Mélange de processus de Dirichlet pour la détection de stades phénologiques par imagerie

Félix MERCIER[1] Nizar BOUHLEL[2] Angelina EL GHAZIRI[2] David ROUSSEAU[1,2]

[2]Institut Agro Rennes-Angers, Université d'Angers, INRAE, IRHS, Angers, France

[1]LARIS, Université d'Angers, Angers, France

Résumé – Dans cette communication, nous proposons un couplage innovant entre l'apprentissage profond, représenté par un réseau de neurones convolutifs (CNN) classique, et les modèles de mélange de processus de Dirichlet (DPMM) dans le cadre de la détection automatique des stades phénologiques de croissance des plantules. Le réseau CNN sert à extraire des caractéristiques pertinentes des données d'entrée, ici des séquences d'images couleur RVB, et à les transformer dans un espace latent. Cet espace latent, qui contient des représentations compactes et informatives des données d'origine, est ensuite utilisé comme entrée pour le DPMM. Ce dernier exploite ces représentations pour effectuer des tâches de classification, en adaptant automatiquement le nombre de classes (stades phénologiques) en fonction de la complexité des données. Nous testons ce couplage CNN-DPMM sur diverses espèces de plantes (colza, tomate et pois) afin d'évaluer les performances de la méthode.

Abstract – In this paper, we propose an innovative coupling between deep learning, represented by a classical convolutional neural network (CNN), and Dirichlet process mixture model (DPMM) in the context of automatic detection of phenological stages of seedling growth. The CNN network is used to extract relevant features from input data, here sequence of color images RGB, and transform them into a latent space. This latent space, which contains compact and informative representations of the original data, is then used as input for the DPMM. The latter exploits these representations to perform classification tasks, automatically adapting the number of classes (phenological stages) according to the complexity of the data. We test this CNN-DPMM coupling on various plant species (rapeseed, tomato and pea) to evaluate the method's performance.

1 Introduction

Les plantes se métamorphosent continuellement tout au long de leur développement, avec des étapes clés telles que la germination et la formation des premières feuilles. Le suivi cinétique à haut débit de la croissance des semis de manière automatique présente plusieurs intérêts majeurs : efficacité et précision, optimisation des ressources, gain du temps, réduction des coûts, amélioration de la qualité des données, etc. Dans nos études récentes [8, 4], ce suivi a été réalisé grâce à des séries temporelles d'images couleur RVB que nous couplons à des modèles d'apprentissage profond. D'un point de vue informationnel, le suivi des semis se traduit comme une classification de chaque image selon un ordre préétabli de stade phénologiques.

Pour déterminer automatiquement le stade phénologique atteint à un moment donné, nous proposons de tester une approche bayésienne non paramétrique (BNP). Un modèle bayésien non paramétrique est un modèle bayésien sur un espace de paramètres de dimension infinie. Les modèles BNP ont récemment été appliqués à une variété de problèmes d'apprentissage automatique, notamment la régression, la classification, le clustering, la modélisation de variables latentes, la segmentation d'images, la séparation de sources, etc. [7]. Les généralisations BNP des modèles de mélanges finis fournissent une approche permettant d'estimer simultanément le nombre de composants d'un modèle de mélange et les paramètres des composants individuels du mélange à partir des données [7].

Un exemple populaire de modèles BNP est le modèle de mélange de processus de Dirichlet (DPMM) introduit par Antoniak [1] pour le clustering, qui adapte le nombre de clusters à la complexité des données [7]. Un tel modèle utilise un processus de Dirichlet (DP) comme prior pour la distribution de mélange [3]. Un DPMM combine ainsi un modèle de mélange avec le prior du processus de Dirichlet. Un DP est un processus stochastique qui génère des distributions de probabilité discrètes à composantes infinies. Cela entraîne l'apparition de paramètres de composantes qui se répètent pour former des clusters, c'est-à-dire des groupes de points de données qui partagent les mêmes paramètres. Un modèle de mélange de processus de Dirichlet (DPMM) largement utilisé est le modèle de mélange gaussien infini (IGMM) [11]. Ce modèle surmonte la nécessité de spécifier le nombre de composantes dans les modèles de mélange gaussien traditionnels en supposant que les données proviennent d'un modèle de mélange gaussien (GMM) avec un nombre infini de composantes. Le DPMM peut être représenté explicitement par les trois constructions suivantes : le processus de restaurant chinois (CRP), le Stick-breaking [9] et le schéma d'urne de Polya [2]. Ces constructions donnent lieu à différents algorithmes tels que les méthodes de Monte Carlo par chaîne de Markov (MCMC) (Gibbs/Metropolis-Hasting), les méthodes de Bayes variationnelles (VB), les filtres à particules, la maximisation de l'espérance (EM), etc. Grâce aux progrès des méthodes de simulation et d'échantillonnage, comme les méthodes MCMC, les modèles BNP ont connu des développements rapides pour une variété d'analyses de données pratiques. Sachant que l'échantillonnage de Gibbs est l'un des algorithmes MCMC, Neal a présenté dans [6] quelques algorithmes d'échantillonneur de Gibbs basés sur l'urne de Blackwell-MacQueen, sur le CRP et sur l'échantillonneur de Gibbs effondré.

