusivamente dalle emissioni inquinanti dirette (traffico veicolare, attività industriali), ma è fortemente influenzata dalle condizioni meteorologiche che ne regolano i processi di dispersione, diluizione e accumulo nell'atmosfera. Variabili come temperatura, umidità relativa e pressione atmosferica, pur non costituendo direttamente degli inquinanti, rappresentano indicatori fondamentali delle dinamiche atmosferiche che governano il comportamento degli inquinanti.

La **temperatura** influisce direttamente sulla stabilità dell'atmosfera e sui processi di convezione. Elevate temperature accompagnate da condizioni di calma di vento favoriscono il ristagno degli inquinanti negli strati bassi dell'atmosfera,

mentre variazioni termiche significative possono indicare fenomeni di rimescolamento atmosferico che favoriscono la dispersione degli inquinanti. Durante i mesi estivi, le alte temperature promuovono inoltre reazioni fotochimiche che portano alla formazione di inquinanti secondari come l'ozono troposferico.

L'**umidità relativa** gioca un ruolo determinante nella dinamica del particolato atmosferico. In condizioni di elevata umidità, le particelle di aerosol assorbono vapore acqueo dall'atmosfera, aumentando le proprie dimensioni e persistendo più a lungo in sospensione. Le variazioni di umidità relativa possono inoltre segnalare il passaggio di sistemi frontali che alterano significativamente la capacità dispersiva dell'atmosfera. Ad esempio, una diminuzione rapida dell'umidità relativa spesso coincide con l'arrivo di masse d'aria più secche e ventilate che favoriscono la dispersione degli inquinanti.

La **pressione atmosferica** costituisce un eccellente indicatore dei processi meteorologici su scala sinottica. I sistemi di alta pressione (anticicloni) sono tipicamente associati a condizioni di stabilità atmosferica, scarsa ventilazione e conseguente accumulo di inquinanti. Al contrario, i sistemi di bassa pressione (cicloni) portano instabilità meteorologica, venti sostenuti e precipitazioni che contribuiscono alla rimozione degli inquinanti dall'atmosfera attraverso processi di deposizione umida. Un calo repentino di pressione atmosferica rappresenta spesso un chiaro indicatore dell'approssimarsi di un fronte freddo, che tipicamente determina un rapido miglioramento delle condizioni di qualità dell'aria.

L'analisi delle tendenze temporali di temperatura, umidità relativa e pressione atmosferica può quindi fornire informazioni preziose sull'evoluzione futura della qualità dell'aria. Quando elaborate attraverso algoritmi di machine learning, queste variabili meteorologiche possono costituire una valida alternativa alla misurazione diretta degli inquinanti, offrendo un approccio più sostenibile dal punto di vista energetico ed economico.

2. Progettazione della stazione meteorologica e metodologia operativa

Il sistema sviluppato si basa su un microcontrollore **ESP32** che costituisce il nucleo centrale della stazione meteorologica. Il firmware implementato gestisce l'acquisizione coordinata dei dati provenienti da diversi sensori e la loro trasmissione verso un database remoto per l'archiviazione e l'analisi.

2.1 Componenti hardware: sensori e attuatori

La stazione meteorologica integra i seguenti sensori, interfacciati con l'ESP32 attraverso protocolli di comunicazione I2C e UART:

- **MH-Z19B** (interfaccia UART): Sensore a infrarossi non dispersivo (NDIR) per la rilevazione della concentrazione di anidride carbonica (CO₂) e temperatura interna del sensore.

- **AHT20** (interfaccia I2C): Sensore digitale ad alta precisione per la misurazione di temperatura e umidità relativa, utilizzato come riferimento primario e per la compensazione termica di altri sensori.
- **BMP280** (interfaccia I2C): Sensore barometrico per la misurazione della pressione atmosferica in hPa e temperatura ausiliaria.
- **ENS160** (interfaccia I2C): Sensore a ossido metallico (MOX) per la stima dell'Indice di Qualità dell'Aria (AQI), della CO₂ equivalente (eCO₂) e dei composti organici volatili totali (TVOC).
- **SPS30** (interfaccia I2C): Sensore laser per la misurazione delle concentrazioni di massa del particolato atmosferico nelle frazioni PM1.0, PM2.5, PM4.0 e PM10.0.

