

# Machine Learning nella Diagnostica Prenatale: un'applicazione ai Cardiotocogrammi

Miori Davide, Ondeì Alice, Sauro Letterino, Venturi Alessandro

CdLM Data Science, Università degli studi di Milano Bicocca

Con “Diagnostica Prenatale” ci si riferisce all’insieme di tecniche atte a verificare lo stato di salute del feto. La cardiotocografia (CTG) è un valido metodo per indagare se il feto si sta sviluppando secondo le aspettative, mediante la misurazione del battito cardiaco, dei movimenti del feto e delle contrazioni uterine materne. La verifica delle condizioni di salute tramite i valori raccolti è un’operazione complessa e delicata, soprattutto se eseguita manualmente. Un’analisi computerizzata, effettuata mediante la costruzione di modelli di classificazione, può produrre un’interpretazione automatica e ridurre il tasso di mortalità prenatale, il manifestarsi di patologie o il rischio di interventi di parto cesareo non richiesti.

## Introduzione

Nel 1986 la *Federazione Internazionale di Ginecologia e Ostetricia* (FIGO) ha stabilito le linee guida per il monitoraggio fetale; secondo tali norme i tracciati cardiotocografici possono individuare tre stati di salute del feto: *Normale*, *Sospetto* o *Patologico*. Fondamentali alla fine di una corretta diagnosi sono soprattutto i valori di quattro parametri: battito cardiaco basale del feto (BL), accelerazione (ACC), decelerazione (DCL) e variabilità - dove con variabilità si intendono le fluttuazioni del BL non valutate come accelerazione o decelerazione. Monitorando tali parametri è possibile identificare l’ipossia fetale - carenza di ossigeno - la quale, se prolungata, provoca alte probabilità di disabilità nel neonato e, nei casi peggiori, può condurlo alla morte. Dal punto di vista medico, inoltre, un record CTG è considerato *Patologico* quando si presentano almeno due tra i seguenti sintomi [1]:

- Il BL supera i 180 bpm o scende al di sotto di 100 bpm.
- La variabilità scende al di sotto di 5 bpm per un periodo superiore a 90 minuti.
- Si presenta una variabile atipica o avvengono decelerazioni tardive che durano oltre 30 minuti.
- Si verifica una singola decelerazione prolungata e dura più di 3 minuti.
- Il cuore del feto inizia a battere con andamento sinusoidale per più di 10 minuti o si verifica una bradicardia prolungata.

Tramite la stessa raccolta di dati provenienti da un esame cardiotocografico è inoltre possibile discriminare la frequenza cardiaca fetale individuando dieci diverse condizioni.

Scopo del presente elaborato è costruire dei modelli di classificazione che automatizzino l’interpretazione dei CTG e forniscano rapidamente delle risposte al medico. In particolare si sviluppano tre diversi modelli di classificazione:

- **Modello di classificazione binaria:** ha l’obiettivo di riconoscere lo stato del feto rispetto alle due classi: *Normale* o *Anomalo*, dove per *Normale* si intende un feto che non ha bisogno di ulteriori approfondimenti e/o cure mentre per *Anomalo* si intende un feto che potrebbe aver bisogno di ulteriori approfondimenti e/o cure.
- **Modello di classificazione a tre classi:** ha l’obiettivo di riconoscere lo stato del feto rispetto alle tre classi: *Normale*, *Sospetto* o *Patologico*, dove alla classe *Normale* è data la stessa interpretazione riportata sopra, con la classe *Sospetto* si intende un feto con valori anomali del CTG ma non necessariamente riconducibili a uno stato patologico mentre con la *Patologico* si intende un feto i cui valori del CTG sono riconducibili a uno stato che pone in rischio la sua salute.
- **Modello di classificazione a dieci classi:** ha l’obiettivo di riconoscere lo stato del feto rispetto a dieci possibili classi, riportate nella [Nota 1](#).

Data la complessità della lettura di tale esame diagnostico, l’automazione porta vantaggi sia in termini di efficacia che di efficienza, rendendo l’interpretazione più affidabile, più rapida e meno dipendente dalle disponibilità degli specialisti.

## Il Dataset

Il dataset utilizzato è stato creato presso la facoltà di medicina dell’Università di Porto ed analizzato mediante il software SisPorto; contiene 2126 osservazioni e 23 attributi, i quali sono elencati nella Tabella 1. Ogni record è stato manualmente valutato da tre ostetriche esperte le quali, sulla base dei valori assunti dai parametri, hanno assegnato un’etichetta ad ogni istanza. Non deve essere considerato come un fattore secondario la necessità di avere il parere di ben tre ostetriche esperte per poter imputare il valore degli attributi target, ciò infatti sottolinea come il parere di una sola figura tecnica, seppur esperta, potrebbe non essere affidabile, così come il parere di due. In questo contesto potersi servire di un modello automatizzato porterebbe a grandissimi vantaggi.

