

Exploring the utility of artificial intelligence of intrapartum cardiotocography: a systematic review

Mohamed Salih (✉ mohamed.salih@monash.edu)

Monash University

Ritesh Rikain Warty

Monash University

Hamsaveni Kalina Murday

Monash University

Arjun Kaushik

VitalTrace Pty Ltd

Yuxin Zhang

Monash University

Fabrizio Horta

Monash University

Sandeep Reddy

Deakin University

Beverley Vollenhoven

Monash University

Hamid Rezafofighi

Monash University

Wenlong Cheng

Monash University

Vinayak Smith

Monash University

Research Article

Keywords: Cardiotocography, Artificial Intelligence, Machine Learning, Obstetrics, Interpretation, Decision Support

Posted Date: October 9th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3405992/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background:

Cardiotocography (CTG) interpretation is complex and highly subjective. Misinterpretation can result unnecessary, late, or inadequate intervention; potentially harming the mother or fetus. Artificial intelligence (AI) could play a role in reducing the likelihood of these incidents.

Purpose:

To identify the current state-of-the-art in AI models for CTG interpretation and provide clinicians and AI developers alike with an overview of this landscape and guide the development of future models.

Methods:

We searched PubMed, EMBASE, Ovid Medline, and IEEE Xplore for studies published from 01/06/2005 to 07/06/2020. Studies focused on AI applications to CTG were included, with the performance metrics (accuracy, sensitivity, and specificity) being extracted for analysis. PROSPERO registration: CRD42021250394.

Results:

38 articles were eligible for inclusion, though all studies were focused on pre-clinical performance evaluation. The types of AI techniques employed included support vector machines (11/38), decision trees (10/38), random forests (8/38), neural networks (23/38), and novel custom algorithms (11/38). Each model demonstrated promise in a pre-clinical setting although true clinical value is presently uncertain. Other issues included the classification systems used by AI developers, as well as the limited scope of these classification systems when compared to the more comprehensive clinical AI systems already implemented clinically in similar medical applications.

Conclusion:

AI shows promise as an adjunct surveillance tool in CTG interpretation. However, currently, it is too early to conclusively determine its implementation value in a clinical setting. To do so, these AIs need to be developed for and validated in high quality prospective clinical evaluations.

1. Introduction

Obstetric practice is widely known to be highly litigious across various high-income countries [1, 2]. In the 2017/18 year, the UK National Health Service (NHS) reported a net indemnity payment of over £4.5 billion, with obstetric claims accounting for 48% of that value despite representing only 10% of all claims [3]. It has been estimated that around 75% of all obstetricians have encountered litigation at least once, with common causes for lawsuits being due to medical errors or negligence in diagnosis, counselling, and

treatment [4]. A major concern for obstetricians and midwives is with regard to fetal monitoring. This is because misinterpretation of fetal heart rate (FHR) data or the failure to recognise fetal decompensation can lead to devastating consequences such as birth asphyxia. However, prompt recognition of these changes can expedite delivery of the fetus thereby averting these consequences [4, 5].

Cardiotocography (CTG) is one of the most utilised tools in obstetrics to assess fetal well-being with over 90% of women utilising the device during pregnancy. The fundamental principle of CTG relies upon the continuous monitoring of FHR in an attempt to correlate patterns in FHR changes with evolving hypoxia in the fetus. Through accurate identification, the intention is to expedite delivery in an effort to prevent hypoxic injury (birth asphyxia) in the newborn [6]. Although developed in the 1960s with the best of intentions, the evidence suggests that CTG is yet to achieve its lofty aims. Moreover, what remains concerning is that it also suggests that CTG has in fact increased the intervention rates in labour with no changes in neonatal outcome [7]. There are several schools of thought surrounding contributory causes to the current situation. For one, CTG interpretation is still done with the human eye by individuals (Fig. 1). This introduces several issues with its assessment such as difficulties in pattern recognition, issues interpreting the CTG in the clinical context, poor interobserver agreement on the classification (normal, suspicious, pathological), technical issues, or failure by the clinical team to perform additional assessments or consistently manage abnormal traces [8]. Additionally, there is an assumption that all fetuses display the same FHR changes when challenged with hypoxia, and there is an inability to perceive and compute long term changes due to its usage as a point of care test in the delivery rooms. Further complicating this issue, in 50% of fetuses with a non-reassuring CTG, there is no evidence of acidosis in the neonate [9].

Artificial intelligence (AI) is a growing field that has demonstrated significant promise in healthcare. With encouraging results in radiology, pathology, and many aspects of clinical medicine [11], engineering researchers and AI specialists have started developing AI technologies that are targeting obstetrics and gynaecology, with specific areas of interest in CTG monitoring, prognosis determination and prediction of 5-year survival likelihood for gynaecological cancers, and medically-assisted reproduction eg. IVF [5, 12]. For CTG interpretation especially, these include clinical decision support and improved signal processing, the intent being to facilitate objectivity, reproducibility, and consistency in CTG interpretation [13, 14].

2. Objectives

This systematic review aims to describe the techniques that have been discussed in the literature to inform AI developers and clinicians alike of the present AI methods employed and the efficacy of AI in CTG monitoring. In doing so, we hope to raise awareness of the current landscape in CTG applications of AI, provide an overview of existing models that have been developed, and discuss possible directions for future development.

3. Material and Methods

This systematic review was conducted in accordance with the PRISMA guidelines [15]. The study protocol and review were registered with PROSPERO (CRD42021250394) [16].

3.1. Search strategy and selection criteria

For the following review, a systematic search was undertaken independently by MS and RRW across PubMed, EMBASE, Ovid Medline, and IEEE Xplore from the 1st of June 2005 up to and including the 7th of June 2020. Included articles were restricted to those written in the English language.

Search terms utilised across all databases for the study were: (artificial intelligence OR machine learning OR deep learning OR neural networks) AND (fetal heart rate OR fetal monitoring OR cardiotocography). A detailed analysis of the search strategy may be found in Appendix A.

The inclusion criteria of studies for the review were articles that focused on applications of AI in CTG tasks (e.g., pattern recognition, state prediction, and condition interpretation). Examples of AI techniques that were analysed include: Support Vector Machines, decision trees, random forest, neural networks (NNs). Articles from reference lists of studies being screened were also considered suitable for assessment.

Citations were independently screened by MS, RW, and AK for suitability of inclusion and a set of eligible articles was created in accord. Following shortlisting, the full-text articles were read thoroughly by MS, RW, AK, HKM and VS. Inclusion of an article for the review was based on consensus between these authors. Studies that did not relate AI techniques with CTG, focused on doppler ultrasound, or focused on digital signal or image processing were excluded.

3.2. Data analysis

Data for the following study was extracted manually for analysis by MS and RW. Due to the heterogeneity of the included studies, variations in AI techniques, and parameters employed; pooling of data for meta-analysis was not possible and not considered appropriate. As such, a narrative approach was followed for this review.

3.3. Quality assessment

Quality assessment was independently performed by MS and RW using the Modified Down and Black checklist (Appendix B). HKM acted as a third assessor in the event a consensus could not be reached. This is a tool which has been used widely to rate the quality of observational studies and is rated on a 10 point scale [17]. The scores were calculated as the consensus between MS and RW and the questionnaire is included in the Appendix for reference.

3.4. Summary of measures and synthesis of results

The summary measures for the following study were presented in tabular form. Importance was given to the AI techniques implemented in the studies as well as their mode of classification for clinical

interpretation, and performance (accuracy, sensitivity, and specificity). The equations to calculate these performance metrics are:

$$Accuracy = \frac{TruePositive + TrueNegative}{TruePositive + FalsePositive + TrueNegative + FalseNegative}$$

1

.....

$$Sensitivity = \frac{TruePositive}{TruePositive + FalseNegative}$$

2

.....

$$Specificity = \frac{TrueNegative}{FalsePositive + TrueNegative}$$

3

.....

3.5. Statistical methods

Though the data was considered too heterogenous for a formal meta-analysis, for each machine learning technique, the means and standard deviations (SD) for each performance metric was calculated to determine if there is superiority between techniques. All other variables were expressed as a percentage.

Statistical analysis was completed using Excel version of Office 365.

4. Results

4.1. Overview

A total of 38 studies were included in this review. All 38 studies were conducted and published from 2005 to 2020. The article selection process is outlined in Fig. 2.

The characteristics of the models included in this review are reported in Table 1. There was considerable variety in the machine learning techniques employed and some variation in the classification of the CTG data with 18/38 (47.4%) of studies employing a binary classification system (normal vs abnormal) and the remaining 20/38 (52.6%) employing a 3-classification model (normal, suspected, and pathological). Of interest is that all the included studies were of a pre-clinical nature where 10/38 (26.3%) recorded clinical CTGs to create their own databases and 28/38 (73.7%) used publicly available databases. Of these publicly available databases, 18/28 (64.3%) were sourced from the UCI Machine Learning Repository, and 10/28 (35.7%) were sourced from the CTU-UHB database [18, 19].

Table 1
Characteristics of AI models for each included study.

Author, Year	Acquisition	AI used	CTG Classes
Georgoulas et al., 2006 [20]	The recordings were collected in the context of the Research Project POSI/CPS/40 153/2001	SVM	Normal, At-risk
Valensise et al., 2006 [21]	Recorded at University of Rome and University of Modena and Reggio Emilia between 1999 and 2003.	Various NN	Normal, Abnormal
Jezewski et al., 2007 [22]	Recorded on-site	ANN	Abnormal, Normal
Jadhav et al., 2011 [23]	UCI Machine Learning Repository	MNN	Normal, Suspect, Pathological
Dash et al., 2012 [24]	FHR records from 9 subjects at the Stony Brook University Medical Center.	BN from K2 structure learning	Normal, Indeterminate, Abnormal
Hongbiao et al., 2012 [25]	UCI Machine Learning Repository	combining PSO with BP	Normal, Atypical, Abnormal
Yilmaz et al., 2013 [26]	UCI Machine Learning Repository	LS-SVM, PSO, BDT	Normal, Suspect, Pathological
Ocak et al., 2013 [27]	UCI Machine Learning Repository	SVM, GA	Normal, Pathological
Haweel et al., 2013 [28]	UCI Machine Learning Repository	Volterra NN	Normal, Suspect, Pathological
Xu et al., 2013 [14]	7,568 deliveries in John Radcliffe hospital between 20 Apr 93 - 28 Feb 08.	SVM, GA	Normal, Adverse
Stylios et al., 2014 [29]	Collected in the University Hospital of Porto in Portugal	kNN, MLP	Normal, Abnormal
Shah et al., 2015 [30]	UCI Machine Learning Repository	DT	Normal, Suspicious, Pathological

