

## Développement d'un outil de reconnaissance automatique des espèces ligneuses tropicales

**Auteur :** Kaddouri, Marjane

**Promoteur(s) :** Doucet, Jean-Louis; Lejeune, Philippe

**Faculté :** Gembloux Agro-Bio Tech (GxABT)

**Diplôme :** Master en bioingénieur : gestion des forêts et des espaces naturels, à finalité spécialisée

**Année académique :** 2021-2022

**URI/URL :** <http://hdl.handle.net/2268.2/15391>

---

### *Avertissement à l'attention des usagers :*

*Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.*

*Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.*

---

---

DÉVELOPPEMENT D'UN OUTIL DE  
RECONNAISSANCE AUTOMATIQUE DES  
ESPÈCES LIGNEUSES TROPICALES

---

Kaddouri Marjane

Travail de fin d'études présenté en vue de l'obtention du diplôme de  
Master bioingénieur en Gestion des forêts et des espaces naturels

Année académique 2021-2022

Co-promoteurs : Professeur Doucet Jean Louis et Professeur Lejeune Philippe

“Any reproduction of this document, by any means whatsoever, may only be made with the authorization of the author and the academic authority of Gembloux Agro-Bio Tech.”

“This document is the sole responsibility of its author.”

“Toute reproduction du présent document, par quelque procédé que ce soit, ne peut être réalisée qu’avec l’autorisation de l’auteur et de l’autorité académique de Gembloux Agro-Bio Tech. “

“Le présent document n’engage que son auteur. “

## Remerciements

J'aimerais exprimer ma reconnaissance à mes co-promoteurs, Monsieur Jean-Louis Doucet et Monsieur Philippe Lejeune pour m'avoir donné l'opportunité de réaliser ce projet ainsi que pour m'avoir encadré et conseillé tout au long de sa réalisation.

Je remercie Alexandre Delplanque pour son encadrement, sa pédagogie et l'ensemble du savoir qu'il m'a transmis.

Je remercie aussi grandement la société Pallisco-Cifm ainsi que Monsieur Paul Lagoutte pour m'avoir permis de mener à bien ma mission. Je remercie particulièrement Stéphane Tchakoudeu et Isaac Blaise Mbe de m'avoir encadrée et d'avoir mis l'ensemble des moyens disponibles à ma disposition. Un grand merci au reste de l'équipe scientifique ainsi qu'à Papa Ismaël et International sans qui les trajets en voiture n'auraient pas été les mêmes. Je remercie de même Monsieur Elie et mon amie Kelly pour leur accueil.

Je suis également reconnaissant envers l'Ares pour son financement, cette participation au voyage m'a permis de réaliser mon rêve, réaliser un mémoire au sein de la forêt tropicale.

De plus, j'aimerais aussi toucher un mot à ma famille, pour qui la vitamine S a toujours été le maître mot, en particulier, Yasmine et Maman alias « Chris » pour leur soutien infaillible.

Pour finir, je remercie mes amis pour ces cinq années passées à vos côtés. J'aimerais en particulier remercier Clara, Tom et Florian pour leur soutien.



## Abstract

Tree recognition is a prerequisite for the sustainable management of tropical forests. However, it is made complex by the multitude of species and the difficulty to observe closely the main discriminating organs. Deep learning associated with an easily accessible tool such as the telephone, adapted for taking images of bark, could be a promising technology for automatic recognition. The objective of this work will be to develop a preliminary artificial intelligence tool for recognising tropical tree species. To this end, a database of bark images of central African tree species was created. Based on this dataset, a classification model using the ResNet18 convolutional neural network was produced. This resulted in a database of 56,000 images containing 20 classes. The recognition performance of the model is high with 81.3% accuracy with a tendency to confusion for species with similar morphological characteristics. This tool has high potential for larger and more complex databases. However, drastically increasing the number of species will require an increase in the complexity of the used technologies.

### **Mots clés**

Deep learning, convolutional neural network, automatic recognition, bark, tropical botanic

## Résumé

La reconnaissance des arbres est un préalable à la gestion durable des forêts tropicales. Elle est toutefois rendue complexe par la multitude d'espèces et la difficulté d'observer de près les principaux organes discriminants. Le Deep learning associé à un outil simple d'accès tel que le téléphone, adapté pour la prise d'images d'écorces pourrait être une technologie particulièrement prometteuse. L'objectif de ce travail est donc la création d'un outil préliminaire d'intelligence artificielle de reconnaissance d'espèces d'arbres tropicaux. Pour ce faire, une base de données d'images d'écorces de 20 espèces d'arbres d'Afrique centrale a été créée. Sur base de ce jeu de données, un modèle de classification utilisant le réseau de neurones convolutifs ResNet 18 a été créé. Il en résulte une base de données de 56 000 images contenant 20 classes. Les performances de reconnaissance du modèle sont élevées avec 81,3% d'exactitude avec une tendance à la confusion pour les espèces ayant des caractéristiques morphologiques similaires. Cet outil montre des potentialités d'avenir élevées pour des bases de données plus larges et plus complexes. Cependant, augmenter drastiquement le nombre d'espèces étudiées demandera une complexification des technologies utilisées.

### **Mots clés**

Deep learning, réseaux de neurones convolutifs, reconnaissance automatique, écorce, botanique tropicale

# Table des matières

1.	Introduction et objectifs .....	1
1.1.	Introduction .....	1
1.1.1.	Connaissances botaniques : accès et démocratisation.....	1
1.1.2.	Avancées technologiques de l'information en Afrique centrale .....	1
1.1.3.	Reconnaissance automatique des espèces forestières .....	2
1.2.	Objectifs du travail .....	7
2.	Matériel et méthode.....	8
2.1.	Zone d'étude.....	8
2.1.1.	Localisation .....	8
2.1.2.	Site de recherche .....	9
2.2.	Facteurs écologiques liés à la zone d'étude.....	11
2.3.	Espèces étudiées.....	12
2.4.	Création d'une base de données et d'un outil de reconnaissance automatique par CNN	15
2.4.1.	Création d'une base de données.....	15
2.4.2.	Développement d'un outil d'intelligence artificielle.....	25
2.4.3.	Analyse des causes d'erreurs du modèle .....	33
2.5.	Contribution personnelle.....	34
3.	Résultats .....	35
3.1.	Création d'un outil de reconnaissance automatique.....	35
3.2.	Performance du modèle développé.....	35
3.3.	Analyse des causes d'erreurs du modèle.....	39
3.3.1.	La présence d'images révélant des artefacts .....	39
3.3.2.	Etude de l'impact des artefacts : carte de chaleurs.....	42
4.	Discussion.....	45
4.1.	Performance du modèle .....	45
4.1.1.	Comparaison des performances à la littérature disponible.....	45
4.2.	Etude du lien entre la qualité de la base de données et les performances du modèle.....	46
4.2.2.	Impact des artefacts sur les performances du modèle .....	50
4.2.3.	Erreur humaine .....	52
4.3.	Le critère de l'écorce est-il suffisamment discriminant?.....	52
4.4.	Limites de l'étude.....	53

4.4.1.	Méthodologie de création du modèle .....	53
4.4.2.	Limite du matériel de terrain.....	54
4.5.	Identification botanique, lien entre l'humain et la technologie.....	55
5.	Conclusion et perspectives.....	56
6.	Bibliographie.....	58
7.	Annexes.....	63

## Table des figures

Figure 1 : Schéma d'un "Convolutional neural network" (Elgendy, 2020).....	3
Figure 2 : Localisation de la société Pallisco-CIFM.....	8
Figure 3 : Zone de recherche au sein de la concession avec un agrandi sur les deux parcelles de recherche principale.....	10
Figure 4 : Situation et diagramme ombrothermique de la zone d'étude au Cameroun. ....	11
Figure 5 : Représentation des écorces relatives aux espèces étudiées.....	14
Figure 6 : Représentation des artefacts possiblement présents au sein de la base de données .....	18
Figure 7 : Feuille de <i>Rauvolfia macrophylla</i> dont le nombre de nervures permet sa différenciation vis à vis de <i>Rauvolfia vomitoria</i> .....	19
Figure 8 : Tranche réalisée dans le but de reconnaissance d' <i>Alstonia boonei</i> grâce à son latex .....	19
Figure 9 : Tronc et houppier de <i>Terminalia superba</i> .....	19
Figure 10 : mesure du diamètre d'un individu.....	20
Figure 11 : prise d'image d'un individu .....	20
Figure 12 :Histogramme de répartition des classes de diamètre au sein du jeu total de données .....	22
Figure 13 : Histogramme de la proportion des individus présentant un artefact sur l'ensemble des individus d'une même classe.....	23
Figure 14 :Comparaison d'une image avant et après réduction d'un facteur 8 du nombre de pixels.....	24
Figure 15 : Organigramme des étapes de création de l'outil d'intelligence artificielle .....	25
Figure 16 : Représentation de l'ajustement du modèle par la descente du gradient de la fonction de coût. Les graphiques du haut sont l'évolution de la fonction de coût en fonction du temps et ceux du bas sont l'optimisation de la droite représentant le modèle .....	27
Figure 17 : Représentation d'une matrice de confusion.....	30
Figure 18 : Schématisation du rappel et de la précision avec l'exemple de l'espèce <i>Afzelia bipindensis</i> . .....	32
Figure 19 : Représentation d'une carte de chaleur d'une image d'écorce.....	33
Figure 20 : Matrice de confusion des performances du modèle .....	36
Figure 21: Proportions d'images relatives aux 20 classes choisies présentant des artefacts .....	39
Figure 22 : Histogramme du pourcentage d'artefacts lié à l'opérateur en fonction de la classe .....	40
Figure 23 : Histogramme des artefacts présents sur le tronc en fonction de la classe .....	41
Figure 24 : Carte de chaleurs des zones d'intérêts du modèle pour la classification ( A : Exemple sans artefacts, B et C : Exemple avec mousse et lichens ).....	43
Figure 25 : Carte de chaleurs des zones d'intérêts du modèle pour la classification ( A : Image floue, B : Termitière, C et D : arrière-plan ) .....	44

Figure 26 : Représentation d'images d'écorces impactées par les conditions climatiques. Les trois images du haut représentent *Periscopsis elata* exposé ou non à des conditions d'exposition élevées. Les trois images du bas correspondent à l'espèce *Trichilia welwitschii* avec le tronc mouillé suite à un épisode pluvieux ou un tronc sec.....49

## Table des tableaux

Tableau 1 : Base de données d'images d'écorces existantes (Carpentier et al.,2018 ; Misra et al., 2020 ; Kim et al., 2022).....	5
Tableau 2 : Coordonnées des deux parcelles de 4 ha. Les coordonnées sont exprimées en mètres dans le système de coordonnées WGS 84/UTM zone 33N (EPSG : 32633) (Dupuis, 2019) .....	9
Tableau 3 : Description morphologique de l'écorce des espèces étudiées ( Meunier et al, 2015 ; Wilk et Issembée, 2000; Prota4u, 2022; Vivien & Faure, 2001 ; Hallé, 1966).....	13
Tableau 4 : Intervalle de diamètre en centimètre pour l'ensemble des espèces étudiées (Doucet communication personnelle, 2022 ; Meunier et al.,2015; PROTA4U, 2022 ; Vivien & Faure, 2011).....	16
Tableau 5 : Nombre d'individus et d'images par classe.....	21
Tableau 6 : Résumé des performances du modèle pour chaque espèce étudiée.....	38
Tableau 7 : Comparaison du nombre d'images en fonction des performances ( F-1 score) par classe .....	47
Tableau 8 : Comparaison des performances de 2 classes ayant le même nombre d'individus .....	48
Tableau 9 : Table des Intervalles de performance (F1-score) pour les classes de variabilités des diamètres .....	50
Tableau 10 : Comparaison des hyperparamètres de cette étude à ceux utilisés par Carpentier et al.(2018) et Kim et al.(2022). Seuls les hyperparamètres disponibles en littérature ont été comparés.....	54

## Table des annexes

Annexe 1 : Cartographie des arbres d'intérêts répertoriées dans la Parcelle 1 .....	63
Annexe 2 : Cartographie des arbres d'intérêts répertoriées dans la Parcelle 2 .....	63
Annexe 3 : Illustration des différentes morphologies d'écorces (Letouzey,1972).....	64
Annexe 4 : Capture de l'interface Albuementations permettant le test des différentes augmentations de données.....	65
Annexe 5 : Nombre d'individus par classe et catégorie de division du jeu de données.....	66
Annexe 6 : Cartographie de la distribution des individus de chaque espèce par une cartographie de la densité ( chaque pixel représente 25 hectares).....	67
Annexe 7 : Histogrammes du nombre d'arbres par classe de diamètres pour l'ensemble des espèces ( Page A : Espèces en A, Page B : Espèces de C à I , Page C : Espèces de K à Po, Page D : Espèces de Pt à S, Page E : Espèces de T à X) .....	77



# 1. Introduction et objectifs

## 1.1. Introduction

### 1.1.1. Connaissances botaniques : accès et démocratisation

Dès 1909, le *Sylloge Florae Congolanae* (Durand & Durand 1909) résumait l'état des connaissances sur la flore du Congo. Ensuite, de nombreuses flores décrivant les arbres tropicaux du Bassin du Congo ont été publiées telles que la « Flore du Gabon » comprenant 58 volumes publiés entre 1961 et 2022 (Jardin botanique de Meise, 2022b) et « La flore d'Afrique centrale », cette flore contenant plus de 11 000 plantes vasculaires recensées en République Démocratique du Congo, au Rwanda et au Burundi (Jardin botanique de Meise, 2022a). Ces ouvrages, très précieux pour les botanistes, se basent principalement sur les organes reproducteurs pour séparer les espèces. Les termes utilisés dans ces ouvrages sont souvent complexes, ce qui les placent hors de portée du grand public et des utilisateurs de terrain (Joly et al., 2016).

