

METODO E SISTEMA PER PREVEDERE IL FALLIMENTO DI UN'ENTITÀ MONITORATA

Original

METODO E SISTEMA PER PREVEDERE IL FALLIMENTO DI UN'ENTITÀ MONITORATA / Perboli, Guido. - (2021).

Availability:

This version is available at: 11583/2867715.2 since: 2021-01-26T16:48:42Z

Publisher:

Published

DOI:

Terms of use:

openAccess

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)



Ministero dello Sviluppo Economico

Ricevuta di presentazione

per

Brevetto per invenzione industriale



Domanda numero: 102021000001424

Data di presentazione: 26/01/2021

DATI IDENTIFICATIVI DEL DEPOSITO

Ruolo	Mandatario
Depositante	Antonio Di Bernardo
Data di compilazione	26/01/2021
Riferimento depositante	P1335IT00
Titolo	METODO E SISTEMA PER PREVEDERE IL FALLIMENTO DI UN'ENTITÀ MONITORATA
Carattere domanda	Ordinaria
Esenzione	NO
Accessibilità al pubblico	NO
Numero rivendicazioni	10
Autorità depositaria	

PRIVACY

Autorizzo il trattamento dei dati personali, inseriti all'interno del deposito, ai sensi del GDPR (Regolamento UE 2016/679) e del Decreto Legislativo 30 giugno 2003, n. 196 "Codice in materia di protezione dei dati personali"

RICHIEDENTE/I

Natura giuridica	Persona giuridica
Denominazione	ARISK S.r.l.
P.IVA/CF	10061420963
Tipo Società	societa' a responsabilita' limitata
Nazione sede legale	Italia
Comune sede legale	Milano (MI)
Indirizzo	Via Giuseppe Revere
Civico	16

CAP	20123
Telefono	
Fax	
Email	
Pec	
Quota percentuale	100.0%

DOMICILIO ELETTIVO

Cognome/R.sociale	Thinx S.r.l.
Indirizzo	P.le Luigi Cadorna 10
Cap	20123
Nazione	Italia
Comune	Milano (MI)
Telefono	02 - 82783531
Fax	02 - 21113254
Email\PEC	thinx@legalmail.it

MANDATARI/RAPPRESENTANTI

Cognome	Nome
Di Bernardo	Antonio
Belloni	Giancarlo
Caruti	Filippo
Filippini	Deborah
Mati	Silvia

INVENTORI

Cognome	Nome	Nazione residenza
Perboli	Guido	Italia

CLASSIFICAZIONI

Sezione	Classe	Sottoclasse	Gruppo	Sottogruppo
---------	--------	-------------	--------	-------------

NUMERO DOMANDE COLLEGATE

DOCUMENTAZIONE ALLEGATA

Tipo documento	Riserva	Documento
Disegni	NO	P1335IT00_20200126 Tavole disegno.pdf.p7m hash: e815fa9080de4eb101a6a899f304dd4f
Lettera di Incarico	NO	P1335IT00_20210125lett incarico firmata.pdf.p7m hash: c81b4e9086c9ac79c68bee3f6f36ebb
Riassunto	NO	P1335IT00_20200126 riassunto.pdf.p7m hash: 0a3a6522c1e2167ee6feca7c8d9482a8
Rivendicazioni	NO	P1335IT00_20200126 rivendicazioni.pdf.p7m hash: 950b6b7cfb01e7fc4979314de02ef781
Rivendicazioni in inglese	NO	P1335IT00_20200126 Rivendicazioni_EN.pdf.p7m hash: 0efd467ec5874d971fe760168bf6ce2e
Descrizione in italiano*	NO	P1335IT00_20200126 descrizione.pdf.p7m hash: c730d2d202bed31f217b798441497e46

PAGAMENTI

Tipo	Identificativo	Data
Bollo	01200691981509	25/01/2021

PAGAMENTI PAGO PA

Codice IUUV	UIBM_IUV_20213159839
Tipologia	Tasse/Diritti Primo deposito

DOVUTO

Gli importi indicati non tengono conto delle eventuali esenzioni applicabili

Importo Tasse:

€ 50,00

Importo Imposta Bollo:

€ 20,00

NOTE

**METODO E SISTEMA PER PREVEDERE IL FALLIMENTO DI UN'ENTITÀ
MONITORATA**

RIASSUNTO

5 La presente invenzione riguarda un metodo (100, 200) per predire un malfunzionamento di un'entità (4). In particolare, l'entità (4) è caratterizzata da una pluralità di parametri.

Il metodo (100, 200) prevede di predire (200) un malfunzionamento dell'entità (4) mediante un sistema di intelligenza artificiale (1) cui sono forniti dati relativi ai parametri dell'entità (4), in cui il sistema di intelligenza artificiale (1) è allenato
10 mediante dei dati di allenamento.

Vantaggiosamente, i dati di allenamento sono selezionati secondo i seguenti passi:

- a. acquisire (101-106) un insieme di dati di allenamento, laddove l'insieme di dati di allenamento comprende una pluralità di gruppi di dati di entità,
15 ciascun gruppo di dati di entità comprendendo valori riferiti a parametri caratteristici di una rispettiva entità di allenamento,
- b. selezionare (107) dati di entità compresi in gruppi di dati di entità di un primo sottoinsieme di entità di allenamento dell'insieme di dati di allenamento,
- c. allenare (109) il sistema di intelligenza artificiale mediante i dati di entità del
20 primo sottoinsieme di entità di allenamento selezionati,
- d. determinare (112) una distribuzione delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di dati di allenamento, eseguendo il sistema di intelligenza artificiale (1) allenato con i dati di entità del primo sottoinsieme di entità di allenamento,
- 25 e. determinare (113) i dati di allenamento selezionando dati di entità di entità di allenamento di un secondo sottoinsieme di entità di allenamento dell'insieme di dati di allenamento in modo tale che il secondo sottoinsieme di entità di allenamento presenti una distribuzione delle probabilità di malfunzionamento uguale a quella calcolata al punto d.

RIVENDICAZIONI

1. Metodo (100, 200) per predire un malfunzionamento di un'entità (4), detta entità (4) essendo caratterizzata da una pluralità di parametri, il metodo (100, 200) prevedendo di:
 - 5 predire (200) un malfunzionamento dell'entità (4) mediante un sistema di intelligenza artificiale (1) cui sono forniti dati relativi ai parametri dell'entità (4), in cui il sistema di intelligenza artificiale (1) è allenato mediante dei dati di allenamento caratterizzato dal fatto che
 - 10 i dati di allenamento sono selezionati secondo i seguenti passi:
 - a. acquisire (101-106) un insieme di dati di allenamento, laddove l'insieme di dati di allenamento comprende una pluralità di gruppi di dati di entità, ciascun gruppo di dati di entità comprendendo valori riferiti a parametri caratteristici di una rispettiva entità di allenamento,
 - 15 b. selezionare (107) dati di entità compresi in gruppi di dati di entità di un primo sottoinsieme di entità di allenamento dell'insieme di dati di allenamento,
 - c. allenare (109) il sistema di intelligenza artificiale mediante i dati di entità del primo sottoinsieme di entità di allenamento selezionati,
 - d. determinare (112) una distribuzione delle probabilità di malfunzionamento
 - 20 delle entità di allenamento dell'insieme di dati di allenamento, eseguendo il sistema di intelligenza artificiale (1) allenato con i dati di entità del primo sottoinsieme di entità di allenamento,
 - e. determinare (113) i dati di allenamento selezionando dati di entità di entità di allenamento appartenenti a un secondo sottoinsieme di entità di allenamento
 - 25 dell'insieme di dati di allenamento in modo tale che il secondo sottoinsieme di entità di allenamento presenti una distribuzione delle probabilità di malfunzionamento uguale a quella calcolata al punto d.
2. Metodo (100; 200) secondo la rivendicazione 1, ulteriormente comprendente il passo di:
 - 30 - determinare un primo gruppo di entità di allenamento dell'insieme dati di allenamento che hanno subito un malfunzionamento e un secondo gruppo di entità di allenamento dell'insieme dati di allenamento che non hanno subito un malfunzionamento, laddove il secondo gruppo è molto maggiore del secondo gruppo:
 - 35 il passo di determinare (112) una distribuzione delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di dati di allenamento prevede di:

- determinare la distribuzione delle probabilità di malfunzionamento delle entità di allenamento del secondo gruppo, e
in cui il passo di determinare (113) i dati di allenamento selezionando dati di entità di un secondo sottoinsieme di entità di allenamento dell'insieme di dati di allenamento prevede di:
- 5
- selezionare un terzo sottoinsieme di entità di allenamento del secondo gruppo in modo tale che il terzo sottoinsieme di entità di allenamento presenti una distribuzione delle probabilità di malfunzionamento uguale a quella calcolata al punto d, e
- 10
- definire il secondo sottoinsieme di entità di allenamento dell'insieme di dati di allenamento come unione delle entità di allenamento del primo gruppo e del terzo sottoinsieme.
3. Metodo (100; 200) secondo la rivendicazione 1 o 2, in cui il passo di determinare (112) una distribuzione delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di allenamento prevede di:
- 15
- i. definire una pluralità di valori di soglia di probabilità;
 - ii. determinare almeno uno tra un numero e una lista di entità di allenamento dell'insieme di dati di allenamento la cui probabilità di malfunzionamento è minore o uguale a un valore di soglia selezionato, eseguendo il sistema di intelligenza artificiale,
- 20
- iii. selezionare un nuovo valore di soglia di probabilità di detta pluralità di valori di soglia di probabilità,
 - iv. iterare i passi i - iii, fino a quando sono stati selezionati tutti i valori di soglia di probabilità, e
- 25
- v. calcolare la distribuzione delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di allenamento sulla base dei risultati forniti ad ogni iterazione del passo ii.
4. Metodo (100; 200) secondo la rivendicazione 3, ulteriormente prevede i passi di:
- 30
- ii_a. selezionare dati di entità di un sottoinsieme intermedio di entità di allenamento, detto sottoinsieme intermedio di entità di allenamento comprendendo entità comprese nella lista di entità di allenamento determinata al passo ii ed entità di allenamento dell'insieme di allenamento che hanno subito un malfunzionamento;
- 35
- ii_b. allenare il sistema di intelligenza artificiale mediante i dati di entità del sottoinsieme intermedio di entità, prima di eseguire i passi iii - v.