Come elemento di **attuazione**, il sistema incorpora un LED di stato che fornisce un feedback visivo immediato in caso di rilevazione di condizioni di qualità dell'aria critiche.

2.2 Firmware: logica di acquisizione e processamento dati

Il firmware è stato progettato seguendo principi di robustezza ed efficienza energetica per garantire un funzionamento continuo e affidabile.

2.2.1 Fase di inizializzazione del sistema

All'avvio del dispositivo, il firmware esegue una sequenza di inizializzazione che comprende:

1. **Connessione alla rete Wi-Fi** e sincronizzazione dell'orologio interno attraverso protocollo NTP, seguita dalla verifica della connettività verso il server InfluxDB di destinazione.
2. **Inizializzazione sequenziale dei sensori** con verifica della loro presenza e funzionalità sui rispettivi bus di comunicazione. In caso di mancata risposta di un sensore, il sistema continua il funzionamento utilizzando esclusivamente i dispositivi attivi.
3. **Periodo di stabilizzazione** di 60 secondi per consentire ai sensori, in particolare quelli basati su tecnologie MOX (ENS160) e NDIR (MH-Z19B), di raggiungere condizioni operative stabili.

2.2.2 Ciclo operativo principale

Il sistema entra successivamente in un ciclo operativo continuo con frequenza di campionamento di 5 minuti:

1. **Acquisizione coordinata dei dati:** Il firmware interroga sequenzialmente tutti i sensori attivi acquisendo le rispettive misurazioni.
2. **Elaborazione e validazione:** I dati grezzi vengono sottoposti a processamento per migliorarne l'affidabilità. Le misurazioni di temperatura e umidità relativa

dell'AHT20 vengono utilizzate per la compensazione ambientale delle letture dell'ENS160. Viene inoltre calcolata una temperatura media aggregando i dati provenienti dai diversi sensori, mentre le letture considerate anomale vengono automaticamente scartate.

3. **Attuazione e trasmissione dati:** La fase finale del ciclo comprende:

- **Sistema di allerta:** Il LED di stato viene attivato quando l'Indice di Qualità dell'Aria fornito dall'ENS160 raggiunge o supera il valore 4, oppure quando la concentrazione di PM2.5 supera la soglia di $25 \mu\text{g}/\text{m}^3$ o quella dei PM10 superi i $10 \mu\text{g}/\text{m}^3$. Questi limiti sono stati scelti prendendo come riferimento il D.Lgs.155/2010.
- **Invio verso InfluxDB:** In presenza di connettività di rete, i dati validati vengono trasmessi al database per l'archiviazione e l'analisi successiva.

2.3 Evoluzione verso l'attuazione basata su machine learning

L'attuale sistema di allerta, basato su soglie fisse predefinite, costituisce una soluzione di feedback immediato. L'obiettivo a lungo termine prevede il superamento di questo approccio attraverso l'implementazione di un sistema più sofisticato. La raccolta sistematica dei dati è finalizzata all'addestramento di un modello di machine learning in grado di identificare correlazioni significative tra i parametri meteorologici a basso consumo energetico (temperatura, umidità relativa, pressione atmosferica) e le concentrazioni degli inquinanti atmosferici (PM, TVOC, CO_2).

Nel caso in cui le correlazioni identificate risultino statisticamente significative, l'attuazione non dipenderà più dalle misurazioni dirette degli inquinanti, ma sarà guidata dalle predizioni del modello. Questo approccio consentirebbe lo sviluppo di dispositivi di monitoraggio e attuazione caratterizzati da costi e consumi energetici drasticamente ridotti, capaci di stimare la qualità dell'aria e attivare sistemi di purificazione in modalità predittiva.

