

*UNIVERSITÀ POLITECNICA DELLE MARCHE*

*FACOLTÀ DI INGEGNERIA*



---

*Corso di Laurea Magistrale in  
Ingegneria Informatica e dell'Automazione*

***Relazione tecnica Progetto Manutenzione Preventiva Per  
La Robotica e l'Automazione Intelligente***

Studenti:

DANIELE PALLINI 1107326  
MATTEO ABBRUZZETTI 1108842  
ANDREA GIULIANI 1114482

Docenti:

ALESSANDRO FREDDI

ANNO ACCADEMICO 2023-2024

# Indice

<b>1</b>	<b>Introduzione</b>	<b>3</b>
1.1	Obiettivi . . . . .	3
1.2	Dataset . . . . .	3
1.3	Organizzazione . . . . .	4
<b>2</b>	<b>Task 1: Distinzione dei casi in guasti e normali</b>	<b>6</b>
2.1	Preparazione del dataset . . . . .	6
2.2	Diagnostic Feature Designer . . . . .	6
2.3	Classification Learner . . . . .	8
2.4	Test e risultati . . . . .	9
<b>3</b>	<b>Task 2: Distinzione della tipologia di guasto</b>	<b>11</b>
3.1	Preparazione del dataset . . . . .	11
3.2	Diagnostic Feature Designer . . . . .	11
3.3	Classification Learner . . . . .	12
3.4	Test e risultati . . . . .	13
<b>4</b>	<b>Task 3: Individuazione delle locazioni dei guasti bubble anomaly</b>	<b>15</b>
4.1	Preparazione del dataset . . . . .	15
4.2	Diagnostic Feature Designer . . . . .	15
4.3	Classification Learner . . . . .	16
4.4	Test e risultati . . . . .	17
<b>5</b>	<b>Task 4: Individuazione delle valvole guaste</b>	<b>19</b>
5.1	Preparazione del dataset . . . . .	19
5.2	Diagnostic Feature Designer . . . . .	19
5.3	Classification Learner . . . . .	20
5.4	Test e risultati . . . . .	21
<b>6</b>	<b>Task 5: Determinazione delle percentuali di apertura delle valvole guaste</b>	<b>23</b>
6.1	Preparazione del dataset . . . . .	23
6.2	Diagnostic Feature Designer . . . . .	23
6.3	Regression Learner . . . . .	24
6.4	Test e risultati . . . . .	25
<b>7</b>	<b>Conclusioni</b>	<b>27</b>



Inoltre, è presente una tabella che contiene informazioni relative ai singoli casi, così strutturata:

- **Case:** Contiene il numero del caso a cui fanno riferimento le informazioni.
- **Spacecraft:** Contiene il numero dello Spacecraft a cui fanno riferimento le informazioni.
- **Condition:** Assume il valore *Normal* se il caso non è relativo a un guasto, *Fault* se è presente un *Solenoid Valve Fault*, *Anomaly* se è presente un guasto di tipo *Bubble Anomaly*.
- **Solenoid Valves Opening Ratio (%):** Contiene la percentuale di apertura delle valvole SV1 ÷ SV4. Per i casi Normal la percentuale è del 100%.
- **Bubble:** Indica se è presente o meno un guasto di tipo bubble anomaly su BP1 ÷ BP7, BV1.

Case#	Spacecraft#	Condition	Solenoid valves Opening Ratio /%				Bubble							
			SV1	SV2	SV3	SV4	BP1	BP2	BP3	BP4	BP5	BP6	BP7	BV1
1	1	Normal	100	100	100	100	No	No	No	No	No	No	No	No
2	1	Normal	100	100	100	100	No	No	No	No	No	No	No	No

Figura 3: Estratto Dataset

### 1.3 Organizzazione

Si è scelto di suddividere il lavoro in Tasks e di addestrare modelli di classificazione o regressione specifici, secondo il seguente schema:

- **Task 1:** Addestramento di un modello di classificazione per la distinzione dei casi in esame in *guasti* e *normali*.
- **Task 2:** Addestramento di un modello di classificazione per la distinzione dei casi rilevati come *guasti* (nel Task 1) in *bubble anomaly*, *solenoid valve fault* e *unknown fault*.
- **Task 3:** Addestramento di un modello di classificazione per individuare la locazione (BP1 ÷ BP7, BV1) del guasto nei casi identificati come *bubble anomaly* (nel Task 2).
- **Task 4:** Addestramento di un modello di classificazione per individuare la valvola guasta (SV1 ÷ SV4) nei casi identificati come *solenoid valve fault* (nel Task 2).

- **Task 5:** Addestramento di un modello di regressione per determinare la percentuale di apertura delle valvole guaste nei casi identificati come *solenoid valve fault* (nel Task 2).

## 2 Task 1: Distinzione dei casi in guasti e normali

Il Task 1 riguarda l'addestramento di un modello di classificazione per la distinzione dei casi in esame in *guasti* e *normali*.

### 2.1 Preparazione del dataset

Per il dataset di training sono stati considerati tutti i 177 casi disponibili. Essi sono divisi equamente tra i tre spacecrafts (59 casi ognuno). Ad ogni caso è stato associato il relativo Faultcode, che può assumere due diversi valori:

- **0**, se il caso corrisponde ad un funzionamento normale.
- **1**, se il caso corrisponde ad un guasto.

#### Struttura della tabella di training

La tabella di training (Figura 4) è quindi composta da tre colonne:

- **Casi**: identifica il caso in esame.
- **Rilevazioni**: contiene le rilevazioni effettuate sulla finestra temporale di 1.2 secondi.
- **FaultCode**: contiene il Faultcode associato.

	1 Casi	2 Rilevazioni	3 FaultCode
1	"Caso1"	1201x8 table	0
2	"Caso2"	1201x8 table	0

Figura 4: Estratto tabella di training Task 1

### 2.2 Diagnostic Feature Designer

La tabella di training è stata caricata nel tool di MATLAB *Diagnostic Feature Designer* impostando FaultCode come *Condition Variable*. Sono stati generati e osservati i grafici dei segnali relativi ai profili di pressione dei vari sensori.