NR: Not reported; ML: Machine Learning; SVM: Support vector machine; DT: Decision tree; RF Random forest; GA: Genetic Algorithm; NN: Neural network; MNN: Modular NN; CNN: Convoluted NN; MCNN: Multimodal CNN; BN: Bayesian network; ELM: Extreme learning machine; RNN: Recurrent NN; ANN: Artificial NN; MLP: Multi-layer perceptron; RBF: radial basis function; kNN: k-nearest neighbour; NB: Naive Bayes; GDI: Gini's diversity index; PCA: Principal component analysis; STI: Short-term irregularity; LTI: Long-term irregularity; STV: Short-term variability; LTV: Long-term variability; II: Interval index; WT: Wavelet transform, LS-SVM: Least Square-SVM, PSO: Particle Swarm Optimization, BDT: Binary Decision Tree, BP: Back Propagation

Author, Year	Acquisition	AI used	CTG Classes
Comert et al., 2016 [31]	UCI Machine Learning Repository	ANN	Normal, Suspicious, Abnormal
Warrick et al., 2016 [32]	Dataset of 72 CTG recordings	hidden semi-Markov model and Long-short term memory RNN	Noise, non-Acceleration, Acceleration
Batra et al., 2017 [33]	UCI Machine Learning Repository	SVM, DT, RF	Normal, Suspicious, Pathological
Permanasari et al., 2017 [34]	UCI Machine Learning Repository	DT	Normal, Suspicious, Pathological
Nagendra et al., 2017 [35]	UCI Machine Learning Repository	SVM, RF	Normal, Suspect, Pathological
Fergus et al., 2017 [36]	CTU-UHB	DNN	Normal, Pathological
Mazumdar et al., 2017 [37]	UCI Machine Learning Repository	ANN	Normal, Suspect, Pathological
Tang et al., 2018 [38]	Micro fetal heart monitor collected from more than 20 hospitals	SVM, RF, MKNet, MKRNN	Healthy, Abnormal
Feng et al., 2018 [39]	CTU-UHB	Deep Gaussian Processes	Normal, Abnormal
Abry et al., 2018 [40]	CTU-UHB (BDB), LDB	Sparse SVM	Normal, Acidotic
Fergus et al., 2018 [41]	CTU-UHB	RF, SVM	Reassuring, non-reassuring, abnormal
Ramla et al., 2018 [42]	UCI Machine Learning Repository	DT	Normal, Pathological, Suspect

NR: Not reported; ML: Machine Learning; SVM: Support vector machine; DT: Decision tree; RF Random forest; GA: Genetic Algorithm; NN: Neural network; MNN: Modular NN; CNN: Convoluted NN; MCNN: Multimodal CNN; BN: Bayesian network; ELM: Extreme learning machine; RNN: Recurrent NN; ANN: Artificial NN; MLP: Multi-layer perceptron; RBF: radial basis function; kNN: k-nearest neighbour; NB: Naive Bayes; GDI: Gini's diversity index; PCA: Principal component analysis; STI: Short-term irregularity; LTI: Long-term irregularity; STV: Short-term variability; LTV: Long-term variability; II: Interval index; WT: Wavelet transform, LS-SVM: Least Square-SVM, PSO: Particle Swarm Optimization, BDT: Binary Decision Tree, BP: Back Propagation

Author, Year	Acquisition	AI used	CTG Classes
Petrozziello et al., 2019 [43]	Oxford Data, The Signal Processing and Monitoring (SPAM) In Labor Workshop 2017 Database, CTU-UHB	MCNN	Normal, Acidemia
Agrawal et al., 2019 [44]	UCI Machine Learning Repository	DT, SVM, NB	Normal, Suspect, Pathological
Iraji et al., 2019 [45]	UCI Machine Learning Repository	Multi-layer ANFIS topology	Normal, Suspect, Pathological
Ma'sum et al., 2019 [46]	CTU-UHB	CNN DenseNet, SVM	Normal, Hypoxic
Huddar et al., 2019 [47]	UCI Machine Learning Repository	Multi-tasking network	Normal, Suspect, Pathological
Hoodbhoy et al., 2019 [48]	UCI Machine Learning Repository	MLP, XGBoost, DT, RF, Logistic regression, SVM linear kernel, SVM RBF kernel, kNN, NB, AdaBoost	Normal, Suspect, Pathological
Zhao et al., 2019 [49]	CTU-UHB	CNN	Normal, Abnormal
Zhao et al., 2019 [50]	CTU-UHB	CNN	Normal, Abnormal
Signorini et al., 2020 [51]	Hewlett Packard CTG fetal monitors (series 1351A)	DT, GA, SVM	Normal, Abnormal
Gavrilis et al., 2015 [52]	UCI Machine Learning Repository	kNN, Parzen density estimator	Normal, Non-Normal
Das et al., 2020 [53]	CTU-UHB	Bland–Altman plot, Fleiss Kapp, Kendell's coefficient of concordance, Fuzzy Logic	Normal, Stage 1, Stage 2
Tsoulos et al., 2006 [54]	137 of them were acquired using an HP 1350 fetal monitor	BFGS Variant of Powell's	Normal, At Risk

NR: Not reported; ML: Machine Learning; SVM: Support vector machine; DT: Decision tree; RF Random forest; GA: Genetic Algorithm; NN: Neural network; MNN: Modular NN; CNN: Convoluted NN; MCNN: Multimodal CNN; BN: Bayesian network; ELM: Extreme learning machine; RNN: Recurrent NN; ANN: Artificial NN; MLP: Multi-layer perceptron; RBF: radial basis function; kNN: k-nearest neighbour; NB: Naive Bayes; GDI: Gini's diversity index; PCA: Principal component analysis; STI: Short-term irregularity; LTI: Long-term irregularity; STV: Short-term variability; LTV: Long-term variability; II: Interval index; WT: Wavelet transform, LS-SVM: Least Square-SVM, PSO: Particle Swarm Optimization, BDT: Binary Decision Tree, BP: Back Propagation

Author, Year	Acquisition	AI used	CTG Classes
Ravindran et al., 2015 [55]	UCI machine learning repository	kNN-SVM, BN, ELM, SVM	Normal, Pathological, Suspect
Comert et al., 2019 [56]	CTU-UHB	kNN, DT, GDI, SVM	Normal, Hypoxic
NR: Not reported; ML: Machine Learning; SVM: Support vector machine; DT: Decision tree; RF Random forest; GA: Genetic Algorithm; NN: Neural network; MNN: Modular NN; CNN: Convoluted NN; MCNN: Multimodal CNN; BN: Bayesian network; ELM: Extreme learning machine; RNN: Recurrent NN; ANN: Artificial NN; MLP: Multi-layer perceptron; RBF: radial basis function; kNN: k-nearest neighbour; NB: Naive Bayes; GDI: Gini's diversity index; PCA: Principal component analysis; STI: Short-term irregularity; LTI: Long-term irregularity; STV: Short-term variability; LTV: Long-term variability; II: Interval index; WT: Wavelet transform, LS-SVM: Least Square-SVM, PSO: Particle Swarm Optimization, BDT: Binary Decision Tree, BP: Back Propagation			

4.2. Bias Assessment

Figure 3 presents the results of the quality assessment that was performed using the Modified Down and Black's checklist. As may be observed from the figure, only 9/38 (23.7%) studies scored greater than 7 out of 10. This is likely since this checklist was created with the intent of analysing clinical studies rather than preclinical AI validation studies. However, given the generalisability of this tool, it still provides a reasonable assessment of bias for the included studies.

4.3. Synthesis of results

The performance of all included AI models may be found in Table 2. The key performance characteristics that were commonly reported by the included studies included accuracy, sensitivity, and specificity. Although accuracy reflects important information about algorithms and models, it is also important to report sensitivity and specificity to understand how the models perform with true positive and false negative samples.

Table 2
Overall performance across models.

Support Vector Machine	Accuracy	Sensitivity	Specificity
Batra et al., 2017 [33]	93.41	86.7	94
Georgoulas, G., 2006 [20]	78.75		
Yilmaz, E. 2013, LS-SVM-PSO-BDT [26]	91		
Nagendra, V., 2017 [35]	98		
Agrawal et al., 2019 [44]	92.39		
Tang, H., 2018 [38]	83	83	83
Feng, J., 2018 [39]	86	82	82
Xu, L., 2013, GA + SVM classifier [14]	73.58	66.83	81.13
Signorini, M. G., 2020, SVM-radial kernel; SVM-polynomial kernel [51]	86	93	88
Abry, P., 2018, Sparse-SVM [40]	78	64	80
Comert, Z., 2019 [56]	88.58	77.4	93.86
Decision Trees	Accuracy	Sensitivity	Specificity
Batra et al., 2017[33]	90.58	80.85	89.3
Permanasari et al., 2017 [34]	81.6		
Agrawal et al., 2019 [44]	91.54		
Signorini, M. G., 2020 [51]	91.1	87.1	95
Ma'sum, M. A., 2019, Bagging [46]	82		
Ramla, M., 2018, Gini Index [42]	90	89	91
Ramla, M., 2018, Entropy calculation [42]	88.87	89	89
Shah, S. A. A., 2015, REPTree[30]	91.98	87.77	91.4
Comert, Z., 2019[56]	79.34	52.32	92.26
Batra et al., 2017[33]	95.85	93.6	97.4
Hoodbhoy, Z., 2019, XGBoost[48]	87.9		
Neural Networks	Accuracy	Sensitivity	Specificity
Batra et al., 2017 [33]	94.16	89.92	96.84
Comert, Z., 2016, ANN [31]	91.84	94.91	90.66

Support Vector Machine	Accuracy	Sensitivity	Specificity
Petrozziello et al., 2019, MCNN, Stacked MCNN [43]	92		
Warrick, P.A. 2016, GA group (HSMM, LTSM) [32]		74.8	
Tsoulos, I. 2006, MLP [54]	82.5		
Ocak, H., 2013, Genetic Algorithms [27]	99.3		
Tang, H., 2018, MKNet[38]	94.7	94.68	94.71
Tang, H., 2018, MKRNN [38]	90.3	90.28	90.33
Valensise, H., 2006, MLP [21]	86	56	91
Feng, J., 2018, Deep Gaussian Processes [39]		91	82
Haweel, T. 2013, Volterra based neural networks [28]			
Jezewski, M., 2007 [22]	97	88	84
Fergus, P., 2017 [41]	99.9	93.78	90.99
Ma'sum, M. A., 2019, DenseNet[46]	82		
Mazumdar, S., 2017 [37]	99.9		
Huddar, P. P., 2019, Modified Deep NN [47]	74.6		
Jadhav, S., 2011, Modular NN [23]	99		
Hongbiao, Z., 2012, PSO-BP [25]	97.43		
Comert, Z., 2019, ANN [56]	77.7	66.1	83.2
Zhao, Z., DeepFHR_ intelligent prediction, 2019, CNN [50]	98.3	98.2	94.8
Zhao, Z., Computer-Aided Diagnosis System, 2019, CNN [49]	98.69	99.2	98.1
Random Forest	Accuracy	Sensitivity	Specificity
Batra et al., 2017[33]	93.41	85.02	93.3
Nagendra, V., 2017 [35]			
Tang, H., 2018 [38]	84.5	84.5	84.49
Xu, L., 2013 [14]	72.64	67.92	77.36
Signorini, M. G., 2020 [51]	91.1	90.2	91.9
Fergus, P., 2017 [36]	98.12	92.91	91.85
Ma'sum, M. A., 2019 [46]	81		
Shah, S. A. A., 2015 [30]	94.73	88.37	92.87

