

MANUTENZIONE PREVENTIVA PER LA ROBOTICA E L'AUTOMAZIONE INTELLIGENTE

Progetto: Data Challenge – PHM Asia Pacific 2023

Studente: Francesco DI Bernardo, 1102495

Sommario

Introduzione	5
Scenario sperimentale	5
Descrizione anomalie e guasti	7
Anomalie dovute a bolle.....	7
Guasti elettrovalvole	7
Anomalia sconosciuta	7
Obiettivi	7
Dataset.....	9
Training Data	9
labels.xlsx	9
Test Data (data per inferenza).....	9
labels_spacecraft.xlsx	9
Introduzione script Matlab.....	10
ETL	11
[0.1] caricamento csv e xlsx.....	11
[0.2] creazione tabelle train e inferenza	11
[0.3] correzione tabella <i>table_labels_train</i>	11
[0.3.1] creazione e aggiunta colonna 'normal_abnormal'	11
[0.3.2] creazione e aggiunta colonna 'anomaly_fault'	11
[0.3.3] creazione e aggiunta colonna 'anomaly_position_BP_BV'	12
[0.3.4] creazione e aggiunta colonna 'fault_position_SV'	12
[0.3.5] creazione e aggiunta colonna 'fault_position_SV_percent'	12
[0.4] tabelle finali per training	13
[0.4.1] selezione colonne per 'normal_abnormal'	13
[0.4.2] selezione colonne per 'anomaly_fault'	13
[0.4.3] selezione colonne per 'anomaly_position_BP_BV'	14
[0.4.4] selezione colonne per 'fault_position_SV'	14
[0.4.5] selezione colonne per 'fault_position_SV_percent'	15
Estrazione e selezione features [Diagnostic Features Design] e creazione modello [Classifier Lerner] ..	16
Features normal e abnormal	17
Caratteristiche modello di Classificazione.....	17
Features anomaly e fault.....	20
Caratteristiche modello di Classificazione:.....	20
Features position anomaly	23
Caratteristiche modello di Classificazione:.....	23

Features position fault.....	26
Caratteristiche modello di Classificazione:.....	26
Features percent opening fault.....	29
Caratteristiche modello di Classificazione:.....	29
Inferenza.....	32
[1] Analisi 'normal e abnormal'	32
[2] Analisi 'anomaly e fault'	32
[3] Analisi 'position anomaly'	33
[4] Analisi 'position fault'	33
[5] Analisi 'percent opening fault'	34
Risultati.....	35
Risultati normal_abnormal.....	35
Risultati anomaly o fault.....	35
Risultati anomaly position	35
Risultati fault position	35
Risultati fault position percent.....	35
Conclusioni	35

Figura 1 Schematizzazione del sistema di propulsione simulato	6
Figura 2 Tipico profilo delle pressioni.....	6
Figura 3 tabella table_labels_train_normal_abnormal.....	13
Figura 4 tabella table_labels_train_anomaly_fault.....	13
Figura 5 tabella table_labels_train_anomaly_position_BP_BV	14
Figura 6 tabella table_labels_train_fault_position_SV	14
Figura 7 tabella table_labels_train_fault_position_SV_percent.....	15
Figura 8 matrice confusione modello per normal abnormal	18
Figura 9 ROC per modello normal e abnormal.....	19
Figura 10 esempio risultato classificazione normal_abnormal	32
Figura 11 esempio risultato classificazione anomaly_fault.....	32
Figura 12 esempio risultato classificazione anomaly_position_BP_BV	33
Figura 13 esempio risultato classificazione fault_position_SV	33
Figura 14 esempio risultato classificazione fault_position_SV_percent.....	34

Introduzione

Il progetto è stato sviluppato prendendo spunto da una “Data Challenge” proposta dalla AJAX (Japan Aerospace Exploration Agency).

Nello specifico, la AJAX con questa consegna lancia una sfida nel cercare di migliorare la tecnologia PHM per la propulsione dei propri veicoli spaziali di nuova generazione, chiedendo di diagnosticare situazioni di anomalia, bolle e guasti su elettrovalvole sconosciute utilizzando dati generati da un simulatore per un sistema di propulsione sviluppato dalla stessa AJAX.

Scenario sperimentale

Il sistema di propulsione sperimentale è mostrato in Figura 1, esso lavora con acqua pressurizzata a 2Mpa e possiamo notare la presenza di 4 elettrovalvole solenoidi (SV1 – SV4) che simulano i propulsori e 7 sensori di pressione (P1 – P7).

La simulazione di funzionamento del sistema è la seguente: viene simulata l’apertura e chiusura delle elettrovalvole, così da osservare fluttuazioni di pressione all’interno della rete. Con i dati raccolti si vuole rilevare guasti dovuti alla presenza di bolle all’interno della rete o elettrovalvole che si aprono non correttamente, il quale si traduce in una riduzione del volume del fluido che attraversa l’elettrovalvola.

Le elettrovalvole correttamente funzionanti si aprono e si chiudono con un rapporto di apertura rispettivamente del 100% e dello 0% (aperte o chiuse), inoltre ci sono otto posizioni in cui è possibile verificare la presenza di bolle indesiderate, BV1 e da BP1 a BP7 come mostrato in Figura 1.

I dati tipici delle serie temporali sono riportati in Figura 2: L’elettrovalvola si apre in 100 ms e si chiude in 300 ms. Per tenere conto delle differenze individuali dell’elettrovalvola che compaiono nell’attrezzatura reale, il movimento della valvola ha un’incertezza di 1 ms. I tempi di apertura e chiusura sono di 400 ms, anche con l’incertezza, e la sequenza di apertura e chiusura vengono eseguite tre volte di fila, avendo una durata totale su 1200 ms.

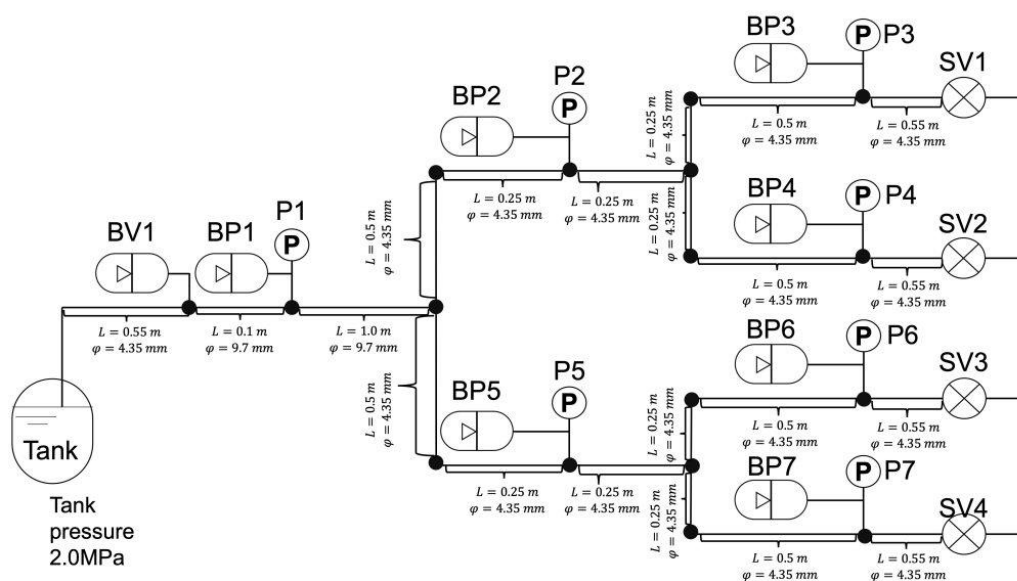


Figura 1 Schematizzazione del sistema di propulsione simulato

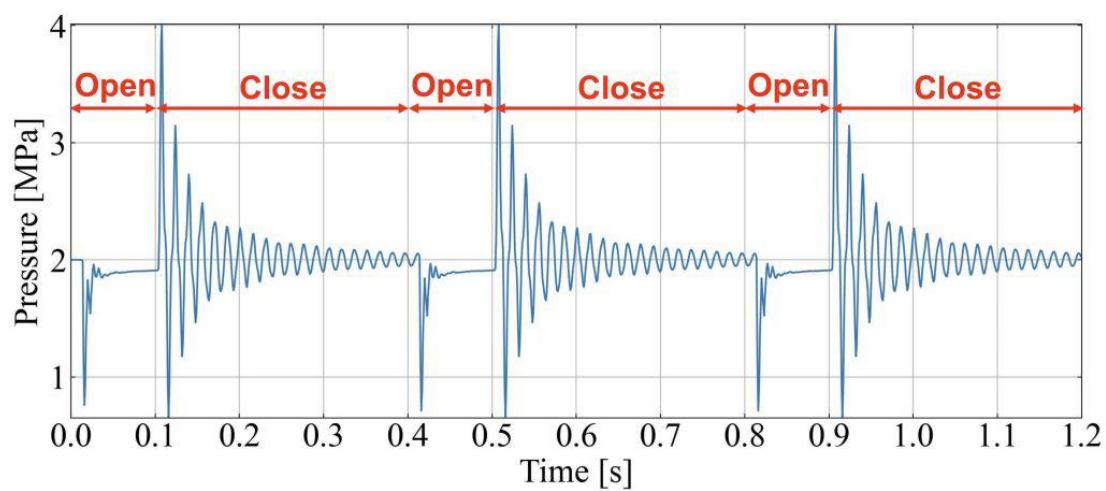


Figura 2 Tipico profilo delle pressioni

Descrizione anomalie e guasti

In questa sfida, le anomalie sono dovute alla presenza di bolle e malfunzionamento nell'apertura delle elettrovalvole, a seguire si descrivono le singole situazioni:

Anomalie dovute a bolle

Occasionalmente compaiono bolle d'aria nei tubi durante il funzionamento effettivo del veicolo spaziale. L'esistenza di bolle provoca lievi variazioni nelle fluttuazioni di pressione.