Pour y remédier, des outils de vulgarisation ont été créés, par exemple « Les arbres de la Guinée équatoriale » par Wilk et Issembé en 2000 ou « Les arbres de forêt dense d'Afrique centrale » de Vivien & Faure (2011) ou encore « Les Arbres utiles du Gabon » par Meunier et al. (2015). Plus facilement accessibles, ces ouvrages souffrent cependant d'un manque d'exhaustivité.

Dans une même volonté de rendre accessibles les outils de reconnaissance botanique, les avancées technologiques actuelles en termes d'équipement, de stockage d'informations et de puissance de calcul rendent la production de support multimédia plus simple et accessible (Joly et al., 2016). Les médias rendent accessibles les connaissances botaniques et permettent donc d'innover et de s'ouvrir à de nouveaux publics.

### 1.1.2. Avancées technologiques de l'information en Afrique centrale

En Afrique centrale, la fracture numérique a été déplorée depuis les années 2000. Cependant, depuis ces dix dernières années, la technologie numérique devient de plus en plus prometteuse étant donné la généralisation du téléphone portable et de l'internet mobile sur le continent africain (Al Dahdah, 2019).

L'explosion de la téléphonie mobile en Afrique sub-saharienne peut être mise en lien avec la problématique du présent travail. En effet, les technologies de l'information montrent de grandes perspectives comme appui au développement en cours en Afrique (Loukou, 2012). En effet, ces nouvelles technologies pourraient apporter une réponse au besoin des forestiers tropicaux, voire à l'ensemble de la société et offrir un outil de reconnaissance botanique accessible sur téléphone mobile. Cette avancée représenterait un appui substantiel au développement du secteur forestier tout en multipliant de manière exponentielle le nombre d'acteurs ayant accès à l'information (Doumenge et al., 2016).



En Occident, une application de reconnaissance automatique de plantes (Pl@ntNet) a déjà montré son utilité comme outil de surveillance écologique et de sensibilisation du grand public à l'environnement. En parallèle, l'utilisation d'identification multimédia et le partage de données entre utilisateurs sont considérés comme les solutions les plus prometteuses à l'avenir pour centraliser l'information botanique (Jolly et al, 2015).

L'automatisation de la reconnaissance des espèces d'arbres n'est pas une tâche simple au vu du nombre d'espèces (4500 à 6000 espèces d'arbres tropicaux africains selon Slick et al. (2015) et de l'inaccessibilité des organes (Jendoubi, 2018). Il est donc indispensable de tester les perspectives d'utilisation de technologies d'automatisation sur les espèces d'arbres tropicaux.

Pour ce faire, le Deep learning est un outil intéressant et potentiellement très puissant, car il permet de réaliser des tâches de reconnaissance d'images avec une précision élevée et permet d'automatiser des processus laborieux (Misra et al, 2020).

### 1.1.3. Reconnaissance automatique des espèces forestières

#### 1.1.3.1. Convolutional neural network (CNN)

Le Deep learning est un type d'intelligence artificielle (IA) compris au sein du Machine learning, fonctionnant sur le principe d'apprentissage automatique. La spécificité du Deep Learning est que son apprentissage est basé sur les réseaux de neurones artificiels. Un "Convolutional neural network" est une sous-catégorie du Deep Learning qui a démontré de grandes perspectives en analyse d'images (Chen et al., 2021 ; Dhruv & Naskar, 2020). Il a été utilisé pour identifier des personnes, objets, des animaux ou différentes races d'animaux, ainsi que pour différencier divers types d'agriculture (Elgendy, 2020). Le CNN apprend à reconnaître les caractéristiques importantes d'éléments d'images afin de pouvoir les détecter, et ce, sur base d'un jeu de données d'entraînement (Elgendy, 2020).

Cet apprentissage est réalisé en valorisant de façon progressive (Figure 1) :

- La couche d'entrée correspondant à l'image d'entrée que l'utilisateur veut caractériser.
- Les couches de convolution permettant l'extraction des caractéristiques de l'image au travers de "feature maps" correspondant à des filtres de plus en plus spécifiques. Les "feature maps" seront les informations extraites via l'application d'un filtre à une couche précédente.
- La classification au travers d'un réseau neuronal appelé couches interconnectées (« fully connected »).
- La couche de sortie correspondant à la réponse de l'algorithme. Cette réponse est donc la classe de l'image.

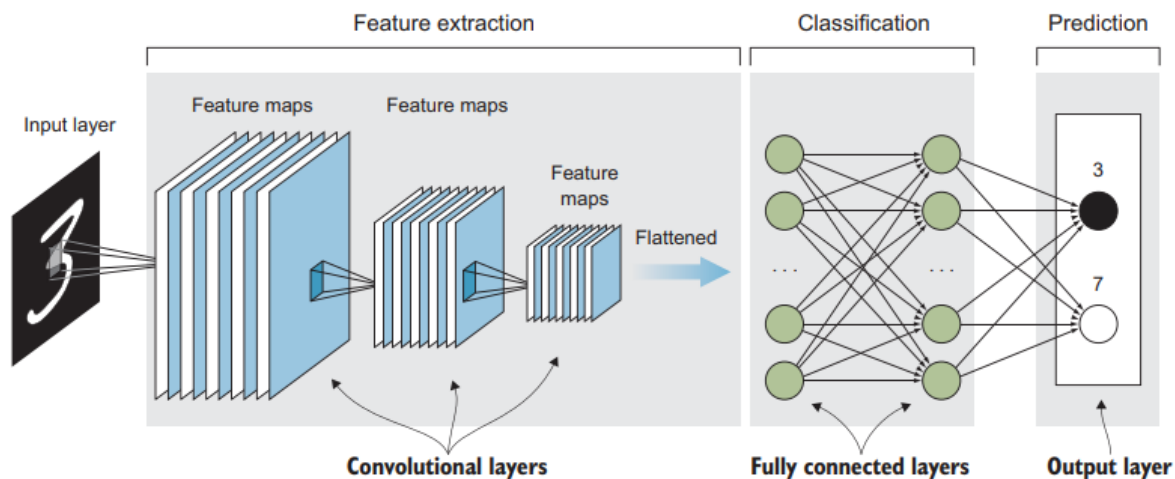


Figure 1 : Schéma d'un "Convolutional neural network" (Elgendy, 2020)

Entre chaque couche de convolution, les dimensions comprenant l'information sur l'image d'entrée seront réduites. Le nombre de "feature maps" correspondant à une couche d'informations sur l'image augmentera jusqu'à avoir une suite de "feature maps" à une dimension en sortie. Cette suite d'informations sera ensuite utilisée pour réaliser la classification de l'image au travers des couches "fully connected". L'information extraite de la classification de l'image sera la réponse de l'algorithme. Par exemple, dans le cas de la figure 1, le nombre trois sera l'information de sortie (Elgendy, 2020).

L'algorithme va donc catégoriser toutes les images du jeu de données d'entraînement et ensuite comparer les résultats avec les informations fournies par un humain, sous la forme d'annotations, pour calculer une erreur. Cette erreur est ensuite utilisée pour ajuster les paramètres du réseau durant une phase dite d'entraînement (minimisation de l'erreur). Une fois le modèle entraîné, il doit être validé sur un jeu de données, nommé "de validation". Ce dernier est composé d'images de même type mais n'étant pas connues par l'algorithme (Chollet, 2018).

### 1.1.3.2. CNN et reconnaissance d'espèces

De nombreux travaux de classification d'espèces par Deep learning ont été réalisés sur une multitude d'organes des plantes.

Les feuilles sont fréquemment utilisées dans la reconnaissance automatique de plantes (Barré et al., 2017; Grinblat et al, 2016; Lee et al., 2017).

Grinblat et al. (2016) ont optimisé un programme de CNN utilisant les caractéristiques de nervures des feuilles Barré et al., (2017) ont développé un algorithme CNN appelé LeafNet. Ils ont montré que les réseaux de neurones convolutifs procurent une meilleure représentation des caractéristiques d'image des feuilles par rapport à celles annotées à la main. Lee et al. (2017) ont mené une étude permettant l'analyse des caractéristiques des feuilles optimisant la reconnaissance par vision artificielle.

Les organes floraux ont été étudiés par Nilsback et al. (2008) et Gogul et Kumar (2017), les premiers ajoutant une localisation de la fleur sur l'image.

La technologie de reconnaissance par CNN a de même été utilisée dans le cadre de l'agriculture de précision dans le but d'optimiser la production en reconnaissant les différentes espèces cultivées (Abdullahi et al., 2017).

L'étude réalisée au cours de ce mémoire utilise l'écorce comme critère de reconnaissance, car elle est un attribut non saisonnier. De plus, la texture de l'écorce ne change pas sur de courtes périodes, le processus de vieillissement étant relativement lent (Ratajczak et al., 2019). La texture est donc observable toute l'année, indépendamment de la phénologie de l'arbre (Misra et al., 2020).

En addition, l'écorce est généralement facile d'accès contrairement à certains organes qui peuvent être inatteignables sur des arbres de grande hauteur (Misra et al., 2020). Toutefois, l'écorce est un critère pouvant être complexe pour une reconnaissance par l'œil humain. La faible différence entre les espèces et la haute variabilité au sein d'une même espèce rend la différenciation subtile (Misra et al., 2020). Généralement, la tranche au travers de l'aubier et les feuilles sont les deux critères privilégiés par les prospecteurs en région tropicale. L'étude de l'écorce devrait apporter un complément utile à une identification précise des espèces ligneuses.

### 1.1.3.3. Ecorce et CNN : état de l'art

L'application de réseaux de neurones convolutifs sur un jeu de données d'images d'écorces en milieu tempéré ou urbain a fait l'objet de trois études récentes.

Carpentier et al. (2018) ont étudié la reconnaissance d'espèces d'arbres par l'écorce sur le jeu de données BarkNet 1.0 contenant 23 000 images d'écorce pour 23 classes. Ce programme a démontré la faisabilité de la procédure via le Deep learning avec un taux d'exactitude correspondant à l'efficacité du modèle de 93,88% avec un choix unique et une exactitude de 97,81 % en utilisant la technique de vote à la majorité (la classe ayant obtenu le plus de voix est celle retenue).

Misra et al. (2020) ont utilisé un réseau de neurones convolutifs (patch-based CNNs) pour la reconnaissance d'espèces d'arbres par l'écorce. Le but de cette étude était de montrer que l'utilisation d'un outil d'intelligence artificielle offrait des performances plus élevées que les techniques de filtres anciennement utilisées. Le jeu de données utilisé était le Bark-101 (Tableau 1). Ce jeu de données n'étant pas suffisant, un redimensionnement a été réalisé sur les images durant le processus d'extraction des patchs afin de pallier le manque de données. Pour obtenir un choix de classification pour une image, l'image est divisée en sous-images. Une prédiction est émise sur chaque sous-image. La prédiction finale est déterminée par un vote majoritaire sélectionnant la classe ayant le nombre de prédictions le plus élevé (Misra et al., 2020).

