POLITECNICO DI TORINO
Repository ISTITUZIONALE

Intelligenza Artificiale Generativa per la produzione di dati sintetici nella diagnosi di macchine rotanti

Original
Intelligenza Artificiale Generativa per la produzione di dati sintetici nella diagnosi di macchine rotanti / Di Maggio, Luigi Gianpio; Brusa, Eugenio; Delprete, Cristiana. - ELETTRONICO. - (2023). (Intervento presentato al convegno AIAS2023 - 52° Convegno tenutosi a Genova nel 6-9/9/2023).

Availability:
This version is available at: 11583/2982252 since: 2023-09-18T10:00:16Z

Publisher:
AIAS

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

15 May 2024

AIAS 2023 - 1982

INTELLIGENZA ARTIFICIALE GENERATIVA PER LA PRODUZIONE DI DATI SINTETICI NELLA DIAGNOSI DI MACCHINE ROTANTI

L. G. Di Maggio, E. Brusa, C. Delprete

Politecnico di Torino – Dipartimento di Ingegneria Meccanica e Aerospaziale,
Corso Duca degli Abruzzi 24, 10129 Torino,
e-mail: luigi.dimaggio@polito.it, eugenio.brusa@polito.it, cristiana.delprete@polito.it

Sommario
L'Intelligenza Artificiale (IA) sta acquisendo crescente rilevanza nell'ambito della diagnosi del macchinario rotante, grazie alla sua capacità di correlare con accuratezza i parametri diagnostici con lo stato di salute. Tuttavia, gli approcci di IA richiedono una quantità considerevole di dati di addestramento, che risultano difficilmente reperibili per macchinari industriali con danni o difetti. Questo studio propone una IA generativa basata su una Rete Generativa Avversaria (GAN) con funzione di perdita a consistenza ciclica (*cycleGAN*) per generare dati sintetici sostitutivi. Il modello è addestrato per convertire immagini wavelet di segnali vibrazionali simulati in dati reali relativi a macchinari operanti con cuscinetti danneggiati. L'utilizzo della *Maximum Mean Discrepancy* (MMD) evidenzia una notevole similitudine tra dati sintetici e reali. I dati sintetici si dimostrano efficaci per allenare algoritmi di diagnosi mediante *Transfer Learning* (TL) su Reti Neurali Convoluzionali (CNNs). Lo studio è validato attraverso un banco prova per cuscinetti industriali disponibile presso il Dipartimento di Ingegneria Meccanica e Aerospaziale del Politecnico di Torino.

Abstract
The subject of rotating machinery diagnostics is witnessing a growing significance of Artificial Intelligence (AI) as it demonstrates a remarkable capability to establish correlations between diagnostic parameters and the overall health condition. Nevertheless, the use of AI methodologies necessitates a substantial amount of training data, posing challenges in acquiring such data for damaged industrial machineries. This research presents a novel approach for generating synthetic data using Cycle-Consistent Generative Adversarial Networks (cycleGANs). The proposed model is designed to transform wavelet-based images of simulated vibration signals into real data obtained from machines exhibiting bearing faults. The Maximum Mean Discrepancy (MMD) demonstrates a noteworthy resemblance between synthetic and real data. Synthetic data result effective for training Convolutional Neural Networks (CNNs) by means of Transfer Learning (TL). The research is conducted using the test rig for industrial bearing located at the Mechanical and Aerospace Engineering Department of Politecnico di Torino.

Parole chiave: meccatronica strutturale, progettazione dinamica dei rotori, intelligenza artificiale, cuscinetti volventi, monitoraggio, diagnosi, reti generative avversarie, reti neurali convoluzionali, Transfer Learning.

1. INTRODUZIONE

Oggigiorno l'Intelligenza Artificiale (IA) riveste un ruolo primario nella diagnosi del macchinario rotante [1]–[4]. L'interesse verso gli approcci di *Machine Learning* (ML) e *Deep Learning* (DL) può essere attribuito al potenziale di tali algoritmi di acquisire conoscenza in modo altamente automatizzato, stabilendo connessioni significative tra variabili di ingresso, spesso derivate dai segnali di monitoraggio [5], e gli output diagnostici. L'accuratezza raggiungibile attraverso tali metodologie rende l'IA uno strumento prezioso per gli obiettivi di manutenzione predittiva. In questo scenario, il coinvolgimento e le conoscenze degli operatori sono ridimensionati, in quanto dipendenti dalle informazioni acquisite da algoritmi intelligenti mediante i dati di addestramento. Tale approccio manutentivo è spesso associato

con la diagnosi dei cuscinetti volventi [6]–[8], essendo componenti meccanici in grado di accentrare informazioni relative a tutto il macchinario e alle sue prestazioni complessive [9]–[11].
Nell'ultimo decennio, sono stati adottati diversi approcci per la diagnosi intelligente del macchinario rotante. Nell'ambito delle metodologie ML, le caratteristiche diagnostiche sono estratte manualmente e trattate mediante algoritmi di apprendimento automatico come il *Support Vector Machine* (SVM) [12]–[14] e il *k-Nearest Neighbors* (kNN) [15]–[17]. Il DL si distingue invece per le sue capacità di estrazione automatica delle caratteristiche di diagnosi e l'apprendimento gerarchico delle stesse [18]. Il DL è stato applicato principalmente mediante le Reti Neurali Convoluzionali (CNNs) [19]–[21]. In genere, nell'ambito della diagnosi delle macchine rotanti, i dati sono trasformati in forma di immagini a partire da segnali accelerometrici, come nel caso di spettrogrammi e trasformate wavelet continue (CWT) [22]. Ciononostante, l'efficacia delle tecniche ML e DL è subordinata alla presenza di un numero adeguato di dati di allenamento [23,24] che siano rappresentativi delle possibili condizioni operative della macchina e delle potenziali difettosità della stessa. Nella pratica industriale, però, i dati non sono sempre reperibili in quantità sufficiente e i dataset pubblici attualmente esistenti [25]–[28] non sono in grado di rappresentare una distribuzione generica e significativa di possibili danneggiamenti. Per questo motivo, una porzione consistente della letteratura scientifica del settore si concentra sull'applicazione di metodologie volte a ridurre la numerosità dei dati di addestramento. In quest'ambito, le tecniche di *Transfer Learning* (TL) [3,29] sono utilizzate per trasferire la conoscenza pre-acquisita da modelli di IA a differenti scenari diagnostici, ridimensionando la richiesta di dati di addestramento.
Nell'ambito dell'apprendimento automatico, il TL rappresenta una metodologia in cui un modello viene inizialmente addestrato su un vasto insieme di dati e successivamente impiegato come base pre-addestrata per affrontare compiti analoghi o con affinità, richiedendo minori quantità di dati e tempo di addestramento rispetto all'allenamento ex-novo. In letteratura sono riportati esempi di trasferimento delle capacità di diagnosi verso nuove condizioni operative dello stesso macchinario [30,31] e verso differenti macchinari [32]. Diverse metodologie sono inoltre presenti per reimpiegare modelli pre-allenati per il riconoscimento di immagini [33] e suoni [34]–[36] nella diagnosi del macchinario rotante. Le reti generative [37,38], invece, sono applicate nella diagnosi [39]–[41] per sviluppare modelli generativi in grado di produrre dati sintetici simili a quelli rilevati dai macchinari in funzione. I dati generati artificialmente vengono impiegati per ampliare il dataset di addestramento, permettendo al modello di apprendere una rappresentazione più completa delle condizioni operative. I dati sintetici risultano applicabili efficacemente per la diagnosi di sistemi rotanti. Inoltre, le reti generative possono simulare diverse condizioni di guasto o situazioni anomale, fornendo al modello una gamma più ampia di informazioni su possibili scenari di malfunzionamento.
Il presente lavoro si propone di generare dati sintetici che possano essere impiegati per ampliare o sostituire i dati di danneggiamento necessari per l'addestramento di algoritmi di IA per la diagnosi di cuscinetti volventi. Diversamente dalla letteratura preesistente, il lavoro si propone di generare dati sintetici anche per condizioni operative per le quali il modello generativo non sia stato esplicitamente istruito. A tale scopo, è stata implementata una Rete Generativa Avversaria (GAN) con funzione di perdita a consistenza ciclica (*cycleGAN*). La rete è stata allenata per convertire immagini CWT di segnali vibrazionali simulati in segnali reali e successivamente utilizzata per generare dati sintetici. Questi ultimi sono stati impiegati per allenare modelli di diagnosi mediante la tecnica del TL. I risultati sono stati validati mediante l'attività sperimentale condotta sul banco prova cuscinetti industriali disponibile presso il laboratorio del Dipartimento di Ingegneria Meccanica e Aerospaziale del Politecnico di Torino [42].