E' desiderabile rilevare l'aspetto della bolla e la sua posizione, e possiamo distinguere 8 stati possibili all'interno della rete, BV1 e da BP1 a BP7.

Guasti elettrovalvole

Questa è una delle principali modalità di guasto nei sistemi di propulsione dei veicoli spaziali. È necessario trovare quali elettrovalvole si sono guastate e il loro rapporto di apertura.

Le elettrovalvole si aprono e si chiudono con un rapporto di apertura rispettivamente del 100% e dello 0%.

In caso di guasto, le elettrovalvole si apriranno di una percentuale compresa tra 0% e 100%, il che si traduce in una riduzione del volume di fluido attraverso l'elettrovalvola.

Anomalia sconosciuta

Nel funzionamento pratico possono verificarsi anomalie o guasti del tutto imprevisti e sconosciuti. È inoltre necessario distinguere le anomalie sconosciute senza confonderle con le anomalie note e difetti.

Poiché le elettrovalvole hanno differenze individuali, come i tempi di apertura e chiusura, i dati delle serie temporali acquisiti dal sistema di propulsione riportano differenze individuali nel singolo veicolo spaziale. Nei dataset forniti vengono utilizzati 4 veicoli spaziali (n. 1 a 4). I dataset di addestramento comprendono i veicoli dal n.1 a n.3 e il dataset di test i veicoli n.1 e n.4.

Obiettivi

Gli obiettivi del progetto sono stati estratti dagli stessi dichiarati dalla challenge AJAX, quindi si possono trovare nello specifico in: [/dataset/readme.md](#) a pagina 3 sotto la voce '**Prediction Goals**'

A seguire sono riportati in modo sintetico:

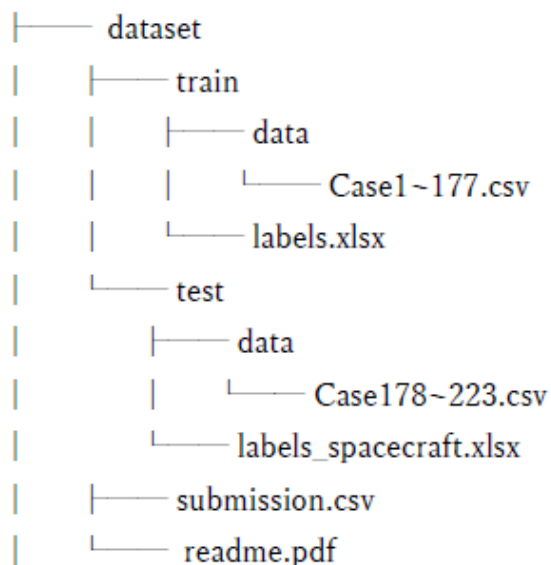
1. **Determinare situazioni normali e abnormali** per tutti i dati di test.
2. Per i dati rilevati come anomali, **determinare se si tratta di un'anomalia** dovuta a contaminazione di bolle, **guasto** dell'elettrovalvola o guasto sconosciuto
3. Per i dati identificati come contaminazione da bolle, **determinare la posizione della bolla** nelle otto posizioni, BV1 e da BP1 a BP7.
4. Per i dati identificati come guasto dell'elettrovalvola, **determinare quale delle quattro elettrovalvole** (da SV1 a SV4) **si è guastata**.
5. Per l'elettrovalvola identificata come guasta, **prevedere il rapporto di apertura**.

Possiamo notare come il **dataset Train** sia necessario per addestrare modelli intelligenti ed in fine il **dataset Test** sia necessario per la fase di inferenza (contro intuitivamente al nome dato).

Quindi, nel procedere all'analisi si inizierà con una breve fase di **ETL** per proseguire con la costruzione di più classificatori per finire con l'inferenza dei dati in Test.

Dataset

Sintesi della directory contenente i datasets.



A seguire vengono descritti i vari file presenti nel dataset.

Training Data

- Numero di dati normali per Spacecraft 1: 35
- Numero di dati normali per Spacecraft 2: 35
- Numero di dati normali per Spacecraft 3: 35
- Numero di dati normali per Spacecraft 1: 24
- Numero di dati normali per Spacecraft 2: 24
- Numero di dati normali per Spacecraft 3: 24

labels.xlsx

Descrive informazioni dettagliate sui dati di addestramento come: un numero identificativo, numero del veicolo spaziale e condizioni sperimentali. Si prega di notare che non ci sono dati Spacecraft-4 nei dati di addestramento.

Test Data (data per inferenza)

- Numero di dati normali per Spacecraft 1: 23
- Numero di dati normali per Spacecraft 4: 23

labels_spacecraft.xlsx

Descrive le informazioni relative al nome del caso e al numero del veicolo spaziale. Il rapporto di apertura della valvola viene scelto casualmente tra 0% e 100% in caso di guasto.

Introduzione script Matlab

Nello script Matlab, sono riportate le fasi di analisi, a seguire vengono introdotte brevemente per poi essere descritte in dettagli nei capitoli successivi.

Lo script è commentato cercando di accompagnare il lettore nella comprensione del codice. Il codice stesso è ordinato seguendo gli obiettivi della challenge e ogni porzione è identificata con un numero ben preciso.

- **[0] Fase di ETL** – inizia con il caricamento dei vari file **CSV** e **XLSX** per passare alla loro manipolazione, così da facilitare le successive fasi.
- **[1] Analisi 'normal e abnormal'** – consiste nel creare un classificatore che possa determinare situazioni normali e non, nella stessa porzione di codice è presente l'inferenza sui dati da classificare.
- **[2] Analisi 'anomaly e fault'** - consiste nel creare un classificatore che possa determinare situazioni di anomaly o fault, nella stessa porzione di codice è presente l'inferenza sui dati da classificare.
- **[3] Analisi 'position anomaly'** - consiste nel creare un classificatore che possa determinare la posizione dove si è verificata l'anomalia, nella stessa porzione di codice è presente l'inferenza sui dati da classificare.
- **[4] Analisi 'position fault'** - consiste nel creare un classificatore che possa determinare la posizione dove è verificato il fault, nella stessa porzione di codice è presente l'inferenza sui dati da classificare.
- **[5] Analisi 'percent opening fault'** - consiste nel creare un classificatore che possa determinare la percentuale di apertura delle elettrovalvole, nella stessa porzione di codice è presente l'inferenza sui dati da classificare.

Nei capitoli a seguire verranno descritti le fasi di analisi, la descrizione sarà dettagliata facendo riferimento al numero identificato della porzione di codice presente nello script Matlab.

Per ogni capitolo verrà descritto il compito svolto e le decisioni prese, per facilitare la spiegazione saranno usati immagini e tabelle riassuntive.

ETL

Fase identificata con id [0].

[0.1] caricamento csv e xlsx

Vengono caricati i file XLSX: **labels.xlsx** e **labels_spacecraft.xlsx** corrispettivi ai datasets contenenti info per il training e info degli elementi che andranno classificati.

A seguire vengono importati i vari CSV. Sono corrispettivamente 177 elementi per il Training e 45 per l'inferenza.

labels.xlsx è così strutturato:

- **Case** si riferisce ad un identificato per ogni singola misurazione
- **Spacecraft** fa riferimento alla navicella specifica
- **Condition** riporta se si tratta di una condizione di Normalità o Non-normalità
- **Solenoid valves opening ratio** riporta la percentuale di apertura delle singole elettrovalvole, in base al grado apertura è possibile capire quale valvola è mal funzionante
- **Bubble** riporta per ogni BP1-BP7 e BV1 la presenza di una bolla nella posizione specifica nella rete

labels_spacecraft.xlsx è così strutturato:

- **Case** come nel caso precedente riporta un identificato
- **Spacecraft** come nel caso precedente riporta la navicella di riferimento

La struttura dei CSV risulta identica per ogni file, vengono riportate le misure di 7 pressioni per ognuna delle P1-P7 (controllare Figura1) ad intervalli regolari di 100ms per un arco temporale di 1200ms.

[0.2] creazione tabelle train e inferenza

Considerando le strutture dei file importati si è pensato di integrare i XLSX con le misure delle pressioni riportate nei CSV. Si è proceduto alla creazione della tabella **table_labels_train** e **table_labels_test**.

[0.3] correzione tabella **table_labels_train**

[0.3.1] creazione e aggiunta colonna 'normal_abnormal'

Conoscendo il tipo di classificazione che si vuole eseguire per l'anomaly detection, si è pensato di tradurre la colonna **Condition** in una nuova chiamata **normal_abnormal** dove sono riportati con 0 i casi **normal** e 1 le situazioni di fault o anomaly (**abnormal**).

[0.3.2] creazione e aggiunta colonna 'anomaly_fault'

Per quanto riguarda i casi di fault e anomaly si fa ancora riferimento alla colonna **Condition** (di **table_labels_train**) per ottenere una nuova colonna che sintetizzasse i casi di **fault** e **anomaly** ed inoltre è stata aggiunta **anomaly_fault** (a **table_labels_train**) dove:

- I casi di **normal** sono associati al numero 0
- I casi di **fault** sono associati al numero 3

- I casi di **anomaly** sono associati al numero 2

[0.3.3] creazione e aggiunta colonna 'anomaly_position_BP_BV'

Per quanto riguarda i casi di fault e anomaly si voluto usare la colonna **Bubble** (di table_labels_train) per ottenere una nuova colonna che sintetizzasse i casi di anomalie e la loro posizione.