3. Analisi dei dati sperimentali

Il periodo di rilevazione, compreso tra il 9 giugno e il 29 giugno 2025, è stato caratterizzato da condizioni meteorologiche relativamente stabili dominate dalla presenza di un sistema anticiclonico di origine africana, con temperature e pressione atmosferica che hanno mantenuto valori sostanzialmente costanti.

L'andamento della temperatura (Figura 1) mostra una sostanziale stabilità termica interrotta solamente dal passaggio di una perturbazione che ha causato un temporaneo abbassamento delle temperature. La distribuzione bimodale evidenziata in Figura 2 riflette chiaramente l'escursione termica giornaliera caratteristica del periodo, con due picchi distinti corrispondenti alle temperature minime notturne e alle massime diurne.

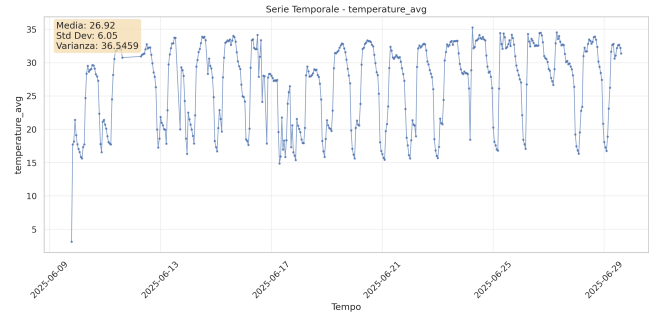


Figura 1. Andamento temporale della temperatura durante il periodo di osservazione

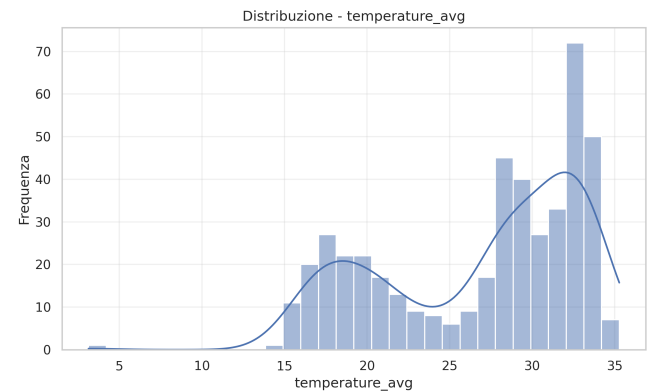


Figura 2. Distribuzione di frequenza della temperatura

La pressione atmosferica (Figura 3) ha mantenuto valori elevati e sostanzialmente stabili durante tutto il periodo di osservazione, con una lieve tendenza alla diminuzione registrata verso la fine del periodo di monitoraggio.

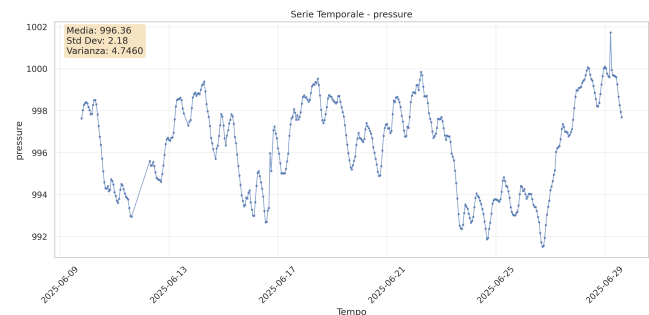


Figura 3. Andamento temporale della pressione atmosferica

L'umidità relativa (Figura 5) ha presentato alcune anomalie significative, con letture prossime allo 0% in condizioni ambientali caratterizzate da bassa umidità, probabilmente attribuibili a un malfunzionamento del sensore. Per evitare che tali anomalie compromettessero l'addestramento del modello predittivo, e per rendere quest'ultimo general-purpose, è stato imposto un valore minimo del 20% per l'umidità relativa. Va considerato che per via della posizione geografica dell'Italia, è estremamente raro trovare tassi di umidità inferiori al 20% per via della natura peninsulare del paese e del suo bacino