5. Metodo (100; 200) secondo la rivendicazione 1 o 2, in cui il passo di determinare (112) una distribuzione delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di allenamento prevede di:
- calcolare una probabilità di malfunzionamento per ciascuna entità di allenamento dell'insieme di dati di allenamento, eseguendo il sistema di intelligenza artificiale, e
 - combinare le probabilità di malfunzionamento calcolate per determinare la distribuzione delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di dati di allenamento.
6. Metodo (100; 200) secondo una qualsiasi delle rivendicazioni precedenti, in cui il passo di determinare (112) una distribuzione delle probabilità di malfunzionamento prevede di:
- calcolare una pluralità di distribuzioni delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di allenamento, laddove ciascuna distribuzione delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di allenamento è calcolata per mezzo di gruppi di dati delle entità di allenamento riferiti a un medesimo intervallo di tempo, e
- in cui il passo di determinare (113) i dati di allenamento prevede di:
- selezionare i dati di entità di un secondo sottoinsieme di entità di allenamento dell'insieme di dati di allenamento in modo tale che il secondo sottoinsieme di entità di allenamento presenti una pluralità di distribuzioni delle probabilità di malfunzionamento uguale alla pluralità di distribuzioni delle probabilità di malfunzionamento calcolata per le entità di allenamento dell'insieme di dati di allenamento.
7. Metodo (100; 200) secondo una qualsiasi delle rivendicazioni precedenti, in cui il passo di selezionare (107) dati di entità di un primo sottoinsieme di entità di allenamento dell'insieme di dati di allenamento prevede di:
- determinare un primo gruppo di entità di allenamento dell'insieme dati di allenamento che hanno subito un malfunzionamento;
 - campionare casualmente un secondo gruppo di entità di allenamento dell'insieme dati di allenamento che non hanno subito un malfunzionamento, laddove il secondo gruppo comprende un numero di entità di allenamento compreso tra 1 e 5 volte un numero di entità di allenamento del primo gruppo, e
 - comporre il primo sottoinsieme di entità di allenamento dell'insieme di dati di allenamento come unione delle entità di allenamento del primo gruppo e del secondo gruppo.

8. Metodo (100; 200) secondo una qualsiasi delle rivendicazioni precedenti, in cui il passo di predire (200) un malfunzionamento dell'entità (4) mediante un sistema di intelligenza artificiale (1) prevede di utilizzare un algoritmo di apprendimento automatico, selezionato tra:
- 5 - Random Forest,
- XGBoost;
- Logistic Regression, e
- una rete neurale.
9. Metodo (100;200) secondo una qualsiasi delle rivendicazioni precedenti, in cui il passo di predire (200) un malfunzionamento dell'entità (4) mediante un sistema di intelligenza artificiale (1) prevede di:
- 10 - calcolare una pluralità di probabilità di malfunzionamento dell'entità (4), ciascuna probabilità di malfunzionamento dell'entità (4) essendo riferita a un rispettivo intervallo di tempo futuro.
- 15 10. Sistema (1) per predire un malfunzionamento di un'entità (4), detta entità (4) essendo caratterizzata da una pluralità di parametri, il sistema (1) comprendendo:
- un modulo di interfaccia (11) configurato per ricevere dati relativi a detta entità (4) e fornire un risultato di predizione di malfunzionamento;
- 20 - un modulo di elaborazione dati (12) configurato per acquisire ed elaborare dati relativi a entità di allenamento e i dati relativi a detta entità (4) forniti al modulo di interfaccia (11), e
- un modulo di intelligenza artificiale (13) configurato per eseguire un algoritmo di intelligenza artificiale che utilizza i dati elaborati dal modulo di
- 25 elaborazione dati (12) per predire una probabilità di malfunzionamento dell'entità (4),
laddove i moduli (11, 12, 13) del sistema (1) sono configurati per implementare il metodo secondo una qualsiasi delle rivendicazioni precedenti.

METODO E SISTEMA PER PREVEDERE IL FALLIMENTO DI UN'ENTITÀ MONITORATA

DESCRIZIONE

CAMPO TECNICO

- 5 La presente invenzione si riferisce al settore dei sistemi informatici. In dettaglio, la presente invenzione riguarda un metodo e un sistema per la predizione di un fallimento di un'entità monitorata.

STATO DELL'ARTE

- 10 In vari settori di attività sono implementati sistemi di grande complessità, ciascuno suddiviso in una pluralità di unità a loro volta suddivise in una pluralità di componenti. Esempi di tali sistemi complessi comprendono impianti di produzione industriale, sistemi di trasporto pubblico, reti di telecomunicazione, reti informatiche, data center, ecc.

- 15 La complessità di questi sistemi rende estremamente difficile e laborioso il controllo e la verifica dello stato delle varie unità e componenti che costituiscono ciascun sistema. Di conseguenza, sono spesso implementate operazioni di manutenzione periodica a prescindere dall'effettiva necessità della stessa e/o i sistemi sono implementati con una considerevole ridondanza per evitare che il malfunzionamento di un componente critico comprometta l'intero sistema.
- 20 Come è evidente tali soluzioni, limitano solo parzialmente gli svantaggi menzionati e portano a incrementano sostanziale del costo di produzione ed esercizio di questi sistemi, così come la complessità degli stessi.

- 25 Nel tentativo di risolvere tali problemi nell'arte sono stati proposti metodi basati su un'analisi automatica dei sistemi complessi, in particolare basata su intelligenza artificiale, al fine di predire un possibile malfunzionamento degli stessi.

- Per esempio, WO 2019/125445 e US 6,125,105 propongono metodi basati su un'intelligenza artificiale destinati a prevedere malfunzionamenti in una rete di sistemi informatici, in particolare per quanto riguarda lo scambio di dati tra
- 30 elementi della rete.

US 2013/132001 e US 2012/025997 propongono invece metodi basati su intelligenza artificiale destinati a prevedere il malfunzionamento di apparati meccanici o elettromeccanici.

Inoltre, sono noti metodi basati su intelligenza artificiale destinati a predire situazioni di malfunzionamento di sistemi complessi non solamente influenzati da fattori meccanici e/o elettrici, ma anche fattori economici/sociali. Per esempio, JP 2020/095583 descrive un metodo basato su intelligenza artificiale
5 destinato a fornire una previsione della probabilità bancarotta di un'azienda.

La Richiedente ha osservato che i metodi proposti, per quanto basati su algoritmi di intelligenza artificiale robusti, sono fortemente influenzati dalla composizione dai dati utilizzati durante l'allenamento degli algoritmi di intelligenza artificiale.

In particolare, la Richiedente ha rilevato che le tecniche di elaborazioni dei dati
10 di allenamento note non permettono di definire un insieme di dati di allenamento ottimale, in special modo nel caso di grandi insiemi di dati non uniformi, con un conseguente effetto negativo sulla qualità delle predizioni di funzionamento generate dai sistemi noti.

SCOPI E RIASSUNTO DELL'INVENZIONE

15 È scopo della presente invenzione quello di superare gli inconvenienti dell'arte nota.

In particolare, è scopo della presente invenzione fornire un metodo e un relativo sistema di predizione di un malfunzionamento ottimizzato.

20 Un ulteriore scopo della presente invenzione è quello di proporre un metodo e sistema di predizione di un malfunzionamento in grado di fornire predizioni di malfunzionamento affidabili su una pluralità di periodi di tempi futuri.

25 Un ulteriore scopo della presente invenzione è quello di proporre un metodo e sistema di predizione di un malfunzionamento in grado di creare in modo automatico un insieme di dati di allenamento ottimale a partire da un insieme di dati disponibili.

Questi ed altri scopi della presente invenzione sono raggiunti mediante un sistema incorporante le caratteristiche delle rivendicazioni allegate, le quali formano parte integrante della presente descrizione.