E' stata aggiunta la colonna **anomaly_position_BP_BV** (a table_labels_train) dove:

- I casi di **BV1** sono associati al numero 8
- I casi di **BP1** sono associati al numero 1
- I casi di **BP2** sono associati al numero 2
- I casi di **BP3** sono associati al numero 3
- I casi di **BP4** sono associati al numero 4
- I casi di **BP5** sono associati al numero 5
- I casi di **BP6** sono associati al numero 6
- I casi di **BP7** sono associati al numero 7
- I casi senza anomalie sono associati al numero 0

[0.3.4] creazione e aggiunta colonna 'fault_position_SV'

Per quanto riguarda i casi di fault e anomaly si voluto usare la colonna **Solenoid valves opening ratio** (di table_labels_train) per ottenere una nuova colonna che sintetizzasse i casi di fault e la loro posizione.

E' stata aggiunta **fault_position_SV** (a table_labels_train) dove:

- I casi di **SV1** sono associati al numero 1
- I casi di **SV2** sono associati al numero 2
- I casi di **SV3** sono associati al numero 3
- I casi di **SV4** sono associati al numero 4
- I casi senza fault sono associati al numero 0

[0.3.5] creazione e aggiunta colonna 'fault_position_SV_percent'

Per quanto riguarda i casi di fault e anomaly si voluto usare la colonna **Solenoid valves opening ratio** (di table_labels_train) per ottenere una nuova colonna che riportasse per ogni fault la percentuale di apertura della elettrovalvola di riferimento.

E' stata aggiunta **fault_position_SV_percent** (a table_labels_train) dove:

- I casi di **0%** sono associati al numero 0
- I casi di **25%** sono associati al numero 25
- I casi di **50%** sono associati al numero 50
- I casi di **75%** sono associati al numero 75
- I casi senza fault sono associati al numero 100


[0.4] tabelle finali per training

Per facilitare l'estrazione e la selezione dei features si è partiti dalla tabella **table_labels_train** (dopo averla aggiornata come descritto in precedenza) estrapolando singole tabelle per ogni modello da costruire.

[0.4.1] selezione colonne per 'normal_abnormal'

La tabella **table_labels_train_normal_abnormal** contiene 3 colonne, ovvero l'identificativo dei singoli elementi, il riferimento ai segnali associati e una label per distinguere normal e abnormal.

A seguire sono riportate le prime 5 righe della tabella **table_labels_train_normal_abnormal**

 177x3 [table](#)


	1 Var1	2 normal_abnormal	3 Case_Table
1	1	0	1201x8 table
2	2	0	1201x8 table
3	3	0	1201x8 table
4	4	0	1201x8 table
5	5	0	1201x8 table

Figura 3 esempio tabella table_labels_train_normal_abnormal

[0.4.2] selezione colonne per 'anomaly_fault'

La tabella **table_labels_train_anomaly_fault** contiene 3 colonne, ovvero l'identificativo dei singoli elementi, il riferimento ai segnali associati e una label per distinguere anomaly e fault.

A seguire sono riportate le prime 5 righe della tabella **table_labels_train_anomaly_fault**

 72x3 [table](#)


	1 Var1	2 anomaly_fault	3 Case_Table
1	106	3	1201x8 table
2	107	3	1201x8 table
3	108	3	1201x8 table
4	109	3	1201x8 table
5	110	3	1201x8 table

Figura 4 esempio tabella table_labels_train_anomaly_fault

[0.4.3] selezione colonne per 'anomaly_position_BP_BV'

La tabella **table_labels_train_anomaly_position_BP_BV** contiene 3 colonne, ovvero l'identificativo dei singoli elementi, il riferimento ai segnali associati e una label per distinguere la presenza di una bolla.

A seguire sono riportate le prime 5 righe della tabella **table_labels_train_anomaly_position_BP_BV**

 24x3 [table](#)


	1 Var1	2 anomaly_position_BP_BV	3 Case_Table
1	154		1 1201x8 table
2	155		2 1201x8 table
3	156		3 1201x8 table
4	157		4 1201x8 table
5	158		5 1201x8 table

Figura 5 esempio tabella table_labels_train_anomaly_position_BP_BV

[0.4.4] selezione colonne per 'fault_position_SV'

La tabella **table_labels_train_fault_position_SV** contiene 3 colonne, ovvero l'identificativo dei singoli elementi, il riferimento ai segnali associati e una label per distinguere la posizione dove si è verificato il fault.

A seguire sono riportate le prime 5 righe della tabella **table_labels_train_fault_position_SV**

 48x3 [table](#)

	1 Var1	2 fault_position_SV	3 Case_Table
1	106		1 1201x8 table
2	107		1 1201x8 table
3	108		1 1201x8 table
4	109		1 1201x8 table
5	110		2 1201x8 table

Figura 6 esempio tabella table_labels_train_fault_position_SV

[0.4.5] selezione colonne per 'fault_position_SV_percent'

La tabella **table_labels_fault_position_SV_percent** contiene 3 colonne, ovvero l'identificativo dei singoli elementi, il riferimento ai segnali associati e una label per con la percentuale di apertura delle elettrovalvole.

A seguire sono riportate le prime 5 righe della tabella **table_labels_train_fault_position_SV_percent**

48x3 [table](#)

	1 Var1	2 fault_position_SV_percent	3 Case_Table
1	106	0	1201x8 table
2	107	25	1201x8 table
3	108	50	1201x8 table
4	109	75	1201x8 table
5	110	0	1201x8 table

Figura 7 esempio tabella table_labels_train_fault_position_SV_percent

Estrazione e selezione features [*Diagnostic Features Design*] e creazione modello [*Classifier Lerner*]

Una volta ottenute le tabelle contenenti le informazioni necessarie, si è passati all'estrazione e selezione dei features. Per questa operazione si è utilizzato l'app fornita da matlab ***Diagnostic Features Design [DFD]***.

Possiamo iniziare dicendo, che vedendo come sono strutturati i segnali delle pressioni, si nota una periodicità di 400ms ripetuta per 3 volte, quindi per ogni sessione DFD si è sempre lavorato con una **FRAME POLICY** che utilizzasse un **frame size** e **rate pari a 400ms**, così da seguire l'andamento della periodicità del segnale.

Inoltre, per l'estrazione dei features si è lavorato nel dominio del tempo e nel dominio delle frequenze. Nel dominio del tempo si è deciso di estrarre le **IMPULSIVE FEATURES** così da avere le seguenti caratteristiche per ogni segnale.

- Crest Factor
- Impulsive Factor
- Clearance Factor
- Peak factor

Per il dominio delle frequenze si è prima ottenuto un modello **autoregressivo** (AR model) per ogni segnale per poi estrarre, considerando come 1 il numero massimo di picchi, le seguenti caratteristiche:

- Band Power
- Peak Amplitude
- Peak Frequency

In ogni sessione si sono analizzate i features considerando 3 metriche di ranking: **T-test**, **ROC** e **one-way ANOVA**. Per poi scegliere la features che avevano un punteggio migliore con un netto distacco dalle altre, o nel caso di situazione molto regolare di scegliere le prime 10 caratteristiche, per poi esportarle in **Classification Learner** per passare alla costruzione del classificatore.

A seguire verranno riportate i features che si è deciso selezionare e l'algoritmo utilizzato per la creazione del modello di classificazione. Riportando l'elenco dei features e le caratteristiche del modello.

Per la fase di validazione e testing si è fatto affidamento alle utility messe a disposizione da Classifier Learner, per validazione si è scelto una **Cross-Validation** con 5 folds e come testing si è deciso di usare una **Hold-Out** con un 10% per il test.

La scelta sul modello da utilizzare è stata fatta considerando la sua percentuale di accuratezza in test, la sua matrice di confusione e curva di ROC.

Features normal e abnormal

I features selezionate considerando le prime 10 su ranking T-test:

	SEGNALE IN TEMPO O SPETTRO	FEATURES
1	Case_Table_sigstats	ClearanceFactor
2	Case_Table_sigstats	CrestFactor
3	Case_Table_sigstats	ImpulseFactor
4	Case_Table_sigstats	ClearanceFactor
5	Case_Table_ps_spec	PeakValue
6	Case_Table_ps_1_spec	BandPower
7	Case_Table_ps_4_spec	BandPower
8	Case_Table_ps_4_spec	ClearanceFactor
9	Case_Table_ps_6_spec	PeakAmp1
10	Case_Table_ps_6_spec	PeakFreq1

Per maggiori dettagli consultare ***features_train_normal_abnormal.mat***

Caratteristiche modello di Classificazione

L'algoritmo selezionato per la creazione del classificatore è ***essemble boosted trees*** con una accuratezza in validazione del 91% e in test del 90.6%.