Support Vector Machine	Accuracy	Sensitivity	Specificity
Custom Algorithm	Accuracy	Sensitivity	Specificity
Batra et al., 2017 [33]	99.25	98.8	99.3
Gavrilis et al., 2015, Nearest Neighbour [52]	75		
Gavrilis et al., 2015, Parzen Density Estimator [52]	52.8		
Dash, S. 2012, Bayesian NN [24]		80	60
Agrawal et al., 2019, Naive Bayes [44]	85.57		
Jezewski, M., 2007, MLP [22]	89	83	85.1
Comert, Z., 2019, KNN [56]	70.47	55.9	74.77
Signorini, M. G., 2020, Logistic Regression [51]	86.7	90	83.3
Iraji, M. S., 2019, Deep learning [45]	96.7	91.15	95.84
Iraji, M. S., 2019, ANFIS-6-4 [45]	95.37	85.37	93.7
Iraji, M. S., 2019, Deep SSAEs [45]	99.5	99.716	97.5
Stylios, I. C., 2014, K-Nearest [29]	73.3		
Fergus, P., 2017, fisher's linear discriminant analysis [36]	78.75	69.73	78.75
Fergus, P., 2018, FLDA + SVM + RF [41]	96	87	90
Das, S., 2020, Fuzzy rule identification [53]	93.8		
Shah, S. A. A., 2015, J48 [30]	93.56	87.67	90.33
Ravindran, S., 2015, Improved Adaptive Genetic Algorithm [55]	93.61		
Comert, Z., 2016, ELM [31]	93.42		

4.4. Support Vector Machine (SVM):

28.9% (11/38) of studies focused on SVM [14, 20, 26, 33, 35, 38–40, 44, 51, 56]. 4 out of the 11 (36.4%) studies used a different combination of SVM. Yilmaz et al. used least squares SVM with Particle Swarm Optimization and Binary Decision Tree producing the 91% accuracy [26]; Xu et al. used Genetic Algorithm followed by SVM classifier delivering 73.58% accuracy [14]; Signorini et al. employed a radial and a polynomial kernel to deliver 86% accuracy [51]; and Abry et al. used a sparse version of SVM to deliver 78% accuracy [40].

4.5. Decision Trees (DT) and Random Forests (RF):

26.3% (10/38) of studies discuss decision trees whilst 21.1% (8/38) of the studies employed random forests [14, 30, 33–36, 38, 42, 44, 46, 48, 51, 56]. While RF studies used a basic algorithm with one study

missing value output, DT on the other hand had 4 studies employing a variation of the DT algorithm. Ma'sum et al. used Bagging algorithm to achieve an 82% accuracy [46]. Shah et al. delivered 91.98% accuracy using REPTree [30]; while Hoodbhoy et al. obtained 87.9% accuracy using XGBoost [48]. Ramla et al. employed two different adaptations - Gini Index and Entropy calculation - to produce accuracies of 90% and 88.87%, respectively [42].

4.6. Neural Networks (NN):

55.3% (21/38) of studies employed different types of NN [21–25, 27–29, 31–33, 37–39, 41, 43, 44, 46, 47, 49, 50, 54, 56]. Out of 21 studies; 10 studies used generic forms of Neural Networks (ANN, CNN, RNN, Modular, etc.), while 2 studies used genetic algorithms. CNN was used by 4 studies, while RNN was used by only 1 study [38].

4.7. Custom Algorithm:

36.8% (14/38) of studies developed novel algorithms, typically using modified ML techniques or combinations of the above [22, 29–31, 33, 36, 41, 44, 45, 51–53, 55, 56]. Agrawal et al. utilised Naive bayesian classifiers to produce 85.57% accuracy, while Stylios et al. and Comert et al. utilised K-Nearest Neighbour classifiers to deliver 73.3% and 70.47% accuracy respectively [29, 44, 56].

4.8. Comparison of performance measures

When comparing the performance measures of accuracy, sensitivity, and specificity; the discussion is often more conclusive if the AI models which are compared employed the same database during the training process. In this section, we divide the performance measures for each model based on the public databases that were employed by the AI developers (ie. UCI Machine Learning Repository and CTU-UHB Intrapartum Cardiotocography Database v1.0.0), and then calculated the mean performance for each machine learning technique for models that employed these respective databases.

4.8.1. UCI Machine Learning Repository:

When evaluating the machine learning techniques that were developed using the UCI Machine Learning Repository (refer to Fig. 4), it is discernible that RF performed the best (mean: 94.07; SD: 0.93), although there was little difference between RF, NN (mean: 93.75; SD: 8.95), and SVM (mean: 93.70; SD: 3.03). Decision trees performed the worst of all the established machine learning techniques (mean: 88.8; SD: 15.34). In relation to sensitivity, SVM, DT, and RF scored similarly with respective mean sensitivities of SVM mean: 86.7 (SD: SD), DT mean: 86.66 (SD: 3.91), RF mean: 86.7 (SD: 2.37). On the other hand, NN and custom algorithms performed highly at producing true positives (NN mean: 92.42 [SD: 3.53], custom algorithm mean: 92.54 [SD: 6.48], respectively). When it came to true negatives, the custom algorithms performed the best (mean: 95.33; SD: 3.48), whilst SVM had a mean of 94.00, NN (mean: 93.75; SD: 4.37), RF (mean: 93.09; SD: 0.3), and DT (mean: 90.18; SD: 1.2). It should be noted however, that for all the machine learning techniques applied on this specific database, the mean sensitivity and specificity were not reported by most studies. For example, the sensitivity and specificity were only reported by 1 out

of 2 SVM models, 5/8 DT, 2/8 NN, 2/3 RF, and 5/10 models using custom algorithms. This incomplete reporting will influence the means for each performance metric and will limit the generalisability of discussions about machine learning technique reliability for these two metrics.

4.8.2. CTU-UHB Intrapartum Cardiotocography Database v1.0.0:

From Fig. 5, we see that although the custom algorithms had the highest mean accuracy (mean: 89.52; SD: 9.39) out of all the machine learning techniques, the difference was not considerably different to random forest (mean: 89.5; SD: 12.11) or neural networks (mean: 87.84; SD: 12.74). Whilst decision trees scored the lowest (mean: 80.67; SD: 1.88), SVM performed slightly better with a mean of 84.19 (SD: 5.52). Although decision trees appear to also have the lowest mean sensitivity, only one study reported this metric for the CTU-UHB database [56]. In comparison, SVM had a sensitivity of mean: 74.47 (SD: 9.35), whilst neural networks, random forest, and other custom algorithms performed better at identifying true positives (NN mean: 84.03 [SD: 18.37], RF mean: 92.91; and custom algorithms mean: 78.37; [SD: 12.21]). Interestingly, decision trees performed better at identifying true negatives (mean: 92.26;) than SVM (mean: 85.29; SD: 7.49), NN (mean: 87.31; SD: 8.81), RF (mean: 91.85;), and other custom algorithms (mean: 84.38; SD: 7.95). As above, it should be noted that for all the machine learning techniques applied on this database, the mean sensitivity and specificity were not reported by a some of the studies. For example, the sensitivity and specificity were reported by 3 out of 3 SVM models, 1/2 DT, 5/6 NN, 1/2 RF, and 3/4 models using custom algorithms. This incomplete reporting will influence the means for each performance metric.

5. Discussion

5.1. AI-based CTG interpretation

This review discusses the present literary landscape of various machine learning techniques in relation to the interpretation of intrapartum CTG signals. To do this, we analysed the current state-of-the-art and identified several techniques which have been applied. These included AIs using the following base algorithms:

5.2. Support Vector Machine (SVM)

SVM makes a good choice for a classifier for the CTG databases that were employed and can be divided into normal and abnormal states. However, without having developed an adequate hyperplane and margin to differentiate the two classes, the system will lose precision while learning the difference between normal and abnormal states [14, 20, 40, 57]. This could have clinical implications as this loss of precision could result in unnecessary intervention due to false classification of borderline normal fetuses into the abnormal class or increase missed diagnoses as borderline abnormal fetuses get misclassified as normal, resulting in a fetal or neonatal morbidity or mortality. In addition to the resultant morbidity and mortality, this would open the hospital or legal manufacturer of the AI to litigation (depending on which

party is legally liable for malpractice)[58–60]. Though this limitation applies for all ML techniques, the inherent binary classification capacity of SVM places it at greater risk of incidence.

While the mean accuracy of the SVM was the weakest in the CTU-UHB database (84.19%), it performed well when applied in the UCI Machine Learning Repository (mean accuracy: 94.00%). After pre-processing Nagendra et al. scored the highest accuracy of all SVM-based classifiers (98%) with a more specified feature vector [35]. If extra variables were employed, they could have carried meaningful information that would influence the overall outcome and introduce the possibility for a third class[35]. A good example for a better application of SVM while maintaining more than two classes is Harimoorthy et al., where three different diseases with common symptoms have been investigated [61]. However, the system was modified to employ a more improved SVM-radial. Therefore, to define SVM's relevance for future development, if the task at hand has two scissile classes, SVM can be a suitable choice. However, if it requires multi-class classification then it is recommended to use an alternate form of SVM or a different classification technique.

5.3. Decision Trees (DT) and Random Forests (RF)

In CTG interpretation, these technologies can be used to predict fetal state using signal changes and the probability that a particular branch of the DT/RF model matches with the signal [62]. This provides both, DT and RF, with an advantage over SVM as developers can incorporate more categories to represent the fetal state (eg. 3: abnormal, suspect, normal). When compared with each other across the CTU-UHB and UCI databases, DT was seen to have performed poorly whilst RF performed extremely well. This suggests that RF may be the superior technique for CTG classification of fetal state.