30 Secondo un primo aspetto, la presente invenzione è diretta a un metodo per predire un malfunzionamento di un'entità. In particolare, l'entità è caratterizzata da una pluralità di parametri.

Il metodo prevede di predire un malfunzionamento dell'entità mediante un

sistema di intelligenza artificiale cui sono forniti dati relativi ai parametri dell'entità, in cui il sistema di intelligenza artificiale è allenato mediante dei dati di allenamento.

Vantaggiosamente, i dati di allenamento sono selezionati secondo i seguenti
5 passi:

- a. acquisire un insieme di dati di allenamento, laddove l'insieme di dati di allenamento comprende una pluralità di gruppi di dati di entità, ciascun gruppo di dati di entità comprendendo valori riferiti a parametri caratteristici di una rispettiva entità di allenamento;
- 10 b. selezionare dati di entità compresi in gruppi di dati di entità di un primo sottoinsieme di entità di allenamento dell'insieme di dati di allenamento;
- c. allenare il sistema di intelligenza artificiale mediante i dati di entità del primo sottoinsieme di entità di allenamento selezionati;
- d. determinare una distribuzione delle probabilità di malfunzionamento delle
15 entità di allenamento dell'insieme di dati di allenamento, eseguendo il sistema di intelligenza artificiale allenato con i dati di entità del primo sottoinsieme di entità di allenamento, e
- e. determinare i dati di allenamento selezionando dati di entità di entità di allenamento appartenenti a un secondo sottoinsieme di entità di allenamento
20 dell'insieme di dati di allenamento in modo tale che il secondo sottoinsieme di entità di allenamento presenti una distribuzione delle probabilità di malfunzionamento uguale a quella calcolata al punto d.

Nella presente l'espressione "entità di allenamento dell'insieme di dati di allenamento" indica una (o più, secondo i casi) entità di allenamento di cui
25 almeno un rispettivo gruppo di dati di entità di allenamento sia compreso nell'insieme di dati di allenamento.

Grazie a questa soluzione è possibile ottenere predizioni di malfunzionamento particolarmente precise anche quando l'insieme dati di allenamento utilizzato per l'allenamento dell'intelligenza artificiale è altamente sbilanciato. In
30 particolare, è lo stesso sistema di intelligenza artificiale che crea i dati di allenamento ottimali in modo completamente automatico.

In una forma di realizzazione, il metodo ulteriormente comprende il passo di:
- determinare un primo gruppo di entità di allenamento dell'insieme dati di allenamento che hanno subito un malfunzionamento e un secondo gruppo di
35 entità di allenamento dell'insieme dati di allenamento che non hanno subito un malfunzionamento.

Quando il secondo gruppo è molto maggiore del primo gruppo il passo di determinare una distribuzione delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di dati di allenamento prevede di determinare la distribuzione delle probabilità di malfunzionamento delle entità di allenamento del secondo gruppo.

5
Inoltre, il passo di determinare i dati di allenamento selezionando dati di entità di un secondo sottoinsieme di entità di allenamento dell'insieme di dati di allenamento prevede di:

10 - selezionare un terzo sottoinsieme di entità di allenamento del secondo gruppo in modo tale che il terzo sottoinsieme di entità di allenamento presenti una distribuzione delle probabilità di malfunzionamento uguale a quella calcolata al punto d, e

15 - definire il secondo sottoinsieme di entità di allenamento dell'insieme di dati di allenamento come unione delle entità di allenamento del primo gruppo e del terzo sottoinsieme.

Nella presente, l'espressione "molto maggiore" indica una proporzione uguale o maggiore di 1:5.

20 In questo modo è possibile ottenere dati di allenamento bilanciati a partire da un insieme di dati di allenamento fortemente sbilanciato in modo rapido ed efficiente.

In una forma di realizzazione, il passo di determinare una distribuzione delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di allenamento prevede di:

- 25
- i. definire una pluralità di valori di soglia di probabilità;
 - ii. determinare almeno uno tra un numero e una lista di entità di allenamento dell'insieme di dati di allenamento la cui probabilità di malfunzionamento è minore o uguale a un valore di soglia selezionato, eseguendo il sistema di intelligenza artificiale,
 - 30 iii. selezionare un nuovo valore di soglia di probabilità di detta pluralità di valori di soglia di probabilità,
 - iv. iterare i passi i - iii, fino a quando sono stati selezionati tutti i valori di soglia di probabilità, e
 - v. calcolare la distribuzione delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di allenamento sulla base dei
 - 35 risultati forniti ad ogni iterazione del passo ii.

Preferibilmente, sono previsti anche i passi di prevede i passi di:

- ii_a. selezionare dati di entità di un sottoinsieme intermedio di entità di allenamento, detto sottoinsieme intermedio di entità di allenamento comprendendo entità comprese nella lista di entità di allenamento determinata al passo ii ed entità di allenamento dell'insieme di allenamento che hanno subito un malfunzionamento;
- 5
- ii_b. allenare il sistema di intelligenza artificiale mediante i dati di entità del sottoinsieme intermedio di entità, prima di eseguire i passi iii - v.

10 In una forma di realizzazione, il passo di determinare una distribuzione delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di allenamento prevede di:

- calcolare una probabilità di malfunzionamento per ciascuna entità di allenamento dell'insieme di dati di allenamento, eseguendo il sistema di intelligenza artificiale, e
- 15 - combinare le probabilità di malfunzionamento calcolate per determinare la distribuzione delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di dati di allenamento.

In una forma di realizzazione, il passo di determinare una distribuzione delle probabilità di malfunzionamento prevede di:

- 20 - calcolare una pluralità di distribuzioni delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di allenamento, laddove ciascuna distribuzione delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di allenamento è calcolata per mezzo di gruppi di dati delle entità di allenamento riferiti a un medesimo intervallo di tempo.

25 Inoltre, il passo di determinare i dati di allenamento prevede di:

- selezionare i dati di entità di un secondo sottoinsieme di entità di allenamento dell'insieme di dati di allenamento in modo tale che il secondo sottoinsieme di entità di allenamento presenti una pluralità di distribuzioni delle probabilità di malfunzionamento uguale alla pluralità di distribuzioni delle probabilità di malfunzionamento calcolata per le entità di allenamento dell'insieme di dati di allenamento.
- 30

La Richiedente ha determinato che le metodologie sopra proposte rendono possibile determinare in modo efficace la distribuzione di probabilità fallimento dell'insieme in modo particolarmente efficace.

35 In una forma di realizzazione, il passo di selezionare dati di entità di un primo sottoinsieme di entità di allenamento dell'insieme di dati di allenamento prevede

di:

- determinare un primo gruppo di entità di allenamento dell'insieme dati di allenamento che hanno subito un malfunzionamento;
- campionare casualmente un secondo gruppo di entità di allenamento dell'insieme dati di allenamento che non hanno subito un malfunzionamento, laddove il secondo gruppo comprende un numero di entità di allenamento compreso tra 1 e 5 volte un numero di entità di allenamento del primo gruppo, e
- comporre il primo sottoinsieme di entità di allenamento dell'insieme di dati di allenamento come unione delle entità di allenamento del primo gruppo e del secondo gruppo.

In questo modo, è possibile ottenere una riduzione iniziale di un insieme di dati di allenamento sbilanciato che permetta di allenare in modo funzionale il sistema di intelligenza artificiale.

In una forma di realizzazione, il passo di predire un malfunzionamento dell'entità mediante un sistema di intelligenza artificiale prevede di utilizzare un algoritmo di apprendimento automatico, selezionato tra:

- Random Forest,
- XGBoost;
- Logistic Regression, e
- una rete neurale.

La Richiedente ha determinato che questi algoritmi di apprendimento automatico, in particolare l'algoritmo Random Forest, permettono di ottenere prestazioni ottimali. Per esempio, l'algoritmo Random Forest permette di ottenere risultati caratterizzati da un fattore AUC (acronimo di Area Under Curve - area sotto curva ROC, acronimo di Receiver Operating Characteristic – caratteristica operativa di ricevitore) compreso tra 88% - 93%.

In una forma di realizzazione, il passo di predire un malfunzionamento dell'entità mediante un sistema di intelligenza artificiale prevede di calcolare una pluralità di probabilità di malfunzionamento dell'entità, ciascuna probabilità di malfunzionamento dell'entità essendo riferita a un rispettivo intervallo di tempo futuro.

Le previsioni di malfunzionamento fornite dal sistema risultano altamente affidabili anche su periodi di tempo futuri molto estesi per esempio compresi tra un anno e cinque anni.

Un differente aspetto della presente invenzione riguarda un sistema per predire un malfunzionamento di un'entità. In particolare, l'entità è caratterizzata da una pluralità di parametri, il sistema comprendendo:

- 5 - un modulo di interfaccia configurato per ricevere dati relativi a detta entità e fornire un risultato di predizione di malfunzionamento;
- un modulo di elaborazione dati configurato per acquisire ed elaborare dati relativi a entità di allenamento e i dati relativi a detta entità forniti al modulo di interfaccia, e
- 10 - un modulo di intelligenza artificiale configurato per eseguire un algoritmo di intelligenza artificiale che utilizza i dati elaborati dal modulo di elaborazione dati per predire una probabilità di malfunzionamento dell'entità.