Per maggiori dettagli consultare ***classification_train_normal_abnormal_T-test.mat***

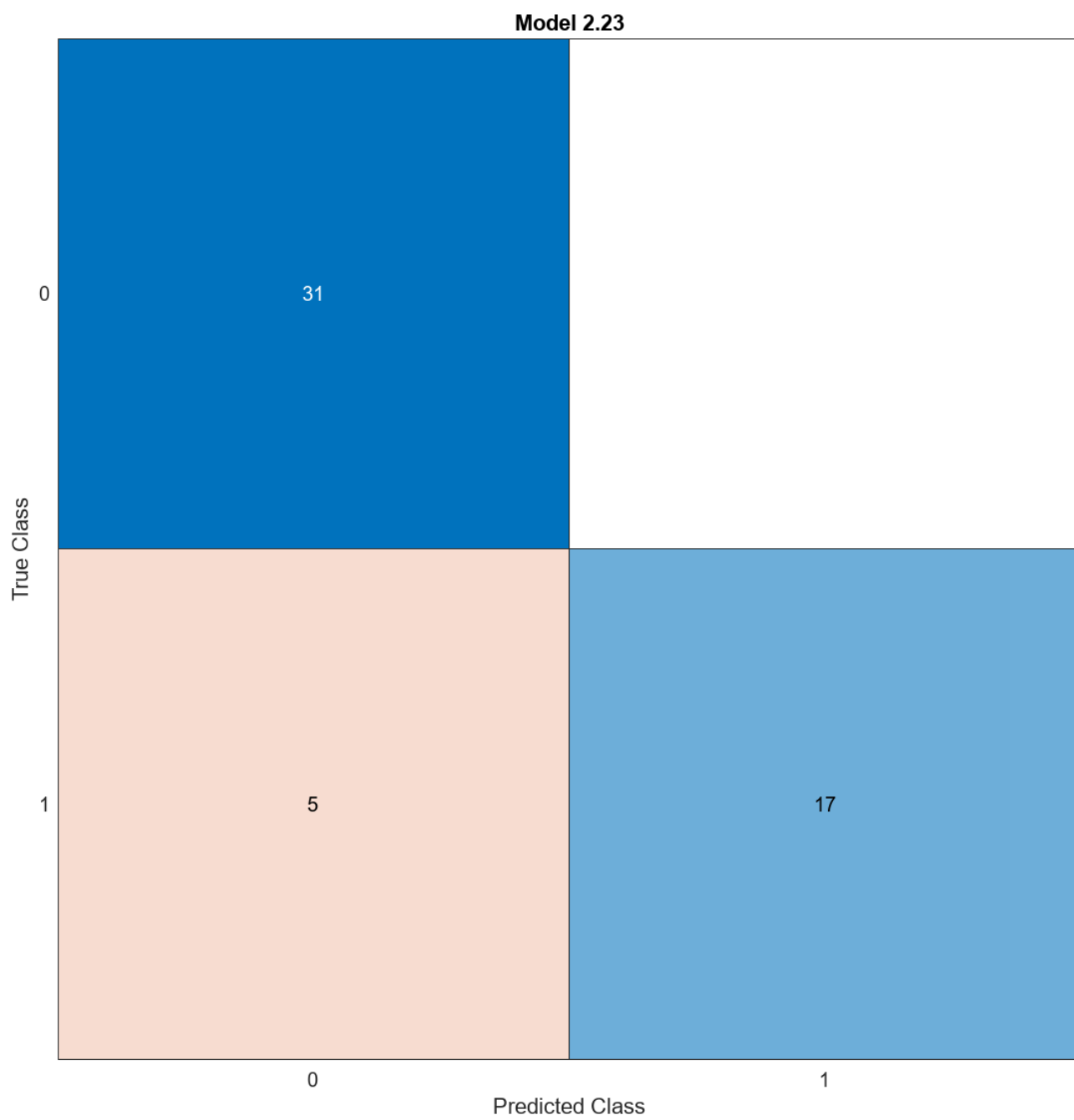


Figura 8 matrice confusione modello per normal abnormal

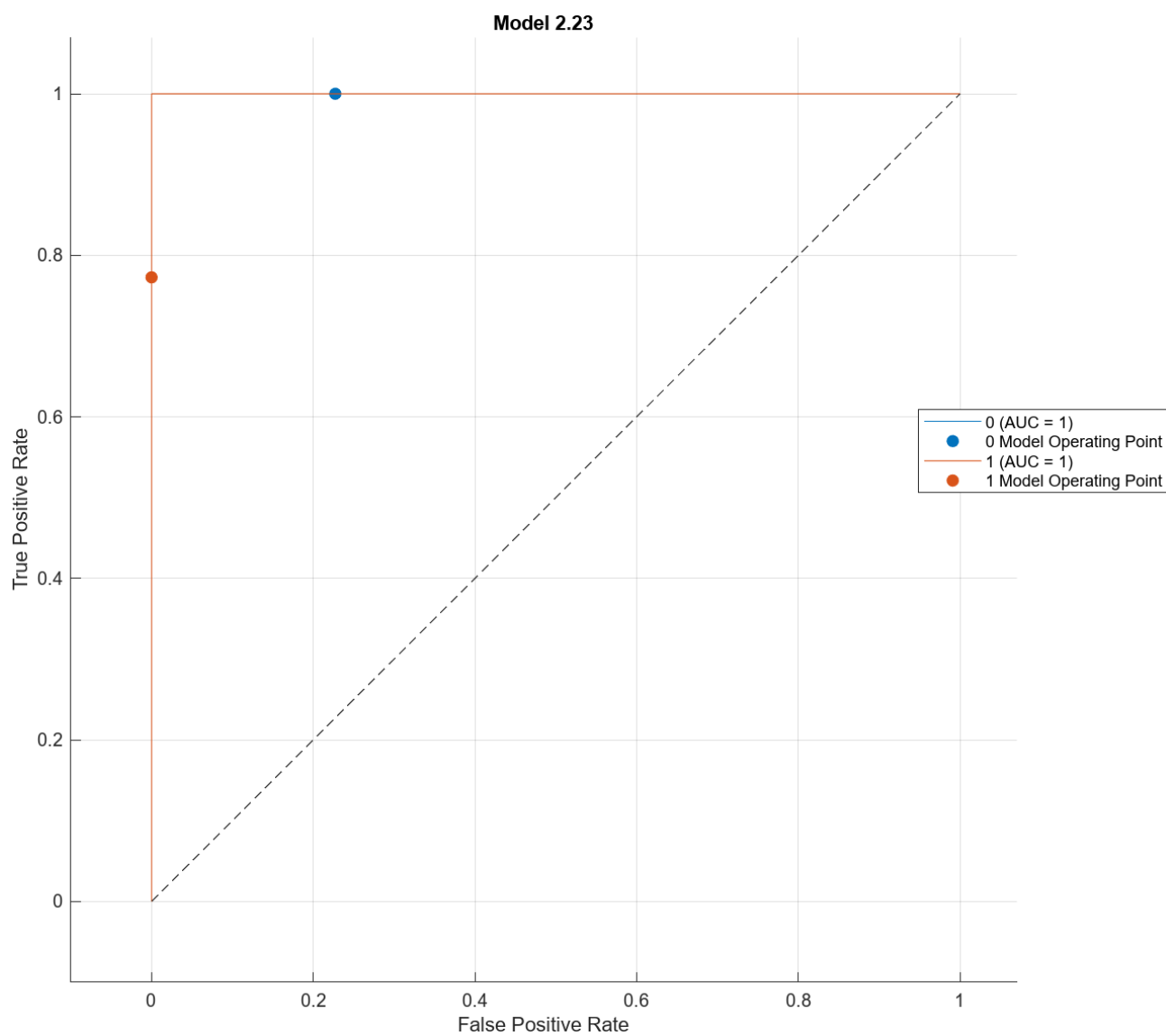


Figura 9 ROC per modello normal e abnormal

Features anomaly e fault

I features selezionate considerando le prime 10 su ranking T-test:

	SEGNALE IN TEMPO O SPETTRO	FEATURES
1	<i>Case_Table_ps_4_spec</i>	BandPower
2	<i>Case_Table_ps_4_spec</i>	PeakFreq1
3	<i>Case_Table_ps_spec</i>	BandPower
4	<i>Case_Table_sigstats</i>	ClearanceFactor
5	<i>Case_Table_sigstats</i>	CrestFactor
6	<i>Case_Table_sigstats</i>	ImpulseFactor
7	<i>Case_Table_sigstats</i>	PeakValue
8	<i>Case_Table_sigstats_6</i>	ClearanceFactor
9	<i>Case_Table_sigstats_6</i>	CrestFactor
10	<i>Case_Table_sigstats_6</i>	ImpulseFactor