2. INTELLIGENZA ARTIFICIALE GENERATIVA E DIAGNOSI DI MACCHINE ROTANTI

Le implementazioni pratiche della IA generativa per la diagnosi dei sistemi rotanti si realizzano principalmente attraverso l'impiego diffuso delle Reti Generative Avversarie (GANs), inizialmente proposte nel 2014 da Goodfellow et al. [38] nell'ambito dell'IA generativa. Le GANs sono costituite da due strutture convoluzionali che competono in un gioco a "somma zero" [43,44]. Un esempio di GAN è mostrato in Fig. 1. Nello specifico, il generatore G tenta di produrre immagini replicando una certa distribuzione Y di immagini, a partire dal rumore bianco X. Il discriminatore D consiste in una CNN che agisce da classificatore binario per distinguere quali immagini siano reali e quali fittizie, ovvero prodotte dal generatore mediante la trasformazione $X \rightarrow G(X)$.

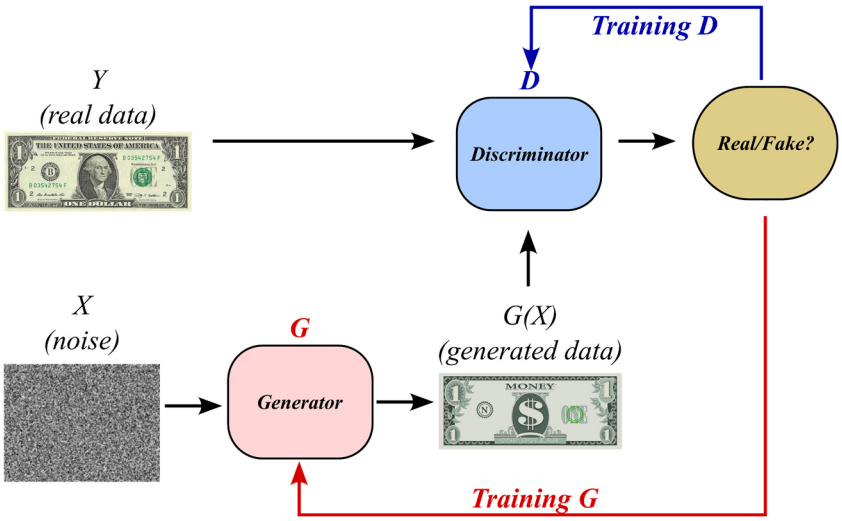


Figura 1. Esempio di Rete Generativa Avversaria (GAN).

L'interazione competitiva fra le due strutture convoluzionali è matematicamente rappresentata dalla funzione obiettivo $\mathcal{L}_{GAN}$ riportata nell'Eq. (1), dove $\mathbb{E}[\cdot]$ è l'operatore valore atteso, y è un generico dato appartenente al dominio Y con distribuzione di probabilità p_y, x è un generico dato appartenente al dominio X con distribuzione di probabilità p_x, $G(\cdot)$ è l'output del generatore e $D_Y(\cdot)$ è l'output del discriminatore, ovvero la probabilità che l'immagine data in input a quest'ultimo sia reale. Dal punto di vista del discriminatore è conveniente massimizzare la funzione $\mathcal{L}_{GAN}$, mentre per il generatore è opportuno minimizzarla. Tale dicotomia dà luogo a un gioco a somma zero. L'allenamento della GAN si concretizza nella ricerca di un equilibrio di Nash [43], [44] tra generatore e discriminatore (Eq. (2)).

$$\mathcal{L}_{GAN} = \mathbb{E}_{y\sim p_y}[\log D_Y(y)] + \mathbb{E}_{x\sim p_x}[\log(1 - D_Y(G(x))] \tag{1}$$

$$\min_G \max_D \mathcal{L}_{GAN} \tag{2}$$

Le reti generative cycleGANs sono state introdotte nel 2017 da Zhu et al. [45] e rappresentano un'evoluzione delle reti GANs. Le cycleGANs sono costituite da due generatori G e F e due discriminatori D_Y e D_X (Fig. 2) collegati matematicamente dalla funzione di perdita a consistenza ciclica riportata nell'Eq. (3). La funzione obiettivo globale è riportata invece nell'Eq. (4), dove λ è il peso assegnato alla funzione $\mathcal{L}_{cyc}$.

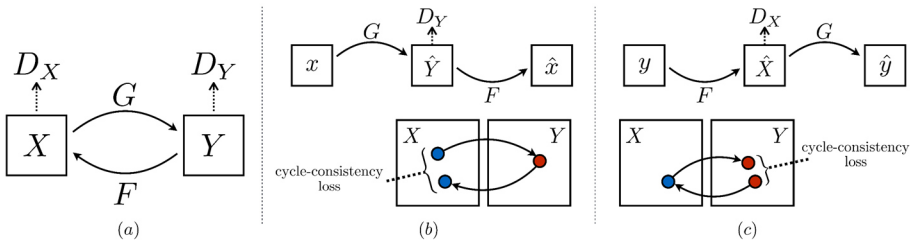


Figura 2. Schema del funzionamento di una cycleGAN [45].

$$\mathcal{L}_{cyc} = \mathbb{E}_{x\sim p_x}\left[\left\|F(G(x)) - x\right\|_1\right] + \mathbb{E}_{y\sim p_y}\left[\left\|G(F(y)) - y\right\|_1\right] \tag{3}$$

$$\mathcal{L}_{tot} = \mathcal{L}_{GAN(x\to y)} + \mathcal{L}_{GAN(y\to x)} + \lambda\mathcal{L}_{cyc} \tag{4}$$

La funzione $\mathcal{L}_{cyc}$ permette di stabilire una corrispondenza biunivoca tra due domini di immagini X e Y, anche quando le coppie di dati di allenamento $\{x_i, y_i\}$ non hanno il proprio corrispettivo esatto nel domino accoppiato. Difatti, la funzione $\mathcal{L}_{cyc}$ dell'Eq. (3) vincola il modello a ritornare alla medesima immagine x dopo aver effettuato la trasformazione $x \to G(x) \to F(G(x)) \approx x$. Tale implementazione rende non necessaria la presenza di un'immagine di training y_i che sia perfettamente accoppiata con x_i. La Fig. 3 mostra alcuni esempi di applicazione di cycleGANs relativi alla conversione di immagini tra due differenti domini.

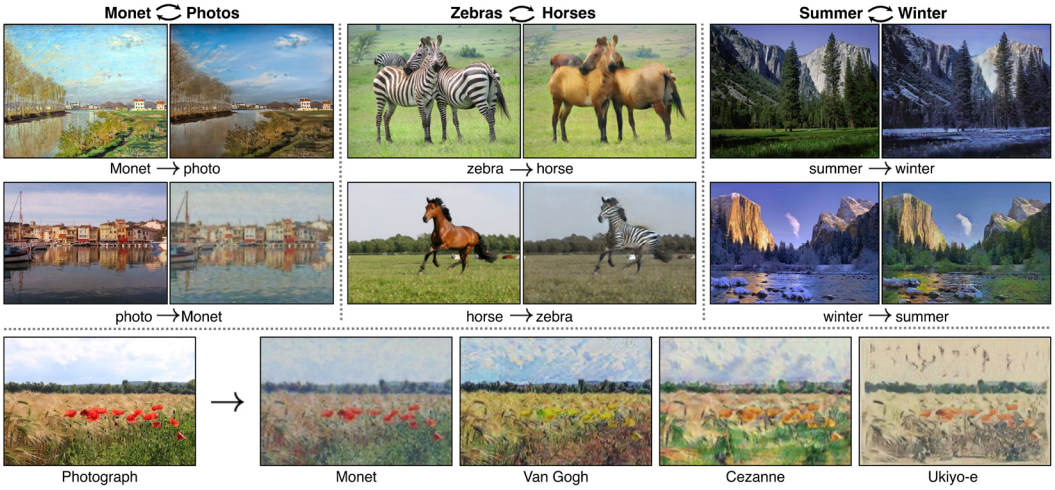


Figura 3. Esempi di traslazione di immagini da un dominio X ad un dominio Y [45].