Vantaggiosamente, i moduli del sistema sono configurati per implementare il metodo secondo una qualsiasi delle rivendicazioni precedenti.

15 Ulteriori caratteristiche e scopi della presente invenzione appariranno maggiormente chiari dalla descrizione che segue.

BREVE DESCRIZIONE DEI DISEGNI

L'invenzione verrà descritta qui di seguito con riferimento ad alcuni esempi, forniti a scopo esplicativo e non limitativo, ed illustrati nei disegni annessi. Questi disegni illustrano differenti aspetti e forme di realizzazione della presente
20 invenzione e, dove appropriato, numeri di riferimento illustranti strutture, componenti, materiali e/o elementi simili in differenti figure sono indicati da numeri di riferimento similari.

La Figura 1 è uno schema a blocchi di un sistema di predizione di malfunzionamento secondo una forma di realizzazione della presente
25 invenzione predisposto per fornire una o più predizioni di malfunzionamento di un'entità monitorata;

la Figura 2 è un diagramma di flusso di una procedura di allenamento implementata dal sistema di predizione di malfunzionamento di Figura 1;

30 la Figura 3 è un grafico qualitativo di una distribuzione di probabilità di malfunzionamento calcolata dal sistema di predizione di malfunzionamento di Figura 1, e

la Figura 4 è un diagramma di flusso di una procedura di predizione di malfunzionamento implementata dal sistema di predizione di malfunzionamento di Figura 1.

DESCRIZIONE DETTAGLIATA DELL'INVENZIONE

Mentre l'invenzione è suscettibile di varie modifiche e costruzioni alternative, alcune forme di realizzazione preferite sono mostrate nei disegni e saranno descritte qui di seguito in dettaglio. Si deve intendere, comunque, che non vi è
5 alcuna intenzione di limitare l'invenzione alla specifica forma di realizzazione illustrata, ma, al contrario, l'invenzione intende coprire tutte le modifiche, costruzioni alternative, ed equivalenti che ricadano nell'ambito dell'invenzione come definito nelle rivendicazioni.

L'uso di "ad esempio", "ecc.", "oppure" indica alternative non esclusive senza
10 limitazione a meno che non altrimenti indicato. L'uso di "include" significa "include, ma non limitato a" a meno che non sia altrimenti indicato.

Con riferimento alla Figura 1 un sistema di predizione di malfunzionamento basato su intelligenza artificiale secondo una forma di realizzazione della presente invenzione – indicato come 'sistema 1' nel seguito, per brevità –
15 comprende un modulo di interfaccia 11, un modulo di elaborazione dati 12 e un modulo di intelligenza artificiale 13, collegati tra loro per scambiare dati – per esempio, in formato binario.

Il modulo di interfaccia 11 è connesso a uno o più dispositivi utente 2 (per esempio, computer o sistemi informatici) al fine di scambiare dati con gli stessi.
20 Inoltre, il modulo di interfaccia 11 è connesso a uno o più sensori 3. In particolare, ciascun sensore 3 è configurato per fornire una misura di una grandezza fisica associata al funzionamento di un'entità monitorata 4.

A tale scopo, il modulo di interfaccia 11 comprende risorse hardware necessarie, come uno o più processori, memorie volatili, memorie non-volatili, modem ecc.

25 Per esempio, l'entità monitorata 4 è un impianto industriale comprendente una pluralità di stazioni di lavorazione automatizzate. In tale caso, i sensori 3 possono comprendere uno o più tra sensori di temperatura, sensori di pressione, sensori di peso, fotocamere, termocamere, sensori di tensione e/o corrente, sensori di portata, ecc. secondo la tipologia di impianto industriale e/o le stazioni
30 di lavorazione comprese nello stesso. In aggiunta, i sensori 3 possono comprendere dispositivi di elaborazione come PLC, microcomputer, microcontrollori, ecc. In tale caso, tali dispositivi di elaborazione sono configurati per fornire informazioni di funzionamento dell'impianto industriale e/o della stazione di lavorazione dello stesso cui sono connessi – per esempio, volume di

produzione, tempo di attività dei componenti compresi nelle stazioni di lavorazione, ecc.

Il modulo di interfaccia 11 è configurato per fornire i dati ricevuti dai dispositivi utente 2 e dai sensori 3 al modulo di elaborazione dati 12. Inoltre, il modulo di interfaccia 11 è collegato al modulo di intelligenza artificiale 13 per scambiare dati con lo stesso. Vantaggiosamente, il modulo di interfaccia 11 è configurato per trasmettere al modulo di intelligenza artificiale 13 richieste di predizione ricevute da un dispositivo utente 2 e/o dati di ingresso da utilizzare per le previsioni come descritto nel seguito. In aggiunta, il modulo di interfaccia 13 è configurato per ricevere dati di output generati dal modulo di intelligenza artificiale 12 da inoltrare a uno o più corrispondenti dispositivi utente 2.

Il modulo di elaborazione dati 12 è connesso a uno o più banche dati 5, o più in generale sorgenti di dati, al fine di scambiare dati con le stesse. In particolare, le banche dati 5 comprendono una pluralità di dati di entità DB. Ciascun dato di entità DB comprende valori di una pluralità di parametri caratteristici di una rispettiva entità (non illustrata).

Il modulo di elaborazione dati 12 è configurato per ricevere, combinare ed elaborare i dati acquisiti dalle banche dati 5 e dal modulo di interfaccia 11, in modo che possano essere utilizzati in modo adeguato dal modulo di intelligenza artificiale 13 al fine di fornire una previsione di malfunzionamento. A tale scopo, il modulo di elaborazione dati 12 comprende risorse hardware necessarie, come uno o più processori (CPU), processori grafici (GPU), processori di segnale (DSP), memorie volatili, memorie non-volatili, modem ecc.

Il modulo di intelligenza artificiale 13 è configurato per elaborare dati processati dal modulo elaboratore 12 e/o forniti dal modulo di interfaccia 11 al fine di predire una probabilità di malfunzionamento dell'entità monitorata 4. Il modulo di intelligenza artificiale 13 è configurato per eseguire un algoritmo di intelligenza artificiale AI, preferibilmente, un algoritmo di apprendimento automatico – o Machine Learning (ML). In una forma di realizzazione preferita della presente invenzione, il modulo di intelligenza artificiale 13 è configurato per eseguire l'algoritmo di apprendimento automatico, preferibilmente, un algoritmo a foresta casuale – o Random Forest.

A tale scopo, il modulo di intelligenza artificiale 13 comprende risorse hardware necessarie, come uno o più processori (CPU), processori grafici (GPU), processori di segnale (DSP), memorie volatili, memorie non-volatili, modem ecc.

Il sistema 1 appena descritto è configurato per eseguire un metodo per predire un malfunzionamento di un'entità secondo una forma di realizzazione della presente invenzione.

5 In particolare, il metodo comprende una procedura di apprendimento 100 e una procedura di predizione 200.

La procedura di apprendimento 100 ha lo scopo di addestrare l'algoritmo di intelligenza artificiale AI al fine di eseguire predizioni di malfunzionamento affidabili.

10 In dettaglio, la procedura di apprendimento 100 prevede di acquisire un insieme di dati di entità DB riferiti a entità di allenamento dalle banche dati 5 (blocco 101). Ciascuna entità di allenamento ha caratteristiche e uno scopo sostanzialmente corrispondenti a, o almeno confrontabili con, le caratteristiche e lo scopo dell'entità monitorata 4. Per esempio, le entità di allenamento sono impianti industriali che producono un medesimo, o simile, lavorato o semilavorato al
15 lavorato o semilavorato prodotto dall'impianto industriale corrispondente all'entità monitorata 4.

Vantaggiosamente, i dati di entità di allenamento acquisiti sono riferiti sia a entità di allenamento che hanno subito un malfunzionamento sia entità di allenamento che non hanno subito un malfunzionamento. Per ciascuna entità di allenamento
20 è acquisito almeno un gruppo di dati di entità. Preferibilmente, per ciascuna entità sono acquisiti gruppi di dati di entità riferiti a due o più intervalli di tempo predeterminati – per esempio, cinque gruppi di dati di entità riferiti all'operazione dell'entità di allenamento durante cinque anni consecutivi, dieci dati riferiti ad altrettanti periodi di esercizio dell'entità di allenamento separati
25 da un periodo di fermo, *ecc.*

Vantaggiosamente, ciascun gruppo di dati di entità comprende valori riferiti a parametri caratteristici della rispettiva entità di allenamento, sia tempo varianti sia tempo invarianti. Per esempio, i parametri indipendenti dal tempo comprendono parametri strutturali (dimensioni, peso, lista delle stazioni e/o dei
30 componenti, *ecc.*) e parametri nominali (potenza massima assorbita/generata, frequenza massima di operazione, carico massimo, MTBF di uno o più componenti dell'impianto e/o delle stazioni dell'impianto, *ecc.*). Diversamente, i parametri dipendenti dal tempo comprendono parametri di funzionamento dell'entità di allenamento (potenza istantanea assorbita, giri al minuto di una
35 componente meccanica/elettromeccanica, quantità di materia prima in ingresso,

quantità di prodotto lavorato o semilavorato in uscita, ecc.). In particolare, i parametri dipendenti dal tempo possono comprendere una pluralità di valori acquisiti in diversi istanti di tempo, un andamento del valore dei parametri in detto intervallo di tempo, una variazione del valore dei parametri tra due o più istanti di tempo e/o simili informazioni.

