



UNIVERSITÀ  
POLITECNICA  
DELLE MARCHE

FACOLTÀ DI INGEGNERIA  
**Corso di Laurea in Ingegneria Biomedica**  
Dipartimento di Ingegneria dell'Informazione

***Classificazione dei tracciati cardiocografici con Random Forest***

*Classification of cardiocograph traces with Random Forest*

Relatore:

Prof.ssa **Burattini Laura**

Correlatore:

Dott. **Sbrollini Agnese**

Tesi di Laurea di:

**Rotariu Diana-Maria**

**A.A. 2020 / 2021**

## Abstract

La gravidanza è la condizione fisiologica in cui si trova la donna durante la quale, dentro di lei, si sviluppa un nuovo essere umano. In particolare, il cuore umano si forma nei primi giorni di gestazione; verso la fine della gravidanza, un feto può trovarsi in condizioni di sofferenza come l'asfissia o l'ipossia che indicano una diminuzione dell'ossigeno. Questa diminuzione può causare una variazione del battito cardiaco fetale. Il battito cardiaco fetale insieme alle contrazioni uterine vengono registrate da un cardiocotografo. La lettura del tracciato e la classificazione in una delle classi della linea guida della classificazione dei tracciati cardiocotografici mette in chiaro se e come si deve intervenire per salvaguardare la salute di un feto. Questa classificazione viene ancora fatta manualmente ma si è auspicabile introdurre un classificatore automatico come Random Forest per poter avere dei risultati oggettivi.

Random Forest è un metodo di Machine Learning capace di affrontare sia compiti di classificazione che di regressione. In particolare, Random Forest usa gli alberi decisionali per classificare i tracciati cardiocotografici. Ogni albero decide in quale delle 3 classi dovrebbe stare un tracciato. La decisione finale è data dal voto di maggioranza degli alberi decisionali. Random Forest ha una sensibilità molto grande sia per la classe normale che quella patologica come hanno evidenziato gli studi di H. Sahin e A. Subasi.

Uno dei primi modelli di classificatori Random Forest usato per classificare i tracciati cardiocotografici è stato fatto da P. Tomáš et al. Essi hanno evidenziato quanto sia importante dividere il set di dati iniziali in modo da avere risultati soddisfacenti.

A differenza degli altri metodi del Deep Learning, Random Forest ha un'accuratezza compresa tra 93.3% e 99%. Il modello di S. Santhiya et al. che è composto da un classificatore Random Forest che utilizza 300 alberi decisionali ha un'accuratezza di 94.92%. Le prestazioni del modello di M. Arif sono minore di quelle di S. Santhiya et al. ma garantiscono la possibilità di inserire un classificatore Random Forest in tutti gli ospedali.

Un modello Random Forest che può avere buoni risultati è formato da 300 alberi decisionali e da 21 parametri. Quindi il modello di M.M. Imran Molla et al. del 2021 risulta il modello che dovrebbe essere preso in considerazione per futuri lavori.

Quindi, in conclusione è auspicabile introdurre un classificatore Random Forest che è formato da 300 alberi e che tiene conto di tutti i parametri per poter diminuire la mortalità prenatale.

## Indice

I. Introduzione	5
1. Sviluppo fetale e gravidanza	7
1.1. Cenni di organogenesi	7
1.2. Circolazione fetale	8
1.3. Circolazione placentare	10
1.4. Travaglio e parto	11
1.5. Sofferenza fetale	11
2. La cardiotocografia	13
2.1. Descrizione di un cardiotocografo	13
2.2. Altri metodi di monitoraggio del benessere fetale	14
2.3. Caratteristiche fondamentali del CTG	15
2.4. Valutazione del tracciato cardiotocografico	17
2.5. Cardiotocografia in periodo espulsivo	19
3. Gli algoritmi Random Forest	21
3.1. Introduzione al Machine Learning	21
3.2. Alberi decisionali	22
3.3. Bagging	25
3.4. Stima dell'errore Out Of Bag	26
3.5. Random Forest	27
4. Previsione dello stato del benessere fetale attraverso gli algoritmi Random Forest	29
4.1. Studio di S. Santhiya et al.: Classificazione della sofferenza fetale con l'uso della cardiotocografia	29
4.1.1. Introduzione	29
4.1.2. Materiali e metodi	30
4.1.3. Risultati	35
4.1.4. Discussione e conclusione	36
4.2. Studio di M. Arif: Classificazione dei tracciati cardiotocografici con Random Forest e selezione dei parametri importanti di un tracciato cardiotocografico	37
4.2.1. Introduzione	37
4.2.2. Materiali e metodi	38
4.2.3. Risultati	44

4.2.4. Discussione e conclusione	45
4.3. Studio di M.M. Imran Molla et al.: Classificazione dei tracciati cardiocografici con Random Forest basato su algoritmi di Machine Learning	46
4.3.1. Introduzione	46
4.3.2. Materiali e metodi	47
4.3.3. Risultati	50
4.3.4. Discussione e conclusione	52
4.4. Studio di P. Tomáš et al: Classificazione dei tracciati cardiocografici con Random Forest	
4.4.1. Introduzione	53
4.4.2. Materiali e metodi	54
4.4.3. Risultati	56
4.4.4. Discussione e conclusione	57
4.5. Altri metodi di classificazione dei tracciati cardiocografici	58
4.5.1. Studio di H. Sahin e A. Subasi: Classificazione dei tracciati cardiocografici utilizzando metodi di Machine Learning per anticipare i rischi fetali	60
4.5.1.1. Introduzione	60
4.5.1.2. Materiali e metodi	60
4.5.1.3. Risultati	62
4.5.1.4. Discussione e conclusione	63
4.5.2. Studio di Z. Hoobhoy et al.: Uso del Machine Learning per la predizione della salute fetale usando tracciati cardiocografici	63
4.5.2.1. Introduzione	63
4.5.2.2. Materiali e metodi	63
4.5.2.3. Risultati	64
4.5.2.4. Discussione e conclusione	65
4.6. Come e perché si dovrebbero usare gli algoritmi Random Forest per classificare i tracciati cardiocografici	65
II. Conclusione	69
III. Bibliografia e Sitografia	70
IV. Ringraziamenti	72

## *I. Introduzione*

Durante la gravidanza, non è semplice acquistare direttamente informazioni che riguardi la salute del feto. Per questo già dal 1970 si è cominciato ad usare in tutti gli ospedali una macchina capace di registrare i battiti cardiaci fetali e le contrazioni uterine: il cardiotocografo.

Attraverso l'uso del cardiotocografo si possono registrare dei tracciati capaci di offrire al personale medico informazioni fondamentali delle condizioni del feto. Questo apparecchio tiene conto di tanti parametri del battito cardiaco fetale come, ad esempio, la linea di base (che deve rimanere sempre in un range di 110 bpm a 160 bpm), le accelerazioni, le decelerazioni e altri. Tutti analizzati in funzione delle contrazioni uterine, perché quest'ultime possono influenzare il giudizio finale della salute del bambino. Tenendo conto di tutte queste informazioni che si ottengono, il personale medico decide di classificare lo stato del feto in una delle tre classi: Normale, Sospetta, Patologica (per la classificazione FIGO) o Categoria I, Categoria II e Categoria III (per la classificazione ACOG).

La classificazione fatta in questo modo può essere soggetta all'errore umano. Per contrastare questa possibilità si può introdurre un classificatore automatico. Il deep learning usa algoritmi matematici e dati registrati dal cardiotocografo per offrire una diagnosi. I dati registrati vengono caricati in un computer e attraverso diversi processi si possono individuare tutti i parametri necessari a classificare i tracciati. I dati in ingresso saranno suddivisi in dati di allenamento (training dataset) e dati per testare l'approccio usato (test dataset). Le prestazioni di un simile approccio sono date dal confronto dei tracciati classificati con questi algoritmi e i tracciati classificati inizialmente dal personale medico.

L'intelligenza artificiale è stata usata in diverse ricerche come predire la recidività di un tumore, predire se una persona soffre di Alzheimer e altre ottenendo buoni risultati. In tutte queste ricerche i dati erano molto lunghi esattamente come si ha per i tracciati cardiotocografici. Ecco perché, l'idea di classificarli con un algoritmo di Machine Learning, chiamato Random Forest, capace di tenere in considerazione tutte le informazioni raccolte dai pazienti sembra ideale per questo lavoro.

L'algoritmo Random Forest usa molti alberi decisionali che classificano il tracciato in modo diverso tenendo conto dei parametri essenziali. Ogni albero decisionale decide in quale classe appartiene un tracciato. Se la maggior parte degli alberi decisionali classifica un tracciato in una specifica

classe, Random Forest darà come risultato finale della classifica proprio la classe predetta dagli alberi decisionali. Quindi si ha una visione e un'analisi molto più estesa delle informazioni ricevute dal cardiocardiografo.

In questa tesi si è voluto evidenziare come la possibilità di classificare i tracciati cardiocardiografici con Random Forest fosse l'approccio ideale per contrastare la mortalità infantile causata dall'errore umano. Le prestazioni di questo algoritmo trovate in letteratura sono molto soddisfacenti. Allo stesso tempo si proverà ad offrire diverse possibilità per diminuire il carico di informazioni ricevute inizialmente dal cardiocardiografo senza però diminuire le prestazioni. Si spera in questo modo di poter inserire un simile classificatore in tutti gli ospedali e quindi anche in luoghi remoti (dove non ci sono molti specialisti per classificare un tracciato) e contrastare i casi estremamente dolorosi di mortalità infantile.

## *Capitolo 1. Sviluppo fetale e gravidanza*

Gravidanza: condizione biologica in cui si trova la donna, dal giorno del concepimento al parto<sup>1</sup>. Questa definizione è una descrizione generale di quello che significa la gravidanza, ma non ci possiamo limitare solo a questo in quanto nelle 40 settimane di gestazione si ha un quadro molto più complesso riguarda non solo la donna ma anche la formazione del feto. E in un periodo così ampio si deve tenere conto di ogni aspetto che riguardi la salute sia della madre che del bambino. Ed in particolare si guarderà più ampiamente a come si può capire di più della salute del feto analizzando i battiti cardiaci che esso ha durante il periodo fetale.

### *1.1. Cenni di organogenesi*

Il cuore deriva embriologicamente dal mesoderma. Infatti, durante lo sviluppo embrionale si formano tre strati germinativi: l'ipoblasto, il mesoderma e l'ectoderma, che daranno origine a differenti parti dell'organismo. In particolare, dalla splancopleura che è il foglietto interno o viscerale della piastra laterale del mesoderma, si formerà il cuore, i grossi vasi sanguigni e i mesenterii (porzioni di peritoneo che collegano la splancopleura di rivestimento dell'intestino e la somatopleura).

Il sistema circolatorio funzionale è una necessità biologica, in quanto i tessuti dei mammiferi non possono far crescere più di pochi strati di cellule senza un apporto di sangue attivo. Ecco perché dopo solo 22-23 giorni dalla fecondazione il tubo cardiaco primitivo inizia a battere.

Il tubo cardiaco primitivo presenta restringimenti che lo dividono in cinque parti. Come si può notare in Figura 1.1, a partire dall'estremità cefalica si distinguono: il bulbo, diviso in tronco arterioso primitivo, cono e ventricolo primitivo destro; il ventricolo sinistro; l'atrio; il seno venoso. Da questa situazione, attraverso fenomeni di allungamento, torsione e sepimentazione, si arriva al cuore definitivo caratterizzato da un atrio e un ventricolo di destra separati da una valvola, e un atrio e un ventricolo di sinistra anch'essi separati da una valvola. I ventricoli destro e sinistro vengono separati da un setto interventricolare proveniente dalla superficie inferiore del bulbo. Il cono e il tronco arterioso vengono separati in due canali, polmonare e aortico. La divisione della cavità atriale rimane incompleta per tutta la durata dello sviluppo embrionale e fetale. Il foro tra i due atri viene denominato foro ovale di Botallo che alla nascita si occlude.<sup>2</sup>

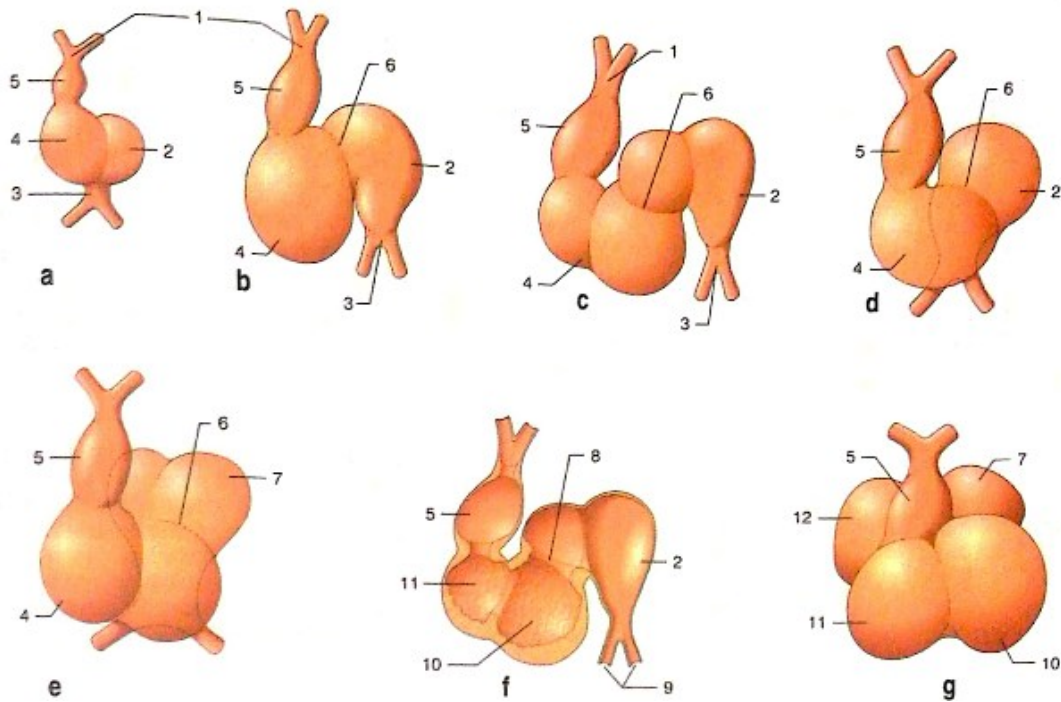


Figura 1.1. Sviluppo del cuore (28°-34° giorno). Si dimostrano le modificazioni di forma e di posizione che le diverse parti dell'abbozzo cardiaco subiscono in stadi successivi dello sviluppo (da a a g). 1, Tronco arterioso; 2, atrio; 3, seno venoso; 4, ventricolo; 5, bulbo cardiaco; 6, canale atrio ventricolare; 7, atrio sinistro; 8, orifizio del canale atrioventricolare; 9, vene della circolazione generale; 10, ventricolo sinistro; 11, ventricolo destro; 12, atrio destro. (G. Ambrosi, 2008)<sup>3</sup>

### 1.2. Circolazione fetale

La circolazione prenatale è molto diversa dalla circolazione postnatale (Figura 1.2), principalmente perché i polmoni non sono ancora in uso. Tutti gli scambi gassosi e l'apporto di materiali nutritizi si realizzano attraverso un annesso fetale, la placenta.

Come già detto, i due atri sono collegati tra di loro attraverso il forame ovale di Botallo, presente nel setto interatriale. Quindi, nella circolazione fetale non esiste una netta separazione tra sangue venoso e sangue arterioso.

Il sangue della madre entra nella placenta dove cede al feto i nutrienti per pressione idrostatica che vengono riversati in una vena fetale, la vena ombelicale. La vena ombelicale, insieme alle due più piccole arterie ombelicali costituisce il cordone ombelicale che arriva al feto nel punto in cui si formerà l'ombelico alcuni giorni dopo la nascita, dopo la caduta del cosiddetto moncone cordonale. L'ossigeno trasportato dai globuli rossi materni viene trasferito ai globuli rossi del feto, che contengono emoglobina più affine proprio all'ossigeno.<sup>4</sup>



## Circolazione fetale

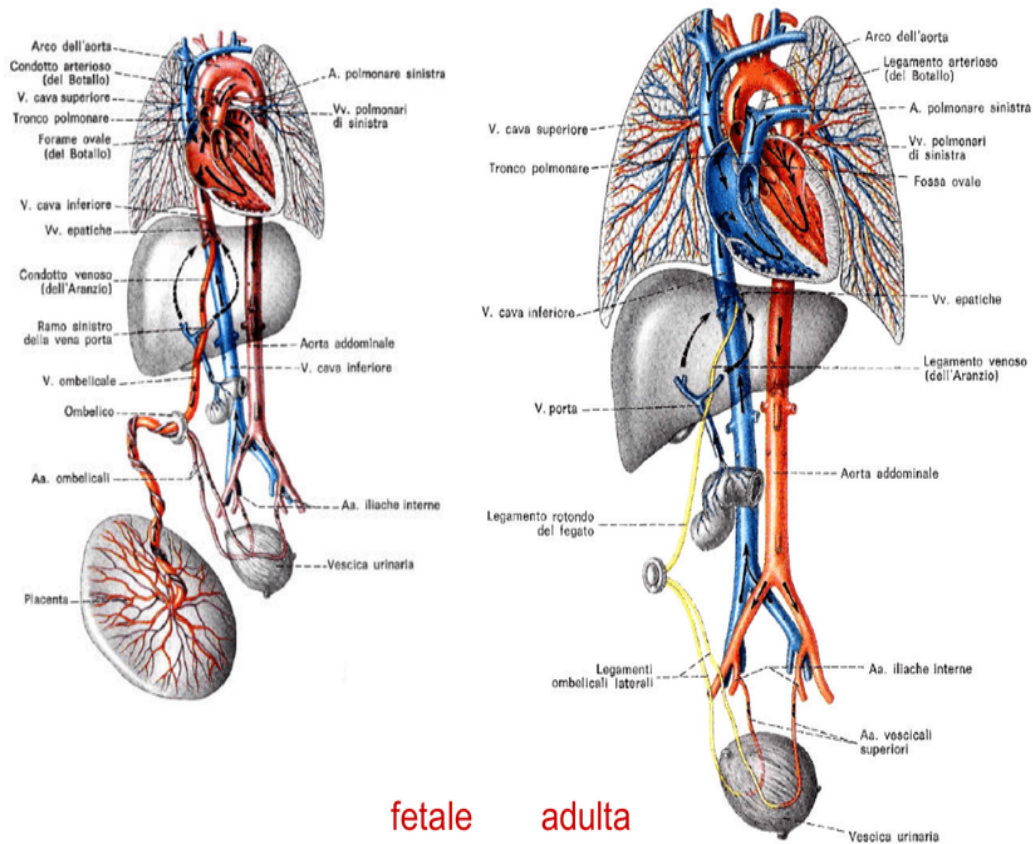


Figura 1.2 Organizzazione generale della circolazione fetale (sinistra) e organizzazione generale del sistema circolatorio dopo la nascita (destra). In rosso è indicato il sangue arterioso, in blu il sangue venoso, in giallo i vasi fetali che, obliterandosi, si trasformano in legamenti.

La vena ombelicale sale fino ad immergersi nella vena porta. Quest'ultima, nell'adulto entra nel fegato dove il sangue viene modificato. Dato che il sangue materno è già passato attraverso il fegato, nel feto questa tappa viene saltata grazie al dotto venoso di Aranzio, che collega la vena porta alla vena cava inferiore. Quest'ultima a sua volta trasportando sangue misto prevalentemente ossigenato, sbocca nell'atrio destro del cuore. All'atrio destro del cuore giunge anche il sangue 'venoso' della vena cava superiore, che passa nel ventricolo destro. A questo punto il sangue prende la via dell'arteria polmonare, ma i polmoni non svolgono le funzioni che sono proprie ad essi nella vita postnatale; quindi, l'arteria polmonare è collegata all'aorta mediante il dotto arterioso di Botallo, che permette al sangue di evitare il circolo polmonare e di passare direttamente nella circolazione generale.

Subito dopo la nascita il cordone ombelicale viene reciso; il forame ovale di Botallo, con i primi atti respiratori si chiude; i vasi ombelicali si obliterano e si trasformano in legamenti e la placenta viene espulsa subito dopo il parto.<sup>5</sup>

### *1.3. Circolazione placentare*

La placenta è un organo vascolare che connette il feto alla parete uterina, premettendo al primo il nutrimento, la termoregolazione, l'eliminazione dei prodotti di scarto e lo scambio gassoso.

La placenta si definisce strutturalmente nell'arco della quarta settimana di vita embrionale. Il massimo sviluppo avviene intorno al quinto mese di gravidanza, quando anche l'attività di barriera è massima. Dopo il quinto mese si ha un invecchiamento della placenta e l'attività di barriera diminuisce. Anche l'area di contatto tra sangue materno e fetale si riduce.

La placenta è costituita da una parte materna e una parte fetale, che è ricoperta dalle membrane amniocoriali che formano il sacco amniotico in cui il feto è avvolto e immerso nel liquido amniotico.

Da una ramificazione dei vasi sanguigni placentari derivano i villi che sono la più piccola unità funzionale della placenta.