Questo set di dati è pubblicamente disponibile presso “*The Data Mining Repository of University of California Irvine (UCI)*” [2].

Nome	Descrizione
LB	Battito cardiaco basale (in battiti al minuto)
AC	N. di accelerazioni al secondo
FM	N. di movimenti fetali al secondo
UC	N. contrazioni uterine al secondo
DL	N. di lievi decelerazioni al secondo
DS	N. di gravi decelerazioni al secondo
DP	N. di decelerazioni prolungate al secondo
ASTV	Percentuale di tempo con variabilità anormale a breve termine
MSTV	Valore medio della variabilità a breve termine
ALTV	Percentuale di tempo con variabilità anormale a lungo termine
MLTV	Valore medio della variabilità a lungo termine
Width	Range dell'istogramma di FHR (Fetal Heart Rate)
Min	Valore minimo dell'istogramma di FHR
Max	Valore massimo dell'istogramma di FHR
Nmax	N. di picchi nell'istogramma di FHR
Nzeros	N. di zero nell'istogramma di FHR
Mode	Valore modale dell'istogramma di FHR
Mean	Valore medio dell'istogramma di FHR
Median	Valore mediano dell'istogramma di FHR
Variance	Varianza dell'istogramma di FHR
Tendency	Tendenza dell'istogramma di FHR
CLASS	FHR pattern class code: da 1 a 10 (così come descritto nella <a href="#">Nota 1</a> )
NSP	Fetal state class code (N = Normale; S = Sospetto; P = Patologico)

Tabella 1. Attributi

## Modelli di classificazione e misure di performance

I modelli di classificazione possono essere raggruppati in quattro classi:

- **Modelli Euristici:** modelli basati sul *trade-off* tra parsimonia e grado di approssimazione, ne fanno parte il *Decision Tree* e il *Random Forest*;
- **Modelli Regressivi:** modelli basati sulle tecniche di regressione, ne fa parte la *Regressione Logistica*;
- **Modelli di Separazione:** modelli basati sul partizionamento dello spazio delle variabili esplicative, ne fa parte la *Support Vector Machine*;
- **Modelli Probabilistici:** modelli basati sulle distribuzioni di probabilità, ne fa parte il *Modello Naive Bayes*.

Nonostante le diverse nature e i diversi approcci risolutivi di questi modelli, è possibile servirsi delle stesse misure di performance per valutarne la bontà ed eseguire dei confronti. Dato un modello e un'osservazione da questo classificata, si hanno quattro possibilità:

- True Positive<sup>1</sup> (TP): l'osservazione con classe positiva è predetta dal modello come positiva;
- False Negative (FN): l'osservazione con classe positiva è predetta dal modello come negativa;
- True Negative (TN): l'osservazione con classe negativa è predetta dal modello come negativa;

<sup>1</sup>È prassi chiamare come positiva la classe minoritaria, cioè quella verso cui si pone maggior interesse nella ricerca

- False Positive (FP): l'osservazione con classe negativa è predetta dal modello come positiva.

Si definisce *Accuracy* la percentuale di predizioni corrette del modello:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (1)$$

Pur essendo l'*Accuracy* un ottimo indicatore di performance, non sempre indica correttamente la bontà del modello, ad esempio nei casi in cui le classi sono particolarmente sbilanciate (Numero *Positivi* < Numero *Negativi*). In tali casi allora si ricorre ad altre misure:

$$Recall = \frac{TP}{TP + FN} ; Precision = \frac{TP}{TP + FP} \quad (2)$$

dove la *Recall* indica la percentuale di record positivi classificati correttamente, mentre la *Precision* tiene in conto unicamente i record classificati come positivi e di questi considera la percentuale di predizioni corrette. È importante sottolineare come nel campo della diagnostica medica si pone solitamente massima attenzione ai valori della *Recall*, i quali devono essere estremamente elevati poiché deve essere minimo il numero dei malati che sfuggono al test diagnostico. Una sintesi di *Recall* e *Precision* è rappresentata dalla *F<sub>1</sub> Measure*, cioè la media armonica delle due, espressa come:

$$F_1 = 2 \frac{Recall \times Precision}{Recall + Precision} \quad (3)$$

Infine è utilizzata l'*Area Under Curve* (AUC), l'area sottesa dalla Curva ROC, quest'ultima è la curva che descrive l'equilibrio tra False Positives Rate e True Positives Rate, maggiore è l'AUC e migliore è il modello. *Accuracy*, *Recall*, *Precision*, *F<sub>1</sub>* e *AUC* sono tutte misure che assumono valori tra 0 e 1. La classificazione multiclasse è eseguita tramite tecnica "*One vs All*", per cui la procedura di valutazione è ripetuta tante volte quanto il numero delle classi, in ciascuna di esse una classe è considerata *Positiva* mentre le restanti sono considerate *Negative*, in questo modo si ottengono valori di *Recall*, *Precision*, *F<sub>1</sub> Measure* e *AUC* per ogni classe, mentre l'*Accuracy* rimane un valore globale per il modello.

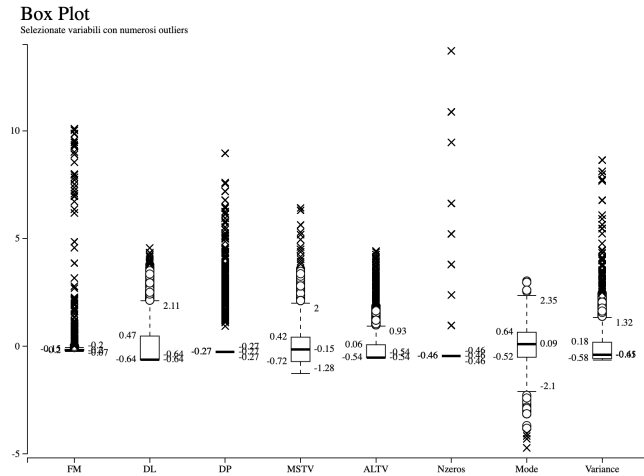
## Analisi Esplorativa

La prima fase dello studio è rappresentata dall'analisi esplorativa, ovvero un'indagine preliminare delle caratteristiche del dataset e degli attributi, attraverso tecniche statistiche.

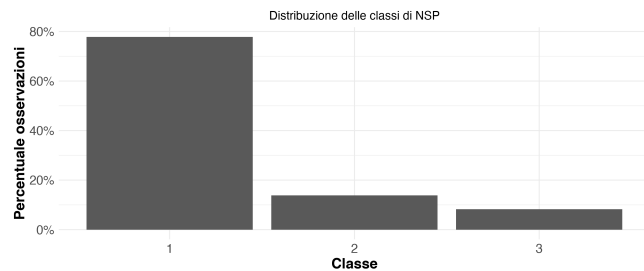
La prima cosa che si nota è l'assenza di *missing data*.

Si procede così con la rappresentazione delle distribuzioni degli attributi. Poiché le unità dei vari attributi non sono tutte uguali, si hanno scale molto diverse tra di loro, è quindi opportuno normalizzare gli attributi numerici. In questa fase si nota che l'attributo *DS* assume valore 0 per tutte le osservazioni meno cinque casi, ciò comporta un valore poco informativo per questo attributo che è quindi rimosso. Dall'analisi delle distribuzioni si nota inoltre una forte presenza di valori anomali, così come si osserva dalla Figura 1 dove sono rappresentate solamente le distribuzioni degli attributi con più

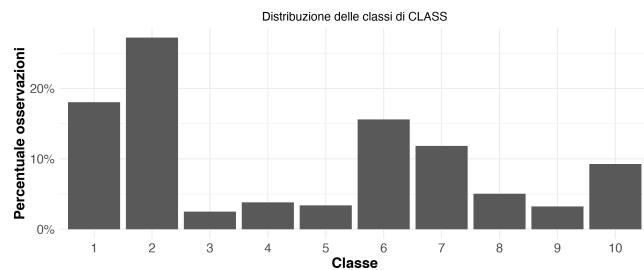
outliers. La presenza di tali outliers non è però da ricondurre necessariamente a degli errori di misurazione, poiché la ricerca di casi patologici è per sua natura legata alla presenza di questo tipo di valori. Non essendo quindi possibile distinguere quali outliers siano associabili a errori di osservazione, si decide di non escludere alcuna istanza dall'analisi.



**Figura 1.** Distribuzioni degli attributi, normalizzati, che presentano numerosi outliers. Tali outliers non dovranno essere esclusi dall'analisi.



**Figura 2.** Distribuzione in frequenza delle classi dell'attributo *NSP* (1 : *Normale*, 2 : *Sospetto*, 3 : *Patologico*). Si nota forte sbilanciamento.