#### Frame policies

Sono state individuate due configurazioni interessanti per le frame policies:

- **Open**: riguarda la fase di apertura delle valvole; è costituita da una frame size di 0.1s e un frame rate di 0.4s.

- **Open + Close:** considera l'intero ciclo di apertura e chiusura; è costituita da una frame size di 0.4s e un frame rate di 0.4s.

Entrambe sono state utilizzate per la generazione delle features, nel dominio del tempo e in frequenza. In questo modo, ogni caso è stato suddiviso in 3 frames e il calcolo delle features avviene su ognuno di essi.

### Generazione ed estrazione delle features

Per la generazione delle features nel tempo, di seguito sono elencate quelle selezionate:

- **Signal Features:** Mean, RMS, Standard Deviation (Std), Shape Factor, Kurtosis, Skewness, Crest Factor, Peak Value, Impulse Factor, Clearance Factor.
- **Time series Features:** Minimum, Median, Maximum, First Quartile (Q1), Third Quartile (Q3), IQR.
- **Model-based Features:** MSE, MAE, Akaike's Information Criterion (AIC), Residual Mean, Residual Variance, Residual RMS, Residual Kurtosis.

Per quanto riguarda le features in frequenza, sono stati generati gli spettri dei segnali per ogni sensore tramite *Welch's Method*, lasciando invariati i parametri di default. Successivamente, per ogni sensore sono state calcolate le features *Peak Amplitude*, *Peak Frequency* e *Band Power*, prendendo in considerazione gli intervalli di frequenza e i picchi più interessanti. Ad esempio, per il sensore P1 è stato considerato l'intervallo di frequenza  $40 \div 250$  Hz ed un numero di picchi pari a 2 (Figura 5).

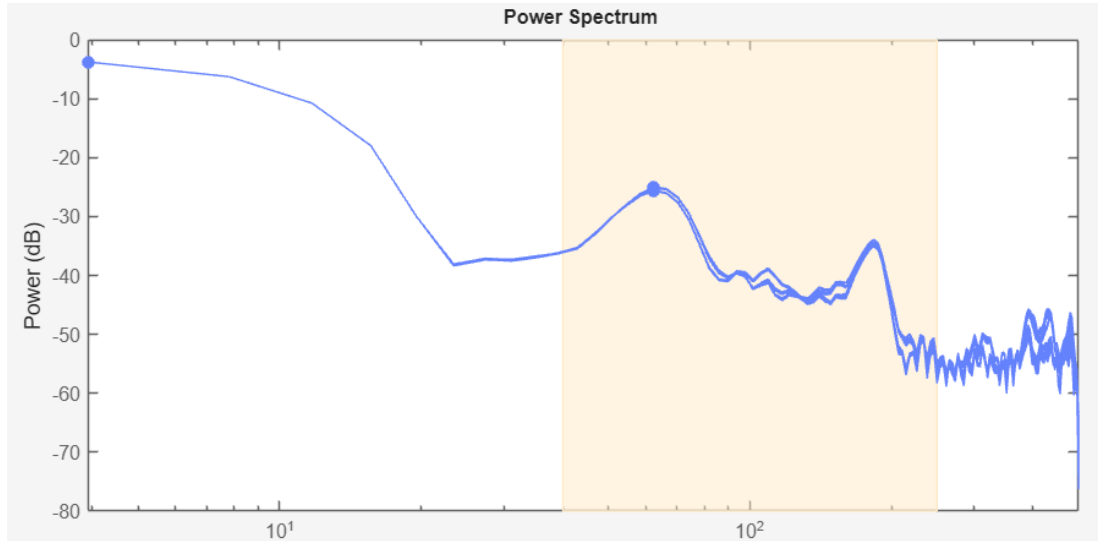


Figura 5: Esempio taglio frequenze per P1 (è evidenziata l'area selezionata)

Le features generate sono state ordinate in base al valore della metrica T-Test (Figura 6) e sono state estratte quelle dal valore più elevato. La scelta della frame policy è ricaduta sulla **Open + Close** in quanto i valori di T-Test risultavano essere più alti.

Feature	T-Test
Cases_res_2_tsmodel/AIC	25.6962
Cases_res_6_tsmodel/AIC	21.7788
Cases_res_2_sigstats/Std	17.3412
Cases_res_5_tsmodel/AIC	16.8937
Cases_res_2_sigstats/RMS	16.6082

Figura 6: Ranking features per T-Test Task 1

Le features nel dominio del tempo selezionate sono: **P1/RMS**, **P1/AIC**, **P2/Std**, **P2/MAE**, **P3/Std**, **P3/MAE**, **P4/PeakValue**, **P4/MAE**, **P5/RMS**, **P5/AIC**, **P6/Kurtosis**, **P6/MAE**, **P7/Kurtosis**, **P7/MAE**.

Quelle in frequenza sono invece: **P1/BandPower**, **P2/BandPower**, **P3/BandPower**.

Queste features sono state importate nel tool *Classification Learner* di MATLAB per l'addestramento di modelli di classificazione.

### 2.3 Classification Learner

Nel tool Classification Learner è stata importata la tabella di training: ogni riga corrisponde ad un frame e contiene le features associate ed il Faultcode corrispondente. In questa operazione sono stati rimossi i campi relativi all'identificatore del caso e quelli riguardanti l'inizio e la fine del frame. Si avranno quindi 3 occorrenze per ogni caso, per un totale di 531 righe. Al fine di mitigare l'overfitting si è scelto di applicare una 10-Fold Cross-Validation e di riservare il 20% del dataset alla validation. La predizione del modello equivale a **0** se il frame analizzato corrisponde ad un funzionamento normale; al contrario, equivale ad **1** se relativo ad un guasto. Sono stati addestrati tutti i modelli disponibili tramite la funzione *All Train* e selezionato il modello **Medium Tree** (Figura 7), il migliore in base alle metriche *Accuracy* (10-Fold Cross-Validation) e *Accuracy* (Validation sul 20% dei dati di training).