Finalement, Kim et al. (2022) ont étudié la reconnaissance d’espèces d’arbres de la base de données Bark-Net 1.0 issue de la publication de Carpentier et al. (2018) complétée d’images d’espèces urbaines communes prises à Séoul en suivant le protocole de Carpentier et al. (2018). Il en résulte un jeu de données de 3672 images pour 20 espèces associé à un jeu de données de 5703 images pour 23 espèces. La question de recherche de cette étude était de comprendre comment les machines interprètent et distinguent les motifs des écorces. Pour se faire, deux algorithmes CNN associés à l’application d’une carte de chaleur ont permis l’identification de 90% des images. De plus, cette étude a démontré que le modèle moins complexe utilisait des caractéristiques plus générales et mieux adaptées, tandis que le modèle le plus performant, doté d’un nombre de couches beaucoup plus important, indiquait des caractéristiques plus spécifiques et moins pertinentes dans ce contexte.

#### 1.1.3.4. Bases de données existantes

La création d’un programme de Deep learning nécessite une base de données d’images conséquente pour permettre un entraînement performant et un taux de réussite élevé. La référence communément admise est un minimum de 1000 images par classe, à l’instar de la base de données ImageNet. Cette base de données est actuellement la plus importante base de données d’images labellisées et peut servir au pré-entraînement de modèles (Chollet, 2020 ; Deng et al., 2009).

L’un des principaux obstacles à la création d’outils de reconnaissance automatique est le faible nombre de bases de données existantes, de plus, peu sont en libre accès (Tableau 1) (Carpentier et al., 2018).

Tableau 1 : Base de données d’images d’écorces existantes (Carpentier et al.,2018 ; Misra et al., 2020 ; Kim et al., 2022)

Référence	Nom	Nombre de classes	Nombre d’images	Libre accès
Chi et al. (2003)	/	8	200	Non
Huang et al.(2006)	/	17	300	Non
Svab (2014)	Trunk12	12	393	Oui
Bressane et al.(2015)	/	5	540	Non
Blaanco et al.(2016)	/	23	920	Non
Fiel et Sablatnig (2011)	AFF	11	1183	Non
Carpentier et al.(2018)	BarkNet 1.0	23	23 000	Oui
Misra et al.(2020)	Bark-101	101	2587	Oui
Kim et al.(2022)	/	23	3672	Non

### 1.1.3.5. Reconnaissance d'espèces et amélioration de la technologie CNN

De nombreuses recherches ont été menées sur les CNN telle que celle de Figueroa et al. (2020) utilisant un “Convolutional siamese Network” qui est une combinaison de CNN. Mukherjee et al. (2017) et Wang et al. (2017) proposent aussi l'utilisation de modèles CNN pré-entraînés tels que GoogleNet et VGG-16.

Pour répondre au défi que représente la reconnaissance automatique de plantes en milieu naturel, l'environnement dans lequel évoluent les plantes doit être pris en compte. Un modèle à deux focales, nommé en anglais « Two-way attention model » utilisant le réseau CNN permet de répondre à cette problématique. Comme son nom l'indique, ce modèle se focalise sur deux éléments. En premier, il se focalise sur la taxonomie générale de la plante, et a comme objectif de reconnaître et labelliser la famille dont la plante est issue. En second, il associe ces informations à la reconnaissance des zones déterminantes (« features ») pour la classification par une carte de chaleurs (Zhu et al., 2019).

Il est finalement intéressant de noter que les techniques utilisées peuvent être combinées. Ben Ameer (2018) a en effet opté pour la méthode de fusion multimodale. Celle-ci permet de combiner plusieurs sources de données afin d'aider une intelligence artificielle dans des processus de reconnaissance automatique. Cette méthode conduit à des performances potentiellement élevées : les données issues des images de feuilles sont fusionnées avec celles des images d'écorce pour proposer à l'utilisateur une liste des espèces les plus probables (Ben Ameer, 2018).

Un exemple pertinent et reconnu de fusion multimodale est l'application de science participative « Plant@net », comprise dans le projet LifeCLEF. Sa base de données se compose de 14.756.033 images pour 37.499 espèces (Plantnet, 2022). Cette application qui compte plus de 90.000 utilisateurs facilite, par son accessibilité, l'alimentation de la banque d'images et l'accès à un réservoir de connaissances botaniques au grand public (Joly et al., 2016).

Cependant, cette base de données ne couvre pas tous les écosystèmes de manière égale. Les écosystèmes tropicaux ne sont que faiblement représentés, LifeCLEF s'appuie sur des données majoritairement localisées en Guyane ou en forêt amazonienne du nord. De plus, les performances de la reconnaissance automatique de leurs algorithmes sont plus performants sur les données européennes ou d'Amérique du nord que sur leurs données amazoniennes (Goëau et al., 2019).

L'ensemble de ces études ont été réalisées pour la grande majorité sur des espèces de forêts tempérées. L'utilisation de ces technologies dans un contexte tropical n'a, à notre connaissance, pas encore mené à des publications et mérite une investigation exploratoire. C'est l'objectif de la présente recherche.

## 1.2. Objectifs du travail

L'objectif principal de ce travail de fin d'études est la création et l'exploitation d'une base de données d'images d'écorces afin de reconnaître automatiquement, de manière la plus fiable possible, diverses espèces ligneuses tropicales. La démarche a comme finalité la création d'un outil original et efficace d'intelligence artificielle adapté à la problématique évoquée.

Il en découle trois objectifs secondaires qui représentent l'ossature du travail.

Premièrement, il s'agit de réaliser une base de données représentative d'images d'écorces d'espèces d'arbres tropicaux.

Deuxièmement, la robustesse et les caractéristiques de la base de données seront analysées. Les performances du programme de reconnaissance automatique permettront de valider ou non l'hypothèse selon laquelle l'écorce est un critère suffisamment discriminant pour la création d'un outil de reconnaissance automatique à partir du jeu de données. Les résultats de la classification permettront aussi de tester la robustesse du jeu de données, et d'évaluer s'il est nécessaire d'augmenter la variabilité ou la quantité d'images pour la réalisation d'un tel outil.

La troisième est d'optimiser les paramètres du modèle pour permettre la reconnaissance des espèces. La détermination de paramètres chiffrés sera aussi envisagée afin d'objectiver la qualité des identifications. L'optimisation des hyperparamètres du modèle sera présentée.

## 2. Matériel et méthode

### 2.1. Zone d'étude

#### 2.1.1. Localisation

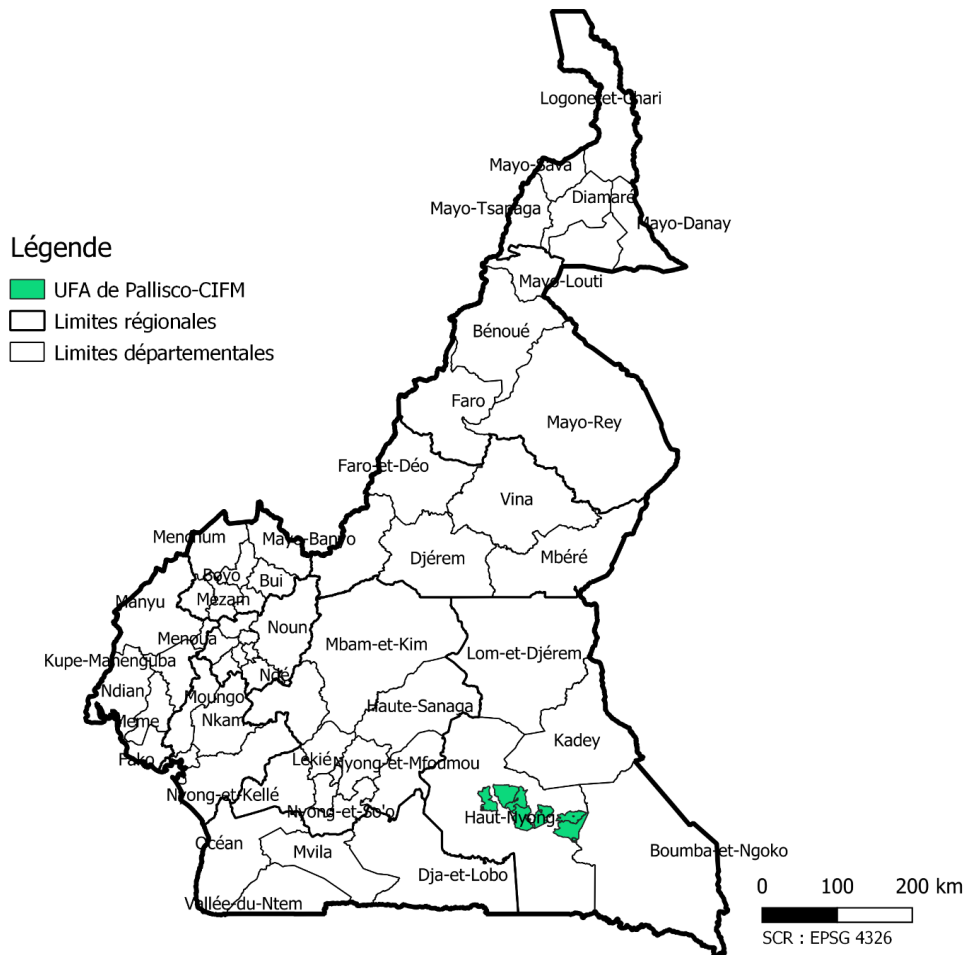


Figure 2 : Localisation de la société Pallisco-CIFM

La zone d'étude (Figure 2) est la concession forestière attribuée à la société Pallisco-CIFM localisée au Cameroun. La concession se situe au sein de l'arrondissement du Dja, dans la localité de Mindourou se trouvant dans le département du Haut-Nyong à l'Est du Cameroun. La société présente 7 unités forestières d'aménagement (UFA) couvrant une surface de 388 949 ha (Pallisco-CIFM, 2022).

### 2.1.2. Site de recherche

La zone de recherche est située au sein des UFA 41,42,44, 39, 30 et 31 (Figure 3). La zone principalement étudiée correspond à deux parcelles de 4 hectares de l'UFA 30-31. La parcelle au sud-ouest sera dénommée "Parcelle 1" et la seconde, au nord-est, "Parcelle 2" (Figure 3). Leurs coordonnées sont renseignées dans le Tableau 2. Le choix de ces parcelles est justifié par une identification et localisation préalable des arbres par Dupuis (2019).

Suite au nombre insuffisant d'arbres des espèces ciblées au sein des deux parcelles de 4 hectares, la zone de recherche a été étendue : parcours phénologiques, parcelles d'inventaires d'exploitation et le long des routes. Les zones où des arbres ont été photographiés sont représentées en vert (Figure 3).

