Universitat Oberta
de Catalunya

uoc.edu

Machine Learning in materno-fetal ultrasound images for early detection of late-onset placental insufficiency

Master’s in Data Science
Master’s thesis
Medicine Area

January 2024

Gwendolin Herrera Carballido (author)
Xavier Paolo Burgos Artizzu (tutor)

Content

Introduction: Placental Insufficiency

BACKGROUND

- Oxygen and nutrients not sufficiently transferred to fetus via placenta
- Affects ~10% of pregnancies

CHALLENGES

- Positive predictive value of existing diagnostic method barely exceeds 50%
- Placental disease not suspected until complications occur

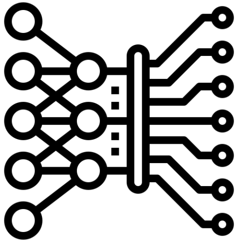
COMPLICATIONS

- Preterm labor
- Preeclampsia
- Fetal growth restriction
- Stillbirth

Introduction: Main approach

Ultrasound images
weeks 27 – 29 of pregnancy

Deep Learning



Computer Vision

Diagnosis

State of the art: Computer Vision

Histogram of Oriented Gradients (HOG)

Object Detection:

1. Extracts gradient orientations for localized object detection.
2. Robust to varying lighting conditions.

Computer Vision Standard:

1. Widely used in tasks like pedestrian detection.
2. Provides reliable feature representation.

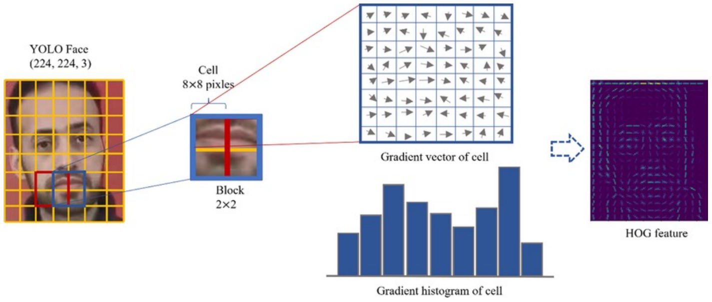
Gray-Level Co-occurrence Matrix (GLCM)

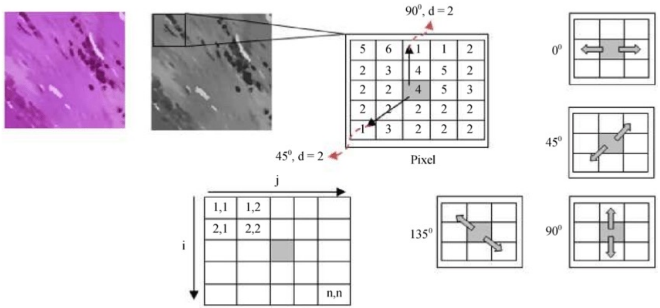
Texture Analysis:

1. Statistically analyzes pixel intensity relationships for texture patterns.
2. Captures spatial dependencies in images.

Versatile Applications:

1. Used in medical imaging for tissue characterization.
2. Applied in remote sensing for land cover classification.

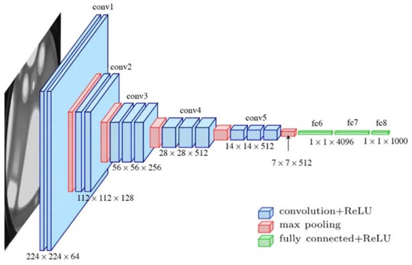




State of the art: Deep Learning (Convolutional Neural Networks)

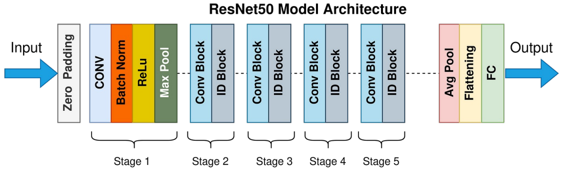
VGGNet

- ❖ Top choice for image feature extraction.
- ❖ Simple and effective architecture.
- ❖ Small 3x3 convolutional filters predominantly.
- ❖ 138 million parameters.