Interestingly, DT and RF can be combined with SVM to create a broader system to introduce more variables into the system, but that has been shown to compromise accuracy [46]. Fergus et al. employed a deep learning approach for the random forest. In doing so, they were able to incorporate more information that enhanced their accuracy (98.12%)[36]. This signifies that the greater the training dataset and appropriateness of the classifier could make RF models clinically relevant and thus useful in guiding the development of future DT/RF models.

5.4. Neural Networks (NN)

In this study, we identified that NN-based AIs were often utilised, with 21 different models developed: thirteen of which having accuracies above 90%. When we compare the mean performance metrics for NN with the other techniques in both databases, it performed very well. Additionally, two studies reported a 99.9% accuracy though it is uncertain if this was tested on a training dataset or a separate validation dataset[36, 37]. If the latter, this becomes an indicator that the system can associate each class to an almost perfect score and that these models hold considerable promise clinically if it can maintain a high standard in clinical evaluations.

Another advantage of using neural networks is relative alteration depending on the application. Artificial neural networks can be used for deep and shallow networks[21–25, 27–29, 31, 33, 37–39, 41, 44, 47, 54,

56]. Convolutional neural networks help look at the shape of signals image processing focus which helps predict accelerations[43, 46, 49, 50], and recurrent neural networks can perform time-based predictions using trends of previous samples of a signal to analyse current and future samples of the same signal[32, 38].

As NNs improve with time, hyperparameters will future proof the concept and increase the relevance of the technique in applications like CTG interpretation. Using hyperparameters, the network structure can be optimised before training or bias has been introduced into the model[63]. This could improve the accuracy and reliability of NN-based models through the incorporation of hidden layers, or even denser, bigger layers via the addition of hidden units. Furthermore, hyperparameters permit different optimizers, regularisation, activation, enhanced learning, and the reduction of overfitting (dropouts)[64]. All these capabilities highlight the promise for NN future NN applications in CTG.

5.5. Custom Algorithms

The category of custom algorithms related to the models that used novel techniques and either combinations or modified versions of the SVM, DT, RF, or NN. Modified versions of NNs can be beneficial as they are less computational hungry than traditional NNs whilst also not compromising the accuracy of the model. Although these models did not score the highest accuracies, they offer a good option for clinical applications where limited computational power and IT resources are common, for example in low-middle income countries, regional hospitals, small clinics, etc., where access to the up-to-date computers or high-powered computers are limited. That said, combinations of different AI techniques can help increase the accuracy. In Fergus et al., Fishers' Linear Discriminant Analysis, SVM and RF were used to produce their final model which performed at 96% accuracy and required little computational power[41].

5.6. Clinical Implications

As mentioned, most of the included studies used publicly available databases which is an inadequate representation of common practices and the general population. Also, these databases make no distinction of the condition of the maternal, fetal, or placental factors (eg. placental insufficiency or fetal growth restriction which can influence fetal risk to hypoxic damage. However, these databases provide sufficient samples to justify developing and testing modern machine learning techniques. In any CTG trace, there are two streams of signals that are detected: one corresponding to the fetal heart rate and the other to the mother's uterine contractions. Additionally, an event tracker may be provided to the mother to mark on the CTG trace, when the mother detects a fetal movement. The samples of these signals are stored in the form of values which makes it easy to load as a table for processing and set-up as an input for the models we have discussed. However, for interpretation, the signals are given to medical staff and parents as a graph print out which is time dependent[65]. Such graph printouts are a good dataset for time-series models such as recurrent neural networks, Warrick et al. and Tang et al. set out to test [32, 38]. Alternatively, the printouts have been seen as 2D images inputs to convolutional neural networks, where various features and sequences have been successfully observed by the models.

Another point that must be raised is that almost all the studies in this review focused on a 2 (normal and abnormal) or 3 (normal, suspect, and pathological) classifier system. The latter system emulates the current clinical standard of the FIGO classification system where CTGs are also designated as either normal, suspect, or pathological[65]. That said, CTG classification does not necessarily correlate to neonatal outcome [9]. However, the capabilities of AI in clinical decision support presently supersede this and as such, there is potential, particularly with NN-based models, to go one step further and provide a clinical diagnosis. Indeed, in cardiology wards, many commercial electrocardiography (ECG) telemetry units are connected to a central screen which can detect when a patient has a sinus rhythm or is experiencing arrhythmias such as tachycardia, ventricular fibrillation, and atrial flutter to name a few to then alarm and alert clinicians[66–69]. In this same way, AI-based CTG interpretation could go further to identify and mark CTG features such as accelerations, early and late decelerations, variability; and potentially identify or aid managing fetal or maternal issues that can be associated with fetal heart rate, such as: congenital abnormalities, fetal compromise, chorioamnionitis, fetal immaturity at gestational age, acute and chronic hypoxemia, and fetal growth restriction[70–72]. In doing so, AI will help change existing point of care CTG based interpretation to a more constant and longitudinal based assessment.

One well known clinical decision support system in the field of fetal monitoring is the INFANT [73]. The system was built to suggest decisions to be carried out by the clinicians during intrapartum periods of the pregnancy. Though the system was still in development [74] the clinical trial was deemed failed by the team, as the system failed to perform better than the professional staff [75]. The system's patent supplied has shown that the system runs time-based neural network (RNN) where each sample relies on the previous sample to predict the current state of the fetus while employing a sigmoid based classifier [76]. Similar systems in this review have shown to score 90% accuracy in separating different fetal states with minimal pre-processing of the samples. In addition, all the information found regarding the INFANT system focused on the signal analysis, while neglecting the innovative side of the machine learning technique.

Based on the findings of this review, AI has demonstrated the potential to distinguish different fetal states with high accuracy. However, none of these systems have been proven to provide diagnostic level details for clinical implementation, as the training datasets are skewed by a lack of understanding of the complexity of the clinical challenge. At the current stage of development, the models have only been tested internally, thus have not been validated externally. As such, there is no evidence that the model will perform at the same accuracy in a clinical setting. Although the validation is promising and justifiable to test for development, more concrete clinical trials are required before considering whether implementation is feasible.

5.7. The evolving technological paradigm in FHR monitoring

There is no doubt that CTG has significant clinical utility for obstetricians, midwives, and pregnant women. It is a technology that has come to represent the standard-of-care for non-invasive electronic fetal monitoring. However, with technological advancements and the growing interest in fetal electrocardiography (fECG), the future of electronic fetal monitoring is likely to place decreasing reliance

on CTG as the industry shifts to fECG. This is because fECG can provide more accurate information than CTG and does not have the physical challenges of losing signal due to transducer placement, limiting maternal mobility, inconvenient attachments, limitations with use for women with high BMIs, etc [7, 77–79]. Indeed, large industry players such as Philips Healthcare [80] and GE Healthcare [81] are already introducing fECG-based devices, and numerous small-to-medium sized enterprises are developing technologies using a similar foundation.

Whilst CTGs are unlikely to be replaced immediately, there will likely be a slow, phased transition towards the implementation of fECG technologies. This is a consideration for those developing AI models for CTG interpretation as their target user will shift towards rural clinics and developing markets over the next decade or so[82].

5.8. Avenues for Future Research Direction

With the growing interest in digital technologies and artificial intelligence in healthcare, fetal monitoring and CTG interpretation is certainly an area that is promising, yet still requires a lot of work. Based on the findings of this review, the authors believe that there are several opportunities for improvement.

Firstly, there is the need for more accurate and reliable models to be developed and evaluated. Though many of the models performed extremely well in their pre-clinical validation, it must be anticipated that clinical performance may be considerably lower, and this should be factored in the development. Along this line is the need for the models to be evaluated in a clinical scenario under trial circumstances. As such, clinical evaluation studies should be pursued to determine if the high accuracies reported by the studies included in this review are indeed representative of how the model will perform in a clinical setting. In turn, such studies could assist in the implementation of promising models as potential adopters will become more aware of the technologies[83].

The second, as previously discussed, is in expanding the endpoints of the AI tools beyond simple classification as per FIGO guidelines alone, to providing more comprehensive clinical decision support and in the future, perhaps fully automated interpretation. Hopefully, this will enable a more complete overview of the fetus and establish an individualised risk model for hypoxic injury. However, to achieve this, developers must develop a better understanding of CTG traces and its link with metabolic acidemia in the fetus. This will also require an understanding of the underlying maternal, fetal, and placental factors which contribute towards evolving antenatal fetal risk; the link between CTG abnormalities and neonatal outcomes; the link between observed changes on CTG and the subsequent clinical management; and the needs of the clinicians who will use this information in their day-to-day practice. For the last point, the needs of the clinician is highly important as alarm fatigue and other human factors considerations can play a significant role in the design of the model and the broader software package [84].

Lastly, with growing interest in alternative fetal monitoring technologies such as fECG, there is the potential that development of AIs for fECG may prove to be a promising new research direction soon.

Particularly as fECG can provide more detailed information about the fetuses' wellbeing [85].

5.9. Limitations of this study

The findings of this review should be interpreted considering the following limitations.

Firstly, none of the papers presented in this study discussed the application of these AIs to the clinical setting without the external validation of a clinical evaluation. This could be a limitation that is associated with the readiness or adequacy of these AIs for clinical evaluation, that the research in this field is progressing at a slow rate, or more likely a sign that the techniques are not mature yet given that only 3 studies were published prior to 2010. However, an associated limitation with this is that most studies used one of two databases (CTU-UHB or UCI Machine Learning Repository) as the source of data for training and validating their respective AI models. Though these databases form a good starting point for AI developers, this raises potential issues with external validity and generalisability of the models to the general population. Particularly as the UCI Machine Learning Repository did not present any inclusion/exclusion criteria or demographic data for patients, whilst the CTU-UHB database only included patients with gestational ages above 36 weeks and experienced a Stage 2 labour duration of < 30 minutes.

Secondly, being of a primarily pre-clinical nature, the studies included in this review were deemed to be at high risk of bias due to their inherent design and methodological shortcomings. For this reason, meta-analysis was ruled inappropriate, but this limitation should be considered when interpreting the findings.

Lastly, whilst the databases were largely the same across all studies, differences in skill, experience, and training of the developer can influence the accuracy and reliability of the AI model.

Despite these limitations, it must be highlighted that the included studies have demonstrated promise and have outlined the merit in further evaluating their ability to interpret CTG accurately and reliably.

6. Conclusion

This review identified that the application of AI and machine learning techniques to CTG interpretation is still in its early days, as it is an area of health informatics that is still being developed. The findings of this review illustrate that the models performed relatively well in pre-clinical evaluations signifying potential, however, the true clinical promise has yet to be realised as there were no clinical trials identified from our search. The most common machine learning tools that have been applied included algorithms based on SVM, decision trees, random forests, and neural networks; with some custom algorithms being developed as well. At present, most of these models were developed in line with current clinical CTG classification, which often follows FIGO guidelines. However, the authors believe that AI has the capability to go beyond the basic CTG interpretation functionality described by the included studies to result in the development of more comprehensive clinical decision support systems. In doing so, developers will help enable a precision medicine approach that can address current clinical limitations with CTG interpretation.