La metodologia di produzione dei dati sintetici presentata in questo studio sfrutta la tecnologia sottesa alle reti cycleGANs al fine di convertire immagini della trasformata CWT di segnali vibrazionali simulati nel loro corrispettivo reale. Gli esempi riportati in Fig. 3 dimostrano le capacità delle cycleGANs di tradurre l'immagine di un dipinto in una foto oppure un paesaggio estivo in un paesaggio invernale. Questo studio impiega un'architettura cycleGAN per traslare il risultato di un modello di simulazione nel suo corrispettivo reale (Fig. 4). In questo lavoro, si ipotizza altresì che la proprietà di consistenza ciclica possa essere utilizzata per generare dati relativi a condizioni operative del macchinario rotante per il quale il modello di conversione cycleGAN non sia stato esplicitamente addestrato.

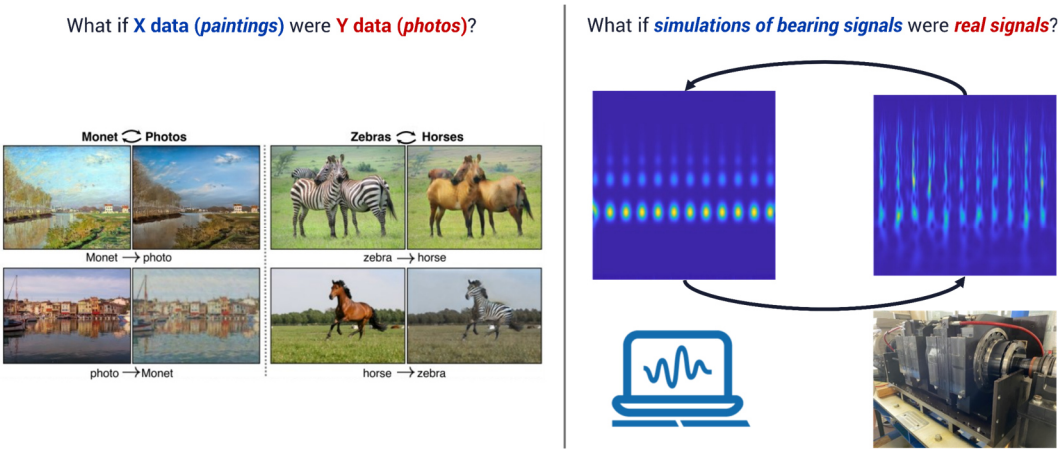


Figura 4. Applicazione di IA generativa cycleGAN per la diagnosi di sistemi rotanti.

3. METODOLOGIA DI PRODUZIONE DEI DATI SINTETICI

La procedura per la generazione dei dati sintetici illustrata in questo paragrafo coinvolge la realizzazione di una campagna sperimentale volta a estrarre segnali vibrazionali da cuscinetti di dimensioni medie, sia in condizioni di salute sia di danneggiamento. In seguito, si procede con la simulazione analitica del comportamento vibrazionale. I segnali ottenuti da entrambe queste fonti vengono elaborati per renderli idonei all'addestramento di un sistema di IA generativa. Quest'ultima viene istruita a convertire i risultati ottenuti da un modello di simulazione in dati che presentano una somiglianza molto marcata con quelli osservabili in una situazione sperimentale reale. Il modello generativo addestrato in questa modalità ha la capacità di generare dati che possono fungere da sostituti per quelli derivanti dall'esperimento, risultando utili per l'addestramento di modelli diagnostici.

3.1. Attività sperimentale su banco prova per cuscinetti industriali

L'attività sperimentale è relativa al banco prova per cuscinetti industriali di media taglia disponibile presso il laboratorio del Dipartimento di Ingegneria Meccanica e Aerospaziale del Politecnico di Torino [36,42]. Il banco prova, mostrato in Fig. 5(a), è in grado di ospitare fino a 4 cuscinetti con diametro esterno compreso tra 280 mm e 420 mm. Il motore trifase da 30 kW alimenta il moto dell'albero principale mediante un giunto elastico PERIFLEX®. I carichi radiali e assiali sono forniti in maniera indipendente fino a 200 kN attraverso attuatori idraulici agenti sugli adattatori che ospitano i cuscinetti in prova (Fig. 5(b)). Gli attuatori idraulici sono attivati da pompe di conversione aria-olio connesse al sistema pneumatico disponibile in laboratorio. I cuscinetti in prova sono lubrificati mediante un sistema di ricircolo esterno con olio ISO VG150 iniettato con una portata di 2.5 l/min a una pressione di 6 bar. L'architettura del banco è racchiusa nella "self-contained box". Tale configurazione permette di bilanciare i carichi di prova mediante la deformazione elastica della box, esaurendo il circuito di carico all'interno della stessa. Uno dei vantaggi di questa configurazione è rappresentato dalla possibilità di gestire carichi di considerevole entità senza ricorrere a cuscinetti di banco di grande taglia. Ciascun adattatore ospita un sensore di accelerazione/temperatura SKF CMS 2200T connesso a un sistema di acquisizione LMS Scadas III. Per una descrizione di maggior dettaglio del banco prova si rimanda ai lavori [35,42].

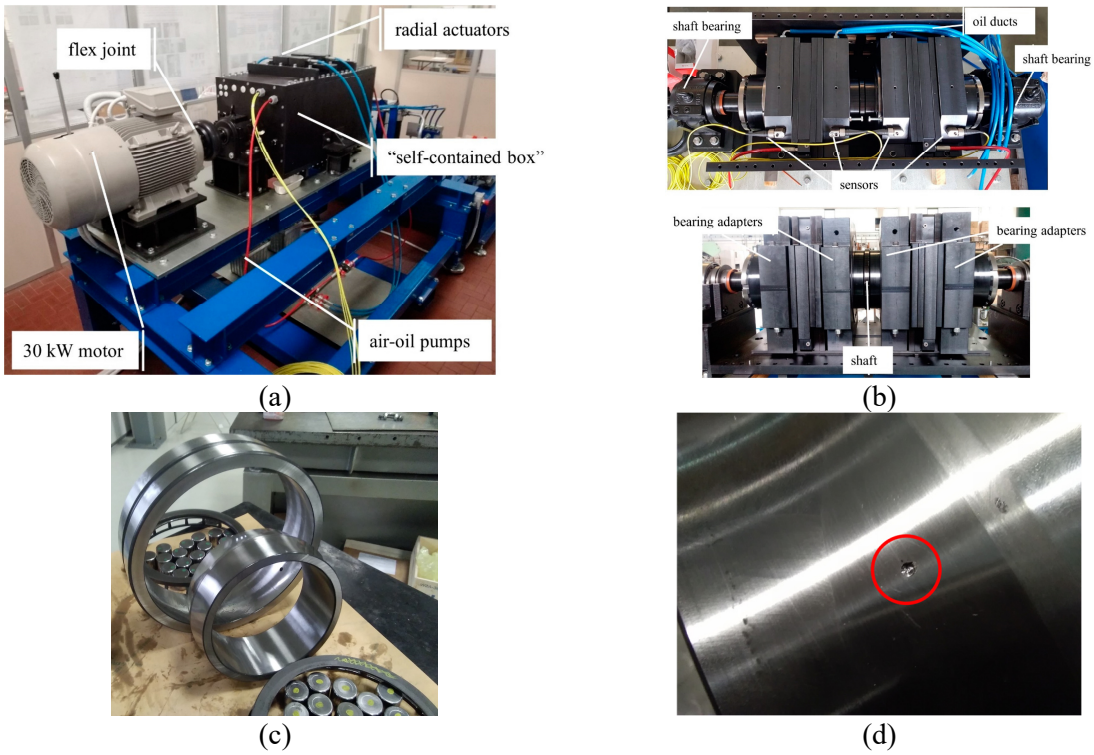


Figura 5. Attività sperimentale: (a) banco prova cuscinetti industriali [42]; (b) componenti interni al banco prova [42]; (c) smontaggio cuscinetto; (d) danneggiamento su pista interna (IR).

