



Politecnico di Torino

Corso di Laurea

A.a. 2024/2025

Digital Twin per il controllo della qualità: un approccio sperimentale alla diagnostica dei difetti

Relatori:

Prof. Maurizio Galetto
Prof.ssa Elisa Verna

Candidati:

Ilaria Catillo

Indice

Elenco delle tabelle	III
Elenco delle figure	V
1 Introduzione	1
2 Il Digital Twin per il controllo qualità	3
2.1 Il Digital Twin: concetti generali	3
2.1.1 Confronto con modelli tradizionali e livelli di maturità	4
2.1.2 Architettura e dimensione digitale	6
2.1.3 L'architettura MDT a cinque dimensioni	7
2.1.4 Approcci di modellazione del Digital Twin	10
2.2 Dal monitoraggio alla decisione: il DT come strumento per la qualità	12
2.3 Le misure numeriche nel controllo qualità tramite Digital Twin: opportunità e limiti	13
2.4 L'integrazione delle variabili categoriche per la classificazione dei difetti	14
3 Analisi della letteratura	19
3.1 Approccio metodologico e criteri di selezione della letteratura	19
3.2 Analisi descrittiva della letteratura raccolta	26
3.3 Classificazione sistematica dei casi di studio	31
3.3.1 Tipologia di dati utilizzati	31
3.3.2 Tecniche di classificazione adottate	32
3.3.3 Ambito di applicazione	33
3.3.4 Struttura del Digital Twin	34
3.3.5 Modalità di costruzione dell'output categorico	36
3.4 Considerazioni generali e implicazioni per il caso di studio	37
4 Implementazione di un Digital Twin per la diagnostica dei difetti	39
4.1 Introduzione al caso studio	39

4.1.1	Configurazione del caso studio e flusso dati adottato	41
4.2	Preparazione ed etichettatura del dataset	42
4.2.1	Caricamento dei dati	42
4.2.2	Etichettatura dei difetti	42
4.2.3	Separazione delle variabili di input e di output	44
4.3	Integrazione dei modelli predittivi nel Digital Twin	45
4.3.1	Architettura del Digital Twin	47
4.3.2	Implementazione dei modelli predittivi	49
4.3.3	Il problema dello sbilanciamento: tecniche di oversampling .	52
4.4	Sviluppo e Addestramento del Modello	53
4.4.1	Suddivisione del dataset	53
4.4.2	Pre-elaborazione dei dati	57
4.5	Processo di valutazione	59
5	Valutazione dei Risultati e Discussione	65
5.1	Metodologia di valutazione	65
5.2	Risultati dei modelli	70
5.2.1	Valutazione del modello Random Forest	72
5.2.2	Valutazione del modello Multilayer Perceptron	76
5.2.3	Valutazione del modello Multinominal Linear Regression .	80
5.2.4	Valutazione del modello LightGBM	85
5.3	Analisi e discussione dei risultati	89
5.3.1	Performance dei modelli	89
5.3.2	Confronto con altri approcci nel controllo qualità	91
6	Conclusioni e implicazioni	93
6.1	Conclusioni	93
6.2	Implicazioni pratiche, limiti del lavoro e sviluppi futuri	94
Glossario		97
Acronimi		101
Bibliografia		102

Elenco delle tabelle

2.1	Confronto tra modelli digitali nei sistemi produttivi: SCADA/PLC/MES, CAD/CAE e Digital Twin.	5
2.2	Gradi di evoluzione del Digital Twin [8].	6
3.1	Raccolta degli articoli scientifici selezionati per la revisione bibliografica	20
3.2	Differenze tra Digital Twin Passivo e Digital Twin Intelligente.	35
4.1	Descrizione delle variabili di input presenti nel dataset.	44
4.2	Estratto delle previsioni del modello di classificazione sul gruppo di lavorazione TP = 3.	62
4.3	Estratto delle previsioni del modello MLP sul gruppo di lavorazione TP = 3.	63
4.4	Estratto delle previsioni del modello MLR sul gruppo di lavorazione TP = 3.	63
4.5	Estratto delle previsioni del modello LightGBM sul gruppo di lavorazione TP = 3.	63
5.1	Struttura della matrice di confusione.	66
5.2	Confronto delle metriche principali tra ROS e SMOTENC per ciascun modello.	70
5.3	Matrice di confusione del modello Random Forest sul set di test. . .	72
5.4	Prestazioni del modello Random Forest per ciascuna classe.	73
5.5	Metriche complessive del modello Random Forest.	73
5.6	Matrice di confusione del modello MLP sul set di test.	76
5.7	Prestazioni del modello MLP per ciascuna classe.	77
5.8	Metriche complessive del modello MLP.	77
5.9	Matrice di confusione del modello MLR sul set di test.	80
5.10	Prestazioni del modello MLR per ciascuna classe.	81
5.11	Metriche complessive del modello MLR.	82
5.12	Matrice di confusione del modello LightGBM sul set di test.	85
5.13	Prestazioni del modello LightGBM per ciascuna classe.	85

5.14 Metriche complessive del modello LightGBM.	86
5.15 Metriche di performance dei modelli riportate nello studio di Boettjer e Ramanujan (2023). * indica che la metrica è ponderata.	90

Elenco delle figure

2.1	Modello generale del Digital Twin per un prodotto.	7
2.2	Architettura del sistema Digital Twin utilizzato per il monitoraggio della qualità nella saldatura, basato su tecniche di deep learning e segnali acustici.	8
2.3	Flusso diagnostico multi-modale integrato in un Digital Twin per la classificazione automatica di difetti [15].	15
2.4	Architettura sperimentale del sistema di saldatura robotizzata con Digital Twin [4].	16
2.5	Classificazione degli stati di penetrazione del bagno di saldatura [4].	16
2.6	Diversi campioni di piccoli difetti superficiali nel set di dati DeA, dove i rettangoli verdi si riferiscono ai difetti di deformazione [21]. .	17
3.1	Distribuzione temporale delle pubblicazioni (2020–2025)	27
3.2	Distribuzione geografica degli articoli pubblicati	28
4.1	Rappresentazione del test part utilizzato come provino per lo studio [26].	40
5.3	Curva ROC del modello Random Forest.	74
5.4	Curva Precision-Recall del modello Random Forest.	75
5.5	Curva ROC del modello MLP.	78
5.6	Curva Precision-Recall del modello MLP.	79
5.7	Curva ROC del modello MLR.	83
5.8	Curva Precision-Recall del modello MLR.	84
5.9	Curva ROC del modello LightGBM.	87
5.10	Curva Precision-Recall del modello LightGBM.	88

Capitolo 1 – Introduzione

Negli ultimi anni, il concetto di Digital Twin (DT) si è affermato come uno dei pilastri fondamentali della trasformazione digitale in ambito industriale, offrendo nuove prospettive per il monitoraggio, la previsione e l'ottimizzazione dei processi produttivi. Attraverso la creazione di una replica virtuale dinamica di un sistema fisico, il DT consente di osservare in tempo reale il comportamento di un processo, simulare scenari alternativi e supportare le decisioni operative in modo più rapido e informato, riducendo inefficienze e tempi di risposta. Grazie alla sua capacità di integrare e correlare informazioni provenienti da sensori, macchine e sistemi informativi aziendali, il DT rappresenta oggi uno strumento strategico per raggiungere gli obiettivi dell'industria 4.0 e, più in generale, per la transizione verso paradigmi produttivi intelligenti e adattivi.

Tra i diversi ambiti applicativi, il controllo qualità si configura come uno dei settori più promettenti per l'adozione del DT: in questo contesto, il DT, può contribuire in modo decisivo alla riduzione degli scarti, al miglioramento della diagnosi dei difetti e all'ottimizzazione delle attività di ispezione e manutenzione, trasformando la gestione della qualità da un processo reattivo a un approccio predittivo e proattivo. L'integrazione di modelli di intelligenza artificiale e di apprendimento automatico ha ulteriormente ampliato le potenzialità del DT, permettendo di estrarre conoscenza dai dati di processo e di adattare il comportamento del sistema in funzione delle condizioni operative reali.

Tradizionalmente, i sistemi di verifica della qualità si sono basati su misure quantitative, ovvero su grandezze numeriche continue come dimensioni geometriche, vibrazioni, temperature o pressioni. Tuttavia, in molti contesti industriali, le decisioni legate alla qualità non dipendono esclusivamente da tali valori numerici, ma anche da informazioni qualitative o categoriche, come la presenza o assenza di un difetto, la tipologia dell'anomalia riscontrata o lo stato operativo di una macchina. L'inclusione di queste variabili all'interno dei modelli di DT rappresenta un'evoluzione significativa, poiché consente di riprodurre in modo più realistico la complessità dei processi produttivi, dove le decisioni operative derivano dall'integrazione di dati numerici, discreti e semantici. Nonostante la crescente diffusione dei DT in ambito industriale, l'integrazione di variabili categoriche nei modelli

di controllo qualità è ancora un ambito poco esplorato, specialmente in contesti caratterizzati da elevata variabilità e complessità dei dati.

In questa prospettiva si inserisce il seguente lavoro di tesi, che ha come obiettivo lo sviluppo di un DT sperimentale per il controllo della qualità, capace di elaborare non solo misure numeriche ma anche informazioni qualitative e categoriche. Il modello proposto si focalizza sulla diagnostica dei difetti nella produzione di stampi ad alta precisione, un settore in cui la qualità del processo dipende fortemente dalle condizioni operative e dalle deviazioni geometriche dei componenti. A differenza degli approcci tradizionali, il DT sviluppato in questa tesi combina variabili numeriche e qualitative per restituire un output diagnostico di tipo categorico (difettoso/non difettoso), offrendo una rappresentazione più interpretabile e utile per il processo decisionale. L'obiettivo è esplorare come i DT possano essere impiegati non soltanto per analisi numeriche tradizionali, ma anche per la diagnostica dei difetti e la classificazione qualitativa dello stato dei componenti.

La ricerca si è articolata in due fasi principali. In primo luogo, è stata condotta un'analisi bibliografica sistematica dei principali contributi scientifici che trattano l'applicazione del DT al controllo qualità, con particolare attenzione agli studi che producono un output categorico (ad esempio difettoso/non difettoso). Questa fase ha permesso di individuare le tendenze emergenti, le metodologie più ricorrenti e le lacune ancora presenti nella letteratura, in particolare riguardo all'integrazione di dati misti e all'adattabilità dei modelli in scenari produttivi reali.

Sulla base delle evidenze raccolte, è stato poi sviluppato un modello sperimentale ispirato a un caso di studio accademico di riferimento, adattato al contesto della produzione di stampi ad alta precisione. Quest'ultimo ha rappresentato un contesto applicativo per la progettazione e validazione di un DT in grado di diagnosticare in modo automatico la conformità dei componenti.

Il lavoro presentato in questa tesi offre un contributo metodologico e applicativo al tema del DT per il controllo qualità, proponendo un approccio in grado di combinare rigore analitico, adattabilità e fedeltà alle condizioni operative reali, orientato a migliorare la rappresentatività dei dati, la robustezza del modello e la coerenza con le reali dinamiche di processo. Nei capitoli successivi vengono approfonditi gli aspetti teorici e sperimentali del progetto: il Capitolo 2 illustra le basi concettuali e i principali approcci presenti in letteratura riguardanti i DT e la loro applicazione al controllo qualità; il Capitolo 3 presenta un'analisi sistematica della letteratura sui DT con output categorici, evidenziando le principali direzioni di ricerca; il Capitolo 4 descrive la progettazione del modello sperimentale e le scelte metodologiche adottate; il Capitolo 5 presenta i risultati della sperimentazione e la loro interpretazione; infine, il Capitolo 6 raccoglie le conclusioni e le prospettive future, con particolare riferimento all'evoluzione verso DT adattivi e cognitivi, capaci di apprendere in modo continuo dai dati di processo e di restituire valore decisionale in tempo reale.

Capitolo 2 – Il Digital Twin per il controllo qualità

2.1 Il Digital Twin: concetti generali

Il Digital Twin è una tecnologia digitale che consente di replicare virtualmente un'entità fisica (che può essere un oggetto, processo o sistema), mantenendo con esso un legame dinamico e aggiornato in tempo reale[1]. A differenza di modelli digitali statici, come quelli impiegati nelle tradizionali simulazioni offline, il DT riceve continuamente dati dal sistema reale e, allo stesso tempo, è in grado di restituire indicazioni operative o predittive [1]. Questa connessione bidirezionale tra fisico e virtuale permette al DT di evolversi con il sistema monitorato, abilitando funzioni come il controllo remoto, la simulazione di scenari alternativi e la manutenzione predittiva. Si tratta di un approccio orientato non solo alla rappresentazione, ma anche alla comprensione e ottimizzazione del comportamento del sistema lungo il suo intero ciclo di vita [1].

Originariamente introdotto da Grieves nel 2003 come un'idea concettuale per la gestione del ciclo di vita dei prodotti (PLM), il DT è stato formalmente impiegato nel contesto aerospaziale a partire dal 2011 [2]; da allora, l'interesse è cresciuto rapidamente, con un'accelerazione significativa a partire dal 2017, portando a molti studi e casi applicativi [2]. Inizialmente focalizzata su aspetti teorici e architettonici, la ricerca si è via via spostata verso applicazioni specifiche in ambiti industriali complessi [2]. Nel campo dell'assemblaggio di motori aeronautici, ad esempio, i DT sono stati impiegati per l'analisi geometrica in tempo reale, la valutazione della qualità, la pianificazione del processo e la previsione di deviazioni dimensionali, integrando modelli CAD, misurazioni fisiche e algoritmi di simulazione [2]. Questo sviluppo evidenzia come il DT, da concetto teorico, si sia evoluto in uno strumento operativo avanzato per il controllo qualità e l'ottimizzazione dei processi produttivi [2]. Le tecnologie abilitanti che rendono possibile la realizzazione e l'operatività del Digital Twin sono molteplici e comprendono: la modellazione CAD/CAE che consente la rappresentazione virtuale della geometria e del comportamento del sistema; sistemi SCADA per il monitoraggio e la supervisione in tempo reale

dei processi; Internet of Things (IoT) che abilita la raccolta di dati continui dal mondo fisico tramite sensori connessi; e infine le tecniche di intelligenza artificiale e machine learning, utilizzate per l’elaborazione avanzata dei dati, la previsione e la diagnosi. Alcuni studi hanno evidenziato l’impiego di reti neurali per supportare decisioni in tempo reale basate su dati acustici o termici, come nel caso del controllo della qualità di saldature robotizzate, dimostrando la crescente capacità del DT di adattarsi a contesti produttivi complessi e variabili [3][4].

2.1.1 Confronto con modelli tradizionali e livelli di maturità

Negli ultimi decenni, la progressiva digitalizzazione dei processi produttivi ha portato alla nascita di diversi strumenti per la rappresentazione virtuale dei sistemi fisici. Dai modelli tridimensionali sviluppati nei software CAD/CAE, fino alle simulazioni di processo e ai sistemi di monitoraggio industriale, tali soluzioni hanno progressivamente arricchito la capacità delle imprese di comprendere, analizzare e ottimizzare le proprie operazioni. In questo scenario, il Digital Twin si colloca come un’evoluzione naturale ma più ambiziosa: non si limita infatti a descrivere o simulare un oggetto fisico, bensì ne riproduce il comportamento in tempo reale, creando un collegamento continuo e bidirezionale tra mondo reale e mondo digitale [5]. Mentre i modelli 1D/3D tradizionali offrono una rappresentazione statica o semi-dinamica (ad esempio nei sistemi CAD o CAE), il DT si distingue per la capacità di aggiornarsi costantemente sulla base dei dati provenienti dal sistema fisico, reagire agli eventi e predire l’evoluzione del processo grazie all’integrazione di sensori e algoritmi di analisi avanzata. Allo stesso modo, rispetto alle simulazioni convenzionali – spesso utilizzate per scenari ipotetici o in fase progettuale – il DT mantiene una validità operativa lungo l’intero ciclo di vita del prodotto o dell’impianto [5].

In ambito industriale, questa tecnologia non sostituisce i sistemi digitali esistenti, ma si innesta su di essi potenziandone le funzionalità. Strumenti come SCADA, PLC e MES, rappresentano da anni la base dell’automazione e del controllo dei processi. Tuttavia, essi operano prevalentemente in modo descrittivo o reattivo, mentre il DT mira a fornire una visione predittiva e decisionale, integrando i dati eterogenei provenienti da tali sistemi in un unico modello coerente e dinamico [6][7]. Per chiarire ulteriormente le differenze tra il Digital Twin e i sistemi digitali tradizionali impiegati nei contesti produttivi, la Tabella 2.1 confronta alcune dimensioni chiave tra i modelli SCADA/PLC/MES, CAD/CAE e Digital Twin.

Tabella 2.1: Confronto tra modelli digitali nei sistemi produttivi: SCADA/PLC/-MES, CAD/CAE e Digital Twin.

Caratteristica	SCADA/PLC/MES	CAD/CAE	Digital Twin
Connessione con il mondo reale	Parziale / unidirezionale	Nessuna o simulata	Bidirezionale e continua
Capacità predittiva	Limitata	Presente, ma offline	Avanzata, in tempo reale
Aggiornamento in tempo reale	Limitato	Assente	Presente
Integrazione di dati multisorgente	Limitata	Solo input definiti	Elevata (sensori, AI, simulazioni)
Interattività con l'utente	Interfacce statiche	Interfacce di simulazione	Dashboard dinamiche, analisi di scenari ipotetici (“what-if”)
Ambito di utilizzo	Controllo e supervisione operativa	Progettazione e test	Controllo qualità, manutenzione, simulazione e supporto decisionale

Come si può notare, il DT si distingue per la connessione continua e bidirezionale con il sistema reale, per la capacità di elaborare previsioni in tempo reale e per l'elevato grado di interattività con l'utente, grazie a dashboard dinamiche e simulazioni decisionali. A differenza dei modelli CAD, che sono statici e confinati alla fase progettuale, o dei sistemi SCADA e MES, che hanno funzionalità operative ma limitate capacità predittive, il Digital Twin integra dati multisorgente ed è operativo lungo l'intero ciclo di vita del prodotto. Questa flessibilità lo rende particolarmente adatto ad applicazioni complesse come il controllo qualità e la manutenzione predittiva. Negli ultimi anni, diverse aziende e consorzi industriali, tra cui Siemens, IBM e PTC, hanno proposto modelli progressivi per descrivere i livelli di adozione e maturità dei DT. Tra i modelli più diffusi, si ritrova uno schema articolato in cinque stadi, che vanno da una semplice rappresentazione statica del sistema fino a un gemello digitale cognitivo e autonomo. Una sintesi concettuale di questi livelli è descritta nella Tabella 2.2:

Tabella 2.2: Gradi di evoluzione del Digital Twin [8].

Livello	Descrizione
0	Nessuna replica digitale: il sistema fisico non è digitalizzato.
1	Modello digitale statico: esiste un modello CAD o simulativo non connesso in tempo reale.
2	Digital Twin con flusso unidirezionale: il modello riceve dati dal sistema fisico, ma non è interattivo.
3	Digital Twin interattivo: il sistema virtuale elabora i dati e può influenzare o correggere il comportamento del sistema reale.
4	Digital Twin cognitivo: il DT apprende, si adatta e agisce in modo autonomo grazie a tecniche di intelligenza artificiale.

Questi livelli sono frequentemente citati nella documentazione tecnica di PTC (2021), IBM (2021) e Siemens (2020), nonché richiamati nella letteratura scientifica [8].

2.1.2 Architettura e dimensione digitale

Nella sua forma più generale, il Digital Twin, si articola in tre elementi fondamentali [5][9]: Il prodotto fisico (*physical entity*), che genera dati durante le sue fasi di progettazione, produzione, utilizzo, manutenzione e dismissione. Le entità fisiche, grazie al fatto che hanno caratteristiche, comportamenti e performance differenti durante il ciclo di vita, generano una grande quantità di dati [5].

Il modello virtuale (*virtual entity*), che riflette digitalmente struttura, comportamento e prestazioni dell'oggetto reale. Comprende non solo i modelli geometrici, ma anche tutte le regole e i comportamenti, come le proprietà dei materiali, l'analisi meccanica e il monitoraggio della salute [5].

Il collegamento dati (*data link*), ovvero l'insieme di flussi informativi che permettono lo scambio continuo e bidirezionale tra mondo fisico e digitale. I dati connessi includono sottoinsiemi di dati fisici e virtuali, nonché nuovi dati acquisiti tramite l'integrazione, la fusione e l'analisi dei dati fisici e virtuali [5].

La Figura 2.1 tratta dall'articolo "*Digital twin-driven product design framework*" di Fei Tao et al., mostra sinteticamente la struttura del Digital Twin, evidenziando i suoi principali elementi.

Questa visione supera l'approccio unidirezionale dei tradizionali strumenti di progettazione digitale, introducendo un dialogo costante tra le due dimensioni: da un lato, il modello virtuale può simulare, monitorare e prevedere il comportamento del prodotto fisico; dall'altro, i dati provenienti dal campo permettono al DT di aggiornarsi e riflettere accuratamente lo stato reale del sistema [10].

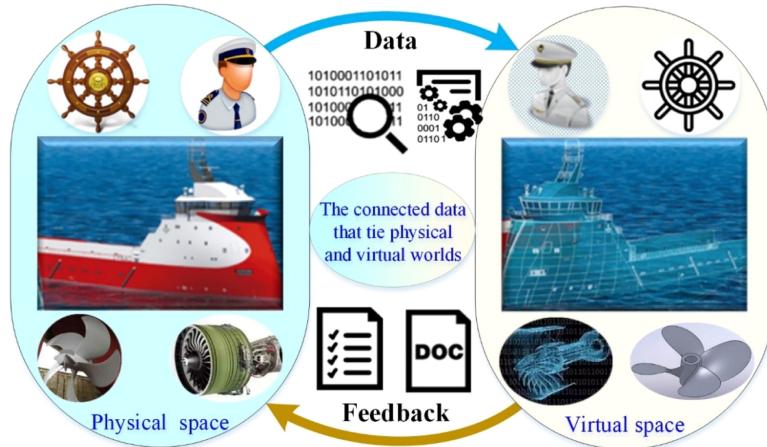


Figura 2.1: Modello generale del Digital Twin per un prodotto.

L’architettura del DT, secondo l’evoluzione concettuale proposta, non è più rigida o lineare ma si presenta come un modello a ciclo chiuso, capace di integrare continuamente nuove informazioni e supportare processi decisionali complessi [10].

Il modello è in grado di raccogliere ed elaborare dati durante l’intero ciclo di vita del prodotto, trasformandoli in informazioni utili per attività come progettazione, diagnostica, manutenzione e miglioramento continuo. Inoltre, consente un’interazione in tempo reale con il mondo fisico, anche grazie all’utilizzo di meccanismi di intervento che permettono modifiche operative dirette [5].

Questo modello tridimensionale ha rappresentato a lungo la struttura di riferimento per il DT, soprattutto nei contesti produttivi e ingegneristici. Tuttavia, con l’aumento della complessità dei sistemi, della varietà dei dati e della necessità di fornire servizi intelligenti in tempo reale, è emersa una nuova formulazione a cinque dimensioni [3], che estende quella originaria e che verrà descritta nella sezione successiva.

2.1.3 L’architettura MDT a cinque dimensioni

Il modello Multi-dimensional Digital Twin (MDT), formalizzato come segue [3]:

$$MDT = (PE, DE, Ss, DD, CN)$$

rappresenta un’estensione del classico schema a tre componenti, includendo due dimensioni aggiuntive fondamentali: i dati digitali (*DE*) e i servizi intelligenti (*Ss*). La Figura 2.2, tratta dall’articolo *"Deep Learning-Empowered Digital Twin Using Acoustic Signal for Welding Quality Inspection"* di Tao Ji e Norzalilah Mohamad Nor, illustra l’interazione tra entità fisiche e digitali, la raccolta dei dati (ambientali,

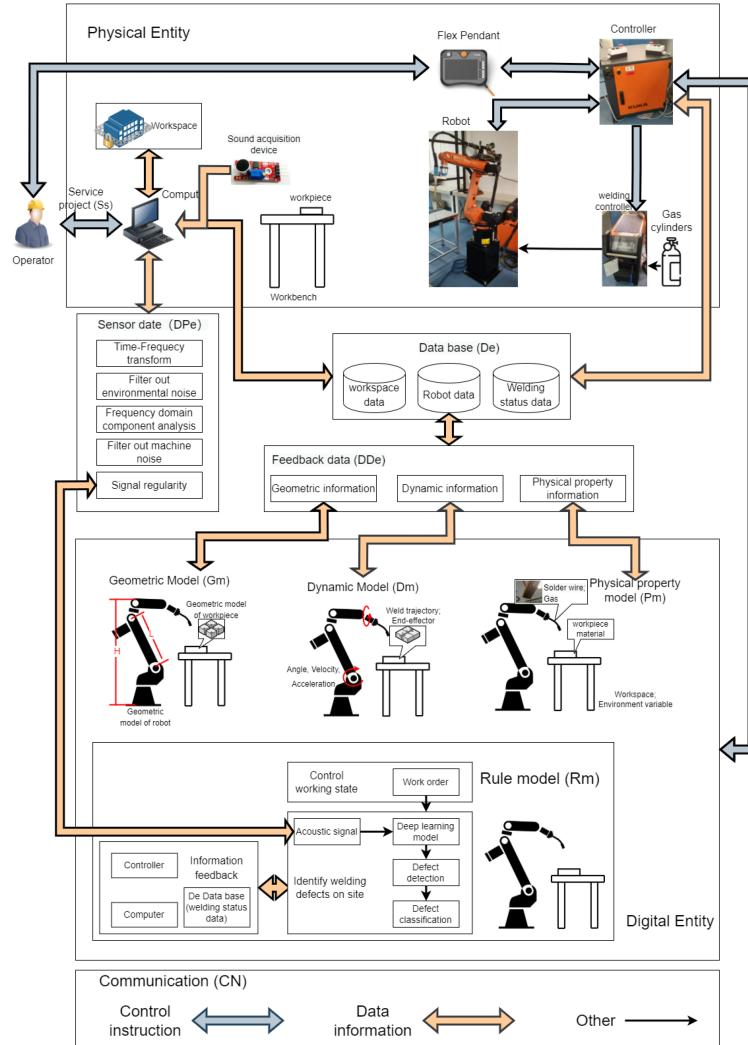


Figura 2.2: Architettura del sistema Digital Twin utilizzato per il monitoraggio della qualità nella saldatura, basato su tecniche di deep learning e segnali acustici.

del robot e dello stato di saldatura) e l'elaborazione tramite modelli geometrici, dinamici e fisici, supportati da un modello a regole per il riconoscimento dello stato di saldatura e la classificazione dei difetti.

Ciascuna dimensione del modello svolge un ruolo specifico all'interno dell'architettura[3]:

1. Physical Entity (PE)

La componente fisica che rappresenta l'oggetto reale da monitorare, come una macchina, un impianto, un robot o un ambiente di produzione. Include tutti

gli elementi che esistono nel mondo fisico e che sono oggetto di digitalizzazione.

2. Digital Entity (DE)

La rappresentazione virtuale altamente integrata e strutturata dell'entità fisica.
Include quattro sottocomponenti:

$$DE = (Gm, Pm, Dm, Rm)$$

- *Geometric model (Gm)*: il modello CAD tridimensionale della macchina o del componente;
- *Physical model (Pm)*: proprietà fisiche e materiali;
- *Dynamic model (Dm)*: informazioni cinematiche, traiettorie, dinamiche operative;
- *Rule model (Rm)*: modello di regole decisionali, spesso basato su algoritmi AI, come il SeCNN-LSTM nel caso studio per classificare i difetti di saldatura.

3. Service System (Ss)

Definisce lo scopo e le funzionalità del Digital Twin. È descritto come:

$$Ss = (Fun, Input, Output, Qua, St)$$

- *Fun*: la funzione principale del DT (es. identificazione dei difetti in saldatura);
- *Input / Output*: dati in ingresso e in uscita, come segnali acustici, posizione del robot, feedback;
- *Qua*: metriche di qualità del servizio, come accuratezza diagnostica o efficienza di riconoscimento;
- *St*: stato operativo della macchina o del sistema.

4. Digital Twin Data (DD)

La componente dati che comprende tutte le fonti informative del sistema, suddivise in:

$$DD = (DPe, DDe, DSs, De)$$

- *DPe*: dati provenienti dall'entità fisica (sensori, robot);
- *DDe*: dati di feedback dell'entità digitale (simulazioni, predizioni);
- *DSs*: dati generati dai servizi;
- *De*: conoscenza esperta, modelli esperti o basi di conoscenza.

5. Communication Network (CN)

Rappresenta l'infrastruttura per il trasferimento bidirezionale delle informazioni. È formalizzato come:

$$CN = (PE_DD, BiDE_DD, PE_Ss, BiDE_Ss, BiDD_Ss)$$

Dove ciascuna sigla rappresenta una specifica connessione (mono o bidirezionale) tra i blocchi funzionali del Digital Twin:

- *PE_DD*: flusso unidirezionale dai sensori dell'entità fisica al modulo dati;
- *BiDE_DD*: scambio bidirezionale tra l'entità digitale e il modulo dati, che consente aggiornamenti e predizioni;
- *PE_Ss*: canale diretto tra la componente fisica e il sistema di servizi, ad esempio per attivare azioni correttive automatiche;
- *BiDE_Ss*: interazione continua tra il gemello digitale e i moduli decisionali;
- *BiDD_Ss*: aggiornamento dinamico tra base dati e servizi.

Queste connessioni non sono solo astratte, ma rappresentano canali di comunicazione reali tra i vari componenti del sistema, implementati tramite protocolli Ethernet, interfacce fisiche, sensori acustici e controller industriali [3]. L'architettura supporta sia comunicazioni unidirezionali, come nel caso di sensori acustici esterni, che trasmettono dati al DT senza ricevere feedback, sia connessioni bidirezionali, come quelle tra il robot industriale e il controller che consente l'adattamento in tempo reale del processo [3].

Questo modello MDT introduce dunque una visione strutturata e modulare del DT, capace non solo di replicare un sistema fisico, ma anche di apprendere da esso, agire in tempo reale e offrire servizi intelligenti per l'ottimizzazione del processo. L'articolazione delle connessioni tra i blocchi, così come l'integrazione tra dati eterogenei e modelli intelligenti, rende questa architettura particolarmente adatta a scenari industriali complessi [3][10].

2.1.4 Approcci di modellazione del Digital Twin

Tra le caratteristiche fondamentali del DT si evidenziano, oltre che l'interconnessione continua tra fisico e virtuale, la capacità di integrazione di modelli geometrici, comportamentali e fisici, e la possibilità di effettuare simulazioni predittive in tempo reale [11]. Questi modelli possono essere costruiti secondo approcci differenti [11], tra cui esistono i modelli white-box, basati su relazioni causa-effetto e modellazione fisica. Sono tipicamente espressi attraverso equazioni differenziali, modelli termodinamici o relazioni ingegneristiche ben definite, e risultano particolarmente utili in domini in cui il comportamento del sistema è governato da leggi fisiche esplicite (es. modelli FEM, termici o meccanici);

I modelli black-box (*data-driven*) sono, invece, orientati ai dati, apprendono correlazioni statistiche e funzionali a partire da dataset osservazionali, spesso tramite tecniche di machine learning. Pur offrendo ottime performance predittive, questi modelli sono meno trasparenti e risultano più sensibili alla qualità e quantità dei dati disponibili.

Infine, i modelli gray-box (*ibridi*) combinano la robustezza e l'interpretabilità dei modelli white-box con la capacità di adattamento e la flessibilità dei modelli black-box. Questa categoria rappresenta una direzione particolarmente promettente per l'implementazione di Digital Twin in contesti industriali complessi e dinamici.

Nel contesto del controllo qualità e della diagnostica dei difetti, la scelta del tipo di modello da integrare nel Digital Twin dipende da molteplici fattori: il livello di conoscenza del sistema fisico, la disponibilità e la qualità dei dati, la necessità di interpretabilità, e il grado di variabilità dell'ambiente operativo [11]. Ad esempio, sistemi di ispezione basati su immagini o segnali acustici impiegano tipicamente modelli black-box (es. reti neurali convoluzionali), mentre i processi fisici noti, come la simulazione della deformazione meccanica o delle sollecitazioni termiche, si prestano a modellazioni white-box. L'approccio gray-box, invece, è particolarmente adatto per la manutenzione predittiva o per contesti in cui è necessario aggiornare in modo dinamico un modello inizialmente fisico con dati in tempo reale provenienti da sensori o da feedback operativi.

Nel contesto produttivo, il DT si configura come un elemento cardine della transizione verso la fabbrica intelligente [6]. Le sue applicazioni si estendono dall'ingegneria di prodotto alla produzione, fino alla manutenzione predittiva e al fine vita, offrendo supporto alla pianificazione, al controllo e alla gestione delle risorse [6]. Nel settore dei materiali compositi, ad esempio, l'utilizzo di DT potenziati da tecniche di deep learning consente la classificazione automatica di difetti, anche non visibili, attraverso l'analisi di segnali acustici e immagini CT (Computed Tomography). Queste ultime, ottenute tramite tomografia a raggi X, permettono di visualizzare in dettaglio l'interno dei materiali senza danneggiarli, e rappresentano una fonte preziosa per alimentare modelli digitali ad alta fedeltà [12].

La funzione distintiva del DT nella trasformazione digitale risiede proprio nella sua capacità di fornire un controllo qualità continuo, adattivo e basato su analisi predittive. Oltre al monitoraggio in tempo reale, il DT consente di simulare scenari alternativi, ottimizzare le decisioni operative e reagire tempestivamente ai cambiamenti di stato del sistema [1].

Per questo, il suo impiego si sta diffondendo trasversalmente nei sistemi produttivi come paradigma avanzato per un controllo qualità intelligente, sostenibile e resiliente [13].

2.2 Dal monitoraggio alla decisione: il DT come strumento per la qualità

Nel contesto attuale dell'industria, caratterizzato da processi sempre più complessi, interconnessi e soggetti a continui cambiamenti, il controllo qualità ha assunto un ruolo centrale. Non si tratta più solo di una verifica a fine linea, ma di intervenire in modo continuo e proattivo. In questo scenario, il Digital Twin si sta affermando come uno strumento innovativo: non solo replica virtualmente ciò che accade nella realtà, ma consente di avere una visione in tempo reale del processo produttivo, rendendo possibile un controllo qualità distribuito, diffuso lungo tutta la filiera. In questo tipo di architetture, il DT funge non solo da replica passiva ma anche da nodo decisionale attivo, capace di integrare dati da più sensori, generare classificazioni in tempo reale e attivare il routing dei prodotti lungo percorsi alternativi in base alla qualità rilevata [6].

L'integrazione del DT nei processi produttivi permette di simulare, monitorare e intervenire su ogni fase del ciclo produttivo in tempo reale. Questo consente, ad esempio, di individuare tempestivamente eventuali anomalie e di reagire subito: se un prodotto presenta difetti, può essere deviato automaticamente verso la riparazione o lo scarto, evitando sprechi inutili e ottimizzando l'uso delle risorse. È stato dimostrato, ad esempio, che l'introduzione di controlli qualità intermedi in ciascuna stazione consente di prevenire l'accumulo di errori, migliorando l'efficienza del sistema e riducendo l'uso di materiale grezzo e rilavorazioni [6].