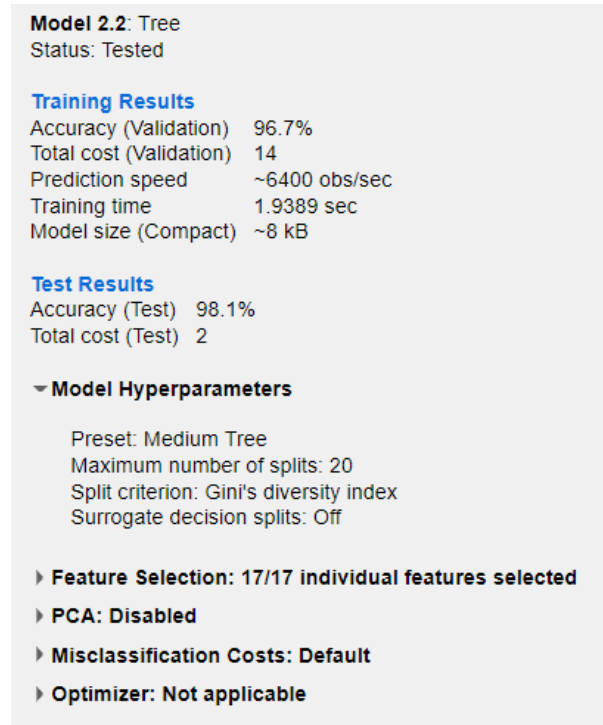


Figura 7: Parametri Medium Tree

## 2.4 Test e risultati

Al fine di testare il modello selezionato, sono stati preparati i 46 casi di test a disposizione attraverso la costruzione di una tabella contenente le features necessarie al modello di classificazione. Per costruzione, il modello effettua la predizione sul singolo frame. Si è reso dunque necessario raggruppare i risultati per caso. In presenza di discordanze tra predizioni di frame appartenenti allo stesso caso è stata adottata la politica Worst case, andando a classificare il caso come guasto se almeno uno dei frame corrispondenti è stato predetto come **1**. Per valutare le prestazioni del modello, sono state confrontate le predizioni con i Faultcode reali presi dal file **answer.csv**. La matrice di confusione ottenuta è la seguente:

	True 0	True 1
Predicted 0	20	0
Predicted 1	0	26

Figura 8: Matrice di confusione Task 1

Come si può notare, il modello predice correttamente tutti i casi di test; le metriche *Accuracy*, *Precision*, *Recall*, *F1 Score* e *Specificity* sono tutte pari al 100%. I 26 casi identificati come guasti sono passati al Task 2, per la distinzione nei 3 tipi possibili.

### 3 Task 2: Distinzione della tipologia di guasto

Il Task 2 riguarda l'addestramento di un modello di classificazione per la distinzione dei casi rilevati come guasti in *bubble anomaly*, *solenoid valve fault* e *unknown fault*. Si è pensato di addestrare un modello di classificazione binaria per la distinzione nei primi due guasti citati; in seguito si procederà ad analizzare e ad interpretare i risultati in modo da individuare i guasti di tipo unknown, casistica non presente nei dati di training.

#### 3.1 Preparazione del dataset

Per il dataset di training sono stati considerati esclusivamente i 72 casi guasti. Ad ogni caso è stato associato il relativo Faultcode, che può assumere due diversi valori:

- **1**, se il caso corrisponde ad un guasto di tipo *solenoid valve fault*.
- **2**, se il caso corrisponde ad un guasto di tipo *bubble anomaly*.

La struttura della tabella di training è identica a quella del Task 1 (Sezione 2.1).

#### 3.2 Diagnostic Feature Designer

La tabella di training è stata caricata nel tool di MATLAB Diagnostic Feature Designer impostando FaultCode come Condition Variable. Le due configurazioni delle frame policies considerate sono le medesime presentate nel Task 1 (Sezione 2.2).

##### Generazione ed estrazione delle features

Per la generazione delle features nel tempo e in frequenza, è stata seguita la medesima procedura presentata nel Task 1 (Sezione 2.2). Le features generate sono state ordinate in base al valore della metrica T-Test (Figura 9) e sono state estratte quelle dal valore più elevato. La scelta della frame policy è ricaduta sulla **Open + Close** in quanto i valori di T-Test risultavano essere più alti.

Feature	T-Test
Cases_res_1_tsfeat/Q1	37.1471
Cases_res_4_tsfeat/IQR	37.1405
Cases_res_4_tsfeat/Q1	36.6084
Cases_res_6_tsfeat/Q1	36.1110
Cases_res_1_tsfeat/IQR	34.9483

Figura 9: Ranking features per T-Test Task 2

Le features nel dominio del tempo selezionate sono: **P1/Q1, P1/Q3, P1/IQR, P2/IQR, P2/Q1, P2/Q3, P3/IQR, P4/Std, P5/Q1, P5/Q3, P5/IQR, P6/Q1, P7/Q1**. In questo caso non sono state esportate features in frequenza, in quanto i valori di T-Test risultavano essere notevolmente più bassi.

### 3.3 Classification Learner

Nel tool Classification Learner è stata importata la tabella di training: ogni riga corrisponde ad un frame e contiene le features associate ed il Faultcode corrispondente. In questa operazione sono stati rimossi i campi relativi all'identificatore del caso e quelli riguardanti l'inizio e la fine del frame. Si avranno quindi 3 occorrenze per ogni caso, per un totale di 216 righe. Al fine di mitigare l'overfitting si è scelto di applicare una 10-Fold Cross-Validation e di riservare il 20% del dataset alla validation. La predizione del modello equivale a **1** se il frame analizzato corrisponde ad un guasto di tipo *solenoid valve fault*; al contrario, equivale ad **2** se relativo ad un guasto di tipo *bubble anomaly*. Sono stati addestrati tutti i modelli disponibili tramite la funzione *All Train* e selezionato il modello **Efficient Linear SVM** (Figura 10), il migliore in base alle metriche *Accuracy* (10-Fold Cross-Validation) e *Accuracy* (Validation sul 20% dei dati di training).

