Model	Weighted Average	Precision			Recall			F1-score		
		N	S	P	N	S	P	N	S	P
Base algorithm										
DT	99%	99%	97%	98%	100%	93%	98%	99%	95%	98%
RF	98%	98%	97%	98%	100%	85%	98%	99%	90%	98%
Ensembler										
DT with Bagging	99%	99%	97%	98%	100%	93%	98%	99%	95%	98%
RF with Bagging	97%	97%	97%	98%	100%	83%	96%	99%	90%	97%
Boosting										
AdaBoost with DT	99%	99%	97%	98%	100%	93%	96%	99%	95%	97%
AdaBoost with RF	99%	99%	97%	98%	100%	92%	98%	99%	94%	98%
XGBoost	**99%**	**99%**	**96%**	**98%**	**100%**	**95%**	**98%**	**99%**	**96%**	**98%**
Resampling techniques										
Undersampling										
DT + Random Undersampling	98%	97%	97%	98%	100%	82%	96%	98%	89%	97%
DT + Condensed Nearest Neighbor (CNN)	89%	94%	**25%**	**40%**	**58%**	**58%**	98%	**72%**	**35%**	**57%**
DT + Tomac Link (TL)	99%	99%	97%	98%	100%	93%	98%	99%	95%	98%
DT + One Sided Selection (OSS)	91%	89%	96%	98%	100%	**29%**	98%	94%	**44%**	98%
DT + Edited Nearest Neighbour (ENN)	99%	99%	97%	98%	100%	93%	98%	99%	95%	98%
DT + Repeated Edited Nearest Neighbour (RENN)	99%	99%	97%	98%	100%	93%	98%	99%	95%	98%
DT + All KNN	99%	99%	97%	98%	100%	93%	98%	99%	95%	98%
DT + Neighborhood Cleaning rule	99%	99%	97%	98%	100%	93%	98%	99%	95%	98%

Table 2a: Precision, Recall, and F1-score resulted from the prediction of each class of fetal health of normal (N), suspicious (S), pathologic (P) using different classifiers, Decision Tree (DT) and Random Forest (RF) with entropy as criterion using CTG records and pattern diagnosis attributes

Model	Weighted Average	Precision			Recall			F1-score		
		N	S	P	N	S	P	N	S	P
Resampling techniques										
Undersampling										
RF + Random Undersampling	99%	99%	98%	95%	99%	94%	98%	99%	96%	97%
RF + Condensed Nearest Neighbor (CNN)	58%	97%	**0%**	19%	63%	**0%**	100%	76%	**0%**	**32%**
RF + Tomac Link (TL)	99%	99%	97%	98%	100%	93%	96%	99%	95%	97%
RF + One-Sided Selection (OSS)	90%	89%	96%	98%	100%	29%	95%	94%	44%	96%
RF + Edited Nearest Neighbour (ENN)	98%	98%	97%	98%	100%	90%	96%	99%	94%	97%
RF + Repeated Edited Nearest Neighbour (RENN)	98%	98%	97%	98%	100%	87%	95%	99%	92%	96%
RF + All KNN	98%	98%	97%	98%	100%	87%	96%	99%	92%	97%
RF + Neighborhood Cleaning rule	98%	98%	97%	98%	100%	89%	98%	99%	93%	98%
Oversampling										
DT + Random Oversampling	99%	99%	97%	98%	100%	93%	98%	99%	95%	98%
DT + SMOTE	98%	98%	97%	98%	100%	86%	96%	99%	91%	97%
DT + Borderline SMOTE 1	95%	94%	96%	98%	100%	**64%**	98%	97%	**77%**	98%
DT + Borderline SMOTE 2	94%	95%	83%	98%	98%	**68%**	98%	96%	**75%**	98%
RF + Random Oversampling	98%	99%	95%	98%	99%	93%	98%	99%	94%	98%
RF + SMOTE	98%	99%	95%	97%	99%	93%	98%	99%	94%	97%
RF + Borderline SMOTE 1	97%	97%	96%	98%	100%	82%	93%	98%	88%	95%
RF + Borderline SMOTE 2	96%	96%	93%	98%	99%	**76%**	96%	98%	84%	97%

Table 2a (Continue): Precision, Recall, and F1-score resulted from the prediction of each class of fetal health of normal (N), suspicious (S), pathologic (P) using different classifiers, Decision Tree (DT), and Random Forest (RF) with entropy as criterion using CTG records and pattern diagnosis attributes