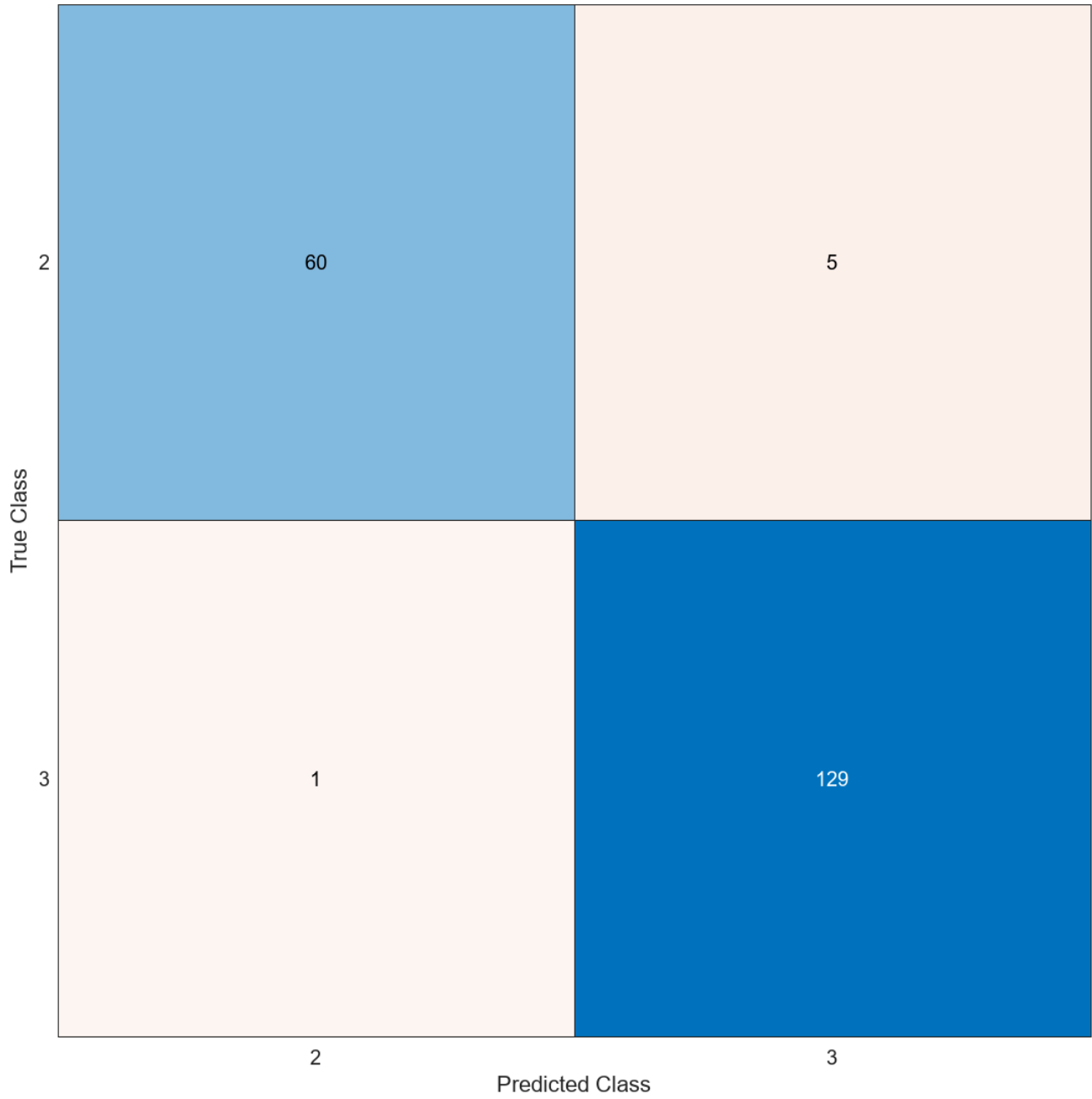
Per maggiori dettagli consultare ***features_train_anomaly_fault.mat***

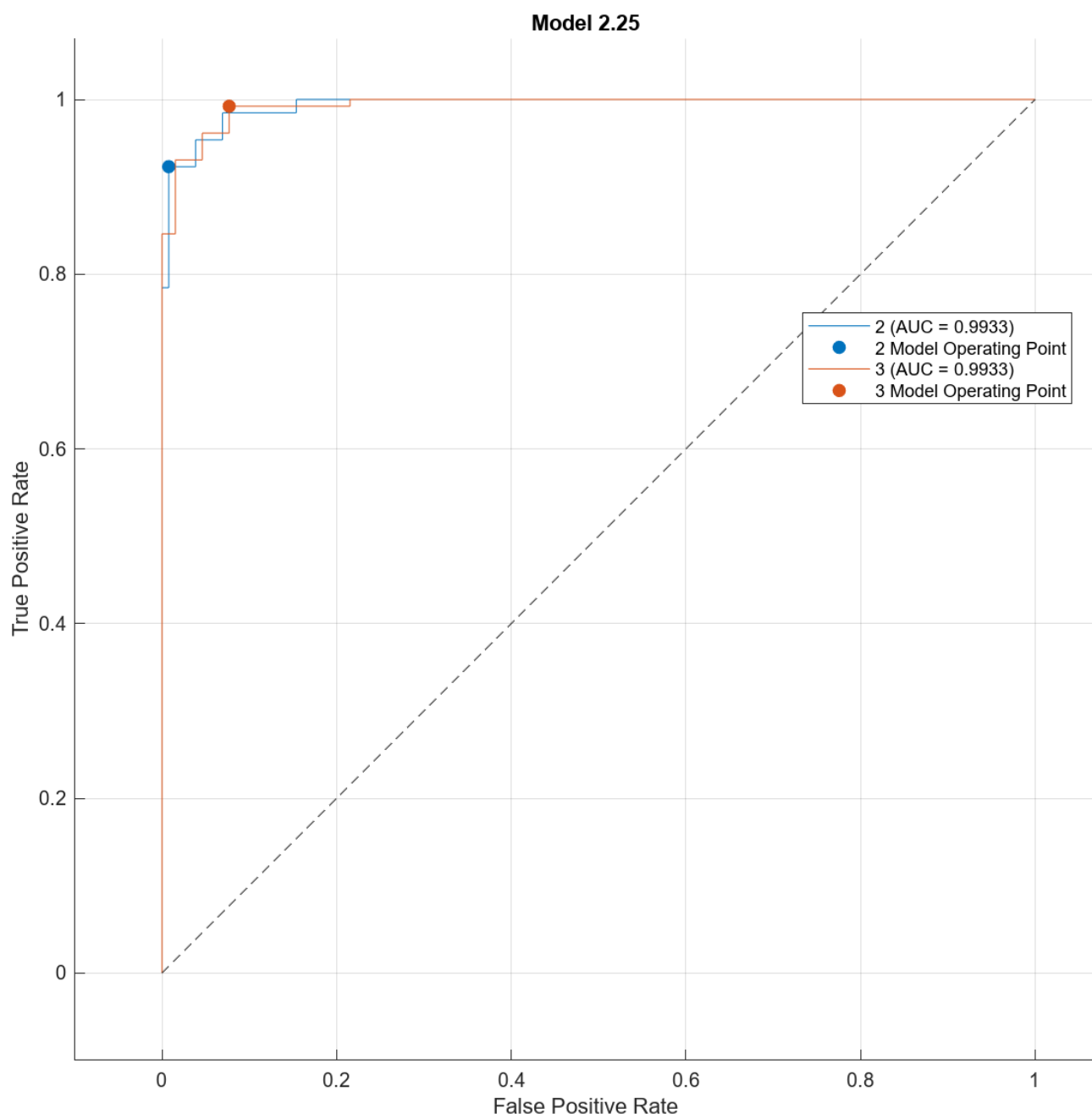
Caratteristiche modello di Classificazione:

L'algoritmo selezionato per la creazione del classificatore è ***essemble subspace discriminant*** con una accuratezza in validazione del 96% e in test del 95.2%.

Per maggiori dettagli consultare ***classification_train_anomaly_fault_T-test.mat***

Model 2.25





Features position anomaly

I features selezionate considerando le prime 10 su ranking ANOVA:

	SEGNALE IN TEMPO O SPETTRO	FEATURES
1	Case_Table_ps_2_spec	PeakFreq1
2	Case_Table_ps_4_spec	BandPower
3	Case_Table_ps_6_spec	PeakAmp1
4	Case_Table_ps_6_spec	PeakFreq1
5	Case_Table_ps_spec	BandPower
6	Case_Table_ps_spec	PeakAmp1
7	Case_Table_sigstats_6	ClearanceFactor
8	Case_Table_sigstats_6	CrestFactor
9	Case_Table_sigstats_6	ImpulseFactor
10	Case_Table_sigstats_6	PeakValue

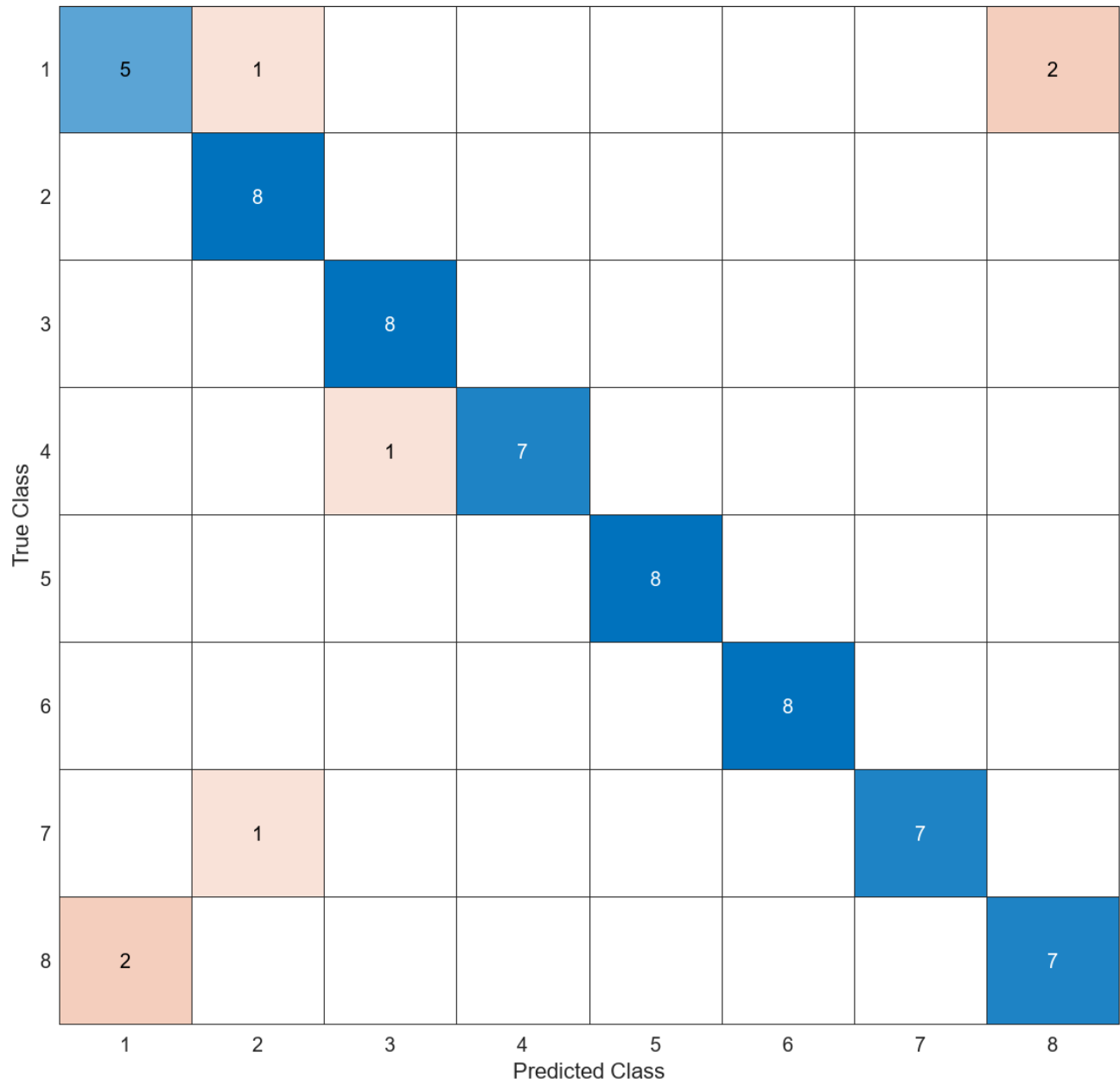
Per maggiori dettagli consultare ***features_train_anomaly_position_BP_BV.mat***