Questa capacità è stata dimostrata, ad esempio, nell'implementazione di DT in sistemi manifatturieri reali, dove gli algoritmi integrati hanno permesso di identificare in modo tempestivo i pezzi difettosi e instradarli automaticamente, migliorando così la tracciabilità, l'efficienza e la sostenibilità complessiva del processo. I DT abilitano inoltre simulazioni di scenari "what-if", che permettono di valutare in anticipo l'effetto di modifiche nei flussi produttivi o nei criteri di qualità, supportando decisioni più informate e mirate. Tali simulazioni sono particolarmente utili in fase di progettazione di nuove linee, nella definizione delle soglie di accettabilità e nella valutazione preventiva di scenari di guasto o congestione operativa [6].

Oltre al controllo qualità in sé, come anticipato precedentemente, il DT ha un ruolo rilevante anche nella manutenzione predittiva: grazie alla raccolta continua di dati da sensori distribuiti, è in grado di riconoscere pattern anomali e prevedere l'insorgenza di guasti prima che si verifichino, riducendo notevolmente i tempi di fermo macchina e i costi associati agli interventi correttivi non programmati [14][11].

Un altro aspetto cruciale è la capacità del DT di gestire la supervisione distribuita [6]. A differenza dei tradizionali sistemi di monitoraggio centralizzati, i DT consentono una visione frammentata ma interconnessa del sistema, in cui

ciascun nodo del processo può agire autonomamente sulla base delle informazioni disponibili. Questo consente di mantenere sotto controllo, in tempo reale, una molteplicità di parametri e condizioni operative, anche in impianti complessi e multi-prodotto. Questo approccio decentralizzato è stato sperimentato anche in scenari di produzione multi-livello, in cui la supervisione è suddivisa tra nodi locali e controllo centrale, con il DT che svolge una funzione di sincronizzazione e previsione tra stazioni eterogenee [6].

In definitiva, il Digital Twin sta trasformando il modo in cui pensiamo al controllo qualità: da semplice verifica finale a una presenza attiva, intelligente e continua all'interno del processo produttivo. La sua forza sta nella capacità di “dialogare” in tempo reale con ciò che accade fisicamente in fabbrica, adattandosi ai cambiamenti e offrendo supporto concreto alle decisioni operative.

2.3 Le misure numeriche nel controllo qualità tramite Digital Twin: opportunità e limiti

Nell'ambito del controllo qualità, le grandezze fisiche misurabili, come pressione, temperatura, vibrazioni e parametri geometrici, rappresentano da tempo la base per il monitoraggio delle condizioni operative di macchine e prodotti. Anche nell'ambito del Digital Twin, tali variabili rappresentano un elemento fondamentale di partenza: attraverso sensori distribuiti, vengono raccolti in modo continuo dati numerici che alimentano i modelli digitali, consentendo di simulare e confrontare in tempo reale il comportamento atteso con quello effettivamente osservato nel sistema fisico [15][14].

Gli approcci classici di analisi si basano su modelli fisico-matematici o sull'uso di soglie prestabilite: quando una variabile supera un certo limite, viene generato un allarme o attivata un'azione correttiva. In alcuni casi, sono utilizzati anche modelli di regressione, come nella saldatura robotizzata, dove la correlazione tra parametri come corrente e velocità di saldatura e la profondità di penetrazione viene modellata attraverso tecniche di regressione kernel non lineare [4]. Questo approccio ha mostrato buone prestazioni in condizioni controllate, ma la sua efficacia tende a diminuire in presenza di rumore di fondo, variabilità del materiale o condizioni ambientali mutevoli, fattori comuni in ambienti industriali reali. Dunque l'affidabilità di queste misure numeriche presenta alcuni limiti, specialmente in ambienti produttivi complessi, dinamici o scarsamente strutturati [4].

In molte situazioni, le variabili fisiche non sono sufficientemente esplicative per identificare difetti complessi o per predire anomalie in modo preciso. Per esempio, il solo dato vibratorio o termico può risultare ambiguo in caso di fenomeni sovrapposti o condizioni di carico variabili [14]. Inoltre, modelli costruiti su singole variabili

tendono a sottostimare la complessità del contesto operativo reale, dove entrano in gioco variabili difficilmente quantificabili in forma continua [14].

Un altro limite risiede nella natura “puntuale” di molte misure fisiche, che riflettono stati locali e temporanei, ma non sempre riescono a cogliere l’interazione tra sottosistemi o l’evoluzione di una criticità nel tempo [16]. Alcuni modelli DT provano a superare questo vincolo combinando più segnali (es. vibrazione + temperatura), ma in mancanza di dati qualitativi o contesto operativo, l’interpretazione resta ambigua [16]. Inoltre, molti approcci di regressione e soglia richiedono dataset ampi, bilanciati e rappresentativi: una condizione non sempre soddisfatta nel mondo produttivo, dove i guasti sono eventi rari, eterogenei e scarsamente etichettati [14].

Infine, l’efficacia di questi approcci dipende fortemente dalla qualità e dalla quantità dei dati disponibili. In scenari dove la frequenza di guasti è bassa o le condizioni di esercizio sono molto eterogenee, l’affidabilità dei modelli predittivi basati solo su misure numeriche può essere ridotta [14][17]. Ciò rende evidente la necessità di integrare nuove tipologie di dati, come le variabili categoriche o simboliche, che consentano di rappresentare e classificare in modo più articolato i fenomeni osservati [14], un tema che verrà approfondito nel prossimo paragrafo.

2.4 L’integrazione delle variabili categoriche per la classificazione dei difetti

Come già evidenziato nel paragrafo precedente, sebbene l’uso delle sole variabili numeriche sia essenziale, non risulta sempre sufficiente per cogliere l’intera complessità di un sistema produttivo reale [18][19].

In questo contesto, l’integrazione delle variabili categoriche nei modelli di Digital Twin si rivela strategica per rafforzare le capacità di classificazione dei difetti [14]. Per variabili categoriche si intendono dati discreti o simbolici (come “difetto superficiale”, “penetrazione eccessiva”, “errore di tipo A”, “macchina ferma”, “stato normale”) che descrivono lo stato di un’entità in modo qualitativo piuttosto che quantitativo [14][4][20].

Un esempio rappresentativo di questa transizione concettuale è illustrato in Figura 2.3, dove un flusso diagnostico multi-modale integra dati numerici raccolti da sensori acustici e visivi, elaborati tramite una rete GAN e un classificatore neurale (MobileViG), per generare etichette categoriche diagnostiche che guidano il processo decisionale del Digital Twin.

Nel dominio del DT, queste variabili sono utilizzate per semplificare la rappresentazione di eventi complessi e fornire una codifica logica a fenomeni difficili da modellare solo attraverso misure continue. Il vantaggio principale è che permettono di associare una semantica chiara ai dati raccolti e, quindi, facilitano l’automazione

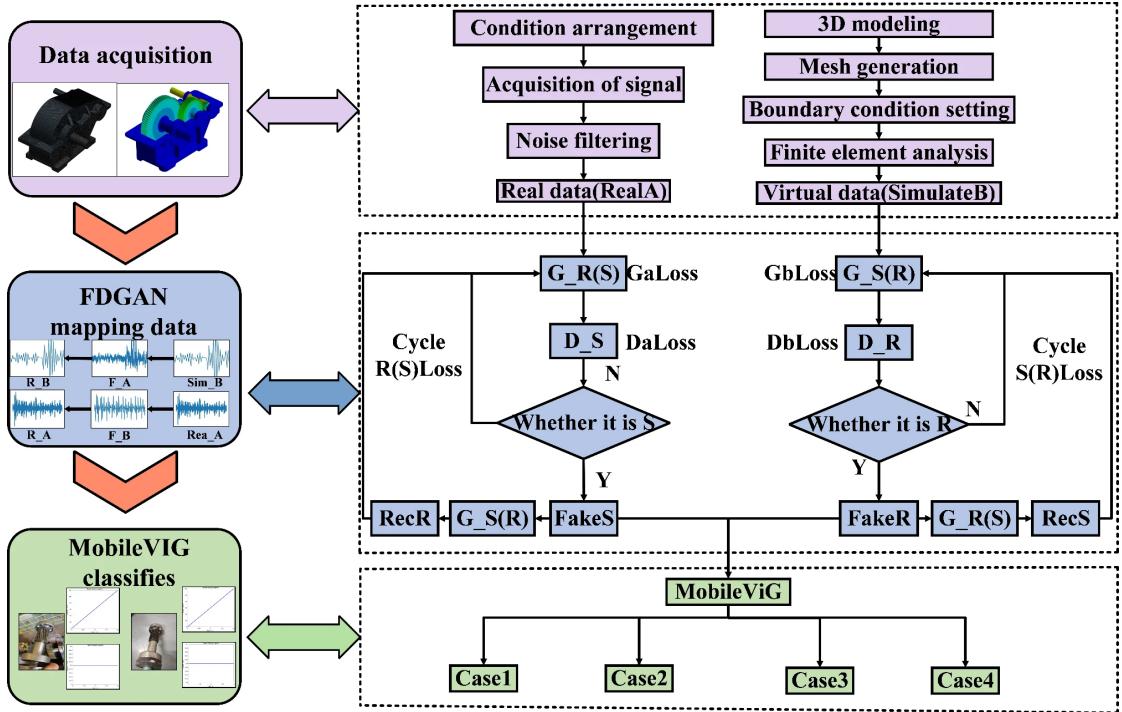


Figura 2.3: Flusso diagnostico multi-modale integrato in un Digital Twin per la classificazione automatica di difetti [15].

delle decisioni operative e l'attivazione di interventi correttivi su base logica e diretta [20].

Un ulteriore ambito applicativo riguarda i modelli utilizzati per classificare stati di penetrazione della saldatura: “incompleta”, “completa”, “eccessiva” e “bruciatura”. Queste classi discrete, emerse, come nel caso precedente, da un sistema di riconoscimento multi-modale (acustico + visivo), sono state integrate nel DT per regolare automaticamente corrente e velocità di saldatura, consentendogli di interagire con l’ambiente reale non solo in funzione di valori numerici, ma anche di etichette diagnostiche apprese [4].

Le Figure 2.4 e 2.5 mostrano rispettivamente l’architettura sperimentale del sistema di saldatura monitorato (con sensori acustici e visivi) e la classificazione degli stati di penetrazione del bagno di saldatura. Tali etichette diagnostiche costituiscono variabili categoriche utilizzate dal DT per l’adattamento automatico dei parametri di processo.

Analogamente, nel contesto della diagnostica di inverter connessi alla rete, la classificazione multi-classe di guasti è stata implementata tramite modelli LightGBM

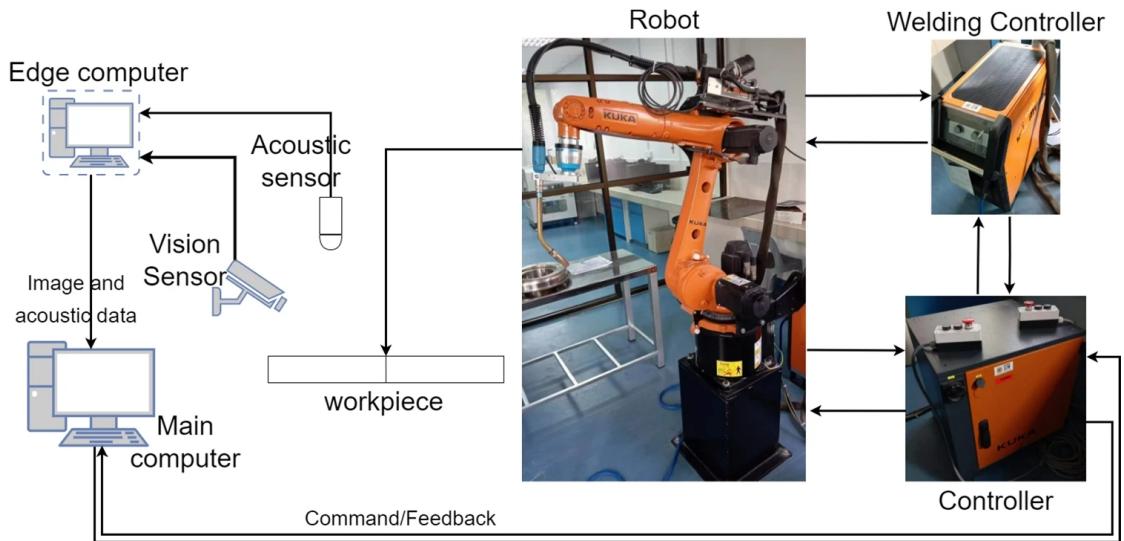


Figura 2.4: Architettura sperimentale del sistema di saldatura robotizzata con Digital Twin [4].

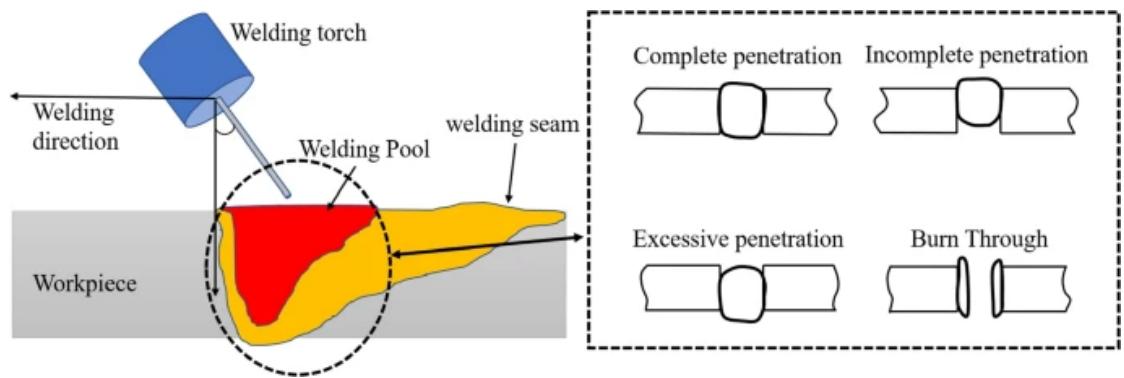


Figura 2.5: Classificazione degli stati di penetrazione del bagno di saldatura [4].

per classificare ogni osservazione in una delle seguenti categorie di guasto: “line-to-line fault”, “sensor fault”, “single-phase sag”, “three-phase fault”. L’integrazione di questi modelli all’interno del DT ha reso possibile una classificazione accurata anche in condizioni operative eterogenee, contribuendo a ridurre i falsi positivi e a rendere più efficiente la manutenzione. In questo caso, le diverse classi di guasto funzionano come variabili categoriche apprese, che alimentano il processo decisionale del gemello digitale e ne rafforzano l’affidabilità diagnostica, anche in presenza di dataset ridotti [20].

Un altro contributo rilevante è dato dai sistemi di ispezione superficiale. Nello studio presentato dall’articolo “Digital Twin of Intelligent Small Surface Defect

Detection with Cyber-manufacturing Systems", i modelli deep learning sono stati addestrati per riconoscere immagini etichettate con classi di difetti (es. "porosità", "crack", "bava") su superfici metalliche. Anche qui, le variabili categoriche fungono da interfaccia semantica tra l'output visivo del sistema e le decisioni automatizzate del DT [21]. A tal fine, è stato utilizzato un dataset DeA (Defect Annotated), una raccolta di immagini industriali ad alta risoluzione contenenti difetti localizzati in zone specifiche, evidenziate tramite riquadri annotati. Il sistema impiega un'architettura multi-modale per l'estrazione e la fusione di caratteristiche 2D/3D, classificando i difetti in categorie discrete. Come mostrato in Figura 2.6, le immagini annotate del dataset DeA forniscono le variabili categoriche diagnostiche (regioni difettose localizzate) che vengono integrate nel DT per supportare il processo decisionale automatizzato.

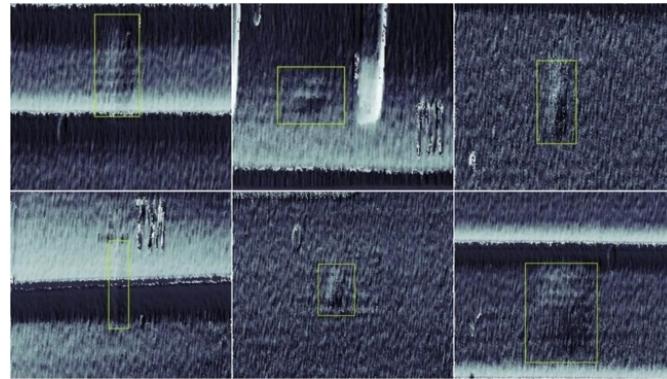


Figura 2.6: Diversi campioni di piccoli difetti superficiali nel set di dati DeA, dove i rettangoli verdi si riferiscono ai difetti di deformazione [21].

Infine, nel contesto dell'ispezione tridimensionale assistita da realtà aumentata, è stato implementato un meccanismo di classificazione semantica basato su segmentazioni 3D ricorsive. Questo approccio consente di analizzare le geometrie rilevate sul campo confrontandole con il DT di riferimento, assegnando a ciascun componente una "label dominante", una variabile categorica che ne identifica lo stato qualitativo (ad esempio, difettoso o conforme). Tale classificazione, integrata nel modello di DT, permette di attribuire significato operativo ai difetti rilevati anche in ambienti non strutturati, supportando il sistema nell'assegnazione di priorità di intervento e nell'interazione semantica con l'operatore [22].

In sintesi, l'introduzione delle variabili categoriche rappresenta un elemento abilitante per portare il DT oltre alla sola replica numerica del sistema fisico. Grazie alla possibilità di mappare informazioni discrete, qualitative e logiche, il DT diventa non solo più interpretabile, ma anche più efficace nel gestire la varietà e l'incertezza che caratterizzano la realtà produttiva. Ciò permette di costruire

modelli più flessibili, robusti e pienamente integrati nei processi decisionali e diagnostici dell'industria avanzata.

Capitolo 3 – Analisi della letteratura

3.1 Approccio metodologico e criteri di selezione della letteratura

L’obiettivo principale di questa revisione della letteratura è quello di individuare, classificare e confrontare gli approcci basati su Digital Twin che impiegano output categorici per il controllo qualità. In particolare, l’analisi mira a far emergere come i Digital Twin vengano implementati per supportare processi decisionali nei quali la qualità del prodotto è descritta non solo attraverso misure numeriche continue, ma anche attraverso variabili simboliche o qualitative, come la presenza o il tipo di difetto. Per garantire rigore metodologico e trasparenza nella selezione delle fonti, la revisione è stata condotta seguendo il metodo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), ampiamente riconosciuto per la strutturazione e conduzione delle revisioni sistematiche della letteratura[23].

Per raccogliere un insieme significativo di contributi, sono state consultate principalmente le banche dati scientifiche Scopus e Google Scholar, selezionando articoli pubblicati prevalentemente tra il 2020 e il 2025, periodo in cui l’interesse per l’applicazione dei DT in ambito manifatturiero ha registrato una crescita esponenziale.

Per una migliore consultazione e per fornire una panoramica immediata della ricerca in esame, i risultati sono stati organizzati e riassunti nella seguente tabella in Tabella 3.1, la quale riporta le informazioni chiave di ogni articolo:

Analisi della letteratura

Tabella 3.1: Raccolta degli articoli scientifici selezionati per la revisione bibliografica

TITOLO	AUTORE	RIVISTA	ANNO DI PUBBLICAZIONE	NOTE	PAESE
A Digital Twin-based on-site quality assessment method for aero-engine assembly	Li, Jinyue Zhao, Gang Zhang, Pengfei Xu, Maocheng Cheng, He Pengfei, Han	Journal of Manufacturing Systems	2023	L'articolo analizza un'applicazione avanzata del Digital Twin nell'assemblaggio dei motori aeronautici, con focus sul monitoraggio qualità in tempo reale tramite sensori, simulazioni e intelligenza artificiale. Pur basandosi su dati principalmente quantitativi, mostra come il DT possa rilevare, classificare e prevedere difetti, offrendo una base coerente per approfondire l'integrazione di variabili qualitative nei sistemi di controllo qualità.	Cina
A BIM-IoT and intelligent compaction integrated framework for advanced road compaction quality monitoring and management	Han, Tao Ma, Tao Fang, Zhou Zhang, Yang Han, Chengjia	Computers and Electrical Engineering	2022	L'articolo descrive un framework che utilizza Digital Twin per migliorare il controllo qualità nelle costruzioni stradali, integrando BIM, IoT e tecnologie di compattazione intelligente. Il sistema consente il monitoraggio in tempo reale del processo di costruzione, l'identificazione automatica dei difetti e l'ottimizzazione dinamica delle operazioni in cantiere.	Cina
Digital Twin of Intelligent Small Surface Defect Detection with Cyber-manufacturing Systems	Wu, Yirui Cao, Hao Yang, Guoqiang Lu, Tong Wan, Shaohua	ACM Transactions on Internet Technology	2023	L'articolo presenta un sistema di controllo qualità basato su Digital Twin per la rilevazione automatica di difetti superficiali nei processi produttivi. Integrando IoT, scanner 3D e algoritmi di deep learning in un'architettura Edge-Cloud, il sistema consente il monitoraggio in tempo reale, l'analisi multimodale e l'ottimizzazione continua della produzione.	Cina
Material twins generation of woven polymer composites based on ResL-U-Net convolutional neural networks	Song, Yingying Qu, Zhaoliang Liao, Haitao Ai, Shigang	Composite Structures	2022	Lo studio propone un Digital Twin per materiali compositi basato su immagini µCT ad alta risoluzione, segmentate tramite reti neurali profonde (ResL-U-Net) per rilevare con precisione fibre e difetti interni. La ricostruzione 3D del materiale, unita a simulazioni FEM, consente di prevedere il comportamento meccanico e ottimizzare il controllo qualità riducendo test fisici.	Cina
Deep Learning-Empowered Digital Twin Using Acoustic Signal for Welding Quality Inspection	Ji, Tao Mohamad Nor, Norzalilah	Sensors	2023	Lo studio propone un sistema di controllo qualità basato su Digital Twin che rileva difetti di saldatura in tempo reale analizzando i segnali acustici raccolti durante il processo. I dati vengono filtrati con trasformate wavelet e classificati tramite modello SeCNN-LSTM, una rete neurale che riconosce i pattern sonori associati ai difetti, raggiungendo un'accuratezza del 91%.	Malaysia

Analisi della letteratura

Digital Twin Model Quality Optimization and Control Methods Based on Workflow Management	Luo, Ruiping Sheng, Buyun Lu, Yingkang Huang, Yuzhe Fu, Gaocai Yin, Xiyan	Applied Sciences (Switzerland)	2023	l'articolo propone una metodologia per il controllo qualità dei modelli Digital Twin, integrando valutazioni e ottimizzazioni in un workflow automatico. Viene utilizzato un sistema a due livelli (macro e micro) per identificare indicatori di bassa qualità e correggerli tramite soluzioni mirate. Il sistema rileva i difetti monitorando 15 indicatori di qualità (es. efficienza, accuratezza, flessibilità); se uno di questi è sotto soglia, viene attivata un'ottimizzazione automatica ispirata al sistema immunitario. Il DT diventa così un sistema auto-valutante e migliorabile, garantendo affidabilità, efficienza e riutilizzabilità del modello digitale.	Cina
A novel inspection technique for electronic components using thermography (Nitect)	Liu, Haochen Tinsley, Lawrence Lam, Wayne Addepalli, Sri Liu, Xiaochen Starr, Andrew Zhao, Yifan	Sensors (Switzerland)	2020	l'articolo propone un sistema di controllo qualità non distruttivo per componenti elettronici, integrando simulazioni FEM e termografia pulsata in un approccio Digital Twin. I difetti strutturali interni vengono rilevati confrontando le simulazioni con immagini termiche reali, analizzate tramite PCA. Il DT permette la diagnosi automatica di difetti come variazioni nel die o nel materiale, migliorando efficienza e affidabilità.	Regno Unito (United Kingdom)
Digital twin-driven machine learning: Ball bearings fault severity classification	Farhat, Mohamed Habib Chiementin, Xavier Chaari, Fakher Bolaers, Fabrice Haddar, Mohamed	Measurement Science and Technology	2021	lo studio propone l'uso del Digital Twin per simulare segnali vibrazionali di cuscinetti difettosi, sostituendo in parte i dati reali nella fase di addestramento dei modelli predittivi. Il DT viene aggiornato per riflettere accuratamente il sistema fisico, consentendo la generazione di dati etichettati realistici. Questi segnali simulati vengono trasformati in feature diagnostiche (es. RMS, kurtosi), utilizzate per addestrare algoritmi di machine learning (MSVM, KNN) in grado di classificare automaticamente il tipo e la gravità del difetto.	Tunisia
Design Technology and AI-Based Decision Making Model for Digital Twin Engineering	Orlova, Ekaterina V	Future Internet	2022	lo studio propone una metodologia strutturata per progettare Digital Twin (DT) capaci di supportare il controllo qualità e la manutenzione preventiva, anche in fase di progettazione. Il DT simula guasti e condizioni anomale in tempo reale e, integrando dati sensoriali e un modello fuzzy, fornisce diagnosi operative intelligenti (riparare, spegnere, ecc.). È un sistema decisionale avanzato che combina simulazione, intelligenza artificiale e logica qualitativa per anticipare e prevenire guasti.	Russia
Improving production efficiency with a digital twin based on anomaly detection	Trauer, Jakob Pfingstl, Simon Finsterer, Markus Zimmermann, Markus	Sustainability (Switzerland)	2021	Lo studio propone un Digital Twin diagnostico applicato a una linea di estrusione di alluminio, che rileva anomalie nella curva di pressione del pistone durante il ciclo produttivo. Il DT utilizza un modello probabilistico basato su Gaussian Process per apprendere il comportamento "normale" del processo e, durante la produzione, confronta in tempo reale le curve misurate con il modello appreso. Se la probabilità che una nuova curva appartenga al comportamento sano è troppo bassa, il sistema classifica il prodotto come anomalo (output binario). I risultati sono visualizzati su un'interfaccia operatore e possono essere usati per attivare dispositivi di scarto automatico. Il sistema è stato testato in un ambiente industriale reale e ha dimostrato alti tassi di accuratezza.	Germania

Analisi della letteratura

A Digital Twin Assisted Framework for Quality Assurance in Mould Manufacturing.	Boettjer, T., Ramanujan, D.	Proceedings of the ASME 2023 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (IDETC-CIE 2023)	2023	Lo studio presenta un framework Digital Twin per l'assicurazione qualità nella produzione di stampi. Il sistema integra dati fisici e virtuali tramite un data lake centralizzato, con l'obiettivo di ridurre rilavorazioni e ispezioni inutili. Sono impiegati due modelli in cascata: il primo, basato su clustering (k-Means) delle forze di taglio, stima l'usura degli utensili; il secondo, un classificatore (Random Forest, MLP, MLR), utilizza le informazioni del primo modello insieme ai parametri geometrici per prevedere la qualità dei pezzi (classificazione binaria: dentro/fuori tolleranza). Il sistema consente di identificare le aree critiche da ispezionare, migliorando efficienza e precisione nel controllo qualità.	Danimarca
In-process surface quality monitoring of the slender workpiece machining with digital twin approach	Lu, K., Li, Z., Longstaff, A.	Journal of Intelligent Manufacturing	2025	Lo studio presenta un Digital Twin ibrido che combina modellazione analitica e approccio data-driven per il monitoraggio in-process della qualità superficiale durante la tornitura di pezzi a bassa rigidità. L'architettura integra un modello predittivo degli errori diametrali e un indicatore per la rilevazione precoce del chatter, basato su trasformata wavelet e sensori di spostamento. Il Digital Twin restituisce un output classificatorio, distinguendo tra condizioni di processo stabili e instabili, fornendo così un'informazione categoriale fondamentale per il controllo qualità in tempo reale. I risultati sperimentali confermano l'efficacia del metodo rispetto alle tecniche tradizionali.	Cina
Integrating Digital Twin Technology to Achieve Higher Operational Efficiency and Sustainability in Manufacturing Systems	Matevž Resman Niko Herakovič Mihael Debevec	Systems	2025	Lo studio presenta un Digital Twin integrato con la linea produttiva, aggiornato in tempo reale tramite sincronizzazione sensori-software, per monitorare e ottimizzare il ciclo produttivo. Il DT simula dinamicamente le operazioni (assemblaggio, incollaggio, ispezione) e indirizza autonomamente i prodotti in base alla qualità rilevata (buono, riparabile, scarto). Nei casi studio, il DT implementa logiche condizionali e reindirizzamento automatico, riducendo gli sprechi fino al 75%. Il sistema gestisce varianti di prodotto (Type_N, Type_B) adattando i parametri e visualizzando inefficienze operative via dashboard interattiva.	Slovenia
Multi-modal recognition control system for real-time robot welding penetration control and quality enhancement	Ji, T. Nor, N.M. Abdullah, A.B.	International Journal of Advanced Manufacturing Technology	2024	Lo studio presenta un Digital Twin multimodale per il controllo in tempo reale della penetrazione nella saldatura robotizzata. Il DT integra sensori visivi e acustici, edge computing e una rete neurale per classificare lo stato della saldatura (incompleta, completa, eccessiva, burn-through). L'output classificatorio guida un controller adattivo model-free che regola corrente e velocità. L'intero ciclo (acquisizione, classificazione, controllo) è sincronizzato dal DT, che funge da replica dinamica del processo fisico per ottimizzare qualità e stabilità.	Malesia
Digital twin and IoT technology for secure manufacturing systems	Lisa Gopal Harbaksh Singh Mounica Panguluri N. Mohankumar Nagendra Panini Challa P. Jayaraman	Measurement: Sensors	2023	Il lavoro propone un Digital Twin basato su IoT per la manutenzione preidattiva e la sicurezza in ambienti manifatturieri. Il sistema raccoglie dati real-time da sensori intelligenti (vibrazione, temperatura, umidità), che vengono elaborati tramite modelli LSTM per classificare lo stato del sistema (normale, stressato, guasto). Il DT integra architetture Big Data (Spark, Hadoop) per l'analisi su larga scala e genera output categoriali per guidare decisioni operative e manutentive. La performance è validata tramite metriche di classificazione (confusion matrix, AUC-ROC).	India

Analisi della letteratura

Digital Twin of Intelligent Small Surface Defect Detection with Cyber-manufacturing Systems	Wu, Yirui Cao, Hao Yang, Guoqiang Lu, Tong Wan, Shaohua	ACM Transactions on Internet Technology	2023	l'articolo presenta un Digital Twin per il rilevamento intelligente di piccoli difetti superficiali, integrato in un'architettura Cyber-Manufacturing con struttura Edge–Cloud. Il DT riceve dati 2D e 3D da sensori, elaborati tramite CNN e Swin-T per estrarre e fondere caratteristiche multimodali. L'output è una classificazione binaria (difetto/no difetto) con bounding box (cioè le coordinate della zona difettosa) e confidence score (cioè quanto il sistema "crede" nella classificazione fatta.), migliorato da post-processing morfologico. Il sistema è validato su dataset industriali, mostrando alte performance in precisione e richiamo. Il DT abilita decisioni operative immediate e aggiornamento continuo del processo.	Cina
Digital Twin for Training Bayesian Networks for Fault Diagnostics of Manufacturing Systems	Toyosi Ademujimi Vittaldas Prabhu	Sensors	2022	l'articolo propone un Digital Twin multilivello per addestrare reti bayesiane (BN) nella diagnosi dei guasti nei sistemi manifatturieri. Il DT è sviluppato tramite co-simulazione tra RobotStudio (livello macchina) e Simio (livello fabbrica), integrando sensori IoT, log testuali (CAPA e MWO) e tecniche NLP. I guasti vengono simulati virtualmente, permettendo interventi strutturali (SIA) per apprendere graficamente le relazioni causa-effetto. Il DT genera dati sintetici per stimare la struttura e i parametri della BN. I KPI (OEE, TH) sono discretizzati e l'output della BN è categoriale (es. guasto attivo/si-no, efficienza bassa/media/alta).	USA
Optimizing a Digital Twin for Fault Diagnosis in Grid Connected Inverters – A Bayesian Approach	Pavol Mulinka Subham Sahoo Charalampos Kalalas Pedro H.J. Nardelli	Conferenza: 2022 IEEE Energy Conversion Congress and Exposition (ECCE 2022)	2022	lo studio presenta un Digital Twin ottimizzato per la diagnosi di guasti in inverter connessi alla rete, basato su modelli LightGBM e deep learning. Il DT è in grado di classificare lo stato del sistema in 5 categorie (normale + 4 tipi di guasto) e fornisce, per ogni osservazione, una distribuzione di probabilità di appartenenza a ciascuna classe. L'ottimizzazione avviene tramite Ray Tune, focal loss e selezione iterativa delle feature. L'output categoriale-probablistico consente una diagnosi affidabile e interpretabile, utile per guidare decisioni automatizzate nella manutenzione predittiva. Validato su dati simulati, il sistema raggiunge un'accuratezza superiore al 99%.	Spagna, Danimarca e Finlandia.
An Efficient Fault Diagnosis Framework for Digital Twins Using Optimized Machine Learning Models in Smart Industrial Control Systems	Zayed, S.M. Attiya, G. El-Sayed, A. Sayed, A. Hemdan, E.E.-D.	International Journal of Computational Intelligence Systems	2023	l'articolo propone un framework di diagnosi dei guasti basato su Digital Twin per sistemi di controllo industriali intelligenti, utilizzando un modello Decision Tree (CART) ottimizzato con Flower Pollination Algorithm (FPA). Il DT riceve dati da sensori in tempo reale e fornisce output categoriali sul tipo di guasto. Le performance del sistema sono valutate con SVM, KNN, RF e CART, ma FPA-CART risulta il più accurato (fino al 96,8%). Il DT integra la selezione ottimale delle feature per migliorare la diagnosi e ridurre i falsi allarmi.	Egitto
Fault diagnosis of reducers based on digital twins and deep learning	Liu, Weimin Han, Bin Zheng, Aiyun Zheng, Zhi Chen, Shujun Jia, Shikui	Scientific Reports	2024	l'articolo propone un Digital Twin avanzato per la diagnosi di guasti nei riduttori, combinando simulazioni fisiche e deep learning. Viene utilizzato FDGAN per trasferire la distribuzione dei dati reali sui dati simulati, migliorando la coerenza tra ambiente virtuale e fisico. L'output diagnostico è ottenuto tramite MobileViG, un classificatore leggero e accurato basato su Vision Gated MLP, che riconosce diverse tipologie di guasto a partire da spettrogrammi vibrazionali. Il sistema restituisce un output categoriale (es. nessun guasto, fouling, perdita, blocco) con accuratezza elevata (fino al 99,5%), supportando il controllo qualità in tempo reale.	Cina