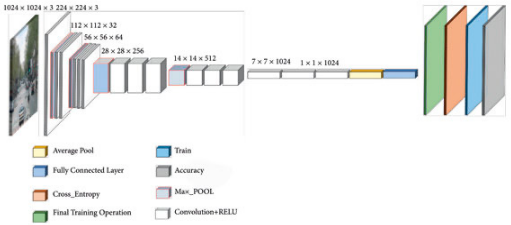
ResNet

- ❖ Introduces skip connections for deep learning.
- ❖ Addresses vanishing gradient problem in deep networks.
- ❖ Proposes residual blocks with identity and shortcut connections.
- ❖ 25 million parameters in ResNet50.



MobileNet

- ❖ Introduced by Google for lightweight deep learning.
- ❖ Significantly reduces the number of parameters.
- ❖ Suitable for real-time applications on resource-constrained devices.
- ❖ 4.2 million parameters.



Implementation: Introduction

Research team

- ❖ Images collected by the Preeclampsia and Intrauterine Growth Restriction Research Unit at the hospital Sant Joan de Deu in Barcelona.
- ❖ Data collection from September 2022 to November 2023.

Dataset

- ❖ Minimum of three ultrasound images of the placenta were acquired for each woman.
- ❖ Masks were created manually for each one of these images, to outline the position of the placenta.
- ❖ Perinatal data was collected after childbirth.

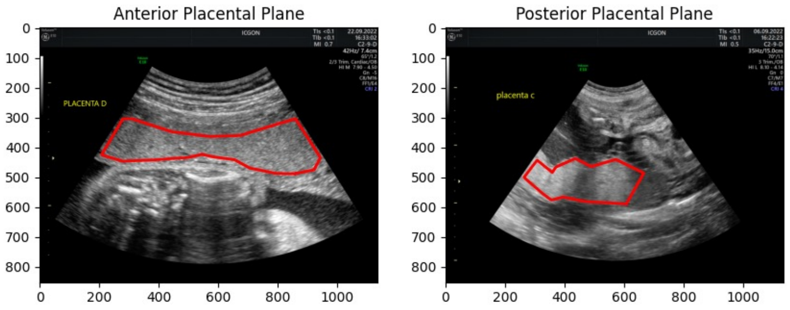
Criteria for inclusion

- ❖ Pregnant women between weeks 27 and 29 of pregnancy.
- ❖ Heightened baseline risk of placental insufficiency identified during routine second-trimester ultrasounds.
- ❖ Single fetus without congenital malformations.
- ❖ All participants were of legal age.

Implementation: Dataset (overview)