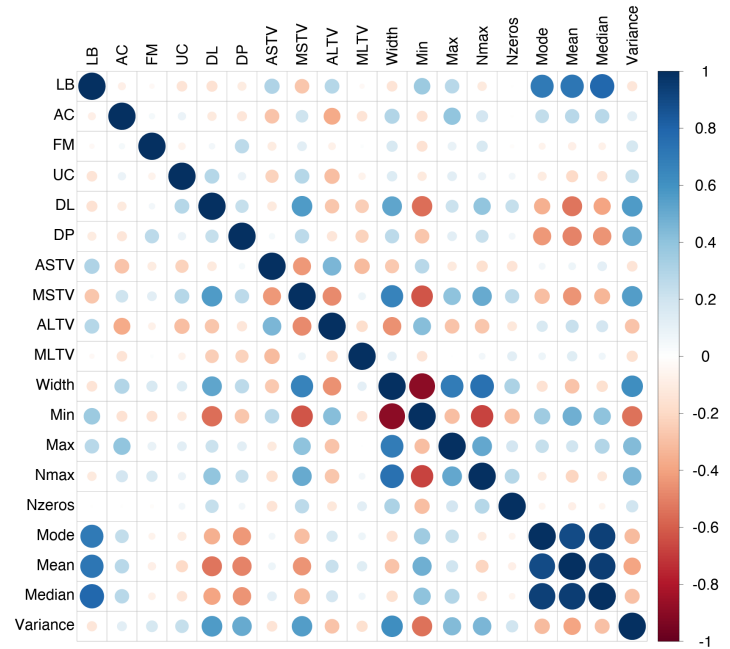


**Figura 3.** Distribuzione in frequenza delle classi dell'attributo *CLASS*. Si nota forte sbilanciamento.

Si indagano poi le distribuzioni di frequenza per le varie classi degli attributi *NSP* e *CLASS*, riportate nelle Figure 2 e 3. Si nota un forte sbilanciamento in entrambe le situazioni, ciò potrebbe rappresentare un problema ai fini di una corretta classificazione, pertanto dovranno essere applicate apposite metodologie in fase di costruzione dei modelli.

Infine viene eseguita un'analisi della correlazione, così come in Figura 4, da cui si nota la presenza di attributi fortemente correlati, in particolare *Width* è fortemente correlato con gli

attributi *Min* e *Max*, così come previsto essendo *Width* la differenza delle due. Anche i tre attributi *Mode*, *Mean*, *Median* sono fortemente correlati tra di loro. Tale correlazione dovrà essere attentamente trattata nella costruzione dei modelli, al fine di scongiurare l'ipotesi di *overfitting*.



**Figura 4.** Correlogramma di tutti gli attributi numerici.

## Costruzione modelli di classificazione

**Modello di classificazione binaria.** Per la classificazione binaria sono stati costruiti cinque modelli (*J48*, *Naive Bayes*, *Regression Logistica*, *Support Vector Machine* e *Random Forest*) e di questi sono state confrontate le performance per individuare infine il modello migliore.

Il primo step è la binarizzazione dell'attributo *NSP*, con cui vengono unite le classi *Sospetto* e *Patologico* in un'unica classe *Anomalo*.

La nuova variabile target presenta quindi due classi: *Normale* e *Anomalo*, la prima si riferisce a tutti quei casi per cui non sono necessari ulteriori approfondimenti perchè il feto è sano, mentre la seconda classe, che sarà considerata la classe *Positiva*, racchiude i casi che necessitano di ulteriori accertamenti circa la presenza di una possibile patologia ai danni del feto.

Successivamente si decide di applicare la normalizzazione *Z-score* agli attributi poiché questi presentano unità di misura differenti. Si ottengono così attributi con media nulla e varianza unitaria.

Come mostrato nella sezione di analisi esplorativa è presente una forte correlazione tra gli attributi, in particolare quelli riferiti all'istogramma FHR. Per eliminare la correlazione si è scelto di utilizzare la *Principal Component Analysis* (PCA), impostando come valore obiettivo il 90% della varianza spiegata. Si ottengono così quattro nuove variabili a partire dalle nove di input. Si può osservare in Figura 5 il nuovo correlogramma, dove si nota che la correlazione è stata in buona