Dans ce travail, nous présentons une méthode de détection des stades phénologiques des plantes avec DPMM basée sur la

représentation CRP du DP sous-jacent. Cette representation est basée sur le concept d'un restaurant comptant un nombre infini de tables vides. Les clients (points de données) arrivent un par un et choisissent d'être associés à une table déjà occupée, avec une probabilité proportionnelle au nombre de clients déjà assis à cette table, ou de choisir une nouvelle table vide (nouveau cluster). Cette représentation montre comment le nombre de tables occupées (clusters) augmente avec le nombre de clients (données), puis tend à se stabiliser. Ainsi, nous adoptons une démarche bayésienne non paramétrique de mélange de Gaussienne infini qui suppose que le nombre de classes augmente au fur et à mesure que de nouvelles données sont observées. La démarche proposée implique le calcul de la distribution a posteriori des classes et des paramètres de chaque classe. La mise à jour des classes et des paramètres correspondants sont assurées par une procédure d'échantillonnage basée sur l'échantillonnage de Gibbs. Pour encoder les données qui seront ensuite traitées par le DPMM, nous utilisons l'espace latent obtenu d'un réseau de neurones convolutifs (CNN) classique. Cet espace latent, qui contient des représentations compactes et informatives des données d'origine, est ensuite utilisé comme entrée pour les DPMM. Les DPMM exploitent alors ces représentations pour effectuer des tâches de classification, en adaptant automatiquement le nombre de classes (stades) en fonction de la complexité des données. Cette approche hybride permet de tirer parti des avantages des deux méthodes : capacités d'extraction de caractéristiques du CNN et adaptabilité des DPMM, offrant ainsi une solution efficace pour des applications telles que la détermination automatique des stades phénologiques de croissance des plantules.

La communication est organisée de la manière suivante. La section II présente notre cas d'usage des mélanges de processus de Dirichlet couplé à un réseau de neurones de type CNN classique. La section III détaille la méthode bayésienne non paramétrique utilisée pour l'estimation des stades phénologiques. La section IV démontre les performances de la méthode avec un ensemble d'images couleur de tomates, pois et colza. Une comparaison est faite avec deux fonctions coût du CNN. La section VI est consacrée à la conclusion.

2 Stades phénologiques

Nous proposons de tester les mélanges de processus de Dirichlet sur un cas d'usage en imagerie des plantes. Nous nous appuyons sur le jeu de donnée récent public [5] (70000 plantes) où des caméras RVB observent en vue de dessus le passage des 4 premier stades phénologiques (voir Figure 1) de plantules [5]. Les images sont acquises toutes les 15 minutes. Il s'agit pour chaque image de déterminer le stade phénologique $\{0, 1, 2, 3\}$. À cette fin, nous utilisons l'encodeur à base de réseau de neurones convolutionnels présenté dans [5] et visible sur la Figure 2 qui permettra la projection des séries d'entraînement dans l'espace latent à deux dimensions (avant-dernière couche FC full connected). Nous avons aussi entraîné ce réseau de neurones avec deux fonctions de coût (Cross-entropie et Center loss [10]) et allons comparer le comportement de la méthode des mélanges de processus de Dirichlet. Lorsqu'une nouvelle image est acquise le nombre de cluster est calculé et associé au stade phénologique $\{0, 1, 2, 3\}$. Dans la section suivante nous présentons le modèle de mélange de processus de Dirichlet

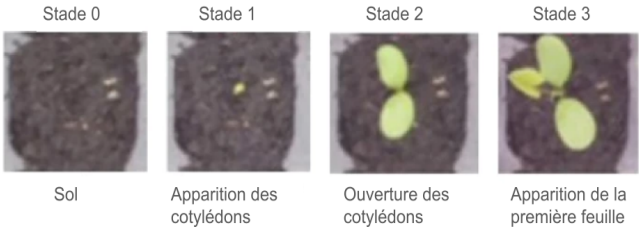


FIGURE 1 : Quatre stades phénologiques observés dans le jeu de données [5] utilisé pour cette communication.

basé sur une représentation de type CRP.