Images

450 women
3 ultrasounds per woman
1 mask per image
2 placental planes



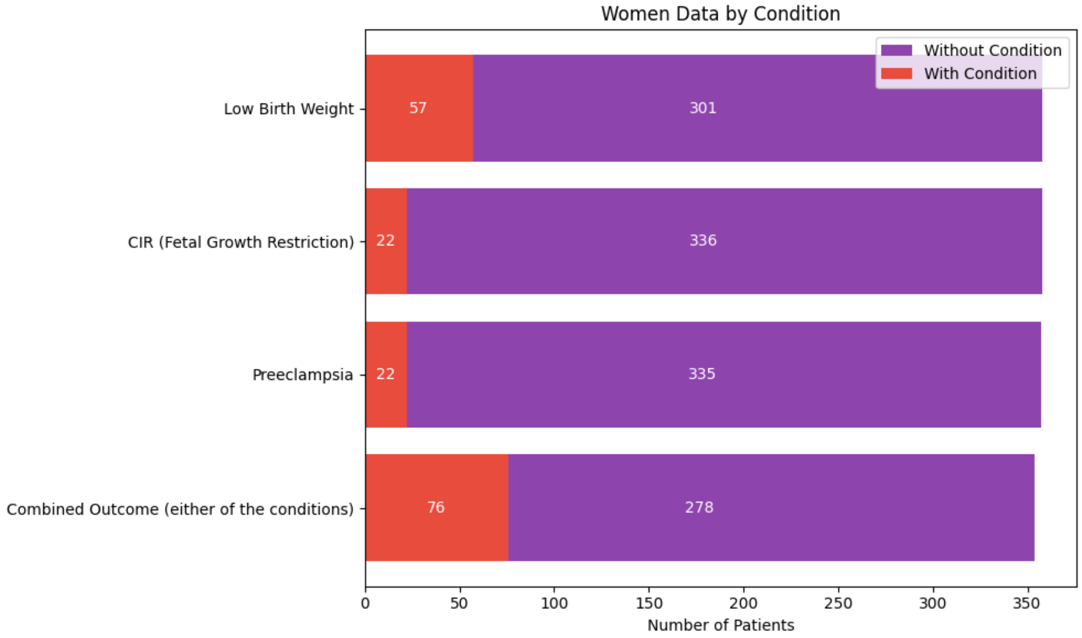
Patient database

Three diagnostic indicators

- Preeclampsia (PRE)
- Prenatal fetal growth restriction (CIR)
- Birth weight percentile < 10 (LBW)

Established after childbirth

Implementation: Dataset (characteristics)



15.9% prevalence

6.1% prevalence

6.2% prevalence

21.5% prevalence

Implementation: Data preparation (splits)

Sample size

(main focus) ▶

Condition	Prevalence	Min. test size	Sample available
Combined outcome	20%	121	354
Preeclampsia	6%	401	357
Fetal growth restriction (CIR)	6%	401	358
Low birth weight (LBW)	16%	151	358

Table 3.1: *Minimum sample size of test from Burderer's method.*

Burderer's params
Sensitivity = 0.5
Specificity = 1 – prevalence
Prevalence = according to condition
Alpha = 0.05
Beta = 0.20

Train, test, val split

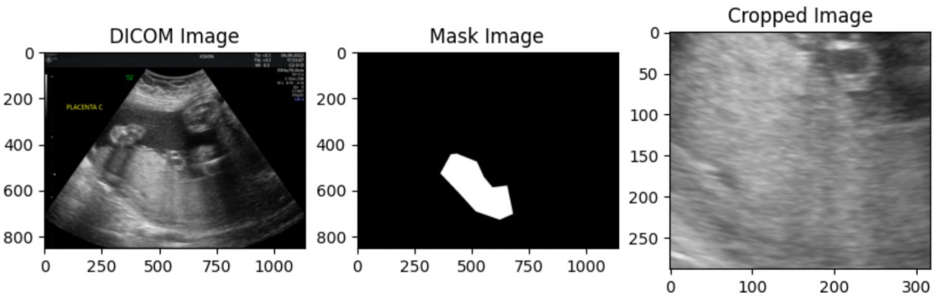
Combined outcome and **LBW** according to Burderer's method

Preeclampsia and **CIR** 70-15-15

*unique patients in each subset
*randomized division
*maintained balance ratio positive/negative

Implementation: Data preparation (images)

- Cropping to focus on placental region
- Two cropping strategies: 1x and 1.3x the mask size



Condition	Mask Size
Combined outcome	Original
Combined outcome	Expanded
Fetal growth restriction	Original
Fetal growth restriction	Expanded
Preeclampsia	Original
Preeclampsia	Expanded
Tenth birth weight percentile	Original
Tenth birth weight percentile	Expanded

Table 3.2: Groups of data

Implementation: Experimental design (Deep Learning)

Experiment	Description	Class Weights	Image Augmentation	Placental planes	Mask size
Experiment 1	Baseline models	-	-	A/P	1x
Experiment 2	Adding normalized class weights	WITH	-	A/P	1x
Experiment 3A	Soft image augmentation	WITH/WITHOUT	WITH	A/P	1x
Experiment 3B	Aggressive image augmentation	WITH/WITHOUT	WITH	A/P	1x
Experiment 4	Expanded mask + image augmentation	WITH/WITHOUT	WITH	A/P	1.3x
Experiment 5	Custom transfer learning	WITH/WITHOUT	WITH/WITHOUT	A/P	1x

Summary trials 1 - 4

Architecture	Configuration
Artificial Neural Network (ANN)	A basic ANN served as a straightforward baseline for comparison.
VGG16	Evaluated with and without transfer learning.
ResNet50	Evaluated with and without transfer learning.
MobileNet	Evaluated with and without transfer learning.
ResNet18	Evaluated with and without transfer learning.

