# Segmentation of Cerebral Tissues in Human Brain MRIs with uncertainty

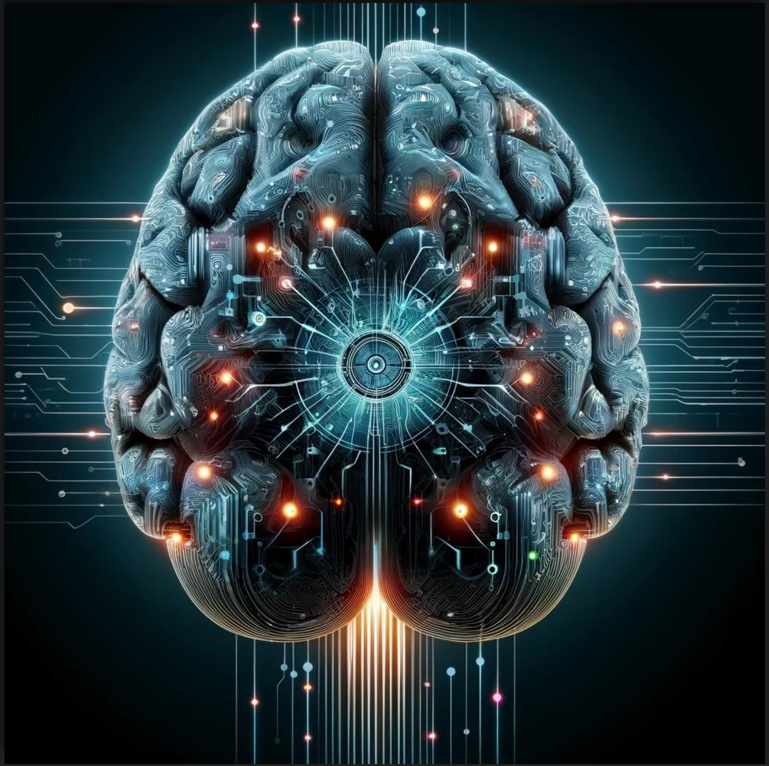
**Yacine BOUREGHDA**
**Supervised by M. Nicolas BOUTRY**

*Seminar Presentation, January 14th 2025*

LABORATOIRE DE RECHERCHE DE L'EPITA

# Segmentation In Medical Imaging with Deep Learning

- **Segmentation of brain MRI's** using deep learning algorithms across various datasets
- **Improved Interpretation**: Quantifying the uncertainty in segmentations produced by Neural Networks enhances interpretation for medical teams
- **Uncertainty in Deep Learning**: Addressing uncertainty remains a crucial and unresolved challenge in the field
- **State of the Art**: The leading approaches to quantify uncertainty include two main methods: Monte Carlo Dropout and Deep Ensembles



LRE
LABORATOIRE DE RECHERCHE DE L'EPITA

# Presentation Outline

- State of the Art and Uncertainty Metrics
- Classification Task on *MNIST*
- iSeg-2017: 6-month Infant Brain MRI Segmentation
- Experimentations
- Related Work : Achievements for this Semester
- Future Directions for the Project

# Quantifying Uncertainty in Deep Learning

| Prediction | Prediction 1 | Prediction 2 | Prediction 3 | Prediction 4 | Prediction 5 | Prediction 6 | Prediction 7 |
|---|---|---|---|---|---|---|---|
| Label 0 | 0.0993 | 0.1861 | 0.0651 | 0.0691 | 0.0480 | 0.1379 | 0.1511 |
| Label 1 | 0.9007 | 0.8139 | 0.9349 | 0.9309 | 0.9520 | 0.8621 | 0.8489 |

**Softmax Output for each prediction :** Represents the probability of belonging to class 1 or class 0

For this distribution, we have the following results for the mean and standard deviation :

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{0.9007 + 0.8139 + 0.9349 + 0.9309 + 0.9520 + 0.8621 + 0.8489}{7} = 0.8919$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \approx 0.0467$$

# State of the art : Quantify uncertainties with Deep Ensembles

***Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles***, Balaji Lakshminarayanan Alexander Pritzel Charles Blundell

## Overview :

- Build and train a set of models to perform several predictions with each network
- Estimate the final prediction by aggregating predictions and computing the mean prediction of the ensemble
- Estimate uncertainty by calculating different metric, such as standard deviation from the mean

## Key Steps :

- **Train five models** with different initializations
- **Assign a unique seed to each model** to control the variations in initialization.
- **Train each U-Net independently** on different subsets of the training dataset

**Algorithm 1** Pseudocode of the training procedure for our method

```
▷ Let each neural network parametrize a distribution over the outputs, i.e. p_θ(y|x). Use a proper
  scoring rule as the training criterion ℓ(θ, x, y). Recommended default values are M = 5 and
  ε = 1% of the input range of the corresponding dimension (e.g 2.55 if input range is [0,255]).
Initialize θ_1, θ_2, ..., θ_M randomly
for m = 1 : M do                                        ▷ train networks independently in parallel
  Sample data point n_m randomly for each net           ▷ single n_m for clarity, minibatch in practice
  Generate adversarial example using x'_{n_m} = x_{n_m} + ε sign(∇_{x_{n_m}} ℓ(θ_m, x_{n_m}, y_{n_m}))
  Minimize ℓ(θ_m, x_{n_m}, y_{n_m}) + ℓ(θ_m, x'_{n_m}, y_{n_m}) w.r.t. θ_m       ▷ adversarial training (optional)
```

***Fig 1 : Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles***, Balaji Lakshminarayanan Alexander Pritzel Charles Blundell [[1612.01474] Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles]

# State of the art : Quantify uncertainties with Monte Carlo Dropout

***Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning***, Yarin Gal Zoubin Ghahramani

**Overview :**

- ❖ **Dropout Regularization**: This technique randomly deactivates a subset of neurons during each forward pass to reduce overfitting in neural networks
- ❖ **Bayesian Approximation**: By treating each forward pass as a sample from a Bayesian posterior distribution, the method allows for uncertainty quantification by applying dropout both during training and testing.
- ❖ **Predictive Distribution**: To capture the model's uncertainty, multiple forward passes ( between 30 and 100) are performed for each input, generating a distribution of predictions

**Key Steps :**

- ❖ **Training**: The model is trained with dropout enabled, often at a rate of around 0.4, to promote robust feature learning
- ❖ **Testing/Inference**: During inference, dropout remains active, and several forward passes (e.g., 100) are executed to gather a range of predictions
- ❖ **Aggregation**: Finally, the mean and variance of these predictions are calculated, providing not only the expected output but also a measure of uncertainty associated with the predictions