Analisi della letteratura

Using Digital Twins for Fault Detection and Root Cause Analysis in Mechanical Systems	Shashank Pasupuleti	International Scientific Journal of Engineering and Management (ISJEM)	2024	<p>l'articolo propone un framework basato su Digital Twin per la rilevazione dei guasti e l'analisi delle cause radice nei sistemi meccanici. Il DT integra modelli fisici e dati real-time da sensori per monitorare continuamente lo stato del sistema. Le tecniche di rilevamento impiegate includono regole a soglia (threshold-based), metodi stocastici e algoritmi di machine learning, che convergono per classificare lo stato operativo del sistema in categorie (es. normale, guasto lieve, guasto critico). L'approccio consente anche simulazioni predittive per valutare scenari futuri e migliorare la gestione della manutenzione.</p>	Francia
Digital Twin for Fault Detection and Diagnosis of Building Operations: A Systematic Review	F. Hodavand L.J. Ramaji N. Sadeghi	Buildings	2023	<p>l'articolo analizza l'impiego del Digital Twin (DT) per la rilevazione e diagnosi dei guasti (FDD) negli edifici, con particolare focus sui sistemi HVAC. Il DT è una replica virtuale sincronizzata del sistema reale, alimentata da dati provenienti da sensori IoT, BIM, BMS e CMMS. Vengono identificati tre approcci FDD: analitico (modelli fisici), knowledge-based (regole esperte, fuzzy logic), e data-driven (machine learning, anche deep). L'output del DT è generalmente categoriale, sotto forma di etichette diagnostiche che indicano lo stato del componente (es. "guasto valvola", "filtro ostruito", "normale"), o livelli di severità (es. lieve, critico). Questo consente un controllo qualità automatizzato e decisioni di manutenzione predittiva.</p>	Iran, USA
A digital twin approach to system-level fault detection and diagnosis for improved equipment health monitoring	Nguyen, Tat Nghia Poncirolli, Roberto Bruck, Paul Esselman, Thomas C. Rigatti, Joseph A. Vilim, Richard B.	Annals of Nuclear Energy	2022	<p>l'articolo propone un Digital Twin physics-based per la diagnosi di guasti a livello di sistema in impianti nucleari, con focus sul sistema di preriscalo dell'acqua di alimentazione. Il DT utilizza modelli parametrici derivati dalle leggi di conservazione (massa, energia, quantità di moto) per costruire sensori virtuali (type-I e type-II), che compensano la carenza di sensori fisici. L'output è una classificazione categoriale (guasti vs. non guasti), basata su residui diagnostici associati a modelli fisici, e interpretata tramite un motore di inferenza causale (reasoning deterministico o probabilistico). Il DT fornisce diagnosi spiegabili e tracciabili, adatte alla manutenzione predittiva e alla validazione degli eventi post-startup.</p>	USA
Fault Detection in Active Magnetic Bearings Using Digital Twin Technology	Hu, Y. Taha, O.W. Yang, K.	Applied Sciences (Switzerland)	2024	<p>Lo studio propone un Digital Twin (DT) per la rilevazione dei guasti negli Active Magnetic Bearings (AMB), integrando modelli fisici e reti neurali convoluzionali (CNN). Il DT simula in tempo reale la dinamica del sistema AMB (campo magnetico, forze, controllo PID), sincronizzandosi con i dati reali acquisiti da sensori (posizione, corrente, controllo). Il sistema genera vibration images da segnali multi-sensore che, processati tramite CNN multi-branch, permettono una classificazione accurata di guasti come imbalance e misalignment. Il DT restituisce, per ogni campione di dati, una previsione categoriale del guasto rilevato, assegnando ciascun input a una specifica classe diagnostica abilitando un monitoraggio continuo e la manutenzione predittiva.</p>	Cina
Fault Diagnosis for Reducers Based on a Digital Twin	Liu, W. Han, B. Zheng, A. Zheng, Z.	Sensors	2024	<p>Il DT sviluppato per la diagnosi di guasti nei riduttori integra modelli fisici, virtuali e ambientali (MG, MA, ME) all'interno di una piattaforma interattiva costruita in Unity3D. Il sistema combina dati simulati (da FEM) e reali (sensori di vibrazione, temperatura, rumore), generando segnali che vengono analizzati tramite trasformate nel dominio del tempo e della frequenza. La similarità tra segnali reali e digitali è valutata con hashing differenziale e distanza di Hamming. L'output del DT è una classificazione categoriale del tipo di guasto (es. allentamento, disallineamento, rottura) o dello stato operativo del sistema (OK/fault), abilitando diagnosi automatica e localizzazione dei difetti.</p>	Cina

Analisi della letteratura

Digital Twin of Rail for Defect Analysis	Ahmad, Mutz, Werth	Proceedings of the 8th International Conference on Virtual and Augmented Reality Simulations	2024	Lo studio propone la creazione di un Digital Twin procedurale del binario ferroviario basato su dati geografici OpenStreetMap e dati di ispezione ultrasonica. Il modello genera una rappresentazione tridimensionale dettagliata del tracciato ferroviario, su cui vengono sovrapposti difetti categorizzati secondo la normativa europea EN 16729-1. Il Digital Twin restituisce un output che associa a ogni difetto una variabile categorica identificativa (tipologia di difetto) e parametri geometrici continui (lunghezza, larghezza, profondità). L'integrazione in ambiente di realtà aumentata (AR) consente un'ispezione interattiva e accurata, migliorando l'efficienza nella manutenzione ferroviaria.	Germania
Welding robot automation technology based on digital twin	Yuhui Kang e Rongshang Che	Frontiers in Mechanical Engineering	2024	Lo studio propone un sistema avanzato per il controllo della qualità della saldatura robotizzata, basato sull'integrazione di tecnologie Digital Twin e algoritmi di deep learning. Il modello sviluppato sfrutta il filtro di correlazione kernel (KCF) migliorato per il tracciamento automatico della traiettoria di saldatura e l'algoritmo YOLOX-s per la rilevazione e la classificazione dei difetti. Ogni difetto viene categorizzato in una delle quattro classi predefinite (spatter, weld bead defects, porosity, slag inclusion), stabilendo quindi un output strutturato su variabili categoriche. Il Digital Twin restituisce una rappresentazione tridimensionale arricchita da informazioni di qualità.	Cina
A digital twin approach based method in civil engineering for classification of salt damage in building evaluation	Guzmán-Torres, J.A. Dominguez-Mota, F.J. Alonso Guzmán, E.M. Tinoco-Guerrero, G. Tinoco-Ruiz, J.G.	Mathematics and Computers in Simulation	2025	Lo studio propone un sistema per la classificazione automatica di danni da sali (Salt damage) nelle infrastrutture civili basato su CNN e Digital Twin di livello 2 (digital shadow). Vengono confrontate sei architetture (tra cui VGG-16, ResNet50 e MobileNet) attraverso transfer learning. La VGG-16, selezionata per la sua stabilità ed efficienza computazionale, viene ottimizzata tramite fine-tuning (dropout, L2 regularization, SGD). Il modello finale raggiunge un'accuratezza del 94,75% in validazione e 98,44% nel test, con classificazione binaria (danneggiato/non danneggiato). L'approccio migliora la diagnosi strutturale riducendo soggettività e costi ispettivi, con possibili applicazioni su robot e droni per ispezioni autonome	Mexico
Few-Shot and Portable 3D Manufacturing Defect Tracking with Enterprise Digital Twins Based Mixed Reality	Tan, Iyyong Send mail to Tan Y.; Banerjee, Bhaskar Ranjan, Rishi	Lecture Notes in Computer Science	2025	Lo studio propone un sistema per la rilevazione automatica di difetti strutturali tramite digital twin 3D e segmentazione di point cloud acquisiti con dispositivi AR (es. HoloLens 2). Utilizzando istogrammi di segmentazione e modelli di machine learning leggeri (GMM, SVM, MLP), il sistema classifica difetti in modo binario e ne stima la severità con metriche statistiche (Wasserstein, Hellinger). Il modello non richiede grandi dataset specifici (few-shot) e funziona anche in ambienti industriali complessi. L'integrazione con il DT consente il confronto continuo con la geometria ideale per rilevare anomalie geometriche anche minime.	India
A local digital twin approach for identifying, locating and sizing cracks in CHS X-joints subjected to brace axial loading	Cheok, Evan Wei Wen Qian, Xudong Send mail to Qian X.; Chen, Cheng Quek, Ser Tong Si, Michael Boon Ing	Engineering Structures	2024	l'articolo propone un Digital Twin locale per la diagnosi remota e autonoma di cricche in giunti saldati CHS, basato esclusivamente su letture di deformazione (strain). Il DT utilizza un modello surrogato composto da tre modelli ML sequenziali (due reti neurali per identificazione generando un output binario e localizzazione, e una regressione GPR per la quantificazione). I dati per l'addestramento derivano da una libreria di modelli FEM generati rapidamente tramite tecniche ROM (substructuring). La soluzione è validata sperimentalmente su due gemelli fisici sottoposti a carichi ciclici, dimostrando alta accuratezza e tempestività nel rilevare e misurare le cricche. Il lavoro evidenzia l'integrazione efficace di ML e ROM all'interno del DT.	Singapore

La strategia di ricerca si è basata su una combinazione di parole chiave quali: “*Digital Twin to detection*”, “*fault diagnosis*”, “*categorical output*”, “*quality control*”, “*defect classification*”, “*smart manufacturing*”, oltre a espressioni più specifiche come “*tool wear monitoring*”, “*welding defect detection*”, “*surface quality monitoring*” e “*in-process inspection*”. In alcuni casi, sono stati inclusi articoli che, pur non trattando direttamente variabili categoriche, descrivevano processi in cui l’output del sistema diagnostico era espresso in forma discreta o simbolica, come etichette di difettosità.

I criteri di inclusione hanno privilegiato ricerche che adottano un approccio integrato tra ambiente fisico e virtuale, rappresentando un sistema reale tramite un Digital Twin. Inoltre, sono stati selezionati gli studi che applicano tecniche di classificazione per la valutazione della qualità, utilizzando sia modelli simbolici che reti neurali o altre architetture di deep learning. Infine, sono stati inclusi i lavori che esplicitano l’utilizzo di dataset etichettati per la diagnosi di difetti o per il monitoraggio dello stato di usura degli utensili, come nel caso del modello a cascata sviluppato per la classificazione dei difetti nei componenti stampati [24].

Sono stati invece esclusi gli studi con finalità puramente simulativa o predittiva non legate a specifici meccanismi di controllo qualità, nonché quelli che si limitavano a discutere il concetto di DT senza proporre implementazioni o casi studio concreti. Inoltre sono stati privilegiati gli articoli che descrivevano approcci riconducibili a scenari industriali reali, come la produzione di componenti saldati, inserti per stampi, cuscinetti o riduttori, nei quali la qualità del prodotto è valutata anche tramite misurazioni qualitative.

Infine, per garantire l’attendibilità dei contributi selezionati, è stato effettuato un controllo incrociato delle fonti e sono stati considerati solo articoli pubblicati su riviste scientifiche sottoposte a revisione tra pari o conferenze di settore rilevanti nel campo della manifattura digitale.

3.2 Analisi descrittiva della letteratura raccolta

La letteratura analizzata mostra una crescita costante dell’interesse per l’uso del Digital Twin nel controllo qualità negli ultimi anni, con una concentrazione significativa di pubblicazioni tra il 2020 e il 2025. Questo andamento riflette l’accelerazione della digitalizzazione industriale post-pandemia e l’adozione sempre più diffusa di modelli di simulazione ibrida per migliorare l’efficienza e la sostenibilità dei processi produttivi.

Per comprendere la dinamica evolutiva della ricerca accademica sull’impiego del Digital Twin nel controllo qualità, è stata condotta un’analisi temporale delle pubblicazioni selezionate. Il grafico a barre riportato in Figura 3.2 mostra la distribuzione degli articoli analizzati in base al loro anno di pubblicazione.

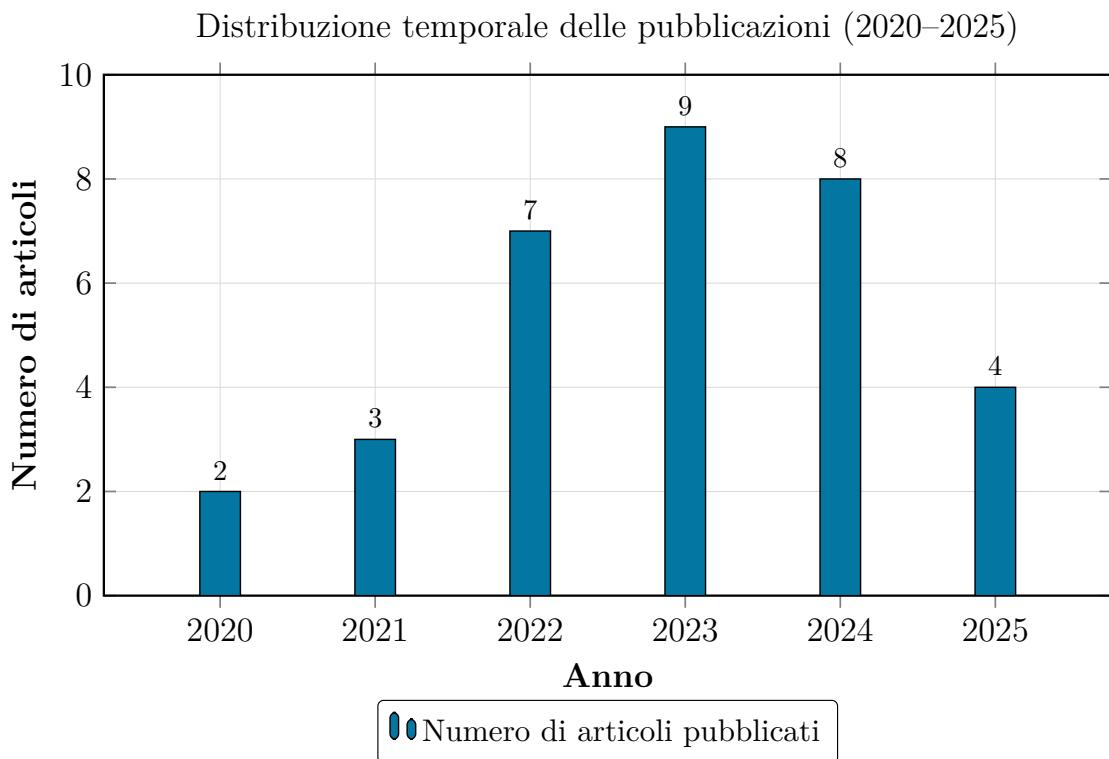


Figura 3.1: Distribuzione temporale degli articoli selezionati sul tema del *Digital Twin* applicato al controllo qualità.

Come evidenziato, il numero di pubblicazioni sul tema ha registrato un incremento significativo a partire dal 2022, con un picco tra il 2023 e il 2024. Questo andamento conferma l’interesse crescente verso i DT come strumento strategico per l’ottimizzazione della qualità in ambito manifatturiero, edilizio ed energetico. Il dato relativo al 2025, pur se parziale, suggerisce un consolidamento della tendenza, alimentato anche dalla maggiore diffusione di approcci ibridi e di modelli basati su deep learning, edge computing e realtà aumentata. La concentrazione temporale delle pubblicazioni negli ultimi tre anni rafforza quindi la rilevanza attuale del tema all’interno del panorama scientifico e industriale.

Come mostrato in Figura 3.3, dal punto di vista geografico la letteratura analizzata mostra una forte concentrazione di contributi provenienti da un numero limitato di aree geografiche, riflettendo la distribuzione globale della ricerca e dello sviluppo tecnologico nel campo dei Digital Twin. In particolare, la maggior parte degli articoli è stata prodotta in paesi asiatici, come Cina e Malesia, seguiti da Europa (Germania, Spagna, Regno Unito, Danimarca) e Nord America. Sono inoltre presenti studi rilevanti provenienti da altre aree, tra cui India, Iran, Tunisia e Finlandia.

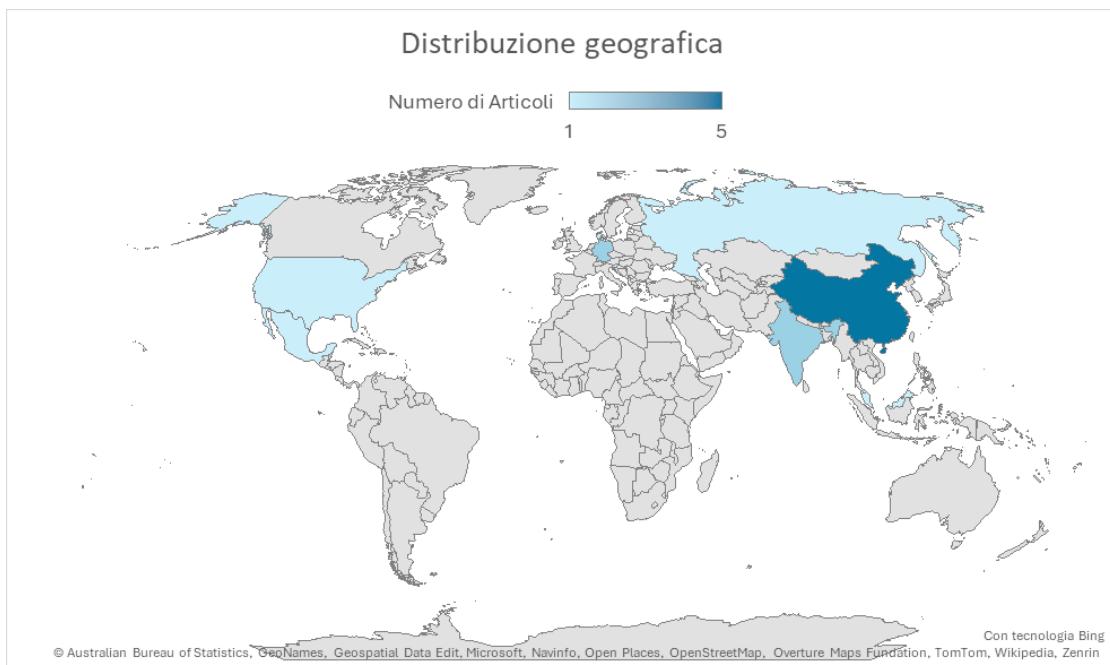


Figura 3.2: Distribuzione geografica degli articoli pubblicati

La figura mostra in modo sintetico il numero di articoli per ciascun paese, evidenziando chiaramente come la Cina rappresenti il principale centro di produzione scientifica sul tema, seguita da altri poli accademici e industriali distribuiti in modo disomogeneo nel panorama globale. La legenda (1–5) indica il numero di articoli pubblicati, con tonalità di colore più scure associate a una maggiore produzione scientifica.

Un ulteriore aspetto rilevante emerso dall’analisi bibliografica riguarda l’individuazione degli autori che risultano maggiormente attivi nella pubblicazione di studi sul tema dei Digital Twin applicati al controllo qualità. In particolare, tra i ricercatori più ricorrenti figurano Wu Yirui, Cao Hao, Yang Guoqiang, Lu Tong e Wan Shaohua, che compaiono in più contributi, sia come autori principali sia come coautori.

Questa concentrazione suggerisce l’esistenza di gruppi di ricerca strutturati e specializzati sul tema, spesso collegati a istituzioni universitarie di riferimento nei rispettivi Paesi.

Infine, l’analisi congiunta dei casi studio raccolti e della letteratura scientifica selezionata, confermano che la tecnologia del Digital Twin viene ormai applicata in svariati settori, dall’aerospaziale alla sanità, dalle smart city alle smart factory, con l’obiettivo di monitorare sistemi fisici e migliorare il controllo qualità. Emergono molteplici ambiti applicativi in cui il Digital Twin viene impiegato con finalità di

controllo qualità, dimostrando una notevole eterogeneità in termini di settori e tipologie di dati coinvolti.

Un ambito rilevante riguarda la saldatura robotizzata, dove il DT viene integrato con sistemi di classificazione multimodale basati su segnali visivi e acustici. Tali modelli permettono di monitorare in tempo reale la penetrazione del giunto saldato e di valutare la qualità complessiva del cordone, favorendo un processo decisionale automatico e tempestivo. In questi scenari, la combinazione tra sensoristica integrata e modelli intelligenti consente una classificazione accurata dello stato di saldatura (ad esempio: completa, incompleta, sovrapposta), come documentato in più casi studio.

Un altro ambito di applicazione è rappresentato dalla manifattura meccanica e dei componenti metallici, in cui il DT viene impiegato per il monitoraggio dell’usura degli utensili, il controllo di processo e la diagnosi predittiva di guasti in riduttori. I dati raccolti (es. segnali di forza, pressione, vibrazioni) vengono elaborati mediante tecniche di classificazione come Random Forest, Clustering o modelli gaussiani, al fine di valutare lo stato di efficienza operativa e prevenire scarti o fermate improvvise.

Meritano attenzione anche la produzione industriale e le smart factory, ovvero ambienti produttivi altamente automatizzati e interconnessi, in cui il DT gioca un ruolo chiave come strumento di coordinamento e ottimizzazione. In questi contesti, il DT non si limita a replicare virtualmente le condizioni operative della linea produttiva, ma agisce come una vera e propria piattaforma decisionale, capace di monitorare costantemente lo stato dei macchinari e dei prodotti, rilevare anomalie, e suggerire azioni correttive in tempo reale. Attraverso l’integrazione con sistemi di raccolta dati (sensori, PLC, SCADA) e modelli logici condizionali o predittivi, il DT consente di classificare dinamicamente lo stato dei prodotti lungo il flusso produttivo (es. prodotto conforme, riparabile, da scartare), migliorando così la tracciabilità e la gestione degli scarti. Inoltre, la capacità di eseguire simulazioni “what-if” permette di testare scenari alternativi senza interrompere la produzione, contribuendo a una maggiore flessibilità, riduzione dei costi e miglioramento della qualità complessiva.

Inoltre, si segnala un’applicazione particolarmente significativa nel settore medico e biomedicale, come nel caso del monitoraggio ambientale all’interno di incubatrici neonatali. In questo contesto, il DT elabora dati sensoriali sia quantitativi (temperatura, umidità, ossigeno) sia qualitativi (input linguistici), ricorrendo a modelli fuzzy per formulare giudizi diagnostici sullo stato del sistema, migliorando l’affidabilità e la sicurezza operativa.

Oltre ai quattro ambiti già evidenziati, esistono numerosi altri contesti applicativi in cui i Digital Twin stanno dimostrando la loro versatilità come strumento per il controllo qualità. L’impiego del DT per il controllo qualità risulta particolarmente

rilevante è quello degli impianti energetici e dei sistemi elettronici complessi, come gli inverter connessi alla rete o i circuiti termoidraulici di centrali nucleari. In questi contesti, pur predominando l'uso di dati numerici continui (come tensioni, correnti, temperature o pressioni), si osserva una crescente integrazione di variabili categoriche per rappresentare gli stati di guasto o le condizioni operative. Ad esempio, nello studio di Mulinka et al. (2022) sull'ottimizzazione di un DT per la diagnosi di inverter di potenza, il problema viene formalizzato come una classificazione multi-classe, in cui ciascun output del modello rappresenta una variabile categorica che distingue tra comportamenti normali e diverse tipologie di guasto (es. line-to-line fault, sensor fault, grid fault). Tali etichette discrete non solo facilitano l'interpretabilità dei risultati da parte degli operatori, ma permettono anche di attivare interventi mirati in modo rapido e automatizzato. Questo approccio evidenzia come anche in domini fortemente numerici, l'adozione di variabili simboliche possa contribuire a una gestione più efficiente, semantica e reattiva della qualità operativa.

Nel campo dell'ingegneria civile e delle infrastrutture, i Digital Twin stanno rivoluzionando le attività di ispezione e manutenzione, migliorando la qualità e la sicurezza strutturale. Un DT di una struttura (ponte, edificio, rete infrastrutturale) consente infatti un monitoraggio continuo dello stato dell'opera e un'analisi predittiva dei degradi strutturali. Un esempio nell'ambito civile è la gestione di edifici intelligenti (smart building), presentato nell'articolo "*Digital Twin for Fault Detection and Diagnosis of Building Operations: A Systematic Review*": i DT vengono utilizzati per sorvegliare impianti come i sistemi HVAC (Heating, Ventilation and Air Conditioning) al fine di garantire condizioni ambientali ottimali e rilevare guasti che potrebbero influire sul comfort o sui consumi energetici. Attraverso modelli data-driven e integrazione con i sistemi di building management, il gemello digitale del fabbricato aiuta i facility manager a diagnosticare malfunzionamenti (ad es. filtri ostruiti, cali di prestazione) e ad attuare manutenzioni predittive, migliorando sia la qualità dell'ambiente interno sia l'efficienza operativa dell'edificio[11]. Sebbene la letteratura non specifichi esplicitamente l'uso di variabili categoriche, in scenari come gli smart building risulta plausibile l'impiego di etichette discrete per rappresentare stati diagnostici (es. filtro ostruito, prestazione degradata, funzionamento normale), al fine di agevolare l'automazione delle decisioni e l'attivazione di interventi manutentivi.

Anche il settore dei trasporti beneficia dell'adozione dei Digital Twin per il controllo qualità e la manutenzione predittiva delle infrastrutture. Un caso rappresentativo è quello ferroviario, dove la complessità e l'estensione della rete impongono strumenti innovativi per garantire l'affidabilità dell'infrastruttura e la sicurezza operativa. In questo ambito, i DT consentono il monitoraggio in tempo reale dello stato dei binari e l'analisi predittiva dei difetti. Un esempio è dato da un modello

procedurale del binario ferroviario, costruito integrando dati GIS (es. OpenStreetMap) con ispezioni tramite sensori visivi e a ultrasuoni. I difetti rilevati, come cricche o usure, vengono classificati e geolocalizzati nel DT, rendendone visibile la posizione e la gravità tramite realtà aumentata. Questa rappresentazione semantica supporta le decisioni di intervento e migliora l'efficienza manutentiva [25]. In questo contesto, i difetti rilevati vengono classificati secondo tipologie discrete (es. cricca, usura, disallineamento), che costituiscono variabili categoriche impiegate per supportare il processo decisionale sulla priorità e modalità degli interventi manutentivi.

Questa varietà settoriale conferma la versatilità del Digital Twin come tecnologia abilitante per il controllo qualità, capace di integrare simulazione digitale, dati reali e logiche intelligenti, in cui l'utilizzo congiunto di variabili numeriche e categoriche risulta cruciale per una diagnosi efficace e reattiva.

3.3 Classificazione sistematica dei casi di studio

Per una comprensione strutturata dell'eterogeneità metodologica e applicativa riscontrata nei casi di studio analizzati, si è proceduto a una classificazione sistematica secondo cinque dimensioni chiave: tipologia dei dati utilizzati, modelli di classificazione, campo di applicazione, struttura architettonica del Digital Twin e modalità di costruzione dell'output categorico. Questa riclassificazione aiuta a mettere in luce alcune ricorrenze, differenze significative e tendenze comuni tra i casi esaminati, offrendo spunti utili e concreti per la progettazione del modello sperimentale che verrà presentato nei capitoli successivi.

3.3.1 Tipologia di dati utilizzati

I casi analizzati evidenziano una notevole varietà nei tipi di dati utilizzati, che possono essere distinti in quattro grandi categorie:

1. Dati quantitativi puri

Si tratta di segnali di pressione, corrente, vibrazioni, temperatura, forze di taglio o dimensioni geometriche. Sono tipici di ambienti altamente sensorizzati (es. estrusione di profili in alluminio, controllo della profondità di penetrazione nella saldatura). Tra i casi considerati rientrano, ad esempio, quelli relativi alla diagnostica di riduttori, alla stima della vita utile residua (RUL) in sistemi rotanti, ovvero prevedere quando un componente rotante inizierà a guastarsi, e alla previsione di guasti in ambienti produttivi complessi basati su misure vibro-acustiche.

2. Dati qualitativi o simbolici

Includono input descrittivi di natura non numerica, come classi diagnostiche predefinite, etichette semantiche o istruzioni in linguaggio naturale. Un esempio emblematico è rappresentato dall'incubatrice neonatale intelligente citato nel paragrafo precedente, in cui le condizioni ambientali vengono valutate attraverso variabili linguistiche interpretate mediante logiche fuzzy. Fanno parte di questa categoria anche quei sistemi che associano in automatico etichette simboliche agli stati operativi del processo, come “buono”, “riparabile” o “non conforme”, fornendo una semantica immediatamente comprensibile e utile per il processo decisionale.

3. Dati multimodali

Comprendono combinazioni di più sorgenti eterogenee (es. immagini + suoni + parametri numerici), utilizzati in sistemi di riconoscimento visivo o acustico avanzato. Rientrano qui il modello MME-BMR per la saldatura, che integra immagini e dati acustici, e il sistema per la classificazione tridimensionale dei difetti mediante segmentazione ricorsiva e point cloud 3D.

4. Dati simulati o sintetici

Sono presenti in casi in cui i dati reali sono incompleti o costosi da acquisire. Il DT è costruito su scenari “what-if” o dataset simulati per supportare inferenze su guasti o stati futuri (es. stampi ad alta precisione, Digital Twin predittivo nella robotica). Questo approccio è stato adottato, ad esempio, per addestrare modelli bayesiani e deep learning su superfici difettose o per generare vibrogrammi sintetici utili alla classificazione, con l'obiettivo di superare la scarsità di dati etichettati.

La ricchezza delle fonti informative conferma l'importanza di progettare DT capaci di gestire dati eterogenei, sia per migliorare la precisione diagnostica, sia per costruire modelli più interpretabili e robusti.

3.3.2 Tecniche di classificazione adottate

Le tecniche di classificazione osservate sono altrettanto diversificate e possono essere classificate secondo il grado di complessità e interpretabilità:

1. Modelli simbolici e basati su regole

Includono logiche fuzzy, regole condizionali del tipo Regole “if-then” e alberi decisionali interpretabili. Questi modelli sono costruiti esplicitamente dall'uomo, spesso tramite conoscenze esperte, e permettono una tracciabilità diretta delle decisioni prese. Proprio per questo, trovano largo impiego in contesti

ad alto rischio o dove è essenziale giustificare ogni scelta, come in ambito biomedicale o aerospaziale. Un esempio tipico è rappresentato da sistemi di produzione che usano logiche condizionali per prendere decisioni sulla base di test qualitativi o soglie prestabilite.

2. Modelli statistici e probabilistici

Si basano su metodi inferenziali e sulla teoria della probabilità per gestire l'incertezza. Ne fanno parte i Gaussian Process (utili per modellare relazioni non lineari e fornire stime con intervalli di confidenza), le Bayesian Network (reti che rappresentano relazioni causali tra variabili) e tecniche di clustering come il K-Means. Questi modelli sono adatti in contesti dove i dati presentano variabilità elevata o rumore, come nell'estruzione di profili metallici o nel monitoraggio di componenti mobili su linee robotiche.

3. Tecniche di machine learning supervisionato

Comprendono algoritmi come Random Forest, Support Vector Machines (SVM), LightGBM o LSTM. Sono modelli addestrati su dataset etichettati e molto efficaci nella classificazione multi-classe. Tra i casi studiati si evidenzia l'uso di LightGBM per classificare guasti negli inverter di rete, LSTM per la manutenzione predittiva, e Random Forest ottimizzate da algoritmi bio-ispirati (come il Flower Pollination Algorithm) per classificazioni avanzate in ambito manifatturiero.

4. Modelli multimodali e deep learning

Questi modelli avanzati integrano diverse sorgenti informative (es. segnali acustici, immagini, vibrazioni) e utilizzano reti neurali profonde, come CNN o architetture ibride (es. CNN–GNN), per apprendere automaticamente caratteristiche complesse dai dati. Sebbene potenti e versatili, richiedono grandi quantità di dati per l'addestramento e risorse computazionali elevate. Un esempio è il modello MME-BMR, che fonde dati audio-visivi per classificare la qualità della saldatura. Altri casi includono l'uso di YOLOX-s per la segmentazione automatica dei difetti su cordoni saldati o MobileVig per segnali rumorosi.

Nel complesso, la tendenza è verso un'integrazione di metodi interpretabili con modelli ad alta capacità predittiva, soprattutto nei contesti produttivi dove è richiesto sia il supporto decisionale automatico che la tracciabilità delle decisioni.

3.3.3 Ambito di applicazione

I casi di studio coprono una gamma di settori eterogenei, riflettendo la trasversalità del concetto del Digital Twin nel controllo qualità: Nel settore biomedicale, il

Digital Twin viene utilizzato per ottimizzare le condizioni ambientali in incubatrici intelligenti, garantendo il monitoraggio e l'adattamento in tempo reale per mantenere condizioni ideali per la crescita cellulare. In ambito metalmeccanico, invece, trova applicazione nel controllo della qualità dei componenti metallici, attraverso il monitoraggio di parametri come pressione, flussi e processi di saldatura, al fine di garantire precisione e integrità.

Nel contesto della produzione industriale avanzata, il Digital Twin è impiegato in ambienti di smart manufacturing, dove integra dati provenienti da macchine, operatori e condizioni ambientali per ottimizzare i flussi di produzione e migliorare la qualità dei prodotti finali. In ambito elettronico ed energetico, il Digital Twin è fondamentale per la diagnostica avanzata di inverter e sistemi a rete, permettendo il monitoraggio in tempo reale delle performance e la diagnosi di eventuali guasti, con conseguente miglioramento dell'affidabilità del sistema.

Nel campo della robotica e dell'automazione, il Digital Twin supporta la supervisione distribuita, la manutenzione predittiva e l'adattamento in tempo reale dei sistemi robotici, incrementando l'efficienza operativa e riducendo i tempi di inattività. Infine, nel settore edilizio e delle infrastrutture, il Digital Twin viene utilizzato per la valutazione strutturale, il monitoraggio dei guasti nei giunti e la diagnosi di difetti in strutture civili, con l'obiettivo di garantire la sicurezza e ottimizzare i processi di manutenzione.

Questo ampio spettro conferma che la logica di classificazione categoriale trova riscontro in domini anche molto diversi tra loro, con requisiti e vincoli specifici.

3.3.4 Struttura del Digital Twin

Come si può osservare, in alcuni di questi casi il Digital Twin assume un ruolo prettamente descrittivo, limitandosi a raccogliere, aggregare e visualizzare dati provenienti dal sistema fisico, spesso come supporto alla supervisione umana. In altri casi, invece, si configura come un sistema intelligente, dotato di logiche di classificazione, algoritmi di machine learning o modelli decisionali capaci di elaborare attivamente le informazioni e restituire una valutazione qualitativa o un'azione operativa. Questa distinzione tra DT “passivo” e DT “intelligente” rappresenta una delle principali traiettorie evolutive del paradigma: mentre il primo agisce come specchio digitale del processo fisico, il secondo diventa un agente attivo e autonomo, in grado di classificare gli stati del sistema e contribuire direttamente alle decisioni di qualità.

La Tabella 3.2 sintetizza le differenze principali tra le due tipologie:

Tabella 3.2: Differenze tra Digital Twin Passivo e Digital Twin Intelligente.