Tableau 2 : Coordonnées des deux parcelles de 4 ha. Les coordonnées sont exprimées en mètres dans le système de coordonnées WGS 84/UTM zone 33N (EPSG : 32633) (Dupuis, 2019)

Coin de la parcelle de 4 ha	Coordonnées x ; y (m)	
	Parcelle 1	Parcelle 2
Sud-Ouest	427309,1 ; 362804,6	429266,3 ; 364715,6
Nord-Ouest	427308,6 ; 362992,4	429271,6 ; 364921,6
Sud-Est	427510,3 ; 362795,1	429468,1 ; 364716,0
Nord-Est	427500,3 ; 362981,8	429470,4 ; 364921,1



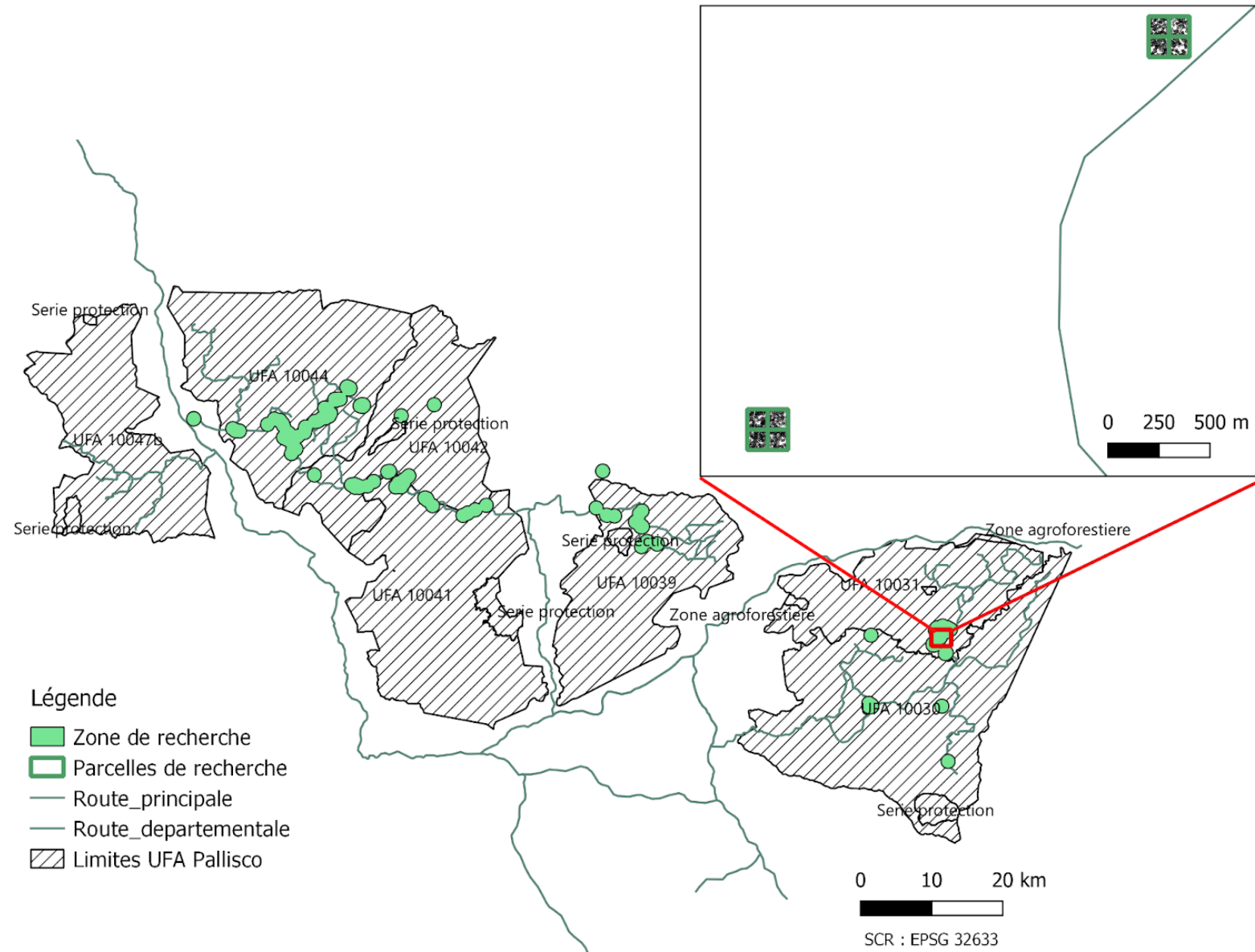


Figure 3 : Zone de recherche au sein de la concession avec un agrandi sur les deux parcelles de recherche principale

## 2.2. Facteurs écologiques liés à la zone d'étude

La concession forestière présente un relief moyennement vallonné. L'altitude varie de 600 à 800 m. En termes de dénivelé, les pentes sont généralement douces avec une succession de collines (Pallisco-CIFM & nature +, 2008).

La zone d'étude jouit d'un climat équatorial chaud et humide de type guinéen. Il présente 4 saisons dont une grande saison des pluies (GSP) et une petite saison des pluies (PSP) (Figure 4). Les précipitations moyennes annuelles à Mindourou sont de 1597 mm avec les précipitations les plus faibles de décembre à février (Pallisco-CIFM & Nature+, 2008).

La température moyenne de la région oscille autour de 24°C avec les moyennes mensuelles les plus élevées enregistrées en avril et les plus basses en juillet (Figure 4). Le rayonnement solaire présente des maxima en février et en novembre et des minima en janvier et en août (Feteke et al, 2017).

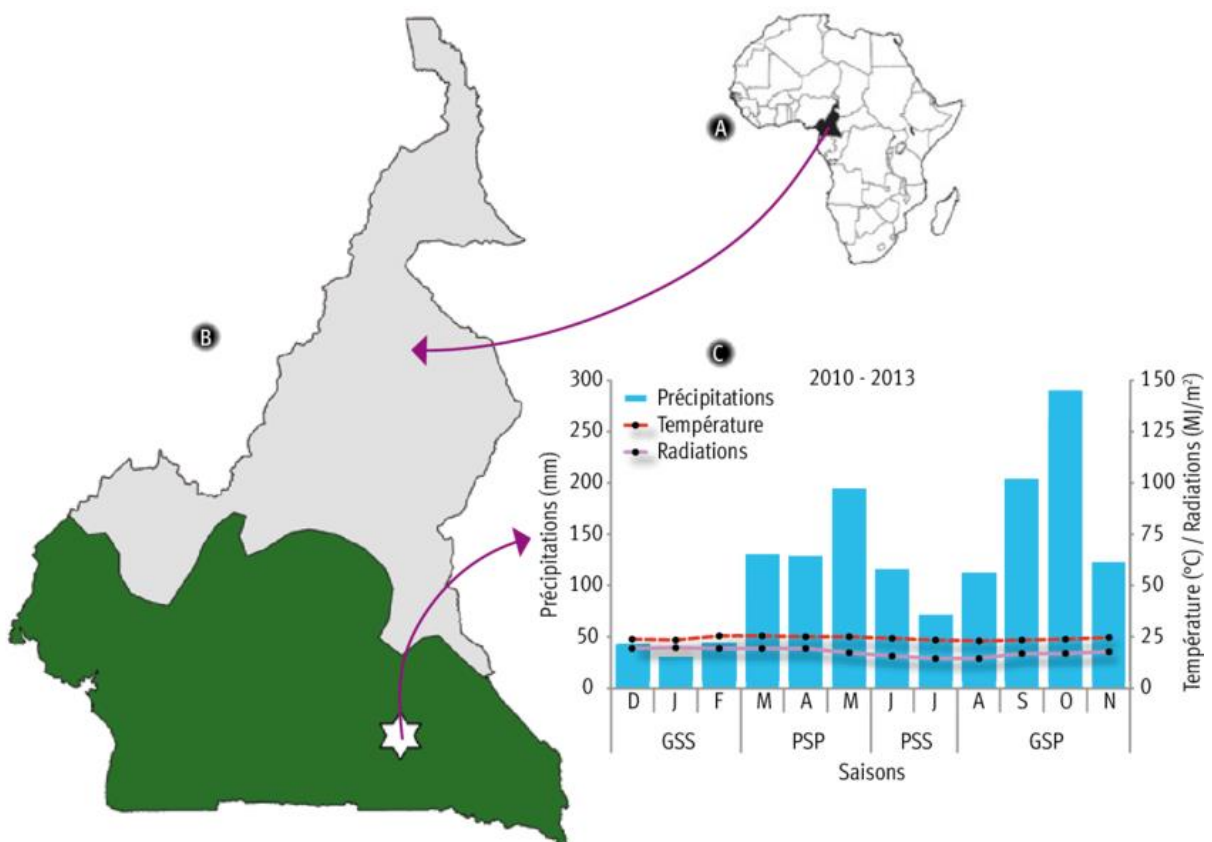


Figure 4 : Situation et diagramme ombrothermique de la zone d'étude au Cameroun.

Sur le plan géologique, la zone d'étude repose sur le bassin atlantique du Congo. Le substrat rocheux est composé de séries de gneiss, de micaschistes et de granites d'anatexie entrecoupées de séries intermédiaires de schistes chloriteux. Les sols sont très poreux, humides et meubles, pauvres en éléments minéraux, peu épais et plus ou moins décomposés (JMN, 2014).

Sur le plan pédologique, les sols rencontrés dérivent de roches métamorphiques. La terre ferme est principalement caractérisée par des sols ferrallitiques (JMN, 2014).

La zone d'étude présente une hydrographie dense combinant de nombreux petits cours d'eau temporaires en fonction des saisons et de plus grands cours d'eau permanents (Pallisco-CIFM & nature+, 2008). Les cours d'eau les plus importants au sein de la concession sont la Bomba, le Dja, le Me, l'Edjé, le Bek, le Ntam et le Médoum (Daïnou et al., 2017).

La concession se situe au sein de la région floristique guinéo-congolaise dans la forêt dense humide de transition entre les types sempervirent et semi-caducifolié (Fayolle et al., 2014 ; Feteke et al., 2017). La présence d'espèces typiques, de *Malvaceae* (*ex Sterculiaceae*) telles que le bété (*Mansonia altissima*), l'eyong (*Eribroma oblongum*) et l'Ayous (*Triplochiton scleroxylon*) et des *Cannabaceae* dont les *Celtis spp.*, permet de définir le couvert végétal comme présentant une forte affinité avec le type semi-caducifolié (Pallisco-CIFM & Nature+, 2008; Letouzey, 1986; Feteke et al.; 2017).

### 2.3. Espèces étudiées

Vingt espèces ont été sélectionnées sur base des trois critères suivants :

1. L'abondance au sein des parcelles pré-identifiées. Les espèces présentant les abondances les plus élevées ont été sélectionnées. En complément, des espèces commerciales ont été choisies.
2. La présence de caractéristiques botaniques reconnaissables (la tranche par exemple) ne nécessitant pas l'avis d'un botaniste. La facilité de reconnaissance botanique est primordiale, car des images d'espèces erronées fausseraient le modèle.
3. L'occurrence de critères morphologiques d'écorce distincts ou semblables à d'autres espèces étudiées. L'importance de ce critère réside dans le fait que la sélection d'espèces devait présenter un jeu de données diversifié avec des caractéristiques morphologiques dissemblables aussi bien que des espèces présentant des aspects difficilement différenciables à l'œil nu. L'illustration des différents types de morphologies d'écorces est disponible en Annexe 3.

Les caractéristiques d'écorce sont colligées dans le tableau 3 et sont illustrées par la Figure 5. Certaines espèces sont très aisément différenciables : *Pericopsis elata*, *Terminalia Superba* et *Entandrophragma cylindrum*. A l'opposé, *Polyalthia suaveolens* (*Greenwayodendron suaveolens*) et *Annickia affinis* sont difficilement discriminables l'une de l'autre.

Tableau 3 : Description morphologique de l'écorce des espèces étudiées ( Meunier et al, 2015 ; Wilk et Issembée, 2000; Prota4u, 2022; Vivien & Faure, 2001 ; Hallé, 1966)

Nom latin	Nom vernaculaire	Famille	Texture de l'écorce	Couleur de l'écorce
<i>Afrostryax lepidophyllus</i>	Ovita / Rondelle	Huaceae	Lenticellée se détachant par plaque	Brun-jaune
<i>Azelia bipindensis</i>	Doussié	Caesalpinoideae	Lenticellée avec des crêtes sinueuses	Rougeâtre à jaunâtre
<i>Alstonia boonei</i>	Emien	Apocynaceae	Lenticellée	Grisâtre à jaunâtre
<i>Annickia affinis</i>	Moambe Jaune	Annonaceae	Plissée horizontalement	Grisâtre à noirâtre
<i>Corynanthe pachyceras</i>	Akela	Rubiaceae	Lisse à légèrement écailleuse	Rosâtre ou ocre
<i>Desbordesia glaucescens</i>	Alep	Irvingaceae	Lisse à écailleuse	Grisâtre
<i>Entandrophragma cylindricum</i>	Sapelli	Meliaceae	Écailleuse, lenticellée et présentant de nombreuses dépressions	Grise à brune
<i>Irvingia gabonensis</i>	Andok	Irvingaceae	Lisse à écailleuse	Grise à gris-jaune
<i>Keayodendron bridelioides</i>	Abip élé	Phyllanthaceae	Écailleuse et fissurée	Brun grisâtre
<i>Meiocarpidium lepidotum</i>	Amvim	Annonaceae	Lisse présentant des écailles argentées	Brun-jaunâtre
<i>Pericopsis elata</i>	Assamela/ Afrormosia	Fabaceae	Écailles fines laissant des tâches rouges-brun	Crème à grisâtre
<i>Polyalthia suaveolens</i>	Otungui	Annonaceae	Présence de plis horizontaux	Grisâtre à noirâtre
<i>Pterocarpus soyauxii</i>	Padouk	Fabaceae	Se détache en fines écailles rectangulaires	Brun-grise à brun-rouge
<i>Rauvolfia macrophylla</i>	Essombi	Apocynaceae	Rugueuse et lenticellée	Gris-brun clair
<i>Strombosia pustulata</i>	Mbang mbazoa afum	Olacaceae	Lisse avec des écailles minces en plaques laissant des taches blanches	Gris clair à gris-brun
<i>Strombosiopsis tetrandra</i>	Edip mbazoa	Olacaceae	Lisse à rugueuse avec des lenticelles	Gris foncé
<i>Tabernaemontana crassa</i>	Etoan	Apocynaceae	Lisse à rugueuse à grosses lenticelles	Grisâtre à brunâtre
<i>Terminalia superba</i>	Limba/fraké	Combretaceae	Écailleuse chez les sujets âgés et lisses chez les jeunes sujets	Grise
<i>Trichilia welwitschii</i>	Ebegbemva osoé	Meliaceae	Fissurée longitudinalement	Brune
<i>Xylopia aethiopica</i>	Oyang	Annonaceae	Fissurée longitudinalement	Grisâtre à brunâtre





Figure 5 : Représentation des écorces relatives aux espèces étudiées

## 2.4. Création d'une base de données et d'un outil de reconnaissance automatique par CNN

Pour sa réalisation, cette étude a nécessité le matériel suivant :

- Un smartphone Samsung A72 pour la prise d'images ;
- Une tablette afin d'avoir les supports cartographiques à disposition ;
- Un GPS garmin 66S;
- Le livre de reconnaissance botanique "Les arbres utiles du Gabon" (Meunier et al., 2015).