Il sangue sistemico della madre viene portato ai cotiledoni dove avvengono gli scambi con il sangue fetale. Malgrado la stretta vicinanza tra il sangue materno e quello fetale, le due circolazioni non entrano mai in contatto (Figura 1.3), separati dalla sottile membrana che circonda i villi.

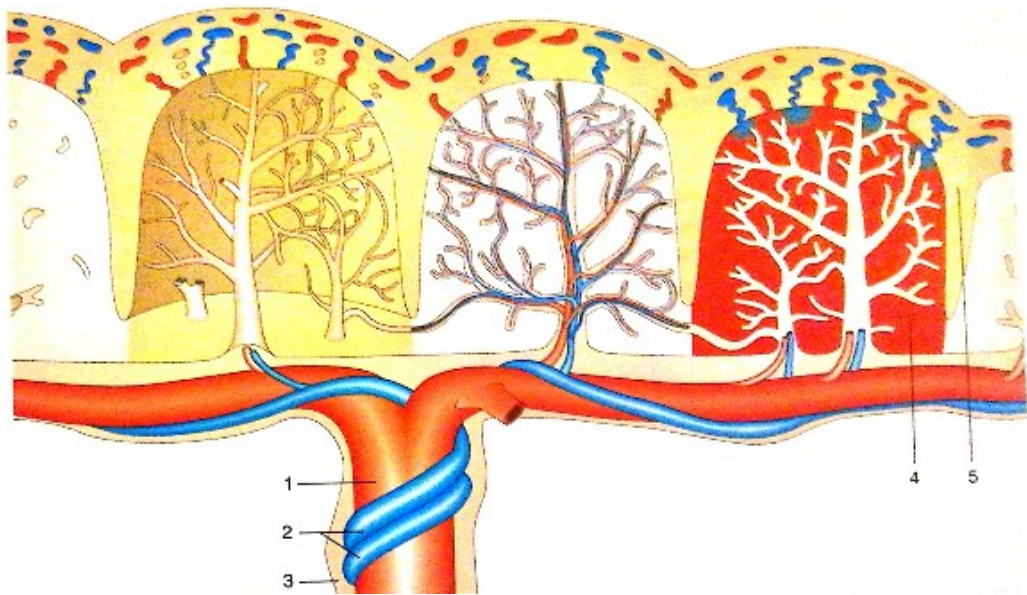


Figura 1.3 Schema della struttura della placenta umana, con la circolazione fetale e uterina che non vengono mai in contatto diretto. (G. Ambrosi, 2008)<sup>6</sup>

#### 1.4. Travaglio e parto

Il travaglio è il periodo in cui il corpo della madre si prepara all'espulsione del feto. La rottura delle membrane corioamniotiche o la comparsa di materiale ematico sono diagnostici per l'inizio del travaglio.

Per facilitare l'espulsione del feto, il collo dell'utero si deve dilatare circa 10 cm. Questo avviene grazie a delle contrazioni. Le contrazioni dette contrazioni di Braxton Hicks sono specifiche nel periodo della preparazione dell'utero al travaglio.<sup>7</sup>

La prima fase del travaglio è rappresentata proprio dalle contrazioni che aiutano alla dilatazione del collo dell'utero. In questo periodo è molto importante tenere in considerazione la frequenza cardiaca fetale che deve essere controllata continuamente con il monitoraggio elettronico.

Il secondo stadio del travaglio è il periodo che va dalla dilatazione completa del collo uterino all'espulsione del feto. Durante questo periodo tutte le strutture muscolari del pavimento pelvico sono sollecitate.

#### 1.5. Sofferenza fetale

Durante la gravidanza, il bambino dipende totalmente dalla madre, e quindi gli scambi che avvengono nella placenta sono essenziali per la sopravvivenza del feto. Un'alterazione di questo sistema può causare un calo della concentrazione di ossigeno a livello del sangue arterioso fetale

(ipossiemia), successivamente a livello tissutale (ipossia) ed infine a livello degli organi centrali (asfissia).

<<Ipossiemia è la fase iniziale del ridotto apporto di ossigeno e si caratterizza per la presenza di una ridotta saturazione di ossigeno nel sangue arterioso. Il feto si adatta ottimizzando i processi di captazione dell'ossigeno e riducendo i propri movimenti per risparmiare energie.

L'ipossia è la riduzione dell'ossigeno negli organi periferici, in conseguenza di un processo di centralizzazione del flusso volto a preservare gli organi nobili (cuore, sistema nervoso centrale).

L'asfissia è la riduzione dell'ossigeno a livello degli organi centrali.>><sup>8</sup>

Nel caso di una gravidanza normale, il battito cardiaco del feto oscilla tra 120 bpm e 160 bpm. Al di fuori di questi limiti si parla di bradicardia e tachicardia. Ci sono casi in cui si considera normale anche una frequenza maggiore di quello limite, quando il feto non ha ancora ben maturato i sistemi simpatico e parasimpatico. Normalmente durante il travaglio si ha una diminuzione del battito cardiaco dovuto anche alla diminuzione dell'area di scambio tra sangue materno e sangue fetale. Anche un aumento della frequenza sopra la linea di base di 15 bpm per più di 15 s ma meno di 600 s viene solitamente considerata normale perché è generalmente associato ai movimenti fetali. Invece se l'incremento avviene dopo una decelerazione si ha un probabile segno dell'inizio della compromissione fetale.

Alcune delle cause di bradicardia, valore inferiore a 110 bpm per più di 600 s sono l'ipotermia materna e il blocco atrioventricolare. Il blocco atrioventricolare è un difetto nel sistema di conduzione del cuore, che si verifica nella porzione tra atrio e ventricolo. In condizioni fisiologiche il nodo senoatriale determina il ritmo cardiaco generando segnali elettrici che si propagano fino al ventricolo. Nel caso, invece, di un blocco atrioventricolare l'impulso elettrico non raggiunge i ventricoli.

Per quanto riguarda la tachicardia, l'iperpiressia materna è la principale causa, essa può essere di origine extra-uterina o intra-uterina. Altre cause meno frequenti sono e le aritmie fetali come la tachicardia sopraventricolare e il flutter atriale.

Nell'ottobre 1998, La classificazione Internazionale delle Malattie identifica come condizione di sofferenza fetale la presenza di acidemia fetale metabolica, escludendo quindi le anomalie del ritmo o della frequenza cardiaca fetale e la tachicardia fetale dalla definizione. <sup>9</sup>

## *Capitolo 2. La cardiotocografia*

Fino alla seconda metà del XX° secolo la valutazione delle condizioni fetali si basava su mezzi molto limitati: la crescita dell'utero e del suo contenuto, i movimenti fetali percepiti dalla madre e l'auscultazione dei battiti cardiaci fetali tramite stetoscopio. Negli anni '50 del 1900 fu inventata un'apparecchiatura in grado di registrare in continuo sia la frequenza cardiaca fetale che le contrazioni uterine. Il metodo inizialmente complesso fu semplificato da Edward Henry Hon e successivamente si rese disponibile la possibilità di usare l'effetto doppler: date le caratteristiche di notevole costanza nella ricezione del segnale ultrasonico anche in condizioni non ideali. Dagli inizi degli anni '70 il cardiotocografo per il monitoraggio elettronico del feto viene utilizzato in tutti gli ospedali.<sup>10</sup>

### *2.1. Descrizione di un cardiotocografo*

La cardiotocografia fetale o CTG permette di monitorare la frequenza cardiaca del feto e le contrazioni uterine. Viene chiamato anche "esame no stress" in quanto è un esame non invasivo, basato sull'utilizzo di una sonda a ultrasuoni posizionata sull'addome della donna.

L'apparecchiatura che viene utilizzato è il cardiotocografo che è composto principalmente da un box centrale; da una sonda ad ultrasuoni (il cardio) per misurare il battito cardiaco fetale; da un trasduttore (il toco) che misura le contrazioni uterine e da carta per stampare. La sonda per misurare il battito cardiaco fetale e il trasduttore che misura le contrazioni uterine sono inserite su delle fasce elastiche che vengono messe sulla gestante al livello dell'addome.

Per rilevare il battito cardiaco fetale viene utilizzata la tecnica Doppler che rileva il segnale ad ultrasuoni di ritorno del battito cardiaco fetale e calcola la frequenza del battito cardiaco in tempo reale al minuto; per poi stamparlo. Si possono usare trasduttori esterni (scelta consigliata per la sicurezza del feto) oppure interni mediante un elettrodo che viene applicato sullo scalpo fetale

Le contrazioni uterine della donna vengono rilevate con un sensore di pressione, che misura le variazioni di pressione relative alle contrazioni uterine. Vengono rilevate sempre attraverso un trasduttore esterno.

L'apparecchio ha diverse velocità di scorrimento della carta: 1, 2 e 3 cm/min. Nella maggioranza dei Paesi Europei si utilizza la velocità di scorrimento della carta di 1 cm/min.<sup>11</sup>

## *2.2. Altri metodi di monitoraggio del benessere fetale*

La cardiotocografia è un esame non stress, ma nel caso in cui i parametri rilevati non sono rassicuranti si può ricorrere al test con stimolo (CST- Contraction Stress Test). Il CST si può effettuare con una stimolazione vibro-acustica con un trasmettitore applicato all'addome materno. Questo test si basa sull'insorgere delle contrazioni uterine per verificare se il feto potrà sopportare il parto. L'insorgenza delle contrazioni uterine comporta una riduzione del flusso ematico e quindi una transitoria ipossiemia. Se dalle contrazioni uterine non compaiono decelerazioni nel battito cardiaco fetale, l'esame viene considerato negativo. Invece se ci sono decelerazioni tardive, ci troviamo nella condizione di sofferenza fetale cronica che associata ad assenza di reattività fetale e a riduzione della variabilità del battito cardiaco, implica l'esistenza dell'acidosi metabolica.<sup>12</sup>

Lo stetoscopio ostetrico o di Pinard ha un diametro alla base di 5-6 cm e una conformazione ad imbuto che permette di raccogliere i toni ed i rumori del battito cardiaco fetale. Permette di percepire una varietà di manifestazioni sonore da attribuire al feto e alla madre. Questo metodo ha molti svantaggi, tra i quali ci sono: la difficoltà dell'ascoltazione durante le contrazioni uterine; valutazione soggettive basata sull'interpretazione del medico; richiede un'adeguata abilità sia nella tecnica di rilevazione che nell'interpretazione dei suoni.

L'ecocardiografia fetale è un esame che ha come finalità il controllo sequenziale dell'anatomia cardiaca allo scopo di evidenziare, o escludere, la presenza di una cardiopatia congenita.

Il sistema STAN (STaN= tratto ST analysis) viene usato solo durante il travaglio di parto; e associa al monitoraggio standard della CTG, la rilevazione e l'analisi dell'elettrocardiogramma fetale (ECG) ottenuto attraverso un apposito elettrodo che viene posizionato in corrispondenza dello scalpo fetale. Il sistema STAN identifica e analizza in maniera automatica le variazioni dell'onda T e del segnale ST provenienti dall'ECG fetale che riflette le correnti elettriche generate dal miocardio (onda P, complesso QRS, onda T).<sup>13</sup>

L'ecografia ostetrica con flussimetria è un esame ecografico usato per valutare la regolarità della crescita fetale e la sua simmetria al fine di escludere sia la presenza di malformazioni congenite o genetiche, sia situazioni di sofferenza fetale, grazie anche ad una valutazione semiquantitativa del liquido amniotico. Attraverso l'applicazione degli ultrasuoni emessi da una sonda, vengono

visualizzati graficamente gli organi interni oggetto d'indagine, la loro forma e le loro eventuali alterazioni.

La stimolazione dello scalpo fetale è un test biofisico associato alla cardiocografia per la diagnosi di acidosi fetale. Consiste nella stimolazione digitale o strumentale dello scalpo fetale. La comparsa di un'accelerazione della frequenza cardiaca fetale in risposta allo stimolo rappresenta un buon indicatore del benessere fetale. La non comparsa di accelerazione non rappresenta un segno certo di acidosi fetale ma si devono fare altri esami. <sup>14</sup>

### *2.3. Caratteristiche fondamentali del CTG*

Per valutare un tracciato CTG si devono esaminare diversi parametri come la linea di base, la variabilità, le accelerazioni, le decelerazioni del battito cardiaco fetale e le contrazioni uterine della madre.

La linea di base o frequenza cardiaca basale media è il livello medio della frequenza cardiaca fetale. Viene valutata in un intervallo di tempo di almeno 10 min ed è espressa in battiti per minuto. Secondo l'ACOG (American College of Obstetricians and Gynecologists) la frequenza cardiaca è normale se è compresa tra 110 e 160 bpm, si parla di tachicardia se al di sopra di 160 bpm e di bradicardia se è al di sotto di 110 bpm. Per RCOG (Royal College of Obstetricians & Gynaecologists) la frequenza cardiaca basale è considerata normale se compresa tra 110 e 160 bpm, sospetta se compresa tra 100 e 110 bpm o tra 160 e 180 bpm e patologica se è al di sotto di 100 bpm o al di sopra di 180 bpm.

La variabilità si riferisce alle oscillazioni della frequenza cardiaca fetale, definita come l'ampiezza media tra il picco più alto e quello più basso in un segmento pari ad un minuto di tracciato. Viene considerata normale se è compresa tra 5-25 bpm; ridotta se è inferiore a 5 bpm per più di 50 min; aumentata (pattern saltatorio secondo l'interpretazione della FIGO-International Federation of Gynecology and Obstetrics) se c'è un valore di ampiezza superiore a 25 bpm per più di 30 min. Secondo la RCOG la variabilità è considerata normale se è al di sopra di 5 bpm; sospetta se è al di sotto di 5 bpm per un periodo di osservazione di 40-90 min; patologica se è al di sotto di 5 bpm per più di 90 min. <sup>15</sup>

Le accelerazioni sono improvvisi aumenti della frequenza cardiaca fetale sopra la linea di base (picco in elevazione in meno di 30 s), maggiore di 15 bpm in altezza, della durata superiore a 15 s ma inferiore a 10 min, con successivo ritorno alla linea di base. Un'accelerazione è definita

prolungata se dura più di 2 min ma termina entro i 10 min, se dura per maggiore tempo, allora si parla di variazione della frequenza cardiaca basale. La maggior parte delle accelerazioni coincidono con i movimenti fetali e rappresentano un segno di attività neurologica fetale, che esclude l'ipossia fetale.

Le decelerazioni sono riduzione della frequenza cardiaca fetale sotto la linea di base, maggiore di 15 bpm di ampiezza e della durata superiore a 15 s. Secondo l'ACOG ci sono quattro tipi di decelerazioni da tenere sotto osservazione:

1. Le prime sono le decelerazioni precoci che sono delle decelerazioni poco profonde, di breve durata che hanno una variabilità normale e sono sincrone con le contrazioni uterine.
2. Le seconde sono le decelerazioni tardive che sono decelerazioni che hanno un inizio graduale dopo una contrazione uterina; hanno un lento ritorno alla linea di base e una ridotta variabilità.
3. Il terzo tipo è quello delle decelerazioni variabili che sono caratterizzate da un rapido calo, buona variabilità, rapido ritorno alla linea di base. Le decelerazioni variabili sono quelle con maggiore presenza durante il travaglio. Di solito non indicano una ipossia fetale grave ma associate ad una ridotta variabilità possono indicare problemi gravi della salute del feto.
4. Il quarto tipo è quello delle decelerazioni prolungate; nel caso in cui hanno una durata maggiore di 3 min possono essere indice di ipossia fetale. Infatti, le decelerazioni che superino i 5 min, con una frequenza cardiaca fetale fissa inferiore a 80 bpm e ridotta variabilità all'interno della decelerazione sono frequentemente associate a ipossia fetale.

Le contrazioni uterine sono molto importanti durante il travaglio e sono caratterizzate da un graduale incremento dell'attività uterina seguita da un decremento che durano insieme approssimativamente da 45 a 120 s. Nel caso in cui c'è un'eccessiva frequenza delle contrazioni si parla di tachisistolia definita come la presenza di più di 5 contrazioni in 10 min. L'identificazione delle contrazioni è molto importante in quanto può influenzare il battito cardiaco fetale. Con il tocodinometro si può valutare con attendibilità solo la frequenza delle contrazioni che possono essere tenute in considerazione mentre si osserva l'andamento della frequenza cardiaca fetale.<sup>16</sup>

Questi parametri sono osservabili in un tracciato cardiotocografico come si può vedere in Figura 2.1.



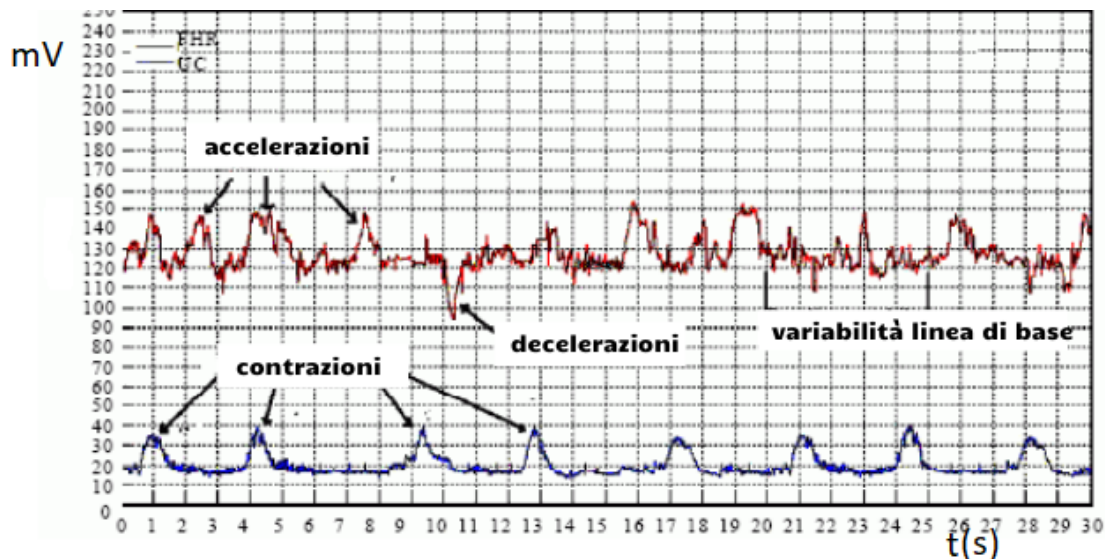


Figura 2.1 Esempio di un tracciato cardiocotografico in cui si possono osservare le accelerazioni, decelerazioni, la variabilità e le contrazioni uterine.

#### 2.4. Valutazione del tracciato cardiocotografico

I diversi parametri si devono tenere in considerazione allo stesso tempo per comprendere a pieno la salute del feto. Secondo l'ACOG ci sono tre categorie di tracciati che possono rivelare il benessere del feto come si può leggere dalla Tabella 1. Nella prima categoria (Categoria I) la frequenza cardiaca basale varia da 110-160 bpm; la variabilità è normale; ci possono essere accelerazioni o decelerazioni precoci ma non le decelerazioni tardive o variabile. La prima categoria si riferisce ad un normale equilibrio acido-base fetale.

La seconda categoria (Categoria II) ha le seguenti caratteristiche:

- la linea di base è minore di 100 bpm (bradicardia) ma non accompagnata da una variabilità legata alla frequenza cardiaca basale o è maggiore di 160 bpm (tachicardia)
- la variabilità è ridotta o aumentata o assente non accompagnata da delle decelerazioni
- le accelerazioni sono assenti quando il feto viene stimolato
- ricorrenti decelerazioni tardive con variabilità basale moderata o decelerazioni variabili con lento ritorno alla linea di base;

Può indicare un equilibrio acido-base anomalo

La terza categoria (Categoria III) ha le seguenti caratteristiche: assenza di variabilità con decelerazioni tardive ricorrenti o decelerazioni variabili ricorrenti o si ha una frequenza cardiaca minore di 100 bpm.

	TIPO 1	TIPO 2	TIPO 3
<b>Linea di base</b>	110-160 bpm	Mancanza di almeno una delle caratteristiche di normalità, ma assenza di segni patologici	< 100 bpm
<b>Variabilità</b>	5-25 bpm		Ridotta variabilità per più di 50 minuti e/o variabilità aumentata per più di 30 minuti
<b>Decelerazioni</b>	Assenza di decelerazioni ripetitive		Decelerazioni ripetitive, tardive o prolungate per più di 30 minuti o in caso di variabilità ridotta per più di 20 minuti; oppure decelerazione prolungata per più di 5 minuti
<b>Interpretazione</b>	Assenza di ipossia/acidosi fetale	Bassa probabilità di ipossia/acidosi fetale	Possibilità di ipossia/acidosi fetale
<b>Gestione clinica</b>	Nessun intervento necessario a migliorare lo stato di ossigenazione fetale	Intervento volto a correggere le cause reversibili di ipossia/acidosi fetale	Intervento immediato volto a correggere le cause reversibili, metodiche aggiuntive per valutare lo stato di ossigenazione fetale.