# Combining Deep Ensembles and Monte Carlo Dropout (MCD)

## Why Combine Deep Ensembles and MCD ?

**Monte Carlo Dropout (MCD) :**

- **Advantages:**
  - Efficient : Generates many predictions from a single trained model
- **Limitations:**
  - Limited diversity : Predictions tend to be similar, reducing the quality of uncertainty estimation

**Deep Ensembles :**

- **Advantages :**
  - Captures a broader and better range of predictive differences.
- **Limitations :**
  - Computationally expensive : Requires training multiple models
  - Limited predictions: Typically only 5 predictions, insufficient for a proper distribution

# Combining Deep Ensembles and Monte Carlo Dropout (MCD)

## Overview of the Combined Method

- Train **5 models** independently as in Deep Ensembles
- For each model, perform **20 stochastic predictions** with Dropout activated
- Combine the predictions to create a **distribution of 100 predictions**
- **Goal:** Capture greater diversity in predictions and establish a more robust uncertainty estimation

## Key Idea of the Combined Approach

- **Leverage the strengths** of both methods :
  - **Deep Ensembles** provide diverse predictive distributions
  - **MCD** generates a large number of predictions for each model
- **Result :**
  - **Improved diversity** in predictions from the ensemble
  - **Higher quality uncertainty estimation** with a more complete predictive distribution

LRE
LABORATOIRE DE RECHERCHE DE L'EPITA

# Quantifying Uncertainty in Deep Learning (1/2)

1. **Variation ratios**
   - For each stochastic forward pass $t \in \{1; T\}$, compute label from softmax probabilities
   - $c^*$: most frequent label over the $T$ passes, with frequency $f_{\boldsymbol{x}}^{c^*}$
   - Compute variation-ratio var-ratio$[\boldsymbol{x}] = 1 - \frac{f_{\boldsymbol{x}}^{c^*}}{T}$
     $\Rightarrow$ **Epistemic uncertainty**
2. **Predictive entropy**: captures the average amount of information contained in the predictive distribution.

$$\hat{\mathcal{H}}[y|x, \mathcal{D}_{train}] = -\sum_c \left(\frac{1}{T}\sum_t p(y = c|x, \hat{w}_t)\right) \log\left(\frac{1}{T}\sum_t p(y = c|x, \hat{w}_t)\right)$$

   $\Rightarrow$ **Aleatoric uncertainty**

3. **Mutual information** : maximise the mutual informations are points on which the model is uncertain on average

$$\hat{\mathcal{I}}[y, w|x, \mathcal{D}_{train}] = \hat{\mathcal{H}}[y|x, \mathcal{D}_{train}] - \frac{1}{T}\sum_{c,t} p(y = c|x, \hat{w}_t) \log p(y = c|x, \hat{w}_t)$$

   $\Rightarrow$ **Epistemic uncertainty**

LRE
LABORATOIRE DE RECHERCHE DE L'EPITA

# Quantifying Uncertainty in Deep Learning (2/2)

## Example of variance ratio :

If a pixel is classified as {1,1,1,1,1,0} over 6 passes,

- The most frequent class is $c* = 1$
- The frequency of $c*$ is $f = 5$

The variance_ratio is calculated as :

$$\text{var\_ratio} = 1 - \frac{f_{c^*}}{T} = 1 - \frac{5}{6} = 0.1667$$

If a pixel is classified as {1,0,1,0,1,0} over 6 passes,

- Both classes $c* = 0$ and $c* = 1$ appear with the same frequency
- The frequency of the most frequent class is $f = 3$

The variance_ratio is calculated as :

$$\text{var\_ratio} = 1 - \frac{f_{c^*}}{T} = 1 - \frac{3}{6} = 0.5$$

## Example of predictive entropy :

If a pixel is predicted with the probabilities : {0.9,0.9,0.9,0.9,0.9,0.9} over 6 passes,

The average predicted probability for class 1 is :

$\bar{p}_1 = 0.9$

The entropy is:

$$H = -\left(\bar{p}_1 \log(\bar{p}_1) + (1 - \bar{p}_1) \log(1 - \bar{p}_1)\right)$$

Substituting $\bar{p}_1 = 0.9$:

$$H = -\left(0.9 \log(0.9) + 0.1 \log(0.1)\right) = 0.2715$$

If a pixel is predicted with the probabilities : {0.5,0.5,0.5,0.5,0.5,0.5} over 6 passes,

The average predicted probability for class 1 is :

$\bar{p}_1 = 0.5$

The entropy is:

$$H = -\left(\bar{p}_1 \log(\bar{p}_1) + (1 - \bar{p}_1) \log(1 - \bar{p}_1)\right)$$

Substituting $\bar{p}_1 = 0.5$:

$$H = -\left(0.5 \log(0.5) + 0.5 \log(0.5)\right) = 0.693$$

# Classification Task on MNIST

- **Convolutional Layers :**
  - 6 channels, kernel size 5, padding 2, ReLU activation
  - Max pooling, kernel size 2
  - 16 channels, kernel size 5, ReLU activation
  - Max pooling, kernel size 2
- **Fully-Connected Layers :**
  - Dropout p = 0.25
  - 120 units, ReLU activation
  - Dropout p = 0.5
  - 10 output units (one per digit class)