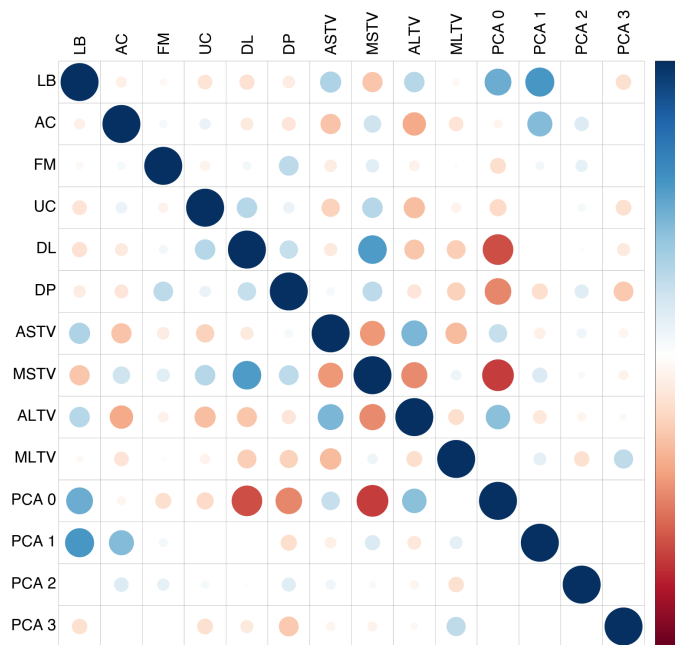


Figura 5. Correlogramma di tutti gli attributi numerici a seguito della PCA.

parte rimossa e permane solo con valori inferiori a 0.8 (in valore assoluto).

Si passa così alla costruzione dei cinque modelli precedentemente introdotti, per ognuno di questi sono state implementate diverse tecniche, sia per evitare *bias* legati alla scelta del *Test set* sia per trattare lo sbilanciamento delle classi. Si sono pertanto utilizzati il metodo *Iterated Holdout* (con 15 iterazioni) per la validazione e il metodo *oversampling* con la tecnica *SMOTE* (Synthetic Minority Over-sampling Technique) per bilanciare le classi. Inoltre il partizionamento *Training set - Test set* è stato eseguito con proporzione 70% - 30% e con la tecnica dello *stratified sampling* per trattare la scarsità dei record associati alla classe *Anomalo*.

Successivamente si procede all'apprendimento e al test dei modelli, estraendo, per ogni modello e per ogni iterazione dell'*Iterated Holdout*, i valori delle principali misure di performance: Accuracy, Recall, Precision, F1-measure e AUC. In tal modo aggregando i risultati di ogni iterazione è possibile ottenere degli intervalli di confidenza per ciascuna di queste misure.

Per il confronto tra i modelli si prende in prima battuta in considerazione la Recall, infatti il principale obiettivo che si pone in fase di valutazione dei CTG è evitare di classificare come *Normali* feti potenzialmente malati e che hanno quindi bisogno di ulteriori accertamenti. È mostrato in Figura 6 il confronto degli intervalli di confidenza della Recall per i vari modelli, si osserva come i 3 modelli migliori siano *Regressione Logistica*, *Random Forest* e *Support Vector Machine*, ma non è possibile spingersi oltre nella scelta. Per selezionare quindi il classificatore migliore tra questi tre si studiano le altre misure di performance: *Accuracy* (riportata in Figura 7), *Precision*, *F1*, *AUC*: si osserva che per ciascuna di queste misure il *Random Forest* primeggia. La scelta finale sul modello da utilizzare ricade infine sul *Random Fo-*

*rest*, con 100 alberi decisionali, che mostra i seguenti valori medi:  $\overline{Accuracy} = 95\%$ ;  $\overline{Recall} = 89\%$ ;  $\overline{Precision} = 87\%$ ;  $\overline{F_1} = 87\%$ ;  $\overline{AUC} = 99\%$ .

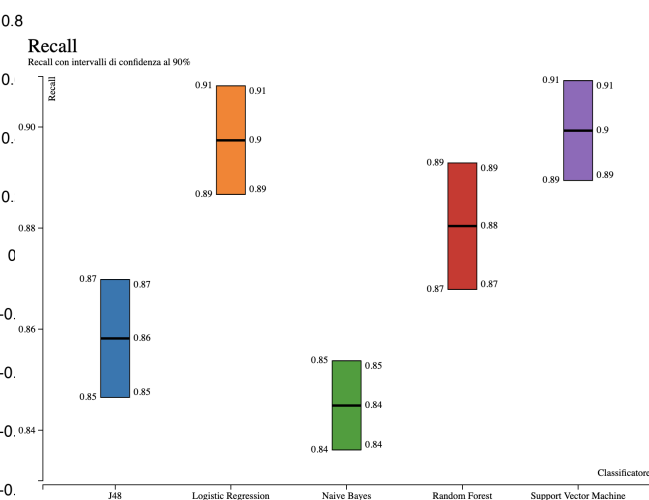


Figura 6. Confronto intervalli di confidenza della Recall, al 90% del livello di confidenza. Tre modelli si distinguono per alte performance.