Tabella 1 (Monitoraggio cardiocografico in travaglio, 2018)<sup>17</sup>

Secondo invece la RCOG i tracciati si possono suddividere in: normale, sospetto e patologico. In questo modo si ha una spiegazione quasi diretta dell'andamento della gravidanza. Un tracciato è considerato normale quando tutti i parametri sono nella norma, è indice di benessere fetale. Un tracciato è considerato sospetto quando un solo parametro è anomalo, ma, se c'è variabilità l'acidosi fetale è molto improbabile. Invece un tracciato viene considerato patologico se sono presenti due o più caratteri non rassicuranti.<sup>18</sup>

Anche secondo NICE (National Institute of Health and Care Excellence) abbiamo tre tipi di tracciati:

1. rassicurante - quando la linea di base è compresa tra 110 e 160 bpm; la variabilità è tra 5 e 25 bpm; non ci sono decelerazioni; non ci sono specifiche riguardante le accelerazioni;
2. non rassicurante – quando la linea di base è compresa tra 161 e 180 bpm o tra 100 e 109 bpm anche se quest'ultima può essere considerata normale se la variabilità è normale e non ci sono decelerazioni variabili o tardive; la variabilità è minore di 5 bpm per più di 30 min ma meno di 50 min o maggiore di 25 bpm per più di 15 min ma meno di 30 min;
3. anomali – se la linea di base è minore di 100 bpm o maggiore di 180 bpm; la variabilità è minore di 5 bpm per più di 50 min o maggiore di 25 bpm per più di 25 min; presenta decelerazioni variabili o tardive.<sup>19</sup>

### *2.5. Cardiotocografia in periodo espulsivo*

Per valutare un tracciato CTG nel periodo espulsivo si utilizza la classificazione di Piquard (Figura 2.2) definita nel 1988. Si considera cinque tipi di tracciati con le seguenti caratteristiche:

1. Tipo 0 – non c'è nessuna anomalia, tutti i parametri del feto sono nella norma. Questo tipo di tracciato è molto raro;
2. Tipo 1 – c'è una decelerazione nel battito cardiaco fetale ad ogni spinta della madre; è un tracciato normale se la variabilità e la frequenza cardiaca fetale sono nella norma. Questo tipo di tracciato è molto frequente, si riscontra in quasi 85% dei casi;
3. Tipo 2 a – presenta delle decelerazioni ad ogni spinta della madre ma a differenza del tracciato di Tipo 1, la frequenza cardiaca basale è compresa tra 90 bpm e 120 bpm, viene considerata una bradicardia lieve se dura più di 10 min. Questo tipo non si riscontra in tanti casi e la sua gravità non è alta;
4. Tipo 2 b – presenta una riduzione della frequenza cardiaca basale al di sotto di 90 bpm con una riduzione nella variabilità (bradicardia grave e prolungata). Questo tipo di tracciato è molto grave in quanto può indicare un'ipossia fetale;
5. Tipo 3 – presenta una bradicardia severa con accelerazioni prolungate durante le contrazioni uterine. Anche i tracciati di Tipo 3 vengono considerati molto gravi;
6. Tipo 4 – presenta una bradicardia terminale severa che dura pochi minuti. Dato che ha una durata di pochi minuti non viene considerato un tracciato preoccupante.

Prima di procedere all'interpretazione del tracciato CTG in periodo espulsivo bisogna sempre tenere conto del più ampio contesto clinico, valutando anche le caratteristiche dei tracciati precedenti.

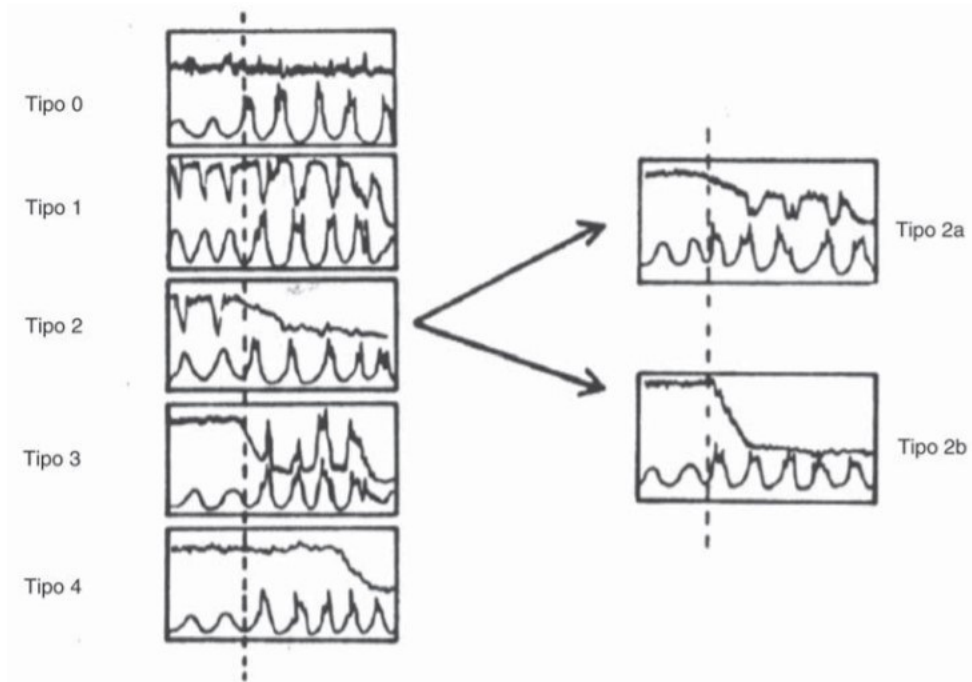


Figura 2.2 Esempi di tracciati classificati secondo Piquard

## *Capitolo 3. Gli algoritmi Random Forest*

Random Forest è un metodo di machine learning, capace di affrontare sia compiti di classificazione che di regressione. È uno dei metodi d'insieme più diffusi, appartenenti alla categoria dei metodi di bagging.

### *3.1. Introduzione al Machine Learning*

Il Machine Learning è una branca dell'intelligenza artificiale che consente di analizzare i dati e migliorare le performance di determinati algoritmi attraverso l'esperienza. Nasce dalla teoria che i computer possono imparare ad eseguire compiti specifici senza essere programmati per farlo.

L'obiettivo principale del Machine Learning è che una macchina sia in grado di svolgere ragionamenti induttivi, generalizzando i dati attraverso l'esperienza.

L'aspetto più importante del machine learning è la ripetitività, perché più i modelli sono esposti ai dati, più sono in grado di adattarsi in modo autonomo. I computer imparano da elaborazioni precedenti per produrre risultati e prendere decisioni che siano affidabili e replicabili.

<<Un programma apprende da una certa esperienza E se: nel rispetto di una classe di compiti T, con una misura di prestazione P, la prestazione P misurata nello svolgere il compito T è migliorata dall'esperienza E.>><sup>20</sup>

Gli algoritmi di Machine Learning sono tipicamente classificati in tre principali categorie, in base alla natura dell'apprendimento:

- **Apprendimento Supervisionato:** è una tecnica di machine learning che mira a istruire un sistema informatico in modo da consentirgli di elaborare automaticamente previsioni sui valori di uscita di un sistema rispetto ad un input sulla base di una serie di esempi ideali, costituita da coppie di input e di output, che gli vengono inizialmente fornite. Il computer è addestrato tramite un training set, un insieme di attributi relativi ad un elemento in input, ai quale è associato un valore di output, che indica il risultato di tale elemento. Il fine di questo approccio è quello di insegnare al computer una regola generale che mappi l'andamento degli attributi di input a quel determinato risultato. Una volta individuato il modello migliore si passa alla fase di testing, in cui l'algoritmo utilizza il modello trovato su una serie di dati di prova che non hanno un risultato associato.

- **Apprendimento Non Supervisionato:** è una tecnica di machine learning che consiste nel fornire al sistema informatico una serie di input che egli riclassificherà ed organizzerà sulla base di caratteristiche comuni per cercare di effettuare ragionamenti e previsioni sugli input successivi. A differenza dell'apprendimento supervisionato, durante l'apprendimento non supervisionato vengono forniti alla macchina solo esempi non annotati, in quanto le classi non sono noti a priori ma devono essere apprese automaticamente.
- **Apprendimento per rinforzo:** Con questo approccio il programma interagisce con un ambiente dinamico, dal quale riceverà stimoli a seconda della correttezza della scelta compiuta. Il sistema riceverà un premio nel caso la scelta sia corretta, mentre riceverà una penalizzazione nel caso non lo sia. L'apprendimento del modello migliore deriverà dal conseguimento del maggior premio possibile.<sup>21</sup>

Un'altra categorizzazione si può fare prendendo in considerazione gli output desiderati:

- **Classificazione:** nella classificazione gli output sono divise in due o più classi e il sistema di apprendimento deve produrre un modello che assegni gli input non ancora visti in una delle classi.
- **Regressione:** la regressione è un processo statistico che cerca di stabilire una relazione tra due o più variabili. Fornito a un modello di regressione un valore  $x$ , questo restituirà il corrispondente valore  $y$  generato dall'elaborazione di  $x$ .
- **Clustering:** nel clustering un insieme di inputi viene diviso in gruppi, ma diversamente da quanto accade per la classificazione, i gruppi non sono noti prima, rendendolo un compito tipicamente non supervisionato.

### *3.2. Alberi decisionali*

Un albero decisionale (Figura 3.1) è composto da una serie di rami ognuno derivante da un nodo e da foglie. In particolare: i rami (o archi) costituiscono l'insieme delle proprietà che determinano il percorso all'interno dell'albero; i nodi rappresentano una variabile; le foglie sono i risultati intermedi o finali, i luoghi dove finiscono i dati una volta separati.

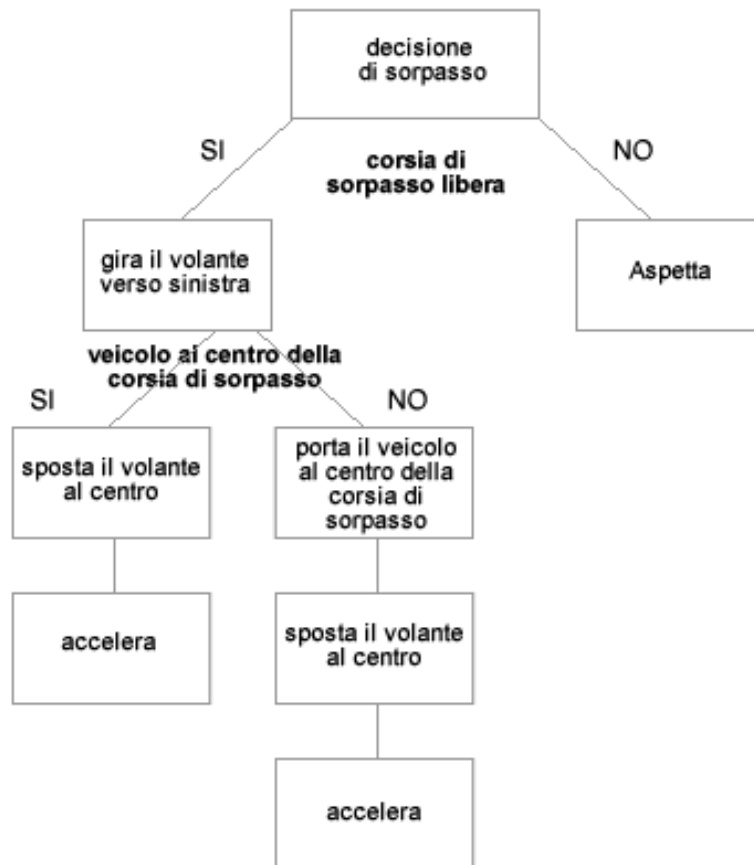


Figura 3.1. Esempio di un albero decisionale

Nel campo del machine learning un albero decisionale è un modello che utilizza una struttura dati ad albero per contenere le informazioni ed effettuare le predizioni. È un algoritmo di tipo supervisionato e un modello predittivo che può essere usato sia per casi di classificazione che per casi di regressione.

In generale l'obiettivo di un albero decisionale è quello di dividere i dati iniziali per il valore di una variabile che permette di creare due gruppi il più omogenei possibile internamente e il più disomogenei possibile tra di loro.

Un albero decisionale può essere immaginato come una struttura gerarchica formata da un insieme di nodi correlati da rami orientati ed 'etichettati'. La classificazione delle istanze dei dati, avviene ordinandole dalla radice ad una certa foglia, la quale rappresenta il risultato predetto a partire dai valori dei vari attributi e quindi la classificazione del dato in input. Ogni nodo effettua un test su un attributo, e ogni ramo uscente da un nodo corrisponde ad uno dei valori che

l'attributo può assumere. Ciascuna regola di predizione è rappresentata dal cammino dal nodo radice al nodo foglia.

La costruzione di un albero decisionale si basa su due principali fasi:

1. La fase di building, ovvero di costruzione. In questa fase avviene l'accrescimento delle dimensioni dell'albero in termini di archi e nodi per poter definire delle proprietà o regole di splitting (archi), in grado di identificare delle classi omogenee in termini di output (nodi). La crescita di questo albero può essere fermata inserendo un criterio di arresto (o criterio di halting); al manifestarsi di condizioni prestabilite si interrompe la crescita dell'albero. La crescita di un albero dipende, essenzialmente, dal numero di nodi che vengono inseriti, al fine di raggiungere una classificazione omogenea dei dati di input. Ad ogni nodo corrisponde un attributo che consente di classificare il set di dati in input, in relazione all'attributo target, ovvero l'informazione che l'albero deve predire. La prima difficoltà per un albero decisionale è la scelta della variabile con cui effettuare la prima divisione del dataset; per questo si considera il concetto di purezza, ed è proprio la sua misura che caratterizza le differenti misurazioni per identificare il criterio con cui effettuare la divisione. Le misurazioni di riferimento sono il *coefficiente di impurità di Gini* e il *guadagno informativo* (o *Information Gain*). Il primo è un indicatore della frequenza con cui un elemento casuale sia indentificato in modo non coretto. Viene calcolato per ciascun nodo fino a raggiungere il suo minimo (zero) quando tutti i casi del nodo rientrano nella stessa categoria. Il *guadagno informativo* è un indicatore di quanto la scelta di uno specifico attributo consenta di ridurre l'entropia. Un attributo con il più alto guadagno informativo si divide per primo.
2. La fase di pruning, ovvero di potatura. Tale fase è finalizzata a ridurre le dimensioni dell'albero per prevenire il rischio di overfitting al fine di ottimizzare le dimensioni dell'albero decisionale.

Gli alberi decisionali usano un metodo chiamato Partizione Ricorsiva, generalmente noto come "divide and conquer" perché in base al valore della variabile suddivide i dati in sottogruppi sempre più piccoli. Cominciando dal nodo radice, che rappresenta l'intero dataset che si ha a disposizione, l'algoritmo sceglie la variabile più accurata per ottimizzare la previsione della variabile target scelta, e i dati vengono suddivisi in rami diversi a seconda delle specifiche richieste dalla variabile.



L'algoritmo procede così per ogni nodo fino a quando si raggiunge la dimensione prefissata o non ci sono più variabili per distinguere i dati.

### 3.3. Bagging

Il bagging è una tecnica di Machine Learning che rientra nella categoria dell'apprendimento d'insieme, la cui caratteristica principale è l'utilizzo di una serie di metodi d'insieme che usano modelli multipli per avere un'accuratezza maggiore e una miglior prestazione predittiva rispetto al modello originale.

Il bagging viene usata in modo particolare con gli alberi decisionali, dove aumenta in modo significativo la stabilità dei modelli nella riduzione della varianza e nel miglioramento della precisione, eliminando l'overfitting. Il bagging richiede una spiegazione iniziale della tecnica bootstrap. Infatti, il termine bagging deriva dalla combinazione delle parole inglesi bootstrap (ovvero il campionamento casuale con rimpiazzo) e aggregation (in riferimento all'aggregazione di più modelli, tipico dell'apprendimento d'insieme).

Il bootstrap è una tecnica statistica di ricampionamento con reimmissione utilizzata per approssimare la distribuzione campionaria di una statistica. Permette di approssimare media e varianza di uno stimatore e di costruire intervalli di confidenza. Si basa sul fatto paradossale che l'unico campione disponibile serve per generarne molti altri e per costruire la distribuzione teorica di riferimento. Partendo da un dataset iniziale attraverso vari ricampionamenti, crea nuovi dataset, ciascuno con proprie caratteristiche, come ad esempio media e deviazione standard.

Il processo funziona come segue:

- La probabilità di assegnare un elemento del dataset finale in uno dei dataset finali è la stessa per ogni elemento
- Il ricampionamento è con reimmissione, quindi è possibile trovare lo stesso elemento più volte nello stesso dataset finale
- La dimensione e il formato dei dataset finali sono gli stessi di quelli iniziali.

È possibile utilizzare il metodo di bootstrap per un'ampia varietà di dati e anche con campioni di bassa numerosità.

Il bagging è una tecnica utile per gli alberi decisionali in quanto se quest'ultimi sono troppo grandi hanno una varianza elevata e se sono piccolo hanno un bias elevato. Il bagging riduce queste due variabili.

Il primo passo del metodo bagging è quello di creare diverse campioni partendo da quelli iniziali, esattamente come è per il metodo bootstrap. Nel caso degli alberi di regressione si calcola la media dei risultati ottenuti mettendo i campioni del primo passo in un albero di regressione. Invece per gli alberi di classificazione, si adatta un "comitato" di alberi, ognuno dei quali esprime un voto per i campioni ottenuti.

### 3.4. Stima dell'errore Out of Bag

La tecnica di Bootstrap tende a utilizzare il 66% degli elementi della popolazione iniziale per la creazione di diversi dataset, il restante viene utilizzato per valutare l'accuratezza del modello e calcolare il livello di errore dell'albero. Le osservazioni che costituiscono questo 34% sono le cosiddette (Out of Bag observations). Su queste osservazioni si farà una valutazione esattamente come per un albero decisionale.

La stima dell'errore finale, dipende dal numero di ricampionamenti effettuati per la creazione dei dataset finali, quindi dal numero di training set risultanti dal bagging e dal numero di alberi generati. Un numero maggiore di alberi generati si traduce in una riduzione dell'errore (Figura 3.2).

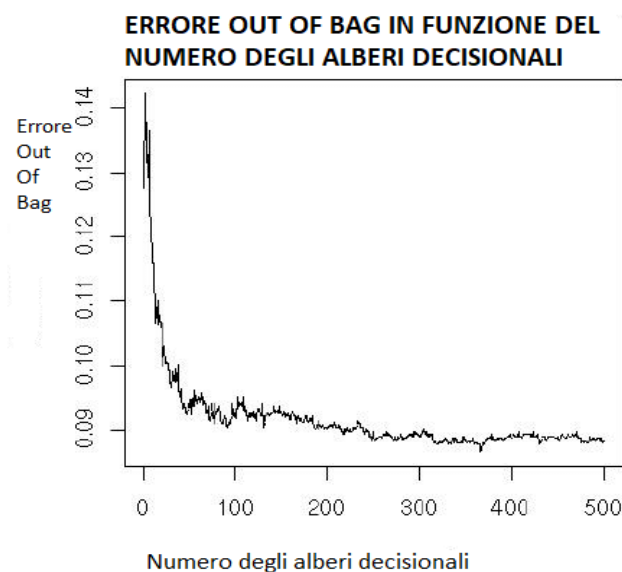


Figura 3.2 All'aumentare del numero degli alberi decisionale l'errore Out of Bag diminuisce

### 3.5. Random Forest

Il Random Forest è un caso particolare del Bagging che usa insieme diversi alberi decisionali. La decisione finale sarà data dalla scelta della maggioranza degli alberi. Offre così molti più punti di vista sullo stesso problema.<sup>22</sup>

Prendendo come esempio la scelta che un liceale deve fare alla fine della maturità si potrà comprendere meglio come funziona la tecnica Random Forest. Il liceale è indeciso fra dieci università, quindi, fa delle domande sia ai suoi professori che alle persone che li sta vicino per aiutarlo a decidere il percorso migliore. Alla fine, decide di scegliere l'università scelta dalla maggior parte delle persone interrogate. In questo esempio le persone sono come interi alberi decisionali appartenenti a Random Forest, le domande sono i rami di ogni albero, la decisione finale è presa analizzando la decisione di ogni albero (Figura 3.3).