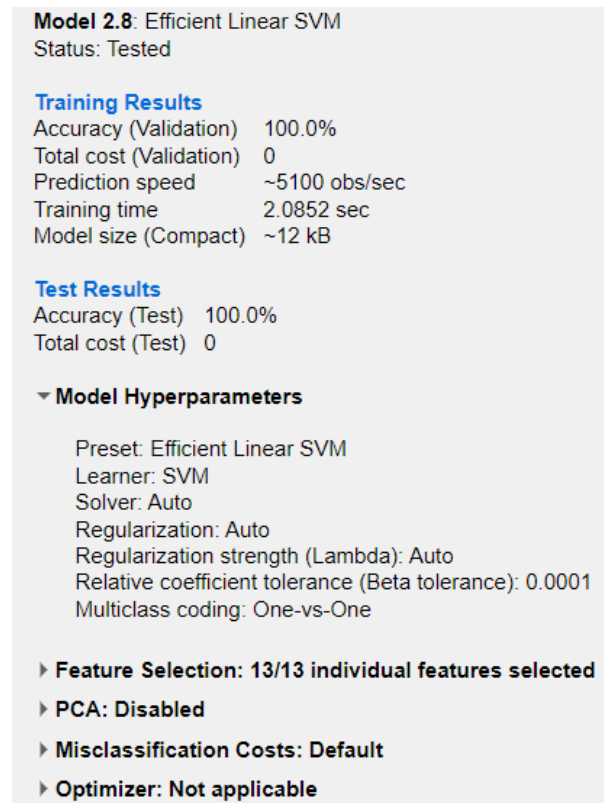


Figura 10: Parametri Efficient Linear SVM

### 3.4 Test e risultati

Al fine di testare il modello selezionato, sono stati preparati i 26 casi predetti come guasti nel Task 1 (Sezione 2.4) attraverso la costruzione di una tabella contenente le features necessarie al modello di classificazione. Per costruzione, il modello effettua la predizione sul singolo frame. Si è reso dunque necessario raggruppare i risultati per caso. In presenza di discordanze tra predizioni di frame appartenenti allo stesso caso, viene adottato il criterio di maggioranza.

#### Riconoscimento dei guasti di tipo unknown

Dal momento che i dati di training non contengono guasti di tipo unknown, è stato necessario definire un criterio per individuarli. Dato che il modello utilizzato per la classificazione è di tipo SVM, l'idea è quella di basarsi sulle distanze delle predizioni dall'iperpiano separatore per individuare gli outliers, che saranno identificati come guasti di tipo unknown. A questo proposito, è stato applicato il modello ai dati di training ed è stata proiettata la distribuzione delle distanze in un boxplot (Figura 11). Non essendo presenti casi guasti di tipo unknown, è possibile visualizzare le distanze medie dall'iperpiano e individuare quelle massime.

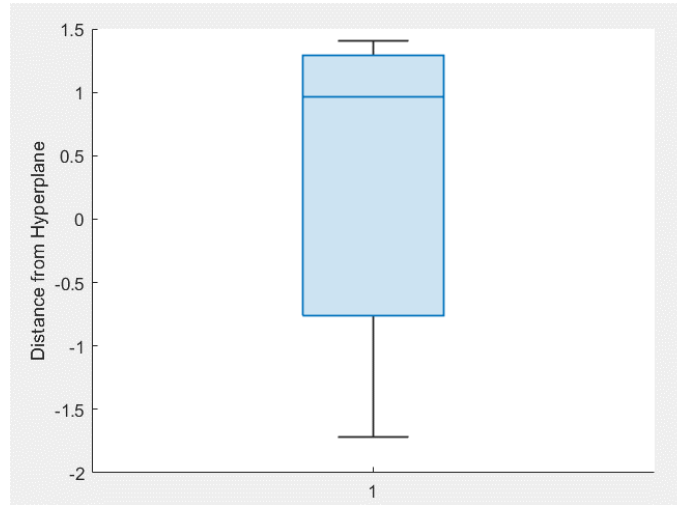


Figura 11: Boxplot casi di training

Come è possibile osservare dalla Figura 11, il valore della distanza massima equivale a circa 1.4 per i guasti di tipo solenoid valve fault (valori positivi), e a circa 1.7 per i guasti di tipo bubble anomaly (valori negativi). Si è quindi ritenuto il valore **2** una soglia sufficiente oltre la quale determinare i casi guasti di tipo unknown. Applicando il modello sui dati di test e integrando la soglia, il boxplot ottenuto è il seguente:

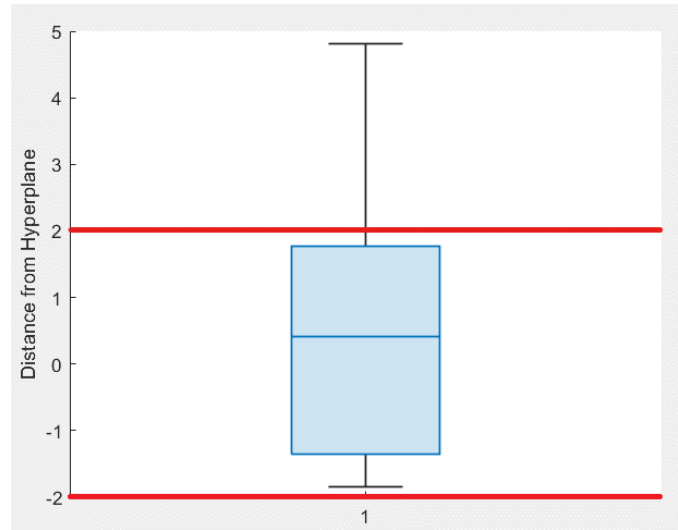


Figura 12: Boxplot casi di test

Si avranno quindi dei casi, predetti dal modello come guasti di tipo solenoid valve fault, le cui distanze sono maggiori del valore di soglia **2** e che dunque saranno classificati come casi guasti di tipo unknown.