Caratteristiche modello di Classificazione:

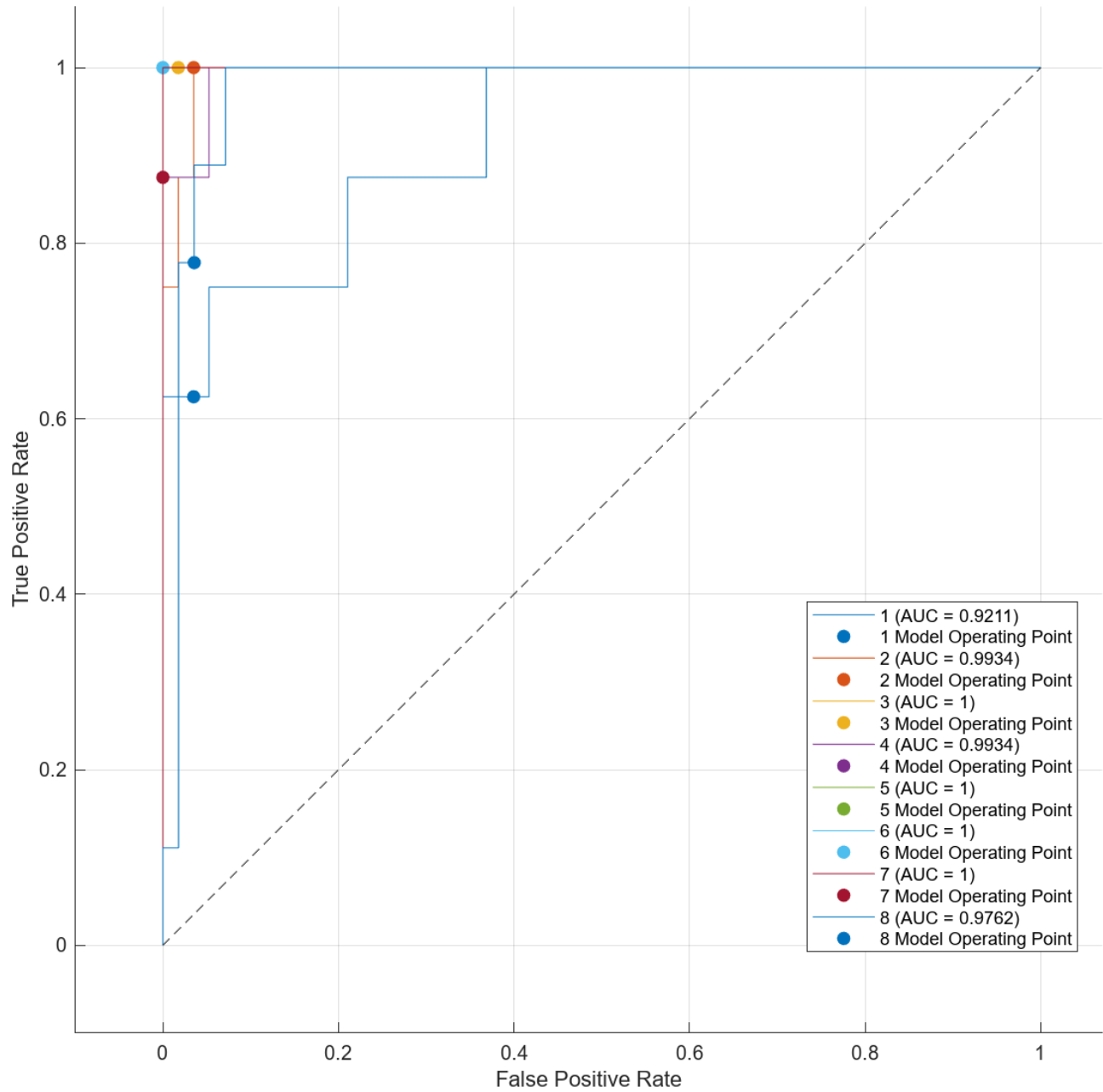
L'algoritmo selezionato per la creazione del classificatore è ***essemble bagged trees*** con una accuratezza in validazione del 89.2% e in test del 85.7%.

Per maggiori dettagli consultare ***classification_train_anomaly_position_BP_BV_ANOVA.mat***

Model 2.23



Model 2.23



Features position fault

I features selezionate considerando le prime 10 su ranking Anova:

	SEGNALE IN TEMPO O SPETTRO	FEATURES
1	Case_Table_ps_2_spec	PeakFreq1
2	Case_Table_ps_3_spec	PeakFreq1
3	Case_Table_sigstats_5	ClearanceFactor
4	Case_Table_sigstats_5	CrestFactor
5	Case_Table_sigstats_5	ImpulseFactor
6	Case_Table_sigstats_5	PeakValue
7	Case_Table_sigstats_6	ClearanceFactor
8	Case_Table_sigstats_6	CrestFactor
9	Case_Table_sigstats_6	ImpulseFactor
10	Case_Table_sigstats_6	PeakValue

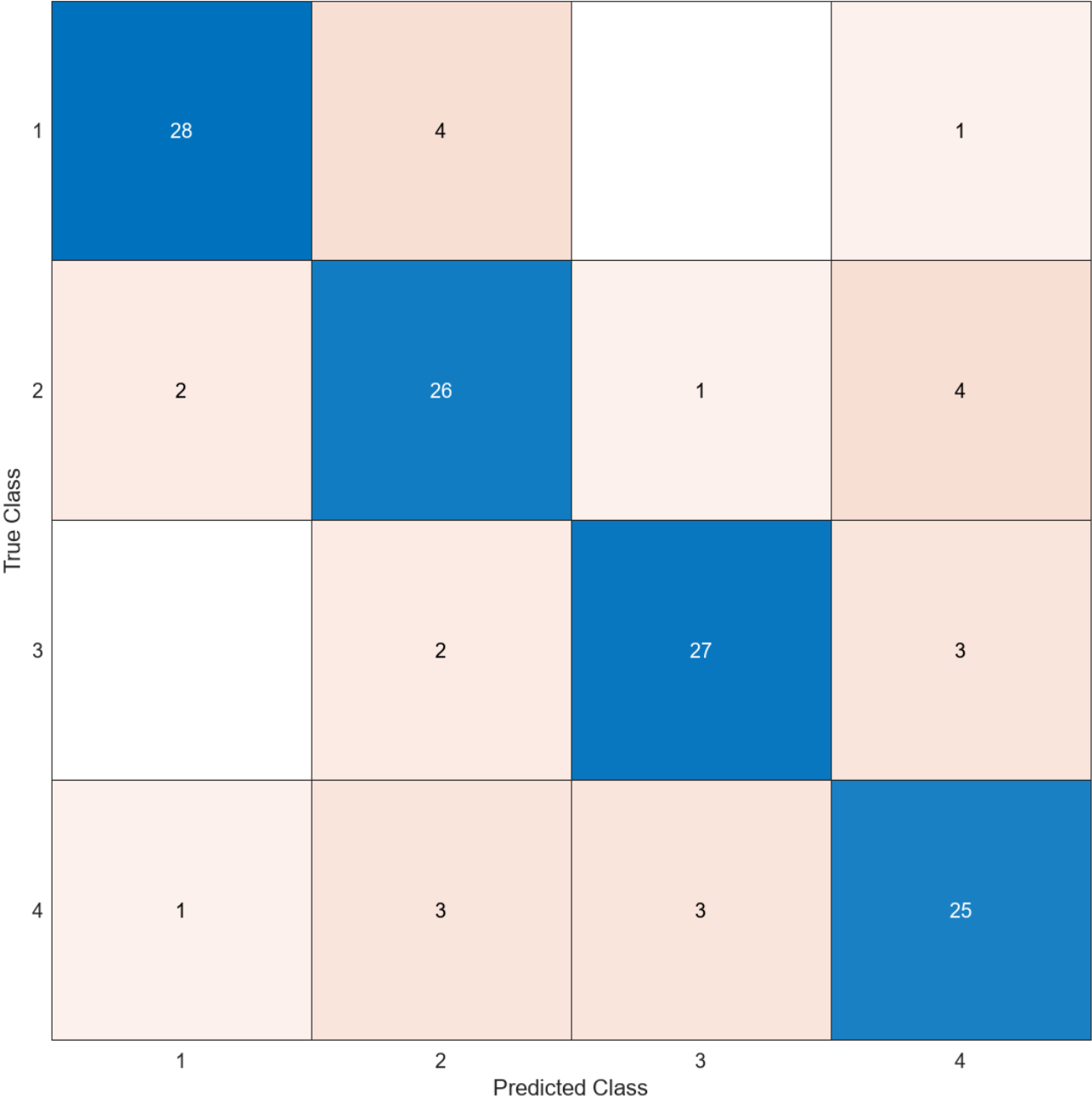
Per maggiori dettagli consultare ***features_train_fault_position_SV.mat***

