

# CALIBRATION DES RÉSEAUX DE NEURONES MODERNES

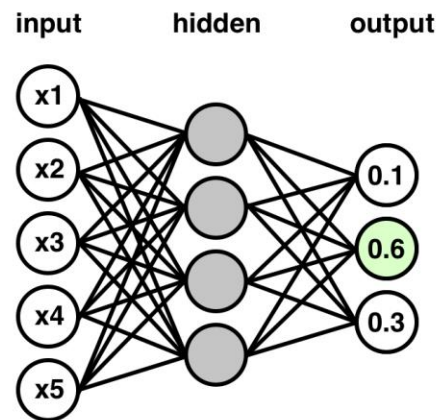
ANASS BEN BOUAZZA

SERGIO BOSSOU



# CONFIANCE DES MODÈLES

- Étant donnée une observation, le réseau fournit sa probabilité d'appartenance à chacune des classes



- Les réseaux doivent non seulement être précis, mais aussi indiquer lorsqu'ils sont susceptibles d'être incorrects  
Ex: Domaine médical : Diagnostic d'un patient

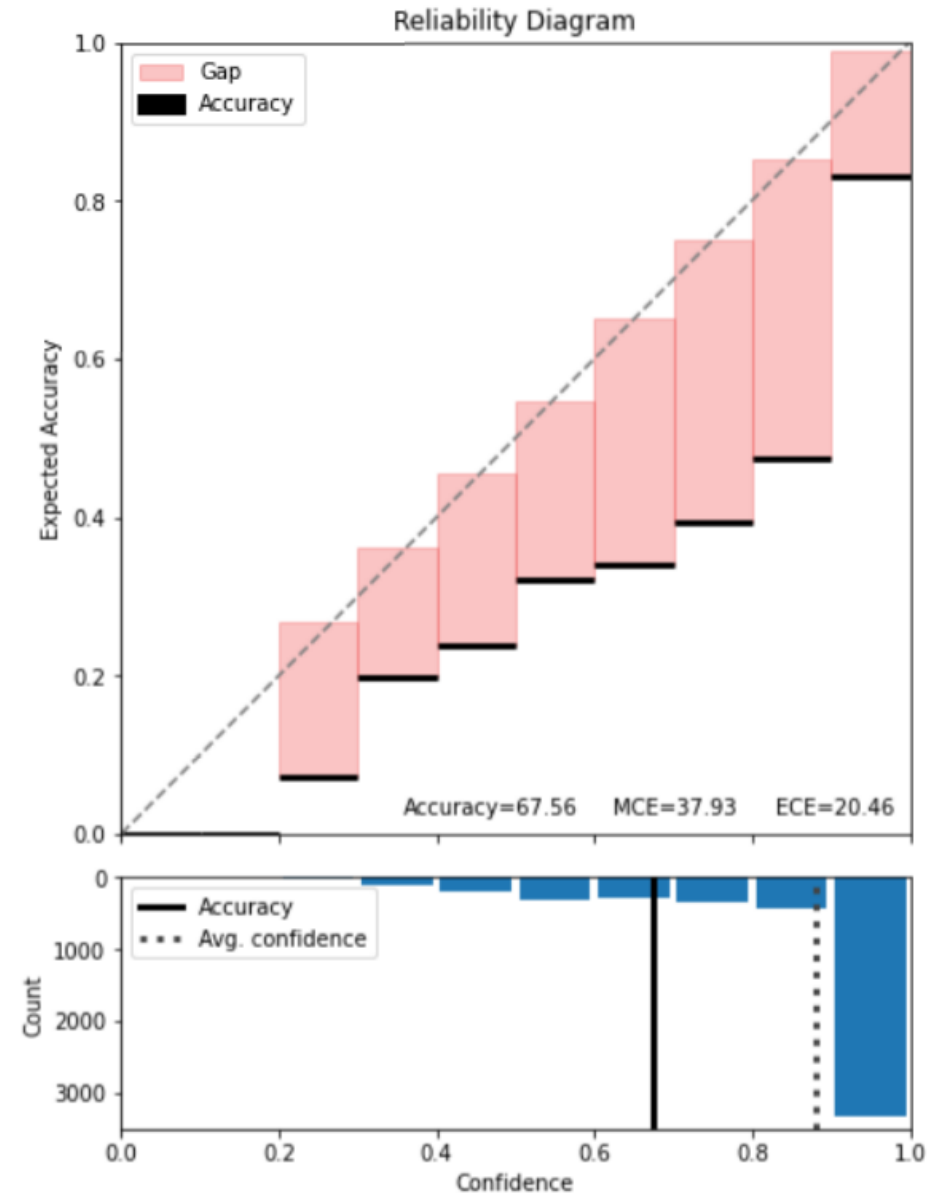
# DIAGRAMME DE CONFIANCE

- Soit  $(B_m)$  l'ensemble des observations dont la prédiction appartient à l'intervalle  $[\frac{m-1}{M}; \frac{m}{M}]$ . On définit :