I dati di entità DB acquisiti sono resi uniformi tra loro (blocco 102). In particolare, i dati di entità sono elaborati in modo che ciascun gruppo di dati di entità comprenda valori riferiti ai medesimi parametri caratteristici dell'entità di allenamento.

10 Per esempio, è eseguita un'operazione di missing value imputation come descritta in Barnard, J., Meng, X.-L., 1999: "*Applications of multiple imputation in medical studies: from aids to nhanes*" *Statistical methods in medical research* 8 (1), 17–36, e una standardizzazione dell'insieme dati. Preferibilmente, qualsiasi valore mancante da un dato di entità è posto a zero ed è eseguita un'operazione di standard scaling.

I dati di entità DB uniformati sono quindi analizzati e separati in dati principali, o core data, e dati secondari, o non-core data (blocco 103). Preferibilmente, i dati principali comprendono gruppi di dati di entità aventi valori riferiti ai medesimi parametri, in modo da consentire un confronto adeguato tra le corrispondenti entità. Diversamente, i dati secondari raccolgono dati riferiti a parametri non contenuti in tutti i gruppi dati di entità dell'insieme di dati di entità.

I dati secondari sono elaborati per generare dati di perturbazione (blocco 104). In particolare, i dati di perturbazione rappresentano un effetto dei dati secondari sui dati principali.

25 I dati principali sono poi suddivisi in un insieme di dati di allenamento DT e un insieme di dati di validazione DV (blocco 105). I dati principali sono suddivisi in modo che l'insieme di dati di allenamento DT sia sostanzialmente maggiore dell'insieme di dati di validazione DV. Preferibilmente, l'insieme di dati di allenamento DT comprende lo 80% dei dati di entità DB e, quindi, l'insieme dei dati di validazione DV comprende il restante 20% dei dati di entità DB.

L'insieme di dati di allenamento DT è poi sottoposto a un processo di riduzione di caratteristiche – o Feature Reduction – (blocco 106). Per esempio, sono rimossi i parametri per cui la presenza o assenza non modifica la precisione dei risultati di un valore maggiore del 1% nell'esecuzione di un'operazione di classificazione.

Preferibilmente, il processo di riduzione delle caratteristiche è un processo di rimozione di caratteristiche iterativo – o iterative feature removal process.

5 Successivamente, è selezionato un sottoinsieme di entità di allenamento associate a corrispondenti gruppi di dati di entità dell'insieme dati di allenamento DT (blocco 107). Per esempio, è eseguito un ridimensionamento (downsizing) dell'insieme dei dati di allenamento, per mezzo di un campionamento casuale delle entità di allenamento dell'insieme di dati di allenamento DT, in modo da ottenere un sottoinsieme di entità di allenamento sostanzialmente ridotto rispetto alla totalità di entità di allenamento dell'insieme di dati di allenamento DT.

10 Nella presente l'espressione "entità di allenamento dell'insieme di dati di allenamento DT" indica un'entità di allenamento di cui almeno un rispettivo gruppo di dati di entità di allenamento sia compreso nell'insieme di dati di allenamento DT.

15 Preferibilmente, nel caso di un insieme di dati di allenamento altamente sbilanciato, è previsto di selezionare separatamente un sottoinsieme delle entità di allenamento dell'insieme di dati di allenamento DT che non hanno subito un malfunzionamento e un sottoinsieme delle entità di allenamento dell'insieme di dati di allenamento DT che hanno subito un malfunzionamento, in particolare con dimensioni differenti, e poi uniti in un unico sottoinsieme di entità di allenamento. Preferibilmente, il sottoinsieme di entità di allenamento ridotto è
20 caratterizzato da proporzione tra le entità di allenamento che hanno subito un malfunzionamento e le entità di allenamento che non hanno subito un malfunzionamento è compresa tra 1:5 e 1:1, preferibilmente, uguale o inferiore a 1:3.

25 L'espressione "insieme di dati di allenamento altamente sbilanciato" indica un insieme dati di allenamento in cui le entità di allenamento che hanno subito un malfunzionamento sono il 20% o meno delle entità di allenamento che non hanno subito un malfunzionamento.

I dati di allenamento associati al sottoinsieme di entità di allenamento selezionate
30 sono utilizzati come primo sottoinsieme di dati di allenamento DT1 (blocco 108) ed è utilizzato per allenare l'algoritmo di intelligenza artificiale AI (blocco 109).

L'algoritmo di intelligenza artificiale AI così addestrato è sottoposto a valutazione (blocco decisionale 110). In particolare, l'insieme di dati di validazione DV e i dati di perturbazione sono forniti in ingresso all'algoritmo di

intelligenza artificiale AI e quindi si verifica la corrispondenza della predizione di malfunzionamento per ciascuna entità di allenamento associata ai dati di validazione DV. Preferibilmente, è verificato che le predizioni di malfunzionamento dell'algoritmo di intelligenza artificiale AI non siano affette da sovradattamento, o overfitting.

In caso la validazione rilevi la produzione di risultati sub-ottimi (ramo di uscita N del blocco 110), gli iperparametri (hyperparameter) dell'algoritmo di intelligenza artificiale sono eventualmente regolati in modo da ottimizzare la precisione e il recupero delle predizioni di malfunzionamento fornite dall'algoritmo di intelligenza artificiale (blocco 111) ed è ripetuto il passo di validazione per verificare che l'algoritmo di intelligenza artificiale AI operi in modo desiderato. Preferibilmente, la regolazione degli iperparametri è eseguita per mezzo di un approccio di regolazione basato su ricerca di griglia (Grid-search) - per esempio, in modo analogo a quanto descritto in Bergstra, J., Bengio, Y., 2012: *"Random search for hyper-parameter optimization"*, Journal of Machine Learning Research 13, 281–305.

Una volta che la validazione rileva la produzione di risultati accettabili (ramo di uscita Y del blocco 110), è previsto di utilizzare l'algoritmo di intelligenza artificiale AI per calcolare una probabilità di malfunzionamento delle entità di allenamento dell'insieme di dati allenamento DT (blocco 112).

Per esempio, l'algoritmo di intelligenza artificiale AI è eseguito più volte. A ogni iterazione l'algoritmo è configurato per fornire in uscita un numero e/o una lista di entità di allenamento dell'insieme dati DT la cui probabilità di malfunzionamento è minore o uguale a un rispettivo valore di soglia.

Preferibilmente, l'algoritmo di intelligenza artificiale AI è eseguito per determinare le entità di allenamento con una probabilità di malfunzionamento uguale o inferiore a 20%, 30%, 40%, 50% e 60%.

Ancor più preferibilmente, l'operazione di calcolare la probabilità di malfunzionamento delle entità di allenamento dell'insieme di dati allenamento DT è eseguita per ciascun intervallo di tempo per cui sono presenti corrispondenti gruppi di dati delle entità di allenamento.

Combinando i risultati delle iterazioni dell'algoritmo di intelligenza artificiale AI si ottiene una distribuzione della probabilità di malfunzionamento delle entità di allenamento dei dati di allenamento DT (blocco 113) (illustrata qualitativamente

in Figura 3). Preferibilmente, è definita una distribuzione della probabilità di malfunzionamento delle entità di allenamento dei dati di allenamento DT per ciascun intervallo di tempo per cui sono presenti corrispondenti gruppi di dati delle entità di allenamento.

- 5 Sulla base della distribuzione di probabilità di malfunzionamento così definita, è selezionato un secondo sottoinsieme di dati di allenamento DT2 a partire dai dati di allenamento DT (blocco 114). In particolare, il secondo sottoinsieme di dati di allenamento SDT2 comprende dati riferiti a un sottoinsieme di entità di allenamento delle entità di allenamento dell'insieme di allenamento DT
- 10 selezionati in modo che il sottoinsieme di entità di allenamento abbia una distribuzione di probabilità di malfunzionamento corrispondente alla distribuzione di probabilità di malfunzionamento della totalità delle entità di allenamento dell'insieme di allenamento. Preferibilmente, il secondo sottoinsieme di dati di allenamento SDT2 è definito in modo che vi sia una
- 15 corrispondenza con ciascuna delle distribuzioni della probabilità di malfunzionamento calcolate per un rispettivo intervallo di tempo.

L'algoritmo di intelligenza artificiale AI è sottoposto a un nuovo allenamento basato sul secondo sottoinsieme di dati di allenamento DT2 (blocco 115).

- Al termine dell'allenamento, l'algoritmo di intelligenza artificiale AI è
- 20 nuovamente sottoposto a valutazione (blocco decisionale 116) e regolazione degli iperparametri (blocco 117) nel caso sia rilevata la produzione di risultati sub-ottimi (ramo di uscita N del blocco 116), in modo analogo a quanto sopra descritto.