Declarations

Statements and Declarations:

The authors did not receive support, grants, or funding from any organization for the submitted work. The authors have no competing interests to declare that are relevant to the content of this article.

Author Contributions:

Conceptualization: Mohamed Salih, Ritesh Rikain Warty, Vinayak Smith; Methodology: Mohamed Salih, Ritesh Rikain Warty, Vinayak Smith; Formal analysis and investigation: Mohamed Salih, Ritesh Rikain Warty, Hamsaveni Kalina Murday, Arjun Kaushik, Vinayak Smith; Writing - original draft preparation: Mohamed Salih, Ritesh Rikain Warty, Hamsaveni Kalina Murday; Writing - review and editing: All Authors; Supervision: Sandeep Reddy, Beverley Vollenhoven, Hamid Rezatofighi, Wenlong Cheng, Vinayak Smith.

References

1. B. M. Nowotny, E. Loh, M. Davies-Tuck, R. Hodges, and E. M. Wallace, Using patient factors to predict obstetric complaints and litigation: A mixed methods approach to quality improvement, *Journal of Patient Safety and Risk Management*, vol. 23, no. 5, pp. 185-199, 2018. <https://doi.org/10.1177/2516043518799020>
2. D. Shaw *et al.*, Drivers of maternity care in high-income countries: can health systems support woman-centred care?, *The Lancet*, vol. 388, no. 10057, pp. 2282-2295, 2016. [https://doi.org/10.1016/s0140-6736\(16\)31527-6](https://doi.org/10.1016/s0140-6736(16)31527-6)
3. Annual report and accounts 2017/18, in "NHS Resolution," National Health Service, UK(2018), Available: <https://resolution.nhs.uk/wpcontent/uploads/2018/08/NHS-Resolution-Annual-Report-2017-2018.pdf>.
4. J. Adinma, Litigations and the Obstetrician in Clinical Practice, *Ann Med Health Sci Res*, vol. 6, no. 2, pp. 74-9, Mar-Apr 2016. <https://doi.org/10.4103/2141-9248.181847>
5. E. I. Emin, E. Emin, A. Papalois, F. Willmott, S. Clarke, and M. Sideris, Artificial Intelligence in Obstetrics and Gynaecology: Is This the Way Forward?, *In Vivo*, vol. 33, no. 5, pp. 1547-1551, Sep-Oct 2019. <https://doi.org/10.21873/invivo.11635>
6. G. German Society of, Obstetrics, G. Maternal Fetal Medicine Study, M. German Society of Prenatal, Obstetrics, and M. German Society of Perinatal, S1-Guideline on the Use of CTG During Pregnancy and Labor: Long version - AWMF Registry No. 015/036, *Geburtshilfe Frauenheilkd*, vol. 74, no. 8, pp. 721-732, Aug 2014. <https://doi.org/10.1055/s-0034-1382874>
7. Z. Alfirevic, D. Devane, G. M. Gyte, and A. Cuthbert, Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour, *Cochrane Database Syst Rev*, vol. 2, p. CD006066, Feb 3 2017. <https://doi.org/10.1002/14651858.CD006066.pub3>

8. ((2021)). *Avoiding CTG Misinterpretation*. Available: <https://www.huntleigh-diagnostics.com/media/avoiding-ctg-misinterpretation>
9. C. Yu and S. Bower, *Fetal Growth*, (2015), pp. 211-222.
10. V. Gintautas, G. Ramonienė, and D. Simanavičiūtė. ((Undated)). *Cardiotocography*.
11. J. H. Thrall *et al.*, Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success, *J Am Coll Radiol*, vol. 15, no. 3 Pt B, pp. 504-508, Mar 2018. <https://doi.org/10.1016/j.jacr.2017.12.026>
12. P. Iftikhar, M. V. Kuijpers, A. Khayyat, A. Iftikhar, and M. DeGouvia De Sa, Artificial Intelligence: A New Paradigm in Obstetrics and Gynecology Research and Clinical Practice, *Cureus*, vol. 12, no. 2, p. e7124, Feb 28 2020. <https://doi.org/10.7759/cureus.7124>
13. Z. Zhao, Y. Zhang, and Y. Deng, A Comprehensive Feature Analysis of the Fetal Heart Rate Signal for the Intelligent Assessment of Fetal State, *J Clin Med*, vol. 7, no. 8, Aug 20 2018. <https://doi.org/10.3390/jcm7080223>
14. L. Xu, Georgieva, A., Redman, C. W., Payne, S. J., Feature selection for computerized fetal heart rate analysis using genetic algorithms, (in eng), *Conf Proc IEEE Eng Med Biol Soc*, vol. 2013, pp. 445-8, 2013. <https://doi.org/10.1109/embc.2013.6609532>
15. D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and P. Group, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, *PLoS Med*, vol. 6, no. 7, p. e1000097, Jul 21 2009. <https://doi.org/10.1371/journal.pmed.1000097>
16. M. Salih, R. Warty, V. Smith, H. Murday, A. Kaushik, and F. Horta, Revolutionising obstetrics: a systematic review on the application of artificial intelligence in cardiotocography., 2021.
17. J. A. Aubut, S. Marshall, M. Bayley, and R. W. Teasell, A comparison of the PEDro and Downs and Black quality assessment tools using the acquired brain injury intervention literature, *NeuroRehabilitation*, vol. 32, no. 1, pp. 95-102, 2013. <https://doi.org/10.3233/NRE-130826>
18. D. Dua and C. Graff, UCI Machine Learning Repository, ed. University of California, Irvine, School of Information and Computer Sciences, (2017).
19. V. Chudacek *et al.*, Open access intrapartum CTG database, *BMC Pregnancy Childbirth*, vol. 14, p. 16, Jan 13 2014. <https://doi.org/10.1186/1471-2393-14-16>
20. D. S. G. Georgoulas, P. Groumpos, Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines, *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 5, pp. 875-884, 2006. <https://doi.org/10.1109/TBME.2006.872814>
21. H. Valensise, Facchinetti, F., Vasapollo, B., Giannini, F., Monte, I. D., Arduini, D., The computerized fetal heart rate analysis in post-term pregnancy identifies patients at risk for fetal distress in labour, (in eng), *Eur J Obstet Gynecol Reprod Biol*, vol. 125, no. 2, pp. 185-92, Apr 1 2006. <https://doi.org/10.1016/j.ejogrb.2005.06.034>
22. M. Jezewski *et al.*, Some practical remarks on neural networks approach to fetal cardiotocograms classification, (in English), *Conference proceedings : ..* vol. Annual International Conference of the

- IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference., pp. 5170-5173, 2007.
23. S. N. S. Jadhav, A. Ghatol, Modular neural network model based foetal state classification, in *2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW)*, 2011, pp. 915-917.
 24. J. G. Q. S. Dash, P. M. Djurić, Learning dependencies among fetal heart rate features using Bayesian networks, in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp. 6204-6207.
 25. Y. G. Z. Hongbiao, Identification of CTG Based on BP Neural Network Optimized by PSO, in *2012 11th International Symposium on Distributed Computing and Applications to Business, Engineering & Science*, 2012, pp. 108-111.
 26. E. Yilmaz and C. Kilicier, Determination of fetal state from cardiotocogram using LS-SVM with particle swarm optimization and binary decision tree, (in English), *Computational & Mathematical Methods in Medicine*, Evaluation Study Validation Study vol. 2013, p. 487179, 2013. <https://doi.org/10.1155/2013/487179>
 27. H. Ocak, A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being, (in eng), *J Med Syst*, vol. 37, no. 2, p. 9913, Apr 2013. <https://doi.org/10.1007/s10916-012-9913-4>
 28. J. I. B. T. I. Haweel, Volterra neural analysis of fetal cardiotocographic signals, in *2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, 2013, pp. 1-5.
 29. V. V. I. C. Stylios, I. Androulidakis, Performance comparison of Machine Learning Algorithms for diagnosis of Cardiotocograms with class inequality, in *2014 22nd Telecommunications Forum Telfor (TELFOR)*, 2014, pp. 951-954.
 30. W. A. S. A. A. Shah, M. Arif, M. S. A. Nadeem, Decision Trees Based Classification of Cardiotocograms Using Bagging Approach, in *2015 13th International Conference on Frontiers of Information Technology (FIT)*, 2015, pp. 12-17.
 31. A. F. K. Z. Cömert, S. Güngör, Cardiotocography signals with artificial neural network and extreme learning machine, in *2016 24th Signal Processing and Communication Application Conference (SIU)*, 2016, pp. 1493-1496.
 32. E. F. H. P. A. Warrick, Antenatal fetal heart rate acceleration detection, in *2016 Computing in Cardiology Conference (CinC)*, 2016, pp. 893-896.
 33. A. C. A. Batra, V. Matoria, Cardiotocography Analysis Using Conjunction of Machine Learning Algorithms, in *2017 International Conference on Machine Vision and Information Technology (CMVIT)*, 2017, pp. 1-6.
 34. A. N. A. E. Permanasari, Decision tree to analyze the cardiotocogram data for fetal distress determination, in *2017 International Conference on Sustainable Information Engineering and Technology (SIET)*, 2017, pp. 459-463.