La sperimentazione è stata effettuata analizzando i campioni vibrazionali estratti dal banco prova oggetto di studio. In particolare, è stato testato il cuscinetto orientabile a rulli SKF 22240 CCK/W33. Il cuscinetto ha un diametro interno di 200 mm con conicità di 1:12 e un diametro esterno di 360 mm. La Fig. 5(c) mostra il sistema cuscinetto in fase di disassemblaggio. Il cuscinetto è stato testato nella condizione di funzionamento normale e nella condizione di danneggiamento sulla pista interna (IR). Il difetto localizzato è stato introdotto mediante lavorazione meccanica e ha un diametro di 2 mm e una profondità di 0.5 mm (Fig. 5(d)). La Tabella 1 riporta le condizioni di prova e le caratteristiche di estrazione del segnale accelerometrico. Il modello generativo è stato allenato utilizzando la condizione di lavoro a 877 rpm. Di converso, i dati relativi alle condizioni di lavoro a 607 rpm e a 997 rpm non sono stati impiegati per allenare il modello cycleGAN.

Tabella 1. Condizioni di prova per cuscinetto SKF 22240 CCK/W33.

Carico radiale (kN)	124.8
Velocità di rotazione (rpm)	607, 877, 997
Tempo (s)	30
Frequenza di campionamento f_s (Hz)	20480
Overlap campioni	0.7
Durata campioni (s)	0.8
Campioni per segnale	250

3.2. Simulazione del segnale vibrazionale di cuscinetti danneggiati

Per la simulazione del comportamento vibrazionale di cuscinetti volventi con difetti localizzati ci si è avvalsi del modello analitico di McFadden e Smith [46]–[48]. Il modello è ben consolidato in letteratura, nonostante sia uno dei primi tentativi volti a replicare il comportamento di cuscinetti difettati. Recenti studi [49] hanno evidenziato che una modellistica più avanzata, che per esempio contempli analisi agli elementi finiti tridimensionali e fenomeni di meccanica del contatto, fornisce informazioni aggiuntive in merito alle interazioni tra i componenti del sistema cuscinetto; tuttavia il comportamento generale del sistema rimane ben rappresentato dal modello di McFadden e Smith.
In questo lavoro, è stato impiegato questo modello in quanto esso non include diverse informazioni relative al sistema meccanico in esame. Per esempio, il modello non tiene conto delle interazioni inerziali tra corpi volventi, lubrificante e piste, né del comportamento elasto-idrodinamico del lubrificante. Ai fini di questo studio, tali mancanze costituiscono un vantaggio poiché consentono di valutare la capacità del modello generativo di incorporare tutte le informazioni mancanti nel modello simulato, specialmente nel caso specifico in cui la complessità del modello simulato sia notevolmente ridotta.
Il modello è riassunto nelle Eq. da (5) a (8), dove la risposta all'impulso $h(t)$ viene ottenuta a partire dal parametro di decadimento $\beta = 500$ Hz e dalla risonanza strutturale $f_{struct} = 1700$ Hz; $d(t)$ è la funzione pettine di Dirac, che contiene un treno di impulsi spaziati del reciproco della frequenza caratteristica del difetto f_{defect}. Per il danno IR praticato sul cuscinetto in esame si considera $f_{defect} = 10.824 \cdot f_r$, dove f_r è la frequenza di rotazione della macchina. Q_ψ è il carico agente sul corpo volvente ubicato alla posizione angolare ψ e noto dalla formulazione di Harris [50], ϵ è il fattore di distribuzione del carico. L'operatore di convoluzione è rappresentato dal simbolo $*$, f_{motor} è la frequenza caratteristica del motore asincrono pari a $6f_{stator}$, dove f_{stator} è la frequenza statorica di alimentazione. I parametri β, f_{struct}, A_1 e A_2 sono stati scelti utilizzando valori di primo tentativo. Il contributo di questi parametri verrà corretto dall'algoritmo generativo, che apprenderà come alterare i dati simulati per renderli coerenti con la realtà sperimentale.

$$h(t) = e^{-\beta t} \sin\left(2\pi f_{struct} t\right) \tag{5}$$

$$d(t) = \sum_k \delta(t - kT), \qquad T = 1/f_{defect} \tag{6}$$

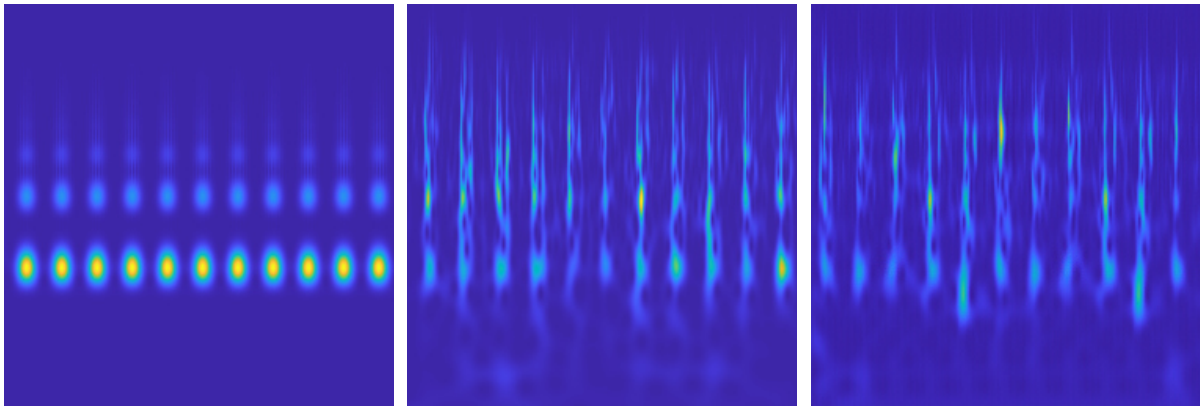
$$Q_\psi = Q_{max}\left[1 - \frac{1-\cos(\psi_i)}{2\epsilon}\right]^{\frac{10}{9}}, -\psi_m < \psi < \psi_m \tag{7}$$

$$s(t) = \left[Q_\psi d(t)\right] * h(t) + A_1 \sin(2\pi f_r t) + A_2 \sin(2\pi f_{motor} t) \tag{8}$$

3.3. Trattamento del segnale e allenamento del modello generativo cycleGAN

I segnali simulati e quelli sperimentali sono stati pre-processati per evidenziare la presenza del danneggiamento nel contenuto armonico. In particolare, le 250 coppie di campioni estratte dai segnali (Tabella 1) sono state normalizzate rispetto al valore medio e alla varianza di ciascun campione. Il processo di normalizzazione è attuato affinché i segnali simulati e reali siano descrivibili su scale compatibili. In seguito, i segnali sono stati filtrati nella banda 1400–2800 Hz ed è stato estratto l'inviluppo $env(t)$ [51]. Infine, è stata applicata la trasformata CWT riportata nell'Eq. (9), dove l'operatore $\psi^*(\cdot)$ rappresenta il complesso coniugato della wavelet Morse e a e b rappresentano rispettivamente il fattore di scala e di traslazione della trasformata wavelet. Le immagini risultanti sono ottenute mediante l'operazione $W^2(a;b)$, che permette di accentuare la traccia del passaggio del corpo volvente dal difetto all'interno dello spettro CWT.
La Fig. 6(a) presenta un esempio di spettro CWT relativo a un segnale simulato a 877 rpm, mentre la Fig. 6(b) mostra il corrispondente segnale sperimentale. Le 250 coppie di immagini sono state utilizzate per allenare un modello cycleGAN implementato in ambiente Matlab®. Il modello è stato allenato per 9 h 26 min avvalendosi di una GPU NVIDIA® T4 disponibile su HPC in *cloud* commerciale Amazon® AWS. Una volta allenato il modello, quest'ultimo è stato impiegato per generare immagini a partire da dati simulati. La Fig. 6(c) mostra un esempio di segnale generato per la condizione a 877 rpm. Si evidenziano marcate similitudini rispetto al segnale reale (Fig. 6(b)).