## Risultati

La matrice di confusione che si ottiene è la seguente:

	True 1	True 2	True 3
Predicted 1	<b>10</b>	<b>0</b>	<b>0</b>
Predicted 2	<b>0</b>	<b>10</b>	<b>0</b>
Predicted 3	<b>0</b>	<b>0</b>	<b>6</b>

Figura 13: Matrice di confusione Task 2

Come si può notare, il modello predice correttamente tutti i casi di test; l'Accuracy è pari al 100%.

I 10 casi identificati come guasti di tipo bubble anomaly sono passati al Task 3 (per l'individuazione della locazione), mentre i 10 casi identificati come guasti di tipo solenoid valve fault sono passati ai Task 4 (per l'individuazione delle valvole guaste) e 5 (per la determinazione della percentuale di apertura delle valvole guaste).

## 4 Task 3: Individuazione delle locazioni dei guasti bubble anomaly

Il Task 3 riguarda l'addestramento di un modello di classificazione multiclasse per individuare le locazioni (BP1 ÷ BP7, BV1) dei guasti di tipo *bubble anomaly*.

### 4.1 Preparazione del dataset

Per il dataset di training sono stati considerati esclusivamente i 24 casi guasti di tipo *bubble anomaly*. Ad ogni caso è stato associato il relativo Faultcode, che può assumere otto diversi valori:

- **1**: se la bolla è situata alla locazione BP1.
- **2**: se la bolla è situata alla locazione BP2.
- **3**: se la bolla è situata alla locazione BP3.
- **4**: se la bolla è situata alla locazione BP4.
- **5**: se la bolla è situata alla locazione BP5.
- **6**: se la bolla è situata alla locazione BP6.
- **7**: se la bolla è situata alla locazione BP7.
- **8**: se la bolla è situata alla locazione BV1.

La struttura della tabella di training è identica a quella del Task 1 (Sezione 2.1).

### 4.2 Diagnostic Feature Designer

La tabella di training è stata caricata nel tool di MATLAB Diagnostic Feature Designer impostando FaultCode come Condition Variable. Le due configurazioni delle frame policies considerate sono le medesime presentate nel Task 1 (Sezione 2.2).

#### Generazione ed estrazione delle features

Per la generazione delle features nel tempo e in frequenza, è stata seguita la medesima procedura presentata nel Task 1 (Sezione 2.2). Le features generate sono state ordinate in base al valore della metrica One-way ANOVA (Figura 14) e sono state estratte quelle dal valore più elevato. La scelta della frame policy è ricaduta sulla **Open** in quanto i valori di One-way ANOVA risultavano essere più alti.

#### 4 TASK 3: INDIVIDUAZIONE DELLE LOCAZIONI DEI GUASTI BUBBLE ANOMALY

Feature	One-way ANOVA
Cases_tsproc_3_sigstats/CrestFactor	3.2103e+03
Cases_tsmodel_5/Coef1	2.7647e+03
Cases_tsproc_1_sigstats/ShapeFactor	2.6996e+03
Cases_tsproc_1_sigstats/PeakValue	2.4992e+03
Cases_tsproc_1_tsfeat/Minimum	2.4992e+03

Figura 14: Ranking features per One-way ANOVA Task 3

Le features nel dominio del tempo selezionate sono: **P1/CrestFactor**, **P1/ImpulseFactor**, **P2/PeakValue**, **P2/Minimum**, **P2/CrestFactor**, **P3/Minimum**, **P5/PeakValue**, **P5/ShapeFactor**, **P5/ ClearanceFactor**, **P5/Minimum**. In questo caso non sono state esportate features in frequenza, in quanto i valori di One-way ANOVA risultavano essere notevolmente più bassi. Queste features sono state importate nel tool *Classification Learner* di MATLAB per l'addestramento di modelli di classificazione.

### 4.3 Classification Learner

Nel tool Classification Learner è stata importata la tabella di training: ogni riga corrisponde ad un frame e contiene le features associate ed il Faultcode corrispondente. In questa operazione sono stati rimossi i campi relativi all'identificatore del caso e quelli riguardanti l'inizio e la fine del frame. Si avranno quindi 3 occorrenze per ogni caso, per un totale di 72 righe. Al fine di mitigare l'overfitting si è scelto di applicare una 10-Fold Cross-Validation e di riservare il 20% del dataset alla validation. La predizione del modello corrisponde a uno degli 8 Faultcodes possibili, in base alla locazione della bolla. Sono stati addestrati tutti i modelli disponibili tramite la funzione *All Train* e selezionato il modello **Bilayered Neural Network** (15), il migliore in base alle metriche *Accuracy* (10-Fold Cross-Validation) e *Accuracy* (Validation sul 20% dei dati di training).