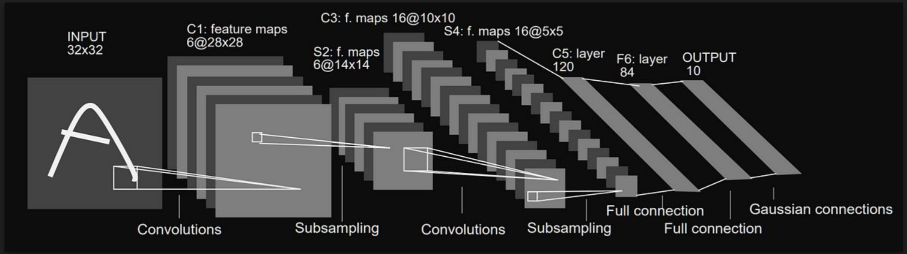



***Fig 2 :*** Architecture of AlexNet for *MNIST* classification

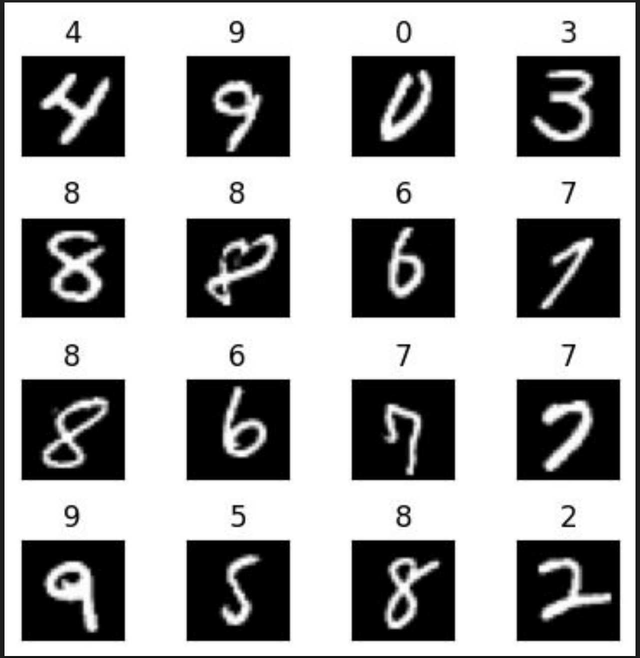



***Fig 3 :*** Sample of digits with both clear and unclear representations from the *MNIST* dataset

# Classification Task on MNIST (1/3)

## Monte Carlo Dropout Method

| Metric | Value |
|---|---|
| Accuracy | 0.9923 |
| Recall | 0.9923 |
| F1 Score | 0.9922 |

***Fig 4 :*** Performance metrics of the model evaluated on the *MNIST* test set

## Deep Ensembles Method

| Model | Accuracy | Recall | F1-Score |
|---|---|---|---|
| **Model 1** | 0.9890 | 0.9890 | 0.9890 |
| **Model 2** | 0.9902 | 0.9902 | 0.9902 |
| **Model 3** | 0.9909 | 0.9909 | 0.9909 |
| **Model 4** | 0.9745 | 0.9745 | 0.9746 |
| **Model 5** | 0.9862 | 0.9862 | 0.9862 |

***Fig 5 :*** Performance metrics (Accuracy, Recall, F1-Score) for the five independently trained models in the ensemble, evaluated on the *MNIST* test set

- Each model was initialized with a different random seed to ensure diversity in predictions

LRE
LABORATOIRE DE RECHERCHE DE L'EPITA

# Classification Task on MNIST (2/3)

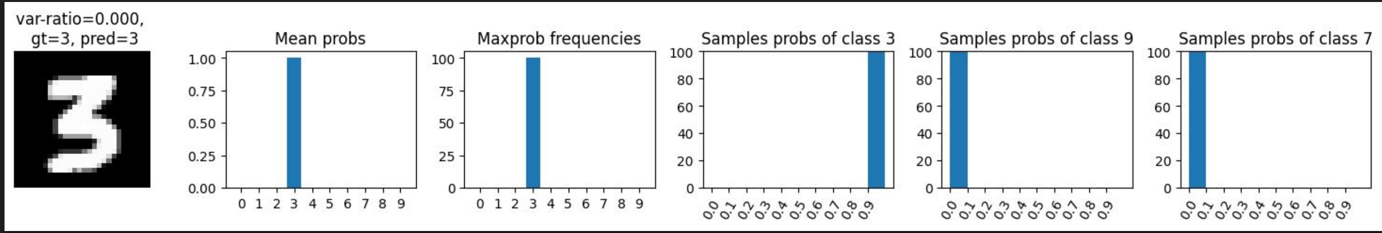



*Fig 6 :* Prediction and uncertainty visualization for a correctly classified *MNIST* digit with *MCD* method. The model predicted **3** while the true label was **3**

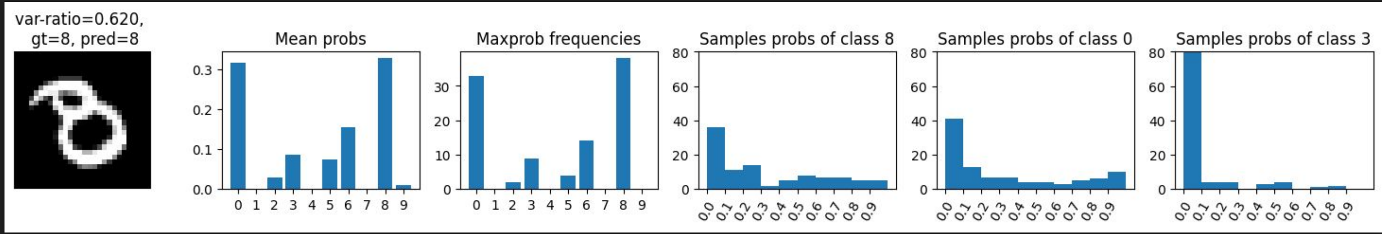



*Fig 7 :* Prediction and uncertainty visualization for a correctly classified *MNIST* digit with *MCD* method. The model predicted **8** while the true label was **8**

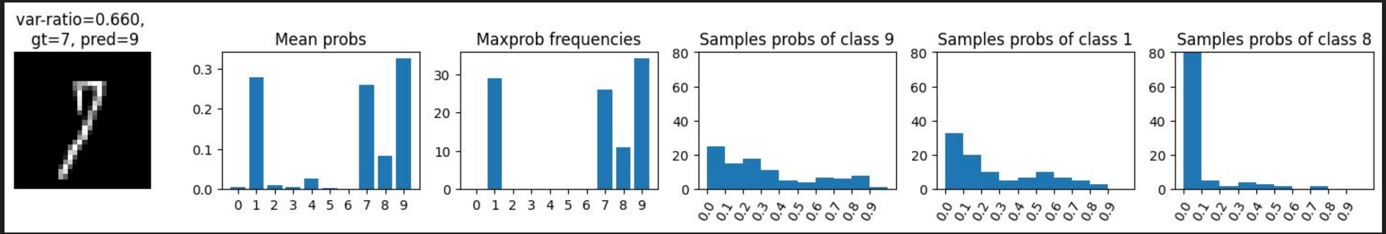



*Fig 8 :* Prediction and uncertainty visualization for a misclassified *MNIST* digit with *MCD* method. The model predicted **9** while the true label was **7**