3 Mélange de processus de Dirichlet

Les GMM peuvent modéliser avec succès la structure statistique multimodale complexe des données. Cependant, le GMM suppose que le nombre de composantes gaussiennes est connu. Le GMM infini n'a pas cette limitation, car il suppose que le nombre de composantes gaussiennes est infini. De plus, il permet que le nombre de clusters soit inféré directement à partir des données par l'intermédiaire des méthodes d'échantillonnage MCMC. Dans notre cadre de DPMM, les observations $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\}$ sont générées à partir d'un GMM avec un nombre infini de composantes gaussiennes et soit c_i la classe correspondante à $\boldsymbol{x}_i, i = \{1, ..., n\}$. Le DP génère des distributions de probabilité sur un espace de paramètres $\boldsymbol{\Theta}$. Le DP est utilisé comme une variable a priori sur la distribution de mélange permettant un nombre infini de composantes. Un PD est défini par deux paramètres : une distribution de base G_0, qui est la distribution sous-jacente à partir de laquelle les paramètres des composantes du mélange ($\boldsymbol{\theta}_k$) sont tirés, et le paramètre de concentration α, qui est un scalaire positif contrôlant la concentration ou la diversité des distributions générées. Ainsi, soit $G \sim \text{DP}(\alpha, G_0)$, où G est une distribution discrète aléatoire générée par le processus de Dirichlet, qui fournit les paramètres pour le nombre infini de composantes. $\boldsymbol{x}_i$ est généré à partir de la distribution gaussienne indiquée par c_i comme suit

$$\boldsymbol{x}_i | c_i = k \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad k = 1, 2, ..., \infty \tag{1}$$

On note par $\boldsymbol{\theta}_k = (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ les paramètres de k^{eme} composante gaussienne. Nous plaçons un a priori conjugué sur $\boldsymbol{\theta}_k$ tels que $\boldsymbol{\theta}_k \sim G_0$, qui est choisi de type distribution normale-inverse-Wishart ($\mathcal{NIW}$).

$$\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k | G \sim \mathcal{NIW}(\boldsymbol{\mu}_0, \kappa_0, \boldsymbol{\Lambda}_0, \nu_0) \tag{2}$$

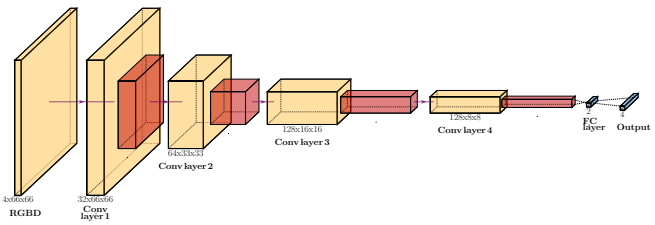


FIGURE 2 : Encodeur utilisé pour réduire la dimension des images de la Figure 1.

avec $\boldsymbol{\mu}_k \sim \mathcal{N}(\boldsymbol{\mu}_0, \frac{\boldsymbol{\Sigma}_k}{\kappa_0})$ et $\boldsymbol{\Sigma}_k \sim \mathcal{IW}(\boldsymbol{\Lambda}_0^{-1}, \nu_0)$ tels que $\boldsymbol{\mu}_0$ est le vecteur moyenne et κ_0 est la précision relative pour l'a priori gaussien sur $\boldsymbol{\mu}_k$ et $\boldsymbol{\Lambda}_0^{-1}$ est la matrice d'échelle et ν_0 est le degrés de liberté de la distribution de Wishart inverse. Notre objectif est d'estimer les paramètres du modèle, à savoir les classes $C = \{c_1, ..., c_n\}$ et les paramètres de chaque classe $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_K\}$ où K est le nombre courant de classes. Cette estimation passe par l'obtention des échantillons de leur distribution postérieure à l'aide de l'échantillonneur de Gibbs. Celui-ci obtient itérativement une nouvelle valeur pour chaque paramètre à partir de sa distribution a posteriori, conditionnée par les valeurs actuelles de tous les autres paramètres. Plus précisément, à chaque itération de l'échantillonnage de Gibbs, $\boldsymbol{\theta}_k$ est mis à jour par échantillonnage de sa distribution a posteriori