$$W(a;b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} env(t)\psi^*\left(\frac{t-b}{a}\right) dt \tag{9}$$

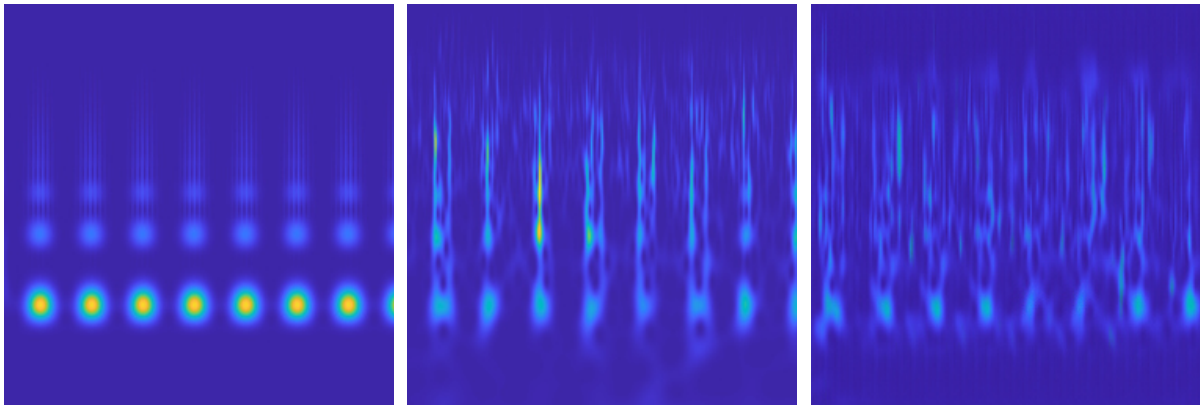


(a) (b) (c)

Figura 6. Trasformata CWT @ 877 rpm: (a) modello numerico; (b) segnale sperimentale; (c) immagine generata mediante cycleGAN.

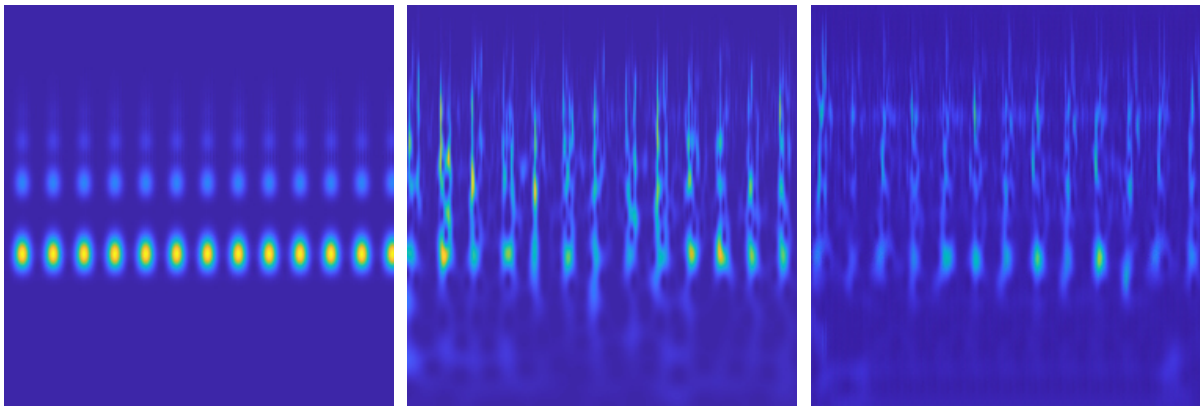
La Fig. 6 è indicativa delle capacità del modello generativo di produrre dati sintetici per condizioni di lavoro per le quali l'algoritmo è stato allenato. Questa competenza può rivelarsi preziosa per ampliare i dati per una condizione operativa per la quale siano già disponibili dei dati di danneggiamento, permettendo così di bilanciare eventuali disparità nei set di dati. Ciononostante, alcune condizioni operative, per esempio relative a differenti velocità di rotazione, potrebbero non essere disponibili dall'attività sperimentale. Sarebbe pertanto auspicabile che il modello generativo fosse in grado di generare anche alcune di queste condizioni.

Con questo proposito, è stato impiegato il modello cycleGAN per generare i dati CWT associati alle velocità di rotazione di 607 rpm e 997 rpm. La scelta di variare le condizioni operative in termini di diverse velocità di rotazione è stata presa con l'obiettivo di mettere alla prova l'algoritmo generativo, richiedendo di generare contenuti in frequenza che non fossero mai stati previamente riscontrati. Al contrario, la selezione di un carico diverso per la stessa velocità di rotazione di 877 rpm non influirebbe sulle frequenze caratteristiche del difetto, le quali sono determinate intrinsecamente da fattori cinematici. Ad ogni modo, le velocità 607 rpm e 997 rpm sono selezionate in prossimità della velocità di addestramento di 877 rpm, in modo che la banda di demodulazione compresa tra 1400 Hz e 2800 Hz conservi ancora rilevanza nell'identificare il danno nel segnale demodulato. Anche in questi casi, le Fig. 7 e 8 rivelano marcate similitudini tra segnali reali e segnali generati.



(a) (b) (c)

Figura 7. Trasformata CWT @ 607 rpm: (a) modello numerico; (b) segnale sperimentale; (c) immagine generata mediante cycleGAN.



(a) (b) (c)

Figura 8. Trasformata CWT @ 997 rpm: (a) modello numerico; (b) segnale sperimentale; (c) immagine generata mediante cycleGAN.

3.4. Validazione della metodologia generativa

La metodologia di generazione di dati sintetici proposta in questo lavoro è stata validata mediante l'utilizzo della metrica *Maximum Mean Discrepancy* (MMD) [52] e attraverso una metodologia di TL appositamente predisposta.

La MMD è una misura non-parametrica della distanza tra due distribuzioni di probabilità. Dati due set di dati $X = \{x_i\}_{i=1}^{N}$ e $Y = \{y_j\}_{j=1}^{M}$, la stima empirica della MMD è definita dall'Eq. (10), dove $\phi(\cdot)$ è il kernel caratteristico della funzione di mappatura non-lineare. Nel caso in analisi è stato adottato un kernel con funzione radiale di base (RBF) [53].

Nel contesto di questo studio, la metrica della MMD è stata utilizzata per valutare la distanza tra le distribuzioni di probabilità relative a dati simulati, dati reali e dati sintetici. Questo permette di quantificare la similitudine che emerge qualitativamente (Fig. 6, 7 e 8) tra i dati reali e quelli sintetici.

Di conseguenza, un valore di MMD inferiore indica una maggiore "similitudine" tra le distribuzioni, il che può suggerire che i dati sintetici possono essere validi sostituti dei dati reali.
Nella Tabella 2 sono riportati i valori di MMD calcolati per le varie condizioni operative. Si osserva che i dati generati tramite l'uso di cycleGAN presentano una maggiore somiglianza con i dati reali rispetto a quelli simulati. Questa affermazione è valida sia per la condizione di addestramento esplicita dell'algoritmo generativo, sia per due diverse condizioni operative.

$$MMD^2(X,Y) = \left\| \frac{1}{N}\sum_{i=1}^{N}\phi(x_i) - \frac{1}{M}\sum_{j=1}^{M}\phi(y_i) \right\|^2 \tag{10}$$

Tabella 2. MMD per differenti condizioni operative.