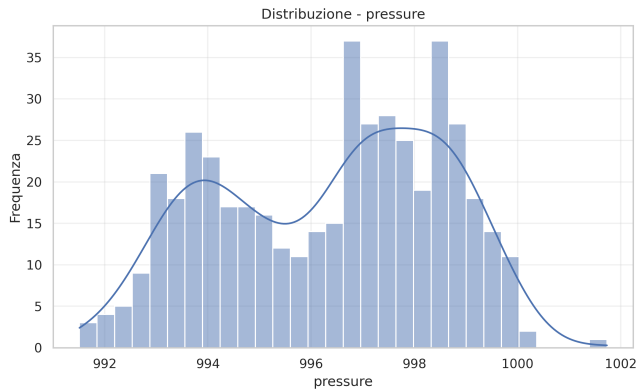


Figura 4. Distribuzione di frequenza della pressione atmosferica

idrografico.

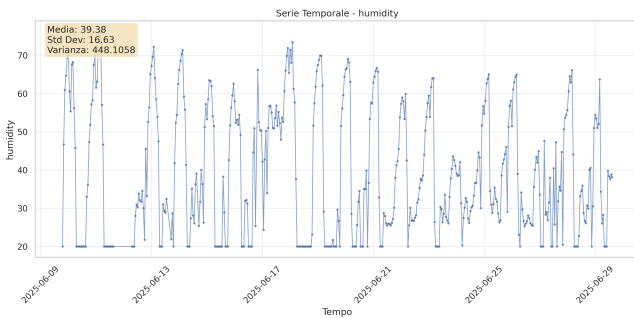


Figura 5. Andamento temporale dell'umidità relativa

Le anomalie registrate nelle misurazioni di umidità relativa hanno influenzato significativamente le stime di CO₂ equivalente (eCO₂) fornite dal sensore ENS160, che utilizza tali dati per la compensazione ambientale, producendo oscillazioni irrealistiche nei valori misurati (Figura 6).

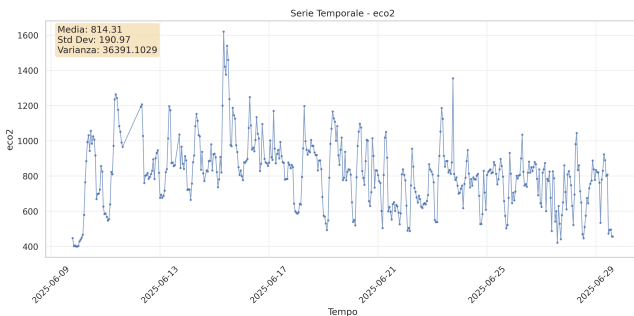


Figura 6. Andamento temporale della CO₂ equivalente stimata dal sensore ENS160

Al contrario, le misurazioni dirette di CO₂ effettuate dal sensore MH-Z19B (Figure 7 e 8) hanno mostrato risultati coerenti e affidabili, con valori compresi nella maggior parte dei casi tra 400 e 500 ppm, in linea con le concentrazioni tipiche di ambienti esterni.