Summary trial 5

- Investigate transfer learning using a tailored classifier.
- 1. Developed **binary brain classifier** using **VGG16** and **ResNet18** architectures on 12,400 **ultrasound images** from **fetal planes** (AUC: 0.9975, F1: 0.98/0.99).
- 2. Replicated brain classifier architectures and weights for placental ultrasound.

Implementation: Experimental design (Computer Vision)

Feature extraction

Histogram of Oriented Gradients (HOG)
- ❖ Resizing
- ❖ Window size
- ❖ Block size
- ❖ Block stride
- ❖ Cell size
- ❖ Number of bins

Gray Level Co-occurrence Matrix (GLCM)
- ❖ Resizing
- ❖ Dissimilarity
- ❖ Correlation
- ❖ Homogeneity
- ❖ Contrast
- ❖ Angular Second Moment
- ❖ Energy

Classifiers
- ❖ Support Vector Classifier (SVC)
- ❖ XGBoost
- ❖ Logistic Regression (LR)

Implementation: Experiments (Computer Vision)

Experiment	Description	Class Weights	Image Augmentation	Placental planes	Mask size
Experiment 1	Baseline models	-	-	A/P	1x
Experiment 2	Adding normalized class weights	WITH	-	A/P	1x
Experiment 3	Image augmentation	WITH/WITHOUT	WITH	A/P	1x
Experiment 4	Expanded mask + image augmentation	WITH/WITHOUT	WITH	A/P	1.3x
Experiment 5	Enhanced feature extraction	WITH/WITHOUT	WITH/WITHOUT	A/P	1x

Results: Evaluation metrics

Focus: maximum **sensitivity** achieved at the specific points on the ROC curve where **specificity** equals **1 - prevalence**

Metric	Description	Formula
Accuracy	Overall correctness of the model	$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$
Sensitivity (Recall)	Proportion of actual positives correctly predicted	$\text{Sensitivity} = \frac{TP}{TP+FN}$
Specificity	Proportion of actual negatives correctly predicted	$\text{Specificity} = \frac{TN}{TN+FP}$
Positive Predictive Value (Precision)	Proportion of predicted positives correctly predicted	$\text{Precision} = \frac{TP}{TP+FP}$
Negative Predictive Value	Proportion of predicted negatives correctly predicted	$\text{NPV} = \frac{TN}{TN+FN}$
Positive Likelihood Ratio	Ratio of the probability of a true positive to the probability of a false positive	$\text{PLR} = \frac{\text{Sensitivity}}{1-\text{Specificity}}$
Negative Likelihood Ratio	Ratio of the probability of a false negative to the probability of a true negative	$\text{NLR} = \frac{1-\text{Sensitivity}}{\text{Specificity}}$

Results: before optimization

Deep Learning

TRIAL	Best Architecture	Max Sensitivity
1	ANN	0.26
2	ResNet18 [TL, CW]	0.32
3A	**ResNet18 [TL, AUG]**	**0.36**
3B	**ResNet18 [TL, AUG]**	**0.36**
4	ResNet18 [TL, CW, AUG]	0.32
5	**ResNet18 [TL, CW, AUG]**	**0.36**

Table 4.1: Best Architecture and Max Sensitivity by Trial. 1 - Baseline; 2 - Class Weights; 3A - Soft image augmentation; 3B - Aggressive image augmentation; 4 - Expanded mask; 5 - Transfer learning.

Computer Vision

TRIAL	Best Architecture	Max Sensitivity
1	HOG + LR	0.26
2	HOG + SVC [CW]	0.25
3	HOG + XGBOOST [CW, AUG]	0.27
4	GLCM + SVC [CW, AUG]	0.29
5	**GLCM + SVC**	**0.32**

Table 4.4: Best Architecture and Max Sensitivity by Trial

Architecture to optimize: ResNet18, soft image augmentation, transfer learning, no class weights, 1:1 mask.

Architecture to optimize: GLCM with enhanced feature extraction and SVC, no augmentation, no class weights, 1:1 mask.