# Classification Task on MNIST (3/3)

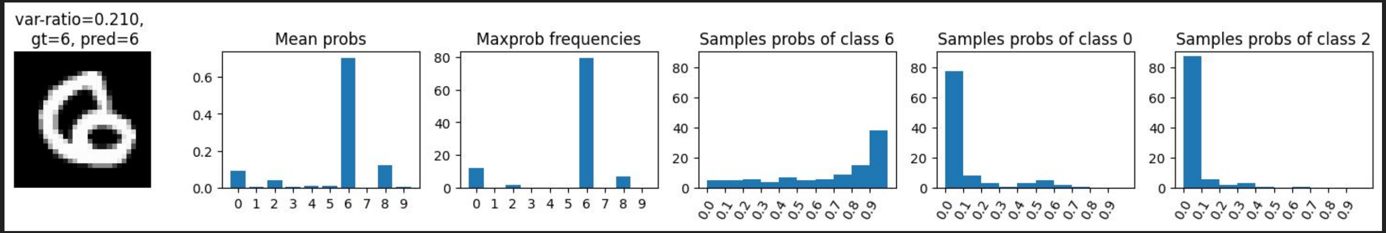



***Fig 9 :*** Prediction and uncertainty visualization for a correctly classified *MNIST* digit with *MCD* method. The model predicted **6** while the true label was **6**

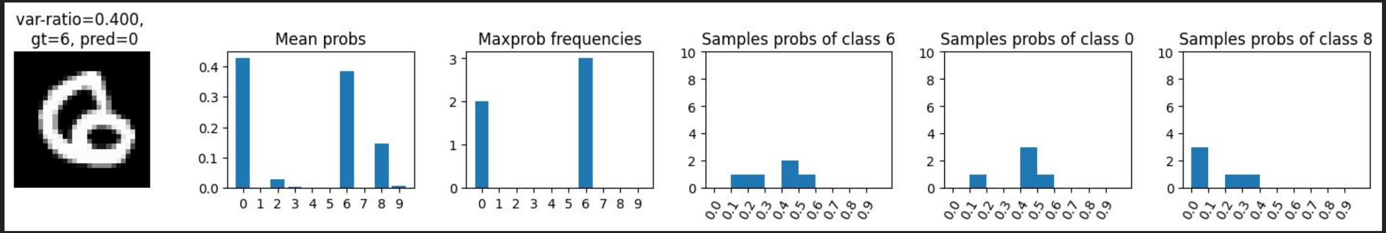



***Fig 10 :*** Prediction and uncertainty visualization for a misclassified *MNIST* digit with *Deep Ensembles* method. The model predicted **0** while the true label was **6**

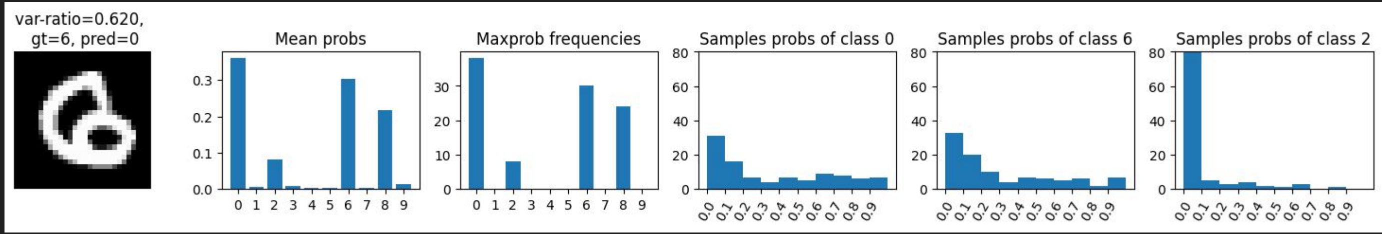



***Fig 11 :*** Prediction and uncertainty visualization for a misclassified *MNIST* digit with *Hybrid method*. The model predicted **0** while the true label was **6**

# iSeg-2017 : 6-month infant brain MRI Segmentation

- iSeg-2017 challenge focuses on comparing semi-automatic algorithms for segmenting 6-month infant brain MRIs using T1 and T2 images
- Critical for studying the dynamic first year of postnatal human brain development and associated cognitive and motor functions
- Intense phase at 6 months presents the lowest tissue contrast, posing significant challenges for accurate segmentation
- Engages researchers to develop and test automatic segmentation algorithms for white matter, gray matter, and cerebrospinal fluid

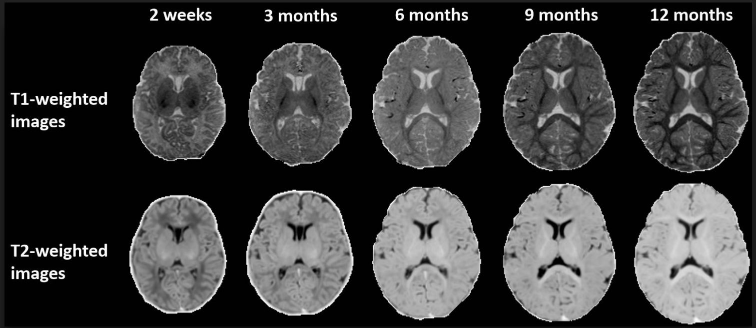



***Fig 12 :*** MIICCAI Grand Challenge on iSeg-2017, 6-Month infant Brain MRI Segmentation, [iSeg-2017]

# Overview of the U-Net Utilized for Segmentation Tasks

**Segmentation Overview :**

- ❖ **Segmentation of Regions** : White matter and gray matter are grouped in relation to other tissues
- ❖ **2D Slices**: Working with cuts along the z-axis of the brain
- ❖ **Slice Filtering**: Area is calculated ; only slices with an area greater than 100 are retained

**U-Net Model :**

- ❖ **Architecture** : 39 layers
- ❖ **Parameters** : 485,885 parameters
- ❖ **Dropout** : Set at 0.5 between the 4th and 5th convolutional layers in the encoding phase
- ❖ **Training** : Utilized the Adam optimizer with 30 epochs and a batch size of 16

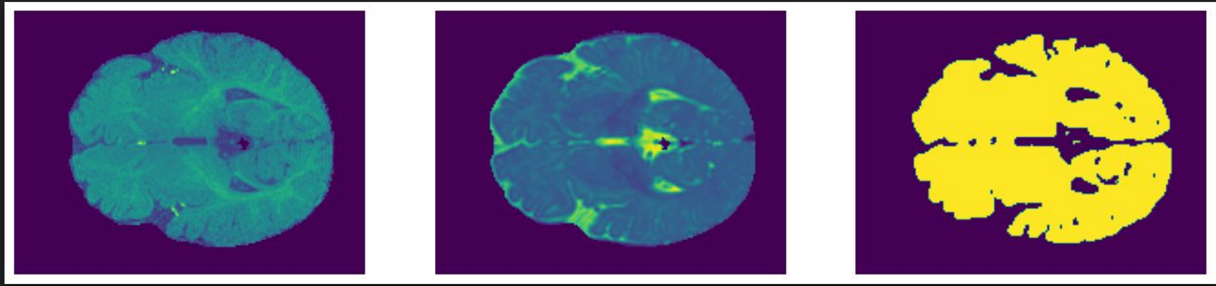


***Fig 13 :*** T1 / T2 / Region to Segment for Slice 128 for patient 0 from the training set

# Summary of Network Performance on the Test Set (X_test)

**Model Performance on Test Set :**

| Metric | Value |
|---|---|
| Loss | 0.095 |
| Accuracy | 0.9886 |
| Dice score | 0.9249 |
| Precision | 0.9121 |
| Recall | 0.939 |