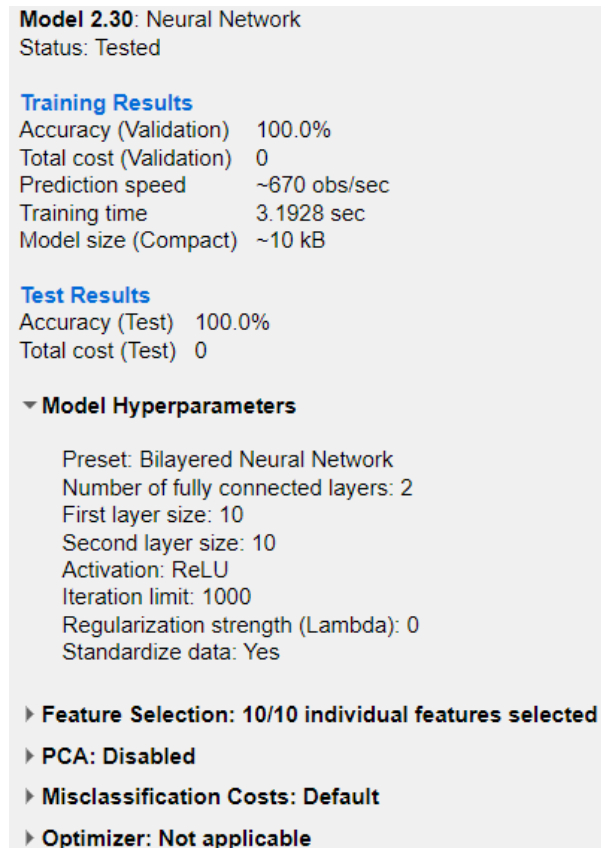


Figura 15: Parametri Bilayered Neural Network

## 4.4 Test e risultati

Al fine di testare il modello selezionato, sono stati preparati i 10 casi predetti come guasti di tipo *bubble anomaly* nel Task 2 (Sezione 3.4) attraverso la costruzione di una tabella contenente le features necessarie al modello di classificazione. Per costruzione, il modello effettua la predizione sul singolo frame. Si è reso dunque necessario raggruppare i risultati per caso. In presenza di discordanze tra predizioni di frame appartenenti allo stesso caso, viene adottato il criterio di maggioranza<sup>1</sup>. Per valutare le prestazioni del modello, sono state confrontate le predizioni con i Faultcode reali presi dal file **answer.csv**. Il risultato ottenuto è il seguente:

---

<sup>1</sup>Nel caso il modello fornisca predizioni tutte differenti per i tre frames, si è deciso di considerare la predizione del primo frame.

Casi	Esito	Real
Caso 1	BP2	BP2
Caso 2	BP6	BP6
Caso 3	BP1	BP1
Caso 4	BP4	BP4
Caso 5	BP7	BP7
Caso 6	BP3	BP3
Caso 7	BP7	BP7
Caso 8	BP1	BP1
Caso 9	BP5	BP5
Caso 10	BP2	BP2

Figura 16: Confronto predizioni locazioni con quelle reali

Come si può notare, il modello predice correttamente tutti i casi di test e la metrica *Accuracy* è pari al 100%.

## 5 Task 4: Individuazione delle valvole guaste

Il Task 4 riguarda l'addestramento di un modello di classificazione multiclasse per individuare le locazioni (SV1 ÷ SV4) dei guasti di tipo *solenoid valve fault*.

### 5.1 Preparazione del dataset

Per il dataset di training sono stati considerati esclusivamente i 48 casi guasti di tipo *solenoid valve fault*. Ad ogni caso è stato associato il relativo Faultcode, che può assumere quattro diversi valori:

- **1:** se il guasto è riferito alla valvola SV1.
- **2:** se il guasto è riferito alla valvola SV2.
- **3:** se il guasto è riferito alla valvola SV3.
- **4:** se il guasto è riferito alla valvola SV4.

La struttura della tabella di training è identica a quella del Task 1 (Sezione 2.1).

### 5.2 Diagnostic Feature Designer

La tabella di training è stata caricata nel tool di MATLAB Diagnostic Feature Designer impostando FaultCode come Condition Variable. Le due configurazioni delle frame policies considerate sono le medesime presentate nel Task 1 (Sezione 2.2).

#### Generazione ed estrazione delle features

Per la generazione delle features nel tempo e in frequenza, è stata seguita la medesima procedura presentata nel Task 1 (Sezione 2.2). Le features generate sono state ordinate in base al valore della metrica One-way ANOVA (Figura 17) e sono state estratte quelle dal valore più elevato. La scelta della frame policy è ricaduta sulla **Open**, in quanto i valori di One-way ANOVA risultavano essere più alti.

Feature	One-way ANOVA
casi_tsmodel_2/Kurtosis	215.1141
casi_tsproc_6_sigstats/Std	135.8539
casi_tsproc_6_sigstats/RMS	135.8539
casi_sigstats_6/Std	132.2577
casi_tsproc_2_sigstats/Std	125.5152

Figura 17: Ranking features per One-way ANOVA Task 4

Le features nel dominio del tempo selezionate sono: **P1/Skewness**, **P1/MAE**, **P2/Skewness**, **P2/MAE**, **P3/Std**, **P3/MSE**, **P4/Std**, **P4/AIC**, **P5/Kurtosis**, **P5/Skewness**, **P6/ShapeFactor**, **P6/Std**, **P7/MSE**, **P7/AIC**.

Quelle in frequenza sono invece: **P1/BandPower**, **P2/BandPower**, **P3/BandPower**, **P4/BandPower**, **P5/BandPower**, **P6/BandPower**, **P7/BandPower**. Queste features sono state importate nel tool *Classification Learner* di MATLAB per l'addestramento di modelli di classificazione.

### 5.3 Classification Learner

Nel tool Classification Learner è stata importata la tabella di training: ogni riga corrisponde ad un frame e contiene le features associate ed il Faultcode corrispondente. In questa operazione sono stati rimossi i campi relativi all'identificatore del caso e quelli riguardanti l'inizio e la fine del frame. Si avranno quindi 3 occorrenze per ogni caso, per un totale di 144 righe. Al fine di mitigare l'overfitting si è scelto di applicare una 10-Fold Cross-Validation e di riservare il 20% del dataset alla validation. La predizione del modello corrisponde a uno dei 4 Faultcodes possibili, in base alla valvola guasta. Prima di addestrare i modelli, si è deciso di filtrare ulteriormente le features nella sezione Feature Selection del tool, in quanto si ritenevano eccessive 21 features. Sono quindi state scartate altre 5 features: **P2/Skewness**, **P2/MAE**, **P1/BandPower**, **P2/BandPower**, **P5/BandPower**. Sono stati addestrati tutti i modelli disponibili tramite la funzione *All Train* e selezionato il modello **Medium KNN** (Figura 18), il migliore in base alle metriche *Accuracy* (10-Fold Cross-Validation) e *Accuracy* (Validation sul 20% dei dati di training).