- La précision : 
$$\text{acc}(B_m) = \frac{1}{|B_m|} \sum_{i \in B_m} \mathbf{1}(\hat{y}_i = y_i),$$

- La confiance : 
$$\text{conf}(B_m) = \frac{1}{|B_m|} \sum_{i \in B_m} \hat{p}_i,$$

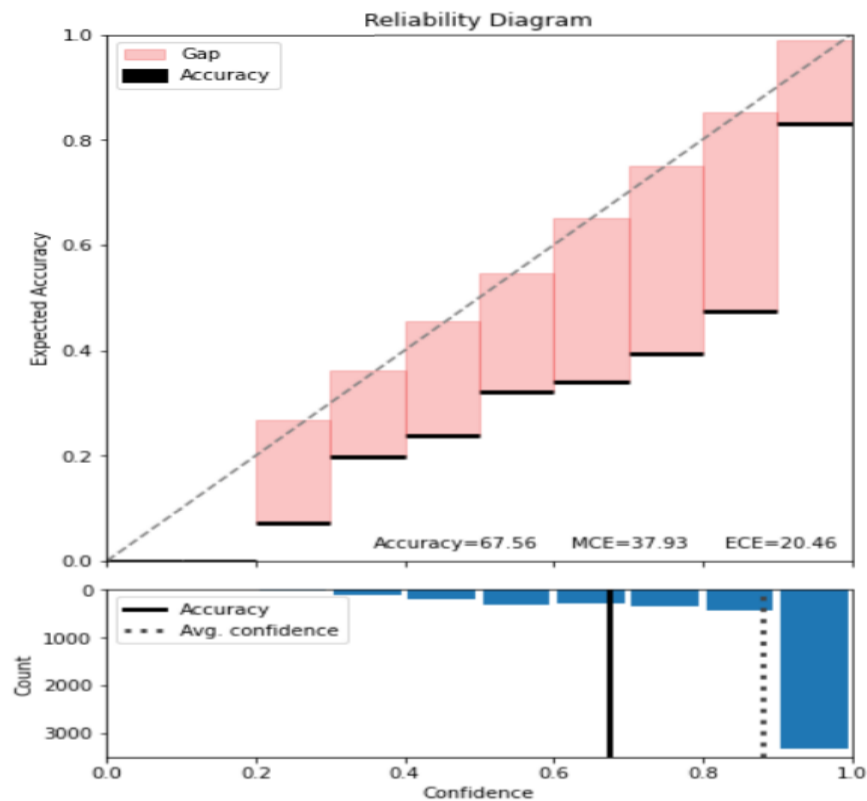
- Pour un réseau parfaitement calibré, ces deux quantités sont égales.



# QUELQUES MÉTRIQUES

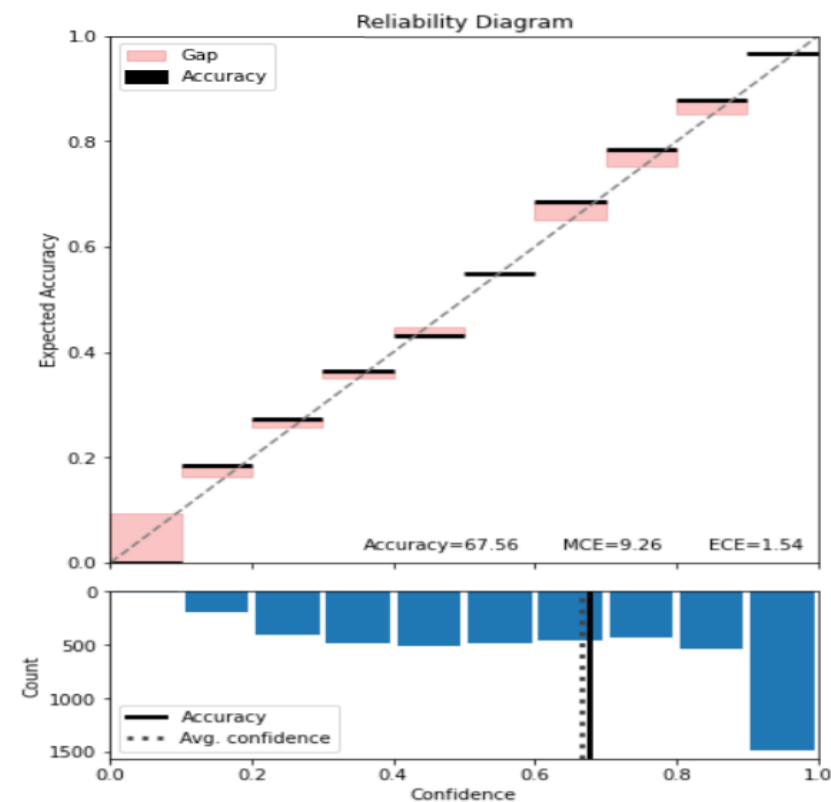
## Expected calibration error (ECE)