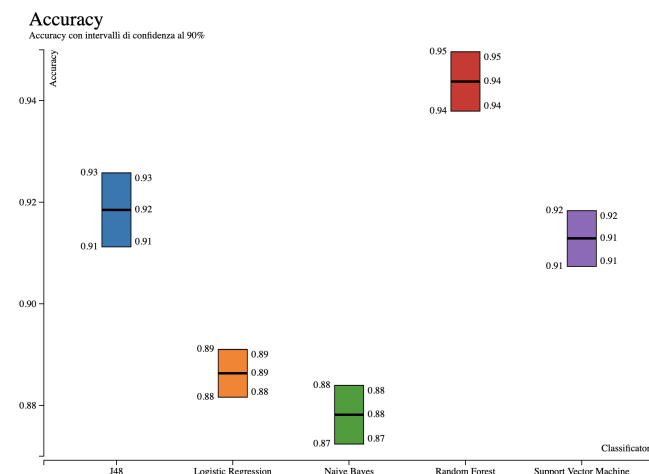
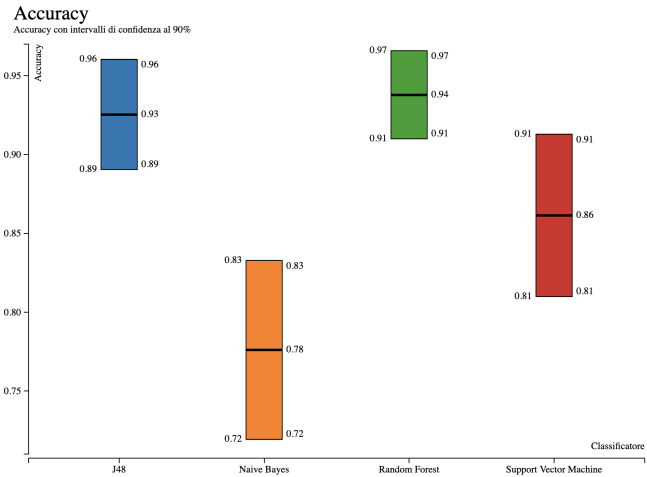


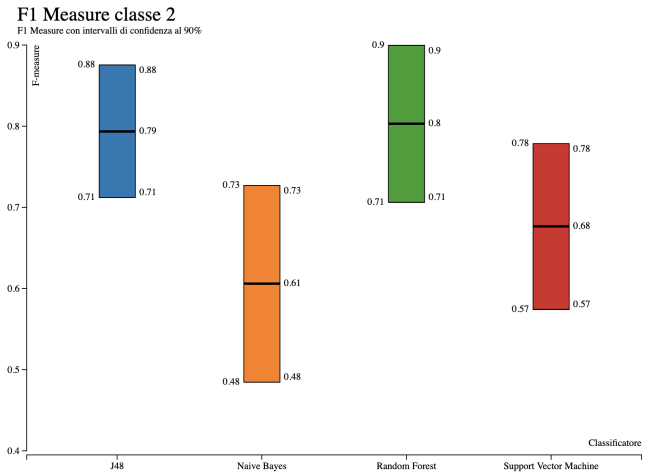
Figura 7. Confronto intervalli di confidenza della Accuracy, al 90% del livello di confidenza. Il *Random Forest* è il modello migliore.

**Modello di classificazione multiclasse: 3 classi.** Ai fini della classificazione a tre classi è stata selezionata *NSP* come variabile target, i cui livelli risultano essere *Normale*, *Sospetto* e *Patologico*. Prima di procedere con l'analisi dei dati è stato rimosso l'attributo *Width*, poiché combinazione lineare di *Max* e *Min*. Alcuni attributi presenti si sono dimostrati fortemente correlati, si decide però di non eliminarne nessuno aprioristicamente poiché verrà effettuata una *Feature Selection* tramite metodo *Wrapper*, a differenza del precedente caso per il quale si era deciso di procedere applicando una PCA. Suddiviso il dataset in due partizioni, *Training set* (75%) e *Test set* (25%), si è proceduto all'applicazione di quattro classificatori, quali: *J48*, *Random Forest*, *Support Vector Machine* e *Naive Bayes*. Mediante *Cross Validation* (*k-fold*), con *k* pari a 10, i precedenti modelli sono stati allenati sui *Training set* e testati sui *Validation set* così ottenuti. Dato l'elevato sbilanciamento, si decide di applicare *oversampling* con tecnica

**SMOTE.** Si procede poi nuovamente con l'applicazione della tecnica di *Cross Validation* costruendo gli intervalli di confidenza relativi alle metriche per ogni classe. Sebbene non sia possibile stabilire quali modelli siano i migliori con un livello di confidenza al 90%, avendo su quasi tutte le metriche la sovrapposizione degli intervalli di confidenza, si decide di proseguire l'analisi utilizzando unicamente i modelli *J48* e *Random Forest*. Nelle Figure 8, 9 e 10 sono riportate rispettivamente l'Accuracy, e le *F1 Measure* delle classi *Sospetto* e *Patologico*.