Le travail a été divisé en deux parties :

1. La création d'une base de données d'images d'écorces. Celle-ci fut réalisée durant une période de 8 semaines avec l'appui d'Isaac Blaise Mbe, technicien de l'équipe de recherche scientifique de la société Pallisco.

2. L'établissement d'un modèle de Deep learning au départ de cette même base de données. Ce travail inclut le prétraitement et le nettoyage de la base de données, l'optimisation du modèle et l'analyse des résultats obtenus.

### 2.4.1. Création d'une base de données

#### 2.4.1.1. Prise en compte de la variabilité

L'objectif de la présente démarche consiste à établir une base de données d'images d'écorces en maximisant à la fois le nombre de clichés par classe et la variabilité intraspécifique dans le but de rendre robuste le système de reconnaissance.

#### **Caractéristiques morphologiques de l'écorce**

L'écorce, bien qu'étant un attribut non saisonnier peut varier en fonction de l'âge de l'arbre ou des conditions écologiques du milieu (Doucet, 2006).

L'âge des arbres peut être évalué au travers du diamètre de ceux-ci. En conséquence, avoir un jeu de données selon les différentes catégories d'âge pour la même espèce ligneuse rend la base de données plus robuste.

En forêt tropicale, certaines espèces peuvent présenter des diamètres très élevés dont les espèces suivantes : *Terminalia superba*, *Alstonia boonei*, *Entandrophragma cylindricum*. D'autres ont une gamme de diamètre plus restreinte à l'exemple de *Tabernaemontana crassa*. Il est donc utile de connaître la gamme de diamètre de chaque espèce au préalable (Tableau 4).

Il est aussi important de savoir que des données peuvent manquer pour certains individus. Cela peut être lié à la présence de contreforts à 1m30, à des diamètres trop élevés ou des déformations de l'arbre ne permettant pas de mesurer le diamètre.

Tableau 4 : Intervalle de diamètre en centimètre pour l'ensemble des espèces étudiées (Doucet communication personnelle, 2022 ; Meunier et al.,2015; PROTA4U, 2022 ; Vivien & Faure, 2011)

<b>Espèce</b>	<b>Classe de diamètre (cm) à 1m30 de</b>
<i>Afrostryax lepidophyllus</i>	[0 - 40]
<i>Afzelia bipindensis</i>	[0 -140]
<i>Alstonia boonei</i>	[0 -150]
<i>Annickia affinis</i>	[0 -90]
<i>Corynanthe pachyceras</i>	[0 -70]
<i>Desbordesia glaucescens</i>	[0 -150]
<i>Entandrophragma cylindricum</i>	[0 - + de 200]
<i>Irvingia qabonensis</i>	[0 -100]
<i>Keayodendron bridelioides</i>	[0 -100]
<i>Meiocarpidium lepidotum</i>	[0 -30]
<i>Pericopsis elata</i>	[0 -130]
<i>Polyalthia suaveolens</i>	[0 -70]
<i>Pterocarpus soyauxii</i>	[0 -150]
<i>Rauwolfia macrophylla</i>	[0 -90]
<i>Strombosia pustulata</i>	[0 -60]
<i>Strombosiopsis tetrandra</i>	[0 -120]
<i>Tabernaemontana crassa</i>	[0 -30]
<i>Terminalia superba</i>	[0 -150]
<i>Trichilia welwitschii</i>	[0 -60]
<i>Xylopia aethiopica</i>	[0 -75]

Le second facteur concerne les conditions écologiques du milieu. L'exposition lumineuse peut faire varier la couleur de l'écorce (Doucet, 2006). En effet, une écorce exposée à la lumière peut être plus lisse et plus claire.

Des conditions de luminosité variables ont été prises en compte par (i) une phase de terrain à cheval entre la saison sèche et la saison des pluies, (ii) la présence de zones ouvertes suite à l'exploitation.

### **Paramètres de prise d'images**

Augmenter le nombre d'images ne doit pas être synonyme d'une répétition de l'information. En conséquence, les images doivent présenter un minimum de recouvrement<sup>1</sup> afin que le modèle ne s'entraîne pas plusieurs fois sur la même information.

Les paramètres de prise d'images tels que l'angle de prise de vue, la distance au tronc ou la qualité de l'image peuvent de même influencer la variabilité du jeu de données.

### **Impact externe à la morphologie de l'écorce**

Sur l'écorce, peuvent cohabiter d'autres organismes tels que le lichen ou la mousse (Figure 6).

La mousse pourrait perturber la classification dans le cas où sa présence empêcherait la visibilité de la couleur et la texture de l'écorce.

Quant au lichen, sa présence est complexe à qualifier. Il peut être récurrent chez certaines espèces et de ce fait, être considéré lors de la description morphologique d'une espèce. A titre d'exemple, le tronc d'*Aucoumea Klaineana* est marbré de lichens (Doucet, 2006). « La composition et la richesse des lichens est fortement corrélées aux espèces d'arbres et aux caractéristiques de l'arbre hôte, telles que la structure de l'écorce et le diamètre des arbres. » (Benítez et al.,2019). Dans d'autres cas, sa présence n'est pas liée à l'espèce, il est donc occasionnellement présent et pourrait affecter la classification de l'image.

En revanche, divers artefacts sont susceptibles de diminuer les performances des modèles notamment : les dégâts sur les écorces ou des variations de luminosité sur même image (Figure 6).

Une proportion trop élevée d'images d'une même espèce présentant le même artefact se traduirait par un entraînement du modèle sur un caractère ne définissant pas l'espèce en elle-même. Cela engendrerait le risque de surentraînement.

---

<sup>1</sup> Le recouvrement correspond au fait que la même portion de l'écorce se situe sur plusieurs photos et donc la même information.



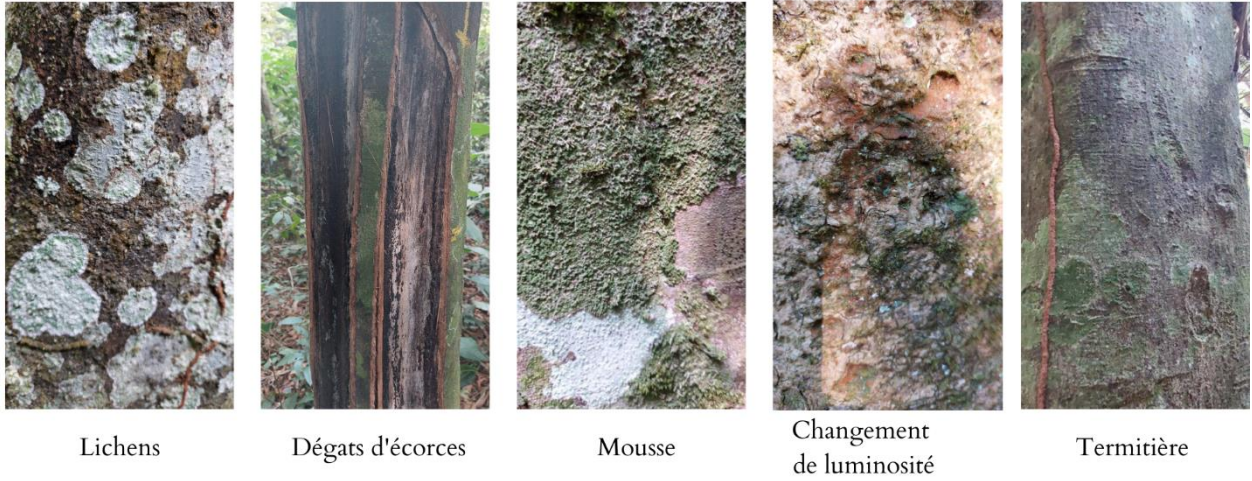


Figure 6 : Représentation des artefacts possiblement présents au sein de la base de données

Dans le but d'une compréhension exhaustive de l'analyse des artefacts, une nuance reste importante à préciser. Un individu, sur lequel est répertorié un artefact, n'aura pas ce même artefact sur l'ensemble de ses photos. Par exemple, une écorce écorchée ou une termitière ne sera observée que sur une portion réduite du tronc.

#### 2.4.1.2. Protocole de création d'une base de données mis en œuvre sur le terrain

Avant la mission de terrain, un protocole ainsi qu'une interface d'encodage à l'aide du logiciel OpenForis v.4.0.42 ont été réalisés.

Le protocole en quatre étapes est le suivant :

##### 1. Localisation des arbres

Au sein des deux parcelles pré-identifiées, un projet cartographique Qgis Field contenant les arbres d'intérêts a été élaboré sur base des données de Dupuis (2019). Il en est de même pour les arbres présents sur les suivis phénologiques ou sur les zones pré-inventoriées (Annexe 1 et 2).

##### 2. Identification botanique de chaque individu. L'identification est réalisée avec l'aide d'un membre de l'équipe scientifique de la société forestière et en se basant sur l'ouvrage de Meunier et al. (2015). L'ensemble des identifications étaient réalisées à l'aide des caractéristiques de tranche et des feuilles (Figure 7,8 et 9).

##### 3. Description de l'individu à l'aide de l'application Openforis Collect v.4.0.42. Elle permet de mettre en place une interface recensant les informations nécessaires. La description comprend le recensement de l'ensemble de ces informations dans un ordre chronologique précis :

- a. Le code unique de l'individu
- b. L'espèce
- c. L'heure et la date
- d. La mesure du diamètre (Figure 10).
- e. La description des artefacts.
- f. La couleur et la texture de l'écorce
- g. Les coordonnées GPS



Figure 7 : Feuille de *Rauvolfia macrophylla* dont le nombre de nervures permet sa différenciation vis à vis de *Rauvolfia vomitoria*



Figure 9 : Tronc et houppier de *Terminalia superba*



Figure 8 : Tranche réalisée dans le but de reconnaissance d'*Alstonia boonei* grâce à son latex



4. La prise d'images de l'écorce de l'individu (Figure 11<sup>2</sup>). Un maximum d'images est pris par arbre à différentes hauteurs.



Figure 10 : mesure du diamètre d'un individu

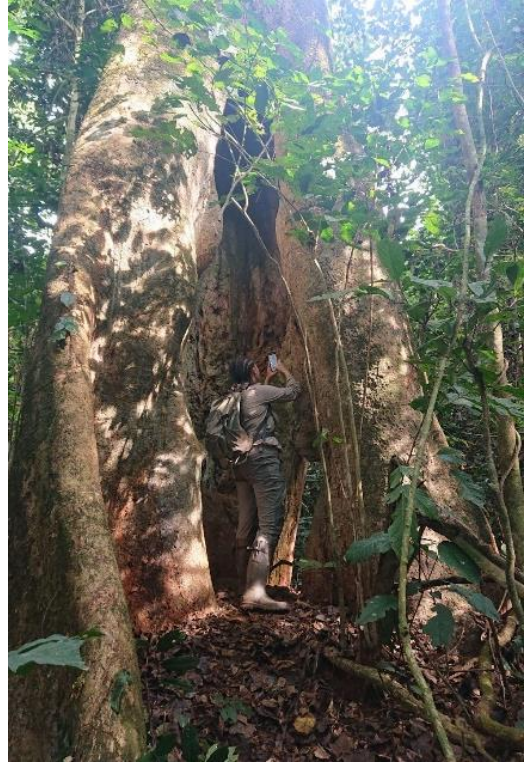


Figure 11 : prise d'image d'un individu

Les tranches dans l'écorce ou du flou ont été conservées au sein du jeu de données afin d'introduire une certaine variabilité dans les prises de vue. L'idée est d'assurer la robustesse du procédé en considérant l'ensemble des prises d'images qu'un opérateur externe pourrait réaliser.