Caratteristiche modello di Classificazione:

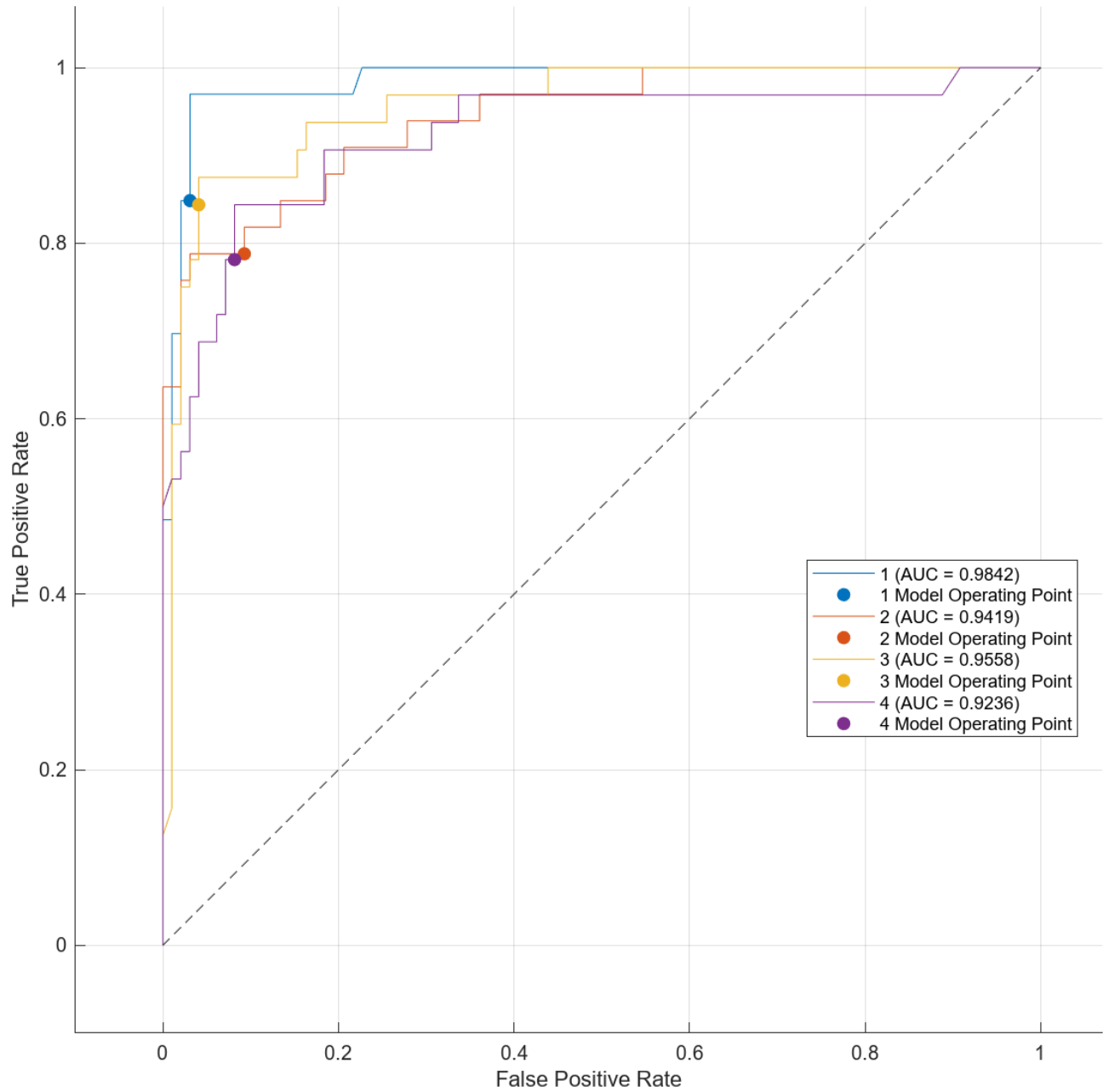
L'algoritmo selezionato per la creazione del classificatore è ***essemble boosted trees*** con una accuratezza in validazione del 81.5% e in test del 92.3%.

Per maggiori dettagli consultare ***classification_train_fault_position_SV_ANOVA.mat***

Model 2.22



Model 2.22



Features percent opening fault

I features selezionate considerando le prime 10 su ranking Anova:

	SEGNALE IN TEMPO O SPETTRO	FEATURES
1	Case_Table_ps_1_spec	BandPower
2	Case_Table_ps_6_spec	PeakAmp1
3	Case_Table_sigstats	ClearanceFactor
4	Case_Table_sigstats	CrestFactor
5	Case_Table_sigstats	ImpulseFactor
6	Case_Table_sigstats	PeakValue
7	Case_Table_sigstats_1	ClearanceFactor
8	Case_Table_sigstats_1	ImpulseFactor
9	Case_Table_sigstats_1	PeakValue
10	Case_Table_sigstats_4	ClearanceFactor

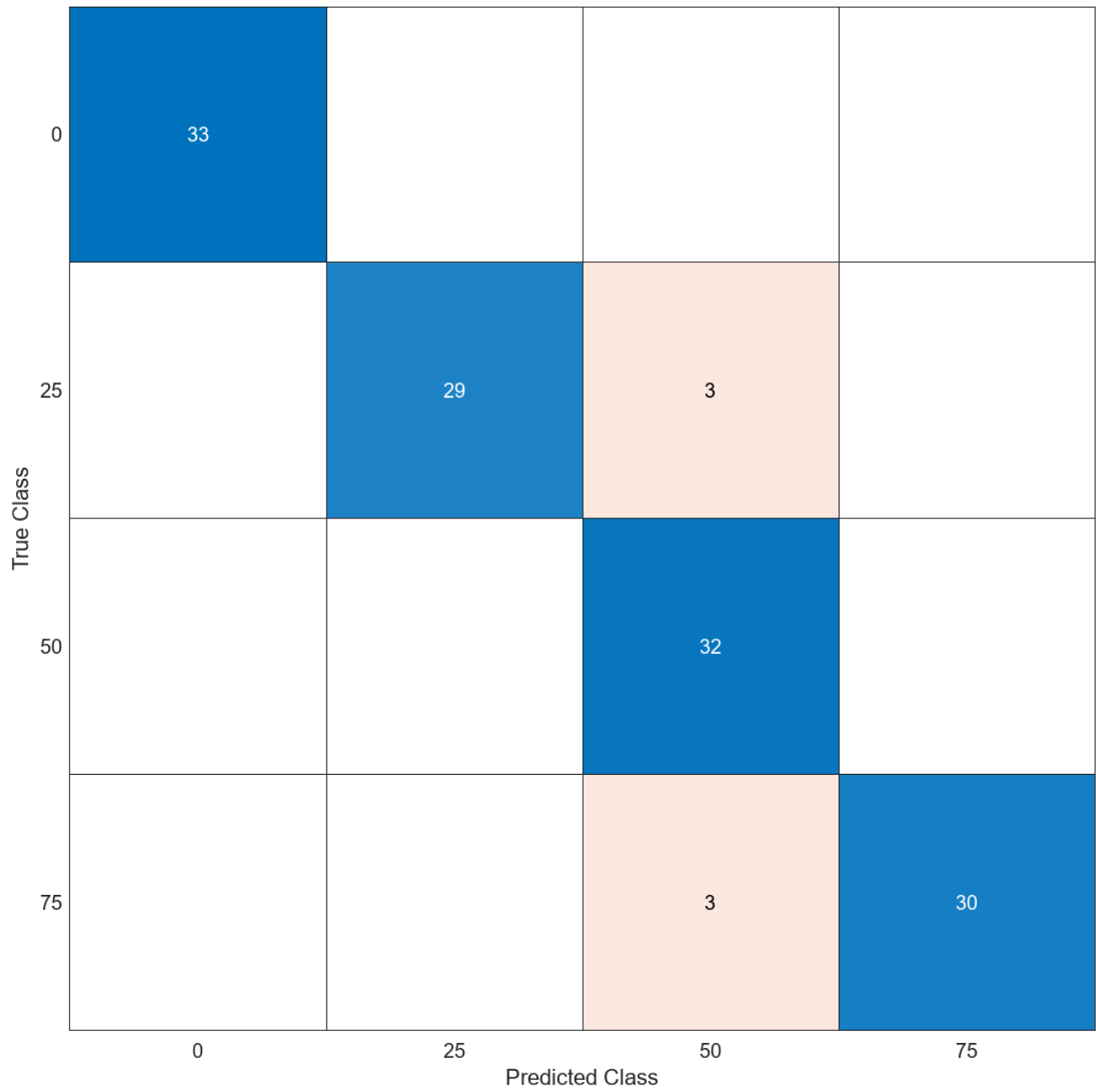
Per maggiori dettagli consultare ***features_train_fault_position_SV_percent.mat***