Distribuzioni dati	**Velocità** [rpm]		
	607	877	997
Reali−Simulati	0.187	0.218	0.202
Reali−Generati	0.106	0.069	0.073

Al fine di valutare se i dati sintetici possano essere effettivamente utilizzati come surrogati dei dati reali, sono stati condotti esperimenti di addestramento su tre differenti modelli diagnostici mediante l'approccio del TL [34]–[36]. In particolare, le reti neurali convoluzionali AlexNet [54], VGG16 [55] e ResNet18 [56], originariamente pre-allenate per il riconoscimento di immagini [57], sono state addestrate mediante *fine-tuning*. Ciò implica che gli strati iniziali e intermedi responsabili dell'estrazione delle caratteristiche discriminanti delle immagini sono rimasti invariati, mentre lo strato finale è stato sostituito e i relativi pesi sono stati riaddestrati per l'obiettivo diagnostico specifico. Questo approccio consente di capitalizzare la conoscenza pregressa di tali modelli nel riconoscimento delle immagini su un ampio dataset come ImageNet [57] e di applicare questa conoscenza generale in un contesto specifico come la diagnosi dei macchinari mediante riconoscimento delle immagini in tempo-frequenza.
Ogni modello è stato sottoposto a test utilizzando esclusivamente dati provenienti dal banco prova, ma è stato addestrato in tre diverse configurazioni: nel primo caso i dati di danneggiamento utilizzati per l'addestramento sono reali, nel secondo caso i dati di danneggiamento per l'addestramento sono stati simulati e infine, nel terzo caso, i dati di danneggiamento per l'addestramento sono stati generati in modo sintetico attraverso il modello generativo cycleGAN.
I modelli sono stati addestrati e valutati utilizzando 900 campioni di addestramento (150 campioni per ogni velocità e 2 condizioni di salute), rappresentanti il 60% del dataset, insieme a 300 campioni di validazione (50 campioni per ogni velocità e 2 condizioni di salute), pari al 20% del totale, e infine 300 campioni di test (50 campioni per ogni velocità e 2 condizioni di salute).

4. RISULTATI E DISCUSSIONE

Le analisi sono state condotte considerando le metriche di accuratezza e tasso di recupero dei test diagnostici. L'accuratezza è riportata nell'Eq. (11), mentre l'Eq. (12) formalizza il tasso di recupero. Nelle equazioni, VP indica il numero di veri positivi, corrispondenti ai dati di danneggiamento IR identificati correttamente come tali, VN è il numero di veri negativi, cioè dati di stati di integrità diagnosticati correttamente, FP è il numero di falsi positivi, ovvero dati normali erroneamente classificati come danneggiati e FN è il numero di falsi negativi, indicando i dati danneggiati erroneamente classificati come normali. Pertanto, il tasso di recupero fornisce una misura di quanti dei campioni di danneggiamento siano identificati come tali dal modello di diagnosi.

$$Accuratezza = \frac{VP + VN}{VP + VN + FP + FN} \tag{11}$$

$$Recupero = \frac{VP}{VP + FN} \tag{12}$$

I risultati di test ottenuti su dati reali per i tre modelli diagnostici sono mostrati in Fig. 9. La figura illustra le prestazioni nei casi di addestramento con dati di danneggiamento reali (indicati in rosso), simulati (in blu) e sintetici (in verde), espressi come media e deviazione standard ottenute da cinque test. Emerge che tutti i modelli dimostrano ottime prestazioni quando addestrati con dati reali, mentre falliscono nell'individuare il danneggiamento (Fig. 9(b)) quando addestrati con dati simulati. Un risultato particolarmente interessante, tuttavia, è che per due dei modelli impiegati, ovvero AlexNet e VGG16, le metriche di diagnosi tornano a mostrare valori significativi quando addestrati con dati sintetici. Le metriche ottenute utilizzando i dati sintetici per l'allenamento sono comparabili con quelle ottenute utilizzando dati reali. Ciò suggerisce che i dati sintetici possano effettivamente fungere da validi sostituti dei dati reali nell'addestramento degli algoritmi di diagnosi. La Fig. 9(c) illustra i tempi di elaborazione richiesti per l'addestramento. Si osserva che il modello AlexNet si distingue per le prestazioni migliori, mentre non emergono significative differenze tra i dataset reali, simulati o sintetici.

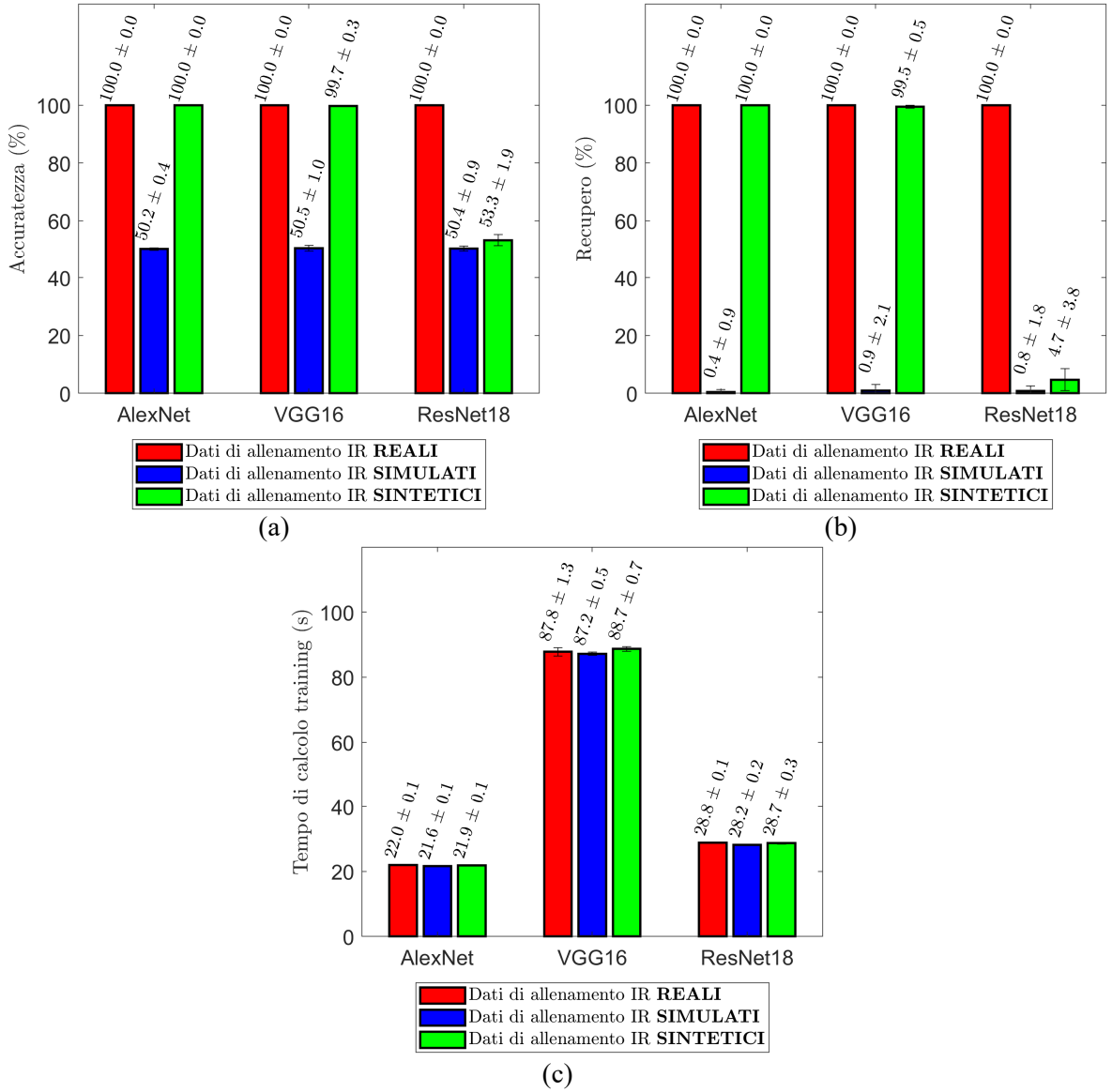


Figura 9. Risultati di test per differenti modelli diagnostici su dati reali, simulati e sintetici: (a) accuratezza; (b) recupero; (c) tempi di calcolo.