---

<sup>2</sup> L'espèce présente sur l'image, en l'occurrence *Triplochiton scleroxylon*, ne fait pas partie de la présente étude

### 2.4.1.3. Base de données

Suite à cette mission de terrain, une base de données de **56 324 images** pour 20 classes a été constituée. Par classe, le nombre d'images photographiées va de 1830 à 5754 et le nombre d'individus varie entre 61 et 235 (Tableau 5).

Tableau 5 : Nombre d'individus et d'images par classe

<b>Espèce</b>	<b>Acronyme</b>	<b>Nombre</b>	<b>Total</b>
<i>Afrostryrax lepidophyllus</i>	AFLE	67	1858
<i>Afzelia bipindensis</i>	AFBI	62	2608
<i>Alstonia boonei</i>	ALBO	64	2851
<i>Annickia affinis</i>	ANAF	117	3314
<i>Corynanthe pachyceras</i>	COPA	84	2364
<i>Desbordesia glaucescens</i>	DEGL	71	2469
<i>Entandrophragma</i>	ENCY	68	3191
<i>Irvingia gabonensis</i>	IRGA	68	2140
<i>Keayodendron bridelioides</i>	KEBR	61	1945
<i>Meiocarpidium lepidotum</i>	MELE	112	2298
<i>Pericopsis elata</i>	PEEL	68	2499
<i>Polyalthia suaveolens</i>	POSU	235	5754
<i>Pterocarpus soyauxii</i>	PTSO	88	2700
<i>Rauwolfia macrophylla</i>	RAMA	60	1834
<i>Strombosia pustulata</i>	STPU	143	4159
<i>Strombosiopsis tetrandra</i>	STTE	65	1830
<i>Tabernaemontana crassa</i>	TACR	182	5025
<i>Terminalia superba</i>	TESU	62	3151
<i>Trichilia welwitschii</i>	TRWE	106	2279
<i>Xylopia aethiopica</i>	XYAE	67	2055
Total		1850	56324

Cette base de données présente une gamme de diamètre allant de 7 à 156 cm. Une majorité des individus étudiés possède un diamètre situé entre 10 et 20 cm. Les individus localisés dans les Parcelles 1 ou 2, dont le diamètre n'a pas pu être mesuré, ont été estimés à l'aide des données de Dupuis (2019) (Figure 12).

Trois niveaux de variabilité de diamètre ont été définis suite à l'analyse de la répartition du diamètre pour chaque espèce.

1. La répartition du diamètre des différents arbres est très variée pour les espèces : *Azelia bipindensis*, *Annickia affinis*, *Desbordesia glaucescens*, *Keayodendron bridelioides*, *Pericopsis elata*, *Polyalthia suaveolens*, *Strombosiopsis tetrandra* (Annexe 7).
2. La répartition est partagée entre différentes classes de grandeur avec plus de 30% des individus appartenant à une classe dominante au sein des espèces suivantes, *Afrostryax lepidophyllus*, *Xylopia aethiopica*, *Corynanthe pachyeras*, *Entandrophragma cylindricum*, *Strombosia pustulata* (Annexe 7).
3. Le diamètre des fûts est très peu variable pour les espèces : *Meiocarpidium lepidotum*, *Rauwolfia macrophylla*, *Irvingia gabonensis*, *Trichilia welwitschii*, *Tabernaemontana crassa* (Annexe 7).

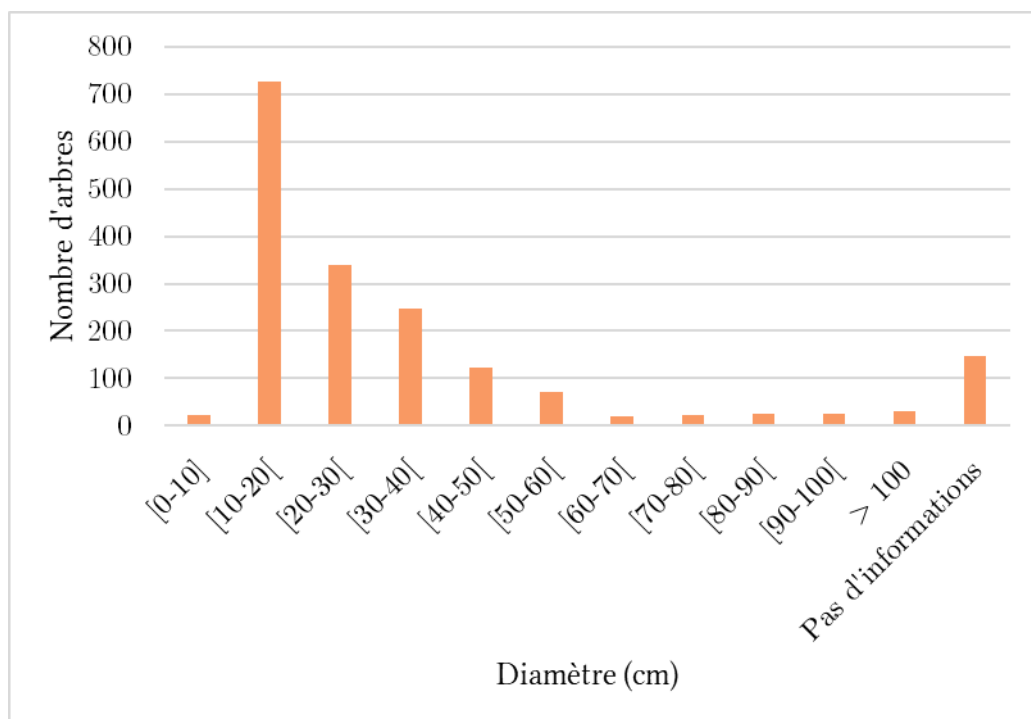


Figure 12 :Histogramme de répartition des classes de diamètre au sein du jeu total de données

Des mousses sont observées sur 40 à 90 % des arbres et des lichens sur 40 à 97% d'entre-eux (Figure 13). L'occurrence de ces organismes génère dès lors une variabilité des images pour des arbres d'une même espèce.

Des termitières sont de même récurrentes sur un grand nombre d'individus (en moyenne 61% par espèce). Quant aux dégâts d'écorces, leur impact est négligeable (Figure 13).

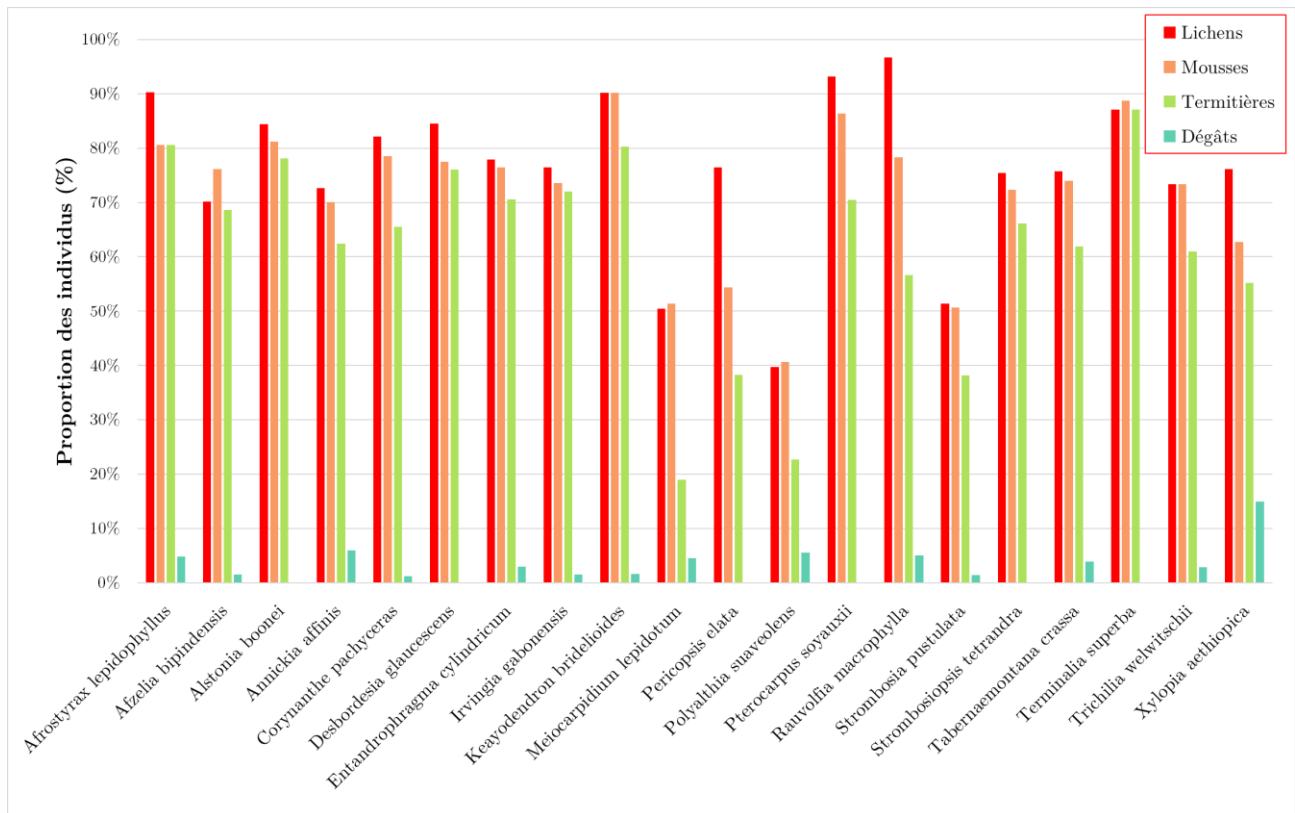


Figure 13 : Histogramme de la proportion des individus présentant un artefact sur l'ensemble des individus d'une même classe

Pour information, la distribution géographique des individus de chaque espèce au sein de la concession forestière est en Annexe 6.

Finalement, le jeu de données a été labellisé à l'aide des heures et dates de prise d'images. Cependant, la base de données initiale contenait des images ayant une taille trop élevée. De ce fait, le modèle prenait un temps d'entraînement conséquent pour des performances faibles. Dès lors, le nombre de pixels a été divisé par 8. Après divers essais, cette valeur (facteur 8) a été sélectionnée car elle permettait d'offrir une qualité d'images pour laquelle les caractéristiques de toutes les écorces restaient toujours bien discriminées tout en ayant un nombre d'octets plus faible (Figure 14).



Image resize 8  
Taille 44 ko



Image initiale  
Taille 3.08 Mo

Figure 14 : Comparaison d'une image avant et après réduction d'un facteur 8 du nombre de pixels

## 2.4.2. Développement d'un outil d'intelligence artificielle

La méthodologie utilisée pour développer l'outil d'intelligence artificielle comprend trois axes centraux et interconnectés (Figure 15) .