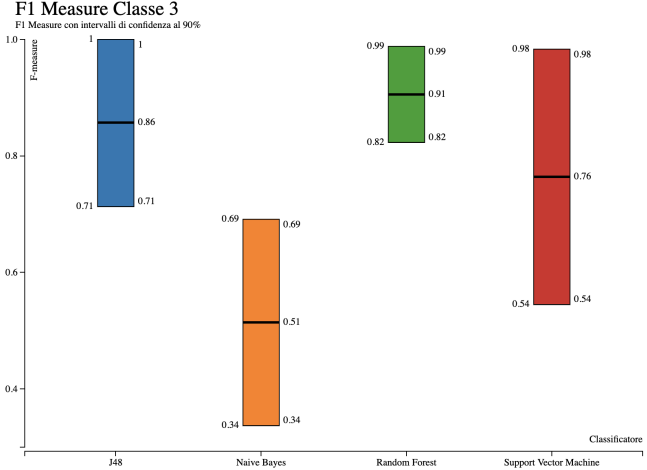


**Figura 8.** Confronto intervalli di confidenza della *Accuracy*, al 90% del livello di confidenza.



**Figura 9.** Confronto intervalli di confidenza della *F1 Measure* della classe *Sospetto*, al 90% del livello di confidenza.

Il passo successivo è l'utilizzo di *Feature Selection* mediante *CfsSubsetEval* e *InfoGain* applicata con i due classificatori scelti. Gli attributi selezionati con il primo metodo sono *AC*, *DP*, *ASTV*, *MSTV*, *ALTV* e *Mean* mentre quelli ottenuti dal secondo, tenendo conto di un contributo maggiore del 19%, sono *AC*, *ASTV*, *MSTV*, *ALTV*, *Mean* e *Variance*. Entrambi i modelli hanno ottenuto lo stesso set di *feature*. Questa procedura è stata effettuata estraendo un *Validation set* corrispondente al 20% del *Training set*. Prima di effettuare la verifica finale sul *Test set* si è deciso di ottimizzare il numero di alberi relativo al *Random Fore-*



**Figura 10.** Confronto intervalli di confidenza della *F1 Measure* della classe *Patologico*, al 90% del livello di confidenza.

*st* sia con le *feature* selezionate tramite *CfsSubsetEval* sia con quelle selezionate tramite *InfoGain* mediante il *Parameter Tuning*. Sempre grazie alla *Cross Validation*, il parametro in esame è stato validato ponendo a priori il valore dello stesso tra 60 e 140. Il numero di alberi ottimale nel caso di *CfsSubsetEval* è pari a 140 mentre relativamente all'altro 60. Il *Training set* utilizzato in questo passaggio è stato ottenuto ricongiungendo i due set utilizzati nella *Feature Selection*. L'analisi si è conclusa andando ad osservare le performance dei due modelli sul *Test set*. La tabella 2 mostra i risultati ottenuti dai quattro classificatori allenati sul *Training set* con gli attributi precedentemente selezionati e con il numero di alberi ottimizzato nel caso del *RandomForest*.

Metrica	J48(Cfs)	RF(Cfs)	J48(infoGain)	RF(InfoGain)
RecS	0.919	0.892	0.824	0.892
RecP	0.955	0.955	0.932	0.932
PreS	0.773	0.815	0.753	0.825
PreP	0.778	0.824	0.661	0.759
F1S	0.84	0.853	0.787	0.857
F1P	0.857	0.884	0.774	0.837

**Tabella 2.** Risultati classificazione *NSP*.

Si nota come i valori relativi al set di *feature* del *CfsSubsetEval* siano ottimali per entrambi i classificatori relativamente alla *Recall*. Sulla base di quest'ultima il *J48 Cfs* performa leggermente meglio nella classe *Sospetto*. Prendendo in considerazione il valore dell'*F1 Measure*, invece, il *Random Forest Cfs* risulta di poco il migliore perciò, sulla base di questi valori, quest'ultimo risulta essere quello selezionato.