Caratteristica	Digital Twin Passivo	Digital Twin Intelligente
Funzione principale	Riproduce lo stato del sistema fisico in modo fedele e aggiornato	Supporta il processo decisionale e può intervenire direttamente sul sistema fisico
Tipo di analisi	Descrittiva o storica	Predittiva (anticipa anomalie) e prescrittiva (suggerisce o attiva azioni correttive)
Modalità di aggiornamento	Sincronizzazione periodica, spesso non in tempo reale	Aggiornamento continuo e dinamico, adattivo rispetto al contesto operativo
Interazione con l'ambiente	Unidirezionale: riceve dati dal mondo fisico ma non agisce su di esso	Bidirezionale: scambia informazioni e può agire in tempo reale sul mondo fisico
Reattività	Limitata o assente	Elevata: reagisce a eventi, deviazioni e segnali critici
Tecnologie integrate	Sensori, SCADA, modelli geometrici statici	Intelligenza artificiale, machine learning, inferenza causale, ottimizzazione
Esempi di applicazione	Monitoraggio di stato, rappresentazione grafica di impianti	Controllo qualità in tempo reale, manutenzione predittiva, rilevamento e risposta a guasti

In questo progetto di tesi, l'attenzione si è concentrata su questa seconda tipologia, ovvero su quei DT che, attraverso l'integrazione di tecniche di classificazione e modelli categoriali, sono in grado di produrre un output diagnostico automatizzato, trasformando la replica digitale in uno strumento decisionale concreto.
Dal punto di vista architettonale, si identificano tre modelli principali di implementazione:

- DT basato su dati reali in tempo reale: in cui sensori fisici applicati al sistema reale alimentano il gemello digitale, abilitando un monitoraggio continuo.
- DT costruiti su dati simulati: utili per analisi predittive, studi di sensitività o quando il sistema reale non è completamente osservabile.
- Architetture ibride Edge–Cloud–IoT–AI: nei casi più avanzati, l'analisi e la classificazione avvengono in modo distribuito: i dati vengono raccolti in periferia (Edge), elaborati in cloud con supporto di AI, e reintegrati nel loop operativo. Questa configurazione è frequente nei casi di supervisione distribuita o ispezione in ambienti complessi.

La struttura architetturale ha implicazioni dirette sulla latenza del sistema, la scalabilità e l'efficienza del processo decisionale [6][17].

3.3.5 Modalità di costruzione dell'output categorico

In tutti i casi analizzati, l'output del DT assume una forma categorica, ma le modalità di generazione variano sensibilmente:

1. Etichette discrete predefinite

Si tratta di classi assegnate in modo deterministico, costruite a priori sulla base della conoscenza del processo o delle specifiche di qualità. Queste etichette vengono impiegate sia in modelli simbolici, sia in classificatori supervisionati, e rappresentano stati qualitativi come “difettoso/conforme”, “penetrazione eccessiva”, “bruciatura” o “saldatura incompleta”. La loro funzione è quella di semplificare la diagnosi, trasformando misurazioni complesse in categorie chiare e azionabili [4].

2. Etichette fuzzy

Rappresentano categorie non mutuamente esclusive, in cui uno stesso stato può appartenere contemporaneamente a più classi con diversi gradi di appartenenza. Questo approccio si rivela particolarmente utile in contesti dove le condizioni operative non sono facilmente separabili in modo netto. Ad esempio, un cordone di saldatura potrebbe appartenere contemporaneamente, ma con gradi diversi, alle classi “accettabile” (0.7) e “borderline” (0.4). Il grado di appartenenza (membership) è espresso con valori tra 0 e 1 e consente di gestire l'incertezza e l'ambiguità tipiche delle valutazioni qualitative [21].

3. Output appreso da modelli AI

In presenza di modelli di deep learning o machine learning supervisionato, l'output categoriale non è definito a priori, ma viene appreso dai dati durante la fase di addestramento. In questi casi, la classificazione finale viene determinata sulla base delle probabilità associate alle classi, restituite dal modello in fase di inferenza. È una modalità tipica nei sistemi basati su reti neurali convoluzionali (CNN), LightGBM o LSTM, in cui il DT si affida a una soglia decisionale per etichettare ogni osservazione come “conforme” o “difettosa” (es. se la probabilità di “difettoso” supera 0.6, si classifica come tale). Questi modelli, pur molto precisi, pongono sfide in termini di interpretabilità [17][15].

4. Output collegato a logiche operative

In alcuni casi, la classe assegnata dal modello non ha solo una funzione diagnostica, ma costituisce un trigger diretto per azioni automatiche all'interno del sistema produttivo. È il caso, ad esempio, di Digital Twin che, in base alla classificazione di un difetto, instradano automaticamente il prodotto verso una

linea di rilavorazione, attivano segnalazioni intelligenti, oppure modificano parametri di processo in tempo reale.

Queste modalità rafforzano il ruolo del DT come motore semantico, in grado di trasformare dati eterogenei in azioni significative per il controllo qualità [3]. La riclassificazione sistematica dei casi di studio ha evidenziato come i modelli di Digital Twin per il controllo qualità stiano evolvendo verso architetture complesse, distribuite e intelligenti, capaci di gestire input multimodali e produrre classificazioni categoriali integrate nel ciclo operativo. La convergenza tra dati numerici e simbolici, modelli interpretabili e AI avanzata, costituisce oggi il fondamento metodologico per la progettazione di DT affidabili, scalabili e reattivi [6].

3.4 Considerazioni generali e implicazioni per il caso di studio

Dopo aver analizzato i principali approcci esistenti e le architetture prevalenti nei Digital Twin per il controllo qualità, è possibile trarre alcune considerazioni generali sulle tendenze emergenti e le implicazioni per il modello sperimentale sviluppato in questa tesi.

Con l'avanzare delle tecnologie abilitanti, il Digital Twin ha progressivamente abbandonato la sua natura statica, assumendo una funzione chiave nei sistemi intelligenti e adattivi. Le tendenze emergenti mostrano un graduale spostamento da implementazioni descrittive a modelli predittivi e prescrittivi, grazie alla crescente integrazione con tecniche avanzate di machine learning, ottimizzazione bayesiana, inferenza causale e intelligenza artificiale distribuita [20].

Un primo filone di sviluppo riguarda l'aumento della fedeltà dei modelli: il ricorso a tecniche di simulazione ad elementi finiti (FEM), modelli dinamici parametrici e reti neurali profonde consente ai DT di rappresentare con sempre maggiore precisione il comportamento del sistema reale, anche in condizioni operative non lineari o soggette a disturbi [15].

Parallelamente, si osserva una crescente attenzione all'ottimizzazione adattiva dei DT, in particolare per scopi diagnostici. Tecniche come l'ottimizzazione Bayesiana stanno diventando fondamentali per il tuning automatico dei modelli digitali anche in contesti con dati limitati, come mostrato in ambito energetico per inverter con controllo VSG [20]. Questo approccio migliora la capacità del DT di adattarsi in tempo reale alle variazioni operative, aumentando l'accuratezza diagnostica e riducendo i tempi di intervento [20].

Un'altra tendenza rilevante è l'integrazione dei DT nei sistemi di controllo qualità, con un impatto diretto sulla sostenibilità dei processi industriali. L'adozione di modelli digitali arricchiti da strategie di controllo qualità distribuito ha dimostrato

di ridurre significativamente gli scarti di produzione e migliorare l'efficienza complessiva, in particolare in sistemi caratterizzati da elevata variabilità e complessità [6].

L'aspetto della portabilità e scalabilità è anch'esso in evoluzione: si stanno affermando modelli digitali leggeri, capaci di operare su dispositivi edge e supportare applicazioni mobili, come nel caso dei sistemi di ispezione dei difetti in ambienti produttivi tramite realtà mista e tecniche di few-shot learning [22].

Infine, si rafforza la tendenza all'utilizzo del DT come strumento integrato di decision making, in grado di supportare non solo il monitoraggio e la diagnosi, ma anche la pianificazione operativa, il controllo adattivo e la gestione predittiva del ciclo di vita di componenti, impianti e processi. In questo contesto, la convergenza con l'Internet of Things, la cyber-physical production e i modelli simbolici-intelligenti (ad es. agenti intelligenti) apre la strada a scenari di produzione realmente autonomi, resilienti e sostenibili [6][19][7].

Sulla base di queste evidenze, si inserisce il caso studio sviluppato nel capitolo successivo, che prende ispirazione dall'articolo "*A Digital Twin Assisted Framework for Quality Assurance in Mould Manufacturing*" (2023). Partendo da questo contributo scientifico, è stato adattato un modello sperimentale al contesto della produzione di stampi ad alta precisione, con l'obiettivo di esplorare come l'integrazione di algoritmi di classificazione all'interno di un'architettura Digital Twin possa supportare il controllo qualità in modo tempestivo e automatico. Il caso è stato personalizzato per evidenziare l'applicabilità della logica di previsione binaria (difettoso/non difettoso) su dati reali di processo, mostrando il potenziale impatto in termini di efficienza operativa e monitoraggio proattivo.

Il caso rappresenta un esempio concreto di come l'integrazione tra tecnologie AI e Digital Twin possa supportare una produzione intelligente, reattiva e data-driven.

Capitolo 4 – Implementazione di un Digital Twin per la diagnostica dei difetti

4.1 Introduzione al caso studio

Numerosi settori industriali, come l'automotive, l'aerospaziale e il medicale, si basano sulla produzione di stampi ad alta precisione. Quando si tratta di questi componenti, sono necessarie tolleranze geometriche estremamente ristrette e controlli di qualità severi, perché anche piccole variazioni geometriche possono compromettere l'efficienza del prodotto finito o causare problemi dopo il processo di produzione [26].

In questo contesto, la fresatura CNC è una delle tecnologie più utilizzate per la produzione di stampi perché consente di lavorare i pezzi con una grande precisione mantenendo contemporaneamente una grande flessibilità e ripetibilità. Si tratta di una lavorazione sottrattiva: l'utensile rotante rimuove gradualmente il materiale dal pezzo grezzo utilizzando macchine a controllo numerico fino a raggiungere la geometria desiderata. Tale tecnologia è uno standard nella produzione di stampi ad alta precisione perché consente di raggiungere elevata accuratezza dimensionale e qualità superficiale anche su materiali complessi [26].

Tuttavia, i metodi tradizionali di controllo qualità, che si basano principalmente su ispezioni manuali o su rilievi metrologici effettuati offline, si rivelano spesso lenti, costosi e difficili da applicare su larga scala, soprattutto nei contesti di produzione ad alto volume o con elevata personalizzazione dei componenti [26].

Per questo motivo, lo sviluppo di soluzioni intelligenti e automatizzate in grado di identificare direttamente i difetti geometrici durante la lavorazione sta diventando sempre più importante. In questo modo è possibile ridurre al minimo gli scarti e migliorare l'efficienza complessiva del processo produttivo.

Il caso studio proposto in questo lavoro si ispira a un contributo accademico presentato alla conferenza ASME/IDETC-CIE 2023 dal titolo "*A Digital Twin Assisted Framework for Quality Assurance in Mould Manufacturing*" (2023). Nello

specifico, il caso studio sperimentale riguarda la produzione di inserti per stampi a iniezione realizzati in acciaio temprato per utensili (46–52 HRC), caratterizzati da tolleranze dimensionali estremamente ristrette. Per lo sviluppo e la validazione del modello basato su Digital Twin sono stati progettati diversi test part, ciascuno caratterizzato da un insieme di geometrie tipiche degli inserti di stampo, con l’obiettivo di riprodurre in modo fedele le condizioni operative reali. Questa varietà ha consentito la rappresentazione di una varietà di scenari di lavorazione.

Come mostrato in Figura 4.1, il test part è stato progettato per includere un insieme rappresentativo di geometrie tipiche degli inserti di stampo a iniezione. Esso presenta 9 tasche (pockets) distribuite sulla superficie: ciascuna tasca è una cavità fresata che riproduce in scala ridotta le condizioni reali di lavorazione di uno stampo. All’interno di queste tasche sono state inserite un totale di 165 elementi geometrici, tra cui raccordi arrotondati (fillet cilindrici), posizionati negli angoli interni delle tasche, con raggi variabili compresi tra 0,35 mm e 2,1 mm, così da coprire un intervallo ampio di condizioni geometriche tipiche nelle lavorazioni di precisione, e superfici inclinate di divisione (parting surfaces) che simulano le linee di divisione dello stampo, realizzate con inclinazioni di 5°, 15°, 25° e 35°, valori scelti perché rappresentativi delle configurazioni comunemente adottate nella produzione industriale.

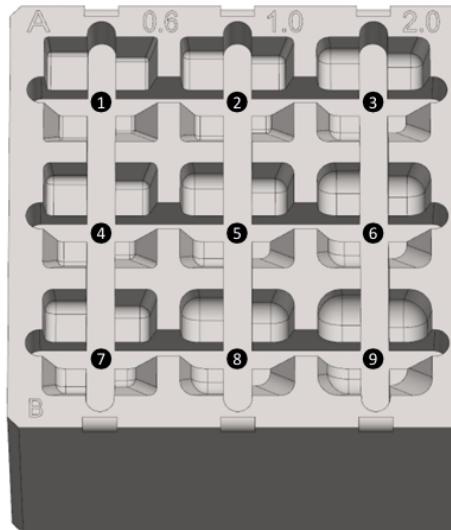


Figura 4.1: Rappresentazione del test part utilizzato come provino per lo studio [26].

Questa combinazione di tasche, raccordi e superfici inclinate rende il test part un pezzo campione complesso e realistico, utile per studiare come le variazioni

geometriche influenzino la qualità finale e per addestrare i modelli predittivi a riconoscere difetti in diversi scenari di lavorazione.

La lavorazione è stata eseguita su una macchina CNC verticale Makino V33i, utilizzando frese a sfera con raggi di taglio nominali pari a 0,59 mm, 0,99 mm e 1,99 mm. Le operazioni hanno previsto tre fasi (sgrossatura, semifinitura e finitura), con particolare attenzione alla fase di finitura, poiché è in questa che si consolidano le deviazioni residue dalle fasi precedenti e che si richiedono le tolleranze più stringenti.

Per la valutazione della qualità sono stati acquisiti dati metrologici tramite scansioni 3D (ATOS ScanBox), confrontando le superfici reali con i modelli CAD di riferimento e trovando il valore della deviazione geometrica per ogni caratteristica. Questo valore è stato quindi utilizzato per etichettare i campioni come difettosi o non difettosi.

In questo lavoro è stato utilizzato il dataset completo messo a disposizione dagli autori dell'articolo, che comprende sia variabili numeriche sia categoriali e mantiene la stessa struttura utilizzata originariamente per l'addestramento del modello predittivo. L'obiettivo è sviluppare un sistema di classificazione binaria (difettoso/non difettoso) che possa supportare la diagnostica in tempo reale integrandosi in un'architettura di Digital Twin.

4.1.1 Configurazione del caso studio e flusso dati adottato

Il dataset utilizzato, messo a disposizione dagli autori dell'articolo [27], comprende un insieme eterogeneo di variabili numeriche e categoriali raccolte durante le operazioni di lavorazione CNC reali, offrendo quindi un contesto sperimentale realistico e rappresentativo delle dinamiche produttive.

A partire da queste informazioni, è stata sviluppata una procedura di apprendimento automatico per addestrare modelli predittivi in grado di classificare ogni componente del prodotto come difettoso o non difettoso. Il processo in questione costituisce la base del Digital Twin virtuale sviluppato per il caso studio, che replica digitalmente le condizioni del processo produttivo per supportare intelligentemente il controllo qualità.

L'architettura proposta dimostra la fattibilità tecnica e metodologica dell'approccio, consentendo la diagnostica automatizzata. In particolare, l'uso di modelli classificatori che restituiscono output categorici (difettoso/non difettoso) rende il sistema comprensibile e direttamente utilizzabile per decisioni operative, riducendo la necessità di ispezioni visive o controlli manuali. Questo migliora la reattività del processo produttivo, aumenta la qualità complessiva e consente la gestione e il controllo più avanzati basati su dati.

4.2 Preparazione ed etichettatura del dataset

4.2.1 Caricamento dei dati

Il dataset utilizzato per l'addestramento dei modelli predittivi, rappresenta un insieme strutturato di misurazioni raccolte durante operazioni reali di lavorazione CNC per la produzione di stampi ad alta precisione.

Come mostrato in Figura 4.2, i dati sperimentali sono stati forniti in formato CSV e caricati in ambiente Python tramite la libreria Pandas, utilizzando il modulo `read_csv()`. Per una gestione più ordinata e portabile dei percorsi dei file è stato impiegato il modulo `pathlib` [28] e, in particolare, l'oggetto `Path`, che consente di definire i percorsi in modo indipendente dal sistema operativo. Nella fase di configurazione sono inoltre stati definiti la cartella di output (`OUT_DIR`), destinata al salvataggio dei risultati e delle metriche, e un valore di *random state* per garantire la riproducibilità degli esperimenti. Dopo l'importazione, la colonna `Dev` è stata convertita esplicitamente in formato numerico mediante la funzione `pd.to_numeric()` con l'opzione `errors="coerce"`, così da gestire eventuali valori anomali o non numerici sostituendoli con `NaN`. Questa operazione ha permesso di assicurare la coerenza tipologica del dataset prima delle successive fasi di analisi e modellazione.

```
1 # Configurazione
2 DATA_PATH = Path("quality_data.csv")
3 OUT_DIR = Path("resultsSMOTECN_FIX")
4 OUT_DIR.mkdir(exist_ok=True)
5 RANDOM_STATE = 42
6
7 # Caricamento dataset
8 df = pd.read_csv(DATA_PATH, sep=";", decimal=",")
9 df[["Dev"]] = pd.to_numeric(df[["Dev"]], errors="coerce")
```

Figura 4.2: Configurazione e caricamento del dataset in Pandas.

Il dataset contiene sia variabili numeriche, come il valore nominale della misura previsto dal progetto CAD/CAM misurato, sia variabili categoriche, relative al tipo di componente, posizione dell'utensile e stato di usura di quest'ultimo. Tale combinazione riflette in modo realistico il tipo di dati disponibili in ambienti produttivi automatizzati.

4.2.2 Etichettatura dei difetti

Per sviluppare un sistema predittivo in grado di classificare ogni componente come “difettoso” o “non difettoso”, è stato necessario definire in modo chiaro la variabile target. A tal fine, per ciascun campione di lavorazione si è calcolata la

deviazione geometrica rispetto alla specifica progettuale, indicata con il campo *Dev*, ottenuta come differenza tra valore reale (misura effettiva rilevata sul pezzo) e valore nominale (misura teorica prevista dal progetto CAD/CAM).

È stata fissata una soglia di tolleranza pari a ± 0.007 mm sulla deviazione secondo il criterio utilizzato nello studio di riferimento. Se la deviazione assoluta è compresa entro questo intervallo ($|Dev| \leq 0.007$ mm), il pezzo è considerato conforme e viene classificato come non difettoso. D'altra parte, quando lo scostamento supera il limite accettabile ($|Dev| > 0.007$ mm), il componente è considerato difettoso.

La soglia di ± 0.007 mm adottata per l'etichettatura dei campioni non è un valore arbitrario, bensì deriva da specifiche industriali e riprende direttamente i criteri impiegati nello studio di riferimento di Böttjer e Ramanujan (2023). In quel contesto, tale intervallo di accettabilità riflette le tolleranze dimensionali tipiche della produzione di stampi ad alta precisione, dove deviazioni anche minime possono compromettere l'intercambiabilità e la funzionalità del componente. Pertanto, la soglia può essere considerata un limite conservativo, progettato per garantire elevata affidabilità del controllo qualità ed evitare falsi negativi, ossia casi in cui un difetto potenzialmente critico non venga rilevato. Al tempo stesso, è importante notare che fissare un margine così stringente aumenta il rischio di classificare come difettosi alcuni campioni che potrebbero essere accettabili in altre condizioni produttive. Questo aspetto evidenzia come la definizione della soglia di tolleranza non sia universale, ma debba essere sempre valutata in base alle specifiche esigenze applicative e sugli standard di qualità richiesti dal settore di riferimento.

Questa procedura consente di trasformare una misura continua in una variabile categorica binaria (0 = non difettoso, 1 = difettoso), che costituisce la variabile target per l'addestramento e la validazione dei modelli predittivi. È importante sottolineare che la variabile *Dev* è utilizzata esclusivamente per generare l'etichetta di classificazione, mentre le variabili di ingresso del modello comprendono soltanto parametri di processo e di progettazione, così da evitare fenomeni di *data leakage*, ovvero evitare che il modello utilizzi informazioni non disponibili in fase produttiva e che porterebbero a una valutazione troppo ottimistica delle prestazioni predittive. Tali misure consentono di garantire l'applicabilità del sistema in un contesto reale di controllo qualità, dove le decisioni devono basarsi soltanto su dati raccolti durante o prima della lavorazione.

La Figura 4.3 mostra l'operazione di etichettatura implementata nel dataset:

```
1 df['difettoso'] = (df['Dev'].abs() > 0.007).astype(int)
```

Figura 4.3: Operazione di etichettatura.

La deviazione, un dato quantitativo continuo, può essere convertita in una variabile categorica binaria utilizzando questa tecnica di etichettatura. Questo è utile per lo sviluppo di modelli di classificazione supervisionata.

4.2.3 Separazione delle variabili di input e di output

Dopo aver definito la variabile target, è stato necessario selezionare le variabili da utilizzare come input per i modelli predittivi. Le variabili di input sono state suddivise in due tipologie come mostrato nella Tabella 4.1:

Tabella 4.1: Descrizione delle variabili di input presenti nel dataset.

Nome variabile	Tipologia	Descrizione
Element	Categorica	Tipo di geometria lavorata (es. <i>Cylinder, Angle</i>)
Pocket	Categorica	La tasca del provino in cui si trova la geometria misurata
tool_state	Categorica	Condizione dell'utensile durante la lavorazione (es. <i>New, Used</i>)
Direction	Categorica	Direzione lungo cui viene effettuata la misura geometrica
Corner	Categorica	Specifica se la misurazione è stata effettuata in un angolo del pocket o in una zona piana
Nominal	Numerica	Valore nominale della misura previsto dal progetto CAD/CAM misurato

Nel dataset utilizzato, i valori nominali sono espressi in millimetri sia per i raccordi (cylinder) sia per le superfici di divisione (angle). Nel secondo caso, invece di riportare l'angolo in gradi, il file CSV del dataset contiene la quota lineare nominale corrispondente alla geometria inclinata, come previsto nei dati di progetto CAD.

Inoltre, nel dataset originale erano presenti anche i campi **Actual**, **Dev** (deviazione = **Actual** – **Nominal**) e **TP**. I primi due valori sono stati utilizzati esclusivamente per la costruzione della variabile target (difettoso/non difettoso) e non come input del modello, questo perché rappresentano misurazioni effettuate a posteriori, ossia dopo la produzione del componente, e quindi non disponibili in fase di previsione. Il valore **TP**, invece, è un identificatore che rappresenta un singolo pezzo lavorato. Ogni valore di **TP** (da 1 a 16 nel dataset) raggruppa più misurazioni acquisite sullo stesso componente, in diverse posizioni o direzioni. Questa variabile è stata usata per aggregare e analizzare i risultati delle predizioni a livello del singolo pezzo.

Un aspetto peculiare del dataset riguarda la variabile **Direction**, che specifica l'orientamento secondo cui viene effettuata la misurazione geometrica (assi X, Y e Z). Per ciascun elemento, come ad esempio un raccordo cilindrico (*Cylinder*), non viene registrata un'unica osservazione, ma più misurazioni lungo direzioni differenti. Questa scelta riflette la natura anisotropa del processo di fresatura: la qualità

geometrica di un componente non è uniforme in tutti gli assi, ma può variare in funzione di fattori quali vibrazioni della macchina, gioco meccanico degli assi o rigidità dell'utensile.

Dal punto di vista metrologico, l'acquisizione multidirezionale è indispensabile per ottenere una caratterizzazione completa della geometria. Una misura effettuata in una sola direzione rischierebbe infatti di trascurare difetti presenti lungo altri assi, restituendo una valutazione parziale della conformità del pezzo. L'inclusione della variabile `Direction` consente dunque di catturare questa variabilità tridimensionale, migliorando la capacità predittiva dei modelli di classificazione e garantendo che il Digital Twin sia in grado di rilevare le non conformità indipendentemente dall'orientamento geometrico in cui si manifestano.

La Figura 4.4 mostra il codice che ha permesso la separazione tra input (X) e variabile target (y):

```

1 features_cat = ["Element", "Pocket", "tool_state", "Direction", "Corner"]
2 features_num = ["Nominal"]
3
4 X = df[features_cat + features_num].copy()
5 y = df["difettoso"]
6 groups = df["TP"]

```

Figura 4.4: Separazione tra input (X) e target (y).

Oltre alla separazione tra variabili di input X e target y, è stata anche definita la funzione `groups = df["TP"]`, permettendo analisi aggregate a livello di pezzo e una valutazione più strutturata delle prestazioni del modello.

Questa distinzione tra le varie tipologie di variabili ha inoltre permesso la creazione di una procedura di pre-processing appropriata. Le variabili categoriche sono state trasformate tramite la funzione One-Hot Encoding, mentre le variabili numeriche sono state standardizzate utilizzando la funzione StandardScaler. L'efficacia dei modelli predittivi addestrati dipende fortemente dal trattamento corretto di queste caratteristiche eterogenee, che consente di gestire in modo coerente dati di natura diversa. Le singole trasformazioni applicate ai dati verranno descritte in dettaglio nella sezione successiva.

4.3 Integrazione dei modelli predittivi nel Digital Twin

Sono stati considerati inizialmente tre modelli di classificazione supervisionata, in linea con l'articolo di riferimento: Random Forest, Multilayer Perceptron (MLP) e Multinomial Linear Regression (MLR). L'obiettivo della scelta di questi algoritmi è quello di esplorare e confrontare diversi approcci, che rappresentano diverse

famiglie di modelli di apprendimento automatico. Ogni famiglia ha i propri punti di forza in termini di predittività, interpretabilità, robustezza e scalabilità.

La Random Forest è un algoritmo basato su una tecnica *ensemble*, ovvero una combinazione di molteplici modelli deboli (in questo caso, alberi decisionali) al fine di migliorare la performance complessiva [29]. Si tratta di una metodologia robusta e ben consolidata, in grado di gestire efficacemente dataset misti, ovvero composti da variabili sia numeriche che categoriche [30]. La sua struttura interna riduce il rischio di overfitting e funziona bene anche quando ci sono dati rumorosi, outlier o caratteristiche non informative [31]. Inoltre, la Random Forest offre una stima diretta dell'importanza delle variabili; questo è vantaggioso nei casi in cui è necessario interpretare i fattori decisivi per la scelta [29].

Il secondo modello adottato è il Multilayer Perceptron (MLP), una rete neurale di tipo *feedforward*, in cui le informazioni fluiscono esclusivamente in avanti: partendo dallo strato di input, vengono elaborate attraverso i livelli nascosti di neuroni (hidden layer) fino a raggiungere lo strato di output. Questa architettura, che non prevede retroazioni interne tra i nodi, appartiene alla famiglia del deep learning ed è in grado di modellare relazioni non lineari complesse nei dati [32]. La sua architettura a strati consente di rappresentare funzioni di decisione altamente flessibili, rendendolo particolarmente adatto a compiti di classificazione complessa. Tuttavia, a differenza dei due modelli precedenti, il modello MLP richiede una fase di addestramento più lunga e può risultare più sensibile all'ottimizzazione degli iperparametri [32].

Il terzo modello Multinomial Linear Regression (MLR) consente di stimare la probabilità di appartenenza alle diverse classi a partire da un insieme di variabili esplicative. Basata sulla funzione `logit`, essa trasforma una combinazione lineare delle variabili indipendenti in una probabilità compresa tra 0 e 1, risultando particolarmente adatta a problemi di classificazione binaria [33]. Rispetto ad algoritmi più complessi, come le reti neurali, la MLR presenta una struttura semplice e interpretabile, permettendo di valutare in modo diretto il contributo delle variabili indipendenti alla determinazione delle classi [34]. Le sue caratteristiche parametriche e la rigorosa formulazione matematica lo rendono un modello di riferimento consolidato, spesso utilizzato come base nei compiti di classificazione e come termine di confronto per approcci più sofisticati [35].

Accanto a questi tre modelli, in questo studio è stato introdotto un quarto classificatore, il Light Gradient Boosting Machine (LightGBM), con l'obiettivo di ampliare l'analisi e individuare un modello capace di potenziare la capacità di generalizzazione del Digital Twin nel riconoscimento dei difetti. LightGBM appartiene alla famiglia dei metodi di boosting basati su alberi decisionali, in cui i modelli vengono costruiti in sequenza, ciascuno con lo scopo di correggere gli errori commessi dai precedenti [36]. Tale approccio consente di catturare interazioni non lineari complesse tra le variabili e di ottenere un compromesso efficace tra

accuratezza, velocità di addestramento e capacità di generalizzazione.

Rispetto alla Random Forest, che aggredisce alberi indipendenti, il boosting ottimizza progressivamente la funzione di perdita, risultando più sensibile alle deviazioni sistematiche e quindi potenzialmente più efficace nel riconoscere pattern sottili legati alla presenza di difetti [37]. Inoltre, LightGBM offre un'elevata efficienza computazionale e una gestione nativa delle variabili categoriche, caratteristiche che lo rendono particolarmente adatto a dataset misti come quello in esame [36].

Ciascun modello è stato inserito in un flusso di lavoro strutturato, che include una fase di pre-elaborazione dei dati attraverso la codifica delle variabili categoriche e normalizzazione, e una fase di addestramento e validazione. Questa tecnica garantisce la coerenza tra le diverse fasi del ciclo di vita del modello e consente un confronto equo delle sue prestazioni, inclusa la nuova versione potenziata con LightGBM.

4.3.1 Architettura del Digital Twin

Questo studio ha sviluppato l'architettura del Digital Twin basata sull'idea di creare un modello digitale che riproduca in modo realistico e dinamico il comportamento del processo di fresatura reale. In questo caso, il Digital Twin è un sistema con componenti fisici e virtuali che possono trasmettere dati e informazioni in modo sincronizzato e continuo. Lo scopo principale è anticipare e rilevare automaticamente i difetti dimensionali nei componenti prodotti, migliorando la tempestività e l'efficienza delle procedure di controllo qualità.

Nella fase sperimentale condotta, non è stato possibile implementare una comunicazione diretta in tempo reale tra la macchina CNC e il Digital Twin tramite sensoristica fisica. Tuttavia, l'intero flusso di dati è stato simulato fedelmente utilizzando un dataset reale messo a disposizione dagli autori Boettjer e Ramanujan, acquisito durante vere sessioni di lavorazione meccanica. Ciò consente di simulare un ambiente industriale realistico, in cui ogni riga del dataset rappresenta una singola operazione di finitura e le variabili registrate (ad esempio direzione, stato dell'utensile, tipo di elemento e deviazione rispetto al valore nominale) derivano da misurazioni sensoristiche reali.

Tre blocchi funzionali principali compongono la logica modulare e scalabile utilizzata nell'architettura implementata. I dati vengono acquisiti nel primo modulo: in un sistema industriale reale, consisterebbe in una rete di sensori installati sulla macchina CNC che rilevano parametri come la direzione dell'utensile, lo stato dell'utensile, il tipo di elemento e la deviazione rispetto alla misura prevista. Tali informazioni sono già incluse nel dataset in questa simulazione e vengono importate utilizzando procedure di caricamento automatizzato in ambiente Python visto in precedenza.

Il secondo modulo è quello di elaborazione e classificazione, che rappresenta il nucleo logico del DT. Qui i dati vengono preprocessati (tramite codifica delle variabili categoriali e normalizzazione dei valori numerici), per poi essere elaborati dai modelli predittivi addestrati. Per ogni componente esaminato, il modello produce un output categorizzato come "difettoso" o "non difettoso". L'infrastruttura di elaborazione può essere modulata, il che consente di sostituire il classificatore con uno più aggiornato o adatto alle modifiche del processo in futuro.

Il terzo modulo è quello di visualizzazione e supporto decisionale, in cui le previsioni ottenute dal DT vengono rappresentate in modo interpretabile per gli operatori o integratori di sistema. La previsione binaria consente il riconoscimento immediato degli elementi non conformi, facilitando l'adozione di contromisure in tempi rapidi, anche in ottica in-process.

Nel caso studio reale descritto da Boettjer e Ramanujan, i dati di processo sono stati acquisiti in tempo reale tramite sensori fisici e connessione OPC-UA al controllore macchina. Tuttavia, l'implementazione del DT è stata condotta in modalità offline, utilizzando i dati raccolti per l'addestramento e la validazione dei modelli di classificazione. Gli autori evidenziano come la successiva estensione del framework in modalità online, con sincronizzazione continua tra ambiente fisico e virtuale, rappresenti un'evoluzione naturale del sistema verso un monitoraggio più proattivo e adattivo.

In questo lavoro, l'architettura del DT è stata riprodotta anch'essa in modalità offline, impiegando un dataset reale acquisito in precedenza e caricato in ambiente Python. Pur in assenza di un collegamento diretto con la macchina CNC, la struttura modulare del framework e la fedeltà del flusso informativo hanno consentito di ricreare in modo coerente la sequenza logica delle fasi di acquisizione, elaborazione e classificazione dei dati, simulando così il comportamento del DT in condizioni realistiche.

L'obiettivo generale dell'architettura è quello di trasformare il controllo qualità da una funzione reattiva a una funzione predittiva e proattiva, in modo che possa intervenire prontamente quando si verificano problemi che compromettono la produzione. Un aspetto particolarmente rilevante è che il Digital Twin, non richiede la conoscenza preventiva di misurazioni complesse come la deviazione geometrica (Dev) o il valore effettivo (Actual), che sono ottenibili soltanto tramite ispezioni successive al processo. Al contrario, esso sfrutta le combinazioni di variabili di progetto e di processo (geometria, pocket, direzione, corner, stato utensile, ecc.) per predire l'esito qualitativo (difettoso/non difettoso).

In questo modo, il sistema può contribuire ad anticipare i difetti, bloccando una lavorazione sospetta, inviando un alert per un controllo manuale o attivando una correzione automatica del processo. Ciò si traduce in una riduzione delle rilavorazioni, degli sprechi e dei fermi macchina, con un impatto diretto sulla produttività e sull'affidabilità del processo se implementato in contesti industriali.

reali.

Infine, è fondamentale sottolineare come l'approccio sviluppato può essere esteso: anche se è stato testato utilizzando la fresatura di stampi, può essere facilmente utilizzato per altri tipi di lavorazione meccanica, come foratura, tornitura o rettifica, e in altri contesti di produzione in cui la qualità dimensionale dei componenti è fondamentale.

4.3.2 Implementazione dei modelli predittivi

Per i tre modelli originariamente considerati nello studio di riferimento, Random Forest, Multilayer Perceptron (MLP) e Multinomial Linear Regression (MLR), i parametri sono stati impostati in modo da riflettere i valori ottimali individuati dagli autori a valle di un processo di validazione incrociata a tre fold (3-fold cross-validation) eseguito sul training set.