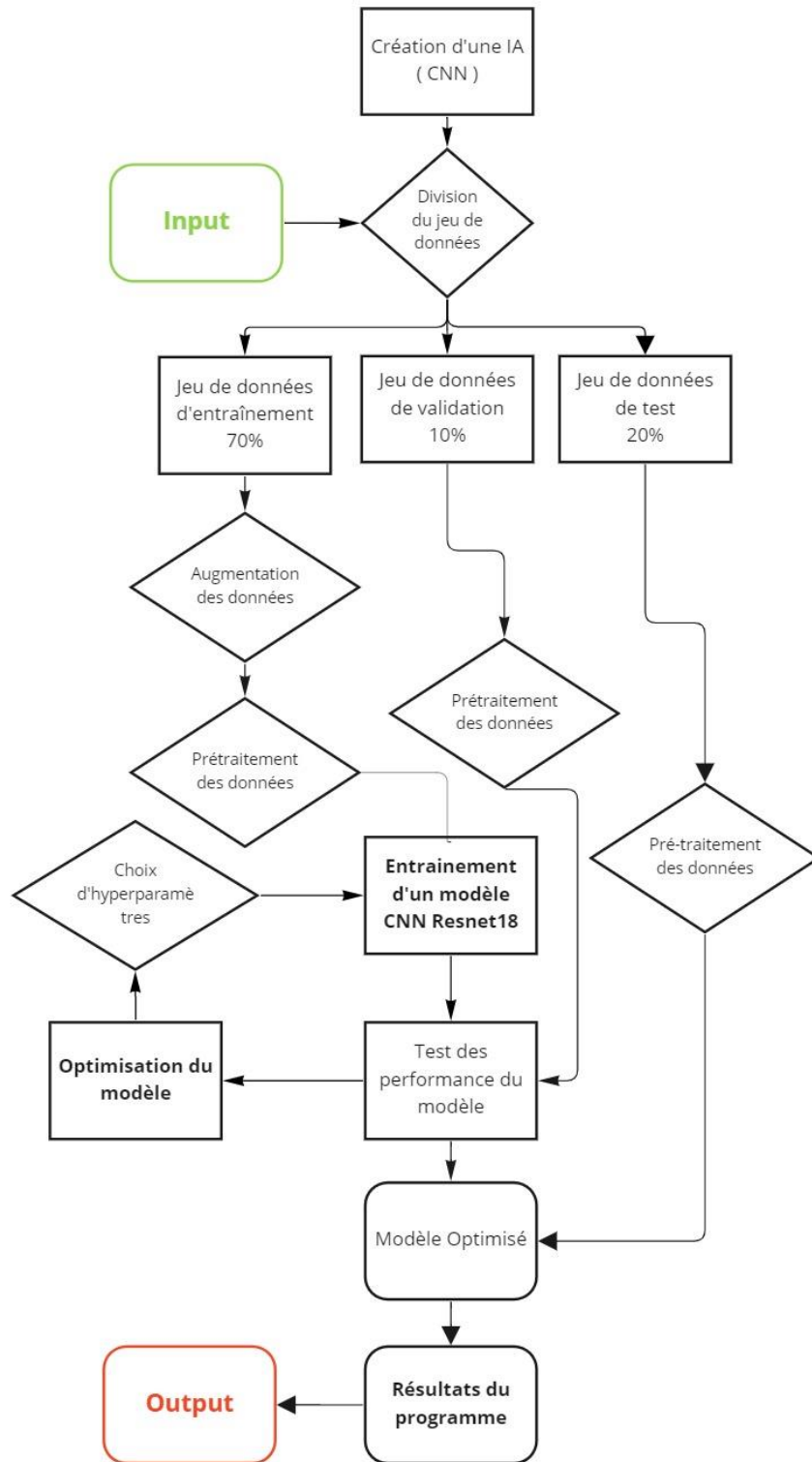


Figure 15 : Organigramme des étapes de création de l'outil d'intelligence artificielle



La première phrase comprend la paramétrisation et l'entraînement du modèle. Cette phase nécessite un jeu de données conséquent appelé jeu de données d'entraînement (Chapitre 4.4.2.2).

En second lieu, intervient la phase d'optimisation. Elle consiste à réitérer une multitude de fois les opérations de paramétrisation et d'entraînement de la phase 1 en testant divers paramètres. A chaque répétition, une analyse de l'impact de ceux-ci sur la performance de classification du modèle est réalisée sur un jeu de données dit « de validation ». Cette phase est réalisée jusqu'à l'obtention d'un modèle final présentant l'exactitude la plus élevée (Chapitre 4.5.2.7).

Une fois un modèle performant obtenu, ses performances sont mesurées au départ du jeu de données de test (Chapitre 4.5.2.8).

L'ensemble de la programmation a été réalisée à l'aide du langage de code Python version 3.7.1. et des packages : *Albumentations* 1.2.1, *matplotlib* 3.4.1 , *numpy* 1.21.5, *opencv* 4.5.5.64, *os* 0.1.4, *pandas* 1.3.5, *Pillow* 9.0.1, *random*, *scikit-learn* 1.1.1, *time* 0.3.0, *torch* 1.11.0, *torchvision* 0.12.0 et *wandb* 0.12.17.

Pour information, le terme classe a été systématiquement utilisé dans la suite de ce travail pour définir un genre et une espèce d'arbre. L'usage de ce terme est judicieux car c'est le terme utilisé dans le contexte de l'intelligence artificielle.

#### 2.4.2.1. Division du jeu de données

Le jeu de données a été divisé en trois parties respectivement pour l'entraînement (70%), la validation (10%) et le test (20%) du dispositif.

Afin d'entraîner au mieux le programme, les deux conditions suivantes ont été associées à la division du jeu de données :

1. La division a été réalisée en respectant les proportions de chaque classe pour empêcher un surentraînement des classes surreprésentées par leur nombre d'images.

2. Un même arbre ne peut pas se retrouver au sein de différents jeux de données. En effet, les images d'un même individu présentent une faible variabilité. De ce fait, entraîner des données et tester les performances du modèle sur le même individu serait un facteur de biais.

### 2.4.2.2. Entraînement du modèle

L'entraînement d'un algorithme de reconnaissance automatique est un processus itératif se déroulant au cours d'un nombre d'époques définies. Dans le cas présent, l'époque est le laps de temps imparti au passage de l'ensemble du jeu de données d'entraînement dans le modèle. Lors d'une époque, les images vont être envoyées par lots aussi appelés « batches » (4, 8, 16 ou 32 images) au modèle et cela, de façon aléatoire. À la suite du passage d'un lot de données, les paramètres du réseau sont ajustés par un **optimiseur** qui ajuste les paramètres du réseau (poids et biais) de manière à minimiser **la fonction de coût** lors de l'apprentissage (Figure 16). La fonction de coût quantifie à quel point le modèle est distant de la solution correcte. Plus le coût est faible, plus la classification du modèle est performante. Quant à l'optimiseur, il quantifie la différence entre deux distributions de probabilités. Dans le cas de cette étude c'est l'entropie croisée qui est utilisée. Elle est définie à l'équation 1.

$$E(W, b) = - \sum_{i=1}^m \hat{y}_i \log(\rho_i) \quad \text{Entropie croisée(1)}$$

Où  $W$  est le poids,  $b$  est le biais,  $y$  est la probabilité cible,  $p$  est la probabilité de la prédiction et  $m$  est le nombre de classes.

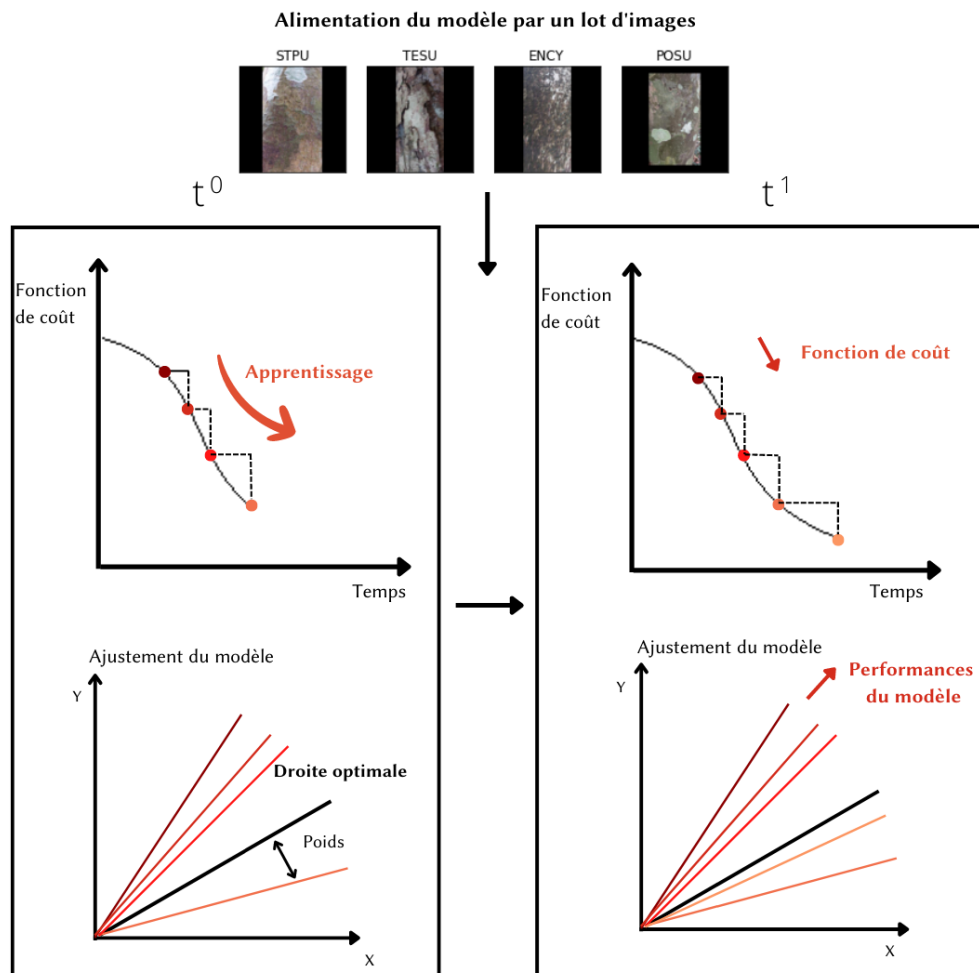


Figure 16 : Représentation de l'ajustement du modèle par la descente du gradient de la fonction de coût. Les graphiques du haut sont l'évolution de la fonction de coût en fonction du temps et ceux du bas sont l'optimisation de la droite représentant le modèle

Cet optimiseur calcule un poids qui représente la distance au modèle cible (100% des prédictions correctes). La droite du modèle sera ensuite ajustée selon l'évolution d'une fonction de coût (Delplanque, 2020; Elgendy, 2020). L'ajustement du modèle réel est plus complexe que l'exemple mentionné, celui-ci est simplifié dans le but d'une meilleure compréhension du processus.

Ce processus s'ajustant après chaque lot d'images, les espèces présentant le plus d'images tendent à être plus entraînées que le reste des classes rendant ainsi la reconnaissance des classes minoritaires plus complexe. Pour répondre à cette problématique, deux traitements ont été appliqués en amont de l'entraînement : l'augmentation des données et la pondération des classes (Delplanque, 2020).

#### 2.4.2.3. Augmentation des données

L'augmentation des données est une pratique courante lors de l'utilisation de CNN. Elle consiste en la création de données artificielles en modifiant des paramètres des images présentes au sein du jeu de données initial (Elgendy, 2020).

Le jeu de données initial a été augmenté en ajoutant des rotations et des retournements des images à angles aléatoires. Chaque traitement a été appliqué à 20% du jeu de données d'entraînement de manière aléatoire. Ce choix a été réalisé arbitrairement mais le nombre d'images sélectionné reste important, de manière implicite cela apporte les éléments de variabilité souhaités. La robustesse du système s'en trouve renforcée.

Les transformations ont été réalisées à l'aide de la librairie « Albumentation ». En relation avec cette librairie Python, un site internet permet la visualisation de l'impact des traitements sur une image (Albumentation, 2022). La représentation de l'interface de l'outil est disponible en Annexe 4.

Le choix des transformations a été fait au travers de cette même interface. Elles correspondent à des angles de prise de vue différents, tel qu'un opérateur extérieur le ferait.

#### 2.4.2.4. Pondération des classes

La pondération des classes est un outil d'équilibrage des données, elle améliore la classification des classes minoritaires en augmentant leur importance pour le modèle. En d'autres termes, l'attribution d'un poids plus important aux classes sous-représentées et plus faible aux classes sur-représentées (He et al., 2012).

#### 2.4.2.5. Prétraitement des données

Lorsqu'on travaille avec des images en 2D, les données sont prétraitées afin de correspondre à un objet mathématique à plusieurs dimensions appelé tenseur que le réseau neuronal peut traiter et interpréter. De ce fait, les valeurs matricielles de chaque image sont normalisées et transformées en tenseur. La normalisation correspond à soustraire à la valeur de chaque pixel,

la valeur moyenne des pixels de l'image et à diviser cette valeur par la division standard (Elgendy, 2020).