35. H. G. V. Nagendra, D. Sampath, S. Corns, S. Long, Evaluation of support vector machines and random forest classifiers in a real-time fetal monitoring system based on cardiotocography data, in *2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 2017, pp. 1-6.
36. P. Fergus, A. Hussain, D. Al-Jumeily, D. S. Huang, and N. Bouguila, Classification of caesarean section and normal vaginal deliveries using foetal heart rate signals and advanced machine learning algorithms, (in English), *BioMedical Engineering Online*, vol. 16 (1) (no pagination), no. 89, 06 Jul 2017. [https://doi.org/ 10.1186/s12938-017-0378-z](https://doi.org/10.1186/s12938-017-0378-z)
37. R. C. S. Mazumdar, A. Swetapadma, An innovative method for fetal health monitoring based on artificial neural network using cardiotocography measurements, in *2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, 2017, pp. 265-268.
38. H. Tang, T. Wang, M. Li, and X. Yang, The Design and Implementation of Cardiotocography Signals Classification Algorithm Based on Neural Network, (in English), *Computational & Mathematical Methods in Medicine*, vol. 2018, p. 8568617, 2018. <https://doi.org/10.1155/2018/8568617>
39. J. G. Q. G. Feng, P. M. Djurić, Supervised and Unsupervised Learning of Fetal Heart Rate Tracings with Deep Gaussian Processes, in *2018 14th Symposium on Neural Networks and Applications (NEUREL)*, 2018, pp. 1-6.
40. P. Abry, Spilka, J., Leonarduzzi, R., Chudacek, V., Pustelnik, N., Doret, M., Sparse learning for Intrapartum fetal heart rate analysis, (in English), *Biomedical Physics and Engineering Express*, vol. 4 (3) (no pagination), no. 034002, 25 Apr 2018. <https://doi.org/10.1088/2057-1976/aabc64>
41. P. Fergus, M. Selvaraj, and C. Chalmers, Machine learning ensemble modelling to classify caesarean section and vaginal delivery types using Cardiotocography traces, (in English), *Computers in Biology & Medicine*, vol. 93, pp. 7-16, 02 01 2018. <https://doi.org/10.1016/j.compbimed.2017.12.002>
42. S. S. M. Ramla, S. Nickolas, Fetal Health State Monitoring Using Decision Tree Classifier from Cardiotocography Measurements, in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2018, pp. 1799-1803.
43. C. W. G. R. A. Petrozziello, A. T. Papageorghiou, I. Jordanov, A. Georgieva, Multimodal Convolutional Neural Networks to Detect Fetal Compromise During Labor and Delivery, *IEEE Access*, vol. 7, pp. 112026-112036, 2019. <https://doi.org/10.1109/ACCESS.2019.2933368>
44. H. M. K. Agrawal, Cardiotocography Analysis for Fetal State Classification Using Machine Learning Algorithms, in *2019 International Conference on Computer Communication and Informatics (ICCCI)*, 2019, pp. 1-6.
45. M. S. Iraj, Prediction of fetal state from the cardiotocogram recordings using neural network models, (in English), *Artificial Intelligence in Medicine*, vol. 96, pp. 33-44, May 2019. <https://doi.org/10.1016/j.artmed.2019.03.005>
46. P. R. D. I. M. A. Ma'sum, W. Jatmiko, A. A. Krisnadhi, N. A. Setiawan, I. M. A. D. Suarjaya, Improving Deep Learning Classifier for Fetus Hypoxia Detection in Cardiotocography Signal, in *2019*

International Workshop on Big Data and Information Security (IW BIS), 2019, pp. 51-56.

47. S. A. S. P. P. Huddar, Acquiring Domain Knowledge for Cardiotocography: A Deep Learning Approach, in *2019 3rd International Conference on Informatics and Computational Sciences (ICICoS)*, 2019, pp. 1-6.
48. Z. Hoodbhoy, Noman, M., Shafique, A., Nasim, A., Chowdhury, D., Hasan, B., Use of Machine Learning Algorithms for Prediction of Fetal Risk using Cardiotocographic Data, (in eng), *Int J Appl Basic Med Res*, vol. 9, no. 4, pp. 226-230, Oct-Dec 2019. https://doi.org/10.4103/ijabmr.IJABMR_370_18
49. Z. Zhao, Y. Zhang, Z. Comert, and Y. Deng, Computer-aided diagnosis system of fetal hypoxia incorporating recurrence plot with convolutional neural network, (in English), *Frontiers in Physiology*, vol. 10 (MAR) (no pagination), no. 255, 2019. <https://doi.org/10.3389/fphys.2019.00255>
50. Z. Zhao, Y. Deng, Y. Zhang, X. Zhang, and L. Shao, DeepFHR: intelligent prediction of fetal Acidemia using fetal heart rate signals based on convolutional neural network, (in English), *BMC medical informatics and decision making*, Review vol. 19, no. 1, p. 286, 30 Dec 2019. <https://doi.org/10.1186/s12911-019-1007-5>
51. M. G. Signorini, N. Pini, A. Malovini, R. Bellazzi, and G. Magenes, Integrating machine learning techniques and physiology based heart rate features for antepartum fetal monitoring, (in English), *Computer Methods and Programs in Biomedicine*, vol. 185 (no pagination), no. 105015, March 2020. <https://doi.org/10.1016/j.cmpb.2019.105015>
52. G. N. D. Gavrilis, G. Georgoulas, A one-class approach to cardiotocogram assessment, in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 518-521.
53. S. Das, S. M. Obaidullah, K. C. Santosh, K. Roy, and C. K. Saha, Cardiotocograph-based labor stage classification from uterine contraction pressure during ante-partum and intra-partum period: a fuzzy theoretic approach, (in English), *Health Information Science and Systems*, vol. 8 (1) (no pagination), no. 16, 01 Dec 2020. <https://doi.org/10.1007/s13755-020-00107-7>
54. G. G. I. Tsoulos, D. Gavrilis, C. Stylios, J. Bemardes, P. Groumpos, Introducing Grammatical Evolution in Fetal Heart Rate Analysis and Classification, in *2006 3rd International IEEE Conference Intelligent Systems*, 2006, pp. 560-565.
55. S. Ravindran, A. B. Jambek, H. Muthusamy, and S. C. Neoh, A novel clinical decision support system using improved adaptive genetic algorithm for the assessment of fetal well-being, (in English), *Computational and Mathematical Methods in Medicine*, vol. 2015 (no pagination), no. 283532, 22 Feb 2015. <https://doi.org/10.1155/2015/283532>
56. Z. Comert, Sengur, A., Budak, U., Kocamaz, A. F., Prediction of intrapartum fetal hypoxia considering feature selection algorithms and machine learning models, (in English), *Health Information Science and Systems*, vol. 7 (1) (no pagination), no. 17, 01 Dec 2019. <https://doi.org/10.1007/s13755-019-0079-z>
57. S. N. Srihari, A. Xu, and M. K. Kalera, Learning strategies and classification methods for off-line signature verification, presented at the Ninth International Workshop on Frontiers in Handwriting

- Recognition, Kokubunji, Japan, 2004.
58. F. Griffin, Artificial Intelligence and Liability in Health Care, *31 Health Matrix: Journal of Law-Medicine* 65-106, 2021.
 59. A. Wong *et al.*, External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients, *JAMA Intern Med*, vol. 181, no. 8, pp. 1065-1070, Aug 1 2021.
<https://doi.org/10.1001/jamainternmed.2021.2626>
 60. S. Gerke, T. Timo Minssen, and G. Cohen, Ethical and legal challenges of artificial intelligence-driven healthcare, *Artificial Intelligence in Healthcare*, pp. 295–336, 2020. <https://doi.org/10.1016/B978-0-12-818438-7.00012-5>
 61. K. Harimoorthy and M. Thangavelu, Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system, *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 3715-3723, 2020. <https://doi.org/10.1007/s12652-019-01652-0>
 62. A. Hassan and K. Zhang, Using Decision Trees to Predict the Certification Result of a Build, presented at the 21st IEEE/ACM International Conference on Automated Software Engineering (ASE'06), 2006.
 63. T. B. Ludermir, A. Yamazaki, and C. Zanchettin, An optimization methodology for neural network weights and architectures, *IEEE Trans Neural Netw*, vol. 17, no. 6, pp. 1452-9, Nov 2006.
<https://doi.org/10.1109/TNN.2006.881047>
 64. M. Claesen and B. Moor, Hyperparameter Search in Machine Learning, *CoRR*, vol. abs/1502.02127, 2015.
 65. D. Ayres-de-Campos, C. Y. Spong, E. Chandrharan, and F. I. F. M. E. C. Panel, FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography, *Int J Gynaecol Obstet*, vol. 131, no. 1, pp. 13-24, Oct 2015. <https://doi.org/10.1016/j.ijgo.2015.06.020>
 66. ((2015)). *IntelliVue Information Center iX*. Available: <https://ek.so-hf.no/docs/pub/DOK34866.pdf>
 67. ((2021)). *CARESCAPE Central Station*. Available: <https://www.gehealthcare.com.au/products/patient-monitoring/patient-monitors/carescape-central-station>
 68. ((2021)). *Infinity@ CentralStation Wide*. Available:
https://www.draeger.com/en_aunz/Products/Infinity-CentralStation
 69. Z. Chen, Z. Lin, P. Wang, and M. Ding, Negative-ResNet: noisy ambulatory electrocardiogram signal classification scheme, *Neural Computing and Applications*, vol. 33, no. 14, pp. 8857-8869, 2021.
<https://doi.org/10.1007/s00521-020-05635-7>
 70. C. Kouskouti, K. Regner, J. Knabl, and F. Kainer, Cardiotocography and the evolution into computerised cardiotocography in the management of intrauterine growth restriction, *Arch Gynecol Obstet*, vol. 295, no. 4, pp. 811-816, Apr 2017. <https://doi.org/10.1007/s00404-016-4282-8>
 71. L. Galli, A. Dall'Asta, V. Whelehan, A. Archer, and E. Chandrharan, Intrapartum cardiotocography patterns observed in suspected clinical and subclinical chorioamnionitis in term fetuses, *J Obstet Gynaecol Res*, vol. 45, no. 12, pp. 2343-2350, Dec 2019. <https://doi.org/10.1111/jog.14133>