- Al termine della procedura di allenamento 100 (ramo di uscita Y del blocco 116)
- 25 il sistema 1 è disponibile per ricevere richieste di predizione di malfunzionamento dell'entità monitorata 4 eseguite tramite un dispositivo utente 2 (blocco 118).

- La procedura di predizione 200 (cui si riferisce il diagramma di flusso di Figura 4), inizialmente, prevede di acquisire una pluralità di dati operativi DE relativi
- 30 all'entità monitorata 4 (blocco 201). In particolare, i dati operativi DE sono acquisiti attraverso i sensori 3 e/o il dispositivo utente 2.

Vantaggiosamente, i dati operativi DE comprendono valori relativi a parametri caratteristici dell'entità monitorata indipendenti dal tempo e/o relativi a parametri caratteristici dell'entità monitorata dipendenti dal tempo in modo

analogo a quanto sopra descritto rispetto ai dati di entità DB.

I dati operativi DE acquisiti sono quindi analizzati e separati in dati principali, o core data, e dati secondari, o non-core data (blocco 202). Preferibilmente, i dati principali comprendono un sottoinsieme di dati operativi DE corrispondenti a
5 dati di allenamento DT utilizzati durante un metodo di allenamento descritto nel seguito della presente. Diversamente, i dati secondari comprendono un sottoinsieme dei dati operativi DE che non corrispondono ai dati di allenamento DT.

Per esempio, il modulo di elaborazione dati 12 comprende un'unità di
10 discriminazione 121 configurata per analizzare e separare i dati operativi DE in dati principali e dati secondari.

I dati secondari sono elaborati per generare dati di perturbazione (blocco 203). Per esempio, il modulo di elaborazione dati 12 comprende un'unità di perturbazione 122 configurata per ricevere in ingresso i dati secondari e generare
15 in uscita corrispondenti dati di perturbazione.

I dati principali e i dati di perturbazione sono forniti in ingresso all'algoritmo di intelligenza artificiale eseguito dal modulo di intelligenza artificiale. L'algoritmo di intelligenza artificiale fornisce una predizione di malfunzionamento PM dell'entità monitorata 4 (blocco 204).

20 Preferibilmente, la predizione di malfunzionamento comprende almeno una probabilità che l'entità, o un suo componente, subisca un malfunzionamento entro un intervallo di tempo futuro a partire dall'acquisizione dei dati operativi DE dell'entità monitorata 4

In una forma di realizzazione preferita, l'algoritmo di intelligenza artificiale è
25 configurato per fornire una sequenza di predizioni di malfunzionamento PM_{1-n} riferiti a intervalli di tempi consecutivi. Per esempio, genera una probabilità di malfunzionamento associata per due o più intervalli di tempo selezionati tra 6 mesi, 12 mesi, 24 mesi, 36 mesi, 48 mesi, 60 mesi a partire dall'acquisizione dei dati operativi DE dell'entità monitorata 4.

30 Inoltre, l'algoritmo di intelligenza artificiale genera un'indicazione degli uno o più parametri caratteristici dell'entità monitorata 4 risultati più rilevanti nel calcolo della probabilità di malfunzionamento per ciascun intervallo di tempo considerato.

È tuttavia chiaro che gli esempi sopra riportati non devono essere interpretati in senso limitativo e l'invenzione così concepita è suscettibile di numerose modifiche e varianti.

5 Per esempio, in forme di realizzazione alternative (non illustrate) i dati operativi relativi all'entità da monitorare sono forniti al sistema 1 unicamente attraverso uno o più dispositivi utente 2 collegati al modulo di interfaccia 11 del sistema 1. In questo caso, le misure fornite dai sensori 3 sono ricevute dai dispositivi utente o, alternativamente, non sono previsti sensori 3 per effettuare una misurazione automatica dei parametri dell'entità monitorata 4, ma i parametri sono
10 determinati sulla base di misure/osservazioni indirette.

Nulla vieta di realizzare il modulo di interfaccia 11, il modulo di elaborazione dati 12 e/o il modulo di intelligenza artificiale 13 come un unico modulo.

Sarà evidente al tecnico del settore che il modulo di interfaccia 11, il modulo di elaborazione dati 12 e/o il modulo di intelligenza artificiale 13 possono essere
15 implementati attraverso sia risorse hardware dedicate sia risorse hardware condivise. Ancora, uno o più moduli possono essere realizzate, almeno in parte, tramite macchine virtuali.

Sebbene l'algoritmo di apprendimento automatico preferito sia del tipo Random Forest, nulla vieta di configurare il modulo di intelligenza artificiale per
20 implementare un differente algoritmo. In particolare, in forme di realizzazione alternative (non illustrate), il modulo di intelligenza artificiale è configurato per eseguire un algoritmo di tipo XGBoost, regressione logistica (Logistic Regression) o una rete neuronale (Neural Network - NN).

Come sarà evidente al tecnico del settore, uno o più passi delle procedure 100 e
25 200 sopra descritte possono essere eseguiti in parallelo tra loro o con un ordine differente da quello sopra presentato. Analogamente, uno o più passi opzionali possono essere aggiunti o rimossi da uno o più delle procedure sopra descritte.

Per esempio, in forme di realizzazione alternative (non illustrate), l'uniformazione dei dati e la separazione dei dati in insieme di allenamento e
30 insieme di validazione sono eseguite in ordine inverso.

In una forma di realizzazione (non illustrata), è omesso il passo di riduzione di caratteristiche nella procedura di allenamento. In aggiunta o in alternativa, nella procedura di previsione la conversione dei dati secondari in dati di

perturbazione può essere omessa.

Ancora, nulla vieta di omettere i passi di validazione e/o i passi di regolazione degli iperparametri.

- In una forma di realizzazione alternativa (non illustrata), è previsto di ridefinire una pluralità di sottoinsiemi di dati di addestramento con cui addestrare sequenzialmente l'algoritmo di intelligenza artificiale al fine di determinare un insieme di dati di allenamento definitivo. In dettaglio, una volta completato l'allenamento dell'intelligenza artificiale AI per mezzo del primo sottoinsieme di dati di allenamento, l'algoritmo di intelligenza artificiale AI è eseguito per elaborare i dati dell'insieme di allenamento DT in modo da determinare le entità di allenamento con una probabilità di malfunzionamento uguale o inferiore a una prima soglia – per esempio, 20%. Successivamente, è previsto di selezionare un nuovo sottoinsieme di entità di allenamento, costituito da un campionamento di entità comprese tra le entità di allenamento con una probabilità di malfunzionamento uguale o inferiore alla prima soglia e le entità di allenamento oggetto di un malfunzionamento. L'algoritmo di intelligenza artificiale AI è allenato con il nuovo sottoinsieme di dati di allenamento e poi è eseguito per elaborare i dati dell'insieme di allenamento DT in modo da determinare le entità di allenamento con una probabilità di malfunzionamento uguale o inferiore a una seconda soglia – per esempio, 30%. I passi appena descritti sono iterati ogni volta per identificare entità di allenamento con una probabilità uguale o inferiore a una soglia maggiore della precedente – per esempio, altre tre iterazioni con soglie paria 40%, 50% e 60%, rispettivamente – fino a che il numero di entità di allenamento individuato ha un valore significativo – per esempio, una soglia oltre la quale il sistema degrada fortemente come oltre il 70% nell'esempio delle iterazioni sopra considerato. I risultati dei passi precedenti sono utilizzati per determinare la distribuzione di probabilità di malfunzionamento delle entità di allenamento dell'insieme di dati di allenamento DT e determinare il sottoinsieme di dati di allenamento definitivo in modo analogo a quanto sopra descritto.
- Come sarà evidente nel caso di un insieme di dati di allenamento altamente sbilanciato, ossia comprendente un gruppo di entità di allenamento dell'insieme dati di allenamento che hanno subito un malfunzionamento molto inferiore di un gruppo di entità di allenamento dell'insieme dati di allenamento che non hanno subito un malfunzionamento – per esempio, in cui le entità di allenamento del primo gruppo sono 5% o meno delle entità di allenamento del secondo gruppo - il peso del primo gruppo di entità di allenamento è sulla distribuzione

della probabilità di malfunzionamento dell'insieme di dati di allenamento è trascurabile.

Di conseguenza, il passo di determinare la distribuzione delle probabilità di malfunzionamento delle entità di allenamento dell'insieme di allenamento DT (blocco 112 della procedura 100) può essere implementato determinando la
5 distribuzione delle probabilità di malfunzionamento delle entità di allenamento del solo secondo gruppo di entità – ossia che non hanno subito un malfunzionamento. Di conseguenza, il passo di determinare una distribuzione della probabilità di malfunzionamento delle entità di allenamento dei dati di
10 allenamento DT (blocco 113) prevede di selezionare un terzo sottoinsieme di entità di allenamento SDT3 del secondo gruppo – ossia, tutte entità che non hanno subito un malfunzionamento – in modo tale che il terzo sottoinsieme di entità di allenamento presenti una distribuzione delle probabilità di malfunzionamento uguale alla distribuzione della probabilità di
15 malfunzionamento delle entità di allenamento dei dati di allenamento DT. Infine, il secondo sottoinsieme di entità di allenamento SDT2 dell'insieme di dati di allenamento DT è, quindi, definito come l'unione delle entità di allenamento del primo gruppo e del terzo sottoinsieme di entità di allenamento SDT3.

