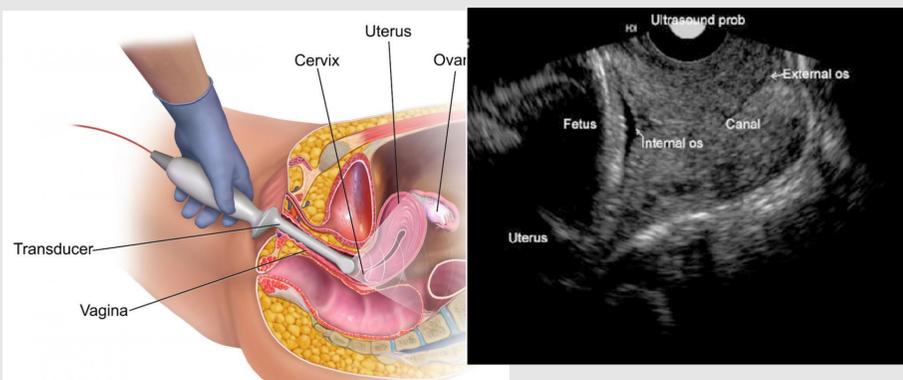


Machine learning models for predicting preterm birth

Introduction

- About 15 millions of babies are affected by preterm birth complications each year, and almost 1 million die[1].
- Children who survive a preterm birth can suffer life long disabilities, learning problems and hearing and visual problems
- Prediction of preterm birth is usually done by measuring the cervical length in the transvaginal ultrasound (US), but it does not report any information about the compression or structural and histological changes of the tissue.
- Hypothesis: the analysis of the texture of the ultrasound image can be correlated with structural changes in the cervix and have a predictive value regarding the outcome of the pregnancy.
- Approach: image processing techniques to study texture features have been used to train machine learning classifiers to predict two classes: preterm or full-term birth.



Results and discussion

Classifier	AUC(mean)	AUC(std)
GaussianNB	0.674904	0.011426
BernoulliNB	0.511534	0.017915
RandomForest_trees_150	0.742723	0.013381
MLP	0.624721	0.020881
SVM_rbf_-1	0.489556	0.048535
SVM_sigmoid_-1	0.509817	0.026982
SGD_log_8000_none	0.66059	0.012086
SGD_modified_huber_r_8000_none	0.655495	0.01008
DecisionTree	0.62037	0.028223
ExtraTrees_trees_150	0.723297	0.013771

Random Forest	
160	35
42	28

Extra Tress	
171	24
45	26

Naive Bayes	
71	124
10	60

Some features have a very high correlation therefore we remove them and keep only 58:

Classifier	AUC(mean)	AUC(std)
GaussianNB	0.773311	0.00237
RandomForest_trees_150	0.767717	0.005132
MLP	0.499872	0.016022
SVM_rbf_-1	0.5	0
SVM_sigmoid_-1	0.485863	0.032375

Naive Bayes Gaussian	
146	29
86	115

Random Forest	
122	53
60	141

Using random forest importance for feature selection (only 8 features):

Classifier	AUC(mean)	AUC(std)
GaussianNB	0.767744	0.009602
BernoulliNB	0.5	0
RandomForest_trees_150	0.762148	0.006804
MLP	0.595454	0.043759
SVM_rbf_-1	0.484285	0.019761
SVM_sigmoid_-1	0.497411	0.030671
SGD_log_8000_none	0.576382	0.011193
SGD_modified_huber_8000_none	0.570467	0.017182
DecisionTree	0.640821	0.02257

Naive Bayes Gaussian	
92	83
34	167

Random Forest	
116	59
58	143

Conclusions

- Cervical US images can help estimating the risk of preterm birth using texture analysis and machine learning
- Future work: validation with larger database and explore deep learning techniques to improve results

References

[1] Althabe, Fernando, Christopher Paul Howson, Mary Kinney, Joy Lawn, and World Health Organization. Born Too Soon: The Global Action Report on Preterm Birth, 2012.

<https://pyradiomics.readthedocs.io/en/>
<http://rasbt.github.io/mlxtend/>
<https://scikit-learn.org/stable/>

Methodology

Materials:

We use retrospective transvaginal ultrasounds acquired in the 20th week of pregnancy from 382 patients, from which 179 had a preterm birth.

Methods:

- Extract 77 texture features from a region of interest, using pyradiomics.
 - Gray Level Co-occurrence Matrix,
 - Gray Level Run Length Matrix,
 - Gray Level Dependence Matrix and Neighboring Gray Tone Difference Matrix
- Train machine learning classifiers: Gaussian naive Bayes, Random Forest, SVM, MLP
- Apply feature selection and hyperparameter selection
- Validate results: 10 fold cross validation, ROC