$$p(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k | \boldsymbol{X}, C) = \mathcal{NIW}(\boldsymbol{\mu}_{n_k}, \kappa_{n_k}, \boldsymbol{\Lambda}_{n_k}, \nu_{n_k}) \quad (3)$$

où

$$\boldsymbol{\mu}_{n_k} = \frac{\kappa_0}{\kappa_0 + n_k}\boldsymbol{\mu}_0 + \frac{n_k}{\kappa_0 + n_k}\overline{\boldsymbol{x}} \quad (4)$$

$$\kappa_{n_k} = \kappa_0 + n_k \quad (5)$$

$$\nu_k = \nu_0 + n_k \quad (6)$$

$$\boldsymbol{\Lambda}_{n_k} = \boldsymbol{\Lambda}_0 + \mathbf{S} + \frac{\kappa_0 n_k}{\kappa_0 + n_k}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^T \quad (7)$$

avec $\overline{\boldsymbol{x}} = \frac{1}{n_k}\sum_{c_i=k}\boldsymbol{x}_i$ et $\mathbf{S} = \sum_{i=1}^{n_k}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T$. n_k est le nombre de données affectées à la classe k.

De la même manière, c_i $(i = 1, ..., n)$ est mis à jour par échantillonnage de sa distribution a posteriori, obtenue en multipliant l'a priori sur les classes par la vraisemblance :

$$p(c_i = k | C_{-i}, \boldsymbol{X}, \boldsymbol{\Theta}, \alpha) \propto p(\boldsymbol{x}_i | c_i = k, \boldsymbol{\Theta}) p(c_i = k | C_{-i}, \alpha) \quad (8)$$

Pour l'a priori sur les classes, nous suivons la représentation de la CRP. Ainsi, $c_i \sim \text{CRP}(\alpha)$ et il est défini par [6]

$$p(c_i = k | C_{-i}, \alpha) = \begin{cases} \dfrac{n_{-i,k}}{n - 1 + \alpha}, & \text{si } k \in C_{-i} \\ \dfrac{\alpha}{n - 1 + \alpha}, & \text{sinon} \end{cases} \quad (9)$$

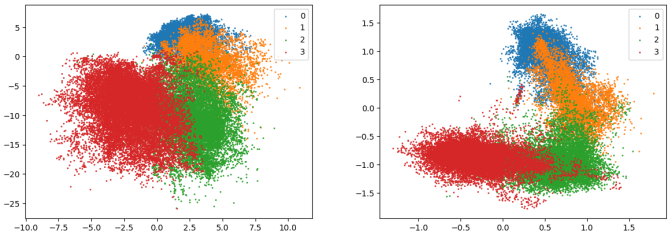
Le $C_{-i} = \{c_1, ..., c_{i-1}, c_{i+1}, ..., c_n\}$ est défini comme l'ensemble des classes sans la classe c_i et $n_{-i,k}$ fait référence au nombre de données de classe k sans compter x_i . Ainsi, une donnée $\boldsymbol{x}_i$ qui arrive est affectée à un cluster non vide avec une probabilité proportionnelle à $n_{-i,k}$ et elle est affectée à un nouveau cluster vide avec une certaine probabilité proportionnelle à α. La vraisemblance est donnée comme suit [6]

$$p(\boldsymbol{x}_i | c_i = k, \boldsymbol{\Theta}) = \begin{cases} p(\boldsymbol{x}_i | \boldsymbol{\theta}_k), & \text{si } k \in C_{-i} \\ \displaystyle\int p(\boldsymbol{x}_i | \boldsymbol{\theta}) p(\boldsymbol{\theta} | G_0) \mathrm{d}\boldsymbol{\theta}, & \text{sinon} \end{cases} \quad (10)$$

L'intégrale précédent correspond à la distribution prédictive postérieure. Dans notre configuration elle correspond à la distribution de Student multidimensionnelle de dimension d

$$t_{\nu_0 - d + 1}\left(\boldsymbol{\mu}_0, \frac{\boldsymbol{\Lambda}_0(\kappa_0 + 1)}{\kappa_0(\nu_0 - d + 1)}\right) . \quad (11)$$

Finalement l'échantillonneur de Gibbs utilisant la représentation CRP se résume dans les étapes suivantes :
- Étant donné $\{\boldsymbol{x}_i\}_{i=1}^n$, itérer



(a) (b)

FIGURE 3 : Quatre stades phénologiques observés dans l'espace latent pour le cas de fonction perte cross entropie (a) et center loss (b).