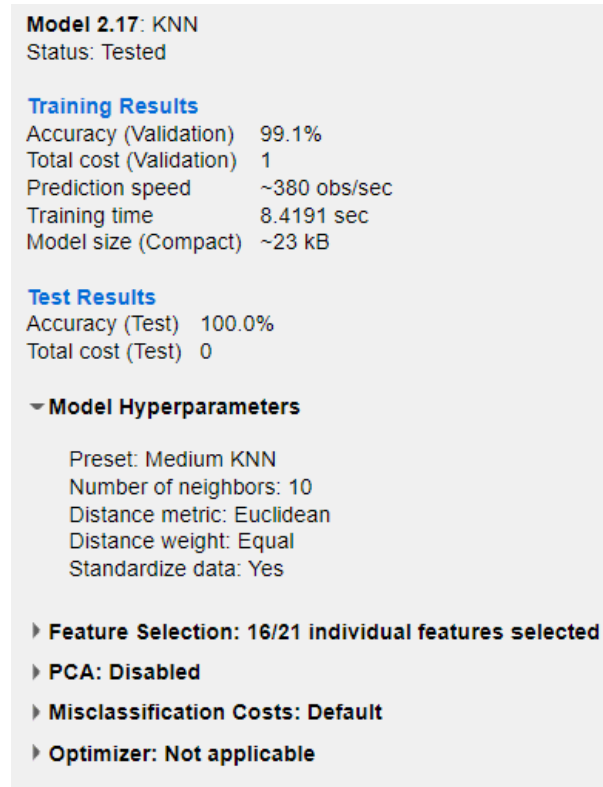


Figura 18: Parametri Medium KNN

## 5.4 Test e risultati

Al fine di testare il modello selezionato, sono stati preparati i 10 casi predetti come guasti di tipo *solenoid valve fault* nel Task 2 (Sezione 3.4) attraverso la costruzione di una tabella contenente le features necessarie al modello di classificazione. Per costruzione, il modello effettua la predizione sul singolo frame. Si è reso dunque necessario raggruppare i risultati per caso. In presenza di discordanze tra predizioni di frame appartenenti allo stesso caso, viene adottato il criterio di maggioranza<sup>2</sup>. Per valutare le prestazioni del modello, sono state confrontate le predizioni con i Faultcode reali presi dal file **answer.csv**. Il risultato ottenuto è il seguente:

---

<sup>2</sup>Nel caso il modello fornisca predizioni tutte differenti per i tre frames, si è deciso di considerare la predizione del primo frame.

Casi	Esito	Real
Caso 1	SV2	SV2
Caso 2	SV4	SV4
Caso 3	SV1	SV1
Caso 4	SV3	SV3
Caso 5	SV4	SV1
Caso 6	SV3	SV3
Caso 7	SV2	SV2
Caso 8	SV4	SV1
Caso 9	SV2	SV2
Caso 10	SV4	SV4

Figura 19: Confronto predizioni valvole guaste con quelle reali

Come si può notare, il modello predice correttamente 8 casi di test su 10; la metrica *Accuracy* è quindi pari all'80%.

## 6 Task 5: Determinazione delle percentuali di apertura delle valvole guaste

Il Task 5 riguarda l'addestramento di un modello di regressione per determinare le percentuali di apertura delle valvole guaste (valore compreso tra 0% e 100%).

### 6.1 Preparazione del dataset

Per il dataset di training sono stati considerati un totale di 60 casi:

- 48 casi guasti di tipo *solenoid valve fault*, a cui sono state associate le percentuali di apertura delle valvole guaste.
- 12 casi non guasti, al fine di fornire al modello informazioni relative al funzionamento normale.

La percentuale di apertura è un valore compreso tra 0% e 99% in caso di guasto, mentre assume il valore di 100% per i casi relativi ad un funzionamento normale.

#### Struttura della tabella di training

La tabella di training (Figura 20) è quindi composta da tre colonne:

- **Casi:** identifica il caso in esame.
- **Rilevazioni:** contiene le rilevazioni effettuate sulla finestra temporale di 1.2 secondi.
- **Percentage:** contiene la percentuale di apertura della valvola. Per i casi relativi ad un funzionamento normale il valore è 100%, per i guasti essa contiene la percentuale di apertura della valvola guasta.

	1	2	3
	Casi	Rilevazioni	Percentage
1	"Caso 1"	<i>1201x8 table</i>	100
2	"Caso 2"	<i>1201x8 table</i>	100

Figura 20: Estratto tabella di training Task 5

### 6.2 Diagnostic Feature Designer

La tabella di training è stata caricata nel tool di MATLAB Diagnostic Feature Designer impostando Percentage come Condition Variable. Le due configurazioni delle frame policies considerate sono le medesime presentate nel Task 1 (Sezione 2.2).

### Generazione ed estrazione delle features

Per la generazione delle features nel tempo e in frequenza, è stata seguita la medesima procedura presentata nel Task 1 (Sezione 2.2). Le features generate sono state ordinate in base al valore della metrica One-way ANOVA (Figura 21) e sono state estratte quelle dal valore più elevato. La scelta della frame policy è ricaduta sulla **Open + Close** in quanto i valori di One-way ANOVA risultavano essere più alti.