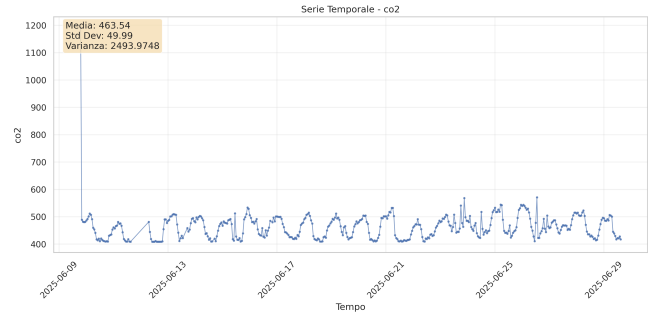


Figura 7. Andamento temporale della CO₂ misurata dal sensore MH-Z19B

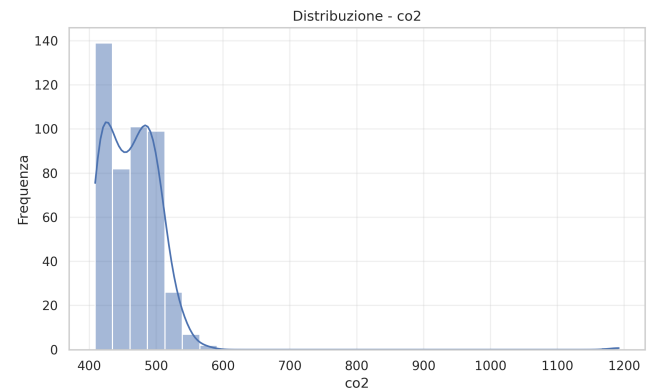


Figura 8. Distribuzione di frequenza della CO₂ misurata dal sensore MH-Z19B

Le concentrazioni di particolato (Figure 9 e 10) mostrano diversi picchi attribuibili a eventi locali, principalmente episodi di combustione di materiale vegetale e sterpaglie. Un picco particolarmente significativo registrato il 15 giugno 2025 è stato attribuito al trasporto a lunga distanza di inquinanti provenienti dagli incendi boschivi in corso in Canada durante quel periodo.

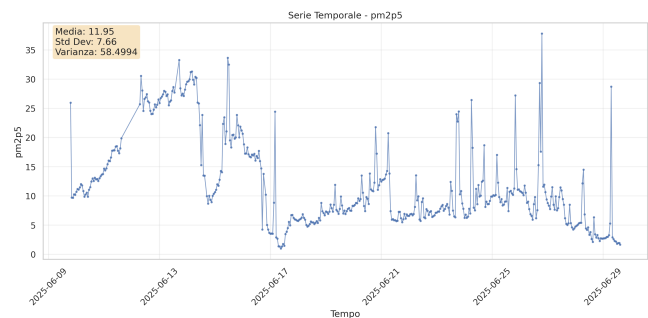


Figura 9. Andamento temporale del particolato PM2.5

I composti organici volatili totali (Figura 11) presentano incrementi significativi in corrispondenza degli stessi eventi che hanno causato i picchi di particolato. A differenza della CO₂ equivalente, questa misurazione non è soggetta a compensazione basata sull'umidità relativa e risulta in teoria quindi più robusta e affidabile.

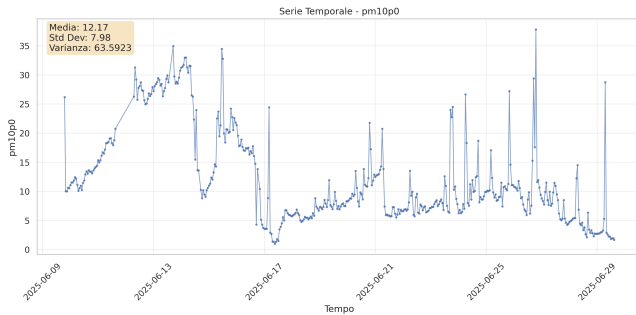


Figura 10. Andamento temporale del particolato PM10

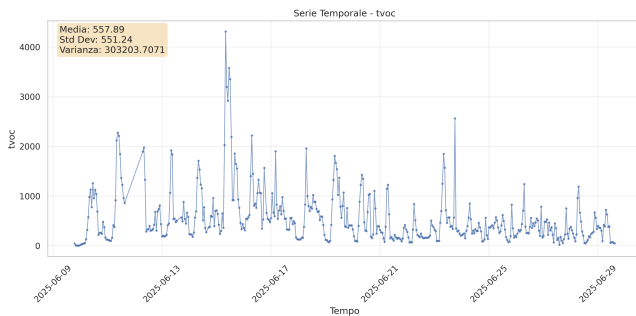


Figura 11. Andamento temporale dei composti organici volatili totali (TVOC)

4. Sviluppo del modello predittivo

Per analizzare quantitativamente la relazione esistente tra i parametri meteorologici e le concentrazioni degli inquinanti atmosferici, è stato sviluppato e addestrato un modello predittivo basato su tecniche di machine learning. La scelta progettuale è ricaduta su una rete neurale convoluzionale (CNN) addestrata utilizzando un approccio basato su finestre temporali scorrevoli.