$$\text{ECE} = \sum_{m=1}^M \frac{|B_m|}{n} \left| \text{acc}(B_m) - \text{conf}(B_m) \right|$$



## Maximum calibration error (MCE)

$$\text{MCE} = \max_{m \in \{1, \dots, M\}} |\text{acc}(B_m) - \text{conf}(B_m)|$$

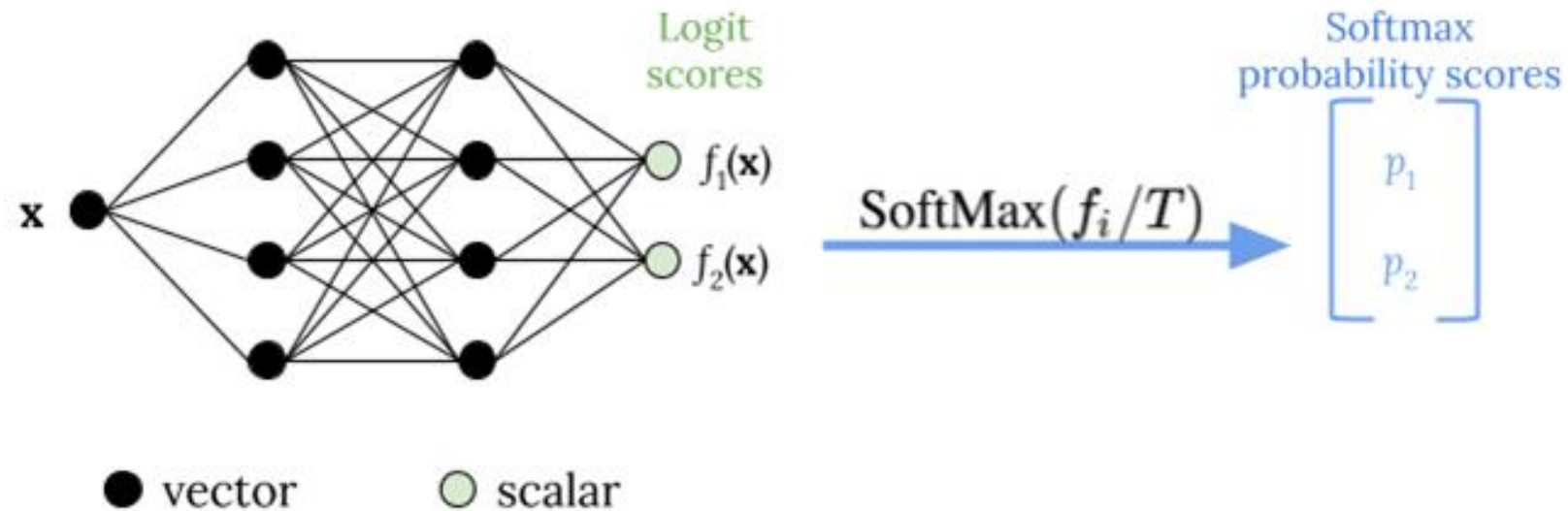


Calibration



# TEMPERATURE SCALING

- Trouver le paramètre  $T$



- On utilise la Negative log-likelihood loss sur les données de validation pour optimiser  $T$ . Heuristiquement, un réseau bien calibré à une NLL loss plus faible.