Feature	One-way ANOVA
casi_res_tsfeat/Median	3.0809e+05
casi_res_tsfeat/Q1	1.6565e+05
casi_ps_spec/PeakAmp1	6.0872e+04
casi_tsfeat/Q3	5.1198e+04
casi_sigstats/Mean	4.7764e+04

Figura 21: Ranking features secondo One-way ANOVA Task 5

Le features nel dominio del tempo selezionate sono: **P1/Median, P1/Minimum, P1/ShapeFactor, P1/Mean, P1/RMS, P2/Q1, P2/IQR, P3/RMS, P3/Median, P3/Q3, P4/RMS, P4/Median, P4/Q3, P5/Median, P5/IQR, P6/Median, P6/Q1, P6/Q3, P7/Median, P7/Q1, P7/Q3.**

Quelle in frequenza sono invece: **P1/PeakAmp1, P1/BandPower, P2/BandPower, P5/BandPower, P5/PeakAmp1.** Queste features sono state importate nel tool *Regression Learner* di MATLAB per l'addestramento di modelli di regressione.

### 6.3 Regression Learner

Nel tool Regression Learner è stata importata la tabella di training: ogni riga corrisponde ad un frame e contiene le features associate e la percentuale di apertura corrispondente. In questa operazione sono stati rimossi i campi relativi all'identificatore del caso e quelli riguardanti l'inizio e la fine del frame. Si avranno quindi 3 occorrenze per ogni caso, per un totale di 180 righe. Al fine di mitigare l'overfitting si è scelto di applicare una 10-Fold Cross-Validation e di riservare il 20% del dataset alla validation. La predizione del modello corrisponde alla percentuale di apertura della valvola guasta. Prima di addestrare i modelli, si è deciso di filtrare ulteriormente le features nella sezione Feature Selection del tool, in quanto si ritenevano eccessive 26 features. Sono quindi state scartate altre 8 features: **P1/Median, P1/Mean, P3/Q3, P5/Median, P5/IQR, P6/Q3, P7/Q1, P1/BandPower.** Sono stati addestrati tutti i modelli disponibili tramite la funzione *All Train* e selezionato il modello **Trilayered Neural Network** (Figura



22), il migliore in base alle metriche *RMSE* sui dati di Training (10-Fold Cross-Validation), *RMSE* sui dati di Validation (20% dei dati di training) e Prediction Speed.

<b>Model 3.5: Neural Network</b>	
Status: Tested	
<b>Training Results</b>	
RMSE (Validation)	9.5321
R-Squared (Validation)	0.93
MSE (Validation)	90.86
MAE (Validation)	5.5055
Prediction speed	~1500 obs/sec
Training time	7.0096 sec
Model size (Compact)	~12 kB
<b>Test Results</b>	
RMSE (Test)	9.9757
R-Squared (Test)	0.90
MSE (Test)	99.516
MAE (Test)	6.2666
<b>Model Hyperparameters</b>	
Preset: Trilayered Neural Network	
Number of fully connected layers: 3	
First layer size: 10	
Second layer size: 10	
Third layer size: 10	
Activation: ReLU	
Iteration limit: 1000	
Regularization strength (Lambda): 0	
Standardize data: Yes	
<b>Feature Selection: 18/26 individual features selected</b>	
<b>PCA: Disabled</b>	
<b>Optimizer: Not applicable</b>	

Figura 22: Parametri Trilayered Neural Network

## 6.4 Test e risultati

Al fine di testare il modello selezionato, sono stati preparati i 10 casi predetti come guasti di tipo *solenoid valve fault* nel Task 2 (Sezione 3.4) attraverso la costruzione di una tabella contenente le features necessarie al modello di regressione. Per costruzione, il modello effettua la predizione sul singolo frame. Si è reso dunque necessario raggruppare i risultati per caso. Ciò è stato effettuato attraverso il calcolo della media delle tre percentuali. Nel caso in cui si abbiano valori percentuali maggiori o uguali a 100, la predizione viene impostata pari a 99%, corrispondente al limite superiore. Per valutare le prestazioni del modello, sono state confrontate le predizioni con le percentuali reali prese dal file **answer.csv**.

6 TASK 5: DETERMINAZIONE DELLE PERCENTUALI DI APERTURA  
DELLE VALVOLE GUASTE

---

Casi	Esito	Real
Caso 1	35	22
Caso 2	73	76
Caso 3	13	5
Caso 4	46	46
Caso 5	96	98
Caso 6	45	44
Caso 7	90	94
Caso 8	99	95
Caso 9	68	70
Caso 10	26	24

Figura 23: Confronto predizioni percentuali con quelle reali

Come si può notare, le predizioni effettuate dal modello si avvicinano ai risultati reali. In particolare, andando a calcolare alcune metriche, si ottiene:

- RMSE: 5.3572
- MAE: 3.9
- MSE: 28.7

## 7 Conclusioni

L'obiettivo del progetto è stato il riconoscimento e la distinzione di guasti in un sistema di propulsione di navicelle spaziali, addestrando e utilizzando modelli di classificazione e regressione. Attraverso un'analisi approfondita dei dati raccolti e l'impiego di strumenti come MATLAB Diagnostic Feature Designer e Classification Learner, si è riusciti a identificare correttamente i casi di guasto e a distinguere tra diverse tipologie di anomalie.

I risultati ottenuti dimostrano un'accuratezza elevata nella classificazione dei guasti e una buona precisione nella stima delle percentuali di apertura delle valvole guaste. Questi risultati confermano l'efficacia dei modelli sviluppati e suggeriscono che tecniche simili potrebbero essere applicate in altri contesti.

Nonostante i risultati positivi, lo studio presenta alcune limitazioni, come la dimensione contenuta del dataset per quanto riguarda le singole tipologie di guasto.

In conclusione, il lavoro svolto può essere considerato come una base solida per possibili ulteriori sviluppi, che potranno focalizzarsi, ad esempio, sull'integrazione di tecniche di ensemble e sul miglioramento della capacità di generalizzazione dei modelli.