**Modello di classificazione multiclasse: 10 classi.** Ai fini dell'analisi in questione sono stati selezionati a priori due modelli: la *Support Vector Machine* e il *Random Forest*. Il primo passo è stato selezionare gli attributi migliori tramite *CorrelationAttributeEval* per entrambi i modelli. Tenendo conto di un contributo maggiore del 15%, sono stati selezionati i seguenti attributi: *AC*, *DL*, *Variance*, *Min*, *ALTV*, *MSTV*, *ASTV*, *Nmax*, *Mean* e *Max*. Così come per il caso precedente è stato ottimizzato per il



*Random Forest* il numero di alberi ottenendo, tramite la tecnica *Cross Validation*, un numero ottimo pari a 70. Per la *Support Vector Machine*, invece, è stato ottimizzato il parametro *C* scegliendo come valori limite 1 e 100. Il valore ottenuto è stato di 61. Si riportano i risultati raggiunti nel *Test set* da parte dei due modelli nella tabella 3.

Metrica	SVM(Corr)	RF(Corr)
Rec10	0.653	0.959
Rec9	0.882	0.882
Rec8	0.852	0.926
Rec5	0.778	0.772

**Tabella 3.** Risultati classificazione CLASS.

Come si evince osservando le performance riportate, il *Random Forest* ottiene dei valori di *Recall* ottimi per le classi 8, 9 e 10 mentre relativamente alla classe 5 il valore non è eccellente. Tuttavia avendo solo 18 osservazioni nel *Test set*, l'errata classificazione di pochi record porta a ridurre in maniera sostanziale il valore della *Recall*. In aggiunta, è necessario sottolineare che le osservazioni mal classificate relative alla classe 5 vengono maggiormente attribuite alla 10, anch'essa facente parte delle classi aventi maggior interesse e raffigurante la necessità di ulteriori accertamenti per il paziente. La *SVM* invece, performa peggio sotto ogni punto di vista, fatta eccezione per la classe 5. Si decide dunque di eleggere il *Random Forest* come miglior modello.

## Conclusioni e sviluppi futuri

Nel presente lavoro sono state eseguite tre classificazioni, con risultati soddisfacenti.

Per la classificazione binaria sono stati testati cinque modelli, con questi sono stati trattati efficacemente sia lo sbilanciamento delle classi sia la ridondanza degli attributi. La scelta finale è ricaduta sul *Random Forest*, il quale è risultato ottimo su tutte le metriche su cui è stato valutato.

Anche la classificazione multiclasse ha condotto alla scelta del *Random Forest*, sia per la 3 classi sia per la 10 classi. Il problema dello sbilanciamento è stato risolto grazie all'utilizzo di *oversampling*; la presenza di attributi ridondanti e poco informative è stata risolta utilizzando *Feature Selection* con metodi *Wrapper*; infine la scelta dei parametri è stata ottenuta con *Parameter Tuning*.

Le performance ottenute nel lavoro svolto possono certamente essere migliorate, tra i principali problemi si evidenzia lo sbilanciamento delle classi considerate, quindi la scarsità di osservazioni, di cui risentono in particolare i modelli multiclasse. Considerando che le etichette fornite sono state imputate manualmente, viene da sé comprendere la difficoltà di una tale operazione, l'introduzione di osservazioni prive di attributo target e la conseguente applicazione di tecniche di *Semi-Supervised Learning* potrebbero essere una buona soluzione a tale problema. Inoltre, la possibilità di un confronto con tecnici del settore potrebbe permettere l'introduzione di una matrice dei costi e il conseguente svolgimento di un'analisi *Cost-Sensitive*.

## Bibliografia

1. Peterek T., Gajdoš P., Dohnálek P., Krohová J. (2014) Human Fetus Health Classification on Cardiotocographic Data Using Random Forests. In: Pan JS., Snasel V., Corchado E., Abraham A., Wang SL. (eds) *Intelligent Data analysis and its Applications, Volume II. Advances in Intelligent Systems and Computing*, vol 298. Springer, Cham. [https://doi.org/10.1007/978-3-319-07773-4\\_19](https://doi.org/10.1007/978-3-319-07773-4_19)
2. <https://archive.ics.uci.edu/ml/datasets/cardiotocography>.

## Supplementary Note 1: Nota 1

Codice	Descrizione
1	Sonno calmo
2	Sonno in fase REM
3	Sveglia calmo
4	Sveglia attivo
5	Pattern con variazioni
6	Pattern di accelerazione/decelerazione (situazione di stress)
7	Pattern di decelerazione (stimolazione vagale)
8	Pattern ampiamente decelerativo
9	Pattern sinusoidale piatto (stato Patologico)
10	Pattern Sospetto

**Tabella 4.** Classi dell'attributo CLASS.