Results: optimization techniques

Deep Learning

- Retrain on anterior placenta images.
- Tune hyperparameters of best model (original or anterior planes) with Optuna.
- Automated process for optimal parameters (learning rate, number of layers, number of neurons, dropout rate)
- Iteratively refine parameters across 10 trials.
- Goal: maximize Sensitivity, maintain minimum Specificity.

Computer Vision

- Optimize using Support Vector Classifier (SVC).
- Employed two strategies: with and without augmentation.
- Retrain on anterior placenta images.
- Chose superior strategy for fine-tuning.
- Used GridSearchCV for hyperparameter tuning.
- Exhaustive search for effective parameter combination (C, gamma, kernel, class_weight)

Results: Overall Best Performers

Criteria	C3	LBW	PRE	CIR
Best Architecture	ResNet18 [TL, AUG]	ANN	GLCM + SVC [CW]	ANN [CW]
Placental Plane	Both	Both	Anterior	Both
Sensitivity	0.36	0.29	**0.56**	0.33
AUC	0.60	0.53	0.76	0.55
F1-Score Pos Class	0.33	0.27	0.53	0.29
Accuracy	0.69	0.77	0.96	0.91
Specificity	0.79	0.86	0.99	0.96
PPV	0.31	0.27	0.67	0.33
NPV	0.82	0.86	0.97	0.94
PLR	1.64	1.96	34.67	6.42
NLR	0.83	0.84	0.56	0.78

Table 4.6: Best Classifiers by Criteria

- Minimum requirements are met only by **Preeclampsia** classifier
- Dataset size not enough for strong statistical significance

Results: Overall Best Performers (Combined Outcome)

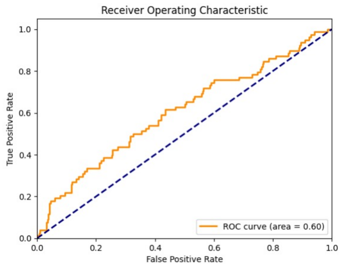


Figure 4.1: C3: ROC curve

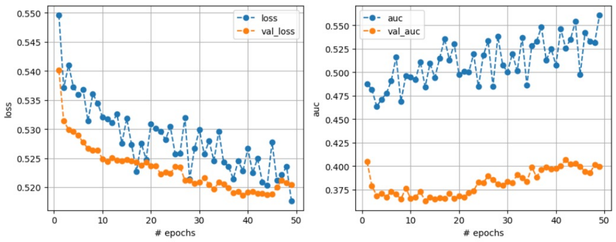


Figure 4.2: C3: training

Architecture: ResNet18 with image augmentation and transfer learning from ImageNet

- Modest performance.
- Max sensitivity at 0.36, indicating limitations.
- F1-Score: 0.33, moderate balance in predictions.
- PPV: 0.31, revealing considerable false positives.
- NPV: 0.82, reliable negative predictions.
- PLR: 1.64, NLR: 0.83, highlighting model limitations.
- Room for improvement; further investigation recommended.

Results: Overall Best Performers (Tenth Percentile Birth Weight)

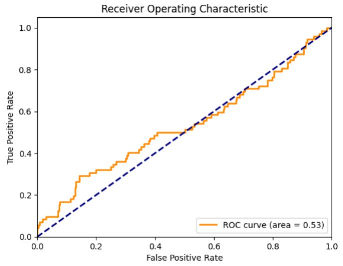


Figure 4.3: LBW: ROC curve

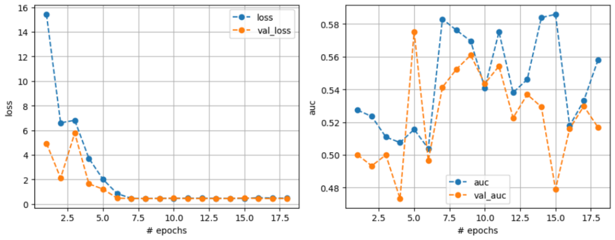


Figure 4.4: LBW: training

Architecture: Custom ANN

- Modest outcomes with maximum sensitivity of 0.29.
- F1-Score: 0.27, moderate balance in predictions.
- Low PPV at 0.31 indicates false positives.
- PLR: 1.96, NLR: 0.84, indicating diagnostic limitations.
- Room for improvement; further investigation recommended.