# AUTRES MÉTHODES DE CALIBRATION

## Platt scaling

- On optimise les matrices  $W$  et  $b$  (généralisation de la temperature scaling).

$$\hat{q}_i = \max_k \sigma_{SM}(\mathbf{W}\mathbf{z}_i + \mathbf{b})^{(k)},$$

## Isotonic regression

- Trouver, pour chaque classe, la bonne fonction continue par morceau qui recalibre la sortie du réseau

$$\min_{\substack{M \\ \theta_1, \dots, \theta_M \\ a_1, \dots, a_{M+1}}} \sum_{m=1}^M \sum_{i=1}^n \mathbf{1}(a_m \leq \hat{p}_i < a_{m+1}) (\theta_m - y_i)^2$$

subject to  $0 = a_1 \leq a_2 \leq \dots \leq a_{M+1} = 1,$   
 $\theta_1 \leq \theta_2 \leq \dots \leq \theta_M.$

# TABLEAU RÉCAPITULATIF

- Réseau : DenseNet-40 entraîné sur 300 epochs
- Données : CIFAR-100

		Accuracy	Avg confidence	ECE	MCE
Validation set	Uncalibrated	67.56	87.34	20.46	37.93
	Temperature scaling	67.56	65.62	1.54	9.26
	Platt scaling	68.06	67.96	1.37	8.44
	Isotonic regression	70.18	70.17	0.07	8.38
Test set	Uncalibrated	65.74	85.50	22.04	40.99
	Temperature scaling	65.74	67.12	1.17	13.19
	Platt scaling	65.88	67.16	1.59	3.59
	Isotonic regression	65.62	72.23	5.34	31.38

# INFLUENCE DE L'ARCHITECTURE

5 réseaux entraînés sur CIFAR100 pour 50 epochs :

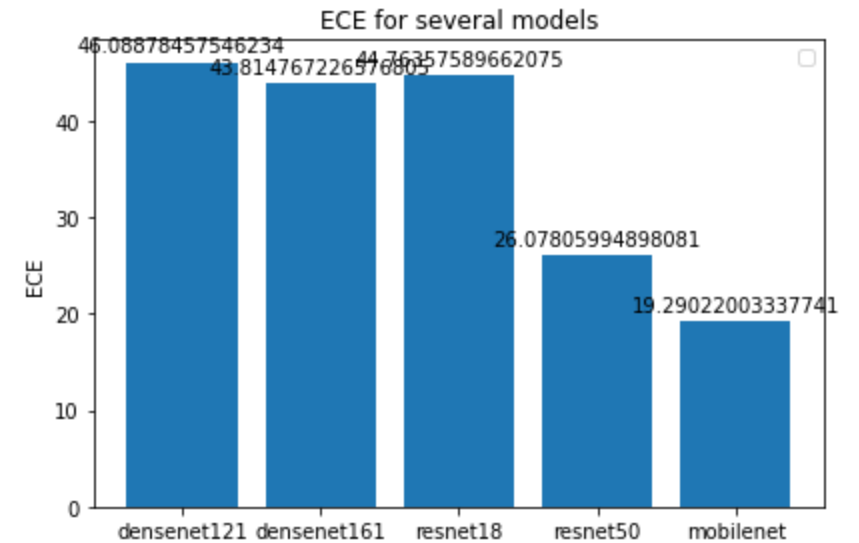
- DenseNet 121 et 161
- ResNet 18 et 50
- MobileNet

Comportement attendu :

MobileNet < DenseNet121 < ResNet18  
< ResNet50 < DenseNet161

Comportement réel :

MobileNet < ResNet50 < DenseNet161  
< ResNet18 < DenseNet121



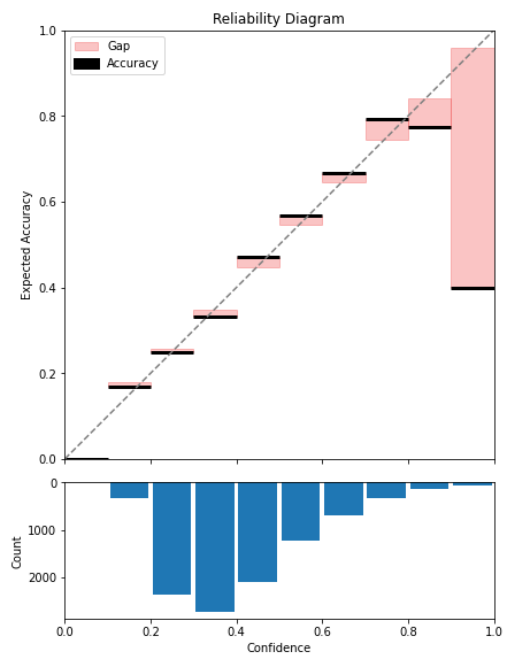


# CALIBRATION PAR FOCAL LOSS

ResNet50 entraîné sur CIFAR10 pour 50 epochs

Cross entropy : Minimise la différence entre probabilités de sortie et labels one-hot encodés

Focal loss :  $FL(p_t) = -(1 - p_t)^\gamma \log(p_t)$   $p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise} \end{cases}$



Avec Temperature scaling