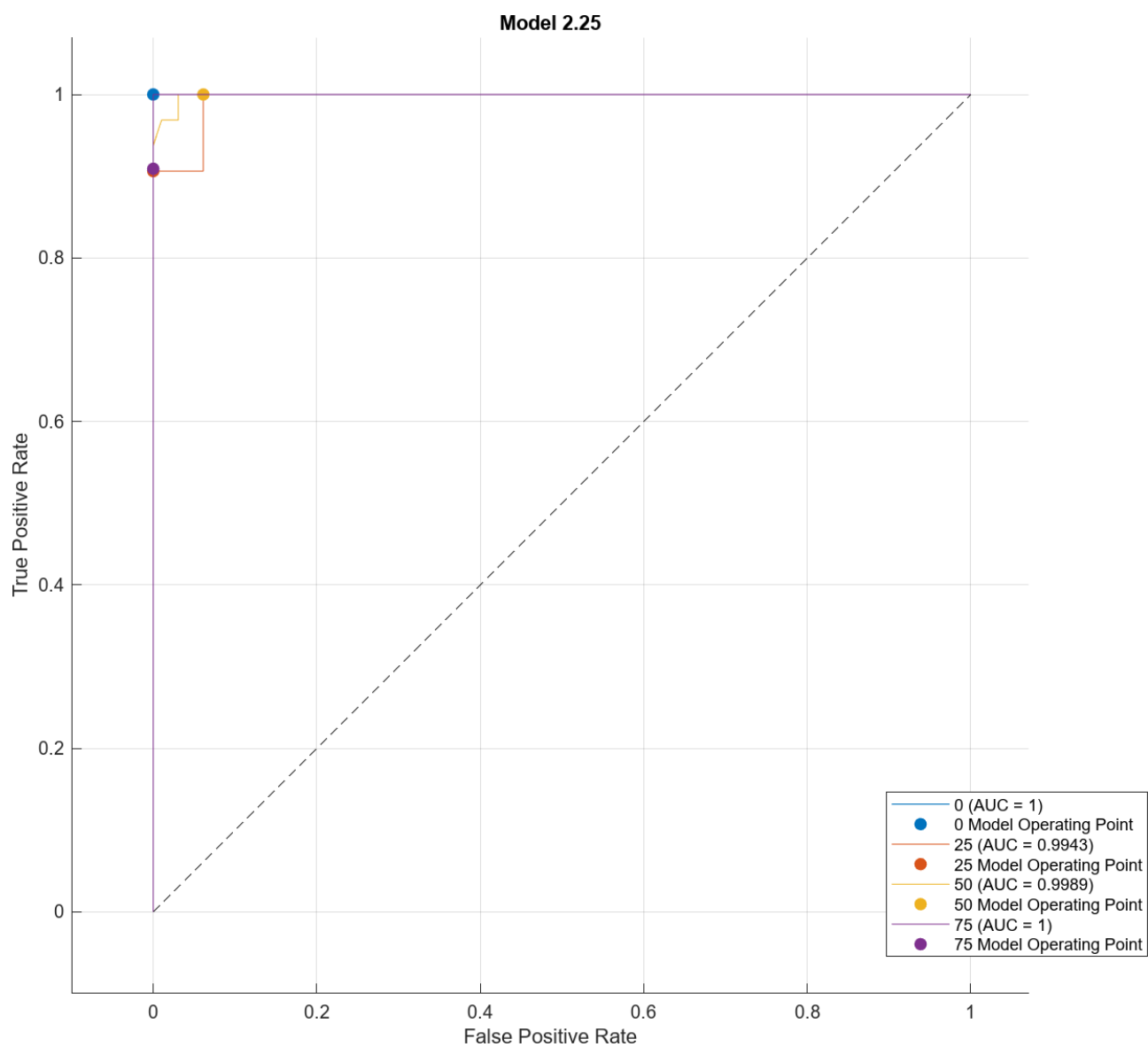
Caratteristiche modello di Classificazione:

L'algoritmo selezionato per la creazione del classificatore è ***essemble boosted trees*** con una accuratezza in validazione del 91% e in test del 90.6%.

Per maggiori dettagli consultare ***classification_train_fault_position_percent_SV_ANOVA.mat***

Model 2.25





Inferenza

Una piccola delucidazione va fatta su come sono stati ottenuti le classi di appartenenza per ogni segnale di pressione.

Il modello determina la classe per ogni elemento che gli viene passato in input. Ma i dati che gli sono stati dati in pasto erano strutturati in modo da avere per ogni segnale di pressione 3 frame, quindi il modello ha restituito un esito per ogni frame e per ottenere un risultato per il singolo segnale si sono contati gli esiti dei corrispettivi frame.

Quindi, perché un segnale potesse essere associato ad una classe doveva avere i suoi 3 frame classificati come con la stessa classe, sennò verrà associato un valore di **unknown (-1)**.

Questa procedura è ripetuta per tutte le analisi effettuate.

[1] Analisi 'normal e abnormal'

Nella porzione di codice identificata con [1] è presenta l'analisi sulla **anomaly detection**, in particolare si è usato il classificatore ottenuto per distinguere tra in dati Test gli elementi **Normali** e **Abnormali**.

I risultati sono presenti nella tabella del workspace in matlab: **result_classification_normal_abnormal**

	1	2	3
	Case_	normal_abnormal	Case_Table
1	178	1	1201x8 table
2	179	1	1201x8 table
3	180	0	1201x8 table
4	181	0	1201x8 table
5	182	0	1201x8 table

Figura 10 esempio risultato classificazione normal_abnormal

[2] Analisi 'anomaly e fault'

Nella porzione di codice identificata con [2] è presenta l'analisi su anomaly e fault, in particolare si è usato il classificatore ottenuto per distinguere tra in dati Test gli elementi anomaly e fault.

I risultati sono presenti nella tabella del workspace in matlab: **result_classification_anomaly_fault**

	1	2	3
	Case_	anomaly_fault	Case_Table
1	178	2	1201x8 table
2	179	3	1201x8 table
3	184	3	1201x8 table
4	186	2	1201x8 table
5	187	3	1201x8 table

Figura 11 esempio risultato classificazione anomaly_fault

[3] Analisi 'position anomaly'

Nella porzione di codice identificata con [3] è presente l'analisi delle posizioni delle anomaly, in particolare si è usato il classificatore ottenuto per determinare le posizioni delle varie anomalie.

I risultati sono presenti nella tabella del workspace in matlab:

result_classification_anomaly_position_BP_BV

	1	2	3
	Case_	anomaly_position_BP_BV	Case_Table
1	178	-1	1201x8 table
2	186	6	1201x8 table
3	193	1	1201x8 table
4	196	4	1201x8 table
5	197	7	1201x8 table

Figura 12 esempio risultato classificazione anomaly_position_BP_BV

[4] Analisi 'position fault'

Nella porzione di codice identificata con [4] è presente l'analisi delle posizioni dei fault, in particolare si è usato il classificatore ottenuto per determinare le posizioni dei vari fault.

I risultati sono presenti nella tabella del workspace in matlab: **result_classification_fault_position_SV**

	1	2	3
	Case_	fault_position_SV	Case_Table
1	179	1	1201x8 table
2	184	3	1201x8 table
3	187	2	1201x8 table
4	188	1	1201x8 table
5	189	-1	1201x8 table

Figura 13 esempio risultato classificazione fault_position_SV

[5] Analisi 'percent opening fault'

Nella porzione di codice identificata con [5] è presenta l'analisi sulle percentuali di apertura delle elettrovalvole, in particolare si è usato il classificatore ottenuto per le percentuali di apertura.

I risultati sono presenti nella tabella del workspace in matlab:

result_classification_fault_position_SV_percent

	1 Case_	2 fault_position_SV_percent	3 Case_Table
1	179	25	1201x8 table
2	184	0	1201x8 table
3	187	75	1201x8 table
4	188	25	1201x8 table
5	189	75	1201x8 table

Figura 14 esempio risultato classificazione fault_position_SV_percent

Risultati

I risultati ottenuti sono consultabili dallo script Matlab facendo riferimento alle corrispettive tabelle.

Ora si vuole dare un resoconto generale degli esiti ottenuti, procedendo in ordine;

Risultati normal_abnormal

D'analisi di **anomaly detection** possiamo notare che tra i 46 elementi classificati, si hanno 36 situazioni di Abnormal e 10 Normal, si può notare come la maggior parte degli elementi siano considerati come anormali, ma era un risultato atteso sapendo che la challenge era incentrata sul trovare situazioni malevoli.

Risultati anomaly o fault

Dai 36 elementi anormali estratti precedentemente, si è passati all'individuare **anomaly** e **fault**, ottenendo 10 anomaly, 11 fault e 15 sconosciuti (unknown -1). Per semplificare l'analisi non si è andato nel dettaglio per discriminare i singoli -1 ma ci si è concentrati sulle situazioni ben definite.

Risultati anomaly position

Dai risultati sulla **posizione** nella rete dove si sono verificate le **anomalie**, si riesce a capire che si ha una sola situazione di incertezza (unknown -1) e si è riuscito ad ottenere una posizione chiara e inequivocabile ad ogni altro elemento.

Risultati fault position

Anche per quanto riguarda il determinare la **posizione dei fault** si è ottenuta un unico sconosciuto (unknown -1), ma per i restanti elementi si è riusciti a determinare una situazione chiara e inequivocabile.

Risultati fault position percent

In ultima analisi, si è ottenuto la classificazione per la **percentuale di apertura** delle elettrovalvole per le situazioni di fault. Si può notare come dagli 11 fault si sono riuscite ad ottenere situazioni chiare e senza indecisioni.

Conclusioni

Il progetto d'esame consisteva nel creare modelli intelligenti che potessero essere usati in secondo momento per classificare situazioni ben precise.

I passaggi che mi hanno portato ai risultati ottenuti sono descritti in questo documento, ponendo attenzione sui dettagli che ritengo di maggior importanza.