In particolare, per il modello Random Forest è stato adottato un numero di alberi pari a 78 e una profondità massima pari a 7, corrispondenti ai valori di miglior performance [38]. L'utilizzo di un numero relativamente elevato di alberi consente di ridurre la varianza del modello e aumentare la stabilità delle previsioni, poiché ogni albero viene addestrato su un sottoinsieme casuale dei dati e delle variabili. Al tempo stesso, la limitazione della profondità massima a 7 livelli evita una eccessiva specializzazione dei singoli alberi sui dati di training, contenendo così il rischio di overfitting. In questo modo, il modello mantiene un buon equilibrio tra capacità predittiva e generalizzazione, risultando sufficientemente complesso da catturare le relazioni non lineari tra le variabili, ma al tempo stesso robusto rispetto al rumore e alle fluttuazioni del dataset. È stato inoltre attivato il bilanciamento automatico delle classi (`class_weight="balanced"`), che in `scikit-learn` applica una pesatura inversamente proporzionale alla frequenza di ciascuna classe, al fine di compensare lo sbilanciamento tra campioni difettosi e non difettosi. In particolare, il peso assegnato alla classe C_i è calcolato secondo la relazione [39]:

$$w(C_i) = \frac{n_{total}}{n_{classes} \cdot n_{C_i}}$$

dove $w(C_i)$ è il peso per la i -esima classe, n_{total} è il numero totale di campioni nel dataset e n_{C_i} è il numero di campioni nella i -esima classe.

Per il modello MLP è stata adottata un'architettura con due strati nascosti da 10 neuroni ciascuno (`hidden_layer_sizes=(10, 10)`), con solver di tipo L-BFGS, learning rate iniziale pari a 0.001 e un massimo di 2000 iterazioni [40]. Questa configurazione riprende fedelmente quella indicata nell'articolo, in cui l'architettura ottimale prevedeva due strati intermedi da 10 neuroni e un solver L-BFGS, scelto per la sua efficienza nei contesti di dimensione ridotta e per la capacità di gestire feature ad alta dimensionalità derivate da codifica categorica. Anche in questo caso è stato mantenuto il bilanciamento automatico delle classi.

Infine, il modello di Multinomial Linear Regression è stato configurato senza penalizzazione (`penalty='none'`) e con solver L-BFGS, in modo da replicare la condizione di assenza di regolarizzazione (parametri elastic net e regolarizzazione nulli) utilizzata dagli autori per valutare l'approccio lineare puro [41]. Anche in questo caso è stato attivato il bilanciamento delle classi.

A completamento del confronto con i modelli originari, è stato introdotto un quarto classificatore, il Light Gradient Boosting Machine (LightGBM), con l'obiettivo di estendere l'analisi e verificare se un approccio di boosting potesse migliorare la capacità di generalizzazione del Digital Twin. I parametri del modello LightGBM sono stati definiti con l'obiettivo di ottenere un compromesso tra accuratezza predittiva, capacità di generalizzazione e stabilità numerica.

In particolare, il numero di alberi (`n_estimators = 300`) è stato scelto per consentire al modello di apprendere in modo progressivo, mantenendo un tempo di addestramento contenuto e limitando il rischio di overfitting [42].

Il learning rate, fissato a 0.05, rappresenta un valore intermedio ampiamente adottato in letteratura, in grado di garantire una convergenza graduale del processo di boosting: valori più elevati comportano un apprendimento troppo rapido e instabile, mentre valori inferiori richiederebbero un numero molto maggiore di iterazioni.

La profondità massima degli alberi è stata impostata a 7 per mantenere la complessità del modello in linea con quella adottata per la Random Forest, facilitando così il confronto tra approcci.

Il numero di foglie per albero (`num_leaves = 31`) è stato definito in modo coerente con la profondità scelta, secondo la relazione empirica:

$$\text{num_leaves} \approx 2^{\max_depth}$$

che consente di rappresentare adeguatamente le interazioni non lineari senza introdurre eccessiva variabilità. Il sottocampionamento delle istanze (`subsample = 0.8`) e delle variabili (`colsample_bytree = 0.8`) è stato introdotto per incrementare la robustezza del modello e ridurre la correlazione tra gli alberi generati nel processo di boosting. In particolare, il parametro `subsample` fa sì che ciascun albero venga addestrato su un sottoinsieme casuale dei campioni, prevenendo che il modello si adatti eccessivamente a specifiche osservazioni del dataset di training. Analogamente, il parametro `colsample_bytree` limita il numero di variabili disponibili per la costruzione di ogni albero, inducendo una maggiore diversificazione tra i modelli deboli e riducendo la varianza complessiva del sistema. Questa combinazione di tecniche di sottocampionamento contribuisce a diminuire la dipendenza reciproca tra gli alberi della foresta, migliorando la capacità di generalizzazione del modello su dati non ancora osservati e rendendolo meno sensibile al rumore presente nei dati di addestramento.

Infine, l'opzione `class_weight = "balanced"` permette di compensare automaticamente lo sbilanciamento tra classi, in analogia con gli altri modelli implementati, evitando che la classe maggioritaria influenzi in modo predominante il processo di ottimizzazione della funzione di perdita.

Questi valori non derivano da un processo di ottimizzazione sistematica dei parametri, ma da una calibrazione empirica eseguita sulla base di test preliminari, in coerenza con gli intervalli più comunemente adottati nella letteratura per compiti di classificazione binaria su dataset industriali eterogenei. Questa estensione è stata sviluppata per valutare la capacità del metodo di boosting di gestire in modo più efficiente la combinazione di variabili numeriche e categoriche e di rafforzare la funzione diagnostica del Digital Twin. Il modello è stato addestrato sullo stesso schema di suddivisione del dataset per gruppi dei pezzi prova (TP) descritto nella sezione successiva, al fine di garantire la piena confrontabilità dei risultati.

La definizione accurata di questi modelli mostrata in Figura 4.5, ha consentito un confronto rigoroso e replicabile tra diversi approcci predittivi.

```

1 base_models = {
2     "RandomForest": RandomForestClassifier(
3         n_estimators=78, max_depth=7,
4         class_weight="balanced", random_state=RANDOM_STATE),
5     "MLP": MLPClassifier(
6         hidden_layer_sizes=(10, 10),
7         solver="lbfgs",
8         learning_rate_init=0.001,
9         max_iter=2000,
10        random_state=RANDOM_STATE),
11    "MLR": LogisticRegression(
12        random_state=RANDOM_STATE,
13        class_weight="balanced",
14        solver="lbfgs",
15        penalty='none'),
16    "LightGBM": LGBMClassifier(
17        boosting_type="gbdt",
18        objective="binary",
19        learning_rate=0.05,
20        n_estimators=300,
21        max_depth=7,
22        num_leaves=31,
23        class_weight="balanced",
24        subsample=0.8,
25        colsample_bytree=0.8,
26        random_state=RANDOM_STATE,
27        verbose=-1)}
```

Figura 4.5: Definizione dei modelli di classificazione.

4.3.3 Il problema dello sbilanciamento: tecniche di oversampling

Il problema dello sbilanciamento delle classi è un aspetto cruciale nelle applicazioni di machine learning, in particolare quando si lavora con dataset dove una classe è significativamente meno rappresentata rispetto all'altra. In questo contesto, si osserva un forte sbilanciamento dovuto al fatto che i pezzi difettosi rappresentano la classe maggioritaria mentre i pezzi non difettosi la classe minoritaria. Nei casi di classi sbilanciate, i modelli tendono a predire la classe maggioritaria con una precisione più alta, ignorando o sottovalutando la classe minoritaria, che potrebbe essere di interesse per confermare la qualità.

Come suggerito dagli autori nell'articolo di riferimento, tecniche di oversampling possono contribuire in modo significativo a migliorare la performance del modello aumentando il numero di esempi della classe minoritaria, consentendo al modello di apprendere in modo più equilibrato e di non ignorare la classe meno rappresentata. Nonostante i modelli come Random Forest e MLP (Multi-Layer Perceptron) siano già dotati di meccanismi automatici per gestire lo sbilanciamento tramite l'uso del parametro `class_weight='balanced'`, che bilancia automaticamente l'importanza delle classi durante l'addestramento, è stato, quindi, deciso di testare l'efficacia di tecniche di oversampling per migliorare ulteriormente la classificazione delle classi minoritarie.

L'oversampling è una tecnica utilizzata per affrontare il problema dello sbilanciamento delle classi, in cui una classe è sotto-rappresentata rispetto all'altra, agendo tramite l'aumento del numero di campioni della classe minoritaria, portando così la distribuzione delle classi a essere più equilibrata [43]. L'obiettivo principale è quello di consentire al modello di apprendere meglio la classe minoritaria, evitando che si concentri prevalentemente sulla classe maggioritaria. Il risultato è una migliore generalizzazione da parte del modello, che è in grado di rilevare anche gli esempi rari e difficili della classe minoritaria [43].

Sono state implementate e confrontate due tecniche: RandomOverSampler (ROS), che genera nuovi campioni minoritari tramite replicazione casuale, il che potrebbe introdurre una ridotta variabilità nei dati addestrati, e SMOTENC (Synthetic Minority Over-sampling Technique for Nominal and Continuous), che applica una generazione sintetica di nuovi campioni, tenendo conto sia delle variabili numeriche che di quelle categoriche nel dataset. L'obiettivo di queste due tecniche è determinare quale delle due soluzioni migliori risulti più efficace nel migliorare le performance del modello, bilanciando in modo ottimale la classe minoritaria e aumentando la capacità predittiva. Entrambe le tecniche di oversampling sono state applicate esclusivamente al set di addestramento, con l'obiettivo di fornire al modello un insieme di dati bilanciato da cui apprendere in modo più efficace. I set di validazione e di test, invece, sono stati mantenuti invariati, in modo da garantire

una valutazione imparziale delle prestazioni e prevenire fenomeni di data leakage dovuti all’alterazione della distribuzione originale delle classi.

4.4 Sviluppo e Addestramento del Modello

4.4.1 Suddivisione del dataset

A differenza dell’articolo di riferimento, in cui si adotta una suddivisione del dataset pari al 90% per l’addestramento e al 10% per il test, con una validazione incrociata interna a tre fold finalizzata all’ottimizzazione degli iperparametri e basata su singole osservazioni di feature, nel presente lavoro è stata invece introdotta una logica di partizione per gruppi di pezzi di prova (TP). Tale approccio, più coerente con la natura dei dati e con l’obiettivo applicativo del modello, prevede che tutte le misurazioni relative a uno stesso TP vengano mantenute all’interno dello stesso sottoinsieme, così da evitare qualsiasi sovrapposizione tra addestramento, validazione e test. È stato pertanto definito uno schema di suddivisione con proporzioni pari al 70% per l’addestramento, 15% per la validazione e 15% per il test, che consente di preservare la rappresentatività dei dati e di garantire una valutazione più realistica della capacità di generalizzazione del modello su componenti mai osservati in fase di addestramento.

Nel caso studio originale, il processo di classificazione è condotto a livello di singola caratteristica di lavorazione (ad esempio raccordi o superfici di separazione), e l’etichetta qualitativa viene assegnata direttamente a tale livello. Gli autori adottano quindi un approccio orientato alla singola osservazione, nel quale le misurazioni appartenenti a uno stesso pezzo possono essere distribuite tra training e test set. Sebbene questa modalità risulti coerente con le finalità sperimentali dello studio di riferimento, essa comporta il rischio di introdurre fenomeni di contaminazione dei dati, poiché caratteristiche peculiari di un stesso pezzo potrebbero essere apprese durante l’addestramento e successivamente riconosciute nel test, producendo così una sovrastima della reale capacità di generalizzazione del modello. In questo lavoro, invece, la suddivisione del dataset è stata eseguita per gruppi di pezzi prova (TP), in modo da preservare l’indipendenza tra le fasi di addestramento, validazione e test, riducendo così il rischio di contaminazione dei dati e garantendo una valutazione più realistica della capacità di generalizzazione del modello. Tuttavia, pur adottando questa suddivisione per gruppi, l’output del Digital Twin rimane definito a livello di singola caratteristica geometrica, restituendo per ciascun elemento lavorato la classificazione binaria difettoso/non difettoso, coerentemente con la struttura diagnostica proposta nello studio di riferimento. Questa impostazione è fondamentale perché consente al Digital Twin di operare a una scala di analisi localizzata, in grado di individuare il difetto non solo a livello di componente complessivo, ma sulla specifica caratteristica geometrica che lo produce. In tal modo il modello non

si limita a fornire una valutazione globale della qualità del pezzo, ma diventa uno strumento diagnostico interpretativo, capace di correlare direttamente le deviazioni di processo con la loro manifestazione geometrica, supportando in modo mirato le attività di controllo e di intervento correttivo.

La suddivisione del dataset è stata realizzata direttamente in ambiente Python, definendo un elenco dei pezzi prova unici e suddividendoli casualmente nei tre sottoinsiemi di addestramento, validazione e test, secondo le proporzioni 70%, 15% e 15%. Per garantire la riproducibilità degli esperimenti, la selezione casuale è stata eseguita utilizzando un seme fisso. In questo modo, ciascun pezzo prova e tutte le sue misurazioni sono stati assegnati in modo coerente a un solo insieme, evitando sovrapposizioni tra le diverse fasi di analisi.

L'adozione di questa logica risulta pienamente coerente con il contesto applicativo industriale della ricerca. I dati a disposizione presentano infatti una forte correlazione interna tra le misurazioni di uno stesso TP, dovuta a fattori condivisi come l'usura dell'utensile, le condizioni operative e le caratteristiche geometriche del pezzo. Ragionare per singola feature, come avviene nell'articolo, è appropriato in una prospettiva puramente dimostrativa, ma non garantisce una valutazione affidabile delle prestazioni del modello in uno scenario operativo. La suddivisione per gruppi di TP consente invece di stimare in modo più conservativo e realistico la capacità del modello di generalizzare a nuovi lotti produttivi o a componenti non ancora osservati, riflettendo più fedelmente le condizioni di utilizzo reale del Digital Twin.

Inoltre, l'introduzione di un set di validazione esplicito pari al 15% dei dati ha permesso di condurre un confronto strutturato tra differenti modelli e strategie di riequilibrio delle classi, mantenendo il test finale completamente separato e non influenzato dalle scelte di tuning. Tale impostazione garantisce la trasparenza del processo di addestramento e la robustezza della valutazione delle prestazioni finali.

Per garantire un confronto equo tra le due tecniche di oversampling, SMOTENC e RandomOverSampler (ROS), è stato fondamentale utilizzare gli stessi gruppi di lavorazione (TP) sia durante la fase di addestramento che nella fase di test. Questa scelta metodologica si è resa particolarmente necessaria in quanto le due tecniche sono state implementate in script Python distinti, uno dedicato a SMOTENC e l'altro a ROS. In questo modo, entrambi i modelli sono stati addestrati e testati sugli stessi campioni di dati, garantendo che le differenze nelle performance fossero attribuibili esclusivamente alla tecnica di oversampling e non a una variazione nella composizione del set di test. Questo permette di isolare l'effetto delle tecniche di oversampling sulle prestazioni del modello, poiché le metriche finali (come F1-score, precision, recall) sono calcolate sugli stessi dati di test, rendendo il confronto tra SMOTENC e ROS diretto e imparziale.

```

1 unique_tp = df[["TP"]].unique()
2 np.random.seed(RANDOM_STATE)
3 np.random.shuffle(unique_tp)
4
5 n_tp = len(unique_tp)
6 n_train = int(0.7 * n_tp)
7 n_val = int(0.15 * n_tp)
8
9 train_tp = unique_tp[:n_train]
10 val_tp = unique_tp[n_train:n_train + n_val]
11 test_tp = unique_tp[n_train + n_val:]
12
13 with open(OUT_DIR / "split_tp_70_15_15.json", "w") as f:
14     json.dump({"train_tp": train_tp.tolist(), "val_tp": val_tp.
15     tolist(), "test_tp": test_tp.tolist()}, f, indent=2)
16
17 train_mask = df[["TP"]].isin(train_tp)
18 val_mask = df[["TP"]].isin(val_tp)
19 test_mask = df[["TP"]].isin(test_tp)
20
21 X_train, X_val, X_test = X[train_mask], X[val_mask], X[test_mask]
22 y_train, y_val, y_test = y[train_mask], y[val_mask], y[test_mask]
23 groups_train, groups_val, groups_test = groups[train_mask], groups
24 [val_mask], groups[test_mask]

```

Figura 4.6: Suddivisione del dataset in train e test fissi per TP.

Il codice riportato in Figura 4.6 illustra la procedura di suddivisione del dataset in insiemi di addestramento, validazione e test adottata nel caso di applicazione della tecnica di oversampling SMOTENC [44]. L’obiettivo di questa fase è duplice: da un lato, garantire la stratificazione rispetto alla variabile target (in modo che la proporzione tra campioni difettosi e non difettosi sia preservata), e, dall’altro, assicurare che tutti i campioni appartenenti allo stesso gruppo TP rimangano uniti all’interno dello stesso insieme.

A tal fine viene estratta la lista dei pezzi prova unici presenti nel dataset, identificati dalla colonna TP. L’utilizzo del metodo `unique()` consente di individuare ciascun pezzo come entità indipendente, a prescindere dal numero di misurazioni a esso associate. Successivamente, viene impostato un seme casuale fisso (`RANDOM_STATE`), che permette di riprodurre esattamente la stessa suddivisione a ogni esecuzione, garantendo la replicabilità del processo sperimentale. La funzione `shuffle()` di NumPy viene poi impiegata per mescolare in modo casuale l’ordine dei pezzi prova, così da evitare che la partizione possa risentire di eventuali ordinamenti preesistenti nel dataset.

Successivamente è stato determinato il numero totale di pezzi prova (`n_tp`) e, in proporzione, il numero di elementi destinati all’addestramento (70%) e alla

validazione (15%), lasciando implicitamente il restante 15% per il test. Questa proporzione è stata scelta per bilanciare l'esigenza di disporre di un set di addestramento sufficientemente ampio con quella di mantenere campioni statisticamente significativi per la validazione e la valutazione finale.

Dopo il mescolamento, l'insieme dei pezzi prova viene diviso in tre sottoinsiemi distinti. In questo modo, ciascun pezzo e tutte le sue misurazioni vengono assegnati in modo coerente a un solo gruppo, evitando che lo stesso componente possa comparire in più insiemi. Tale accorgimento è fondamentale per prevenire fenomeni di contaminazione dei dati e garantire una valutazione imparziale della capacità di generalizzazione del modello.

Successivamente, la suddivisione ottenuta viene salvata in un file in formato JSON [45]: il salvataggio della partizione in un file esterno consente di preservare la tracciabilità del processo, garantendo che i confronti tra i diversi approcci di bilanciamento, come SMOTENC e Random Over Sampling (ROS), avvengano sulle medesime basi di addestramento, validazione e test, evitando che differenze nella composizione dei sottoinsiemi influenzino i risultati sperimentali.

In seguito, sono state definite le maschere logiche utilizzate per selezionare le righe del dataset appartenenti a ciascun sottoinsieme: le funzioni `isin()` restituiscono vettori booleani che indicano, per ogni riga del dataset, se il pezzo corrispondente appartiene all'insieme di addestramento, validazione o test.

Infine, tali maschere vengono applicate per estrarre i dati e le etichette corrispondenti: le variabili indipendenti (X), le etichette di classificazione (y) e i gruppi di appartenenza (groups) vengono suddivisi coerentemente secondo le maschere precedentemente definite. Il risultato è una tripla di insiemi di dati completamente disgiunti, coerenti tra loro e rappresentativi delle rispettive fasi di addestramento, validazione e test.

```

1 #SPLIT FISSO CARICATO DA FILE
2 with open(OUT_DIR / SPLIT_FILE) as f:
3     split = json.load(f)
4
5 train_tp = set(split["train_tp"])
6 val_tp = set(split["val_tp"])
7 test_tp = set(split["test_tp"])
8
9 train_mask = df["TP"].isin(train_tp)
10 val_mask = df["TP"].isin(val_tp)
11 test_mask = df["TP"].isin(test_tp)
12
13 X_train, X_val, X_test = X[train_mask], X[val_mask], X[test_mask]
14 y_train, y_val, y_test = y[train_mask], y[val_mask], y[test_mask]
```

Figura 4.7: Caricamento dello split fisso per la separazione train-test.

La Figura 4.7, invece, mostra la procedura di caricamento dello split fisso precedentemente salvato, adottata nello script dedicato alla tecnica di oversampling ROS (Random OverSampler) [46].

In questo caso, invece di ricreare una nuova suddivisione, viene importato il file json generato nello script SMOTENC, che contiene l'elenco dei gruppi TP assegnati rispettivamente all'addestramento, validazione e test.

Tale scelta consente di mantenere identici insiemi di dati per entrambi gli esperimenti, garantendo che eventuali differenze nelle prestazioni dei modelli siano imputabili esclusivamente alla diversa tecnica di oversampling impiegata, e non a variazioni nei dati di input.

Questo approccio rispetta i principi di riproducibilità sperimentale e confrontabilità diretta dei risultati, due requisiti fondamentali in contesti di ricerca applicata in ambito machine learning.

4.4.2 Pre-elaborazione dei dati

È stato necessario implementare un processo strutturato di pre-elaborazione per rendere il dataset adatto all'addestramento di modelli di apprendimento automatico. Le variabili categoriali presenti (*Element, Pocket, tool state, Direction e Corner*) sono state trasformate utilizzando One-Hot Encoding, una tecnica che consente di convertire ciascuna categoria in un set di variabili binarie, evitando l'introduzione di ordini arbitrari tra le modalità. Questo passaggio è fondamentale per garantire che gli algoritmi possano trattare correttamente la natura qualitativa delle informazioni.

Parallelamente, StandardScaler ha applicato una trasformazione di standardizzazione, utilizzando la media zero e la deviazione standard unitaria, per normalizzare la variabile numerica "*Nominal*". Questo step è particolarmente importante per modelli sensibili alla scala delle variabili in input, come le reti neurali (MLP) e la regressione logistica.

Nel caso delle reti neurali, l'aggiornamento dei pesi avviene tramite la discesa del gradiente: una scala eterogenea delle variabili può causare gradienti instabili, con il rischio di oscillazioni o, al contrario, di un apprendimento eccessivamente lento. Allo stesso modo, anche nella regressione logistica la standardizzazione consente di migliorare la stabilità numerica durante l'ottimizzazione dei coefficienti [32].

Pertanto, grazie alla standardizzazione, tutte le variabili risultano confrontabili, evitando distorsioni numeriche, migliorando la stabilità del processo di addestramento e favorendo una più rapida convergenza verso soluzioni ottimali. L'intera catena di pre-elaborazione è stata integrata in un processo automatizzato, mostrato in Figura 4.8, così da garantire la consistenza delle trasformazioni sia nella fase di addestramento che in quella di inferenza [47].

```

1 features_cat = ["Element", "Pocket", "tool_state", "Direction", "Corner"]
2 features_num = ["Nominal"]
3
4 preprocessor = ColumnTransformer([
5     ("cat", OneHotEncoder(handle_unknown="ignore"), features_cat),
6     ("num", StandardScaler(), features_num),]

```

Figura 4.8: Creazione del pre-processore per la gestione delle variabili categoriche e numeriche.

Il pre-processore così definito è stato integrato all'interno delle pipeline di addestramento e validazione. In questo modo, le trasformazioni sulle variabili categoriche e numeriche vengono eseguite automaticamente in ciascun ciclo di cross-validation, esclusivamente sui dati di training del fold considerato e poi replicate sul corrispondente set di validazione o test. Questo approccio garantisce la coerenza delle trasformazioni, previene fenomeni di leakage informativo e consente di mantenere un flusso unificato che include pre-elaborazione, oversampling e classificazione. Le fasi di addestramento, validazione e test, illustrate attraverso le pipeline implementate, verranno descritte nel dettaglio nel paragrafo successivo [48].

Dopo la definizione del pre-processore, è stata applicata la fase di creazione delle pipeline di addestramento, all'interno delle quali vengono integrate le tecniche di oversampling per la gestione dello sbilanciamento delle classi.

```

1 # Adatta temporaneamente il preprocessor per ottenere gli indici
2 temp_processor = preprocessor.fit(X_train)
3 categorical_features_indices = list(range(len(temp_processor.
4     named_transformers_["cat"].get_feature_names_out())))
5 pipelines = {
6     name: Pipeline([
7         ("pre", preprocessor),
8         ("over", SMOTENC(categorical_features=categorical_indices,
9             random_state=RANDOM_STATE)),
10        ("model", model)])
11     for name, model in base_models.items()}

```

Figura 4.9: Creazione della pipeline con SMOTENC.

La Figura 4.9 mostra il codice in cui è stata adottata la tecnica SMOTENC tramite il parametro `categorical_features`, che consente di specificare le variabili categoriche da trattare in modo diverso rispetto alle numeriche. In particolare, per determinare correttamente gli indici delle variabili categoriche, è stato necessario

eseguire temporaneamente l’adattamento del preprocessore sui dati di addestramento, utilizzando il metodo `fit` per identificare e successivamente applicare la codifica corretta durante la fase di oversampling.

```

1 pipelines = {
2     name: Pipeline([
3         ("pre", preprocessor),
4         ("over", RandomOverSampler(random_state=RANDOM_STATE)),
5         ("model", model)])
6     for name, model in base_models.items()
7 }
```

Figura 4.10: Creazione della pipeline con ROS.

Nel codice in Figura 4.10, invece, è stato scelto il RandomOverSampler (ROS). A differenza di SMOTENC, questa tecnica non necessita di trattare le variabili categoriche in modo esplicito, poiché si limita a duplicare i campioni già presenti nel dataset senza generare nuovi campioni sintetici. La scelta di ROS è stata fatta per confrontare l’efficacia di un approccio meno complesso rispetto a SMOTENC, valutando se la semplice duplicazione dei campioni sia altrettanto efficace nella gestione dello sbilanciamento. Entrambe le tecniche sono state integrate nelle pipeline, permettendo di eseguire il bilanciamento prima della fase di addestramento del modello, con l’intento di confrontare l’efficacia di ciascuna nella gestione dello sbilanciamento.

4.5 Processo di valutazione

Le fasi di addestramento, validazione e test fanno parte di un processo cruciale per allenare e valutare i modelli in modo rigoroso. Ogni fase è fondamentale per garantire che il modello eviti fenomeni di overfitting o underfitting e che i risultati ottenuti siano generalizzabili sui dati reali. Di seguito vengono descritti in dettaglio le tre fasi.

1. Fase di addestramento

L’addestramento è stato condotto sui dati del training set, che rappresentano il 70% del totale, dopo l’applicazione delle operazioni di pre-processing e delle tecniche di riequilibrio delle classi. A seconda della configurazione sperimentale, i dati di addestramento sono stati riequilibrati utilizzando una delle due tecniche considerate (ROS o SMOTENC) applicate esclusivamente in questa fase per non alterare la distribuzione dei set di validazione e test. Il modello, inserito nella pipeline, è stato quindi addestrato invocando il metodo `fit()` sulla coppia (`X_train, y_train`). Ogni pipeline è stata gestita in modo indipendente, permettendo di confrontare le prestazioni di diversi classificatori in condizioni di addestramento coerenti e controllate.

2. Fase di validazione

Una volta completato l’addestramento, ciascun modello è stato valutato sul set di validazione, pari al 15% del dataset complessivo, mantenuto completamente indipendente rispetto ai dati di training. Le predizioni sono state ottenute mediante i metodi `predict()` e, quando disponibile, `predict_proba()`, che consente di calcolare la probabilità di appartenenza alla classe positiva (“difettoso”). Sulla base dei risultati, sono state calcolate le principali metriche di valutazione: F1-score ponderato, precisione ponderata, recall ponderato, ROC AUC e PR AUC. L’uso dell’F1-score ponderato come metrica principale consente di valutare l’equilibrio tra precisione e sensibilità, tenendo conto della distribuzione sbilanciata delle classi. Le metriche aggiuntive, come ROC AUC e PR AUC, forniscono informazioni complementari sulla capacità discriminante del modello e sulla sua affidabilità nelle previsioni probabilistiche. I valori medi di F1-score ponderato ottenuti in validazione, compresi tra 0.784 e 0.907 a seconda del modello e della tecnica di bilanciamento applicata (ROS o SMOTENC), confermano che i classificatori sono in grado di apprendere efficacemente le relazioni tra le variabili e di distinguere entrambe le classi. Ciò suggerisce una buona capacità di generalizzazione già all’interno del training set, senza segnali marcati di overfitting.

3. Fase di test

Infine, i modelli sono stati sottoposti alla fase di test, utilizzando il rimanente 15% dei dati, completamente separato dai set di addestramento e validazione. In questa fase, ciascun modello precedentemente addestrato è stato applicato al test set senza ulteriori modifiche dei parametri. Analogamente alla validazione, le previsioni sono state generate mediante i metodi `predict()` e `predict_proba()`, consentendo di valutare sia le classificazioni discrete sia le probabilità di appartenenza. Le metriche calcolate, accuracy, precisione, recall e F1-score, hanno permesso di ottenere una misura complessiva delle prestazioni, mentre le curve ROC e PR hanno offerto una rappresentazione più dettagliata della capacità del modello di distinguere tra classi in condizioni di incertezza. L’utilizzo di un set di test completamente indipendente ha assicurato una valutazione imparziale e realistica delle prestazioni del modello, fornendo un’indicazione attendibile della sua capacità di generalizzare su nuovi dati e di operare efficacemente in un contesto produttivo reale.

Di seguito, in Figura 4.11, viene riportato il codice che implementa l’intero processo di addestramento, validazione e test:

```

1 summary = []
2
3 for model_name, pipeline in pipelines.items():
4     print(f"\n[SMOTENC BY-TP 70/15/15] - {model_name}")
5
6     #TRAINING SU TRAIN SET (70%)
7     pipeline.fit(X_train, y_train)
8
9     #VALIDAZIONE SU SET DEDICATO (15%)
10    y_val_pred = pipeline.predict(X_val)
11    y_val_proba = pipeline.predict_proba(X_val)[:, 1] if hasattr(
12        pipeline.named_steps["model"], "predict_proba") else None
13
14    val_metrics = {
15        "F1_weighted_val": f1_score(y_val, y_val_pred, average="weighted",
16                                     zero_division=0),
17        "Precision_weighted_val": precision_score(y_val,
18                                                    y_val_pred, average="weighted",
19                                     zero_division=0),
20        "Recall_weighted_val": recall_score(y_val, y_val_pred,
21                                             average="weighted", zero_division=0),
22        "ROC_AUC_val": roc_auc_score(y_val, y_val_proba) if
23        y_val_proba is not None else None,
24        "PR_AUC_val": average_precision_score(y_val, y_val_proba)
25        if y_val_proba is not None else None
26    }
27
28    print(f"Validazione F1_w: {val_metrics['F1_weighted_val']:.3f} | "
29          f"ROC_AUC: {val_metrics['ROC_AUC_val']} if val_metrics['"
30          "ROC_AUC_val'] is not None else 'n/a' | "
31          f"PR_AUC: {val_metrics['PR_AUC_val']} if val_metrics['"
32          "PR_AUC_val'] is not None else 'n/a'}")
33
34    #TEST SU TEST SET (15%) CON LO STESSO MODELLO
35    y_test_pred = pipeline.predict(X_test)
36    y_test_proba = pipeline.predict_proba(X_test)[:, 1] if hasattr(
37        pipeline.named_steps["model"], "predict_proba") else None

```

Figura 4.11: Implementazione delle fasi di training, validazione incrociata e test del modello.

Per facilitare la comprensione del funzionamento dei modelli predittivi e della natura dell'output generato, nelle Tabelle 4.2, 4.3, 4.4 e 4.5 sono riportati tre estratti rappresentativi, contenenti le predizioni sui primi 12 campioni appartenenti al gruppo di lavorazione con codice TP = 3 con l'utilizzo della tecnica SMOTENC. Ogni riga mostra i valori delle variabili originali, il valore della deviazione geometrica rispetto al nominale (Dev), l'etichetta reale (Reale) e quella predetta dal

modello (Predetto), oltre al punteggio di probabilità associato alla classe “difettoso” (Score_difettoso).

Come si può osservare dalle figure, tutti i modelli implementati risultano in grado di tradurre la misura continua della deviazione in un’etichetta binaria (difettoso / non difettoso), mostrando nel complesso una buona capacità di apprendimento della soglia di non conformità. Tuttavia, i classificatori differiscono nel livello di confidenza associato alle predizioni e nel modo in cui stimano la probabilità di appartenenza alle due classi. Il modello MLP tende a restituire probabilità più elevate e concentrate, mostrando un comportamento più “deterministico” e una minore incertezza nelle previsioni. La regressione logistica (MLR) si conferma più conservativa, con valori di probabilità distribuiti su un intervallo più ampio e una maggiore cautela nel distinguere i casi limite. La Random Forest si colloca in una posizione intermedia, bilanciando la stabilità con una buona capacità discriminante. Il modello LightGBM, infine, mostra un profilo di risposta più regolare e coerente tra i diversi gruppi di misura, combinando una buona sensibilità con una distribuzione calibrata delle probabilità. Questa diversità di comportamento conferma come l’analisi comparativa non debba limitarsi alla sola accuratezza, ma debba considerare anche la qualità e la coerenza della confidenza associata alle predizioni, un elemento cruciale per l’affidabilità operativa del Digital Twin in contesti industriali reali.

Tabella 4.2: Estratto delle previsioni del modello di classificazione sul gruppo di lavorazione TP = 3.

Element	Nominal	Actual	Dev	Pocket	TP	Corner	Direction	tool_state	Reale	Predetto	Score_difettoso
Cylinder	0.35	0.38	0.03	pocket1	3	A	X	New	difettoso	difettoso	0.7597199756237435
Cylinder	0.35	0.37	0.02	pocket1	3	A	Y	New	difettoso	difettoso	0.7946518584468486
Cylinder	0.35	0.38	0.03	pocket1	3	A	Z	New	difettoso	difettoso	0.756424041621817
Cylinder	0.35	0.38	0.03	pocket1	3	B	X	New	difettoso	difettoso	0.6417451429927117
Cylinder	0.35	0.39	0.04	pocket1	3	B	Y	New	difettoso	difettoso	0.7197471750234306
Cylinder	0.35	0.38	0.03	pocket1	3	B	Z	New	difettoso	difettoso	0.6195412175740304
Cylinder	0.35	0.37	0.02	pocket1	3	C	X	New	difettoso	difettoso	0.5834415662182862
Cylinder	0.35	0.38	0.03	pocket1	3	C	Y	New	difettoso	difettoso	0.831824871907568
Cylinder	0.35	0.36	0.01	pocket1	3	C	Z	New	difettoso	difettoso	0.5647474621365768
Cylinder	0.35	0.37	0.02	pocket1	3	D	X	New	difettoso	non difettoso	0.4398885653742606
Cylinder	0.35	0.41	0.06	pocket1	3	D	Y	New	difettoso	difettoso	0.6709138757032658
Cylinder	0.35	0.37	0.02	pocket1	3	D	Z	New	difettoso	non difettoso	0.4612496741063868

Tabella 4.3: Estratto delle previsioni del modello MLP sul gruppo di lavorazione TP = 3.