- Pour $i = 1, ..., n$ tirer : $c_i = k | C_{-i}, \boldsymbol{X}, \boldsymbol{\Theta}, \alpha$ selon (8)
- Pour $k \in \{c_1, ..., c_n\}$ tirer : $\boldsymbol{\theta}_k | \boldsymbol{X}, C$ selon (3)

4 Expérimentation et résultats

Nous disposons d'un ensemble d'images RVB pour trois espèces de plantule : Colza, tomate et pois. Quatre types de modèles sont entraînés en fonction des espèces de plantes dans les données d'entraînement : le modèle Rp est entraîné sur les données relatives au colza, le modèle T sur les données relatives à la tomate et le modèle B sur les données relatives au pois. Le quatrième modèle est le modèle multi-espèces (MS), entraîné sur l'ensemble des espèces. Tous les modèles ont utilisé l'architecture du réseau de neurones convolutifs (CNN) présentée dans la Figure 2.

À partir des modèles entraînés, la projection des séries d'entraînement dans l'espace latent à deux dimensions (avant-dernière couche FC) a été effectuée. La figure 3 montre l'espace latent à deux dimensions pour le cas d'une fonction de perte type Cross entropie du CNN (Figure 3.a) et le cas d'une fonction de perte de type Center loss [10] (Figure 3.b). Les quatre stades phénologiques {0,1,2,3} sont marqués par des couleurs différentes. Dans l'étape suivante, le DPMM effectue des tâches de classification, en adaptant automatiquement le nombre de classes (stades) en fonction des données. Le tableau 1 montre les paramètres utilisés dans l'échantillonneur de Gibbs pour la mise à jour des paramètres $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ et des classes c_i pour les différentes espèces de plantule. Il est à noter que 'itérations' désigne le nombre d'itérations de l'échantillonneur de Gibbs et que le paramètre λ désigne un paramètre interne à la fonction perte Center loss. Le vecteur moyenne et la matrice échelle sont choisies respectivenet $\boldsymbol{\mu}_0 = 0$ et $\boldsymbol{\Lambda}_0^{-1} = \mathbf{I}$ (matrice identité).

TABLE 1 : Paramètres utilisés dans la procédure d'échantillonnage du modèle DPMM.

	Itérations	α	κ_0	ν_0	λ
Modèle Colza	300	0.5	1.2	15	0.6
Modèle Tomate	300	0.5	0.6	5	1.2
Modèle Pois	300	0.5	0.9	5	0.9
Modèle Multi-espèces	300	0.5	1.2	5	0.9

Pour évaluer les performances, nous avons examiné la classification en termes de précision sur les données relatives au

colza (Rp-Acc), aux tomates (T-Acc), aux pois (B-Acc) et à l'ensemble des données (O-Acc). Le processus a été répété 10 fois et nous avons calculé la performance moyenne. Les résultats sont présentés dans les tableaux 2 et 3 pour le cas des fonctions cout Center loss et Cross entropie, respectivement. L'incertitude indiquée représente l'intervalle de confiance à 95 % calculé à partir de l'écart-type sous l'hypothèse gaussienne. Il est à noter que les quatre modèles sont appliqués sur les données test de type Colza, tomate, pois et mélange de tout.

En ce qui concerne la précision par espèce, chaque ensemble de test est mieux décrit par son modèle spécifique respectif. Dans le cas du tableau 2, le modèle T atteint une précision de 79,8 % pour les données test tomate, le modèle B atteint 65.1 % pour les données test pois, le modèle Rp atteint 57.6 % pour les données test colza, le modèle MS atteint 73.5 % pour le mélange de toutes les espèces. L'application de modèles monospécifiques à d'autres espèces donne des résultats peu satisfaisants. L'exemple le plus remarquable de transférabilité est le modèle B sur les données de colza, avec une précision de 43.3 %. Pour le cas du tableau 3 où il s'agit de la fonction de perte cross entropie, les performances sont très médiocres par rapport à celles du tableau 2.

Bien que la performance globale du modèle MS soit parfois inférieure à celle de chaque modèle monospécifique, cela représente un gain de généralité. En acquérant un vaste ensemble de données d'espèces mixtes, nous avons développé un modèle capable de traiter plusieurs espèces avec un niveau de précision comparable. Cela démontre la valeur ajoutée de l'ensemble de données fourni.