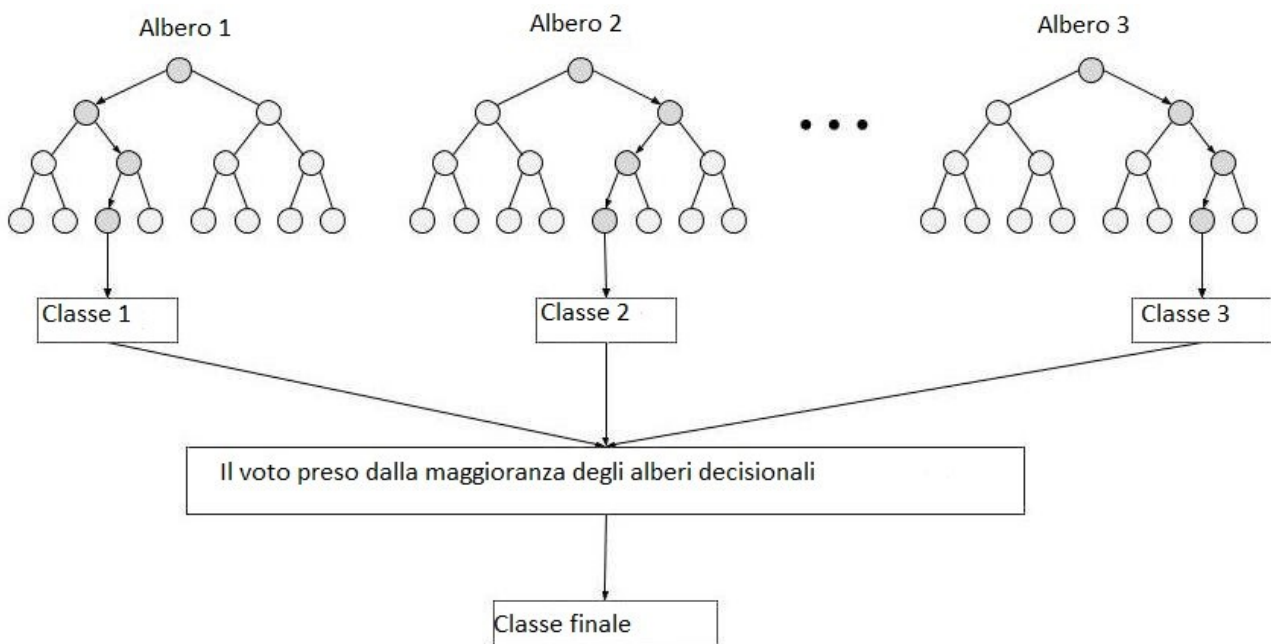


Figura 3.3 Esempio di un albero decisionale

La costruzione di Random Forest ha tre fasi principali:

1. Creazione di un set di dati bootstrap per ogni albero. Principalmente si seleziona in modo casuale i campioni dal set di dati originali. Lo stesso campione può essere selezionato più di una volta. Quindi grazie al metodo bootstrap si avrà diversi set di dati di addestramento.
2. Creazione di alberi decisionali. Nel caso della costruzione di un solo albero si utilizzava il coefficiente di impurità di Gini o il guadagno informativo per fare il primo

splitting. In Random Forest insieme a uno di questi fattori si aggiunge anche una scelta aleatoria per ogni nodo per impedire di includere funzionalità che hanno un potere predittivo molto elevato, creando molti alberi non correlati. In questo modo si riduce la varianza avendo una maggiore diversità tra gli alberi.

3. Si ripete la stessa procedura per tutti gli alberi creati.

Si ottiene in questo modo una foresta casuale in cui tutti i dati vengono inseriti e analizzati in maniera diversa in un albero decisionale, ognuno con un proprio risultato. Il risultato finale sarà dato dal voto preso dalla maggioranza degli alberi. In questo modo si ha un'analisi completa dei dati iniziali.

## *Capitolo 4. Previsione dello stato di benessere fetale attraverso gli algoritmi Random Forest*

La previsione dello stato del benessere fetale partendo dalla lettura di un tracciato cardiocografico può essere fatta sia dal personale medico che con diversi algoritmi del deep learning. Nella maggior parte degli articoli che si prenderanno in considerazione, Random Forest è il classificatore con le migliori prestazioni. Infatti, l'accuratezza di questo metodo oscillerà tra 93.3% e 99.18% (nel caso in cui si prenderanno in considerazione solo due tipi di tracciati: normali e patologici).

Analizzando più in profondità il modo in cui questo tipo di algoritmo viene applicato per classificare i tracciati cardiocografici si evidenzierà quanto un simile approccio può garantire risultati sorprendenti. I seguenti studi mettono in chiaro quanto sia importante l'utilizzo di un classificatore automatico per facilitare il giudizio finale del personale medico.

### *4.1. Studio di S. Santhiya et al.: Classificazione della sofferenza fetale con l'uso della cardiocografia<sup>23</sup>*

S. Santhiya et al.<sup>23</sup> costruiscono un modello di Machine Learning, in particolare un algoritmo Random Forest che classifica i tracciati ottenuti da un cardiocografo in tracciati Normali, Sospetti e Patologici. In particolare, evidenziano se un feto si trova in condizioni di sofferenza oppure no.

Per poter analizzare meglio i tracciati cardiocografici, S. Santhiya et al. hanno diviso il loro lavoro in 3 parti. La prima si basa sull'osservazione dei dati ricevuti da un cardiocografo e l'analisi dei parametri essenziali per poter classificare i tracciati in Normale, Sospetto e Patologico. La seconda parte è l'analisi del modello Random Forest per un CTG. Nell'ultima parte si ha la spiegazione del modello passo dopo passo. Queste parti sono fondamentali per poter creare un classificatore adatto per questi tipi di dati.

#### *4.1.1. Introduzione*

La sofferenza fetale è una condizione di emergenza che si può verificare durante il travaglio e il parto. È una complicazione nella quale un feto può essere soggetto di ipossia o asfissia. Se un feto si trova in queste condizioni la frequenza del suo battito cardiaco può cambiare drasticamente, i suoi movimenti possono rallentare e in condizioni molto gravi possono rimanere danni

permanenti. Il personale medico deve tenere in considerazione tutti i cambiamenti che un feto può avere durante la gravidanza. C'è bisogno di un continuo monitoraggio della salute del feto durante tutta la gravidanza utilizzando il CTG.

L'idea di S. Santhya et al. è quella di analizzare dei dati ricevuti da un CTG per poter creare un classificatore che usa gli algoritmi del Machine Learning per capire se un feto si trova in condizioni di sofferenza e quindi c'è bisogno dell'intervento da parte del personale medico o la gravidanza può proseguire in modo normale.

Gli autori vogliono spiegare in particolare come possono essere usati gli algoritmi Random Forest per classificare un tracciato cardiocografico e quindi osservare se un feto si trova in condizioni pericolose per la sua salute. Allo stesso tempo vogliono provare se l'utilizzo di meno parametri può garantire una classificazione adeguata oppure si dovranno tenere sempre in considerazione tutti i dati ricevuti da un cardiocografo.

#### *4.1.2. Materiali e metodi*

Il set di dati iniziali che sono stati presi in considerazione nel lavoro di S. Santhiya et al.<sup>23</sup> è formato da 2126 tracciati, tutti provenienti da diversi feti. Per ogni tracciato si prendono in considerazione 40 parametri, dei quali 10 sono dei parametri morfologici e 3 sono le tre classi nelle quali si può trovare un feto. Questi parametri possono essere letti nella Tabella 4.1.

Avendo a disposizione il set di dati iniziali si potrà cominciare a ragionare sugli algoritmi Random Forest. Come si sa dai capitoli precedenti questo tipo di metodo di Intelligenza Artificiale è molto utile in quanto, essendo formato da più alberi decisionali degli quali ognuno avrà una decisione finale, la decisione finale di tutto il processo sarà data dal maggior voto ottenuto per una certa classe. Questo tipo di approccio è ideale per i set di dati molto lunghi, come nel caso dei tracciati cardiocografici.

Prima di andare più a fondo per quanto riguarda gli algoritmi Random Forest, si dovrà fare un preprocessing dei dati iniziali per facilitare la costruzione del modello.

<b>Nome del File</b>	Il nome del tracciato cardiocografico	<b>NMAX</b>	Numero dei picchi nell'istogramma
<b>Data</b>	La data dell'acquisizione del tracciato	<b>NZEROS</b>	Numero degli zeri nell'istogramma
<b>b</b>	Istante iniziale	<b>MODE</b>	Modulo dell'istogramma
<b>E</b>	Istante finale	<b>MEAN</b>	Media dell'istogramma
<b>LBE (bpm)</b>	Linea di base (valutata da un medico)	<b>MEDIAN</b>	Mediana dell'istogramma
<b>LB (bpm)</b>	Linea di base (valutata dal programma SisPorto)	<b>VARIANCE</b>	Varianza dell'istogramma
<b>AC (bpm)</b>	Accelerazioni (SisPorto)	<b>TENDENCY</b>	Tendenza dell'istogramma: -1 = asimmetria di sinistra; 0 = simmetria; 1 = simmetria di destra
<b>FM (colpi/minuti)</b>	Movimenti fetali (SisPorto)	<b>A</b>	Sonno tranquillo
<b>UC (MVU)</b>	Contrazioni uterine (SisPorto) (unità Montevideo)	<b>B</b>	Sonno REM
<b>ASTV (%)</b>	% del tempo con variabilità ridotta (SisPorto)	<b>C</b>	Vigilanza tranquilla
<b>MSTV (bpm)</b>	La media della variabilità ridotta (SisPorto)	<b>D</b>	Vigilanza attiva
<b>ALTV (%)</b>	% del tempo con variabilità aumentata (SisPorto)	<b>SH</b>	Shift pattern (classe A o sospetta)
<b>MLTV (bpm)</b>	La media della variabilità aumentata	<b>AD (bpm)</b>	Accelerazioni o decelerazioni (nelle situazioni di stress)
<b>DL (bpm)</b>	Decelerazioni leggere	<b>DE (bpm)</b>	Decelerazioni – stimolazione del nervo vago
<b>DS (bpm)</b>	Decelerazioni severe	<b>LD (bpm)</b>	Ampie decelerazioni
<b>DP (bpm)</b>	Decelerazioni prolungate	<b>FS (bpm)</b>	Stato patologico – pattern sinusoidale piatto
<b>DR (bpm)</b>	Decelerazioni ripetitive	<b>SUSP</b>	Pattern sospetto
<b>WIDTH</b>	Ampiezza dell'istogramma	<b>CLASS</b>	Codice della classe (da 1 a 10) per i parametri da A a SUSP)
<b>MIN (Hz)</b>	Frequenza minima dell'istogramma	<b>NSP</b>	Normale = 1; Sospetto = 1; Patologico = 3
<b>MAX (Hz)</b>	Frequenza massima dell'istogramma		

Tabella 4.1 Parametri presenti nei set di dati del CTG<sup>23</sup>

Il preprocessing dei dati è fatto nel seguente modo:

1. Prima di tutto, i dati iniziali vengono caricati e analizzati

2. Usando la funzione `dropna()` si toglieranno tutti i valori nulli che si incontreranno nel set di dati iniziali
3. Tutti i parametri vengono classificati in ordine di performance
4. Usando la funzione `as.factor()` il parametro NSP diventa una variabile categorica.

Anche facendo il preprocessing dei dati iniziali, si ha comunque a disposizione un numero troppo alto di parametri, quindi l'analisi di tutti i tracciati sarebbe molto complessa. Ecco perché in questo studio sono stati eliminati i parametri per i quali l'eliminazione non cambia l'accuratezza della classificazione finale. Per eliminare i parametri adatti prima di tutto si è costruito un modello Random Forest con valori predefiniti per `mTry` e `nTree`. Dove `mTry` è il numero di variabili campionate casualmente come candidate ad ogni split e `nTree` è il numero di alberi in crescita. Il secondo passo di questo processo è quello di costruire un diagramma (Figura 4.1) con questo modello per poter osservare quale parametro ha una performance maggiore e quale no. Nell'ultimo punto si eliminano i parametri che non hanno un'importanza elevata, per questo si prenderanno in considerazione solo 21 parametri e la classi NSP (Normale, Sospetto, Patologico).

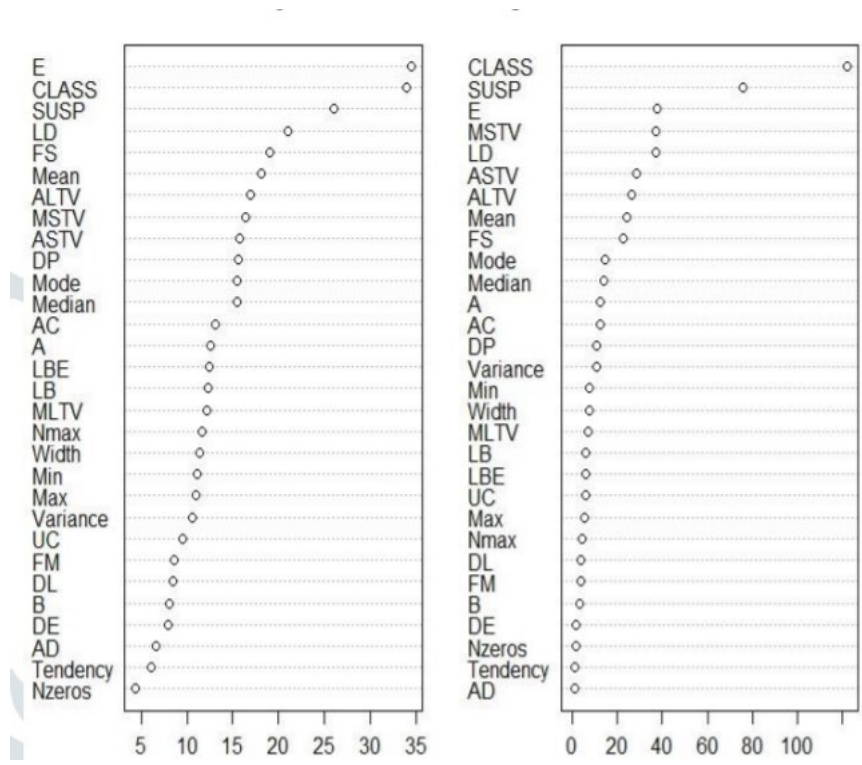


Figura 4.1 Diagramma dei parametri in ordine decrescente di importanza. Il primo diagramma si basa su quanto diminuisce la media dell'accuratezza. Il secondo diagramma si basa su quanto diminuisce il coefficiente di Gini<sup>23</sup>



I 21 parametri sono: LB, AC, FM, UC, DL, DS, DP, ASTV, MSTV, ALTV, MLTV, WIDTH, MIN, MAX, NMAX, NZEROS, MODE, MEDIAN, VARIANCE, TENDENCY.

A questo punto si può iniziare la costruzione del modello Random Forest. Nel primo tentativo si costruirà un modello Random Forest con i parametri predefiniti. Quindi si avrà: numero di alberi = 500, numero di variabili usati per ogni split = 4, e l'errore Out of Bag (OOB) sarà di 5.64% di tutto il data set. Per questo primo modello si potrà osservare la matrice di confusione nella Tabella 4.2. I tracciati che inizialmente erano stati classificati di categoria 1 (Normale) e che sono veramente di questa categoria sono 1140. Invece 16 tracciati di categoria 2 (Sospetto) sono stati erroneamente considerati di categoria 1. Solo 4 tracciati di categoria 3 (Patologico) sono stati considerati di categoria 1. L'errore della prima categoria è di 0.02. Per la seconda categoria l'errore è molto più grande in quanto anche quando i tracciati sono stati inizialmente classificati dai medici, erano considerati o di categoria 1 o di categoria 3.

Previsioni/ riferimento	1	2	3	Errore della classifica
1	1140	16	4	0.02
2	46	164	3	0.23
3	7	8	107	0.12

Tabella 4.2 Matrice di confusione del modello Random Forest con parametri predefiniti<sup>23</sup>

A questo punto si dovranno cambiare i parametri predefiniti. Per prima cosa si dovrà decidere il numero adatto per gli alberi decisionali. A questo scopo gli autori di questo lavoro hanno usato il modello Random Forest rf creando un grafico che offre il valore ottimale che si dovrà usare per il numero degli alberi decisionali. Questo grafico è riportato in Figura 4.2.

Da questo tracciato si deduce che il numero degli alberi decisionali più adatto per questo lavoro è 300.

Per quanto riguarda il numero delle variabili usate per ogni split S. Santhiya et al. hanno guardato la relazione che c'è tra l'errore Out of Bag e mTry, trovando in questo modo che il numero più adatto è 8. Infatti, quando si considerano 8 variabili per ogni split l'errore Out of Bag è molto più piccolo che negli altri casi.

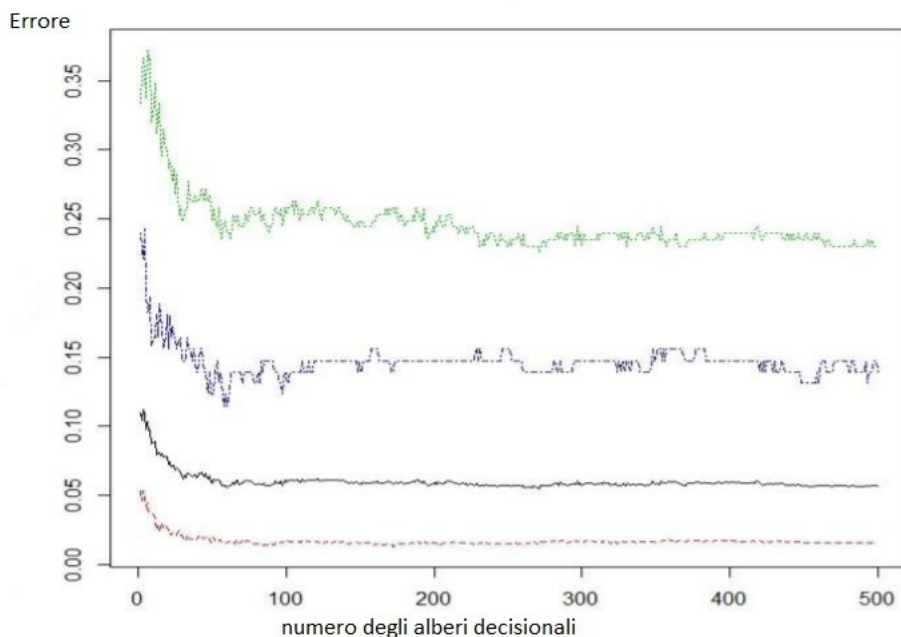


Figura 4.2 Grafico del modello Random Forest rf. La linea verde rappresenta i tracciati patologici, la linea blu rappresenta i tracciati sospetti e quella rossa rappresenta i tracciati normali. La linea nera è l'ottimizzazione dell'errore.<sup>23</sup>

A questo punto si può costruire il modello Random Forest con i nuovi parametri, ottenendo così un errore Out of Bag di 5.08%. La matrice di confusione di questo modello è riportata nella Tabella 4.3.

	1	2	3	Errore della classifica
1	1137	16	7	0.02
2	41	171	1	0.2
3	6	5	111	0.09

Tabella 4.3 La matrice di confusione del modello Random Forest<sup>23</sup>

Per poter analizzare le prestazioni di questo modello si avrà bisogno di definire cosa è l'accuratezza e la sensibilità. L'accuratezza e la sensibilità sono definiti in base a come è stato classificato il tracciato, si avranno tre casi:

1. True positive (TP) – Il tracciato è stato classificato come appartenete ad una specifica categoria e il tracciato appartiene effettivamente a quella categoria. In altre parole, questi sono i tracciati correttamente classificati.
2. False negative (FN) – Il tracciato non è stato classificato nella classe giusta. Ad esempio, se un tracciato è normale (per quanto riguarda la classificazione dei tracciati cardiocografici secondo FIGO) ma non viene classificato nella categoria normale.

3. False positive (FP) – Anche in questo caso si ha che il tracciato non è stato classificato in modo corretto, ma a differenza di False negative, il False positive viene preso in considerazione per il numero totale dei tracciati appartenenti ad una specifica categoria. Ad esempio, se più tracciati vengono classificati come patologici ma in verità sono tracciati normali, il False positive è il numero dei tracciati falsamente classificati come patologici. Quindi per quanto riguarda il numero dei tracciati normali si prenderà in considerazione, per questo esempio, il False negative che è il numero di tracciati che erroneamente non sono stati classificati come normali. Più precisamente, se si ha 100 tracciati dei quali 80 sono normali, 15 sospetti e 5 patologici, ma il classificatore che viene usato considera 79 tracciati normali, dei quali 3 sono in realtà sospetti e 1 è patologico, si avrà True positive = 75, False negative = 5 (5 tracciati non sono stati correttamente classificati come normali), False positive = 4 (3 tracciati sospetti e 1 patologico sono stati classificati come normali).

Accuratezza - è la misura migliore e più efficace per determinare le prestazioni complessive di un sistema perché l'accuratezza è il rapporto tra i risultati positivi delle previsioni e tutti i dati raccolti. L'accuratezza è una misura valida, ma solo quando i valori di FP e FN sono quasi uguali. Di conseguenza, per studiare le prestazioni del modello, è necessario valutare altre variabili di input. L'accuratezza è la misura delle previsioni effettivamente fatte dal modello. L'accuratezza è data dall'equazione (1).

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

La sensibilità è data dall'equazione (2).

$$Sensibilità = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

#### 4.1.3. Risultati

L'accuratezza del modello così creato è di 93.5%. La sensibilità della prima categoria è di 0.98, della seconda categoria è di 0.76 e della terza categoria è 0.91. L'accuratezza delle tre categorie è: 0.91, 0.86 e 0.95.

L'accuratezza dei training data set quando sono stati usati i parametri predefiniti per mTry e nTree è di 94,46%. Dopo la modifica dei parametri l'accuratezza è arrivata a 94,92% per i training dataset

e 93,5% per i test dataset. Il grafico riportato nella Figura 4.3 rappresenta la classificazione fatta dal modello Random Forest dopo la modifica dei parametri predefiniti.