***Fig 14 :*** Segmentation of slice 128 vs ground truth for the patient 0 from the training set

# Experimental Results for Monte Carlo Dropout Method

**Training Parameters for Monte Carlo Dropout :**

- **Iterations** : 100 iterations
- **Dropout** : Set to 0.5 between the 4th and 5th convolutional layers in the encoding phase
- **Pixel-wise Prediction** : Average of the predicted softmax values calculated across each pass through the model

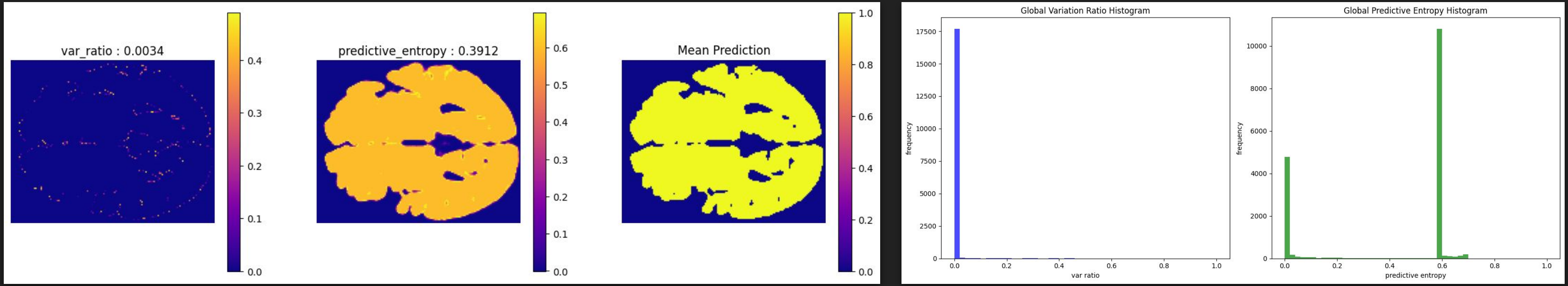



***Fig 15 :*** Uncertainty quantification, average prediction, and uncertainty histogram after 100 iterations on slice 128 of patient 0 from the training set

# Experimental Results for Deep Ensembles Method

**Training Parameters for Deep Ensembles :**

- **Number of Models** : 5
- **Different Seeds** : Trained with different seeds
- **Epochs** : 30
- **Optimizer** : Adam

| Model / Metric | Dice Score | IoU | Precision | Recall |
|---|---|---|---|---|
| Model 1 | 0.9254 | 0.8944 | 0.9243 | 0.9275 |
| Model 2 | 0.9233 | 0.8911 | 0.9035 | 0.9459 |
| Model 3 | 0.9221 | 0.8886 | 0.9241 | 0.9212 |
| Model 4 | 0.9176 | 0.8813 | 0.8967 | 0.9422 |
| Model 5 | 0.9019 | 0.8545 | 0.8601 | 0.9529 |

**Summary Table of the Performance of the 5 Ensemble Models**

***Fig 16 :*** Predictions from the 5 ensemble models for slice 128 for patient 0 from the training set

# Experimental Results for Deep Ensembles Method

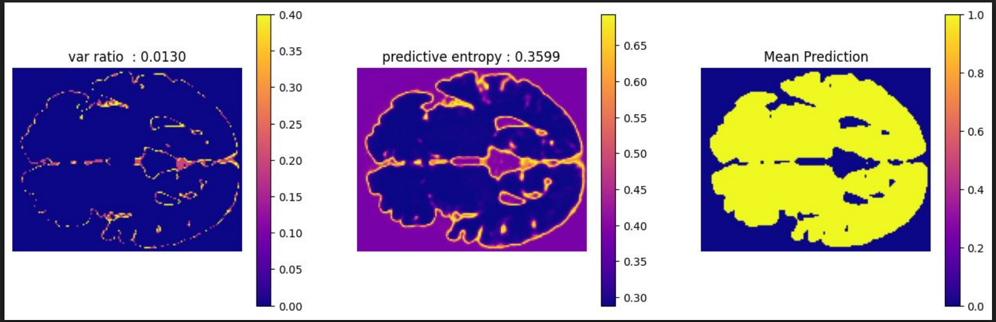



***Fig 17 :*** Uncertainty quantification, average prediction, with Deep Ensembles on slice 128 for patient 0 from the training set

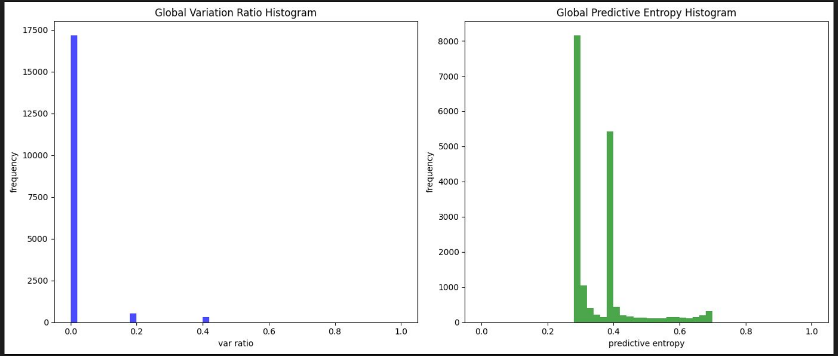



***Fig 18 :*** Uncertainty histograms for Variation Ratio and Predictive Entropy with Deep Ensembles on slice 128 for patient 0 from the training set