4.1 Architettura della rete

La progettazione del modello predittivo si basa su scelte metodologiche specifiche per affrontare efficacemente il problema della predizione di serie temporali multivariate.

4.1.1 Scelta della rete convoluzionale

Sebbene le reti neurali ricorrenti (RNN, LSTM) rappresentino l'approccio standard per l'analisi di serie temporali, le reti neurali convoluzionali costituiscono un'alternativa computazionalmente efficiente e potente per questo tipo di applicazioni. I layer di convoluzione monodimensionale (Conv1D) operano come filtri specializzati nel riconoscimento di pattern locali all'interno delle sequenze temporali dei dati. L'architettura scelta permette al modello di costruire una comprensione gerarchica dei dati, partendo dal riconoscimento di pattern semplici per arrivare a strutture più complesse, mantenendo al contempo un'efficienza computazionale superiore rispetto agli approcci ricorrenti.

4.1.2 Definizione della finestra temporale

Il modello utilizza una finestra di input di 3 ore (corrispondente a 36 campioni con frequenza di 5 minuti) e un orizzonte di predizione di 1 ora (12 campioni futuri). La scelta della finestra di input di 3 ore rappresenta un compromesso ottimale per catturare le dinamiche meteorologiche a breve termine senza appesantire eccessivamente il modello con sequenze temporali troppo estese. L'orizzonte predittivo di 1 ora fornisce un preavviso temporale adeguato per applicazioni pratiche mantenendo un livello di accuratezza accettabile, considerando che predizioni a lungo termine risulterebbero meno affidabili.

4.1.3 Architettura della rete neurale

La rete neurale implementata presenta la seguente struttura:

- **Layer di input:** Accetta sequenze di forma (36, 3), rappresentanti 36 istanti temporali per 3 variabili meteorologiche (temperatura, pressione atmosferica, umidità relativa).
- **Strati convoluzionali:** Due layer Conv1D con 32 filtri ciascuno e kernel di dimensioni diverse (5 e 3) per l'analisi di pattern temporali di diversa lunghezza, seguiti da un layer di MaxPooling per la riduzione dimensionale e l'estrazione delle caratteristiche più significative. Un ulteriore layer Conv1D con 16 filtri completa la fase di estrazione delle caratteristiche.
- **Strati completamente connessi:** Due layer Dense con 50 e 25 neuroni rispettivamente, che combinano le caratteristiche estratte per formulare le predizioni finali.
- **Layer di output:** Produce le predizioni per le 5 variabili target nei 12 istanti temporali futuri, successivamente rimodellate nella forma (12, 5).

Il modello è stato addestrato per 100 epoche utilizzando l'ottimizzatore Adam e la funzione di perdita Mean Squared Error (MSE).

4.2 Risultati sperimentali e valutazione delle performance

I risultati ottenuti sul set di test, riassunti nella Tabella 1, evidenziano performance eterogenee del modello predittivo. Il sistema mostra una capacità predittiva limitata esclusivamente per la CO₂, mentre risulta inefficace per tutte le altre variabili target.

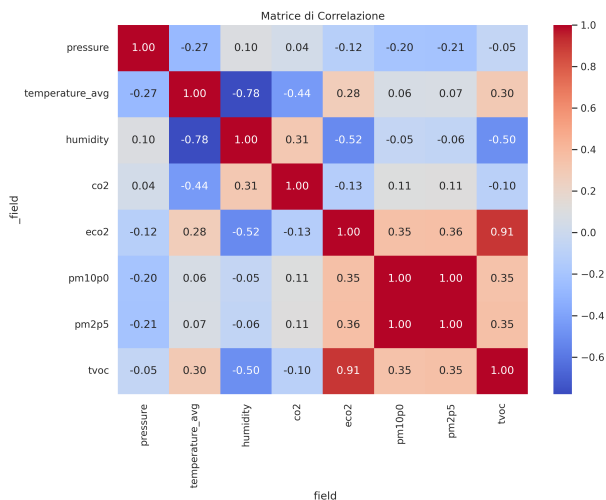
Il coefficiente di determinazione R² rappresenta la metrica chiave per la valutazione delle performance: valori positivi indicano una capacità predittiva del modello, mentre valori negativi (ottenuti per la maggior parte delle variabili) indicano che le predizioni del modello sono meno accurate di una semplice previsione basata sulla media storica dei dati, evidenziando la completa inefficacia del modello per quelle variabili.