In un'ulteriore forma di realizzazione (non illustrata), è previsto di ripetere la
20 procedura di allenamento 100. Per esempio, la procedura di allenamento 100 può essere reiterata trascorso un periodo di tempo predeterminato (come un anno) dall'iterazione precedente della stessa, acquisendo nuovi dati di entità dalle banche dati.

Come sarà evidente, la combinazione delle procedure 100 e 200 sopra presentate
25 forma un complessivo metodo di predizione di malfunzionamento di un'entità monitorata secondo la presente invenzione.

Naturalmente, tutti i dettagli sono sostituibili da altri elementi tecnicamente equivalenti.

Inoltre, sebbene nella descrizione si sia fatto riferimento a un'entità monitorata
30 corrispondente a un impianto industriale, sarà evidente a un tecnico del settore che il metodo e il sistema sopra descritti possono essere adattati per fornire predizioni relative a entità differenti senza richiedere modifiche sostanziale al metodo e sistema secondo la presente invenzione.

Per esempio, in una forma di realizzazione il metodo e il sistema sono configurati

per fornire predizioni di malfunzionamento di un sistema informatico, come un data center, in tale caso i parametri comprendono parametri relativi alle risorse hardware installate, parametri relativi al funzionamento di tali risorse hardware, parametri relativi all'edificio in cui è situato il data center, parametri relativi a un
5 impianto HVAC che regola la temperatura del data center, parametri relativi al funzionamento di detto impianto HVAC, ecc.

Ancora, il metodo e il sistema sono configurati per fornire predizioni di malfunzionamento, ossia il fallimento, di un'entità legale, in particolare un'azienda. In tale caso, il sistema 1 è collegato a banche dati 5 anche di tipo
10 finanziarie pubbliche – per esempio, banche dati delle camere di commercio o simili enti della nazione in cui è operativa l'entità monitorata – e private – per esempio, la banca dati AIDA.

Dalle banche dati 5 sono estratti i dati relativi ad aziende fallite e aziende non fallite con caratteristiche analoghe all'entità monitorata – per esempio, in termini
15 di settore commerciale e volume di affari. In particolare, sono acquisiti un numero predeterminato di anni fiscali – per esempio gli ultimi cinque fiscali di entità. Preferibilmente, sono scartate dall'insieme dati le aziende per cui non siano disponibili un numero di relazioni finanziarie ufficiali maggiore o uguali a un numero di soglia – per esempio, uguale a cinque.

L'insieme di dati così acquisiti sono poi suddivisi in dati principali e dati
20 secondati come sopra descritto. I dati principali sono poi elaborati per uniformare l'insieme dati ed è eseguita una riduzione delle caratteristiche, in modo analogo a quanto sopra descritto.

I dati secondari sono elaborati per generare perturbazioni da fornire in ingresso
25 all'algoritmo di intelligenza artificiale. In tale caso, i dati secondari sono elaborati secondo una tassonomia ad albero basata su modello SHELL (Software, Hardware, Environment, Liveware People and Liveware Environment) come descritto in Cantamessa, M., Gatteschi, V., Perboli, G., Rosano, M., 2018: "*Startups' Roads to Failure*", Sustainability 10 (7), 2346. In particolare, il risultato
30 di tale elaborazione è processato per mezzo di un insieme di regole che mappano l'effetto i risultati dell'elaborazione sui dati principali e creano una matrice di impatto di uscita (risk impact matrix) che è fornita in ingresso all'algoritmo di intelligenza artificiale.

L'insieme dei dati di entità di allenamento derivato dai dati principali risulta
35 altamente sbilanciato, in particolare il numero delle aziende non fallite è

sostanzialmente maggiore del numero di aziende fallite – per esempio, il numero di aziende non fallite è maggiore di due ordini di grandezza rispetto al numero di aziende fallite. L'insieme di dati è campionato come sopra descritto. In una forma di realizzazione, è previsto di campionare solo le aziende non fallite e
5 mantenere la totalità delle aziende fallite o, alternativamente, è previsto di definire un rapporto tra aziende non fallite e aziende fallite compreso tra 5 e 1, più preferibilmente tra 3 e 1.

In modo analogo a quanto sopra descritto, l'algoritmo di intelligenza artificiale è allenato con il sottoinsieme di dati di allenamento e con i dati di perturbazione,
10 validato e, eventualmente gli iperparametri sono regolati.

Dopodiché, l'intero insieme di dati di allenamento è elaborato dall'algoritmo di intelligenza artificiale in modo analogo a quanto sopra descritto. Al termine dell'elaborazione da parte dell'algoritmo di intelligenza artificiale, si ha una probabilità di fallimento per ciascuna delle aziende dell'insieme di allenamento
15 per ciascun intervallo di tempo considerato, per esempio per ciascun dei cinque anni.

Come sopra descritto, le probabilità di fallimento sono combinate per ottenere una distribuzione di probabilità di fallimento per ciascun intervallo di tempo considerato. Di conseguenza, è generato un nuovo sottoinsieme di dati
20 allenamento comprendente dati di allenamento associati a un nuovo campione di aziende dell'insieme di allenamento selezionate in modo da avere una distribuzione di probabilità di fallimento durante ogni intervallo di tempo uguale alla corrispondente distribuzione di probabilità di fallimento per ciascun intervallo di tempo delle aziende dell'insieme di allenamento.

L'algoritmo di intelligenza artificiale è quindi allenato con il nuovo sottoinsieme di dati di allenamento validato e, eventualmente gli iperparametri sono regolati, completando la procedura di allenamento del sistema.
25

Durante la procedura di predizione, tramite il dispositivo utente è possibile richiedere una predizione al sistema sulla probabilità di fallimento di un'azienda monitorata. In particolare, è previsto di ricevere in ingresso dati aziendali e,
30 preferibilmente, una matrice di rischio relativa all'azienda direttamente attraverso il dispositivo utente. Inoltre, i sensori sono omessi o sostituiti da un collegamento a un sistema informatico in grado di fornire informazioni finanziarie e/o di produttività dell'azienda monitorata – per esempio, un
35 andamento in tempo reale di una quotazione sul mercato azionario dell'azienda,

di aziende clienti e/o collegate, e/o un andamento del mercato azionario dei beni prodotti o commerciati dall'azienda monitorata.

5 In modo analogo a quanto sopra descritto, i dati in ingresso sono divisi in dati principali e dati secondari, con i dati secondari che sono elaborati per generare corrispondenti perturbazioni che sono fornite in ingresso all'algoritmo di intelligenza artificiale insieme ai dati principali.

10 L'algoritmo di intelligenza principale fornisce quindi in uscita una predizione di fallimento associato all'azienda monitorato riferita a una pluralità di periodi di tempo futuri. In una forma di realizzazione della presente invenzione, sono fornite cinque predizioni di fallimento, le quali comprendono una probabilità di fallimento dell'azienda a 12 mesi, 24 mesi, 36 mesi, 48 mesi e 60 mesi, dalla richiesta di predizione. Preferibilmente, assieme alla predizione di fallimento sono forniti anche uno o più indicatori di prestazione relativi a regolamenti nazionali e internazionali – per esempio z-score di Altman e/o Indici di allerta
15 Crisi d'impresa italiani (definiti nel Codice della Crisi d'Impresa e dell'Insolvenza - D.Lgs. 14/2019).

In una forma di realizzazione preferita, le predizioni di fallimento e i dati di ingresso sono ulteriormente elaborati in modo da identificare i parametri che hanno maggiore impatto sulla probabilità di fallimento dell'azienda monitorata.

20 Infine, la predizione di probabilità e i parametri critici identificati sono combinati in un rapporto di uscita che è fornito al dispositivo utente che ha eseguito la richiesta.