# Experimental Results for Deep Ensembles combined with Monte Carlo Dropout Method

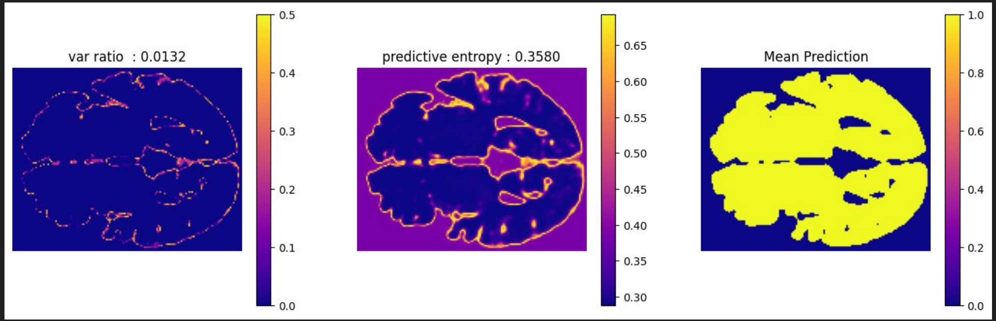



***Fig 19 :*** Uncertainty quantification, average prediction, with Deep Ensembles on slice 128 for patient 0 from the training set

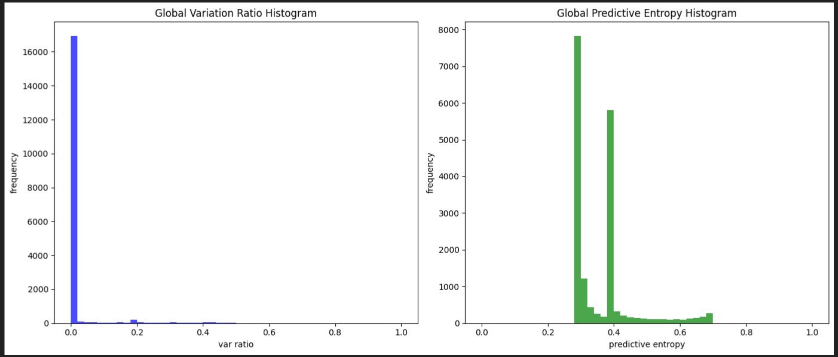



***Fig 20 :*** Uncertainty histograms for Variation Ratio and Predictive Entropy with Deep Ensembles on slice 128 for patient 0 from the training set

# Experimental Results on a blurred image for the 3 methods

**Evaluating Segmentation Uncertainty Metrics Under Gaussian Noise**

- We apply Gaussian noise with a standard deviation of 0.5 to blur the image and introduce noise into the data, aiming to observe the behavior of the uncertainty quantification methods

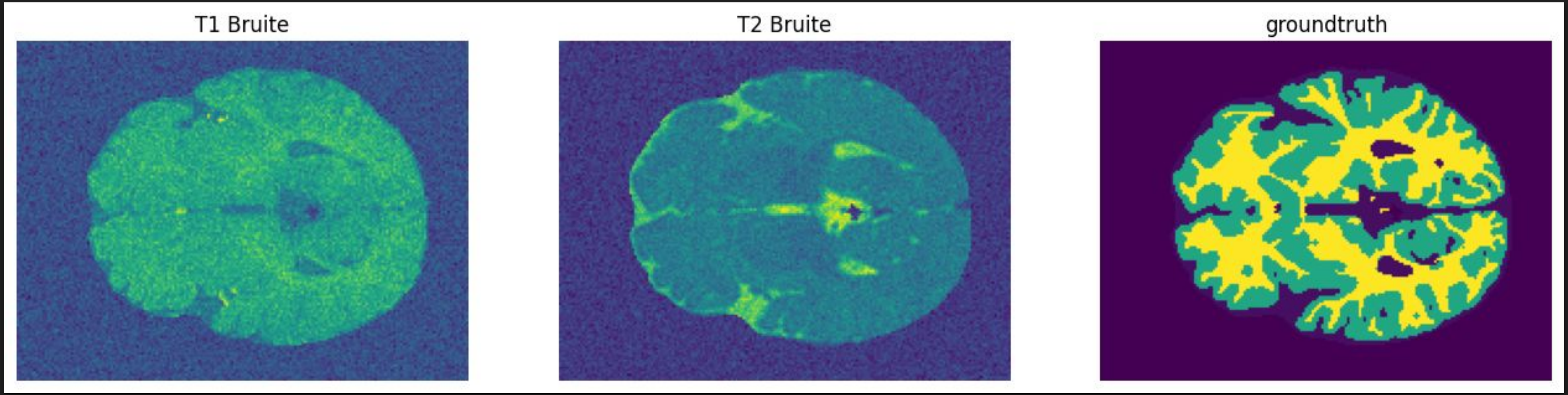



***Fig 21 :*** Noisy T1, T2 and Ground Truth

# Experimental Results on a blurred image for the 3 methods

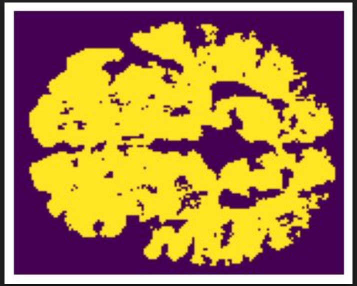


***Fig 22 :*** Mean prediction using the Monte Carlo Dropout (MCD) method

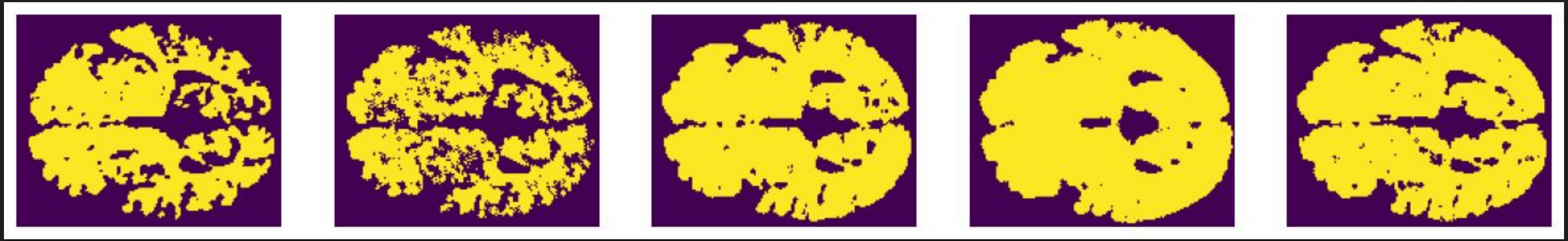


***Fig 23 :*** Predictions from the five models of the ensemble

# Experimental Results on a blurred image for the 3 methods

- We calculate the Variance Ratio to focus on epistemic uncertainty, aiming to measure the variability in predictions caused by model uncertainty rather than data noise