Tabella 1. Risultati della valutazione del modello sul set di test.

Variabile Target	MSE	R ²	Risultato
CO ₂	1041.55	0.2218	Capacità predittiva limitata.
eCO ₂	45376.57	-1.5513	Assenza di capacità predittiva. Performance inferiore alla media storica.
PM10.0	83.44	-1.5139	Assenza di capacità predittiva. Performance inferiore alla media storica.
PM2.5	69.87	-1.1067	Assenza di capacità predittiva. Performance inferiore alla media storica.
TVOC	223494.74	-3.0390	Assenza di capacità predittiva. Performance inferiore alla media storica.

4.3 Analisi delle correlazioni tra variabili

La matrice di correlazione (Figura 12) conferma la debolezza delle relazioni lineari esistenti tra le variabili meteorologiche e gli inquinanti atmosferici, fornendo una spiegazione parziale delle scarse performance ottenute dal modello predittivo.

**Figura 12.** Matrice di correlazione tra le variabili meteorologiche e gli inquinanti atmosferici

Si nota innanzitutto che la pressione per via della sua stabilità, non correla con nessun'altra variabile. La CO₂ ha una correlazione leggera con la temperatura. Oltre a quello si nota che non esiste una correlazione definita tra gli inquinanti e le variabili atmosferiche in questo periodo.

4.4 Limitazioni metodologiche e considerazioni critiche

Le performance limitate del modello evidenziano diverse criticità:

1. **Complessità del fenomeno:** Inquinanti come PM e TVOC sono influenzati da fattori non meteorologici (traffico, industrie) che il modello non può "vedere".
2. **Qualità dei dati:** I problemi con il sensore di umidità hanno probabilmente influenzato notevolmente la qualità dei dati, conseguentemente anche della eCO₂.

3. **Eventi non prevedibili:** Eventi come il trasporto di fumi da incendi lontani, o locali, come quello delle sterpaglie, indeboliscono il modello nelle sue capacità predittive. Situazioni del genere sono casuali e non sono predicibili con certezza.
4. **Orizzonte temporale:** Alcune dinamiche di inquinamento sono più lente potrebbero nascondersi (e.g il particolato è noto seguire una stagionalità).
5. **Periodo di rilevazione:** Il periodo di rilevazione è stato breve e meteorologicamente stabile.

5. Conclusioni

Il modello sviluppato mostra una capacità, seppur molto limitata, di prevedere la CO₂ basandosi solo su dati meteorologici. Tuttavia, la predizione di altri inquinanti come particolato e TVOC si è rivelata inefficace, dimostrando che questo approccio non è sufficiente.

Inoltre, è fondamentale sottolineare nuovamente che il periodo di rilevamento dati è stato limitato. Per catturare la stagionalità e l'impatto di eventi meteorologici specifici, sarebbe necessario estendere la raccolta dati su base annuale o multiannuale, migliorando così la robustezza e la generalizzabilità del modello.

Per migliorare le performance, sarebbe necessario integrare informazioni aggiuntive, come dati su fonti emissive locali e integrare dati da modelli meteorologici su scala più ampia. I risultati suggeriscono che un approccio puramente meteorologico può fornire una stima di base, ma deve essere arricchito sia lato temporale, dal lato qualitativo dei dati e sia dal dominio di informazioni necessarie, per diventare uno strumento predittivo veramente affidabile.