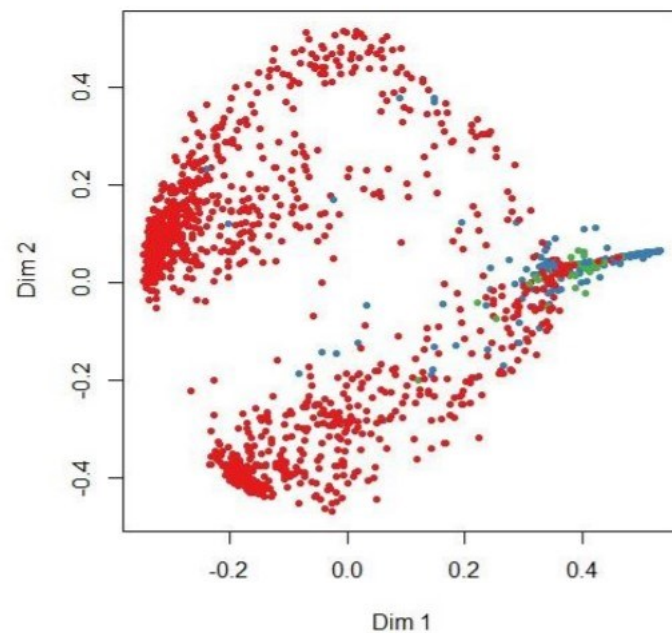


Figura 4.3 Grafico di dispersione dei dati in funzione della classificazione fatta. In Rosso vengono rappresentati i tracciati di Categoria 1 (Normali), in blu i tracciati di Categoria 2 (Sospetti) e in verde i tracciati di Categoria 3 (Patologici).<sup>23</sup>

#### 4.1.4. *Discussione e conclusione*

I grandi progressi che si possono osservare nell'area medica si basano principalmente sull'utilizzo dei computer per decidere la salute di un paziente. In particolare, l'utilizzo del Machine Learning può diminuire i problemi legati all'errore umano. In particolare, in questo articolo si è potuto dimostrare che quando si prendono in considerazione grandi quantità di dati, come nel caso dei tracciati cardiocografici l'approccio del Machine Learning ed in particolare del Random Forest è stato fondamentale per poter capire se un feto si trovava in condizioni di sofferenza fetale oppure no.

In questo studio sono stati usati gli algoritmi Random Forest per classificare i tracciati cardiocografici in 3 categorie: Normale, Sospetta e Patologica. L'eliminazione dei parametri che avevano una bassa importanza e la modifica dei parametri predefiniti della funzione Random Forest ha fatto in modo che l'accuratezza finale del processo fosse elevata.

La classificazione dei tracciati fatta in questo modo non è più soggetta all'errore umana. Come si è potuto osservare anche nella matrice di confusione nella Tabella 4.3, i tracciati giustamente

classificati come normali sono stati 1137, come sospetti sono stati 171 e 111 tracciati sono stati classificati in modo corretto come patologici. Pochi tracciati sono stati classificati in modo errato e in particolare la maggior parte delle classificazioni erroneamente fatte sono state tra i tracciati sospetti perché si è partito proprio con un modello di classificazione nel quale più tracciati sospetti erano considerati normali.

Per questo studio si è arrivato ad un ottimo risultato, mettendo in chiaro che l'approccio del Machine Learning nello studio del CTG è molto vantaggioso e offre al personale medico un aiuto nel decidere se intervenire per salvare il feto oppure il feto si trova in condizioni normali.

L'accuratezza del modello evidenzia che il classificatore Random Forest dovrebbe essere usato in tutti gli ospedali.

#### *4.2. Studio di M. Arif: Classificazione dei tracciati cardiotocografici con Random Forest e selezione dei parametri importanti di un tracciato cardiotocografico<sup>24</sup>*

M. Arif<sup>24</sup> nel suo lavoro ha usato Random Forest per classificare i tracciati cardiotocografici in tracciati: Normali, Patologici e Sospetti sulla base delle caratteristiche osservabili dai cardiotocogrammi. Questo processo ha portato l'accuratezza del modello di M. Arif a 93.6%. In più usando 7 parametri essenziali invece di 21, e quindi semplificando il carico dell'informazione che l'algoritmo Random Forest riceve, l'accuratezza rimane quasi invariata: da 93.6% si passa a 93.3%.

##### *4.2.1. Introduzione*

Un CTG viene usato per monitorare simultaneamente i battiti cardiaci fetali e le contrazioni uterine. Monitorando i battiti cardiaci fetale è possibile identificare se un feto si trova in condizioni di sofferenza, come l'ipossia oppure no. La possibilità che un feto nasce con disabilità permanenti o che muoia nei primi attimi di vita o prima ancora di nascere aumenta all'aumentare del periodo dell'ipossia fetale. Quindi è molto importante identificare le condizioni in cui si trova un feto. La cardiotocografia può essere utilizzata per identificare queste condizioni.

La classificazione dei tracciati cardiotocografici viene fatta manualmente e per questo i risultati possono essere soggetti ad errori umani. Per poter contrastare questo tipo di errore si può introdurre un classificatore automatico capace di classificare i tracciati cardiotocografici in tre classi: Normali, Sospetti e Patologici.

M. Arif nel suo lavoro usa il classificatore Random Forest che offre la possibilità di analizzare molti dati simultaneamente e decidere in quale classe appartiene un tracciato<sup>24</sup>.

L'autore vuole classificare il set di dati ricevuti dalla Clinica di Maternità e Ginecologia in una delle classi sopra elencate e spera di contrastare in questo modo la mortalità infantile causata dall'errore umano.<sup>24</sup>

L'indice di importanza dei parametri è usato proprio per identificare i parametri importanti del dataset che serviranno per classificare i tracciati. Si proverà a dimostrare che una buona accuratezza del classificatore preso in considerazione può essere raggiunta utilizzando solo 7 parametri importanti invece di 21 parametri.

#### 4.2.2. Materiali e metodi

Il set di dati iniziali è dato da 2126 cardiocogrammi ricevuti dalla Clinica di Maternità e Ginecologia (Ospedale Universitario di Porto in Portogalo) (Ayres-de-Campos, Bernades et al. 2000). I tracciati cardiocografici sono stati classificati da tre medici esperti e poi sono stati anche inseriti nel programma SisPorto dove sono stati definiti 21 parametri. I parametri possono essere letti nella Tabella 4.4.

Parametri	Spiegazione dei parametri	Parametri	Spiegazione dei parametri
<b>LB</b>	Linea di base (bpm)	<b>mLTV</b>	Media della variabilità aumentata (bpm)
<b>AC</b>	Accelerazioni(bpm)	<b>DL, DS, DP</b>	Decelerazioni: precoci, variabili e prolungate (bpm)
<b>FM</b>	Movimenti fetali (colpi/minuti)	<b>Width</b>	Ampiezza dell'istogramma
<b>UC</b>	Contrazione uterine (MVU)	<b>Min, Max</b>	La frequenza minima e massima dell'istogramma (H)
<b>ASTV</b>	Percentuale di tempo con ridotta variabilità (%)	<b>Nmax, Nzeros</b>	Numero dei picchi e degli zeri dell'istogramma
<b>mSTV</b>	Media della variabilità ridotta (bpm)	<b>Mode, Mean, Median, Variance</b>	Modulo, media, mediana e varianza dell'istogramma
<b>ALTV</b>	Percentuale di tempo con variabilità aumentata (%)	<b>Tendency</b>	Tendenza dell'istogramma: -1 = asimmetria di sinistra; 0 = simmetria; 1= asimmetria di destra

Tabella 4.4 Spiegazione dei parametri<sup>24</sup>

Per poter classificare i tracciati cardiocografici, M. Arif utilizza gli algoritmi Random Forest.<sup>24</sup>

La Tabella 4.5 indica la classifica dei tracciati cardiocografici fatta dal personale medico.

Per poter creare un algoritmo Random Forest si avrà bisogno della dimensione delle foglie; del numero di alberi decisionali e quanti parametri serviranno per creare un modello soddisfacente.

Classe (Condizione del feto)	Numero dei tracciati
<b>N(Normale)</b>	1655
<b>S(Sospetto)</b>	295
<b>P(Patologico)</b>	176
<b>Totale</b>	2126

Tabella 4.5 La classificazione dei tracciati cardiocografici secondo il personale medico<sup>24</sup>

Per quanto riguarda la dimensione delle foglie che si dovrà usare l'autore calcola l'errore Out Of Bag quando la dimensione delle foglie è: 2, 5, 10 e 15; ottenendo così un grafico (Figura 4.4) nel quale si terrà in conto anche del numero degli alberi decisionali. Come si può osservare (e come si è visto anche nei capitoli precedenti) l'errore Out of Bag decresce all'aumentare del numero di alberi decisionali e diventa quasi costante da quando si prendono in considerazione 50 alberi decisionali in su. Allo stesso tempo, in questo grafico, si evidenzia che la dimensione ideale per quanto riguarda la dimensione delle foglie è due, in quanto l'errore Out of Bag è decisamente più basso che negli altri casi. Quindi M. Arif decide di usare nel suo lavoro la dimensione delle foglie uguale a 2 e il numero degli alberi decisionali uguale a 50.<sup>24</sup>

Per poter analizzare le prestazioni di un classificatore si ha la necessità di definire: l'Accuratezza, la Precisione, il Recupero e F1-score. Per poter definire questi indici si deve prima definire 3 casi nei quali un tracciato può essere classificato come per il lavoro di S. Santhiya et al.

1. True positive (TP) – Il tracciato è stato classificato come appartenente ad una specifica categoria e il tracciato appartiene effettivamente a quella categoria. In altre parole, questi sono i tracciati correttamente classificati.
2. False negative (FN) – Il tracciato non è stato classificato nella classe giusta. Ad esempio, se un tracciato è normale (per quanto riguarda la classificazione dei tracciati cardiocografici secondo FIGO) ma non viene classificato nella categoria normale.
3. False positive (FP) – Anche in questo caso si ha che il tracciato non è stato classificato in modo corretto, ma a differenza di False negative, il False positive viene preso in

considerazione per il numero totale dei tracciati appartenenti ad una specifica categoria. Ad esempio, se più tracciati vengono classificati come patologici ma in verità sono tracciati normali, il False positive è il numero dei tracciati falsamente classificati come patologici. Quindi per quanto riguarda il numero dei tracciati normali si prenderà in considerazione, per questo esempio, il False negative che è il numero di tracciati che erroneamente non sono stati classificati come normali. Più precisamente, se si ha 100 tracciati dei quali 80 sono normali, 15 sospetti e 5 patologici, ma il classificatore che viene usato considera 79 tracciati normali, dei quali 3 sono in realtà sospetti e 1 è patologico, si avrà True positive = 75, False negative = 5 (5 tracciati non sono stati correttamente classificati come normali), False positive = 4 (3 tracciati sospetti e 1 patologico sono stati classificati come normali).

A questo punto si possono definire anche le 4 misure di prestazioni di un certo classificatore:

1. Accuratezza - questa è la misura migliore e più efficace per determinare le prestazioni complessive di un sistema perché l'accuratezza è il rapporto tra i risultati positivi delle previsioni e tutti i dati raccolti. L'accuratezza è una misura valida, ma solo quando i valori di FP e FN sono quasi uguali. Di conseguenza, per studiare le prestazioni del modello, è necessario valutare altre variabili di input. L'accuratezza è la misura delle previsioni effettivamente fatte dal modello. L'accuratezza è data dall'equazione (1).

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

2. Precisione – è la misura data dal rapporto dei tracciati correttamente classificati in una categoria e tutti i tracciati classificati in quella categoria. La precisione è data dall'equazione (3).

$$Precisione = \frac{TP}{TP + FP} \quad (3)$$

3. Recupero – è la misura data dal rapporto dei tracciati correttamente classificati in una categoria e il numero dei tracciati effettivamente appartenenti a quella categoria. Il recupero è dato dall'equazione (4).

$$Recupero = \frac{TP}{TP + FN} \quad (4)$$

4. F-score – è la combinazione della precisione e del recupero. È due volte la precisione per il recupero diviso per la precisione per il recupero. È più efficace dell'accuratezza poiché impone una penalità anche ai campioni classificati erroneamente. F-score è dato dall'equazione (5).



$$F - score = \frac{2 \times Precision \times Recupero}{Precision + Recupero} \quad (5)$$

Per generare il training dataset e il test dataset il set di dati iniziali è stato diviso in modo che 70% dei dati iniziali appartenessero al training dataset e 30% al test dataset. A questo punto si userà il training dataset per addestrare l'algoritmo Random Forest e alla fine i tracciati appartenenti al test dataset saranno classificati in: Normali, Sospetti e Patologici. Tutto questo processo è stato ripetuto per 10 volte. La media e la deviazione standard della Precisione, Recupero e F-score per il training dataset e per il test dataset sono riportati nella Tabella 4.6 e nella Tabella 4.7. Si può osservare che la Precisione, il Recupero e F-score hanno valori molto elevati per i training dataset. Per i test dataset si ha che la classificazione dei tracciati normali (78% di tutto il test dataset) ha una precisione e un recupero maggiore di 0.94 e F-score è uguale a 0.96. La classifica dei tracciati Sospetti ha valori molto più bassi se messi al confronto con le altre due classifiche. Questo perché come per il lavoro di S. Santhya et al.<sup>23</sup> la classificazione iniziale aveva dei tracciati Sospetti confusi come Normali.

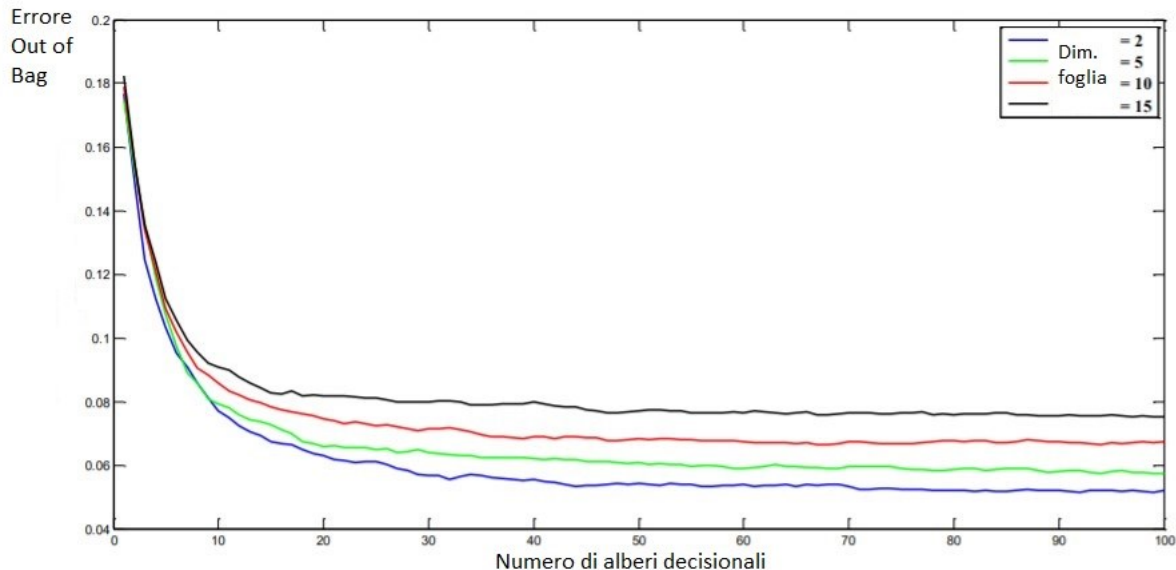


Figura 4.4 Calcolo dell'errore per diverse dimensioni delle foglie<sup>24</sup>

	Precisione	Recupero	F-score
N(Normale)	1	1	1
S(Sospetto)	0.97	0.99	0.98
P(Patologico)	0.98	1	0.99
Media ponderata	<b>0.99</b>	<b>0.99</b>	<b>0.99</b>

Tabella 4.6 Classificazione dei tracciati appartenenti al training dataset (i valori sono riportati come deviazione standard)<sup>24</sup>

	Precisione	Recupero	F-score
N(Normale)	0.98	0.95	0.96
S(Sospetto)	0.73	0.88	0.8
P(Patologico)	0.9	0.93	0.91
Media ponderata	<b>0.94</b>	<b>0.94</b>	<b>0.94</b>

Tabella 4.7 Classificazione dei tracciati appartenenti al test dataset (i valori sono riportati come deviazione standard)<sup>24</sup>

Ci sono 21 parametri per poter classificare i tracciati cardiocografici ma non tutti hanno lo stesso livello di importanza. In questo articolo, sono state studiate le possibili permutazioni per i parametri con un'elevata importanza. Infatti, i valori di un parametro sono permutati casualmente in un albero decisionale e l'incremento dell'errore quadratico medio che sia ha dalla permutazione viene considerato nel gruppo di dati Out of Bag. L'importanza di un parametro sarà data dalla media dell'errore quadratico medio diviso la deviazione standard dell'errore quadratico medio. In Figura 4.5 si può osservare il livello di importanza dei parametri appartenenti al gruppo di dati Out of Bag. Si ha che il parametro 6 (DS) e il parametro 21 (Tendency) sono gli parametri meno importanti, mentre il parametro 10 (ALT) ha il livello più alto di importanza.

Per capire quanti parametri si devono prendere in considerazione, si prenderà una soglia sotto la quale tutti i parametri dovranno essere eliminati in quanto non hanno un'importanza elevata. Per la soglia si terranno in considerazione i valori da 0.1 a 1.5 e si osserverà quanto l'accuratezza e F-score vengono alterati. Queste informazioni vengono riportate nella Tabella 4.8. Il grafico di F-score per le diverse soglie è riportato in Figura 4.6.

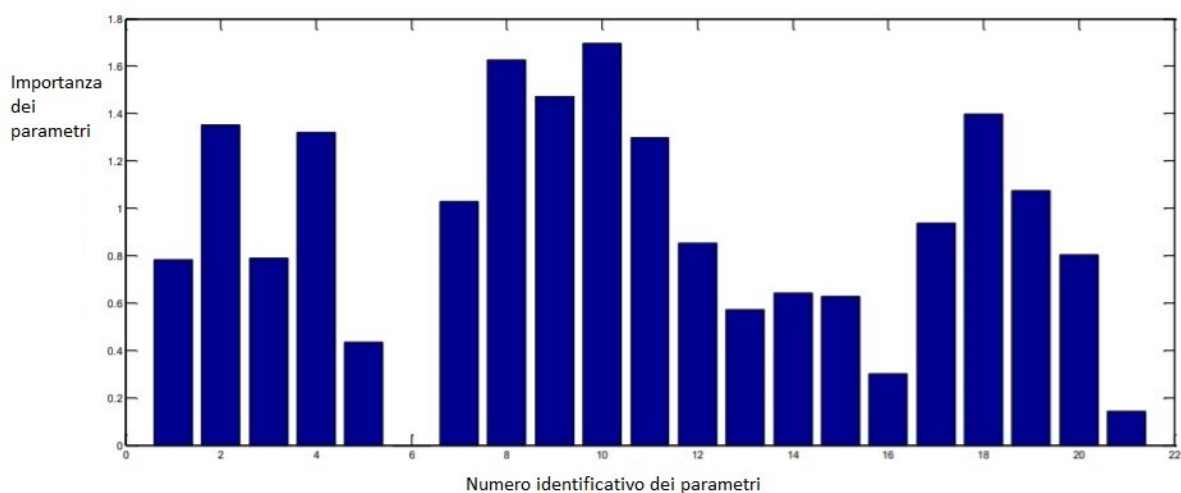


Figura 4.5 Livello di importanza di un parametro <sup>24</sup>

F-score per tutte tre le classifiche					
Soglia	Numero di parametri	Accuratezza	N(Normale)	S(Sospetto)	P(Patologico)
0.1	20	0.94	0.96	0.81	0.88
0.3	18	0.93	0.96	0.78	0.87
0.5	17	0.94	0.96	0.81	0.89
0.7	14	0.94	0.96	0.80	0.90
0.9	10	0.93	0.96	0.80	0.87
1.1	7	0.93	0.96	0.8	0.89
1.3	6	0.92	0.95	0.76	0.86
1.5	2	0.87	0.92	0.7	0.49

Tabella 4.8 Prestazione del modello con diverse soglie e numero di parametri <sup>24</sup>

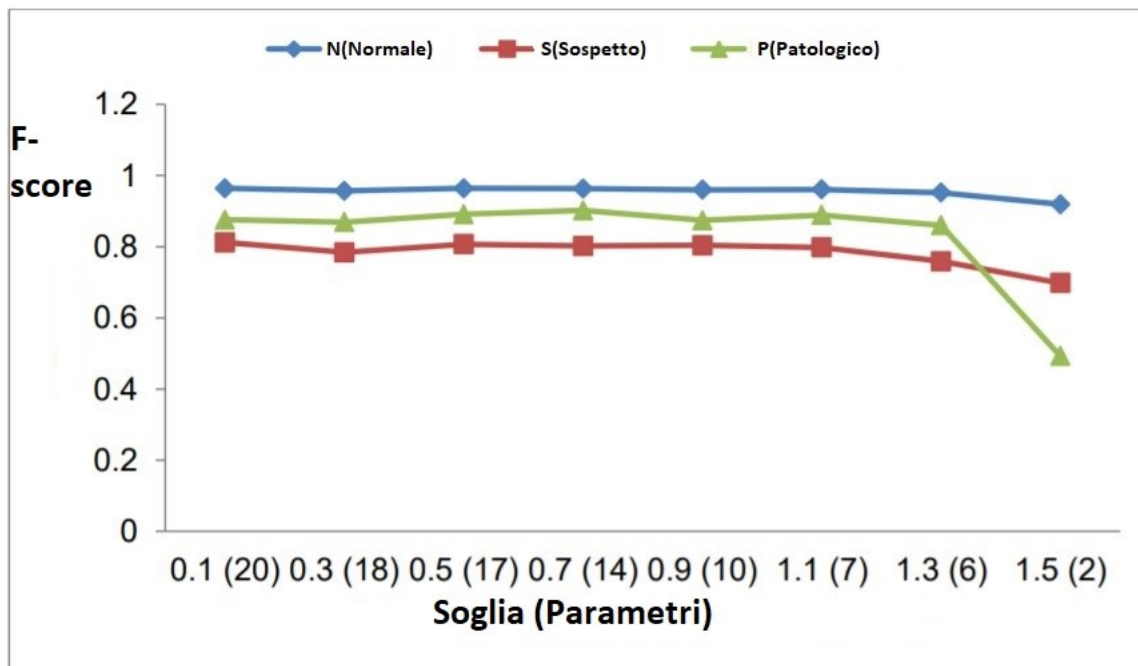


Figura 4.6 Misura di F-score per tutte le tre le categorie (Normale, Sospetto e Patologico) per diversi numeri di parametri<sup>24</sup>

Si può osservare dalla Figura 4.6 che F-score rimane quasi invariato fino alla soglia 1.1 quando vengono considerati solo 7 parametri. I 7 parametri sono quello che hanno un livello di importanza maggiore di 1.1 (Figura 4.5) e sono: AC, UC, ASTV, MSTV, ALTV, MLTV e Mean.