72. M. Eleftheriades, P. Pervanidou, and G. Chrousos, Fetal Stress, *Encyclopedia of Stress*, pp. 46-51, 01/01 2010. <https://doi.org/10.1016/B978-012373947-6.00492-X>
73. ((2017)). *Computerised Interpretation of Fetal Monitoring During Labour*. Available: <https://www.k2ms.com/wp-content/themes/k2ms/documents/infant-guardian/VOL3-Web-Download.pdf>
74. P. Brocklehurst and I. C. Group, A study of an intelligent system to support decision making in the management of labour using the cardiotocograph - the INFANT study protocol, *BMC Pregnancy Childbirth*, vol. 16, p. 10, Jan 20 2016. <https://doi.org/10.1186/s12884-015-0780-0>
75. P. Brocklehurst *et al.*, Computerised interpretation of fetal heart rate during labour (INFANT): a randomised controlled trial, *The Lancet*, vol. 389, no. 10080, pp. 1719-1729, 2017. [https://doi.org/10.1016/s0140-6736\(17\)30568-8](https://doi.org/10.1016/s0140-6736(17)30568-8)
76. R. Keith, Signal analyser, Devon, GB, 2014. Available: <https://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnethtml%2FPTO%2Fsearch-adv.htm&r=2&f=G&l=50&d=PTXT&p=1&S1=8751434&OS=8751434&RS=8751434>.
77. E. W. Abdulhay, R. J. Oweis, A. M. Alhaddad, F. N. Sublaban, M. A. Radwan, and H. M. Almasaeed, Review Article: Non-Invasive Fetal Heart Rate Monitoring Techniques, *Biomedical Science and Engineering*, vol. 2, no. 3, pp. 53-67, 2014/08/12 2021.
78. A. A. Boatin *et al.*, Wireless fetal heart rate monitoring in inpatient full-term pregnant women: testing functionality and acceptability, *PLoS One*, vol. 10, no. 1, p. e0117043, 2015. <https://doi.org/10.1371/journal.pone.0117043>
79. S. M. Vijgen *et al.*, Cost-effectiveness of cardiotocography plus ST analysis of the fetal electrocardiogram compared with cardiotocography only, *Acta Obstet Gynecol Scand*, vol. 90, no. 7, pp. 772-8, Jul 2011. <https://doi.org/10.1111/j.1600-0412.2011.01138.x>
80. ((2021)). *Avalon beltless fetal monitoring solution*. Available: <https://www.philips.com.au/healthcare/product/HC866488/avalon-beltless-fetal-monitoring-solution>
81. ((2021)). *Novii Wireless Patch System*. Available: <https://www.gehealthcare.com.au/products/maternal-infant-care/fetal-monitors/novii-wireless-patch-system>
82. J. Reinhard *et al.*, Intrapartum signal quality with external fetal heart rate monitoring: a two way trial of external Doppler CTG ultrasound and the abdominal fetal electrocardiogram, *Arch Gynecol Obstet*, vol. 286, no. 5, pp. 1103-7, Nov 2012. <https://doi.org/10.1007/s00404-012-2413-4>
83. R. R. Warty, V. Smith, M. Salih, D. Fox, S. L. McArthur, and B. W. Mol, Barriers to the diffusion of medical technologies within healthcare: A systematic review, *IEEE Access*, pp. 1-1, 2021. <https://doi.org/10.1109/access.2021.3118554>
84. S. Sendelbach and M. Funk, Alarm fatigue: a patient safety concern, *AACN Adv Crit Care*, vol. 24, no. 4, pp. 378-86; quiz 387-8, Oct-Dec 2013. <https://doi.org/10.1097/NCI.0b013e3182a903f9>
85. E. Keenan, Modelling techniques to improve the reliability of non-invasive fetal electrocardiography, PhD, Electrical and Electronic Engineering, The University of Melbourne, Australia, (2021).

Figures

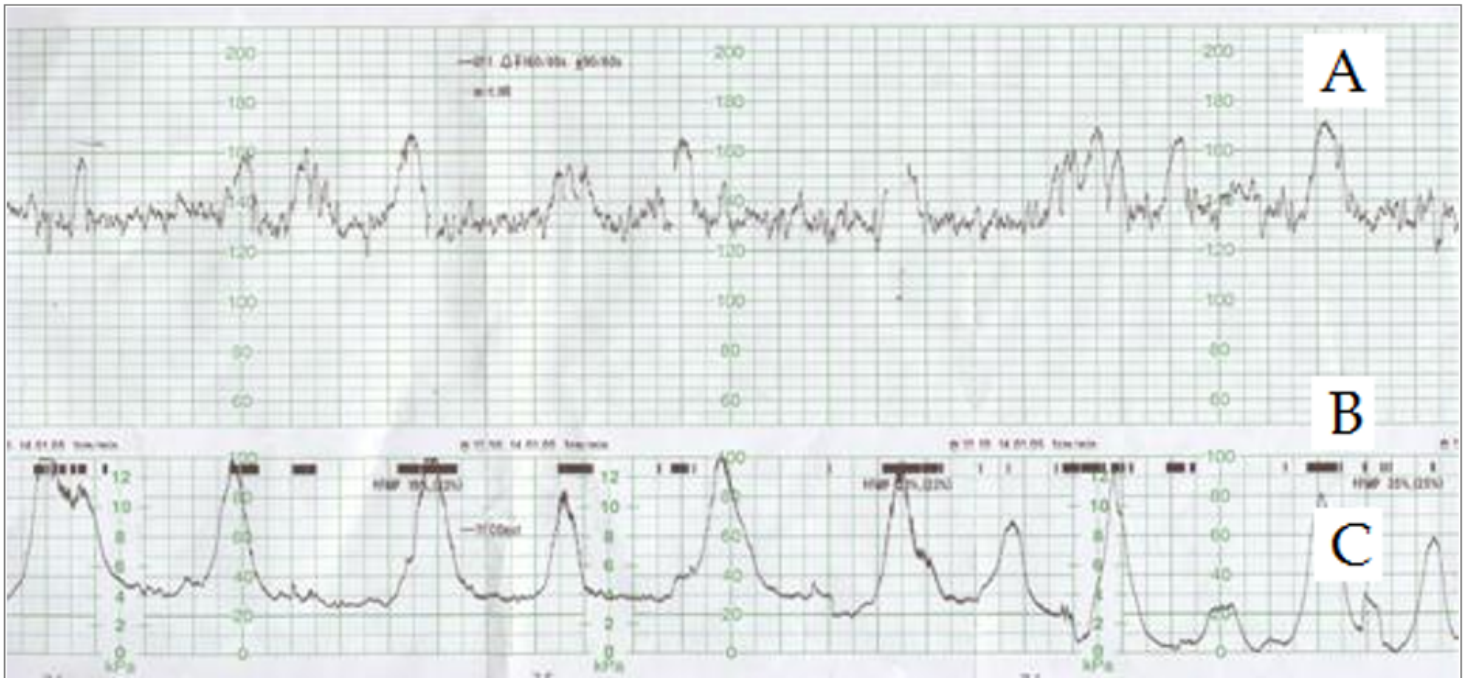


Figure 1

Normal and reassuring CTG trace with a baseline of around 130bpm, accelerations, baseline variability, and no decelerations. A: fetal heart rate; B: fetal movements; C: uterine contractions [10].

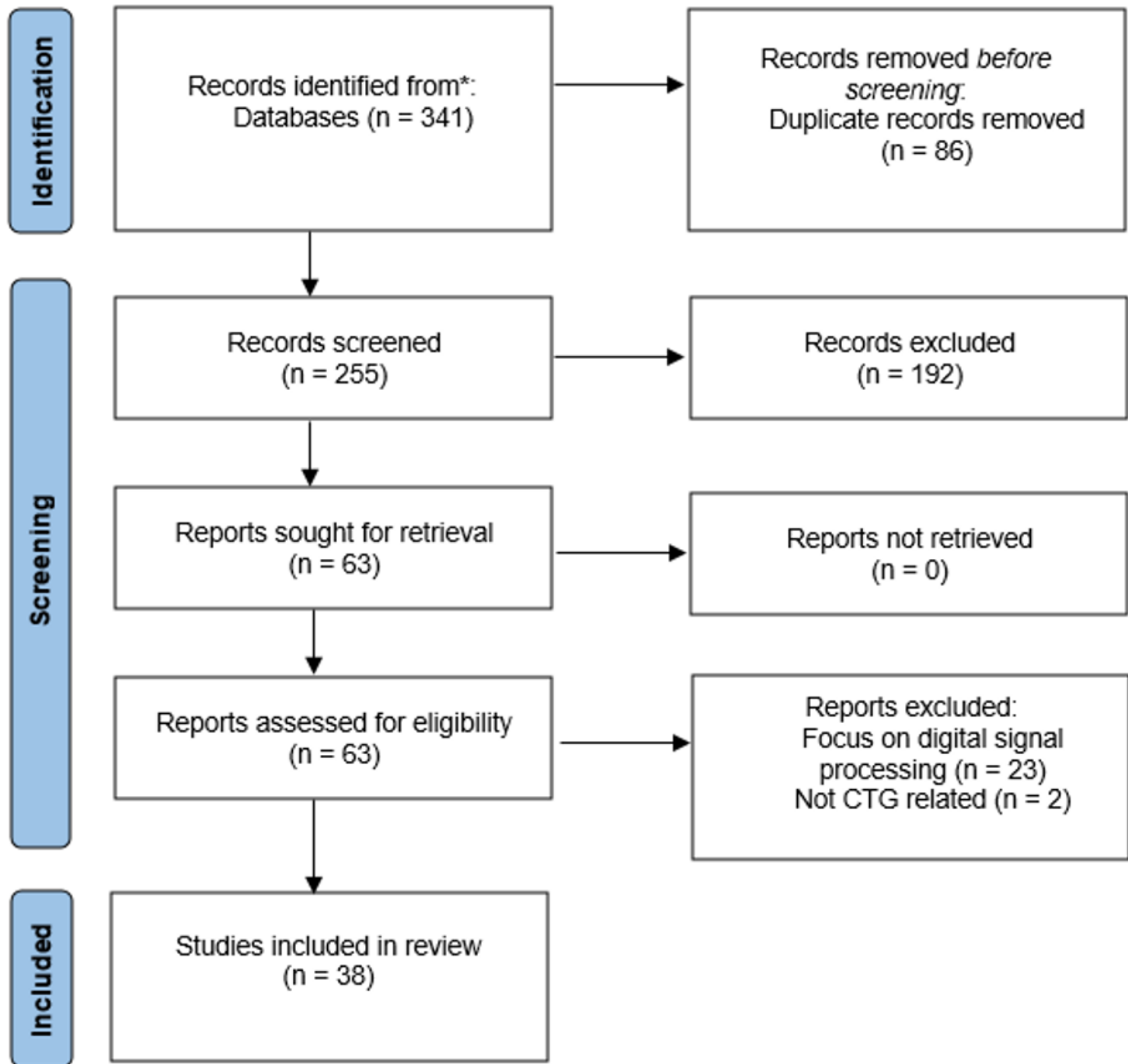


Figure 2

PRISMA flow diagram.