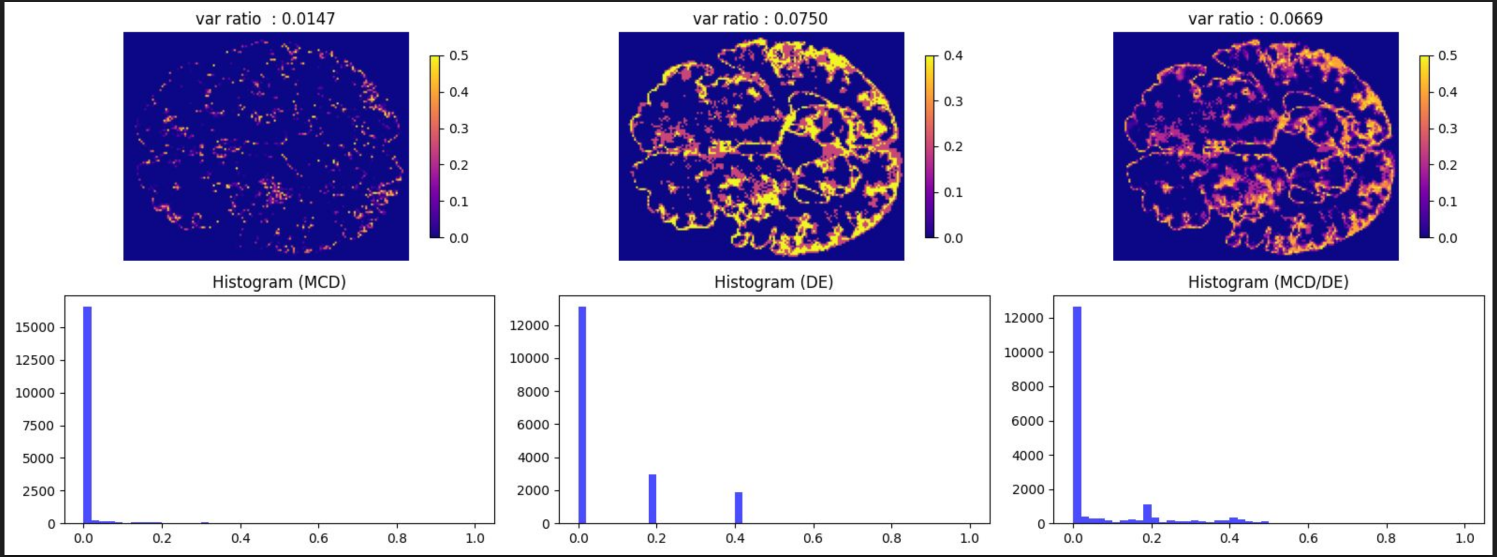



***Fig 24 :*** Variation Ratio maps and histograms for noisy predictions using MCD, Deep Ensembles, and Hybrid methods

# Experimental Results on a blurred image for the 3 methods

- We calculate the Predictive Entropy to focus on aleatoric uncertainty, aiming to measure the variability in predictions caused by data noise rather than model uncertainty

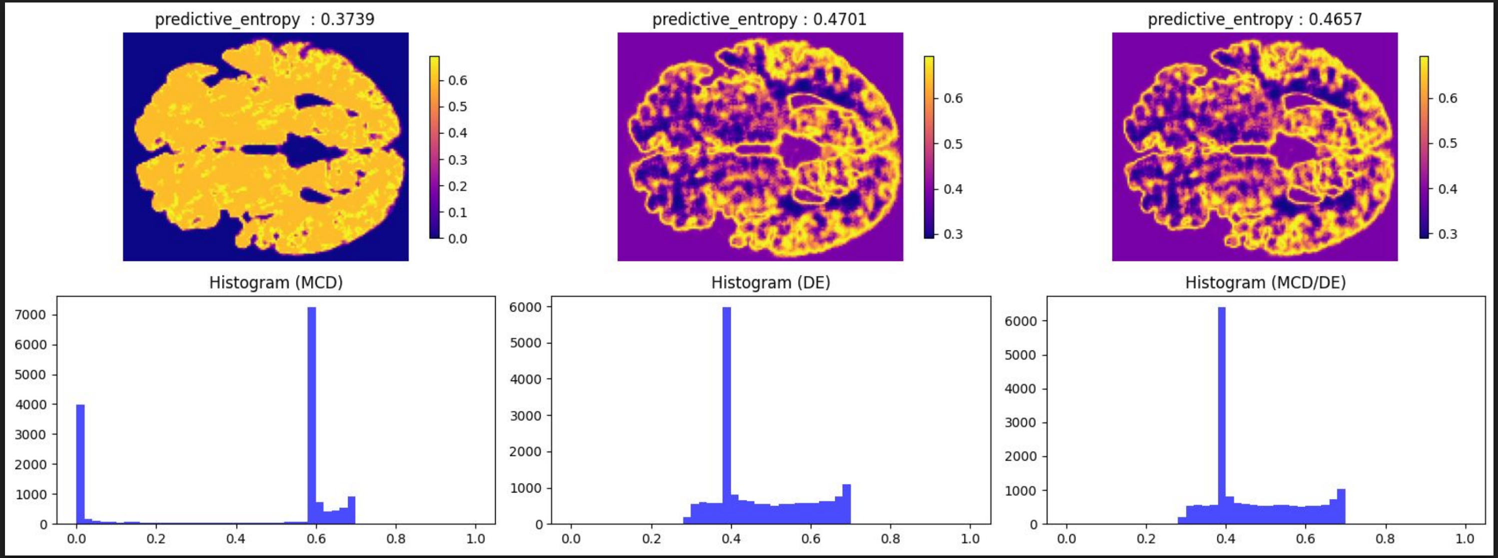



***Fig 25 :*** Predictive Entropy maps and histograms for noisy predictions using MCD, Deep Ensembles, and Hybrid methods