#### 4.2.3. Risultati

La riduzione dei parametri ha evidenziato quanto i parametri che sono rimasti sono importanti. Tra questi si hanno le accelerazioni e le contrazioni uterine, valori che si prendono sempre in considerazione quando si analizza un tracciato cardiocografico.

Il modello così creato ha un'accuratezza di 93.3% e F-score per la classe Normale è di 0.96, per la classe Sospetta è di 0.8 e per quella Patologica è di 0.89.

Le prestazioni del modello sono riportati nella Tabella 4.9.

M. Arif ha identificato solo 7 parametri importanti, ottenendo comunque un'accuratezza quasi uguale all'accuratezza del modello quando si ha a disposizione tutti i parametri.<sup>24</sup>

Tracciato	Precisione	Recupero	F-score
N(Normale)	0.98	0.94	0.96
S(Sospetto)	0.72	0.9	0.8
P(Patologico)	0.88	0.9	0.89
Media ponderata	<b>0.94</b>	<b>0.93</b>	<b>0.93</b>

Tabella 4.9 Prestazioni del modello con 7 parametri invece che 21 per il test dataset<sup>24</sup>

#### 4.2.4. *Discussione e conclusione*

Costruendo il modello M. Arif ha deciso di usare solo 7 parametri invece di 21, 2 dimensioni per le foglie e 50 alberi decisionali, ottimizzando in questo modo il modello.<sup>24</sup> Il set di dati iniziali è stato diviso in modo casuale: 70% per il training dataset e 30% per il test dataset. L'accuratezza finale del modello è di 93.3%. L'accuratezza passa da 93.6% nel caso in cui vengono usati 21 parametri a 93.3% nel caso in cui vengono usati solo 7 parametri importanti. Quindi semplificando in questo modo il modello si ha comunque un'accuratezza elevata e i tracciati vengono classificati in maniera soddisfacente.

In questo lavoro M. Arif ha analizzato le prestazioni di un classificatore Random Forest usato per classificare i tracciati cardiocografici.<sup>24</sup> In particolare, attraverso l'uso della Precisione, del Recupero e F-score è riuscito a dimostrare che i tracciati cardiocografici possono essere classificati con Random Forest, in quanto la classe normale non viene confusa con le altre classi. Infatti, la precisione della classe normale e della classe patologica è di 98% per la prima e di 90% per la seconda.

I risultati così ottenuti confermano che l'utilizzo di un simile classificatore può essere usato per contrastare la mortalità infantile.

I 7 parametri con un'importanza elevata sono le accelerazioni che sono degli aumenti di 15 bpm sopra la linea di base, le contrazioni uterine, la percentuale di tempo con variabilità aumentata e con variabilità ridotta, media della variabilità ridotta e della variabilità aumentata e la media dell'istogramma. Senza questi parametri si è visto che non si può avere una classificazione adeguata. Infatti, nei capitoli precedenti si è evidenziato che nelle classificazioni secondo ACOG o RCOG o NICE un diverso valore dal normale di questi parametri e tenendo conto ovviamente anche del tempo nel quale un parametro non si trova in condizioni standard, indicava che un feto poteva trovarsi in condizioni di sofferenza. La condizione di sofferenza fetale richiede l'intervento

del personale medico. Identificando in maniera veloce e accurata la classi di un tracciato può salvare la vita di un feto.

#### *4.3. Studio di M.M. Imran Molla et al.: Classificazione dei tracciati cardiocografici con Random Forest basato su algoritmi di Machine Learning<sup>25</sup>*

M.M. Imran Molla et al. hanno usato gli algoritmi Random Forest per poter classificare i tracciati cardiocografici in tracciati Normali, Sospetti e Patologici ottenendo un'accuratezza di 94.8%. In questo studio sono stati usati 10 parametri essenziali invece si 21 ottenendo un'accuratezza elevata.

La classificazione di un tracciato cardiocografico da un algoritmo di Machine Learning è molto utile in quanto come spiegato anche nei capitoli precedenti queste classificazioni vengono fatti dal personale medico e quindi è soggetto all'errore umano.

##### *4.3.1. Introduzione*

La cardiocografia viene usata per controllare lo stato della salute del feto. Come si sa dai capitoli precedenti il cardiocografo registra due segnali, quello proveniente dal battito cardiaco fetale e quello proveniente dalle contrazioni uterine. L'obbiettivo del monitoraggio cardiocografico è quello di identificare l'ipossia fetale nella fase iniziale così da intervenire e fare in modo che la gravidanza procedi normalmente e non si abbiano problemi permanenti. A questo scopo è molto importante classificare i tracciati patologici senza confonderli con quelli normali.

La classificazione dei tracciati cardiocografici viene fatta manualmente e quindi può essere soggetto a errori umani. Un classificatore automatico può diminuire questo errore. Gli autori di questo lavoro hanno voluto usare gli algoritmi Random Forest per fare questa classifica. L'utilizzo degli algoritmi Random Forest, basato a sua volta sull'utilizzo degli alberi decisionali può garantire risultati soddisfacenti. Come si sa, il risultato finale del Random Forest è il voto di maggioranza degli alberi decisionali.

Gli autori di questo lavoro intendono usare l'indice di prestazione per indentificare i parametri essenziali per la classificazione. Negli studi precedenti a questo come quello degli autori Yilmaz E e Kılıkçier Ç (2013) si è osservato che si poteva ottenere una buona accuratezza usando solo 10 parametri invece di 21.

### 4.3.2. Materiali e metodi

In Figura 4.7 si può osservare il processo della creazione del modello adatto per classificare i tracciati con Random Forest. La creazione di un modello richiede di avere a disposizione un set di dati iniziali. In questo studio è stato usato il set di dati di C. Sundar, M. Chitradevi, G. Geetharamani (2013)<sup>26</sup>. Questo set di dati è stato raccolto da UCI Machine Learning Repository. Prima di inserire i dati in un classificatore sono stati usati diversi procedimenti per verificare se ci sono dati mancanti o valori sbagliati. Il classificatore Random Forest ha bisogno di dati di allenamento e dati per testare quindi il set di dati iniziali è stato diviso tale che 80% dei dati appartengono al training dataset e 20% al test dataset. Il test dataset è stato testato sulla base del training dataset.

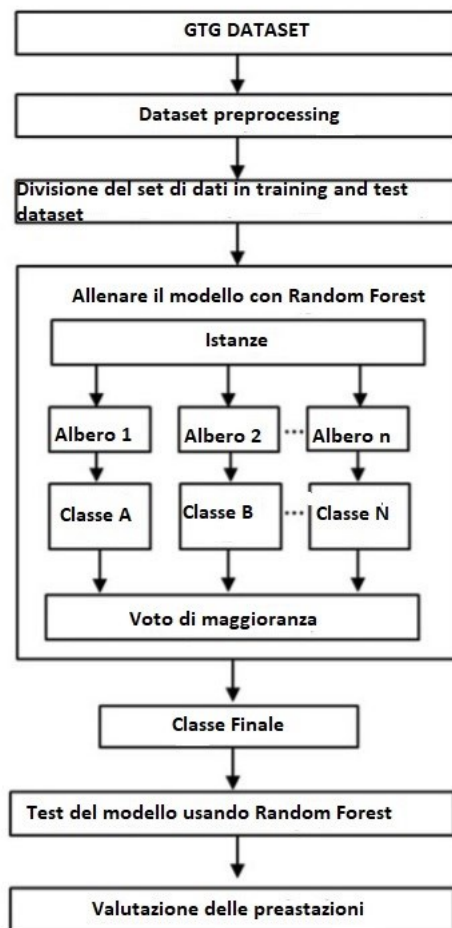


Figura 4.7 Principio di funzionamento di Random Forest per la regressione <sup>25</sup>

Il set di dati iniziale è formato da 2126 tracciati descrivibili grazie a 22 parametri. Questi tracciati sono stati classificati da 3 persone del personale medico in 3 tipi di tracciati: Normale (N), Sospetto (S) e Patologico (P) come in Tabella 4.10.

Condizioni del feto	Classe	Numero della classe	Numero di tracciati
<b>Normale</b>	N	1	1655
<b>Sospetta</b>	S	2	295
<b>Patologica</b>	P	3	176
<b>Totale</b>			2126

Tabella 4.10 Classificazione dei tracciati cardiocografici nelle diverse classi ognuna rappresentata da un numero<sup>25</sup>

Random Forest crea un insieme di alberi decisionali da un sottoinsieme del training dataset e prende la decisione finale dal voto di maggioranza degli alberi decisionali. Ogni albero viene creato come segue:

1. Se il numero di casi all'interno del training dataset è N, si ha N casi campionanti in modo aleatorio con riemissione. Questi campioni saranno il training dataset per la creazione dell'albero.
2. Se ci sono M variabili in ingresso, un numero m molto più piccolo di M indica che per ogni nodo sono scelte m variabili scelte in modo casuale da M. Lo split migliore di queste m variabili viene usato per dividere i nodi. Il valore di m viene mantenuto costante per tutta la crescita dell'albero.
3. Ogni albero cresce fino al suo massimo. Non c'è il purning.

Utilizzando l'errore Out of Bag si può trovare il valore migliore per m. Come si può osservare dalla Figura 4.8 l'errore Out of Bag diminuisce all'aumentare del numero di alberi decisionali e diventa quasi costante quando si ha a disposizione 300 alberi decisionali.

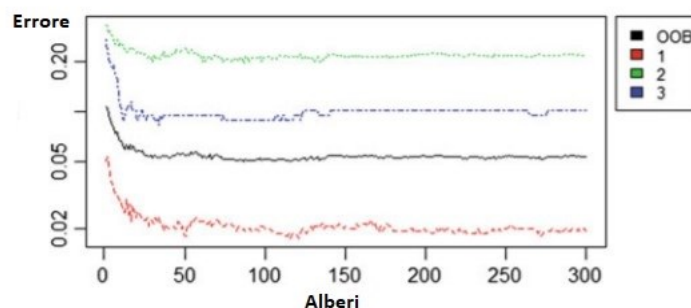


Figura 4.8 Errore Out of Bag per ogni classe in funzione del numero di alberi decisionali<sup>25</sup>



Come per gli altri lavori osservati l'autore definisce, l'accuratezza, la precisione, il recupero e F-score in base a:

1. True positive (TP) – Il tracciato è stato classificato come appartenente ad una specifica categoria e il tracciato appartiene effettivamente a quella categoria. In altre parole, questi sono i tracciati correttamente classificati.
2. False negative (FN) – Il tracciato non è stato classificato nella classe giusta. Ad esempio, se un tracciato è normale (per quanto riguarda la classificazione dei tracciati cardiocografici secondo FIGO) ma non viene classificato nella categoria normale.
3. False positive (FP) – Anche in questo caso si ha che il tracciato non è stato classificato in modo corretto, ma a differenza di False negative, il False positive viene preso in considerazione per il numero totale dei tracciati appartenenti ad una specifica categoria. Ad esempio, se più tracciati vengono classificati come patologici ma in verità sono tracciati normali, il False positive è il numero dei tracciati falsamente classificati come patologici. Quindi per quanto riguarda il numero dei tracciati normali si prenderà in considerazione, per questo esempio, il False negative che è il numero di tracciati che erroneamente non sono stati classificati come normali. Più precisamente, se si ha 100 tracciati dei quali 80 sono normali, 15 sospetti e 5 patologici, ma il classificatore che viene usato considera 79 tracciati normali, dei quali 3 sono in realtà sospetti e 1 è patologico, si avrà True positive = 75, False negative = 5 (5 tracciati non sono stati correttamente classificati come normali), False positive = 4 (3 tracciati sospetti e 1 patologico sono stati classificati come normali).

A questo punto si possono definire anche le 4 misure di prestazioni di un certo classificatore:

- Accuratezza - questa è la misura migliore e più efficace per determinare le prestazioni complessive di un sistema perché l'accuratezza è il rapporto tra i risultati positivi delle previsioni e tutti i dati raccolti. L'accuratezza è una misura valida, ma solo quando i valori di FP e FN sono quasi uguali. Di conseguenza, per studiare le prestazioni del modello, è necessario valutare altre variabili di input. L'accuratezza è la misura delle previsioni effettivamente fatte dal modello. L'accuratezza è data dall'equazione (1).

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

- Precisione – è la misura data dal rapporto dei tracciati correttamente classificati in una categoria e tutti i tracciati classificati in quella categoria. La precisione è data dall'equazione (3).

$$Precisione = \frac{TP}{TP + FP} \quad (3)$$

- Recupero – è la misura data dal rapporto dei tracciati correttamente classificati in una categoria e il numero dei tracciati effettivamente appartenenti a quella categoria. Il recupero è dato dall'equazione (4).

$$Recupero = \frac{TP}{TP + FN} \quad (4)$$

- F-score – è la combinazione della precisione e del recupero. È due volte la precisione per il recupero diviso per la precisione più il recupero. È più efficace dell'accuratezza poiché impone una penalità anche ai campioni classificati erroneamente. F-score è dato dall'equazione (5).

$$F - score = \frac{2 \times Precisione \times Recupero}{Precisione + Recupero} \quad (5)$$

#### 4.3.3. Risultati

Gli algoritmi Random Forest hanno ottenuto un buon risultato per il training dataset, avendo valori molto alti per la Precisione, Recupero e F-score. La media ponderata di questi valori sono riportati nella Tabella 4.11. La Precisione e il Recupero della classe Normale sono di 0.95 con un F-score di 0.95 per il test dataset. La classe Sospetta ha una percentuale molto più bassa se messa a confronto con la classe Normale e quella Patologica. Questo perché come è stato riportato anche per gli altri studi, la classe Sospetta viene inizialmente classificata in modo errato.

Nella Tabella 4.12 si è riportato una matrice di confusione per un test dataset. Molti tracciati appartenenti alla classe Normale sono stati effettivamente classificati come tali. Solo 4 tracciati Sospetti sono stati classificati come Normali e solo un tracciato Patologico è stato confuso come appartenente alla classe dei tracciati Normali.

Questa classificazione è stata fatta usando tutti i 21 parametri. Ma non tutti hanno la stessa importanza. Per questo è capire quale parametro ha un impatto maggiore nella prestazione finale del modello.

L'importanza di 10 parametri basata sulla Media della diminuzione dell'Accuratezza e la Media della diminuzione del coefficiente di Gini sono riportati in Figura 4.9. La media della diminuzione dell'Accuratezza si trova quando viene calcolato l'errore Out of Bag. Un parametro viene considerato molto importante se la sua eliminazione diminuisce l'accuratezza del modello Random Forest. Per questo motivo i parametri che hanno una media della diminuzione dell'Accuratezza molto alta vengono considerati molto importanti. La media della diminuzione del coefficiente di Gini indica la media della diminuzione dell'impurità di un nodo del parametro considerato, pesata per lunghezza che deve percorrere Random Forest in tutti gli alberi decisionali per arrivare a quello specifico nodo. Un parametro con un'elevata media della diminuzione del coefficiente di Gini viene considerato un parametro importante.

Classe	Precisione		Recupero		F-score	
	Train	Test	Train	Test	Train	Test
Normale	1	0.98	1	0.97	1	0.97
Sospetta	1	0.76	1	0.91	1	0.83
Patologica	1.00	0.95	1.00	0.86	1.00	0.90
Media ponderata	<b>1</b>	<b>0.95</b>	<b>1</b>	<b>0.95</b>	<b>1</b>	<b>0.95</b>

Tabella 4.11 Risultati della classificazione fatta per il training dataset e per il test dataset<sup>25</sup>

Classe	Normale	Sospetta	Patologica
Normale	146	4	1
Sospetta	2	19	0
Patologica	1	2	18

Tabella 4.12 Matrice di confusione per un set di dati del test dataset<sup>25</sup>

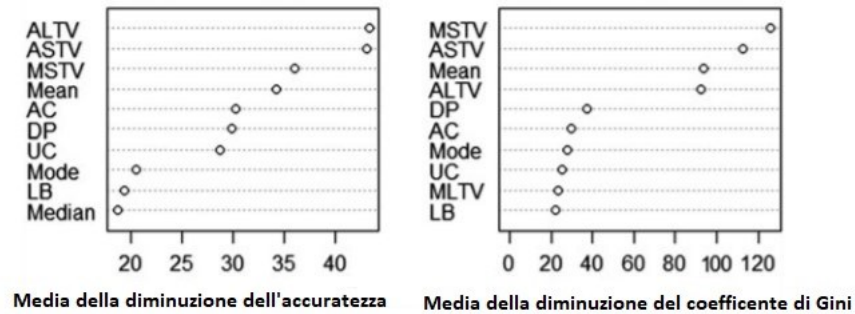


Figura 4.9 I 10 parametri più importanti in funzione della Media della diminuzione dell'Accuratezze e della Media della diminuzione del coefficiente di Gini<sup>25</sup>

#### 4.3.4. *Discussione e conclusione*

I 2126 tracciati raccolti dalla Clinica di Maternità e Ginecologia (Ospedale Universitario di Porto in Portogallo) sono stati classificati inizialmente dal personale medico. Per poter applicare il modello Random Forest il set di dati iniziali è stato diviso in 80-20 dove 80% del set di dati iniziale è stato usato per il training dataset e 20% per il test dataset. Questa suddivisione garantisce una prestazione maggiore del sistema. Infatti, usando più tracciati per allenare l'algoritmo, quest'ultimo riconosce molto più in fretta i tracciati del test dataset.

Per analizzare le prestazioni di questo modello sono stati definiti i seguenti parametri statistici: l'accuratezza, la prestazione, il recupero e F-score. Un valore alto di questi parametri garantisce un'adeguata distinzione dei tracciati normali da quello patologici. L'accuratezza finale del modello è di 94.8% quando vengono usati tutti i parametri.

Grazie all'utilizzo della media della diminuzione dell'accuratezza e della media della diminuzione del coefficiente di Gini sono stati identificati i parametri con una elevata importanza. La percentuale di tempo con una variabilità aumentata è il parametro con la maggior diminuzione dell'accuratezza. Quindi questo dovrebbe essere sempre il primo parametro che si guarda in un tracciato cardiocografico.

Un classificatore come Random Forest può essere di grande aiuto nell'identificare i tracciati correttamente. La decisione degli autori a dividere il set di dati in 80-20 e la loro analisi dell'importanza dei parametri, ha evidenziato come piccoli cambiamenti nel processo possono garantire risultati eccezionali.

Allenando il modello con più dati e prendendo inizialmente solo i parametri con maggior importanza si spera di poter classificare tutti i tracciati correttamente, dando un aiuto importante al personale medico nelle decisioni finale della salute del feto.

#### *4.4. Studio di P. Tomáš et al.: Classificazione dei tracciati cardiotocografici con Random Forest<sup>27</sup>*

##### *4.4.1. Introduzione*

La CTG è un metodo di diagnosi dello stato del benessere fetale. Un tracciato cardiotocografico può essere Normale, Sospetto e Patologico. Questi tre tipi di tracciati possono indicare una normale condizione acido-base del feto (caso Normale), una condizione acido-base sospetta (caso Sospetto) o una condizione acido-base pericolosa (caso Patologico). Lo scopo principale del lavoro di P. Tomáš et al. è di trovare un algoritmo di Machine Learning capace di identificare i tracciati in modo automatico senza l'intervento umano e quindi senza l'errore umano. A questo scopo viene usato il classificatore Random Forest, formato da più alberi decisionali. L'errore che si commette applicando il classificatore Random Forest dipende dalla correlazione di due qualsiasi alberi appartenenti a Random Forest. La correlazione di due alberi aumenta l'errore. Quindi si deve fare in modo che gli alberi sono molto diversi tra di loro per garantire una classificazione adeguata. Per costruire un modello Random Forest c'è bisogno di due parametri molto importanti: il numero di alberi decisionali e il numero delle variabili usate per ogni split. Questi parametri dovrebbero essere ottimizzati per minimizzare l'errore complessivo.