5. CONCLUSIONI

Questo studio mira a generare dati sintetici che possano sostituire dati reali relativi a danneggiamenti di macchinari rotanti per l'addestramento di algoritmi di diagnosi. A questo scopo, è stata impiegata una rete neurale generativa cycleGAN per convertire immagini wavelet appartenenti a segnali simulati nei loro equivalenti reali e riscontrabili sperimentalmente. La validità di questa metodologia generativa è stata confermata tramite l'impiego della metrica MMD.
Successivamente, sono stati addestrati diversi modelli diagnostici attraverso la tecnica del TL, utilizzando dati di danneggiamento reali, simulati o sintetici. In seguito, i modelli di diagnosi sono stati testati su dati reali ottenuti da un'apposita sperimentazione condotta su di un banco prova contenente cuscinetti industriali di media taglia. Si può concludere che:

- le reti cycleGAN sono architetture efficaci per generare dati sintetici sotto forma di immagini CWT che possono ampliare o sostituire i dati di danneggiamento di macchine rotanti;
- i dati sintetici presentano una maggiore somiglianza con i dati reali rispetto ai risultati di un modello di simulazione. Questo enunciato è applicabile sia per le condizioni operative della macchina per cui l'algoritmo generativo è stato esplicitamente addestrato sia per condizioni diverse, grazie alla presenza della coerenza ciclica nel modello generativo;
- i dati sintetici si rivelano idonei all'addestramento di algoritmi di diagnosi, che dimostrano una notevole accuratezza quando testati su dati reali effettivamente ottenuti dal macchinario rotante.

I futuri sviluppi della ricerca prevedono l'esplorazione di diverse condizioni di carico radiale e assiale, nonché varie tipologie di difetti sui macchinari. In aggiunta, si intende condurre un'analisi più approfondita utilizzando modelli di simulazione più dettagliati. Sarà anche importante condurre ulteriori studi per comprendere le ragioni per cui il modello ResNet18 non si dimostra efficace quando addestrato con dati sintetici. Infine, si prevede di investigare i limiti entro cui è possibile generare dati relativi a condizioni operative per le quali il modello generativo non sia stato specificamente addestrato.

BIBLIOGRAFIA

[1] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, e A. K. Nandi, «Applications of machine learning to machine fault diagnosis: A review and roadmap», *Mech. Syst. Signal Process.*, vol. 138, p. 106587, 2020, doi: 10.1016/j.ymssp.2019.106587.

[2] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, e J. Lin, «Machinery health prognostics: A systematic review from data acquisition to RUL prediction», *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, 2018, doi: 10.1016/j.ymssp.2017.11.016.

[3] C. Li, S. Zhang, Y. Qin, e E. Estupinan, «A systematic review of deep transfer learning for machinery fault diagnosis», *Neurocomputing*, vol. 407, pp. 121–135, set. 2020, doi: 10.1016/j.neucom.2020.04.045.

[4] R. Liu, B. Yang, E. Zio, e X. Chen, «Artificial intelligence for fault diagnosis of rotating machinery: A review», *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, 2018, doi: 10.1016/j.ymssp.2018.02.016.

[5] A. R. Mohanty, *Machinery condition monitoring: Principles and practices*. CRC Press, 2014.

[6] M. Alabsi, Y. Liao, e A. A. Nabulsi, «Bearing fault diagnosis using deep learning techniques coupled with handcrafted feature extraction: A comparative study», *JVCJournal Vib. Control*, vol. 27, fasc. 3–4, pp. 404–414, feb. 2021, doi: 10.1177/1077546320929141.

[7] S. Abbasion, A. Rafsanjani, A. Farshidianfar, e N. Irani, «Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine», *Mech. Syst. Signal Process.*, vol. 21, fasc. 7, pp. 2933–2945, 2007, doi: 10.1016/j.ymssp.2007.02.003.

[8] E. Brusa, C. Delprete, e L. G. Di Maggio, «Eigen-spectrograms: An interpretable feature space for bearing fault diagnosis based on artificial intelligence and image processing», *Mech. Adv. Mater. Struct.*, pp. 1–13, lug. 2022, doi: 10.1080/15376494.2022.2102274.

[9] G. Genta, *Dynamics of rotating systems*. Springer Science & Business Media, 2007.

[10] R. B. Randall, *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. John Wiley & Sons, 2011. doi: 10.1007/978-94-007-6422-4_11.

[11] R. B. Randall e J. Antoni, «Rolling element bearing diagnostics-A tutorial», *Mech. Syst. Signal Process.*, vol. 25, fasc. 2, pp. 485–520, 2011, doi: 10.1016/j.ymssp.2010.07.017.

[12] L. M. R. Baccarini, V. V. Rocha e Silva, B. R. de Menezes, e W. M. Caminhas, «SVM practical industrial application for mechanical faults diagnostic», *Expert Syst. Appl.*, vol. 38, fasc. 6, pp. 6980–6984, giu. 2011, doi: 10.1016/J.ESWA.2010.12.017.

[13] C. Cortes e V. Vapnik, «Support-vector networks», *Mach. Learn.*, vol. 20, fasc. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[14] A. Widodo e B.-S. Yang, «Support vector machine in machine condition monitoring and fault diagnosis», *Mech. Syst. Signal Process.*, vol. 21, fasc. 6, pp. 2560–2574, ago. 2007, doi: 10.1016/J.YMSSP.2006.12.007.

[15] T. Cover e P. Hart, «Nearest neighbor pattern classification», *IEEE Trans. Inf. Theory*, vol. 13, fasc. 1, pp. 21–27, 1967, doi: 10.1007/978-0-387-35973-1_862.

[16] Y. Lei e M. J. Zuo, «Gear crack level identification based on weighted K nearest neighbor classification algorithm», *Mech. Syst. Signal Process.*, vol. 23, fasc. 5, pp. 1535–1547, lug. 2009, doi: 10.1016/j.ymssp.2009.01.009.

[17] A. Moosavian, H. Ahmadi, A. Tabatabaeefar, e M. Khazaee, «Comparison of Two Classifiers; K-Nearest Neighbor and Artificial Neural Network, for Fault Diagnosis on a Main Engine Journal-Bearing», *Shock Vib.*, vol. 20, fasc. 2, pp. 263–272, 2013, doi: 10.1155/2013/360236.

[18] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, e R. X. Gao, «Deep learning and its applications to machine health monitoring», *Mech. Syst. Signal Process.*, 2019, doi: 10.1016/j.ymssp.2018.05.050.

[19] L. Wen, X. Li, L. Gao, e Y. Zhang, «A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method», *IEEE Trans. Ind. Electron.*, vol. 65, fasc. 7, pp. 5990–5998, 2018, doi: 10.1109/TIE.2017.2774777.

[20] J. Grezmak, J. Zhang, P. Wang, K. A. Loparo, e R. X. Gao, «Interpretable Convolutional Neural Network Through Layer-wise Relevance Propagation for Machine Fault Diagnosis», *IEEE Sens. J.*, vol. 20, fasc. 6, pp. 3172–3181, mar. 2020, doi: 10.1109/jsen.2019.2958787.

[21] A. Duan, L. Guo, H. Gao, X. Wu, e X. Dong, «Deep Focus Parallel Convolutional Neural Network for Imbalanced Classification of Machinery Fault Diagnostics», *IEEE Trans. Instrum. Meas.*, vol. 69, fasc. 11, pp. 8680–8689, nov. 2020, doi: 10.1109/tim.2020.2998233.

[22] Q. Xiao, S. Li, L. Zhou, e W. Shi, «Improved Variational Mode Decomposition and CNN for Intelligent Rotating Machinery Fault Diagnosis», *Entropy*, vol. 24, fasc. 7, p. 908, giu. 2022, doi: 10.3390/e24070908.

[23] L. Alzubaidi *et al.*, «Review of deep learning: concepts, CNN architectures, challenges, applications, future directions», *J. Big Data*, vol. 8, fasc. 1, p. 53, mar. 2021, doi: 10.1186/s40537-021-00444-8.