Element	Nominal	Actual	Dev	Pocket	TP	Corner	Direction	tool_state	Reale	Predetto	Score_difettoso
Cylinder	0.35	0.38	0.03	pocket1	3	A	X	New	difettoso	difettoso	0.999999990419921
Cylinder	0.35	0.37	0.02	pocket1	3	A	Y	New	difettoso	difettoso	0.9999898545376851
Cylinder	0.35	0.38	0.03	pocket1	3	A	Z	New	difettoso	difettoso	0.9967757439010144
Cylinder	0.35	0.38	0.03	pocket1	3	B	X	New	difettoso	difettoso	0.999999998776574
Cylinder	0.35	0.39	0.04	pocket1	3	B	Y	New	difettoso	difettoso	0.9999991841904019
Cylinder	0.35	0.38	0.03	pocket1	3	B	Z	New	difettoso	difettoso	0.9500670849006286
Cylinder	0.35	0.37	0.02	pocket1	3	C	X	New	difettoso	difettoso	0.8052614307891454
Cylinder	0.35	0.38	0.03	pocket1	3	C	Y	New	difettoso	difettoso	0.99999999472167
Cylinder	0.35	0.36	0.01	pocket1	3	C	Y	New	difettoso	difettoso	0.8487804499183963
Cylinder	0.35	0.37	0.02	pocket1	3	D	X	New	difettoso	non difettoso	0.1867774668600339
Cylinder	0.35	0.41	0.06	pocket1	3	D	Y	New	difettoso	difettoso	0.999999825372755
Cylinder	0.35	0.37	0.02	pocket1	3	D	Z	New	difettoso	non difettoso	0.3102720870714091

Tabella 4.4: Estratto delle previsioni del modello MLR sul gruppo di lavorazione TP = 3.

Element	Nominal	Actual	Dev	Pocket	TP	Corner	Direction	tool_state	Reale	Predetto	Score_difettoso
Cylinder	0.35	0.38	0.03	pocket1	3	A	X	New	difettoso	non difettoso	0.2478488570425544
Cylinder	0.35	0.37	0.02	pocket1	3	A	Y	New	difettoso	non difettoso	0.28965611947243985
Cylinder	0.35	0.38	0.03	pocket1	3	A	Z	New	difettoso	non difettoso	0.14941594822914886
Cylinder	0.35	0.38	0.03	pocket1	3	B	X	New	difettoso	difettoso	0.5005879684382879
Cylinder	0.35	0.39	0.04	pocket1	3	B	Y	New	difettoso	difettoso	0.5536465318856673
Cylinder	0.35	0.38	0.03	pocket1	3	B	Z	New	difettoso	non difettoso	0.3482548290751864
Cylinder	0.35	0.37	0.02	pocket1	3	C	X	New	difettoso	difettoso	0.6978141068112902
Cylinder	0.35	0.38	0.03	pocket1	3	C	X	New	difettoso	difettoso	0.7407698657195362
Cylinder	0.35	0.36	0.01	pocket1	3	C	Y	New	difettoso	difettoso	0.5517735955544518
Cylinder	0.35	0.37	0.02	pocket1	3	D	X	New	difettoso	difettoso	0.6195651498168526
Cylinder	0.35	0.41	0.06	pocket1	3	D	Y	New	difettoso	difettoso	0.668357655780641
Cylinder	0.35	0.37	0.02	pocket1	3	D	Z	New	difettoso	non difettoso	0.46471671182264707

Tabella 4.5: Estratto delle previsioni del modello LightGBM sul gruppo di lavorazione TP = 3.

Element	Nominal	Actual	Dev	Pocket	TP	Corner	Direction	tool_state	Reale	Predetto	Score_difettoso
Cylinder	0.35	0.38	0.03	pocket1	3	A	X	New	difettoso	difettoso	0.9929110302354331
Cylinder	0.35	0.37	0.02	pocket1	3	A	Y	New	difettoso	difettoso	0.9966509965758309
Cylinder	0.35	0.38	0.03	pocket1	3	A	Z	New	difettoso	difettoso	0.9782414060242234
Cylinder	0.35	0.38	0.03	pocket1	3	B	X	New	difettoso	difettoso	0.9900379338638184
Cylinder	0.35	0.39	0.04	pocket1	3	B	Y	New	difettoso	difettoso	0.9961375708746224
Cylinder	0.35	0.38	0.03	pocket1	3	B	Z	New	difettoso	difettoso	0.969756353742098
Cylinder	0.35	0.37	0.02	pocket1	3	C	X	New	difettoso	difettoso	0.7015085508095471
Cylinder	0.35	0.38	0.03	pocket1	3	C	X	New	difettoso	difettoso	0.9972243894414697
Cylinder	0.35	0.36	0.01	pocket1	3	C	Y	New	difettoso	difettoso	0.904565332380854
Cylinder	0.35	0.37	0.02	pocket1	3	D	X	New	difettoso	non difettoso	0.2495314651809319
Cylinder	0.35	0.41	0.06	pocket1	3	D	Y	New	difettoso	difettoso	0.9746396766253924
Cylinder	0.35	0.37	0.02	pocket1	3	D	Z	New	difettoso	non difettoso	0.3161722764694757

Questa fase rappresenta un passaggio cruciale per l’evoluzione del Digital Twin da semplice replica informativa a strumento diagnostico e decisionale. Le procedure sviluppate, attraverso passaggi di pre-elaborazione coerenti e modulari, hanno garantito uniformità nel trattamento dei dati e replicabilità nel flusso di addestramento e inferenza. A differenza dell’approccio proposto da Boettjer e Ramanujan, in cui le singole misurazioni relative a uno stesso pezzo potevano essere distribuite tra training e test, nel presente lavoro la suddivisione del dataset è stata effettuata per gruppi di TP. In questo modo, tutte le feature appartenenti a uno stesso componente sono state mantenute all’interno dello stesso insieme (train, validation o test), evitando contaminazioni tra i set e simulando in modo più realistico l’operatività del Digital Twin su nuovi pezzi non ancora analizzati.

Una volta completato l’addestramento, i modelli sono stati applicati nel contesto dell’architettura di Digital Twin implementata per il caso studio dimostrativo. L’obiettivo non è limitarsi a restituire una previsione statistica complessiva, ma tradurre le misure fisiche e le variabili categoriali raccolte durante la lavorazione CNC in un output diagnostico interpretabile a livello di singola variabile. In questo modo, il DT non si limita a stabilire se un intero componente sia conforme o difettoso, ma permette di localizzare il difetto in una porzione specifica del pezzo, abilitando una diagnosi puntuale e coerente con l’approccio adottato anche dagli autori del caso studio originale.

Questo output categoriale costituisce il collegamento decisionale tra il DT e il sistema reale: una volta integrato nel ciclo operativo, consente di identificare in tempo reale non solo i componenti non conformi, ma anche le variabili critiche da monitorare, attivando così azioni correttive mirate (ad es. scarto del pezzo, segnalazione di manutenzione, aggiornamento dei parametri di processo). In questo senso, la classificazione binaria non rappresenta il punto di arrivo, ma l’elemento chiave che consente al DT di trasformarsi da replica passiva a strumento attivo di supporto al controllo qualità e alla manutenzione predittiva.

I risultati preliminari, mostrati negli estratti di predizione, evidenziano come ciascun modello sia in grado di restituire una classificazione affidabile basata sulle deviazioni dimensionali delle singole feature, offrendo un supporto diretto al processo di controllo qualità. Si osserva come le previsioni non siano sempre coincidenti: ciascun algoritmo, infatti, interpreta i dati secondo logiche differenti e produce talvolta risultati discordanti. Questo aspetto riflette il fatto che non tutti i modelli raggiungono la stessa accuratezza, un risultato che pone le basi alla successiva fase di analisi quantitativa, in cui le prestazioni verranno valutate e confrontate in modo sistematico mediante metriche standard di classificazione, aprendo la strada alla discussione dei risultati e delle implicazioni pratiche che verranno approfondite nel Capitolo 5.

Capitolo 5 – Valutazione dei Risultati e Discussione

5.1 Metodologia di valutazione

Dopo aver descritto l’architettura e la procedura sperimentale del Digital Twin, questa sezione analizza criticamente i risultati ottenuti, confrontando le diverse configurazioni di oversampling e i modelli di classificazione in termini di accuratezza e stabilità diagnostica. Per garantire una valutazione approfondita e comparabile delle prestazioni dei modelli predittivi integrati nel Digital Twin, sono state utilizzate una varietà di metriche di classificazione, ampiamente documentate nella letteratura e utilizzate negli applicativi di machine learning supervisionato. L’obiettivo è fornire una valutazione approfondita della capacità dei modelli di distinguere correttamente tra i due stati considerati (difettoso o non difettoso). Ciò viene fatto prendendo in considerazione sia gli elementi di accuratezza globale che le prestazioni in termini di bilanciamento tra classi.

Le metriche selezionate sono le seguenti: matrice di confusione, Accuracy, Precision, Recall (o sensibilità), F1-Score, ROC-AUC (Receiver Operating Characteristic – Area Under the Curve), PR-AUC (Precision-Recall Area Under the Curve).

L’analisi delle metriche ha innanzitutto consentito di valutare l’efficacia dei due metodi di oversampling nel gestire lo sbilanciamento delle classi, individuando quale approccio garantisse un miglior equilibrio tra le prestazioni dei modelli. In particolare, il confronto ha permesso di evidenziare quale tecnica producesse risultati più stabili e bilanciati sia in termini di accuratezza complessiva, sia nella capacità di riconoscere correttamente entrambe le classi (difettoso e non difettoso). Parallelamente, l’analisi comparativa dei classificatori ha reso possibile identificare il modello più performante nel contesto specifico del Digital Twin, ovvero quello in grado di combinare al meglio robustezza predittiva e affidabilità diagnostica.

1. Matrice di confusione

Rappresenta la tabella di contingenza che mette a confronto le etichette

predette dal modello con quelle reali. È composta da quattro elementi fondamentali:

Tabella 5.1: Struttura della matrice di confusione.

Predizione del modello		
Classe reale	Difettoso	Non difettoso
Difettoso	TP (True Positive)	FN (False Negative)
Non difettoso	FP (False Positive)	TN (True Negative)

La Tabella 5.1 mostra la struttura tipica della matrice di confusione. Ogni cella rappresenta l'esito della classificazione rispetto alla realtà: i True Positive (TP) sono i campioni difettosi correttamente individuati, i True Negative (TN) sono i campioni conformi riconosciuti come tali, mentre i False Positive (FP) e i False Negative (FN) rappresentano rispettivamente i falsi allarmi e i difetti non rilevati.

La matrice di confusione è uno strumento fondamentale per comprendere il tipo e la distribuzione degli errori commessi dal modello e costituisce la base per il calcolo delle principali metriche di valutazione descritte di seguito.

2. Accuracy

Misura la proporzione di campioni correttamente classificati (sia difettosi che non difettosi) rispetto al numero totale di campioni. È definita come:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Un valore elevato di accuracy indica che il modello riesce a distinguere correttamente la maggior parte dei campioni. Tuttavia, nei dataset sbilanciati può risultare fuorviante, perché tende a riflettere soprattutto la classe maggioritaria: per questo motivo è importante considerarla insieme a metriche più robuste.

3. Precision

Misura la proporzione di campioni correttamente classificati come difettosi rispetto al totale dei campioni predetti come difettosi. Formalmente:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Una precisione elevata indica che il modello commette pochi falsi positivi, evitando di etichettare come difettosi prodotti che in realtà sono conformi.

4. Recall (o sensibilità)

Esprime la capacità del modello di identificare correttamente tutti i campioni difettosi, calcolata come:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Un valore elevato di recall è essenziale in scenari di controllo qualità, dove non è accettabile lasciare prodotti difettosi non rilevati.

5. F1-Score

È la media armonica tra precision e recall, fornendo una misura complessiva della capacità del modello di classificazione di individuare correttamente la classe positiva, bilanciando i due tipi di errore possibili: falsi positivi (FP) e falsi negativi (FN):

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Tale indice è particolarmente indicato nei casi di dataset con classi sbilanciate o quando entrambe le tipologie di errore hanno conseguenze significative.

6. Curva ROC-AUC (Receiver Operating Characteristic – Area Under the Curve)

La curva ROC rappresenta la capacità del modello di distinguere tra le classi difettoso e non difettoso al variare della soglia di classificazione. Sull'asse orizzontale è riportato il tasso di falsi positivi (False Positive Rate), mentre sull'asse verticale compare il tasso di veri positivi (True Positive Rate). La forma della curva mostra il compromesso tra sensibilità e specificità del modello, mentre l'area sottesa (AUC) sintetizza la qualità discriminativa complessiva: valori prossimi a 1 indicano un'elevata capacità di separazione tra le due classi, mentre un valore pari a 0.5 corrisponde a un comportamento puramente casuale.

7. Curva PR-AUC (Precision-Recall Area Under the Curve)

La curva Precision-Recall valuta il compromesso tra precision (affidabilità delle predizioni positive) e recall (capacità di individuare correttamente i difettosi). Sull'asse orizzontale è riportato il recall (True Positive Rate), mentre sull'asse verticale compare la precision (Positive Predictive Value). Analizzando la forma della curva è possibile comprendere quanto il modello riesca a mantenere un'elevata precision anche al crescere della recall. L'area sottesa alla curva (PR-AUC) fornisce un indice sintetico di questa relazione ed è particolarmente significativa in presenza di classi sbilanciate, poiché si concentra sulle prestazioni della classe positiva (difettoso). Valori più elevati di

PR-AUC indicano un modello più efficace nel bilanciare sensibilità e affidabilità delle predizioni.

Poichè in questo caso di studio è presente una marcata asimmetria nella distribuzione delle classi, con un numero di campioni “difettosi” nettamente superiore rispetto a quelli “non difettosi”, l'utilizzo diretto delle metriche classiche potrebbe condurre a una sovrastima artificiale delle prestazioni del modello, in quanto quest'ultimo potrebbe ottenere un'elevata accuratezza semplicemente predicendo la classe maggioritaria.

Per ovviare a questo problema, in linea con l'approccio adottato nell'articolo di riferimento, si è scelto di utilizzare le versioni pesate (weighted) delle principali metriche di classificazione: Precision, Recall e F1-Score. Tali versioni calcolano il valore della metrica per ciascuna classe e ne effettuano la media ponderata in base alla numerosità dei campioni appartenenti a quella classe, secondo la formula generale:

$$M_{\text{weighted}} = \sum_{i=1}^C w_i \cdot M_i \quad \text{con} \quad w_i = \frac{n_i}{\sum_{j=1}^C n_j}$$

dove M_i rappresenta il valore della metrica per la classe i , n_i il numero di osservazioni di tale classe e C il numero totale di classi.

L'approccio pesato consente di ottenere una valutazione più realistica e rappresentativa del comportamento del modello sull'intero dataset, mantenendo la proporzionalità rispetto alla distribuzione originale delle classi. In questo modo, le prestazioni non vengono influenzate dalla dominanza della classe maggioritaria, ma riflettono la capacità del modello di riconoscere correttamente entrambe le categorie (“difettoso” e “non difettoso”) in condizioni di squilibrio.

Come mostrato in Figura 5.1, la matrice di confusione è calcolata utilizzando la funzione `confusion_matrix` di `sklearn.metrics`[49], che prende in input le etichette reali (`y_test`) e le etichette predette dal modello (`y_test_pred`). Successivamente, la matrice risultante viene visualizzata e salvata in un file CSV per un'analisi approfondita.

A scopo esemplificativo, il codice riportato fa riferimento allo script in cui è stato applicato l'oversampling SMOTENC, ma questa procedura è stata implementata in modo analogo anche nello script relativo al Random Over Sampling (ROS).

```

1 for model_name, pipeline in models.items():
2     #MATRICE DI CONFUSIONE
3     cm = confusion_matrix(y_test, y_test_pred)
4     # Salva la matrice di confusione in un file CSV
5     cm_df = pd.DataFrame(
6         cm,
7         index=["non difettoso", "difettoso"],
8         columns=["Predetto: non difettoso", "Predetto: difettoso"
9     ])
10    cm_df.to_csv(OUT_DIR / f"confusion_matrix_SMOTENC_{model_name}.csv", index=True)
11    cm_df = pd.DataFrame(cm, index=["non difettoso", "difettoso"],
12        columns=["Predetto: non difettoso", "Predetto: difettoso"])
13    cm_df.to_csv(OUT_DIR / f"confusion_matrix_{ds_name}_{model_name}.csv", index=True)

```

Figura 5.1: Calcolo della matrice di confusione.

Dal calcolo delle matrici di confusione, segue l'implementazione dell'intero processo di valutazione dei modelli di classificazione addestrati nel contesto del Digital Twin. Come rappresentato nel codice in Figura 5.2, sono state calcolate sia le predizioni discrete (y_{test_pred}) sia, se disponibili, le probabilità di appartenenza alla classe positiva (y_{test_proba}). Queste ultime risultano essenziali per la costruzione delle curve ROC e Precision-Recall.

A partire dalle predizioni ottenute, vengono calcolate le principali metriche di classificazione: accuracy, precision, recall, F1-score e le aree sotto la curva ROC (ROC-AUC) e sotto la curva Precision-Recall (PR-AUC). Questi valori vengono salvati e visualizzati per ciascun modello, offrendo una panoramica quantitativa delle rispettive performance[49][50][51].

```

1 for model_name, pipeline in models.items():
2     summary.append({
3         "Modello": model_name,
4         **val_metrics,
5         "Accuracy_test": (y_test_pred == y_test).mean(),
6         "Precision_weighted_test": precision_score(y_test, y_test_pred,
7             average="weighted", zero_division=0),
8         "Recall_weighted_test": recall_score(y_test, y_test_pred,
9             average="weighted", zero_division=0),
10        "F1_weighted_test": f1_score(y_test, y_test_pred, average="weighted",
11            zero_division=0),
12        "ROC-AUC_test": roc_auc_score(y_test, y_test_proba) if
13        y_test_proba is not None else None,
14        "PR-AUC_test": average_precision_score(y_test, y_test_proba)
15        if y_test_proba is not None else None
16    })

```

Figura 5.2: Calcolo delle metriche di classificazione.

L'utilizzo congiunto di queste metriche consente di analizzare in profondità le prestazioni dei modelli, andando oltre la semplice accuratezza complessiva e fornendo una valutazione robusta del loro comportamento in contesti applicativi reali.

5.2 Risultati dei modelli

Durante la fase di valutazione, le prestazioni dei modelli sono state analizzate utilizzando una soglia di classificazione pari a 0,5, valore comunemente adottato nei modelli di classificazione probabilistica. In questa configurazione, le osservazioni con una probabilità predetta superiore a 0,5 vengono classificate come difettose, mentre quelle con probabilità inferiori vengono assegnate alla classe non difettosa. Sebbene tale soglia rappresenti una scelta standard che consente di confrontare in modo omogeneo le prestazioni dei diversi algoritmi, essa non corrisponde necessariamente al punto di massima efficacia del modello; essa può essere successivamente adattata in funzione degli obiettivi applicativi, ad esempio per privilegiare la sensibilità o la precisione in specifici contesti industriali. Tale flessibilità conferisce al modello una maggiore adattabilità operativa, senza pregiudicare la validità delle valutazioni comparative qui presentate.

L'adozione della soglia predefinita comporta un equilibrio diretto tra precision e recall: soglie inferiori avrebbero incrementato la sensibilità (recall) a scapito di un maggior numero di falsi positivi, mentre soglie più elevate avrebbero ridotto i falsi positivi, aumentando però il rischio di falsi negativi.

L'analisi comparativa riportata nella Tabella 5.2 evidenzia come l'utilizzo della tecnica di oversampling SMOTENC consenta in media di ottenere prestazioni superiori rispetto a ROS per tutte le principali metriche operative, in particolare Accuracy, F1-score weighted e Recall weighted, in ciascuno dei modelli considerati.

Tabella 5.2: Confronto delle metriche principali tra ROS e SMOTENC per ciascun modello.

Modello	Oversampling	Accuracy	F1 (w)	Precision (w)	Recall (w)	AUC-ROC	PR-AUC
Random Forest	ROS	0.810	0.833	0.888	0.810	0.857	0.960
	SMOTENC	0.843	0.858	0.891	0.843	0.846	0.955
MLP	ROS	0.773	0.803	0.884	0.773	0.858	0.961
	SMOTENC	0.867	0.869	0.871	0.867	0.860	0.964
MLR	ROS	0.729	0.767	0.860	0.729	0.806	0.947
	SMOTENC	0.788	0.811	0.857	0.788	0.805	0.945
LightGBM	ROS	0.776	0.806	0.885	0.776	0.869	0.969
	SMOTENC	0.871	0.873	0.875	0.871	0.867	0.963

A differenza del ROS, che si limita a duplicare casualmente i campioni della classe minoritaria, SMOTENC genera campioni sintetici più rappresentativi, poiché tiene conto della struttura mista del dataset e delle relazioni tra variabili categoriche e numeriche. In questo modo, il modello riesce ad apprendere confini decisionali più regolari e generalizzabili, riducendo il rischio di overfitting e migliorando la capacità di individuare correttamente i casi difettosi.

Questo significa che, a parità di soglia decisionale, i classificatori addestrati con SMOTENC risultano più bilanciati e meno inclini a trascurare la classe minoritaria, riducendo quindi la probabilità di falsi negativi. Questo aspetto è cruciale nel contesto del controllo qualità, dove l'obiettivo principale è intercettare i difetti e minimizzare il rischio di classificare come “non difettoso” un componente che in realtà presenta delle non conformità. L'incremento dell'F1-score e del recall nei modelli SMOTENC dimostra infatti una migliore capacità del sistema di mantenere l'equilibrio tra sensibilità e precisione, evitando di ottimizzare una singola metrica a scapito dell'altra.

Le metriche di ranking probabilistico (AUC-ROC e PR-AUC) delineano invece un quadro più equilibrato: i valori restano complessivamente elevati per entrambi i metodi, con ROS che in alcuni casi (ad esempio per MLR) mostra un lieve vantaggio marginale nella curva PR-AUC, mentre SMOTENC migliora la coerenza della classificazione binaria, in particolare per i modelli basati su alberi decisionali come Random Forest e LightGBM. Ciò suggerisce che, sebbene ROS possa talvolta garantire una separazione leggermente più marcata tra le distribuzioni probabilistiche, SMOTENC è in grado di tradurre tale separazione in decisioni più affidabili quando viene applicata una soglia operativa, riducendo il rischio di omissioni di casi difettosi.

Nel complesso, l'adozione di SMOTENC si dimostra più coerente con l'obiettivo del Digital Twin, orientato alla rilevazione accurata dei difetti, poiché offre una classificazione più stabile e una minore incidenza di falsi negativi. Pur riconoscendo che ROS possa talvolta mantenere un vantaggio marginale nelle metriche di ranking, tali differenze non si traducono in un effettivo miglioramento della capacità di individuare correttamente i componenti difettosi.

Ciò rende SMOTENC preferibile per applicazioni reali di diagnostica automatizzata, in cui la priorità è ridurre le omissioni di difetti anche a fronte di un lieve incremento dei falsi positivi e, sulla base di tali considerazioni, si è deciso di adottare SMOTENC come tecnica di riferimento per le analisi successive.

Nelle sezioni seguenti, le metriche verranno discusse in modo più approfondito, con l'obiettivo di evidenziare i punti di forza e i limiti emersi per ciascun modello, al fine di identificare quale approccio risulti complessivamente più efficace nella diagnosi dei difetti.

5.2.1 Valutazione del modello Random Forest

La Tabella 5.3 riporta la matrice di confusione del modello Random Forest applicato al set di test. Sono presenti 371 osservazioni correttamente identificate come difettose (True Positive), mentre in basso a destra si contano 59 componenti correttamente classificati come non difettosi (True Negative). Il modello commette 15 falsi positivi, ovvero componenti non difettosi erroneamente etichettati come difettosi, e 65 falsi negativi, cioè casi di difettosità non rilevati dal modello.

L’analisi dei valori evidenzia come la Random Forest mostri una buona capacità di identificare correttamente la classe difettosa, con un numero di veri positivi nettamente superiore ai falsi negativi. Tuttavia, la presenza di 65 falsi negativi indica che in alcuni casi il modello tende a sottostimare la probabilità di difettosità, classificando come “non difettosi” alcuni campioni che presentano effettivamente deviazioni superiori alla soglia.

Dal punto di vista operativo, questa tipologia di errore risulta particolarmente critica in un contesto di controllo qualità, poiché comporta il rischio di non intercettare un difetto reale durante la fase di ispezione. Nonostante ciò, il numero contenuto di falsi positivi suggerisce che il modello mantenga una buona specificità, evitando un eccesso di segnalazioni errate e quindi un inutile scarto di componenti conformi.

Nel complesso, la matrice di confusione mostra che il modello riesce a raggiungere un equilibrio soddisfacente tra sensibilità (capacità di riconoscere i difetti) e precisione (affidabilità delle segnalazioni di difettosità). Questo comportamento indica che la Random Forest, pur con alcune limitazioni nella piena individuazione della classe minoritaria, rappresenta un classificatore robusto e ben calibrato per la diagnostica dei difetti.

Tabella 5.3: Matrice di confusione del modello Random Forest sul set di test.

		Predizione del modello	
		Difettoso	Non difettoso
Classe reale	Difettoso	371	65
	Non difettoso	15	59

Dopo l’analisi della matrice di confusione, sono state calcolate le principali metriche di valutazione per ciascuna classe. I risultati mostrano una netta superiorità delle prestazioni sulla classe difettosa rispetto alla classe non difettosa.

Tabella 5.4: Prestazioni del modello Random Forest per ciascuna classe.

Classe	Precision	Recall	F1-score	N° campioni
Non difettoso	0.476	0.797	0.596	74
Difettoso	0.961	0.851	0.903	436

La Tabella 5.4 riporta le prestazioni del modello Random Forest per ciascuna classe. Si osserva come il classificatore raggiunga un'elevata precision per la classe difettosa (0.961), indicando che la quasi totalità dei componenti segnalati come difettosi è effettivamente tale. Il valore di recall (0.851) conferma inoltre una buona capacità di identificare correttamente la maggior parte dei pezzi difettosi presenti nel set di test. Al contrario, la classe “non difettoso” presenta una precision sensibilmente più bassa (0.476), a indicare che una parte significativa dei campioni classificati come non difettosi sono in realtà difettosi. Tuttavia, il valore di recall pari a 0.797 suggerisce che il modello riesca comunque a riconoscere un'elevata quota di componenti realmente conformi.

Questa asimmetria di comportamento è in parte imputabile allo sbilanciamento del dataset originario, in cui la classe difettosa risulta più numerosa, e al fatto che la Random Forest tende a privilegiare la classe prevalente nei casi ambigui. Nel complesso, l'elevato valore dell'F1-score per la classe difettosa (0.903) evidenzia che il modello risulta particolarmente efficace nel compito principale del Digital Twin, ovvero l'individuazione dei difetti.

Tabella 5.5: Metriche complessive del modello Random Forest.

Metrica	Precision	Recall	F1-score
Macro avg	0.718	0.824	0.749
Weighted avg	0.891	0.843	0.858
Accuracy = 0.843			

La Tabella 5.5 riassume le metriche complessive del modello. La media pesata (weighted average) restituisce valori di precision (0.891), recall (0.843) e F1-score (0.858), in linea con l'accuratezza complessiva pari a 0.843. Tali risultati indicano una buona coerenza tra le diverse metriche e confermano la capacità del modello di mantenere un equilibrio tra sensibilità e specificità. La macro average, che considera le due classi con lo stesso peso indipendentemente dalla loro frequenza, presenta valori inferiori (precisione 0.718 e F1-score 0.749), evidenziando una minore uniformità di prestazioni tra le classi.

In sintesi, la Random Forest si conferma un classificatore robusto e affidabile per la diagnosi dei difetti, capace di mantenere un'elevata precisione sulla classe di interesse principale (difettoso) e di contenere, entro margini accettabili, gli errori di classificazione sui campioni conformi. Le metriche complessive evidenziano un buon equilibrio tra sensibilità e specificità, segno che il modello è in grado di generalizzare efficacemente il comportamento appreso in fase di addestramento. Le differenze di prestazione tra le due classi appaiono principalmente legate allo sbilanciamento intrinseco del dataset e alla distribuzione delle osservazioni, più che a limiti strutturali dell'algoritmo. Ciò suggerisce che un'ulteriore ottimizzazione del pre-processing o un affinamento delle tecniche di bilanciamento potrebbe contribuire a migliorare ulteriormente la capacità discriminativa complessiva del modello.

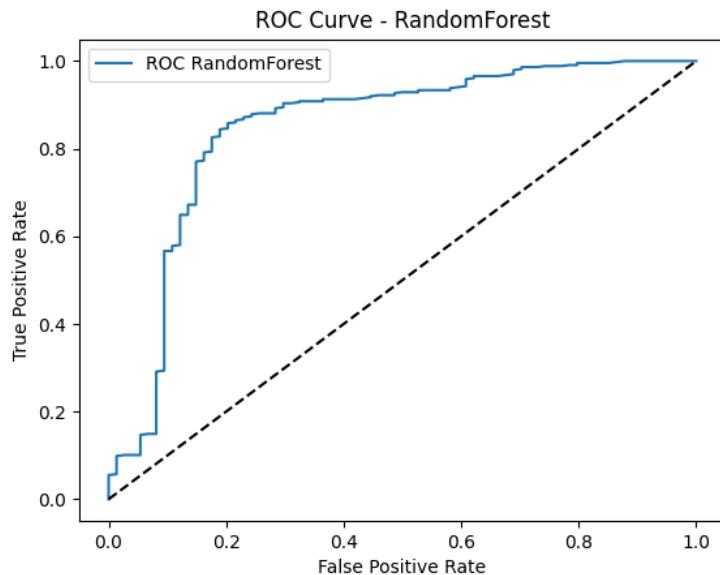


Figura 5.3: Curva ROC del modello Random Forest.

La Figura 5.3 mostra la curva ROC relativa al modello in analisi. La curva rappresenta l'andamento del True Positive Rate (sensibilità) in funzione del False Positive Rate, ossia della proporzione di campioni non difettosi che il modello classifica erroneamente come difettosi, per diversi valori di soglia di classificazione. L'obiettivo è misurare la capacità del modello di distinguere correttamente tra le due classi, “difettoso” e “non difettoso”, indipendentemente dalla soglia scelta.

Nel grafico, la curva del modello si colloca nettamente al di sopra della diagonale di riferimento (linea tratteggiata), che rappresenta il comportamento di un classificatore casuale. Questo andamento indica che la Random Forest possiede una buona capacità discriminante complessiva, riuscendo a distinguere in

modo efficace la maggior parte dei campioni difettosi, pur mostrando una minore accuratezza nella classificazione dei componenti conformi. Tuttavia, l'area sottesa alla curva (AUC-ROC) è pari a circa 0.85, in linea con i valori riportati nella Tabella 5.2, e conferma un buon compromesso tra sensibilità e specificità.

Nelle regioni iniziali della curva, caratterizzate da bassi valori di False Positive Rate, il True Positive Rate cresce rapidamente, segnalando che il modello è in grado di identificare correttamente una quota significativa di componenti difettosi senza introdurre troppi falsi allarmi. All'aumentare del False Positive Rate, la curva tende a saturarsi, evidenziando che il miglioramento in termini di sensibilità diventa progressivamente meno significativo.

Nel complesso, l'andamento regolare e l'ampia area sottesa indicano che il modello mantiene prestazioni stabili, evidenziando una buona affidabilità complessiva nella classificazione binaria.

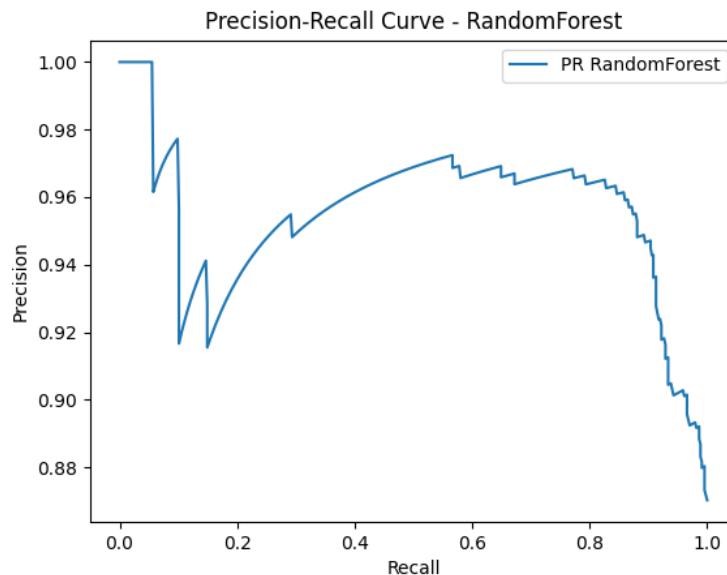


Figura 5.4: Curva Precision-Recall del modello Random Forest.

La Figura 5.4 riporta la curva Precision–Recall (PR), che descrive la relazione tra precision e recall al variare della soglia di classificazione. L'andamento della curva conferma la buona coerenza del modello, con valori di precision generalmente superiori a 0.9 per la maggior parte dell'intervallo di recall. Ciò indica che la Random Forest riesce a mantenere un'elevata affidabilità delle predizioni anche quando aumenta la sensibilità, ovvero quando individua un numero maggiore di campioni difettosi. La parte iniziale della curva, più irregolare, riflette la presenza di campioni a cui il modello assegna probabilità molto elevate di difettosità: in questa

regione, anche piccole variazioni di soglia possono influenzare in modo significativo il bilanciamento tra precision e recall.

L'area sottesa alla curva (PR-AUC), pari a circa 0.96, è coerente con il valore dell'AUC-ROC e conferma la robustezza complessiva del classificatore, capace di mantenere un buon equilibrio tra sensibilità e affidabilità anche in presenza di un dataset moderatamente sbilanciato. Nel contesto del Digital Twin per il controllo di qualità, questi risultati evidenziano la capacità del Random Forest di individuare efficacemente i difetti riducendo il rischio di falsi allarmi, contribuendo a un monitoraggio accurato e proattivo delle condizioni di processo.

5.2.2 Valutazione del modello Multilayer Perceptron

La Tabella 5.6 riporta la matrice di confusione ottenuta per il modello Multilayer Perceptron (MLP) applicato al set di test. Il modello ha classificato correttamente 399 componenti difettosi (True Positive) e 43 componenti non difettosi (True Negative). Tuttavia, sono stati commessi 37 falsi negativi, cioè componenti difettosi non riconosciuti come tali dal modello, e 31 falsi positivi, ovvero campioni non difettosi erroneamente classificati come difettosi.

Il modello MLP, pur mostrando prestazioni superiori rispetto a modelli più semplici come la regressione logistica, conferma il suo comportamento tipico, caratterizzato da una maggiore sensibilità rispetto alla specificità. Ciò comporta che, pur trovando correttamente i difetti, il modello tende a segnalare più difetti di quanti effettivamente ce ne siano, il che può risultare utile nei contesti in cui è preferibile non perdere difetti reali, ma presenta l'inevitabile effetto collaterale di un aumento dei falsi positivi.

Questo aspetto è rilevante nel contesto di applicazioni di controllo qualità, dove la gestione dei falsi positivi potrebbe implicare un maggiore carico operativo per la verifica manuale dei componenti non difettosi segnalati come difettosi. In ogni caso, il modello MLP riesce a garantire una buona capacità di discriminazione, pur con qualche limitazione nella riduzione dei falsi allarmi.

Tabella 5.6: Matrice di confusione del modello MLP sul set di test.