# Experimental Results on a blurred image for the 3 methods (1/2)

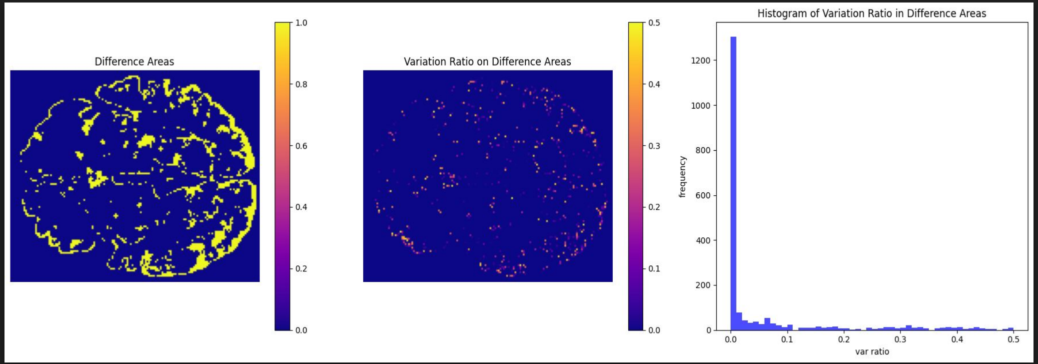



***Fig 26 :*** Difference map between ground truth and prediction using Monte Carlo Dropout, with Variation Ratio analysis and histogram focused on the differing regions

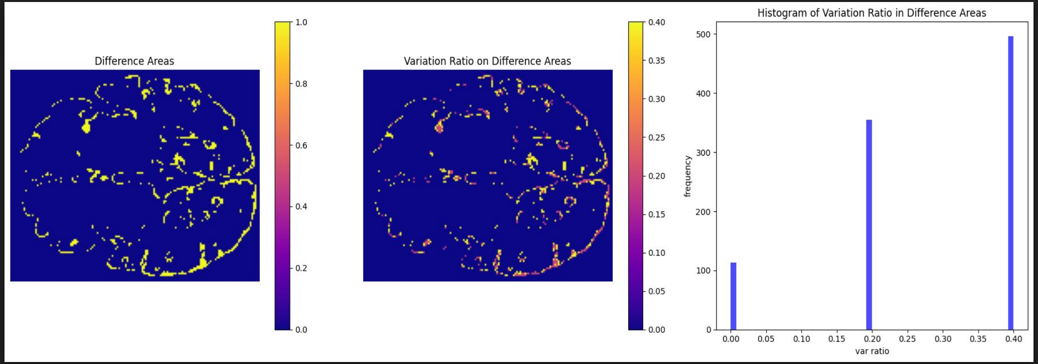



***Fig 27 :*** Difference map between ground truth and prediction using Deep Ensembles, with Variation Ratio analysis and histogram focused on the differing regions

# Experimental Results on a blurred image for the 3 methods (2/2)

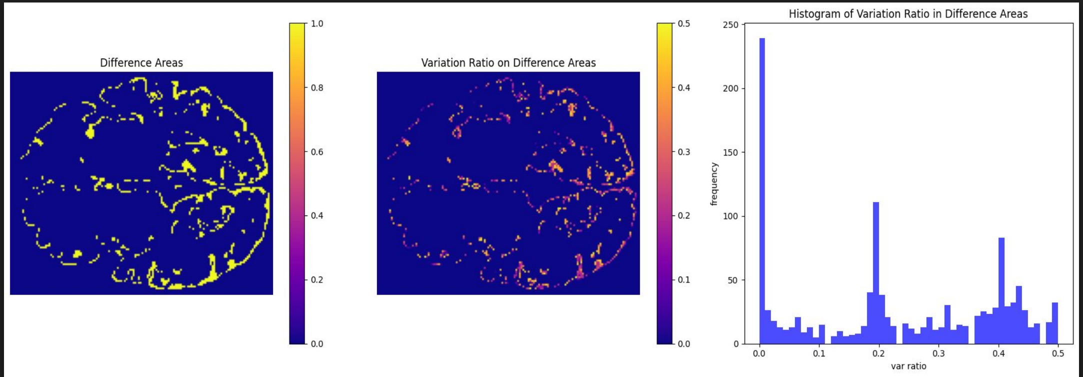



***Fig 28 :*** Difference map between ground truth and prediction using Hybrid Method with Variation Ratio analysis and histogram focused on the differing regions

# Related Work : Achievements for this Semester

- **Hybrid Approach :** Developed a hybrid method combining Deep Ensembles and Monte Carlo Dropout to enhance uncertainty quantification

- **Classification Tasks :** Tested the methods on a classification task using the *MNIST* dataset

- **Uncertainty Quantification Methods :** Implemented and compared Monte Carlo Dropout (MCD) and Deep Ensembles for Uncertainty Estimation

# Future Directions for the Project

- **Exploring more complex Segmentations :** Extend experiments to more complex structures using datasets like MRBrains, requiring finer anatomical segmentation

- **Hybrid Approach Evaluation :** Continue testing the combination of Monte Carlo Dropout and Deep Ensembles, comparing its performance with standard Deep Ensembles alone for uncertainty quantification

# Bibliography (1/2)


[1] : ***Mathématiques et imagerie,*** Bibliothèque Tangente n°77, Edition Pôle, 2022
https://infinimath.com/librairie/pdf/BIB77_sommaire.pdf

[2] : ***Pense-bête de réseaux de neurones convolutionnels***, Afshine Amidi, Shervine Amidi, CS 230 - Pense-bête de réseaux de neurones convolutionnels

[3] : ***Formation EJN – IA Incertitudes et robustesse,*** DE LA RECHERCHE À L'INDUSTRIE, Geoffrey Daniel
https://indico.in2p3.fr/event/25725/contributions/103322/attachments/69771/98691/Incertitudes_robustesse.pdf

[4] : ***Monte Carlo Dropout for Uncertainty Estimation in Deep Learning Model***, Moklesur Rahman, March 22, 2023, Monte Carlo Dropout for Uncertainty Estimation in Deep Learning Model | by Moklesur Rahman

[5] : ***Dropout as a Bayesian Approximation : Representing Model Uncertainty in Deep Learning,*** *Yarin Gal, Zoubin Ghahramani, University of Cambridge* [1506.02142] Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning


LRE
LABORATOIRE DE RECHERCHE DE L'EPITA

# Bibliography (2/2)

[5] : ***Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles***, Balaji Lakshminarayanan Alexander Pritzel Charles Blundell [1612.01474] Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles

[6] : ***MIICCAI Grand Challenge on iSeg-2017, 6-Month infant Brain MRI Segmentation,*** iSeg-2017

[7] : ***D.10 - Score de Brier (BS) et score de performance de Brier (BSS)*** D.10 - Score de Brier (BS) et score de performance de Brier (BSS) - Wikhydro

[8] : ***Sorbonne Université, Computer Science Master Données, Apprentissage et Connaissances (DAC) Bayesian Deep Learning,*** Nicolas Thome RDFIA (M2 SU)