Random Forest funziona in modo efficiente su grandi quantità di dati e stima quali variabili sono importanti nella classificazione.

In questo lavoro gli autori vogliono identificare il numero adeguato di alberi decisionali che devono essere usati per creare un buon classificatore e il numero di variabili che si devono considerare per ogni split.<sup>27</sup> Se si individuano i parametri migliori, l'errore che si commette per ogni tracciato che si classifica diminuisce. Il classificatore migliore è quello con un errore che può essere approssimato a zero.

#### 4.4.2. Materiali e metodi

Il set di dati iniziale è composto da 2126 tracciati, classificato inizialmente da tre ostetrici. Tutti i tracciati sono stati analizzati dal software SiSPorto 2.0 che ha identificato i seguenti parametri: la linea di base (BL), le accelerazioni (ACC), le decelerazioni (DCL), la variabilità, i movimenti fetali (FM) e le contrazioni uterine (UC). Oltre a questi parametri sono stati identificati altri 14 parametri statistici (Tabella 4.13). Tutto il set di dati contiene: 1655 tracciati Normali, 295 tracciati Sospetti e 176 tracciati Patologici.

L'estrazione dei parametri e data preprocessing sono i due passi iniziali per riconoscere il tipo di tracciato. Ci sono due modi per l'estrazione dei parametri adatti: riduzione dei parametri e selezione dei parametri. Il motivo principale di questo approccio è quello di diminuire il carico di dati in ingresso. I metodi per selezionare i parametri fanno in modo che solo i parametri che hanno un'importanza elevata vengono selezionati. La soluzione degli autori è quello di usare la correlazione dei parametri selezionati (CFS). CFS è basato sulla riduzione della dimensione dei parametri. Infatti, l'algoritmo riduce più del 60% delle caratteristiche avute inizialmente. Il CFS calcola la correlazione tra parametri e classe dei tracciati, tra parametri e parametri, ed elimina i parametri che non sono adatti alla classificazione. Applicando questo algoritmo gli autori hanno ottenuto 7 parametri essenziali: AC, DS, DP, ASTV, MSTV, ALTV e Mean.

<i>Contrazioni uterine</i>	<b>UC</b>	<b>Numero di contrazioni uterine per secondo</b>
<i>Linea di base</i>	LB (bpm)	Linea di base dei battiti cardiaci fetale
	FM	Numero dei movimenti fetali per secondo
<i>Variabilità della linea di base</i>	ASTV (%)	Percentuale di tempo con variabilità ridotta
	MSTV (bpm)	Valore medio della variabilità ridotta
	ALTV (%)	Percentuale di tempo con variabilità aumentata
	MLTV (bpm)	Valore medio della variabilità aumentata
<i>Accelerazioni</i>	AC	Numero delle accelerazioni per secondo
<i>Decelerazioni</i>	DL	Numero delle decelerazioni precoci per secondo
	DS	Numero delle decelerazioni severe per secondo
	DP	Numero delle decelerazioni prolungate per secondo
<i>Caratteristiche dell'istogramma</i>	WIDTH (mV)	Ampiezza dell'istogramma
	MIN	Minimo dell'istogramma
	MAX	Massimo dell'istogramma
	NMAX	Numero dei picchi nell'istogramma
	NZEROS	Numero degli zeri nell'istogramma
	MODE	Modulo dell'istogramma
	MEAN	Media dell'istogramma
	MEDIAN	Mediana dell'istogramma
	VARIANCE	Varianza dell'istogramma
	TENDENCY	Tendenza dell'istogramma

Tabella 4.13 Parametri di un tracciato cardiocotografo<sup>27</sup>

A questo punto si possono applicare gli algoritmi Random Forest per classificare i tracciati. Come si sa dai capitoli precedenti, gli algoritmi Random Forest lavorano molto bene per i set di dati molto grandi e grazie all'utilizzo degli alberi decisionali la classifica finale avrà una prestazione elevata.

Le prestazioni del classificatore Random Forest sono date in funzione dell'accuratezza.

- Accuratezza - questa è la misura migliore e più efficace per determinare le prestazioni complessive di un sistema perché l'accuratezza è il rapporto tra i risultati positivi delle previsioni e tutti i dati raccolti. L'accuratezza è una misura valida, ma solo quando i valori di FP e FN sono quasi uguali. Di conseguenza, per studiare le prestazioni del modello, è necessario valutare altre variabili di input. L'accuratezza è la misura delle previsioni effettivamente fatte dal modello. L'accuratezza è data dall'equazione (1).

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

### 4.4.3. Risultati

Tutto il set di dati è stato diviso in training dataset e test dataset. Sono state considerate diverse misure per il training dataset: 10% del set di dati iniziali, 20%, 30%, 40% e 50%. I risultati migliori sono stati ottenuti quando il training dataset aveva a disposizione 50% del set di dati iniziali. L'accuratezza in questo caso è stata di 97.5% per la classe Normale, 75.3% per la classe Sospetta e 91.4% per la classe Patologica (Figura 4.10). La matrice di confusione per questi risultati può essere osservata nella Tabella 4.14. In questa tabella si ha che 792 tracciati normali sono stati identificati come tali, 10 tracciati Sospetti sono stati considerati come Normali e 10 tracciati patologici sono stati considerati erroneamente come Normali. In questo caso 10 feti non hanno avuto le cure adatte essendo considerati della classe sbagliata.

L'accuratezza dell'algoritmo nel caso in cui sono stati usati 600 alberi decisionali è di 93,7%.

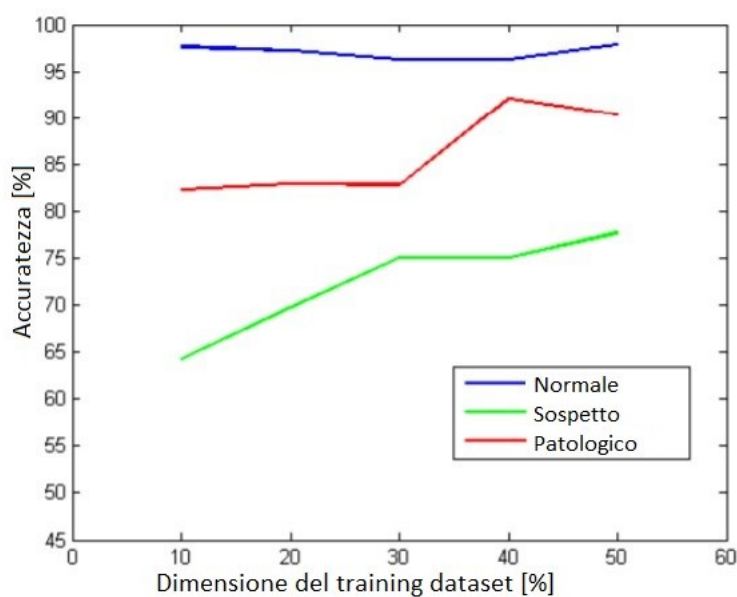


Figura 4.10 L'accuratezza dell'algoritmo Random Forest per ogni classe in funzione della dimensione del training dataset<sup>27</sup>

	Normale	Sospetto	Patologico
Normale	792	10	10
Sospetto	5	119	34
Patologico	3	5	85

Tabella 4.14 Matrice di confusione<sup>27</sup>



#### *4.4.4. Discussione e conclusione*

In questo studio i tracciati cardiocografici sono stati classificati grazie agli algoritmi Random Forest.<sup>27</sup> Prima di tutto il set di dati iniziali è stato analizzato con gli algoritmi CFS per poter identificare i parametri. L'uso di solo 7 parametri invece di 21 ha semplificato esponenzialmente la classificazione dei tracciati.

In questo lavoro sono state prese diverse quantità di dati per il training dataset: 10%, 20%, 30%, 40% e 50%. Ma se si addestrano pochi dati il modello non saprà come classificare i dati del test dataset. Ecco perché il risultato migliore è stato proprio quello in cui il training dataset era formato da 50% dei dati iniziali.

Sono stati usati 600 alberi decisionali, un numero così grande serve per poter classificare in modo più efficace i tracciati. Ma avendo molti alberi il processo viene rallentato e il risultato della classificazione può durare troppo tempo. Tempo che un medico potrebbe non avere a disposizione. Per diminuire il lavoro che ogni albero decisionale fa sono stati usati solo 7 parametri. Infatti, l'utilizzo di meno parametri di cui tutti hanno un'importanza elevata, garantisce un'accuratezza quasi uguale all'accuratezza del modello quando si prendono tutti i parametri.

Gli autori di questo lavoro sono stati tra i primi ad usare gli algoritmi Random Forest per classificare un tracciato cardiocografico. Infatti, a differenza degli altri articoli presi in considerazione in questa tesi, in questo articolo il set di dati iniziali è stato suddiviso più volte per identificare la giusta divisione. Il numero di alberi decisionali presi in considerazione è molto più alto che nel caso in cui si prendono i parametri predefiniti. Ma offre comunque un punto di partenza per identificare i parametri necessari per creare un buon classificatore.

Nel loro lavoro P. Tomáš et al. hanno considerato anche altri due modelli: LDA e REPTree, ma il Random Forest si è dimostrato il classificatore migliore e quindi si dovrebbe inserire questo classificatore in tutti gli ospedali del mondo per poter diminuire la mortalità infantile.<sup>27</sup>

#### *4.5. Altri metodi di classificazione dei tracciati cardiocardiografici*

La classificazione dei tracciati cardiocardiografici può essere fatta con gli algoritmi Random Forest come anche con altri metodi. Tuttavia, il metodo Random Forest per la classificazione dei tracciati ha una prestazione molto maggiore se messa a confronto con i risultati di questi metodi. Nella Tabella 4.15 si possono osservare i diversi metodi di classificazione dei tracciati cardiocardiografici messi al confronto con la classificazione dei tracciati cardiocardiografici con Random Forest studiati nelle pagine precedenti. Come si può osservare l'accuratezza degli algoritmi Random Forest è compresa tra 93.3% (M. Arif (2015))<sup>24</sup> e 94.8% (Imran Molla et al. (2021))<sup>25</sup>. Invece per gli altri studi l'Accuratezza oscilla tra 76.72% (Chamidah (2015)) quando viene usato il metodo SVM e 95.01% (Karabulut e Ibrikci (2014)) quando si usano gli alberi decisionali.

Negli studi di H. Sahin e A. Subasi (2015)<sup>28</sup>, l'accuratezza della classificazione fatta con Random Forest è di 99.2% (per due classi di classificazione: Normale e Patologica) a differenza degli altri tipi di classificatori come, ad esempio, CART che ha ottenuto un'accuratezza di 98.91%.

Riferimenti	Metodo	Risultati
Sundar et al. (2013)	Classificazione basata sulla Rete Neurale Artificiale	Media ponderata della Precisione, Recupero e F-score: Precisione (0.91), Recupero (0.90), e F-score (0.90)
Karabulut e Ibriki (2014)	Alberi decisionali (C4.5)	Accuratezza finale (95.01) (10 volte convalida incrociata)
Menai, Mohder et al. (2013)	Relief F-15	Accuratezza (0.94), Recupero (0.92), Specificità (0.96)
Jezewski, Nski et al.	Classificatore LSVM	Sensibilità: 83%, Specificità: 92% (su tutti i parametri)
Chen et al. (2012)	FG-Kmeans	Precisione (0.76), Recupero (0.81), F-score (0.77)
Zhou e Sun (2014)	Apprendimento attivo su processi Gaussiani	Accuratezza totale: 89% (training dataset formato da solo 140 esempi)
Chamidah (2015)	1. SVM 2. K-Means + SVM	Accuratezza di 1: 76.72% 2: 90.64%
Zhang e Zhou (2017)	PCA e AdaBoost	Accuratezza totale: 93%
Cruz et al. (2014)	Classificatore d'insieme META-DES	Accuratezza totale: 84.6%
Tomáš et al. (2013)	Random Forest (Tutti i parametri)	Accuratezza totale: 93.7%
Imran Molla et al. (2021)	Random Forest (Tutti i parametri)	Precisione, Recupero e F-score sono 0.95 Accuratezza totale: 94.8%
M. Arif (2015)	Random Forest (Tutti i parametri)	Per il training dataset: La Precisione, il Recupero e F-score sono di 0.99 Per il test dataset: La Precisione, il Recupero e F-score sono di 0.94 Accuratezza totale: 93.6%
M. Arif (2015)	Random Forest (Solo sette parametri)	Per il testing dataset: Precisione (0.94), Recupero (0.93) e F-score (0.93) Accuratezza totale: 93.3%

Tabella 4.15 Prestazioni di diversi metodi di Machine Learning usati per la classificazione dei tracciati cardiociclografici

*4.5.1. Studio di H. Sahin e A. Subasi: Classificazione dei tracciati cardiocografici utilizzando metodi di Machine Learning per anticipare i rischi fetali<sup>28</sup>*

*4.5.1.1. Introduzione*

Lo scopo del lavoro di H. Sahin e A. Subasi<sup>28</sup> è quello di analizzare 8 metodi diversi di Machine Learning per classificare i tracciati cardiocografici in due classi: Normale e Patologica. Classificando i tracciati in questo modo H. Sahin e A. Subasi sperano di facilitare la possibilità di riconoscere se un feto si trova in condizione di stress oppure no e di intervenire tempestivamente nel caso di bisogno.

I 8 metodi di Machine Learning usati in questo studio sono: regressione logica (o modello logit), il modello K-NN, il modello basato su una rete di funzione di base radiale (RBFN), ANNs, SVM, CART (Alberi di classificazione e regressione), C4.5 e Random Forest.

*4.5.1.2. Materiali e metodi*

Il set di dati iniziali è composto da 1831 tracciati cardiocografici con 21 parametri. Dopo essere stato analizzato dal software SISPORTO 2.0 il set di dati iniziale viene inserito nei diversi Machine Learning. Alla fine, tutti i tracciati vengono inseriti nell'algoritmo WEKA che darà l'accuratezza, la specificità, la sensibilità e F-score per ogni modello usato.

- Regressione logica - La regressione logistica è un modello statistico (modello logit) usato negli algoritmi di classificazione del machine learning per ottenere la probabilità di appartenenza a una determinata classe. L'algoritmo di classificazione basato sulla regressione logistica è un apprendimento supervisionato. Si basa sull'utilizzo della funzione logistica (sigmoid) che converte i valori reali in un valore compreso tra 0 e 1. Nonostante il nome dell'algoritmo faccia pensare a un algoritmo di regressione, si tratta di un algoritmo di classificazione. In questo lavoro classifica i tracciati in Normali e Patologici.
- K-NN - Dato un numero intero positivo K e un data test x, il classificatore K-Nearest Neighbors (KNN) dapprima identifica i K punti nei training dataset più prossimi a x; poi stima la probabilità che test data appartenga a ciascuna delle classi j della variabile di risposta come la frazione di punti in K il cui valore di risposta è uguale a j. Infine, classifica il data test x alla classe con il valore più alto di probabilità stimata.

- RBFN (Figura 4.11) - è una rete neurale artificiale che usa la funzione di base radiale come funzione d'attivazione. Essa è una combinazione lineare delle funzioni di base radiale.
- ANNs (Reti neurali artificiali) – sono delle strutture ispirate al cervello umano, imitando il modo in cui i neuroni biologici si inviano segnali. Sono composte da livelli di nodi che contengono un livello di input, uno o più livelli nascosti e un livello di output. Ciascun nodo, o neurone artificiale, si connette ad un altro e ha un peso e una soglia associati. Se l'output di qualsiasi singolo nodo è al di sopra del valore di soglia specificato, tale nodo viene attivato, inviando i dati al successivo livello della rete. In caso contrario, non viene passato alcun dato al livello successivo della rete.
- SVM (macchine a vettori di supporto) - sono dei modelli di apprendimento supervisionato associati ad algoritmi di apprendimento per la regressione e la classificazione. Dato un insieme di esempi per l'addestramento, ognuno dei quali etichettato con la classe di appartenenza fra le due possibili classi, un algoritmo di addestramento per le SVM costruisce un modello che assegna i nuovi esempi a una delle due classi, ottenendo quindi un classificatore lineare binario non probabilistico.
- CART (Alberi di classificazione e regressione) - Piuttosto che alberi generici che potrebbero avere più rami, CART utilizza l'albero binario, che ha solo due rami per ogni nodo. CART utilizza Gini Impurity come criterio per dividere il nodo.
- C4.5 – può gestire attributi sia continui sia discreti, può accettare training dataset con valori mancanti, può attribuire una pesatura differenziata a ogni singolo attributo, implementa un criterio di pruning per evitare il rischio di overfitting
- Random Forest - è un metodo di machine learning, capace di affrontare sia compiti di classificazione che di regressione. Il suo principio di funzionamento è quello di usare più alberi decisionali per avere la decisione finale.

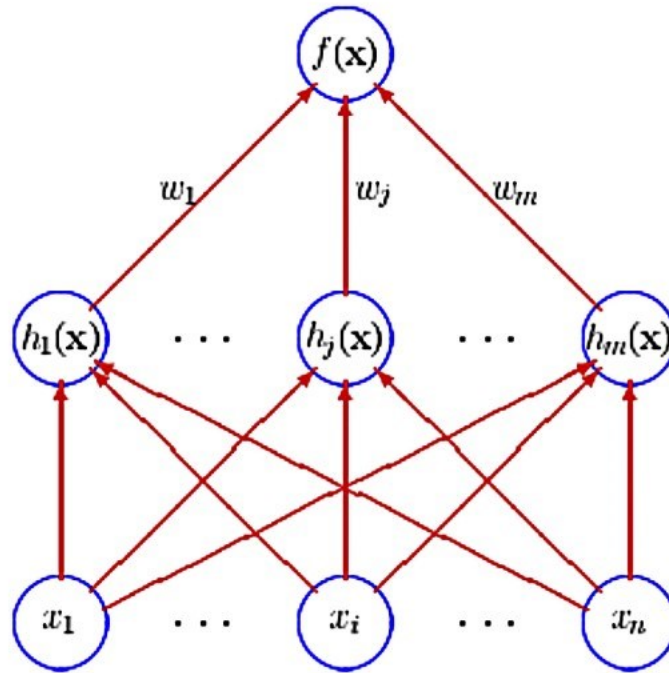


Figura 4.11 Un tipico esempio del modello RBFN. Ognuno degli  $n$  elementi del vettore di ingresso  $x$  invia a  $m$  funzioni base le cui output sono combinazione lineare del peso e dell'uscita di rete. <sup>28</sup>

#### 4.5.1.3. Risultati

L'algoritmo WEKA permette di trovare e cambiare i parametri. Per prima caso i parametri predefiniti sono testati e grazie ai risultati della matrice di confusione si può ottenere l'accuratezza del modello. Identificando in questo modo il modello con la miglior prestazione.

La sensibilità e la specificità sono due misure di controllo del test dataset.

$$\text{Sensibilità} = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

$$\text{Specificità} = \frac{TN}{TN + FP} \times 100\% \quad (6)$$

Nella Tabella 4.16 si possono osservare le prestazioni degli otto metodi di classificazione. Tutti i metodi hanno ottenuto un buon risultato, avendo l'errore della Specificità che varia da 0.26% e 0.67%. Si può notare che Random Forest ha una prestazione molto maggiore degli altri, ottenendo un'Accuratezza di 99.18, una Sensibilità di 99.10 e una Specificità di 99.74.

	ANN	SVM	SL	RBF	C4.5	CART	RF	k-NN
Accuratezza	98.63	98.96	98.74	98.53	99.13	98.91	<b>99.18</b>	98.42
Specificità	99.53	99.61	99.47	99.66	99.68	99.42	<b>99.74</b>	99.33
Sensibilità	90.69	93.59	92.11	88.86	94.10	94.09	<b>94.10</b>	90.69

Tabella 4.16 Confronto tra i diversi metodi di classificazioni (in percentuale)<sup>28</sup>

#### 4.5.1.4. *Discussione e conclusione*

In questo studio è stato usato il set di dati del UCI che ha in verità 2126 tracciati, ma 295 appartenevano alla classe Sospetta che non è stata presa in considerazione. Dato che l'obiettivo principale della classificazione è quello di aiutare il personale medico a decidere se devono intervenire oppure no, la classe Sospetta non influenza questo aspetto.

Il confronto di questo studio con i altri della letteratura non può essere fatta in quanto il set di dati iniziali è diverso. Ma si può comunque evidenziare che il metodo migliore per la classificazione tra gli otto classificatori presi in considerazione in questo lavoro è Random Forest con un'Accuratezza del 99.18%.