Predizione del modello		
Classe reale	Difettoso	Non difettoso
Difettoso	399	37
Non difettoso	31	43

Tabella 5.7: Prestazioni del modello MLP per ciascuna classe.

Classe	Precision	Recall	F1-score	N° campioni
Non difettoso	0.537	0.581	0.558	74
Difettoso	0.928	0.915	0.921	436

La Tabella 5.7 riporta le prestazioni del modello per ciascuna classe, evidenziando prestazioni differenti a seconda della distribuzione dei campioni. Per la classe difettosa, il modello ottiene una precision di 0.928, che indica che una gran parte delle osservazioni etichettate come difettose sono effettivamente tali. La sensibilità (recall) di 0.915 conferma un buon riconoscimento dei campioni difettosi, con pochi falsi negativi, mentre l'F1-score di 0.921 evidenzia un buon equilibrio tra precisione e recall.

Per la classe non difettosa, il modello mostra una precision di 0.537, che suggerisce che una percentuale significativa dei campioni non difettosi è stata erroneamente classificata come difettosa (falsi positivi). La sensibilità (recall) di 0.581 indica che il modello riesce a riconoscere la maggior parte dei campioni non difettosi, ma una parte consistente di essi viene comunque erroneamente classificata come difettosa. L'F1-score di 0.558 conferma la difficoltà del modello nel bilanciare correttamente le due classi.

Tabella 5.8: Metriche complessive del modello MLP.

Metrica	Precision	Recall	F1-score
Macro avg	0.732	0.748	0.739
Weighted avg	0.871	0.867	0.868
Accuracy = 0.867			

La Tabella 5.8 riassume le metriche complessive del modello, con valori di precision media ponderata (0.871) e recall medio ponderato (0.867) che indicano una buona stabilità complessiva del modello. Il F1-score medio ponderato di 0.868 suggerisce che l'MLP riesce ad affrontare in modo equilibrato entrambi i compiti, anche se la differenza tra le prestazioni sulle due classi è ancora evidente. La media macro, che considera ogni classe con uguale peso, mostra valori più bassi (precision 0.732, recall 0.748, F1-score 0.739), a causa della bassa performance sulla classe non difettosa, che ha un numero relativamente ridotto di campioni e risulta meno trattata dal modello.

L'accuratezza complessiva di 0.867 conferma che il modello MLP è abbastanza efficace nel riconoscere correttamente le osservazioni, ma il suo comportamento evidenzia un inclinamento verso la classe difettosa, con una maggiore sensibilità per la classe più numerosa e una precisione inferiore per la classe minoritaria. Tuttavia, l'alta sensibilità per la classe difettosa è positiva per applicazioni di diagnosi di difetti, dove è cruciale non trascurare nessun componente difettoso, pur dovendo bilanciare il rischio di falsi positivi.

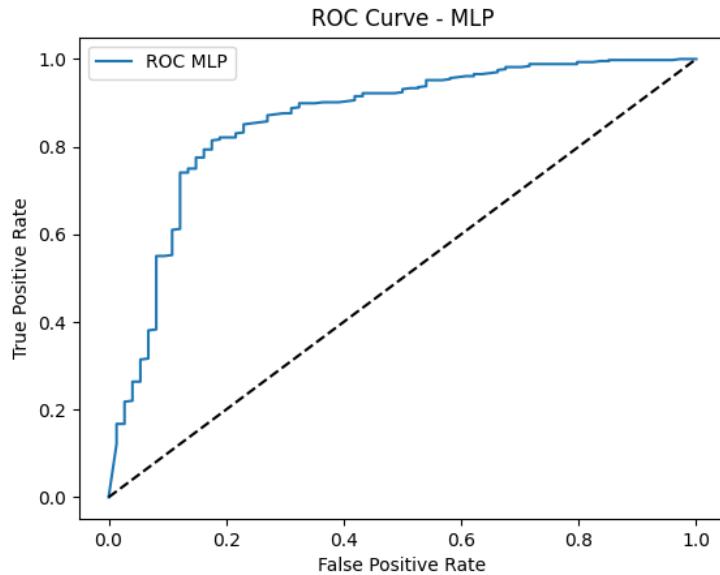


Figura 5.5: Curva ROC del modello MLP.

Come mostrato in Figura 5.5, la curva ROC del modello si posiziona nettamente sopra la linea tratteggiata, indicando una buona capacità di separazione tra le classi. L'area sotto la curva (AUC-ROC) risulta pari a 0.860, un valore che testimonia una buona capacità discriminante del modello, pur non essendo prossima alla perfezione (AUC = 1.0).

In particolare, la forma della curva mostra che il modello riesce a mantenere un elevato tasso di veri positivi (True Positive Rate) anche a fronte di un basso tasso di falsi positivi (False Positive Rate) nelle fasi iniziali, indicando che il classificatore è in grado di individuare efficacemente i campioni difettosi quando la soglia decisionale è alta. Tuttavia, man mano che la soglia viene abbassata, la curva tende ad appiattirsi, segno che l'aumento dei falsi positivi diventa più marcato e la capacità del modello di discriminare correttamente tra le due classi diminuisce.

Questo comportamento conferma la maggiore sensibilità verso la classe difettosa,

in linea con i valori di recall (0.915) e precision (0.928), ma una minore accuratezza nel riconoscimento dei campioni non difettosi (precision 0.537, recall 0.581). Tale asimmetria è attribuibile principalmente alla maggiore rappresentatività e coerenza della classe difettosa nel dataset, che ne facilita l'apprendimento, mentre la variabilità interna della classe conforme rende più complesso il suo corretto riconoscimento.

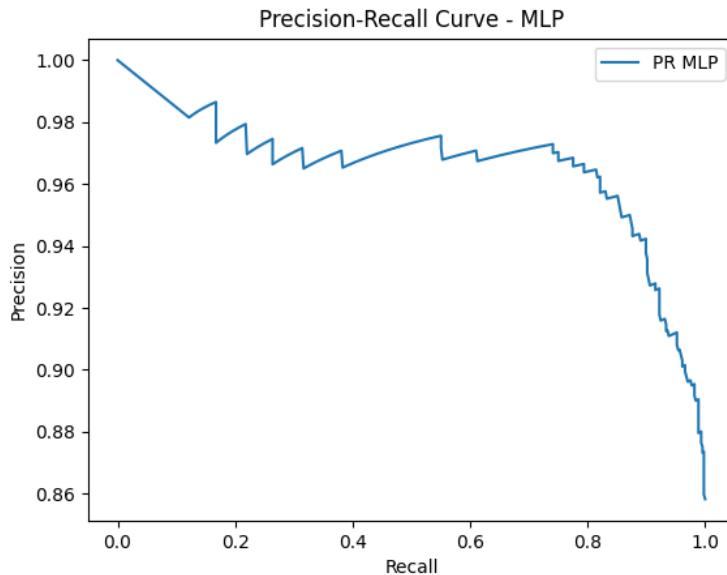


Figura 5.6: Curva Precision-Recall del modello MLP.

Come mostrato in Figura 5.6, la curva Precision-Recall (PR) del modello MLP evidenzia un comportamento complessivamente molto buono, con valori di precisione che si mantengono elevati per la maggior parte dell'intervallo di recall. L'area sotto la curva (AUC-PR) è pari a 0.964, indicando un'eccellente capacità del modello nel distinguere correttamente i campioni appartenenti alla classe positiva (difettosi), anche in presenza di un possibile sbilanciamento tra le classi.

All'inizio della curva, per valori di recall prossimi a zero, la precisione si attesta su valori prossimi a 1.0, a testimonianza del fatto che, quando il modello effettua previsioni molto selettive, quasi tutte le istanze identificate come positive risultano corrette. Man mano che il recall aumenta, la precisione tende a diminuire leggermente, pur mantenendosi stabilmente al di sopra di 0.95 fino a circa 0.8 di recall, e riducendosi solo nelle ultime porzioni della curva.

Questo andamento suggerisce che il modello è altamente affidabile nel riconoscere i campioni difettosi, e conferma quanto affermato in precedenza, ovvero che riesce a mantenere un ottimo equilibrio tra completezza (recall) e affidabilità delle previsioni

positive (precision). Il valore elevato dell’AUC-PR conferma che il classificatore è in grado di massimizzare la precision anche quando l’obiettivo è quello di identificare la maggior parte dei difetti presenti, limitando al minimo i falsi positivi.

Nel complesso, la curva PR e il valore $AUC-PR = 0.964$ indicano che il modello MLP presenta una forte capacità predittiva e discriminante, rendendolo particolarmente adatto per scenari in cui è prioritario individuare correttamente i campioni difettosi.

5.2.3 Valutazione del modello Multinomial Linear Regression

Tabella 5.9: Matrice di confusione del modello MLR sul set di test.

		Predizione del modello	
		Difettoso	Non difettoso
Classe reale	Difettoso	351	85
	Non difettoso	23	51

La Tabella 5.9 mostra la matrice di confusione relativa al modello Multinomial Logistic Regression (MLR), applicato al set di test per valutare la capacità di classificazione tra componenti difettosi e non difettosi. Il modello ha individuato correttamente 351 osservazioni difettose (True Positive) e 51 componenti non difettosi (True Negative). Tuttavia, ha commesso 85 falsi negativi, cioè casi di difettosità non rilevati, e 23 falsi positivi, ovvero componenti conformi erroneamente segnalati come difettosi.

Questo risultato evidenzia un comportamento tipico dei modelli di regressione logistica, che tendono a generare decisioni più conservative. La MLR mostra infatti una discreta precisione nella segnalazione dei pezzi difettosi, ma una sensibilità inferiore rispetto ai modelli non lineari, come Random Forest e LightGBM. In altre parole, il modello riesce a classificare in modo affidabile molti dei campioni positivi, ma tende a sottostimare la probabilità di difettosità nei casi più ambigui, classificandoli come “non difettosi” anche quando presentano deviazioni reali.

L’elevato numero di falsi negativi riflette proprio questa tendenza: pur garantendo una certa stabilità predittiva, la regressione logistica lineare fatica a catturare relazioni complesse e interazioni non lineari tra le variabili categoriche e numeriche. Questa limitazione deriva dal fatto che la MLR impone una relazione lineare tra le variabili e la log-odds della classe positiva, ipotesi che nel contesto di un processo

produttivo reale, caratterizzato da fenomeni multipli e non proporzionali, risulta spesso semplificativa.

D’altro canto, il numero relativamente basso di falsi positivi mostra che la MLR mantiene una buona specificità: il modello raramente classifica in modo errato i campioni conformi come difettosi. Ciò implica che, quando il modello prevede un difetto, tale segnalazione è nella maggior parte dei casi corretta. Questo comportamento la rende adatta a scenari in cui si preferisce minimizzare i falsi allarmi, ad esempio nei processi in cui le verifiche sui pezzi difettosi comportano costi elevati o tempi di fermo significativi. Tuttavia, in applicazioni orientate alla diagnostica precoce dei difetti, dove la priorità è non tralasciare alcun caso critico, una perdita di sensibilità rappresenta un limite rilevante.

Nel complesso, la MLR si conferma un modello interpretabile, stabile e statisticamente coerente, ma meno performante rispetto ai classificatori più avanzati in termini di capacità di generalizzazione. Le sue prestazioni evidenziano la necessità, per compiti di controllo qualità complessi come quelli gestiti dal Digital Twin, di ricorrere a modelli capaci di apprendere pattern non lineari e interazioni tra le variabili, come gli algoritmi basati su alberi decisionali o le reti neurali, che meglio rappresentano la natura multidimensionale dei fenomeni di difettosità.

Tabella 5.10: Prestazioni del modello MLR per ciascuna classe.

Classe	Precision	Recall	F1-score	N° campioni
Non difettoso	0.375	0.689	0.486	74
Difettoso	0.938	0.805	0.867	436

La Tabella 5.10 riporta le prestazioni della Multinomial Logistic Regression per ciascuna classe. Si osserva un comportamento fortemente asimmetrico tra le due classi, tipico dei modelli lineari in presenza di relazioni complesse e distribuzioni sbilanciate. Il modello ottiene un’elevata precisione per la classe difettosa (0.938), segno che la quasi totalità dei componenti segnalati come difettosi è effettivamente tale. Tuttavia, la sensibilità (recall = 0.805) mostra che una parte non trascurabile dei pezzi difettosi non viene riconosciuta, indicando che il modello tende ad essere più selettivo e prudente nella classificazione dei casi positivi.

Per la classe “non difettoso” si rileva invece un comportamento opposto: la precisione scende a 0.375, a indicare che molte delle osservazioni previste come “non difettose” appartengono in realtà alla classe difettosa, mentre il recall (0.689) suggerisce che la maggior parte dei campioni realmente conformi viene comunque riconosciuta correttamente. Questa configurazione riflette la tendenza del modello logistico a privilegiare la classe maggioritaria (difettoso), producendo un numero

relativamente elevato di falsi negativi ma mantenendo un'elevata affidabilità nelle predizioni positive.

Tabella 5.11: Metriche complessive del modello MLR.

Metrica	Precision	Recall	F1-score
Macro avg	0.657	0.747	0.676
Weighted avg	0.857	0.788	0.811
Accuracy = 0.788			

La Tabella 5.11 sintetizza le metriche complessive del modello. Le medie ponderate (weighted averages), calcolate tenendo conto della numerosità delle classi, mostrano valori coerenti con le prestazioni globali. Questi risultati indicano che, pur trattandosi di un modello lineare, la MLR mantiene una discreta capacità discriminante, con una buona coerenza tra le metriche. La macro average assume invece valori più bassi, evidenziando che le prestazioni non sono uniformi tra le due categorie.

Dal punto di vista interpretativo, la regressione logistica si conferma un modello robusto e facilmente interpretabile, capace di restituire probabilità calibrate e di fornire un riferimento teorico per il confronto con modelli più complessi. Tuttavia, i valori inferiori di recall e F1-score rispetto ad algoritmi come Random Forest e LightGBM suggeriscono che la MLR non cattura pienamente le relazioni non lineari presenti nei dati e che la sua efficacia risulta limitata quando il fenomeno da modellare è influenzato da interazioni tra variabili qualitative e quantitative.

In sintesi, la MLR rappresenta una buona base di riferimento per la classificazione binaria nel contesto del Digital Twin, ma mostra limiti strutturali nel riconoscimento completo dei difetti. La sua accuratezza e precisione risultano adeguate, ma la sensibilità ridotta comporta un rischio di sottostima dei casi difettosi, un aspetto critico nelle applicazioni industriali di controllo qualità, dove la priorità è evitare l'omissione di componenti non conformi.

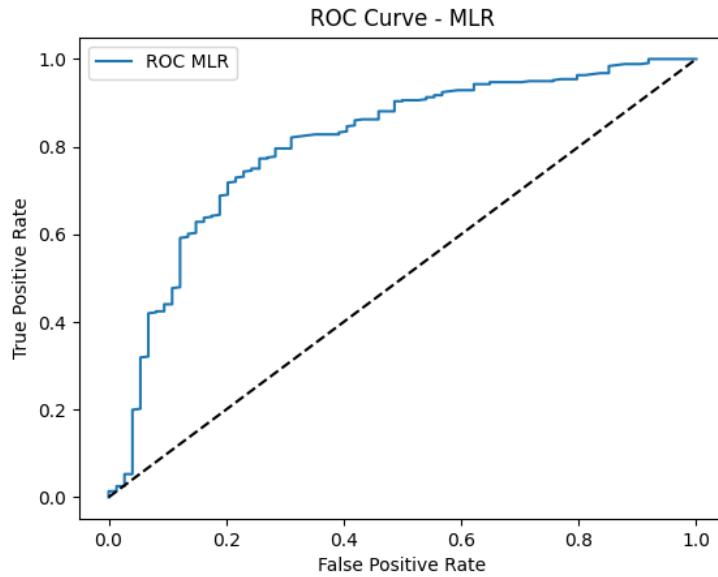


Figura 5.7: Curva ROC del modello MLR.

La Figura 5.7 mostra la curva ROC del modello MLR, la quale si colloca nettamente al di sopra di essa, indicando una buona capacità discriminante.

L'andamento della curva evidenzia che, per la maggior parte delle soglie, il modello mantiene un equilibrio accettabile tra sensibilità e specificità, riuscendo a individuare correttamente una quota significativa di campioni difettosi con un numero relativamente contenuto di falsi allarmi. Il tratto iniziale particolarmente ripido dimostra che già a bassi livelli di False Positive Rate il modello è in grado di raggiungere un'alta percentuale di True Positive Rate, segno di una risposta coerente e stabile nelle zone di classificazione più sicure.

L'area sottesa alla curva (AUC-ROC) è pari a circa 0.80, e questo valore, pur inferiore rispetto a quello ottenuto dai modelli più complessi, conferma che la regressione logistica possiede una discreta capacità di separare le due classi e che la sua perdita di prestazioni è imputabile principalmente alla natura lineare del modello. La curva mostra inoltre una parte finale più piatta, indice di una riduzione dell'efficacia discriminante nelle soglie più alte: in questi casi, per incrementare ulteriormente la sensibilità, il modello dovrebbe accettare un numero crescente di falsi positivi.

Nel complesso, la curva ROC conferma quanto già evidenziato dall'analisi delle metriche: la MLR costituisce un modello di riferimento affidabile e interpretabile per la classificazione binaria, ma manifesta limiti strutturali nel riconoscimento completo dei difetti. La pendenza meno marcata della curva rispetto a quella dei modelli non lineari indica una minor capacità di separazione tra le due classi,

segno che il modello tende a semplificare le frontiere di decisione e a non cogliere pienamente le interazioni complesse tra le variabili. Questa caratteristica si traduce in una sensibilità inferiore, coerente con il rischio di sottostima dei casi difettosi già osservato nelle metriche aggregate.

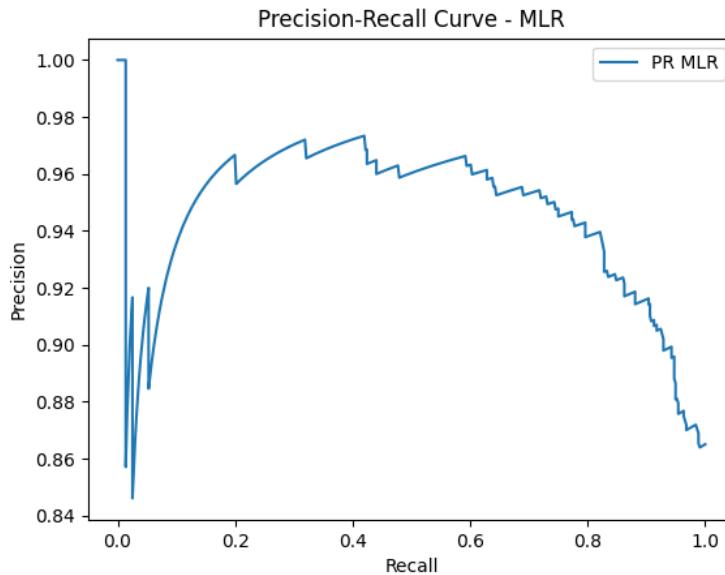


Figura 5.8: Curva Precision-Recall del modello MLR.

La curva Precision-Recall in Figura 5.8 evidenzia una buona stabilità complessiva, con valori di precision costantemente elevati, generalmente superiori a 0.9 per gran parte dell'intervallo di recall. Ciò indica che, anche quando il modello aumenta la propria sensibilità (cioè classifica un numero maggiore di campioni come difettosi), mantiene comunque un'elevata affidabilità nelle predizioni positive. La parte iniziale più irregolare riflette i casi in cui la regressione logistica assegna probabilità estreme (molto vicine a 0 o 1), situazione in cui anche lievi variazioni della soglia possono produrre oscillazioni evidenti nel rapporto tra precision e recall.

L'area sottesa alla curva (PR-AUC), pari a circa 0.95, conferma le buone prestazioni generali del modello e la sua capacità di discriminare in modo coerente la classe difettosa. Tuttavia, la forma complessiva della curva, più piatta e meno espansa rispetto a quella ottenuta dai modelli non lineari, evidenzia una maggiore perdita di precisione nelle regioni ad alto recall. Ciò significa che, quando la soglia viene abbassata per aumentare la sensibilità, la MLR tende a generare un numero crescente di falsi positivi, riducendo la purezza delle predizioni positive.

La regressione logistica, pur garantendo una buona calibrazione probabilistica e prestazioni solide come modello di riferimento, rimane meno flessibile rispetto

a classificatori più complessi. Questa rigidità si traduce in un compromesso tra precision e recall più marcato, in linea con la minore sensibilità già osservata nelle metriche aggregate e con la difficoltà del modello nel cogliere le relazioni non lineari che descrivono i difetti nei processi produttivi reali.

5.2.4 Valutazione del modello LightGBM

Come riportato in Tabella 5.12, la matrice di confusione del modello LightGBM sul set di test mostra una buona capacità di classificazione complessiva. Il modello riesce a identificare correttamente la maggior parte dei campioni difettosi, con 400 istanze classificate correttamente come difettose, mentre 36 campioni difettosi vengono erroneamente riconosciuti come non difettosi. Allo stesso tempo, il modello classifica correttamente 44 campioni non difettosi, commettendo 30 errori di falso allarme, ossia casi in cui un campione non difettoso viene predetto come difettoso.

L'analisi della matrice evidenzia quindi che il modello tende a privilegiare la corretta individuazione dei difetti, mantenendo comunque un controllo accettabile sugli errori di classificazione dei campioni non difettosi.

Nel complesso, la distribuzione dei valori nella matrice suggerisce che il modello LightGBM abbia appreso in modo efficace le caratteristiche distintive delle due classi, mostrando una buona capacità di generalizzazione sul set di test.

Tabella 5.12: Matrice di confusione del modello LightGBM sul set di test.

		Predizione del modello	
		Difettoso	Non difettoso
Classe reale	Difettoso	400	36
	Non difettoso	30	44

Tabella 5.13: Prestazioni del modello LightGBM per ciascuna classe.

Classe	Precision	Recall	F1-score	N° campioni
Non difettoso	0.550	0.594	0.571	74
Difettoso	0.930	0.917	0.923	436

La Tabella 5.13 riporta le prestazioni del modello LightGBM per ciascuna classe. Si osserva che per la classe difettosa, il modello ottiene una precision elevata (0.930), indicando che la maggior parte dei campioni segnalati come difettosi è

effettivamente tale. Il recall di 0.917 conferma una buona capacità del modello di identificare correttamente la classe difettosa, con una bassa incidenza di falsi negativi. L'F1-score per la classe difettosa è molto alto (0.923), indicando un buon equilibrio tra precision e recall.

Per la classe non difettosa, la precision scende a 0.550, il che significa che quasi la metà dei campioni classificati come non difettosi sono in realtà difettosi (False Positive). Tuttavia, la sensibilità (recall) di 0.594 mostra che il modello riesce comunque a identificare correttamente una buona parte dei campioni non difettosi. L'F1-score per la classe non difettosa è 0.571, indicando un equilibrio meno ottimale tra precision e recall rispetto alla classe difettosa.

Tabella 5.14: Metriche complessive del modello LightGBM.

Metrica	Precision	Recall	F1-score
Macro avg	0.740	0.756	0.748
Weighted avg	0.875	0.870	0.873
Accuracy = 0.870			

La Tabella 5.14 riassume le metriche complessive del modello. Le medie ponderate (weighted average) mostrano valori di precision di 0.875, recall di 0.870 e F1-score di 0.873, con un'accuratezza complessiva pari a 0.870. Questi risultati indicano una buona performance globale del modello, sebbene sia presente un margine di miglioramento nella gestione della classe non difettosa, che presenta una precisione inferiore rispetto alla classe difettosa. La media macro, che non tiene conto del numero di campioni per ciascuna classe, mostra valori più bassi (precisione 0.740, recall 0.756, F1-score 0.748), evidenziando che il modello ha difficoltà a trattare equamente le due classi, con una certa penalizzazione per la classe minoritaria (non difettoso).

In sintesi, il modello mostra comunque una buona capacità di discriminare tra le classi e di mantenere elevata la sensibilità verso i difetti, riuscendo a preservare un equilibrio complessivo tra precision e recall migliore rispetto ad altri modelli come MLP o MLR, offrendo un miglior compromesso tra precisione e recall, in particolare nella rilevazione dei difetti.

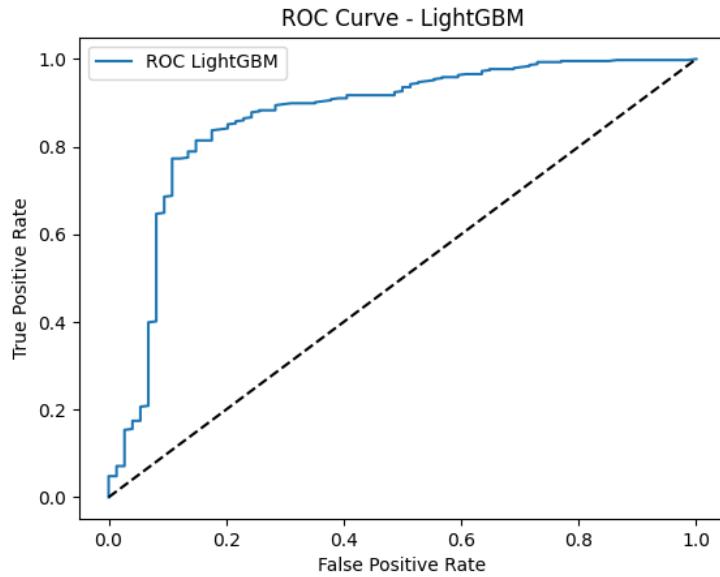


Figura 5.9: Curva ROC del modello LightGBM.

Il grafico in Figura 5.9 mostra la curva ROC ottenuta per il modello LightGBM, la cui area sotto la curva (AUC) risulta pari a 0.867. Essendo che un'AUC pari a 0.5 indica un comportamento casuale, mentre un valore prossimo a 1.0 corrisponde a una classificazione perfetta, l'AUC pari a 0.867 testimonia un'elevata capacità discriminante del modello, che riesce a separare in modo efficace le due categorie riducendo la probabilità di errore sia in termini di falsi positivi sia di falsi negativi.

Nel contesto del Digital Twin sviluppato per il controllo di qualità, il risultato ottenuto conferma la solidità del modello LightGBM e la sua capacità di integrare in modo coerente le informazioni provenienti dalle variabili di processo e dallo stato dell'utensile, pur evidenziando una minore precisione nella classificazione della classe “non difettoso”. L'elevato valore dell'AUC, unito all'ottimo equilibrio tra precision e recall per la classe difettosa, dimostra che il modello è in grado di discriminare efficacemente la presenza di difetti e di fornire previsioni affidabili, rendendolo uno strumento solido per la diagnostica preventiva e il supporto decisionale.

Rispetto ai modelli precedentemente analizzati, il LightGBM mostra un comportamento più bilanciato e una migliore capacità di generalizzazione, grazie all'architettura basata su un approccio di boosting sequenziale, in cui più alberi decisionali vengono combinati per correggere progressivamente gli errori dei modelli precedenti.

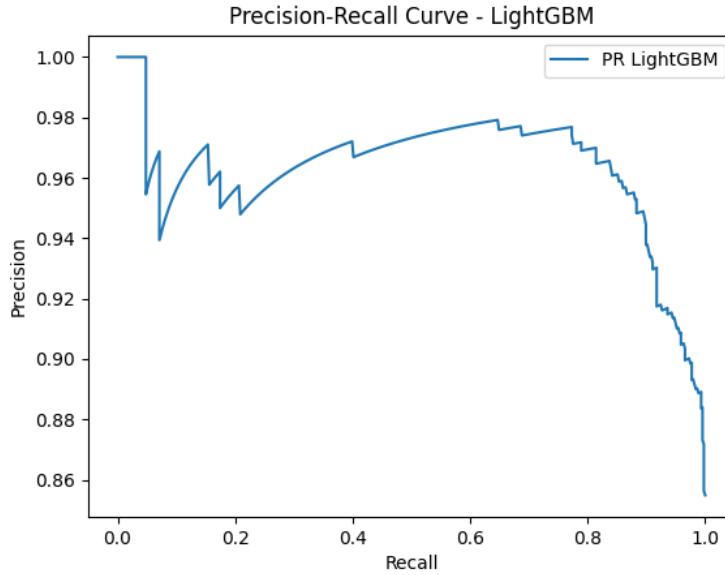


Figura 5.10: Curva Precision-Recall del modello LightGBM.

La Figura 5.10 mostra la curva Precision-Recall ottenuta per il modello LightGBM, la cui area sotto la curva (AUC-PR) è pari a 0.963. Tale valore rappresenta un indicatore particolarmente significativo della capacità del modello di mantenere un'elevata precisione anche per valori di recall elevati, evidenziando un ottimo equilibrio tra la quota di difetti correttamente individuati e la percentuale di predizioni accurate. Il risultato conferma che il LightGBM presenta un comportamento estremamente robusto, riuscendo a preservare livelli di precisione superiori al 95% anche nelle condizioni di massima sensibilità. Ciò dimostra che il classificatore è in grado di individuare la quasi totalità dei difetti reali senza compromettere l'affidabilità complessiva delle previsioni, un aspetto cruciale nel contesto del controllo qualità basato su Digital Twin, dove è essenziale minimizzare gli errori di classificazione sia positivi sia negativi. L'elevata AUC-PR, in accordo con i risultati della curva ROC ($AUC = 0.867$), rafforza ulteriormente l'evidenza dell'ottima capacità predittiva del LightGBM, che si distingue per sensibilità, stabilità e coerenza nelle prestazioni. Nel complesso, tali caratteristiche confermano che il LightGBM rappresenta il classificatore con la miglior capacità discriminante e il compromesso più equilibrato tra sensibilità e specificità, risultando particolarmente adatto all'integrazione nel framework di Digital Twin per la diagnostica dei difetti e per l'implementazione di un controllo qualità proattivo orientato al principio del right first time, secondo cui ogni fase del processo deve garantire la conformità del prodotto fin dal primo ciclo produttivo, riducendo la necessità di rilavorazioni e migliorando l'efficienza complessiva del sistema.

5.3 Analisi e discussione dei risultati

L’analisi complessiva dei risultati evidenzia come l’approccio proposto, basato sull’integrazione di modelli di classificazione supervisionata all’interno di un modello di Digital Twin per il controllo qualità, abbia dimostrato un’elevata efficacia nel riconoscimento dei difetti. Sono stati confrontati quattro modelli di classificazione super visionata (Random Forest, MLP, MLR e LightGBM), con l’obiettivo di valutare le loro capacità predittive nell’ambito del controllo qualità binario, ovvero nella distinzione tra prodotti difettosi e non difettosi. Per gestire lo sbilanciamento intrinseco del dataset, in cui la classe “difettoso” risulta fortemente prevalente, sono state testate due diverse tecniche di oversampling: ROS (Random OverSampling) e SMOTENC (Synthetic Minority Oversampling Technique for Nominal and Continuous features). L’analisi comparativa ha mostrato che, sebbene entrambe le tecniche contribuiscano ad attenuare il problema dello sbilanciamento, SMOTENC ha prodotto risultati mediamente superiori su tutte le principali metriche di valutazione. Tale superiorità è dovuta alla maggiore capacità di SMOTENC di generare istanze sintetiche coerenti con la struttura del dato, preservando le relazioni tra variabili continue e categoriche. Di conseguenza, la scelta finale è ricaduta su SMOTENC come metodo di riferimento per l’addestramento e la validazione dei modelli.

Nel complesso, i modelli si sono dimostrati in grado di trasformare efficacemente le variabili di processo e di progetto in una stima binaria di conformità del componente, traducendo così le misure acquisite in un output diagnostico interpretabile e direttamente utilizzabile nel contesto operativo del Digital Twin. Le analisi delle curve ROC e Precision-Recall hanno permesso di approfondire la capacità dei modelli di distinguere tra le due classi anche al variare delle soglie di classificazione, fornendo una valutazione più robusta rispetto alla sola accuratezza complessiva.

5.3.1 Performance dei modelli

Dal confronto tra i modelli analizzati, il LightGBM emerge come il classificatore con le prestazioni complessivamente più elevate e il miglior equilibrio tra sensibilità e specificità. Con valori di AUC-ROC pari a 0.867 e di AUC-PR pari a 0.963, il modello mostra un’eccellente capacità discriminante, dimostrando di saper distinguere con grande precisione tra componenti difettosi e non difettosi. La precision per la classe difettosa (0.930) e il recall (0.917) confermano che la quasi totalità dei componenti segnalati come difettosi lo sono effettivamente, mentre solo una minima parte dei difetti sfugge alla rilevazione. L’accuratezza complessiva (0.870) e l’F1-score ponderato (0.826) evidenziano inoltre una buona capacità di generalizzazione, mantenuta anche su dati non osservati.

Tali risultati derivano dalla natura del modello, basato su un algoritmo di boosting che costruisce sequenze di alberi decisionali capaci di ottimizzare iterativamente i confini di decisione, garantendo al tempo stesso rapidità di inferenza e stabilità predittiva.

Il Multilayer Perceptron (MLP) presenta risultati molto vicini a quelli del LightGBM, con un'AUC-PR di 0.964 e un'accuratezza di 0.867. Sebbene tenda a generare un numero leggermente superiore di falsi positivi, il MLP garantisce un'elevata copertura dei difetti e un comportamento stabile, rendendolo particolarmente adatto in contesti in cui la priorità è non tralasciare alcuna anomalia. La sua architettura non lineare consente di catturare relazioni complesse tra le variabili, ma richiede tempi di addestramento più elevati e una maggiore sensibilità ai parametri di tuning.

La Random Forest mantiene prestazioni solide e regolari, con un'accuratezza di 0.843 e un'AUC-ROC di 0.846. Il modello mostra una buona capacità di riconoscere la classe difettosa, pur manifestando una leggera perdita di sensibilità rispetto al LightGBM. La robustezza dell'algoritmo, unita alla facilità di interpretazione dell'importanza delle variabili, ne fa comunque un riferimento affidabile per applicazioni industriali.

La Multinomial Logistic Regression (MLR), pur garantendo un comportamento stabile e facilmente interpretabile, si colloca su livelli inferiori di performance, con un recall ridotto e un'elevata incidenza di falsi negativi. L'approccio lineare limita infatti la capacità del modello di cogliere le interazioni non lineari tra le variabili di processo e di geometria.

Nel complesso, il confronto conferma che i modelli basati su alberi decisionali e metodi di boosting, in particolare il LightGBM, offrono le prestazioni più elevate, grazie alla loro capacità di combinare robustezza, adattabilità e generalizzazione. L'introduzione di SMOTENC si è rivelata decisiva nel migliorare il bilanciamento tra le classi e nel ridurre il rischio di sottostima dei casi difettosi, assicurando al Digital Twin una capacità diagnostica accurata e affidabile.

Tabella 5.15: Metriche di performance dei modelli riportate nello studio di Boettjer e Ramanujan (2023). * indica che la metrica è ponderata.