[24] S. L. Brunton e J. N. Kutz, *Data Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2019.

[25] «CWRU Bearing Data Center». https://engineering.case.edu/bearingdatacenter (consultato 3 agosto 2020).

[26] W. A. Smith e R. B. Randall, «Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study», *Mech. Syst. Signal Process.*, vol. 64–65, pp. 100–131, 2015, doi: 10.1016/j.ymssp.2015.04.021.

[27] P. Nectoux *et al.*, «PRONOSTIA : An experimental platform for bearings accelerated degradation tests», in *IEEE International Conference on Prognostics and Health Management*, Denver, CO, USA, 2012.

[28] J. Lee, H. Qiu, G. Yu, e J. Lin, «Bearing data set», *IMS Univ. Cincinnati NASA Ames Progn. Data Repos. Rexnord Tech. Serv.*, 2007.

[29] S. J. Pan e Q. Yang, «A Survey on Transfer Learning», *IEEE Trans. Knowl. Data Eng.*, vol. 22, fasc. 10, pp. 1345–1359, 2010, doi: 10.1109/TKDE.2009.191.

[30] N. Cao, Z. Jiang, J. Gao, e B. Cui, «Bearing state recognition method based on transfer learning under different working conditions», *Sensors*, vol. 20, fasc. 1, p. 234, 2019, doi: 10.3390/s20010234.

[31] F. Pacheco, A. Drimus, L. Duggen, M. Cerrada, D. Cabrera, e R.-V. Sanchez, «Deep Ensemble-Based Classifier for Transfer Learning in Rotating Machinery Fault Diagnosis», *IEEE Access*, vol. 10, pp. 29778–29787, 2022, doi: 10.1109/access.2022.3158023.

[32] L. Guo, Y. Lei, S. Xing, T. Yan, e N. Li, «Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines with Unlabeled Data», *IEEE Trans. Ind. Electron.*, vol. 66, fasc. 9, pp. 7316–7325, 2019, doi: 10.1109/TIE.2018.2877090.

[33] S. Shao, S. McAleer, R. Yan, e P. Baldi, «Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning», *IEEE Trans. Ind. Inform.*, vol. 15, fasc. 4, pp. 2446–2455, 2019, doi: 10.1109/TII.2018.2864759.

[34] E. Brusa, C. Delprete, e L. G. Di Maggio, «Deep Transfer Learning for Machine Diagnosis : From Sound and Music Recognition to Bearing Fault Detection», *Appl. Sci. Switz.*, vol. 11, fasc. 24, 2021, doi: 10.3390/app112411663.

[35] L. G. Di Maggio, «Intelligent Fault Diagnosis of Industrial Bearings Using Transfer Learning and CNNs Pre-Trained for Audio Classification», *Sensors*, vol. 23, fasc. 1, p. 211, dic. 2022, doi: 10.3390/s23010211.

[36] L. G. Di Maggio, «Applicazione di algoritmi di Intelligenza Artificiale per il monitoraggio di cuscinetti volventi», in *Atti del 51° Convegno Nazionale AIAS*, Padova, 7-10/9/2022.

[37] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, e A. A. Bharath, «Generative adversarial networks: An overview», *IEEE Signal Process. Mag.*, vol. 35, fasc. 1, pp. 53–65, 2018, doi: 10.1109/MSP.2017.2765202.

[38] I. Goodfellow *et al.*, «Generative adversarial nets», in *Proceedings of the 27th international conference on neural information processing systems*, Cambridge, MA, USA: MIT Press, 2014, pp. 2672–2680.

[39] S. Cao, L. Wen, X. Li, e L. Gao, «Application of Generative Adversarial Networks for Intelligent Fault Diagnosis», in *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, IEEE, ago. 2018. doi: 10.1109/coase.2018.8560528.

[40] Y. Ding, L. Ma, J. Ma, C. Wang, e C. Lu, «A Generative Adversarial Network-Based Intelligent Fault Diagnosis Method for Rotating Machinery Under Small Sample Size Conditions», *IEEE Access*, vol. 7, pp. 149736–149749, 2019, doi: 10.1109/access.2019.2947194.

[41] X. Li, W. Zhang, e Q. Ding, «Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks», *IEEE Trans. Ind. Electron.*, vol. 66, fasc. 7, pp. 5525–5534, 2018, doi: 10.1109/TIE.2018.2868023.

[42] E. Brusa, C. Delprete, L. Giorio, L. G. Di Maggio, e V. Zanella, «Design of an Innovative Test Rig for Industrial Bearing Monitoring with Self-Balancing Layout», *Machines*, vol. 10, fasc. 54, 2022, doi: 10.3390/machines10010054.

[43] J. F. Nash, «Equilibrium points in *n*-person games», *Proc. Natl. Acad. Sci.*, vol. 36, fasc. 1, pp. 48–49, 1950, doi: 10.1073/pnas.36.1.48.

[44] I. Goodfellow, «NIPS 2016 Tutorial: Generative Adversarial Networks». arXiv, 3 aprile 2017. Consultato: 17 gennaio 2023. [Online]. Disponibile su: http://arxiv.org/abs/1701.00160

[45] J.-Y. Zhu, T. Park, P. Isola, e A. A. Efros, «Unpaired image-to-image translation using cycle-consistent adversarial networks», in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232. doi: 10.1109/ICCV.2017.244.

[46] P. D. McFadden e J. D. Smith, «Model for the vibration produced by a single point defect in a rolling element bearing», *J. Sound Vib.*, vol. 96, fasc. 1, pp. 69–82, 1984, doi: 10.1016/0022-460X(84)90595-9.

[47] Y. Wang e M. Liang, «An adaptive SK technique and its application for fault detection of rolling element bearings», *Mech. Syst. Signal Process.*, vol. 25, fasc. 5, pp. 1750–1764, 2011, doi: 10.1016/j.ymssp.2010.12.008.

[48] S. Ericsson, N. Grip, E. Johansson, L. E. Persson, R. Sjöberg, e J. O. Strömberg, «Towards automatic detection of local bearing defects in rotating machines», *Mech. Syst. Signal Process.*, vol. 19, fasc. 3, pp. 509–535, 2005, doi: 10.1016/j.ymssp.2003.12.004.

[49] C. Sobie, C. Freitas, e M. Nicolai, «Simulation-driven machine learning: Bearing fault classification», *Mech. Syst. Signal Process.*, vol. 99, pp. 403–419, 2018, doi: 10.1016/j.ymssp.2017.06.025.

[50] T. A. Harris, *Rolling bearing analysis*. John Wiley & Sons, 2001.

[51] E. Brusa, F. Bruzzone, C. Delprete, L. Di Maggio, e C. Rosso, «Envelope analysis applied to non-Hertzian contact simulations in damaged roller bearings», in *IOP Conference Series: Materials Science and Engineering*, 2021, p. 012013. doi: 10.1088/1757-899x/1038/1/012013.

[52] D. Garreau, W. Jitkrittum, e M. Kanagawa, «Large sample analysis of the median heuristic». 2017. doi: 10.48550/arXiv.1707.07269.

[53] K. Fukumizu, A. Gretton, X. Sun, e B. Schölkopf, «Kernel measures of conditional dependence», *Adv. Neural Inf. Process. Syst.*, vol. 20, 2007.

[54] A. Krizhevsky, I. Sutskever, e G. E. Hinton, «Imagenet classification with deep convolutional neural networks», *Adv. Neural Inf. Process. Syst.*, vol. 25, pp. 1097–1105, 2012, doi: 10.1145/3065386.

[55] K. Simonyan e A. Zisserman, «Very Deep Convolutional Networks for Large-Scale Image Recognition», 2014, doi: 10.48550/ARXIV.1409.1556.

[56] K. He, X. Zhang, S. Ren, e J. Sun, «Deep Residual Learning for Image Recognition», in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, giu. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[57] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, e Li Fei-Fei, «ImageNet: A large-scale hierarchical image database», *2009 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 248–255, 2010, doi: 10.1109/cvpr.2009.5206848.