TABLE 2 : Précision de classification (en %) pour les modèles entraîner avec Colza (Rp modèle), Tomate (T modèle), Pois (B modèle) et Multi-Species (MS modèle) avec Center loss.

	Rp modèle	**T modèle**	**B modèle**	**MS modèle**
O-Acc	64.3 ± 3.8	70.7 ± 4.2	63.4 ± 4.9	73.1 ± 3.2
Rp-Acc	57.6 ± 5.2	61.7 ± 6.6	43.3 ± 3.0	71.5 ± 4.0
T-Acc	69.4 ± 3.8	79.8 ± 3.8	$71.5 \pm 7.$	78.5 ± 3.6
B-Acc	61.4 ± 8.4	63.9 ± 6.8	65.1 ± 8.2	66.5 ± 4.2

TABLE 3 : Précision de classification (en %) pour les modèles entraîner avec Colza (Rp modèle), Tomate (T modèle), Pois (B modèle) et Multi-Species (MS modèle) avec cross entropie.

	Rp modèle	**T modèle**	**B modèle**	**MS modèle**
O-Acc	36.5 ± 2.0	41.8 ± 2.1	43.4 ± 2.6	52.1 ± 2.2
Rp-Acc	34.7 ± 2.6	32.0 ± 3.5	35.7 ± 6.4	36.5 ± 3.0
T-Acc	40.1 ± 2.5	52.0 ± 3.2	50.9 ± 3.2	64.7 ± 3.3
B-Acc	32.3 ± 3.4	33.8 ± 2.7	37.8 ± 2.9	44.5 ± 4.2

5 Conclusion et perspective

Un couplage a été proposé entre l'apprentissage profond, représenté par un réseau de neurones convolutifs (CNN), et les modèles de mélange de processus de Dirichlet (DPMM) dans le cadre de la détection automatique des stades phénologiques de croissance des plantules. L'approche hybride CNN-DPMM a été testée sur diverses espèces de plantes (colza, tomate et pois) afin d'évaluer les performances de la méthode. Les résultats de classification sont encourageants et peuvent être considérablement améliorés de plusieurs façons. La première façon d'y parvenir consiste à estimer le paramètre α en proposant un a priori sur ce paramètre au lieu de le fixer dans notre article. La deuxième possibilité consiste à utiliser d'autres fonctions coût plus discriminatives.

Références

[1] Charles E. ANTONIAK : Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *The Annals of Statistics*, 2(6):1152–1174, 1974.

[2] David BLACKWELL et James B. MACQUEEN : Ferguson Distributions Via Polya Urn Schemes. *The Annals of Statistics*, 1(2):353 – 355, 1973.

[3] Thomas S. FERGUSON : A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*, 1(2): 209 – 230, 1973.

[4] Hadhami GARBOUGE, Pejman RASTI et David ROUSSEAU : Enhancing the tracking of seedling growth using rgb-depth fusion and deep learning. *Sensors*, 21(24), 2021.

[5] Félix MERCIER, Geoffroy COUASNET, Angelina EL GHAZIRI, Nizar BOUHLEL, Alain SARNIGUET, Muriel MARCHI, Matthieu BARRET et David ROUSSEAU : Deep-learning-ready rgb-depth images of seedling development. *Plant Methods*, 21(16), 2025.

[6] Radford M. NEAL : Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2):249–265, 2000.

[7] Peter ORBANZ et Yee Whye TEH : *Bayesian Nonparametric Models*, pages 81–89. Springer US, Boston, MA, 2010.

[8] Salma SAMIEI, Pejman RASTI, Joseph LY VU, Julia BUITINK et David ROUSSEAU : Deep learning-based detection of seedling development. *Plant Methods*, 16: 1746–4811, 2020.

[9] Jayaram SETHURAMAN : A constructive definition of dirichlet priors. *Statistica Sinica*, 4(2):639–650, 1994.

[10] Yandong WEN, Kaipeng ZHANG, Zhifeng LI et Yu QIAO : A discriminative feature learning approach for deep face recognition. *In Computer Vision - ECCV 2016*, pages 499–515, Cham, 2016. Springer International Publishing.

[11] Hao WU et Saurabh PRASAD : Dirichlet process based active learning and discovery of unknown classes for hyperspectral image classification. *IEEE Trans. on Geosc. and Rem. Sens.*, 54(8):4882–4895, 2016.