Modello	AUC PR	AUC ROC	F1*	P*	R*
RFC	0.662	0.923	0.890	0.896	0.885
MLP	0.805	0.894	0.868	0.900	0.885
MLR	0.600	0.788	0.794	0.791	0.813

Il confronto con i risultati riportati da Böttjer e Ramanujan nell'articolo e riassunti nella Tabella 5.15, mostra che le prestazioni ottenute nel presente studio sono

in linea con quelle del lavoro originale, con alcuni miglioramenti specifici legati all'impiego di tecniche di bilanciamento e a una metodologia di validazione più rigorosa. Nel loro studio, gli autori avevano sviluppato un modello di Digital Twin per il controllo della qualità, testando tre modelli di classificazione — Random Forest (RFC), Multilayer Perceptron (MLP) e Multinomial Logistic Regression (MLR) — applicati alla previsione della conformità dei singoli elementi lavorati. I risultati mostravano già buone capacità di classificazione, con valori di AUC-ROC pari a 0.923 per la Random Forest e 0.894 per il MLP, e valori di AUC-PR compresi tra 0.600 e 0.805.

Nel presente lavoro, le prestazioni ottenute risultano comparabili o leggermente superiori, soprattutto in termini di AUC-PR, che raggiunge valori fino a 0.963 per il modello LightGBM e 0.964 per il MLP. Tali risultati testimoniano una maggiore sensibilità verso la classe difettosa, a fronte di valori di AUC-ROC dello stesso ordine di grandezza (tra 0.846 e 0.867). E' importante sottolineare come il principale avanzamento di questo lavoro risiede nella maggiore solidità metodologica e nella coerenza con le condizioni operative reali, che costituiscono la base per un Digital Twin più affidabile e trasferibile.

5.3.2 Confronto con altri approcci nel controllo qualità

I risultati ottenuti nel presente studio si collocano in continuità con le più recenti evoluzioni nel controllo qualità industriale, ma introducono elementi di sostanziale innovazione rispetto agli approcci tradizionali e ai contributi già presenti in letteratura. Storicamente, i sistemi di controllo qualità si sono basati su metodi statistici come il controllo statistico di processo (SPC), le carte di controllo o i modelli di regressione lineare. Questi strumenti, pur efficaci in scenari con dati continui e ben strutturati, risultano meno adatti a gestire variabili categoriche, interazioni non lineari e fenomeni di dipendenza tipici dei processi manifatturieri complessi.

L'introduzione di tecniche di machine learning ha ampliato significativamente la capacità di modellare relazioni eterogenee e di integrare dati provenienti da fonti diverse. Studi come quello di Böttjer e Ramanujan hanno mostrato l'efficacia di modelli come Random Forest e MLP nel prevedere la qualità delle lavorazioni, raggiungendo risultati di rilievo in termini di AUC-ROC (fino a 0.923) e valori di AUC-PR pari a 0.805. Questi risultati testimoniano una buona capacità discriminante, ma evidenziano anche alcune criticità legate allo sbilanciamento del dataset e alla validazione basata su split 90/10 applicato a livello di singola feature, che può introdurre fenomeni di contaminazione tra training e test set e sovrastimare la capacità di generalizzazione dei modelli.

In questo lavoro vengono affrontate tali limitazioni proponendo un approccio più robusto e realistico. In primo luogo, l'adozione della tecnica di bilanciamento SMOTENC ha migliorato la rappresentatività della classe difettosa, innalzando

sensibilmente l'AUC-PR e riducendo il rischio di sottostima dei difetti. Parallelamente, la definizione di una suddivisione 70/15/15 per gruppi di TP ha permesso di garantire l'indipendenza tra le fasi di addestramento e test, rispecchiando meglio la variabilità dei lotti produttivi reali. Un'ulteriore innovazione è rappresentata dall'introduzione del Light Gradient Boosting Machine (LightGBM), che ha fornito prestazioni superiori rispetto ai modelli originari grazie alla sua capacità di combinare rapidità di addestramento, riduzione della varianza e accuratezza predittiva. Il modello ha mostrato un comportamento più bilanciato tra precision e recall, risultando particolarmente adatto per applicazioni industriali in cui è fondamentale minimizzare sia i falsi positivi sia i falsi negativi.

In sintesi, il confronto con gli approcci esistenti dimostra che l'architettura proposta, basata su modelli di boosting e su una gestione strutturata dei dati, garantisce maggiore robustezza, generalizzazione e coerenza con gli scenari produttivi reali. L'integrazione di variabili categoriche, l'uso di tecniche di bilanciamento avanzate e l'approccio per gruppi di lavorazione rafforzano la capacità del Digital Twin di operare come sistema predittivo e adattivo per il controllo della qualità, superando i limiti dei metodi tradizionali e delle precedenti implementazioni presenti in letteratura.

Capitolo 6 – Conclusioni e implicazioni

6.1 Conclusioni

Il lavoro di ricerca presentato in questa tesi ha avuto come obiettivo lo sviluppo di un Digital Twin per il controllo della qualità basato su variabili categoriche, con applicazione sperimentale alla diagnostica dei difetti nella produzione di stampi ad alta precisione. L'indagine ha permesso di esplorare in modo sistematico la possibilità di integrare, all'interno di un gemello digitale, informazioni eterogenee di natura qualitativa e quantitativa, superando la tradizionale dipendenza dai soli parametri numerici di processo.

La progettazione del modello si basa sul caso di studio sperimentale descritto da Böttjer e Ramanujan nell'articolo “A digital twin assisted framework for quality assurance in mould manufacturing” (2023), in cui gli autori avevano realizzato un modello di Digital Twin per la valutazione della qualità di 16 campioni di inserti metallici sottoposti a lavorazioni CNC. Gli autori avevano evidenziato una delle principali criticità del dataset sperimentale, ovvero lo sbilanciamento tra le classi dei difettosi e non difettosi, in cui la prima risultava essere quella prevalente. Questa asimmetria portava i modelli a favorire la classe maggioritaria, compromettendo la sensibilità complessiva del sistema diagnostico.

Per affrontare tale limitazione, in questo studio sono state testate due tecniche di bilanciamento dei dati: il Random Oversampling (ROS), basato sulla semplice duplicazione delle osservazioni minoritarie, e lo SMOTENC, che genera campioni sintetici preservando la natura mista (numerica e categorica) delle variabili. Per migliorare ulteriormente il lavoro originale, è stata introdotta una suddivisione del dataset per gruppi di provini (TP). Nel dataset sperimentale, infatti, ogni provino è associato a molteplici misurazioni, ciascuna riferita a specifiche variabili geometriche rilevate sullo stesso componente. Se tali misurazioni venissero distribuite casualmente tra i set di addestramento, validazione e test, lo stesso provino potrebbe comparire in più sottoinsiemi, generando una dipendenza statistica tra i dati e fornendo al modello informazioni indirette sul test durante la fase di addestramento.

Per evitare questo effetto e simulare in modo più realistico il comportamento di un sistema produttivo reale, è stato scelto di mantenere tutte le osservazioni relative a uno stesso provino all'interno dello stesso gruppo. In questo modo si garantisce l'indipendenza effettiva tra i tre insiemi di dati e si ottiene una valutazione più solida e rappresentativa delle prestazioni del modello in scenari applicativi reali.

In parallelo, è stato introdotto il modello Light Gradient Boosting Machine (LightGBM), selezionato per la sua efficienza computazionale e per la capacità di modellare relazioni non lineari complesse mantenendo al contempo un'elevata stabilità predittiva. Questo modello si aggiunge ai tre algoritmi originariamente impiegati da Böttjer e Ramanujan — Random Forest, Multilayer Perceptron e Multinomial Logistic Regression — consentendo un confronto diretto tra tecniche tradizionali e un approccio di boosting più recente e avanzato. Il modello proposto si fonda su un flusso di elaborazione automatizzato, realizzato in ambiente Python, che gestisce tutte le fasi del processo dati, dall'acquisizione e pre-elaborazione delle misure di lavorazione fino alla classificazione binaria (difettoso / non difettoso) per ciascun elemento geometrico analizzato, e che ha previsto il confronto tra i quattro modelli di classificazione sopra citati, al fine di valutarne la capacità di individuare correttamente i componenti difettosi. Le metriche di valutazione — Accuracy, F1-score, Precision, Recall, AUC-ROC e AUC-PR — hanno mostrato che l'approccio basato su SMOTENC garantisce una rappresentazione più equilibrata delle classi e una maggiore sensibilità complessiva del sistema, mentre il LightGBM ha evidenziato le prestazioni più bilanciate e la miglior capacità discriminante tra componenti conformi e non conformi, rivelandosi il modello più adatto all'integrazione in un'architettura di Digital Twin orientata al principio del right first time, secondo cui ogni fase del processo deve garantire la conformità del prodotto fin dal primo ciclo produttivo.

Nel complesso, il lavoro ha dimostrato la fattibilità e l'efficacia di un Digital Twin in grado di elaborare simultaneamente variabili numeriche e categoriche, fornendo un output diagnostico interpretabile, aggiornabile e utile per decisioni in tempo reale. Il principale avanzamento rispetto allo studio di riferimento risiede nella solidità metodologica e nella generalizzabilità dell'approccio, che riflettono un trattamento più realistico dei dati e una maggiore aderenza alle condizioni operative industriali.

6.2 Implicazioni pratiche, limiti del lavoro e sviluppi futuri

Dal punto di vista operativo, i risultati dimostrano come un Digital Twin basato su variabili categoriche possa costituire uno strumento concreto per il controllo qualità nei processi di produzione ad alta precisione. L'integrazione di dati eterogenei

consente infatti di anticipare la diagnosi dei difetti e di trasformare il gemello digitale da sistema puramente descrittivo a strumento predittivo e prescrittivo, in grado di supportare le decisioni operative in tempo quasi reale. L'output localizzato del modello offre inoltre un contributo rilevante per la prioritizzazione delle attività di ispezione e manutenzione, permettendo di concentrare le risorse sulle zone più critiche e migliorare l'efficienza complessiva del processo produttivo.

L'applicazione proposta apre anche a prospettive di scalabilità e integrazione con i sistemi informativi aziendali, in particolare con le piattaforme MES (Manufacturing Execution System) e PLM (Product Lifecycle Management). In questa direzione, il Digital Twin può evolvere verso una forma cognitiva e adattiva, capace di apprendere in modo continuo dai dati di produzione e fornire suggerimenti correttivi automatizzati.

Nonostante i risultati promettenti, il lavoro presenta alcuni limiti che delineano le direzioni per futuri sviluppi. Il campione sperimentale considerato è limitato a una specifica tipologia di componente e a condizioni di lavorazione controllate, e sarà quindi necessario estendere la validazione a contesti industriali differenti per verificarne la generalizzabilità. Inoltre, le variabili categoriche utilizzate, sebbene rappresentative, non coprono l'intera gamma di fattori qualitativi che possono influenzare la conformità di uno stampo, come le condizioni ambientali, l'usura degli utensili o le variazioni microstrutturali del materiale. Un ulteriore limite riguarda la natura statica dell'implementazione, che al momento opera in modalità offline, senza aggiornamento dinamico dei parametri o retroazione diretta sul processo produttivo. A ciò si aggiunge la complessità computazionale di alcuni algoritmi, in particolare del LightGBM e dello SMOTENC, che può rappresentare un ostacolo all'utilizzo in tempo reale su sistemi industriali a risorse limitate.

Tuttavia, l'adozione di architetture distribuite o di soluzioni di edge computing potrà in futuro garantire una maggiore efficienza di calcolo e una gestione dinamica dei modelli, rendendo possibile l'implementazione del gemello digitale in ambienti produttivi ad alte prestazioni.

L'articolo conferma la maturità tecnica del paradigma di Digital Twin per la qualità e sottolinea le possibilità di sviluppo verso sistemi più scalabili, adattivi e integrati che possono apprendere in modo continuo e restituire valore decisionale nel ciclo produttivo reale.

Le prospettive di sviluppo di questo lavoro si collocano lungo due strade principali: l'estensione del modello a scenari industriali reali e l'evoluzione verso Digital Twin cognitivi e auto-adattivi. Sul piano tecnico, la futura ricerca dovrà concentrarsi sull'integrazione del DT con i sistemi di produzione in tempo reale, attraverso architetture edge o cloud-based in grado di elaborare i dati direttamente dalle macchine utensili. Questo consentirebbe non solo l'aggiornamento continuo dei modelli di classificazione, ma anche la chiusura del ciclo di controllo, trasformando il DT in un sistema cyber-fisico attivo capace di intervenire automaticamente sui

parametri di processo.

Un altro ambito di evoluzione riguarda l'impiego di tecniche di apprendimento incrementale, che permetterebbero al modello di adattarsi progressivamente a nuove condizioni operative, materiali o configurazioni di macchina senza necessità di riaddestramento completo. Parallelamente, l'integrazione con sensori avanzati e dati non strutturati, come immagini, vibrazioni o segnali acustici, potrebbe ampliare la capacità diagnostica del DT, rendendolo più sensibile a difetti di natura geometrica, termica o microstrutturale.

Dal punto di vista metodologico, la ricerca potrà esplorare l'uso di modelli interpretabili per migliorare la trasparenza delle decisioni e favorire l'accettazione industriale della tecnologia, nonché la combinazione di modelli fisici e data-driven in un approccio ibrido capace di coniugare accuratezza predittiva e interpretabilità ingegneristica.

In questo senso, il lavoro presentato rappresenta un passo preliminare verso DT intelligenti, autonomi e sostenibili, destinati a diventare elementi centrali della fabbrica cognitiva del futuro.

Glossario

Classificazione multimodale Tecnica di apprendimento automatico che integra e analizza dati provenienti da fonti eterogenee (come segnali visivi, acustici, termici o testuali) per assegnare un'etichetta ad un'osservazione, migliorando accuratezza e robustezza della classificazione

Clustering Tecnica di apprendimento non supervisionato che consente di raggruppare automaticamente un insieme di dati in sottoinsiemi (cluster) omogenei, in base a somiglianze o pattern condivisi

Connessione OPC-UA Protocollo standard per la comunicazione sicura e interoperabile tra dispositivi industriali e sistemi informatici.

Decision making Processo di selezione della migliore azione o opzione possibile tra alternative disponibili, basato su criteri, vincoli e informazioni disponibili

Deep learning Sottoinsieme del machine learning basato su reti neurali artificiali con molti strati (deep neural networks), utilizzato per apprendere rappresentazioni complesse dei dati in modo automatico

Digital Twin Rappresentazione virtuale aggiornata dinamicamente di un asset fisico

Edge computing Architettura di elaborazione distribuita che consente di processare i dati vicino al luogo in cui vengono generati (cioè "ai margini" della rete), riducendo la latenza e la dipendenza dal cloud

Elementi finiti Tecnica numerica utilizzata per risolvere problemi complessi di ingegneria strutturale, termica o fluidodinamica, attraverso la discretizzazione del dominio fisico in una mesh di elementi finiti su cui vengono applicate le equazioni differenziali del sistema

Few-shot learning Tecnica di apprendimento automatico che consente a un modello di generalizzare e apprendere concetti da un numero molto ridotto di esempi, spesso utilizzando conoscenza pregressa o strategie di meta-apprendimento

Flower Pollination Algorithm Algoritmo di ottimizzazione ispirato al processo naturale di impollinazione dei fiori, che combina meccanismi di esplorazione globale e sfruttamento locale per risolvere problemi complessi in ambito ingegneristico e computazionale

Inferenza causale Processo statistico e computazionale volto a determinare relazioni di causa-effetto tra variabili, andando oltre la semplice correlazione e cercando di identificare l'impatto di un evento o intervento su un determinato esito

Internet of Things Rete di oggetti fisici dotati di sensori, software e connettività, in grado di raccogliere e scambiare dati con altri dispositivi o sistemi attraverso Internet, con l'obiettivo di monitorare, controllare e ottimizzare processi in tempo reale

Label dominante Etichetta più rappresentativa o frequente assegnata a un'osservazione o a un insieme di dati, utilizzata per semplificare l'interpretazione o la valutazione di modelli di classificazione

LSTM Architettura di rete neurale ricorrente progettata per apprendere e memorizzare dipendenze a lungo termine nelle sequenze temporali, ampiamente utilizzata in ambiti come il riconoscimento di serie temporali, la diagnostica predittiva e l'elaborazione del linguaggio naturale

Machine learning Tecnica dell'intelligenza artificiale che consente ai sistemi di apprendere automaticamente dai dati e migliorare le proprie prestazioni senza essere esplicitamente programmati

Manutenzione predittiva Strategia di manutenzione basata sull'analisi dei dati di funzionamento del sistema, finalizzata a prevedere e prevenire guasti prima che si verifichino

Modello bayesiano Modello statistico basato sul teorema di Bayes, che aggiorna la probabilità di un'ipotesi a partire da nuove evidenze, ed è utilizzato per l'inferenza, la classificazione e la diagnostica in condizioni di incertezza

Modello CAD Rappresentazione digitale tridimensionale di oggetti o sistemi, creata mediante software di progettazione assistita da calcolatore

Modello CAE Modello digitale per l'analisi e la simulazione del comportamento fisico di un sistema, utilizzato nell'ingegneria assistita da calcolatore

Modello fuzzy Modello basato sulla logica fuzzy, che consente di rappresentare conoscenza incerta o approssimata attraverso valori di appartenenza graduati, invece che binari, ed è utilizzato per il supporto decisionale in sistemi complessi

Modello gaussiano Modello statistico che assume che i dati siano distribuiti secondo una distribuzione normale (gaussiana), utilizzato per rappresentare fenomeni naturali, stimare probabilità e identificare anomalie nei dati

Modello LightGBM Algoritmo di machine learning basato su gradient boosting, ottimizzato per efficienza e velocità su grandi dataset, spesso utilizzato per classificazione, regressione e ranking

Modello MME-BMR Architettura di classificazione multimodale che combina l'estrazione di caratteristiche multiple (MME) con una struttura di classificazione a doppia ramificazione (BMR), utilizzata per migliorare l'accuratezza nella diagnosi di difetti

Random Forest Algoritmo di apprendimento automatico basato su un insieme di alberi decisionali addestrati su campioni casuali dei dati, che opera tramite votazione aggregata per migliorare accuratezza e robustezza della classificazione o regressione

Realtà aumentata Tecnologia che sovrappone contenuti digitali (come immagini, dati o modelli 3D) al mondo reale in tempo reale, migliorando la percezione e l'interazione con l'ambiente fisico

Rete neurale Modello computazionale ispirato al funzionamento del cervello umano, costituito da nodi (neuroni artificiali) organizzati in strati, capaci di apprendere relazioni complesse tra dati tramite un processo di addestramento

Rete neurale convoluzionale Architettura di deep learning progettata per elaborare dati strutturati spazialmente, come immagini o segnali, attraverso l'uso di filtri convoluzionali che estraggono automaticamente caratteristiche gerarchiche dai dati in input

Scenario “what-if” Simulazione ipotetica che consente di analizzare l'impatto di modifiche nei parametri di un sistema, esplorando come il comportamento potrebbe cambiare in risposta a determinate condizioni o eventi

Sistema HVAC Insieme di tecnologie per il riscaldamento, la ventilazione e la climatizzazione degli ambienti, impiegate per garantire comfort termico e qualità dell'aria negli edifici civili e industriali

Sistema SCADA Sistema di supervisione, controllo e acquisizione dati utilizzati per monitorare e gestire processi industriali in tempo reale

Smart city Città che utilizza tecnologie digitali e soluzioni intelligenti per migliorare la qualità della vita dei cittadini, l'efficienza dei servizi urbani e la sostenibilità ambientale

Smart factory Sistema produttivo altamente automatizzato e interconnesso, basato su tecnologie digitali come IoT, intelligenza artificiale e Digital Twin, in grado di adattarsi in tempo reale ai cambiamenti e ottimizzare autonomamente i processi

Support Vector Machines Algoritmi di apprendimento supervisionato utilizzati per compiti di classificazione e regressione, che identificano un iperpiano ottimale per separare i dati in classi distinte massimizzando il margine tra esse

Tuning Attività di regolazione o ottimizzazione dei parametri di un sistema, modello o algoritmo, finalizzata a migliorarne le prestazioni in base a criteri specifici o metriche di valutazione

Variabile categorica Variabile che assume valori discreti, spesso qualitativi (es. “difettoso” / “non difettoso”)

“if-then” Regole logiche condizionali utilizzate per rappresentare conoscenza esplicita e formalizzare decisioni nei sistemi esperti, nella logica fuzzy o nei modelli simbolici

Acronimi

CAD Computer-Aided Design

CAE Computer-Aided Engineering

CNC Computerized Numerical Control

CNN Convolutional Neural Network

DT Digital Twin

FEM Finite Element Method

GAN Generative Adversarial Networks

GIS Geographic Information System

IBM International Business Machines Corporation

LightGBM Light Gradient Boosting Machine

MES Manufacturing Execution System

MobileViG Mobile Vision Graph

PLC Programmable Logic Controller

PLM Product Lifecycle Management

PTC Parametric Technology Corporation

RUL Remaining Useful Life

SCADA Supervisory Control and Data Acquisition

Bibliografia

- [1] Ekaterina V. Orlova. «Design Technology and AI-Based Decision Making Model for Digital Twin Engineering». In: *Future Internet* 14.9 (2022). ISSN: 1999-5903. DOI: 10.3390/fi14090248. URL: <https://www.mdpi.com/1999-5903/14/9/248> (cit. alle pp. 3, 11).
- [2] Jinyue Li, Gang Zhao, Pengfei Zhang, Maocheng Xu, He Cheng e Han Pengfei. «A Digital Twin-based on-site quality assessment method for aero-engine assembly». In: *Journal of Manufacturing Systems* 71 (2023), pp. 565–580. ISSN: 0278-6125. DOI: <https://doi.org/10.1016/j.jmsy.2023.10.008>. URL: <https://www.sciencedirect.com/science/article/pii/S0278612523002108> (cit. a p. 3).
- [3] Tao Ji e Norzalilah Mohamad Nor. «Deep Learning-Empowered Digital Twin Using Acoustic Signal for Welding Quality Inspection». In: *Sensors* 23.5 (2023). ISSN: 1424-8220. DOI: 10.3390/s23052643. URL: <https://www.mdpi.com/1424-8220/23/5/2643> (cit. alle pp. 4, 7, 8, 10, 37).
- [4] Tao Ji, Norzalilah Mohamad Nor e Ahmad Baharuddin Abdullah. «Multi-modal recognition control system for real-time robot welding penetration control and quality enhancement». In: *The International Journal of Advanced Manufacturing Technology* 135.9 (2024), pp. 4359–4378 (cit. alle pp. 4, 13–16, 36).
- [5] Fei Tao, Fangyuan Sui, Ang Liu, Qinglin Qi, Meng Zhang, Boyang Song, Zirong Guo, Stephen C.-Y. Lu e A. Y. C. Nee. «Digital twin-driven product design framework». In: *International Journal of Production Research* 57.12 (2019), pp. 3935–3953. DOI: 10.1080/00207543.2018.1443229. eprint: <https://doi.org/10.1080/00207543.2018.1443229>. URL: <https://doi.org/10.1080/00207543.2018.1443229> (cit. alle pp. 4, 6, 7).
- [6] Matevž Resman, Niko Herakovič e Mihael Debevec. «Integrating Digital Twin Technology to Achieve Higher Operational Efficiency and Sustainability in Manufacturing Systems». In: *Systems* 13.3 (2025). ISSN: 2079-8954. DOI: 10.3390/systems13030180. URL: <https://www.mdpi.com/2079-8954/13/3/180> (cit. alle pp. 4, 11–13, 36–38).

- [7] Lisa Gopal, Harbaksh Singh, Panguluri Mounica, N. Mohankumar, Nagendra Panini Challa e P. Jayaraman. «Digital twin and IOT technology for secure manufacturing systems». In: *Measurement: Sensors* 25 (2023), p. 100661. ISSN: 2665-9174. DOI: <https://doi.org/10.1016/j.measen.2022.100661>. URL: <https://www.sciencedirect.com/science/article/pii/S2665917422002951> (cit. alle pp. 4, 38).
- [8] Qinglin Qi e Fei Tao. «Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison». In: *Ieee Access* 6 (2018), pp. 3585–3593 (cit. a p. 6).
- [9] David Jones, Chris Snider, Aydin Nassehi, Jason Yon e Ben Hicks. «Characterising the Digital Twin: A systematic literature review». In: *CIRP Journal of Manufacturing Science and Technology* 29 (2020), pp. 36–52. ISSN: 1755-5817. DOI: <https://doi.org/10.1016/j.cirpj.2020.02.002>. URL: <https://www.sciencedirect.com/science/article/pii/S1755581720300110> (cit. a p. 6).
- [10] Aidan Fuller, Zhong Fan, Charles Day e Chris Barlow. «Digital Twin: Enabling Technologies, Challenges and Open Research». In: *IEEE Access* 8 (2020), pp. 108952–108971. DOI: [10.1109/ACCESS.2020.2998358](https://doi.org/10.1109/ACCESS.2020.2998358) (cit. alle pp. 6, 7, 10).
- [11] Faeze Hodavand, Issa J. Ramaji e Naimeh Sadeghi. «Digital Twin for Fault Detection and Diagnosis of Building Operations: A Systematic Review». In: *Buildings* 13.6 (2023). ISSN: 2075-5309. DOI: [10.3390/buildings13061426](https://doi.org/10.3390/buildings13061426). URL: <https://www.mdpi.com/2075-5309/13/6/1426> (cit. alle pp. 10–12, 30).
- [12] Yingying Song, Zhaoliang Qu, Haitao Liao e Shigang Ai. «Material twins generation of woven polymer composites based on ResL-U-Net convolutional neural networks». In: *Composite Structures* 307 (2023), p. 116672. ISSN: 0263-8223. DOI: <https://doi.org/10.1016/j.compstruct.2023.116672>. URL: <https://www.sciencedirect.com/science/article/pii/S0263822323000168> (cit. a p. 11).
- [13] Ruiping Luo, Buyun Sheng, Yingkang Lu, Yuzhe Huang, Gaocai Fu e Xiyan Yin. «Digital Twin Model Quality Optimization and Control Methods Based on Workflow Management». In: *Applied Sciences* 13.5 (2023). ISSN: 2076-3417. DOI: [10.3390/app13052884](https://doi.org/10.3390/app13052884). URL: <https://www.mdpi.com/2076-3417/13/5/2884> (cit. a p. 11).
- [14] Toyosi Ademujimi e Vittaldas Prabhu. «Digital Twin for Training Bayesian Networks for Fault Diagnostics of Manufacturing Systems». In: *Sensors* 22.4 (2022). ISSN: 1424-8220. DOI: [10.3390/s22041430](https://doi.org/10.3390/s22041430). URL: <https://www.mdpi.com/1424-8220/22/4/1430> (cit. alle pp. 12–14).

- [15] Weimin Liu, Bin Han, Aiyun Zheng, Zhi Zheng, Shujun Chen e Shikui Jia. «Fault diagnosis of reducers based on digital twins and deep learning». In: *Scientific Reports* 14.1 (2024), p. 24406 (cit. alle pp. 13, 15, 36, 37).
- [16] Yefu Hu, Omer W. Taha e Kezhen Yang. «Fault Detection in Active Magnetic Bearings Using Digital Twin Technology». In: *Applied Sciences* 14.4 (2024). ISSN: 2076-3417. DOI: 10.3390/app14041384. URL: <https://www.mdpi.com/2076-3417/14/4/1384> (cit. a p. 14).
- [17] Samar M Zayed, Gamal Attiya, Ayman El-Sayed, Amged Sayed e Ezz El-Din Hemdan. «An efficient fault diagnosis framework for digital twins using optimized machine learning models in smart industrial control systems». In: *International Journal of Computational Intelligence Systems* 16.1 (2023), p. 69 (cit. alle pp. 14, 36).
- [18] Hao Li, Bing Li, Gen Liu, Xiaoyu Wen, Haoqi Wang, Xiaocong Wang, Shuai Zhang, Zhongshang Zhai e Wenchao Yang. «A detection and configuration method for welding completeness in the automotive body-in-white panel based on digital twin». In: *Scientific Reports* 12.1 (2022), p. 7929 (cit. a p. 14).
- [19] Tat Nghia Nguyen, Roberto Ponciroli, Paul Bruck, Thomas C. Esselman, Joseph A. Rigatti e Richard B. Vilim. «A digital twin approach to system-level fault detection and diagnosis for improved equipment health monitoring». In: *Annals of Nuclear Energy* 170 (2022), p. 109002. ISSN: 0306-4549. DOI: <https://doi.org/10.1016/j.anucene.2022.109002>. URL: <https://www.sciencedirect.com/science/article/pii/S0306454922000378> (cit. alle pp. 14, 38).
- [20] Pavol Mulinka, Subham Sahoo, Charalampos Kalalas e Pedro H. J. Nardelli. «Optimizing a Digital Twin for Fault Diagnosis in Grid Connected Inverters - A Bayesian Approach». In: *2022 IEEE Energy Conversion Congress and Exposition (ECCE)*. 2022, pp. 1–6. DOI: 10.1109/ECCE50734.2022.9947986 (cit. alle pp. 14–16, 37).
- [21] Yirui Wu, Hao Cao, Guoqiang Yang, Tong Lu e Shaohua Wan. «Digital Twin of Intelligent Small Surface Defect Detection with Cyber-manufacturing Systems». In: *ACM Trans. Internet Technol.* 23.4 (nov. 2023). ISSN: 1533-5399. DOI: 10.1145/3571734. URL: <https://doi.org/10.1145/3571734> (cit. alle pp. 17, 36).
- [22] Yiyong Tan, Bhaskar Banerjee e Rishi Ranjan. «Few-Shot and Portable 3D Manufacturing Defect Tracking with Enterprise Digital Twins Based Mixed Reality». In: *Pattern Recognition*. A cura di Apostolos Antonacopoulos, Subhasis Chaudhuri, Rama Chellappa, Cheng-Lin Liu, Saumik Bhattacharya e Umapada Pal. Cham: Springer Nature Switzerland, 2025, pp. 17–32. ISBN: 978-3-031-80136-5 (cit. alle pp. 17, 38).

- [23] David Moher, Alessandro Liberati, Jennifer Tetzlaff e Douglas G. Altman. «Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement». In: *International Journal of Surgery* 8.5 (2010), pp. 336–341. ISSN: 1743-9191. DOI: <https://doi.org/10.1016/j.ijsu.2010.02.007>. URL: <https://www.sciencedirect.com/science/article/pii/S1743919110000403> (cit. a p. 19).
- [24] *A Digital Twin Assisted Framework for Quality Assurance in Mould Manufacturing*. Vol. Volume 2: 43rd Computers and Information in Engineering Conference (CIE). International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Ago. 2023, V002T02A070. DOI: 10.1115/DETC2023-114848. eprint: <https://asmedigitalcollection.asme.org/IDETC-CIE/proceedings-pdf/IDETC-CIE2023/87295/V002T02A070/7060993/v002t02a070-detc2023-114848.pdf>. URL: <https://doi.org/10.1115/DETC2023-114848> (cit. a p. 26).
- [25] Waqas Ahmad, Marcel Mutz e Dirk Werth. «Digital Twin of Rail for Defect Analysis». In: *Proceedings of the 2024 8th International Conference on Virtual and Augmented Reality Simulations*. ICVARS '24. Melbourne, Australia: Association for Computing Machinery, 2024, pp. 53–60. ISBN: 9798400709012. DOI: 10.1145/3657547.3657549. URL: <https://doi.org/10.1145/3657547.3657549> (cit. a p. 31).
- [26] Till Böttjer e Devarajan Ramanujan. «A Digital Twin Assisted Framework for Quality Assurance in Mould Manufacturing». In: *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. Vol. 87295. American Society of Mechanical Engineers. 2023, V002T02A070 (cit. alle pp. 39, 40).
- [27] Till Boettjer. «Mould insert part metrology dataset». In: (mar. 2023). DOI: 10.6084/m9.figshare.22251301.v1. URL: https://figshare.com/articles/dataset/Mould_insert_part_metrology_dataset/22251301 (cit. a p. 41).
- [28] Python Software Foundation. *pathlib — Object-oriented filesystem paths*. 2024. URL: <https://docs.python.org/3/library/pathlib.html> (cit. a p. 42).
- [29] Leo Breiman. «Random forests». In: *Machine learning* 45.1 (2001), pp. 5–32 (cit. a p. 46).
- [30] Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.", 2022 (cit. a p. 46).
- [31] Trevor Hastie, Robert Tibshirani, Jerome Friedman et al. *The elements of statistical learning*. 2009 (cit. a p. 46).
- [32] Ian Goodfellow, Yoshua Bengio, Aaron Courville e Yoshua Bengio. *Deep learning*. Vol. 1. 2. MIT press Cambridge, 2016 (cit. alle pp. 46, 57).

- [33] David W Hosmer Jr, Stanley Lemeshow e Rodney X Sturdivant. *Applied logistic regression*. John Wiley & Sons, 2013 (cit. a p. 46).
- [34] Scott Menard. *Applied logistic regression analysis*. SAGE publications, 2001 (cit. a p. 46).
- [35] Gareth James, Daniela Witten, Trevor Hastie e Robert Tibshirani. *An introduction to statistical learning: with applications in R*. Vol. 103. Springer, 2013 (cit. a p. 46).
- [36] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye e Tie-Yan Liu. «Lightgbm: A highly efficient gradient boosting decision tree». In: *Advances in neural information processing systems* 30 (2017) (cit. alle pp. 46, 47).
- [37] Zhi-Hua Zhou. *Ensemble methods: foundations and algorithms*. CRC press, 2025 (cit. a p. 47).
- [38] *RandomForestClassifier*. 2025. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html> (cit. a p. 49).
- [39] *Compute_class_weight*. 2025. URL: https://scikit-learn.org/stable/modules/generated/sklearn.utils.class_weight.compute_class_weight.html (cit. a p. 49).
- [40] *MLPClassifier*. 2025. URL: https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html (cit. a p. 49).
- [41] *LogisticRegression*. 2025. URL: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html (cit. a p. 50).
- [42] *LightGBM*. 2025. URL: <https://lightgbm.readthedocs.io/en/stable/> (cit. a p. 50).
- [43] H He e EA Garcia. *Learning from Imbalanced Data IEEE Transactions on Knowledge and Data Engineering*. 2009 (cit. a p. 52).
- [44] *SMOTENC*. 2025. URL: https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTENC.html (cit. a p. 55).
- [45] Python Software Foundation. *json — JSON encoder and decoder*. 2024. URL: <https://docs.python.org/3/library/json.html> (cit. a p. 56).
- [46] *RandomOverSampler*. 2025. URL: https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.RandomOverSampler.html (cit. a p. 57).
- [47] *Preprocessing data*. 2011. URL: <https://scikit-learn.org/stable/modules/preprocessing.html> (cit. a p. 57).

- [48] *Pipeline*. 2025. URL: <https://imbalanced-learn.org/stable/references/generated/imblearn.pipeline.Pipeline.html> (cit. a p. 58).
- [49] *Metrics and scoring: quantifying the quality of predictions*. 2025. URL: https://scikit-learn.org/stable/modules/model_evaluation.html (cit. alle pp. 68, 69).
- [50] *Roc_curve*. 2025. URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html (cit. a p. 69).
- [51] *Precision_recall_curve*. 2025. URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_curve.html (cit. a p. 69).
- [52] *Matplotlib*. 2025. URL: <https://matplotlib.org/stable/>.