Model	Weighted Average	Precision			Recall			F1-score		
		N	S	P	N	S	P	N	S	P
Base algorithm										
DT	91%	92%	86%	90%	98%	58%	77%	95%	70%	83%
RF	91%	92%	78%	95%	98%	63%	72%	95%	70%	82%
Ensembler										
DT with Bagging	92%	94%	81%	89%	97%	68%	84%	95%	74%	86%
RF with Bagging	91%	93%	78%	95%	98%	64%	71%	95%	72%	82%
Boosting										
AdaBoost with DT	92%	93%	80%	94%	98%	65%	79%	95%	72%	86%
AdaBoost with RF	93%	95%	83%	96%	98%	74%	82%	96%	78%	89%
XGBoost	**95%**	**96%**	**89%**	**93%**	**98%**	**81%**	**91%**	**97%**	**85%**	**92%**
Resampling techniques										
Undersampling										
DT + Random Undersampling	82%	97%	46%	70%	82%	81%	82%	89%	59%	76%
DT + Condensed Nearest Neighbor (CNN)	88%	95%	58%	88%	90%	75%	86%	93%	66%	87%
DT + Tomac Link (TL)	91%	92%	86%	90%	98%	58%	77%	95%	70%	83%
DT + One Sided Selection (OSS)	92%	92%	93%	90%	98%	62%	81%	95%	74%	85%
DT + Edited Nearest Neighbour (ENN)	91%	93%	90%	75%	97%	62%	75%	95%	73%	75%
DT + Repeated Edited Nearest Neighbour (RENN)	91%	93%	90%	75%	97%	62%	75%	95%	73%	75%
DT + All KNN	89%	92%	78%	81%	97%	56%	77%	94%	65%	79%
DT + Neighborhood Cleaning rule	91%	92%	96%	81%	98%	57%	77%	95%	72%	79%

Table 2b: Precision, Recall, and F1-score resulted from the prediction of each class of fetal health of normal (N), suspicious (S), pathologic (P) using different classifiers, Decision Tree (DT) and Random Forest (RF) with entropy as criterion using CTG records only

Model	Weighted Average	Precision			Recall			F1-score		
		N	S	P	N	S	P	N	S	P
Resampling techniques										
Undersampling										
RF + Random Undersampling	87%	98%	60%	65%	86%	87%	91%	92%	71%	76%
RF + Condensed Nearest Neighbor (CNN)	92%	95%	78%	81%	95%	73%	91%	95%	75%	86%
RF + Tomac Link (TL)	92%	93%	81%	96%	98%	64%	75%	95%	72%	84%
RF + One-Sided Selection (OSS)	92%	93%	90%	95%	98%	68%	74%	95%	73%	83%
RF + Edited Nearest Neighbour (ENN)	90%	91%	81%	88%	99%	52%	74%	95%	64%	80%
RF + Repeated Edited Nearest Neighbour (RENN)	90%	91%	81%	88%	99%	52%	74%	95%	64%	80%
RF + All KNN	91%	92%	81%	92%	99%	57%	77%	95%	67%	84%
RF + Neighborhood Cleaning rule	91%	93%	77%	89%	98%	63%	72%	95%	69%	80%
Oversampling										
DT + Random Oversampling	90%	96%	62%	84%	91%	89%	81%	94%	74%	82%
DT + SMOTE	83%	97%	55%	56%	84%	83%	81%	90%	66%	66%
DT + Borderline SMOTE 1	81%	96%	42%	91%	79%	88%	88%	87%	57%	89%
DT + Borderline SMOTE 2	80%	95%	42%	63%	83%	65%	81%	88%	51%	71%
RF + Random Oversampling	90%	98%	61%	87%	91%	89%	84%	94%	73%	86%
RF + SMOTE	90%	98%	60%	87%	91%	88%	84%	94%	71%	86%
RF + Borderline SMOTE 1	89%	97%	59%	83%	90%	82%	88%	93%	69%	85%
RF + Borderline SMOTE 2	83%	99%	44%	68%	83%	73%	96%	90%	54%	80%

Table 2b (Continue): Precision, Recall, F1-score resulted from the prediction of each class of fetal health of normal (N), suspicious (S), pathologic (P) using different classifiers, Decision Tree (DT) and Random Forest (RF) with entropy as criterion using CTG records only