Results: Overall Best Performers (Fetal Growth Restriction)

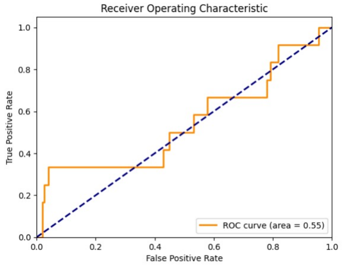


Figure 4.5: CIR: ROC curve

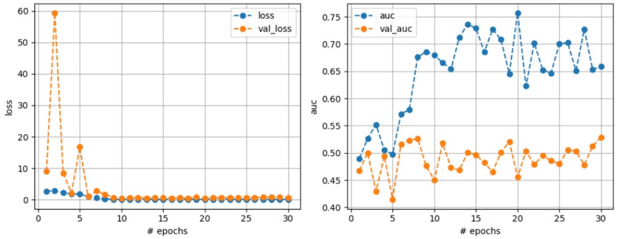


Figure 4.6: CIR: training

Architecture: Custom ANN with class weights

- Modest outcomes with maximum sensitivity of 0.33.
- F1-Score: 0.29, moderate balance in predictions.
- Poor ability to identify true positive cases.

Results: Overall Best Performers (Preeclampsia)

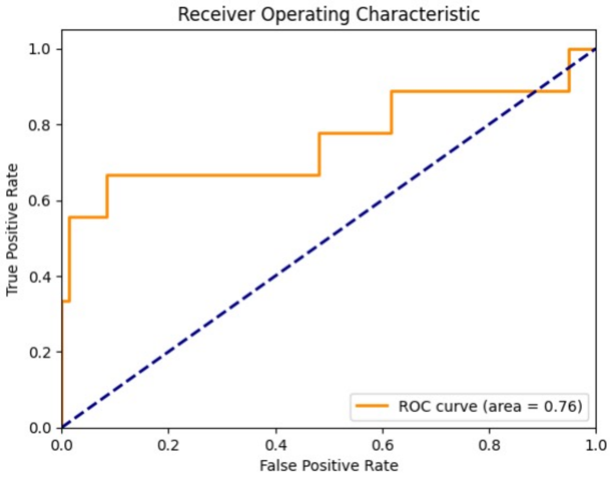


Figure 4.7: PRE: ROC curve

Architecture: GLCM + SVC with class weights

- PLR: 34.67, strong likelihood for preeclampsia identification.
- Maximum sensitivity of 0.56.
- Fairly high PPV at 0.67, encouraging for clinical use.
- Room for improvement to reduce false positive rates.
- Caution: Low prevalence (6%), Burderer's method suggests a minimum sample size of 401 for statistical significance.
- Dataset falls short; reevaluation recommended with dataset expansion.

Conclusions and Future Work

Conclusions

Clinical Impact if good results:

- Accurate models crucial for identifying placental insufficiency.
- High-efficacy models improve maternal-fetal healthcare.
- At-risk women benefit from enhanced monitoring.

Study Outcomes:

- Deep Learning and Computer Vision fall short for clinical tools.
- Preeclampsia model shows moderate strength.
- Insufficient sample size hinders statistical significance.

Challenges:

- Only combination of criteria and low birth weight met size requirements.
- Anterior plane dataset size insufficient for statistical significance.

Summary:

- Models show promise but limited as standalone tools.

Future Work

Enhancements Needed:

- Sensitivity improvements crucial for clinical use.
- Exploration of alternative strategies beyond fine-tuning.

Anterior Plane Consideration:

- Training on anterior planes enhances preeclampsia prediction.
- Acquiring more anterior plane images may improve predictive power.

Image Quality Consideration:

- Ultrasound images may have inherent limitations.
- Improving image quality or exploring alternative medical imaging methods may enhance predictive models.

Holistic Exploration:

- Adjusting expectations and strategies for more extensive exploration.
- In-depth assessment of predictive models' capabilities in various clinical scenarios.

Universitat Oberta
de Catalunya

Thank you for your attention!

UOC.universitat
@UOCuniversidad
UOCuniversitat