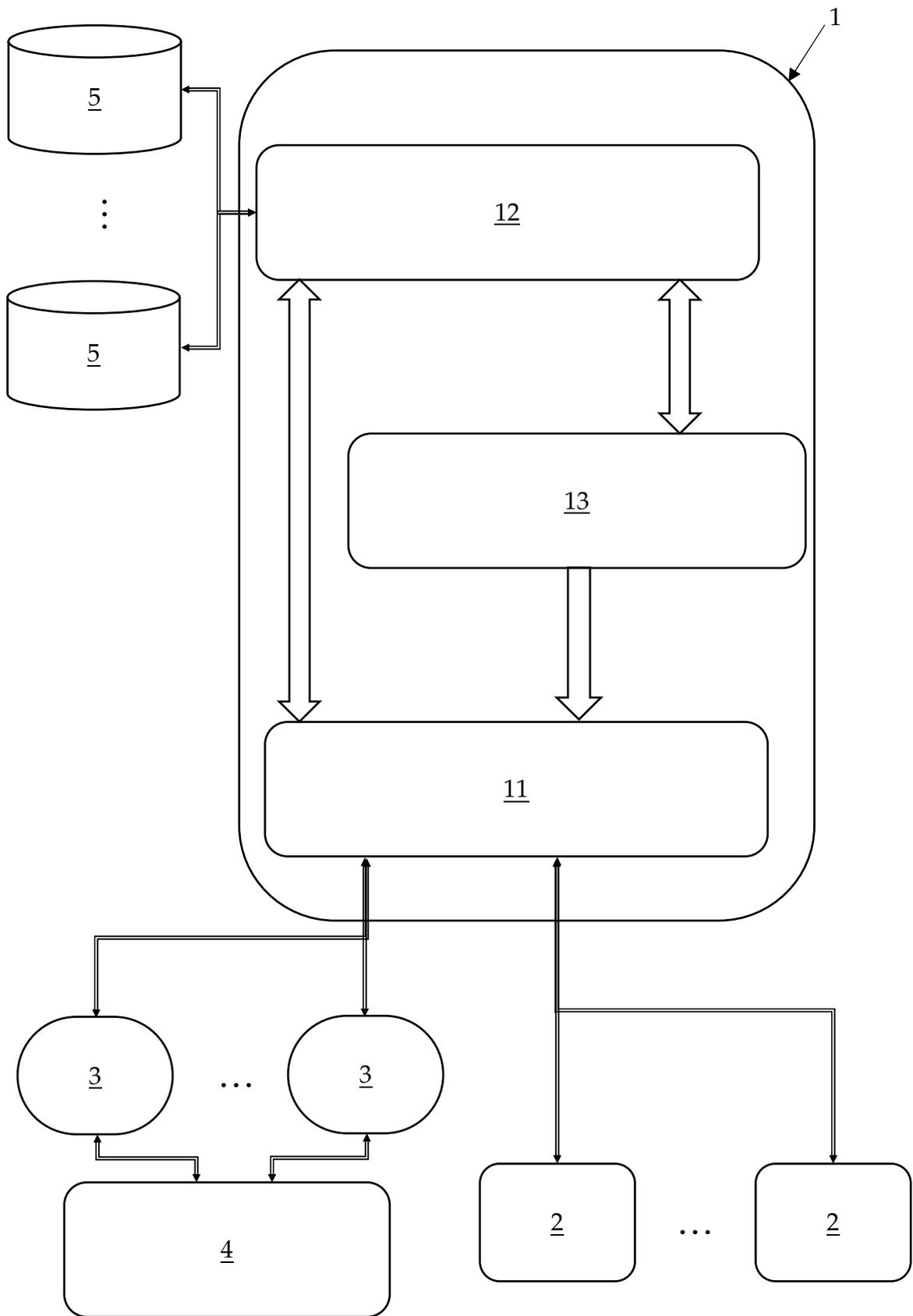


Fig.1

100

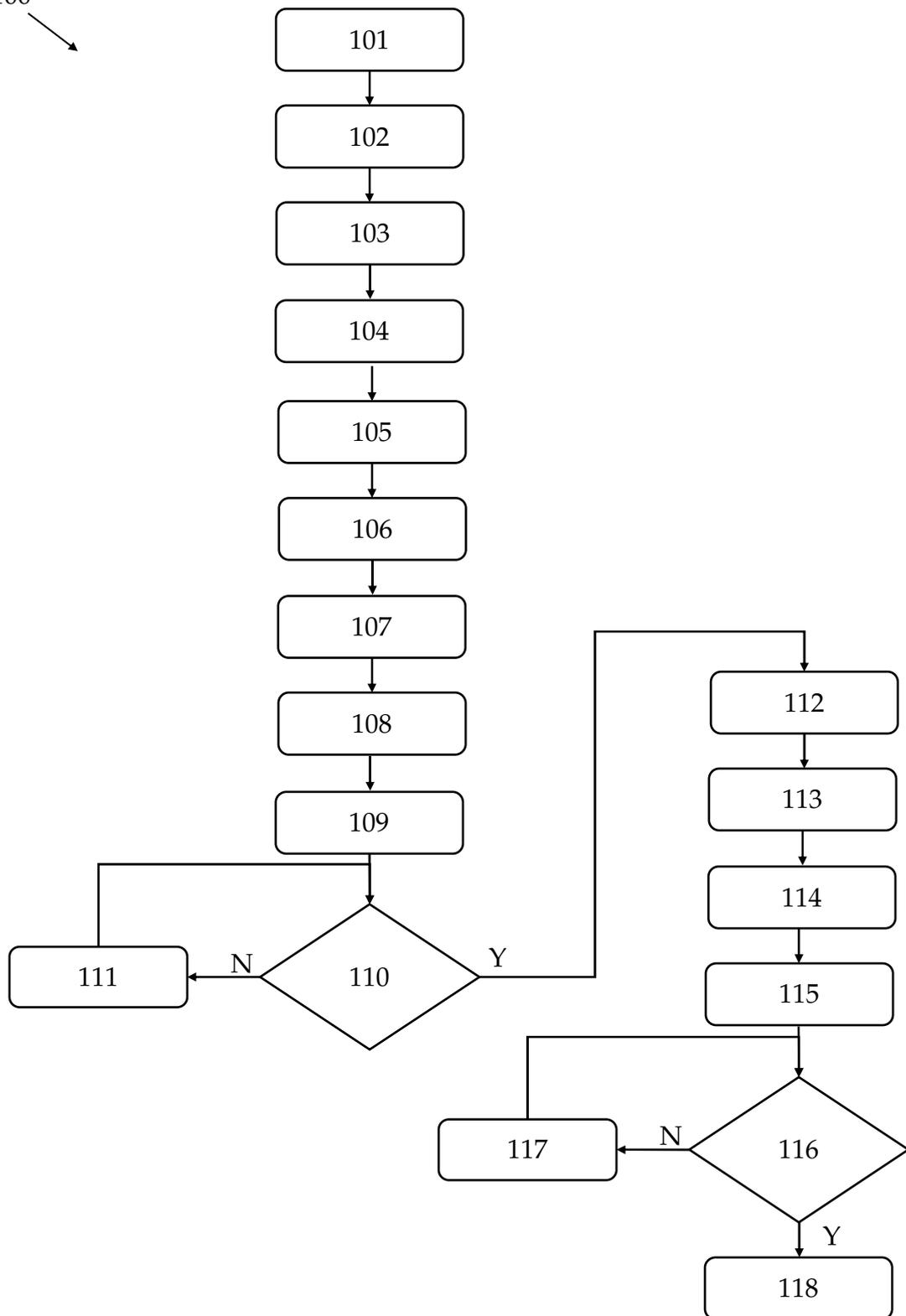


Fig.2

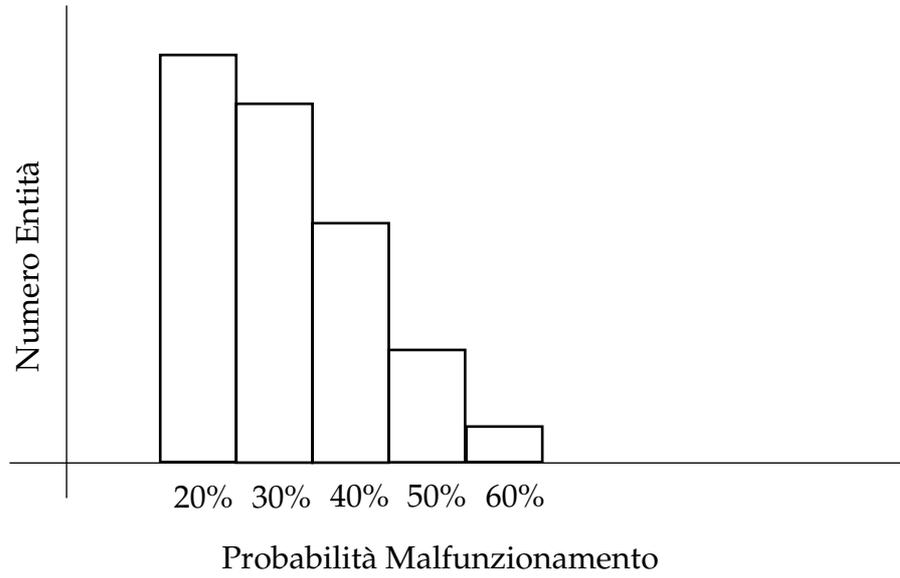


Fig.3

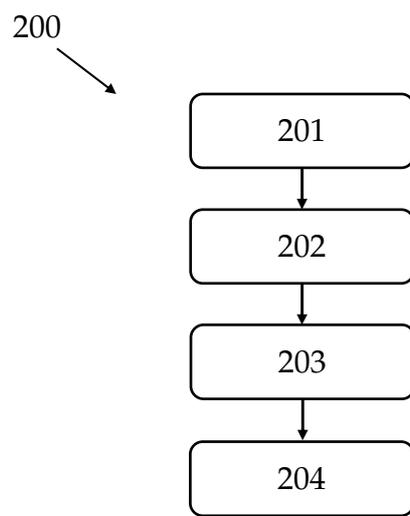


Fig.4