#### 4.5.2. *Studio di Z. Hoobhoy et al.: Uso del Machine Learning per la predizione della salute fetale usando tracciati cardiocografici*<sup>29</sup>

##### 4.5.2.1. *Introduzione*

In questo lavoro si è voluto studiare le prestazioni dei modelli di Machine Learning per classificare i tracciati cardiocografici. 10 diversi algoritmi sono stati addestrati usando il set di dati provenienti da un cardiocografo. Lo scopo principale è quello di riuscire a trovare il miglior classificatore per poter ridurre la mortalità neonatale.

##### 4.5.2.2. *Materiali e metodi*

Il set di dati iniziale è stato ottenuto dal deposito di Machine Learning Irvine dell'Università della California. È composto da 2126 tracciati cardiocografici di donne che si trovavano nel terzo trimestre di gravidanza. Il set di dati ha 21 parametri. I tracciati sono stati classificati inizialmente da 3 ostetrici dopo di che sono stati inseriti nel software SisPorto 2.0, un programma per la valutazione automatica dei tracciati CTG.

Gli algoritmi usati in questo lavoro sono: MLP, SVM linear kernel, SVM RBF kernel, K-NN, classificatore XGBoost, classificatore AdaBoost, Random Forest, regressione logica, Gaussian Naïve Bayes e alberi decisionali.

#### 4.5.2.3. Risultati

Dei 2126 tracciati 70% sono di classe Normale, 20% di classe Sospetta e 10% appartengono alla classe Patologica.

Nelle Tabelle 4.17 e 4.18 sono riportati i risultati dei diversi algoritmi. Si può osservare che XGBoost, alberi decisionali e gli algoritmi Random Forest hanno una precisione maggiore del 96% e una sensibilità maggiore del 99% per il training dataset.

L'Accuratezza totale dell'algoritmo XGBoost è di 93% maggiore se messa al confronto con gli altri metodi analizzati. Nel test dataset XGBoost è riuscito a classificare i tracciati patologici al 92% a differenza del Random Forest che è riuscito a classificarli solo al 87%.

Algoritmo	Precisione			Recupero			F-score		
	N	S	P	N	S	P	N	S	P
MLP	0.88	0.87	0.94	0.87	0.84	0.96	0.87	0.85	0.95
XGBoost	0.99	0.96	1	0.97	0.99	0.99	0.98	0.98	0.99
Alberi decisionali	1	1	1	1	1	1	1	1	1
Random Forest	0.99	0.99	1	0.99	0.99	1	0.99	0.99	1
Regressione logica	0.87	0.77	0.88	0.84	0.79	0.88	0.86	0.79	0.88
SVM linear kernel	0.9	0.8	0.89	0.85	0.83	0.91	0.87	0.81	0.9
SVM RBF kernel	0.98	0.92	0.99	0.92	0.97	0.99	0.95	0.94	0.98
KNN	1	0.95	0.99	0.95	0.99	1	0.97	0.97	0.99
Naïve Bayes	0.88	0.66	0.86	0.76	0.88	0.68	0.82	0.75	0.76
AdaBoost	0.86	0.88	0.988	0.89	0.88	0.95	0.87	0.88	0.97

Tabella 4.17 Confronto tra gli algoritmi di Machine Learning per il training dataset<sup>29</sup>



Algoritmo	Precisione			Recupero			F-score		
	N	S	P	N	S	P	N	S	P
MLP	0.96	0.52	0.7	0.85	0.72	0.89	0.9	0.6	0.77
XGBoost	0.98	0.73	0.92	0.94	0.88	0.92	0.96	0.8	0.92
Alberi decisionali	0.96	0.74	0.87	0.95	0.74	0.92	0.95	0.74	0.89
Random Forest	0.96	0.73	0.86	0.95	0.78	0.88	0.95	0.75	0.87
Regressione logica	0.96	0.48	0.64	0.84	0.75	0.84	0.9	0.58	0.72
SVM linear kernel	0.97	0.49	0.68	0.84	0.79	0.88	0.9	0.6	0.76
SVM RBF kernel	0.98	0.62	0.84	0.91	0.82	0.88	0.94	0.7	0.86
KNN	0.96	0.6	0.82	0.9	0.76	0.87	0.93	0.66	0.84
Naïve Bayes	0.97	0.42	0.46	0.76	0.85	0.67	0.85	0.56	0.54
AdaBoost	0.96	0.58	0.88	0.89	0.81	0.87	0.92	0.67	0.87

Tabella 4.18 Confronto tra gli algoritmi di Machine Learning per il test dataset<sup>29</sup>

#### 4.5.2.4. *Discussione e conclusione*

Questo lavoro è molto importante in quanto mette a confronto 10 algoritmi diversi per classificare i tracciati e propone l'utilizzo dell'algoritmo che ha ottenuto una buona accuratezza sia per i tracciati normali che quelli patologici.

#### 4.6. *Come e perché si dovrebbero usare gli algoritmi Random Forest per classificare i tracciati cardiocografici*

La CTG è un esame essenziale per valutare il benessere di un feto; quindi, i tracciati che si ottengono dovrebbero essere classificati con un algoritmo Random Forest per garantire una classificazione oggettiva.

Partendo dall'idea che un tracciato può essere valutato da algoritmi diversi e quindi l'introduzione di un classificatore automatico è ovviamente molto fattibile, proviamo ad analizzare perché proprio Random Forest dovrebbe essere il classificatore inserito in tutti gli ospedali.

Se ad esempio prendiamo in considerazione il lavoro fatto da H. Sahin e A. Subasi (2015)<sup>28</sup>, dove la classificazione di un tracciato viene fatta in due classi: Normale e Patologica, possiamo osservare che Random Forest identifica i tracciati con un errore di < 1%. Questo garantisce che per un algoritmo Random Forest i tracciati normali sono ben distinguibili dai tracciati patologici. Però, un approccio nel quale vengono considerate solo due classi non potrebbe essere usato in tutti gli ospedali. Si, può comunicare al personale medico se intervenire nel caso patologico, ma se non si

prendono in considerazione anche i tracciati sospetti, o peggio quest'ultimi vengono inseriti in una delle due classi, si potrebbe intervenire in situazioni nelle quale l'unico approccio sarebbe solo quello di fare altri test per capire se la situazione sospetta può peggiorare e quindi rientrare nel caso patologico oppure può migliorare e la situazione è stato un caso dato da altri fattori probabilmente provenienti dalla madre. Considerare solo due classi per classificare i tracciati non è neanche in sintonia con le linee guida per classificare i tracciati cardiocografici secondo ACOG, RCOG, NICE, FIGO.

Sappiamo a questo punto grazie al lavoro di H. Sahin e A. Subasi<sup>28</sup>, che il Random Forest è il classificatore migliore se messo al confronto con la regressione logica (o modello logit), il modello K-NN, il modello basato su una rete di funzione di base radiale (RBFN), ANNs, SVM, CART (Alberi di classificazione e regressione) e C4.5 per classificare i tracciati in due classi.

Ma se la classificazione avviene come spiegato nella linee guida per classificare i tracciati, Random Forest rimane comunque il classificatore migliore se messo al confronto con altri metodi di deep learning? Il lavoro fatto di Z. Hoobhoy et al.<sup>29</sup> dice di no, dichiara che il classificatore ideale per essere usato negli ambienti ospedalieri è il XGBoost. Il XGBoost viene seguito subito da Random Forest e dagli alberi decisionali nella lista dei miglior classificatori. Però se guardiamo la tabella 4.17 si vede che nel caso del training dataset Random Forest ha ottenuto una precisione maggiore di XGBoost. Ovviamente anche gli alberi decisionali hanno risultati ottimali per il training dataset, ma sappiamo che il Random Forest è creato da alberi decisionali; quindi, ottimi risultati per quest'ultimi garantisce ottimi risultati anche per Random Forest. Non sappiamo con quale percentuale i dati iniziali sono stati suddivisi tra il training dataset e il test dataset. Non avendo a disposizione più informazioni del processo della creazione del modello non possiamo garantire che XGBoost è migliore di Random Forest. Sappiamo solo che è riuscito a sorpassare la sensibilità di Random Forest nell'identificare un tracciato cardiocografico. Quindi diciamo che Random Forest è uno degli miglior classificatori che sono stati analizzati fino adesso, quindi perché non usarlo in tutti gli ospedali?

Abbiamo deciso che Random Forest è un classificatore molto buono, capace di classificare i tracciati in tre classi diversi, con una sensibilità soddisfacente.

Vediamo come si devono costruire i modelli basati sulla classificazione con Random Forest. Partendo dal lavoro di P. Tomáš et al.<sup>27</sup> si può capire quanto è importante dividere il set di dati

iniziali in training dataset e test dataset. Si può osservare in Figura 4.12 che l'accuratezza di Random Forest è molto più grande degli altri due metodi presi in considerazione.

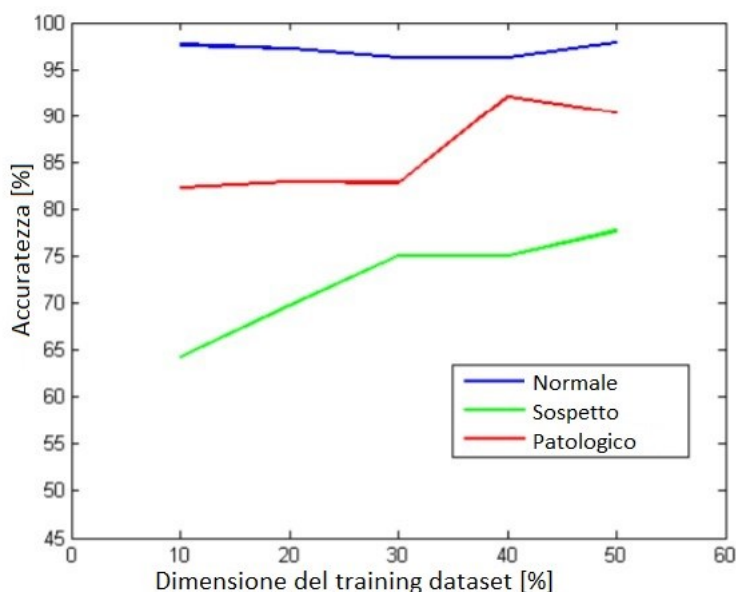


Figura 4.12 Accuratezza [%] di tre diversi classificatori in funzione della dimensione del training dataset<sup>27</sup>

Per creare un modello Random Forest per classificare i tracciati, c'è bisogno di dividere il set di dati iniziali in modo che più del 50% del set di dati iniziali viene inserito nel training dataset. Infatti, nei lavori di M. Arif e M.M. Imran Molla et al., il set di dati è stato diviso in 70-30 per il lavoro di M. Arif<sup>24</sup> e 80-20 per il lavoro di M.M. Imran Molla et al.<sup>25</sup> Per questo motivo, ma non solo, l'accuratezza totale del modello di M.M. Imran Molla et al.<sup>25</sup> è maggiore di quella di M. Arif<sup>24</sup>, come abbiamo potuto notare nella Tabella 4.15.

Si è visto che P. Tomáš et al.<sup>27</sup> hanno usato 600 alberi decisionali. Anche se un numero alto di alberi garantisce una visione molto più estesa del tracciato, può anche rallentare tutto il processo. Un numero perfetto per gli alberi decisionali è 300 come si è visto nei lavori di M.M. Imran Molla et al.<sup>25</sup> e S. Santhiya et al.<sup>23</sup>

Sappiamo che in un tracciato cardiociclografico si devono guardare diversi parametri, come la linea di base, le accelerazioni, le decelerazioni, percentuale di tempo delle variabilità, le variabilità e altri. Ma l'eliminazione di alcuni di questi parametri può garantire una prestazione del modello uguale a quella nella quale si prendono in considerazione tutti i parametri. Negli articoli precedenti sono stati usati diversi processi per identificare quali parametri sono indispensabili. M.M. Imran Molla et al.<sup>25</sup> hanno usato la media della diminuzione dell'accuratezza e la media della diminuzione

del coefficiente di Gini per identificare 10 parametri essenziali. Invece S. Santhiya et al.<sup>23</sup> hanno costruito un modello Random Forest con parametri predefiniti per mTry e mTree per poter estrarre un grafico nel quale si hanno i parametri essenziali. M. Arif<sup>24</sup> considera solo 7 parametri come essenziali. Quindi l'accuratezza del modello di M.M. Imran Molla et al.<sup>25</sup> è maggiore di quella di M. Arif<sup>24</sup> anche perché quest'ultimo ha usato solo 7 parametri.

Quindi per costruire un modello Random Forest capace di classificare un tracciato cardiocografico in una delle tre classi della linea guida FIGO: Normale, Sospetta, Patologica; si deve dividere il set di dati in modo che molti più dati vengono inseriti nel training dataset, si deve avere a disposizione un numero di alberi decisionali al incirca uguale a 300 e se possibile si dovrebbero utilizzare quasi sempre tutti i parametri anche se l'accuratezza non è influenzata di molto.

Il modello così creato avrà la capacità di classificare i tracciati cardiocografici e quindi si potrà diminuire la rata di mortalità infantile e prenatale.

## *II. Conclusione*

Il cardiocardiografo è un metodo economico per monitorare la salute del feto. Un tracciato cardiocardiografico può essere di tre tipi: Normale, Sospetto e Patologico. Uno problema che si può avere quando vengono osservati i tracciati cardiocardiografici è che il giudizio finale della salute del feto è soggetta all'errore umano. Per contrastare questo rischio si è introdotto la possibilità di usare un classificatore automatico. In particolare, si è osservato che l'uso del Machine Learning ed in particolare di Random Forest garantisce una classifica oggettiva dei tracciati.

In particolare, si è preso in considerazione un modello Random Forest composto da 300 alberi decisionali e da 21 parametri. Un modello così creato ha ottenuti risultati sorprendenti e ha evidenziato che questo tipo di classificatore dovrebbe essere inserito in tutti gli ospedali. Infatti, guardando le prestazioni in termini di accuratezza si è visto che quest'ultima era compresa tra 90-99%. Una prestazione così grande garantisce che tutti i tracciati saranno classificati in maniera soddisfacente.

In un primo momento questo modello potrebbe essere inserito solo come una macchina di supporto per la decisione finale del personale medico. Se si riuscirà ad allenare il modello con molti più tracciati e quindi avere un'accuratezza del training dataset molto maggiore dei risultati già buoni che ci sono nella letteratura i classificatori Random Forest potranno essere usati anche negli ospedali remoti dove non ci sono molti medici specialisti per aiutare a identificare i tracciati.

Con questo approccio si spera che la mortalità perinatale possa diminuire e si possa garantire una vita sana del bambino dopo la nascita.

### III. Bibliografia e Sitografia

---

<sup>1</sup> <https://www.treccani.it/vocabolario/gravidanza/>

<sup>2</sup> L. Solnica-Krezel, Conserved Patterns of Cell Movements during Vertebrate Gastrulation (2015), *Current Biology*, Vol. 15: 2013-228

<sup>3</sup> G. Ambrosi, D. Cantino, P. Castano, S. Correr, L. D'Este, R.F. Donato, G. Familiari, F. Fornai, M. Gulisano, A. Iannello, L. Magaudda, M.F. Marcello, A.M. Martelli, P. Pacini, M. Rende, P. Rossi, C. Sforza, C. Tacchetti, R. Toni, G. Zummo, *Anatomia dell'uomo*, Edi.Ermes s.r.l, Milano, 2006, Parte quinta Embriologia e organogenesi, Fig. 25.37

<sup>4</sup> N. Jayne Klossner, *Introductory Maternity Nursing*, Library of Congress Cataloging-in-Publication Data, 2006

<sup>5</sup> G. Ambrosi, D. Cantino, P. Castano, S. Correr, L. D'Este, R.F. Donato, G. Familiari, F. Fornai, M. Gulisano, A. Iannello, L. Magaudda, M.F. Marcello, A.M. Martelli, P. Pacini, M. Rende, P. Rossi, C. Sforza, C. Tacchetti, R. Toni, G. Zummo, *Anatomia dell'uomo*, Edi.Ermes s.r.l, Milano, 2006

<sup>6</sup> G. Ambrosi, D. Cantino, P. Castano, S. Correr, L. D'Este, R.F. Donato, G. Familiari, F. Fornai, M. Gulisano, A. Iannello, L. Magaudda, M.F. Marcello, A.M. Martelli, P. Pacini, M. Rende, P. Rossi, C. Sforza, C. Tacchetti, R. Toni, G. Zummo, *Anatomia dell'uomo*, Edi.Ermes s.r.l, Milano, 2006, Parte quinta Embriologia e organogenesi, Fig. 25.25

<sup>7</sup> <https://www.msdmanuals.com/it-it/professionale/ginecologia-e-ostetricia/travaglio-e-parto-normali/gestione-del-travaglio-fisiologico>

<sup>8</sup> Yli BM, et al. Pathophysiology of foetal oxygenation and cell damage during labour. *Best Pract Res Clin Obstet Gynaecol*. 2016

<sup>9</sup> World Health Organization, *Classificazione statistica internazionale delle malattie e dei problemi sanitari correlati – Decima revisione, quinta edizione*, 2016

<sup>10</sup> D. Gibb, S. Arulkumaran, *Fetal Monitoring in Practice*, Elsevier, 2017

<sup>11</sup> <https://www.gimaitaly.com/DocumentiGIMA/Manuali/IT/M29520IT.pdf>

<sup>12</sup> A.L. Borrelli, D. Arduini, A. Cardone, V. Ventruto, *Medicina dell'età prenatale*, Springer-Verlag, Italia, 2008

<sup>13</sup> G. Straface, A. Ercoli, A. Selmin, G.o Bosco, G. Scambia, Monitoraggio fetale intrapartum, *La Rivista italiana di Ostetrica e Ginecologia* N.31 (2011), 349:354

<sup>14</sup> D. Skupski, C. Rosenberg, G. Eglinton, Intrapartum fetal stimulation tests: a meta-analysis, *Obstetrics and gynecology*(2002)

<sup>15</sup> The 2015 FIGO classification of intrapartum cardiotocography: differences to the STAN classification

<sup>16</sup> SIGO, AOGOI, AGUI, RACCOMANDAZIONI MONITORAGGIO CARDIOTOCOGRAFICO IN TRAVAGLIO (2018), CONSENSUS FIGO

- 
- <sup>17</sup> SIGO, AOGOI, AGUI, Raccomandazioni monitoraggio cardiocografico in travaglio (2018), CONSENSUS FIGO
- <sup>18</sup>Royal College of Obstetricians and Gynaecologists, CTG interpretation and further management
- <sup>19</sup> NICE, Intrapartum care for healthy women and babies (2014)
- <sup>20</sup> T. Mitchell, Machine Learning, McGraw Hill, 1997
- <sup>21</sup> S. J. Russel e P. Norvig, Artificial Intelligence: A Modern Approach.
- <sup>22</sup> L. Breiman, Random Forest, Kluwer Academic Publishers. Manufactured in The Netherlands (2001), Machine Learning: 45, 5–3
- <sup>23</sup> S. Santhiya, G. Kowreesh, G.V. Chitralka, P. Boobalaragavan, Fetal distress classification using cardiocography (2020), JETIR: 39-45
- <sup>24</sup> M. Arif, Classification of cardiocograms using random forest classifier and selection of important features from cardiocogram signal (2015), Biomaterials and Biomedical Engineering, Vol. 2, No. 3: 173-181
- <sup>25</sup> M.M. Imran Molla, J.J. Jui, B.S. Bari, M. Rashid, M.J. Hasan, Cardiocogram Data Classification Using Random Forest Based Machine Learning Algorithm (2021), Proceedings of the 11th National Technical Seminar on Unmanned System Technology 2019: 357-367
- <sup>26</sup> C. Sundar, M. Chitradevi, G. Geetharamani (2013) An overview of research challenges for classification of cardiocogram data. J Comput Sci 9(2):198–206
- <sup>27</sup> P. Tomáš, J. Krohová, P. Dohnálek, P. Gajdoš, Classification of cardiocography records by random forest (2013), IEEE: 620-623
- <sup>28</sup> H. Sahin e A. Subasi, Classification of the cardiocogram data for anticipation of fetal risks using machine learning techniques (2015), Applied Soft Computing volume 33: 231-238
- <sup>29</sup> Z. Hoobhoy, M. Noman, A. Shafique, A. Nasim, D. Chowdhury e B. Hasan, Use of machine learning algorithms for prediction of fetal risk using cardiocographic data (2019), International Journal of Applied and Basic Medical Research: 226-229

---

## *VI. Ringraziamenti*

Ringrazio la professoressa Laura Burattini e la Dottoressa Agnese Sbröllini per avermi dato la possibilità di imparare di più sui tracciati cardiocografici, sull'Intelligenza Artificiale e sull'algoritmo Random Forest.

Ringrazio i miei genitori, mio fratello e Alina per avermi sostenuta e motivata sempre, per avermi abbracciata e incoraggiata anche nei momenti più difficili di questo percorso.

Ringrazio le mie più care amiche: Cristina, Sara e Elisabetta che mi hanno riportata sempre sulla strada giusta anche quando sembrava impossibile.

Grazie a tutte queste persone così importanti della mia vita sono riuscita ad arrivare ad un nuovo traguardo.