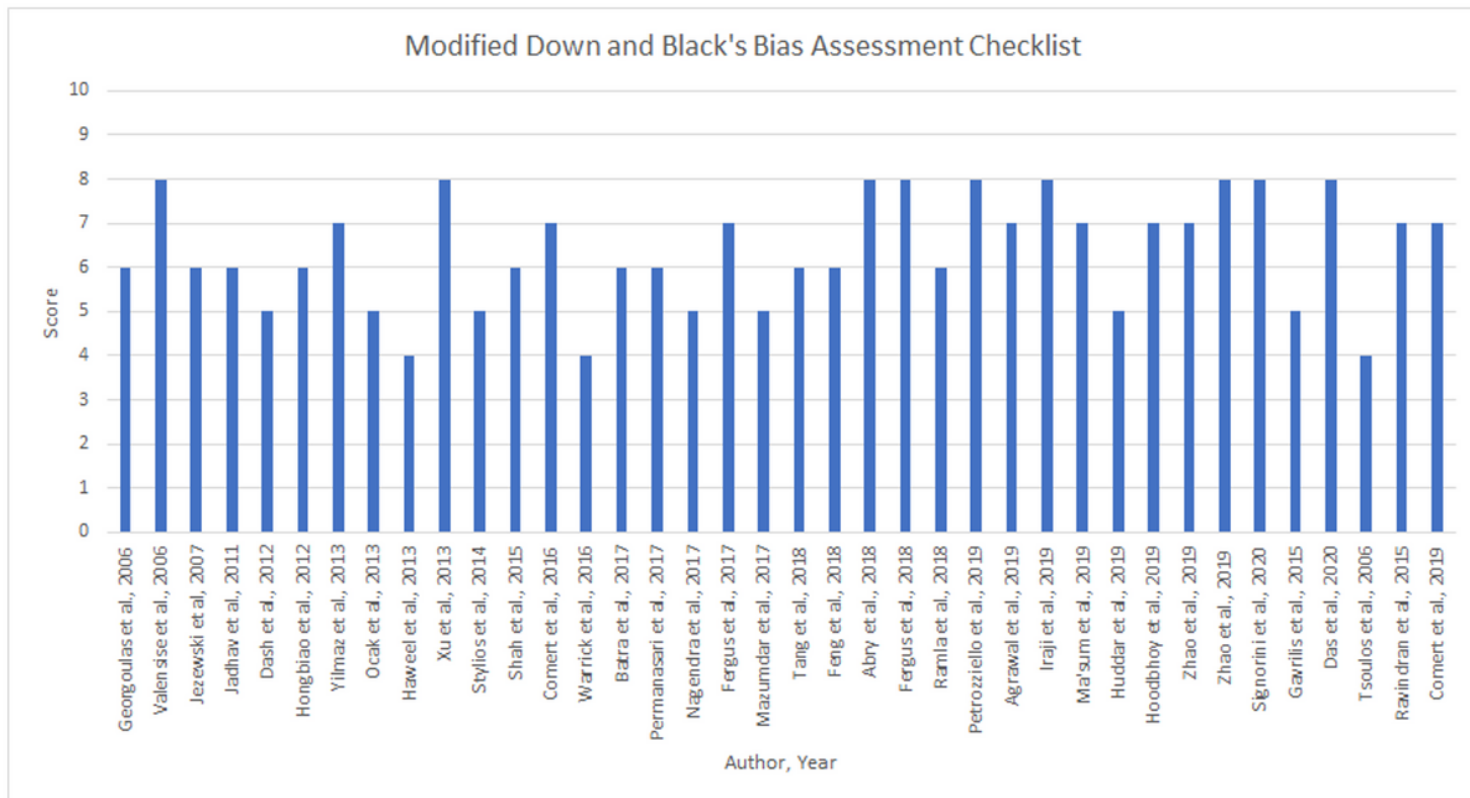


Figure 3

Modified Black and Down's score for quality of studies included in review (Maximum score achievable is 10 points).

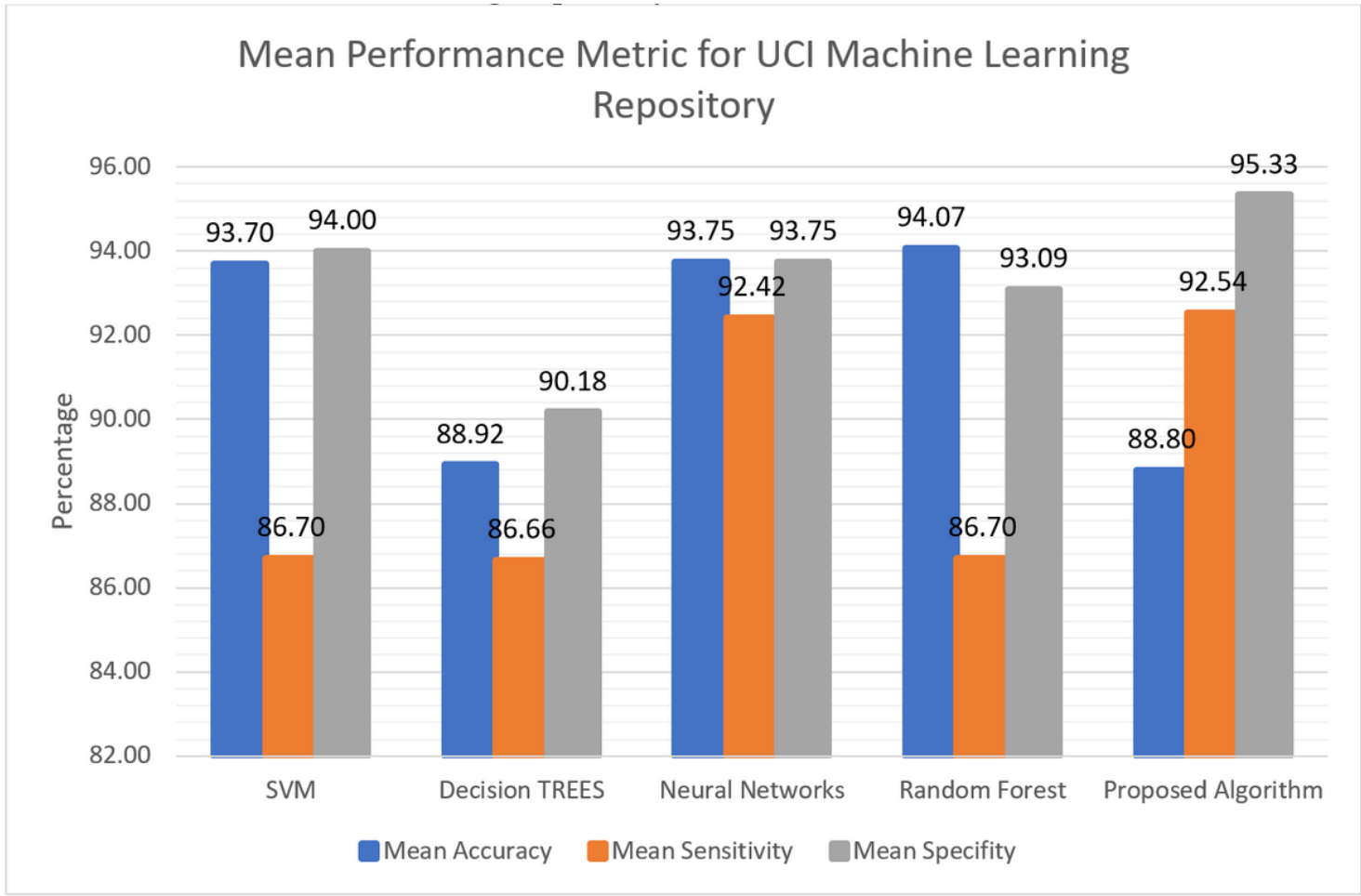


Figure 4

Mean Performance Metric for UCI Machine Learning Repository.

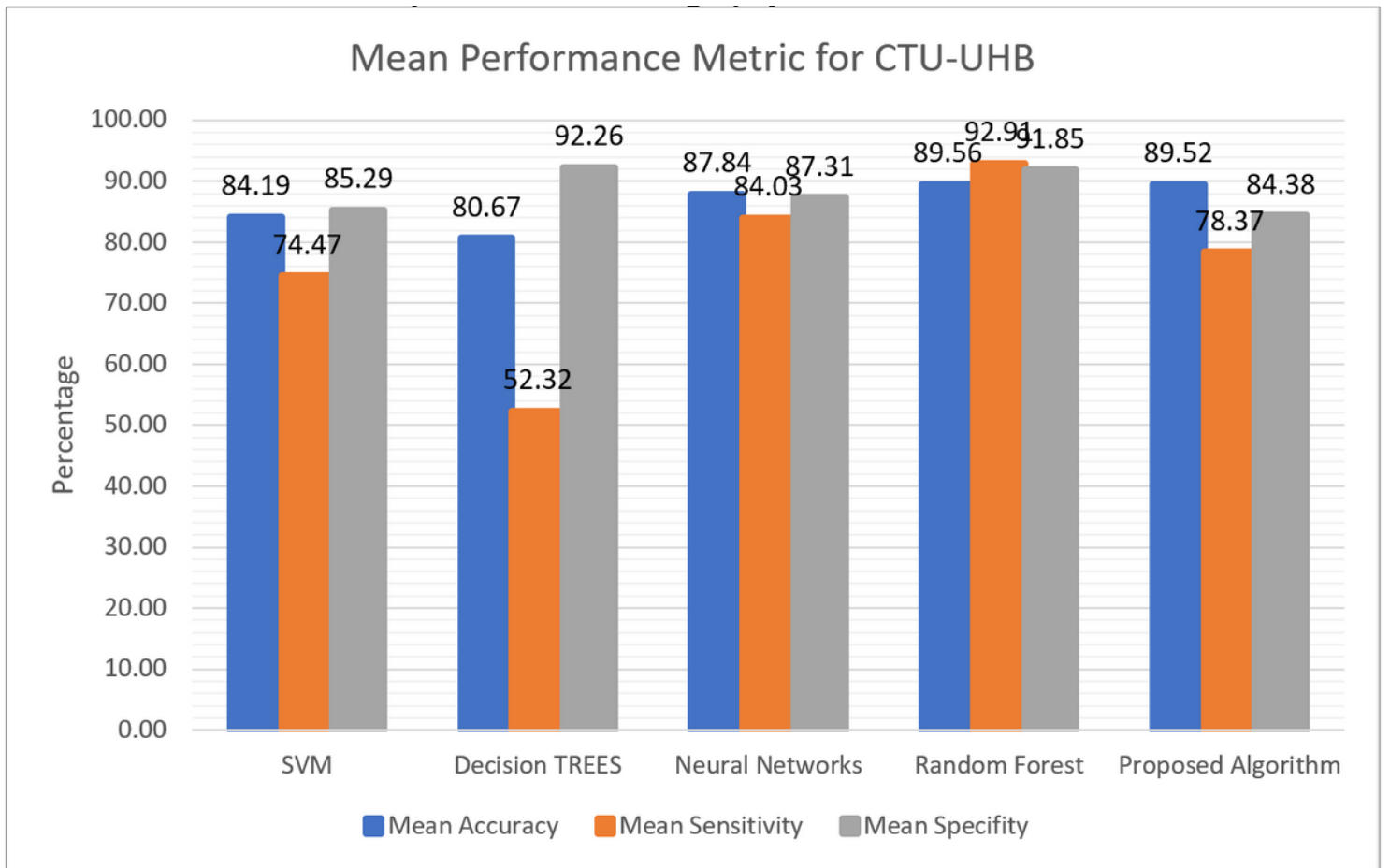


Figure 5

Mean Performance Metric for CTU-UHB Intrapartum Cardiotocography Database.