Au cours de ces prétraitements, aucun recadrage des images n'a été opéré pour deux raisons. Premièrement, le but de ce programme était de simuler au maximum les conditions d'un utilisateur externe. Deuxièmement, Carpentier et al. (2018) démontrent qu'il est plus pertinent en termes de variabilité d'avoir un plus grand nombre d'images le long du tronc que d'avoir une même photographie divisée en plusieurs sous-images (« crop »). Cette division ne permettrait donc que de faibles perspectives d'augmentation de l'apprentissage du modèle.

#### 2.4.2.6. Entraînement du modèle

Le jeu de données d'entraînement est utilisé pour alimenter le modèle de classification. Dans le cadre de cette étude, le réseau de classification est composé d'une couche interconnectée correspondant à un premier traitement de l'information, associée à un Resnet 18 préentraîné sur ImageNet. Le Resnet18 est un modèle créé par He et al. en 2016. Sa spécificité est d'être un modèle simple ayant été optimisé dans le but d'augmenter les performances de reconnaissance et éviter la saturation du réseau lors de son entraînement (He et al, 2016). La capacité de cet algorithme à différencier des espèces d'arbres à l'aide d'images d'écorce a été démontrée empiriquement par Carpentier et al. (2018).

D'un point de vue pratique, une première étape du travail était la prise en main de la technologie CNN et la compréhension de ses composantes. En un second temps, une analyse exploratoire a été effectuée jusqu'à la création de l'algorithme permettant la reconnaissance d'espèces tropicales par l'écorce. Elle consistait en l'entraînement de modèles jusqu'à ce qu'un modèle présente une fonction de coût décroissante avec des valeurs cohérentes (entre 0 et 1).

Lors de cette étape, le temps de traitement a été pris en compte. Par exemple, le premier modèle créé nécessitait plus d'une semaine d'exécution, il aurait été impossible de tester l'ensemble des hyperparamètres avec un temps d'exécution si conséquent.

Lorsque le code a permis la reconnaissance d'écorce en forêt tropicale, la phase d'optimisation des hyperparamètres du modèle a commencé.

#### 2.4.2.7. Optimisation des modèles

En vue de sélectionner le modèle le plus performant, les hyperparamètres de plusieurs modèles ont été ajustés. L'ajustement est réalisé par rapport aux performances du modèle sur le jeu de données de validation. Celles-ci sont quantifiées à l'aide des valeurs d'exactitude.

L'exactitude correspond à la proportion d'images correctement identifiées par le modèle sur l'ensemble des images du jeu de données. Cette métrique est calculée grâce au calcul d'une prédiction et un score de confiance allant de 0 à 1 pour chaque image. Seule la classe présentant le score de confiance le plus haut est prise en compte pour définir si le modèle a rendu une prédiction correcte ou erronée.

Les paramètres suivants ont été testés :

- La taille du lot
- Le taux d'apprentissage
- Les paramètres du 'scheduler' (composante de l'optimiseur) : ils correspondent aux modalités de l'évolution du taux d'apprentissage. Les modalités peuvent évoluer en fonction de l'exactitude maximale ou de la perte minimale.

#### 2.4.2.8. Analyse du modèle final

Les performances du modèle optimisé ont été mesurées sur base du jeu de données de test. L'analyse des métriques suivantes a permis l'étude plus en détails de la qualité de classification du modèle.

La matrice de confusion décrit les performances interclasses d'un modèle. Ce tableau met en relation les prédictions du modèle et les labels corrects, classe par classe (Figure 17).

		Prédictions du modèle		
		A	B	C
Labels corrects	A	Vrai positifs		
	B		Vrai positifs	
	C			Vrai positifs

Figure 17 : Représentation d'une matrice de confusion

Le rappel<sup>3</sup> permet de quantifier le nombre de prédictions correctes sur l'ensemble des prédictions d'une même classe (Elgendy, 2020). Plus la valeur du rappel est élevée, plus le taux de faux positifs est bas. La formule 2 ci-dessous et l'exemple du rappel de la classe *Afzelia bipindensis* schématisent cette métrique (Figure 18).

La précision permet de quantifier la proportion d'images correctement identifiées sur l'ensemble des images d'une même classe théorique (Elgendy, 2020). Une valeur de précision élevée prédit un taux de faux négatifs faibles. Il est défini dans la formule 3. Par exemple, la précision de la classe *Afzelia bipindensis* est représentée en figure 18.

Finalement, le F-score donne une représentation globale des performances du modèle en associant les résultats de rappel et de précision (Formule 4)(Elgendy, 2020).

Pour le calcul des valeurs de rappel et précision, les opérations suivantes ont été réalisées :

1. Division des résultats globaux de prédictions et probabilités du modèle en table de résultats pour chaque classe,
2. Calcul des valeurs des vrais positifs (TP) et des faux positifs (FP) pour chacune des classes (Figure 18),
3. Accumulation des données de TP et FP depuis l'image présentant la plus faible précision à la plus élevée,
4. Utilisation de ces valeurs pour le calcul des valeurs de rappel, précision et F1-score. Ces paramètres sont calculés de la manière suivante selon (Elgendy, 2020) :

$$R = \frac{\sum_{i=1}^n TP[i]}{\sum_{i=1}^n (TP[i] + FN[i])} \quad \text{Rappel(2)}$$

$$P = \frac{\sum_{i=1}^n TP[i]}{\sum_{i=1}^n (TP[i] + FP[i])} \quad \text{Précision(3)}$$

$$F = \frac{2 * P * R}{P + R} \quad \text{F1-score(4)}$$

Où R est le rappel, P est la précision, F est le F1-score, n est le nombre d'images et i est une image.

---

<sup>3</sup> Le « rappel » correspond à la traduction française du terme anglais « recall ».

		Prédictions du modèle	
		Afzelia bipindensis	Autres Classes
Labels corrects	Afzelia bipindensis	Vrais positifs	Faux négatifs
	Autres Classes	Faux positifs	Vrais négatifs

$$\text{Rappel} = \frac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux positifs}}$$

$$\text{Précision} = \frac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux négatifs}}$$

Figure 18 : Schématisation du rappel et de la précision avec l'exemple de l'espèce *Afzelia bipindensis*.

### 2.4.3. Analyse des causes d'erreurs du modèle

#### 2.4.3.1. Analyse des images

Suite au développement du modèle, les images incorrectement reconnues ont été analysées afin de mieux comprendre les points faibles de la base de données et d'envisager des améliorations. L'occurrence des artefacts suivants ont été recherchés :

- Le flou de l'image
- La présence de lichens sur plus de 50 % de l'image
- La présence de mousse sur plus de 50% de l'image
- Le mélange de mousse et de lichens sur plus de 50% de l'image
- La présence d'une tranche
- La présence d'un arrière-plan
- La présence de termitière
- La luminosité de l'image

Ces caractéristiques ont été analysées à l'œil nu, image par image. Ensuite, la proportion d'images par classe contenant un artefact a été quantifiée de même que la proportion de chacun des artefacts séparément.

#### 2.4.3.2. Extraction des cartes de chaleur des images erronées

L'image est définie par un « code » correspondant à ses informations extraites par le réseau de neurones convolutifs comme expliqué en paragraphe 1.1.3. Ce code comprend des zones de focalisation utilisées par le modèle pour la classification. La visualisation de ces zones peut être traduite au travers d'une carte de chaleur. Celle-ci présente un facteur d'importance du pixel pour la classification, il est échelonné de 0 à 1. La figure 19 représente cette carte dont la couleur du pixel varie du blanc (0) au rouge (1) en fonction de son importance.

Ces cartes de chaleur ont été utilisées pour analyser les facteurs d'erreurs. En effet, il est opportun de vérifier l'hypothèse du potentiel impact de ces artefacts.

La programmation nécessaire à la création de ces cartes a été réalisée à l'aide de la librairie capsun version 0.5.0.

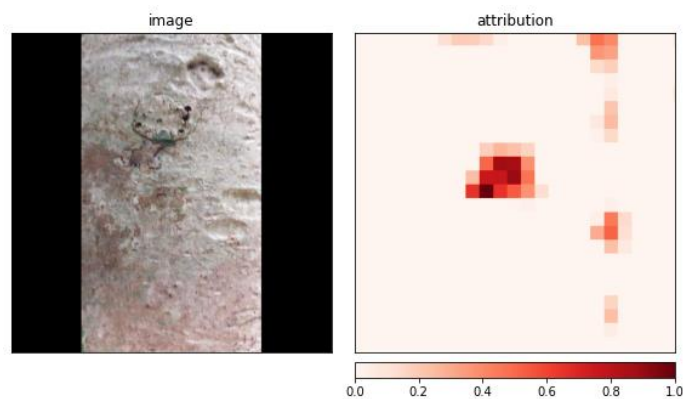


Figure 19 : Représentation d'une carte de chaleur d'une image d'écorce



## 2.5. Contribution personnelle

Dans le but de mener à bien les objectifs de ce travail, l'étudiante en charge de ce mémoire a réalisé un protocole de terrain, elle a ensuite effectué une mission de terrain au Cameroun afin de créer la base de données nécessaire à ce travail. A la suite de son retour en Belgique, la base de données a été nettoyée et labellisée. Finalement, sous l'encadrement de ses promoteurs, un programme d'intelligence artificielle a été mis en œuvre et optimisé afin de répondre à l'objectif de recherche.

Les différentes phases méthodologiques ont respectivement nécessité le temps de travail suivant de la part de l'étudiant :

1. La phase de terrain a duré 8 semaines à raison de 5 jours en forêt par semaine. Les deux jours restants étaient consacrés à la gestion des données.
2. Le nettoyage et la labellisation de la base de données a nécessité 3 semaines.
3. La création du modèle et de ses résultats a pris 7 semaines.
4. L'analyse des sources d'erreurs a duré une semaine.

Méthodologiquement, il était indispensable d'achever une phase avant de démarrer la suivante.

## 3. Résultats

### 3.1. Création d'un outil de reconnaissance automatique

Suite à l'étape de programmation, un outil de reconnaissance automatique d'espèces tropicales a été créé. Le modèle le plus performant, avec 85,5% d'exactitude sur le jeu de données de validation, a été sélectionné. Il présente les hyperparamètres suivants :

- Il est alimenté par lots de 8 images,
- Il est entraîné durant 50 périodes,
- L'optimiseur est une descente de gradient stochastique (SGD). Son taux d'apprentissage est de  $10^{-3}$  et diminue d'un facteur de 0.1 suite à 4 valeurs d'exactitude non maximales,
- Le moment est de 0.9.

Son temps d'entraînement est de 25 heures.

Le code source du programme est disponible dans un environnement en libre accès : [marjanekdr/TFE M Kaddouri: Code source du développement d'un outil de reconnaissance automatique des espèces ligneuses tropicales \(github.com\)](https://github.com/marjanekdr/TFE_M_Kaddouri)

### 3.2. Performance du modèle développé

Le modèle possède une exactitude évaluée à 81,3% sur le jeu de test. Le rappel, la précision et le F-score moyens du modèle sont respectivement de 81,3%, 81,9% et 81,4% (Tableau 6).

Concernant les résultats pour chaque espèce d'arbre, la matrice de confusion permet une première description de la performance du modèle (Figure 20). L'ensemble des classes sont souvent correctement identifiées mais à des proportions diverses. Cette observation est confirmée à l'aide des valeurs de rappel pour chaque espèce qui s'échelonnent entre 65% et 96%. Les espèces les mieux reconnues (F-score > 90%) sont *Pericopsis elata*, *Tabernaemontana crassa* et *Pterocarpus soyauxii*.

La matrice de confusion permet ensuite, d'analyser les risques de confusion inter-classes.

Le modèle présente de la confusion inter-classes. Entre 0 et 87 images de la même classe peuvent être prédites de manière erronée, comme appartenant à une autre classe. La confusion peut être relative à une classe unique ou à plusieurs autres classes en proportion variable.

Le cas de l'espèce *Polyalthia suaveolens* présente un exemple intéressant de confusion élevée. En effet, 3 espèces ont une majorité de leurs images mal classifiées reconnues à tort comme *Polyalthia suaveolens*. En particulier, la confusion entre *Polyalthia suaveolens* et *Annickia affinis* est marquée. 52 images (4,5%) de *Polyalthia suaveolens* sont reconnues comme *Annickia affinis* et 87 images (13%) de *Annickia affinis* sont reconnues comme *Polyalthia suaveolens* par le modèle (Figure 20).

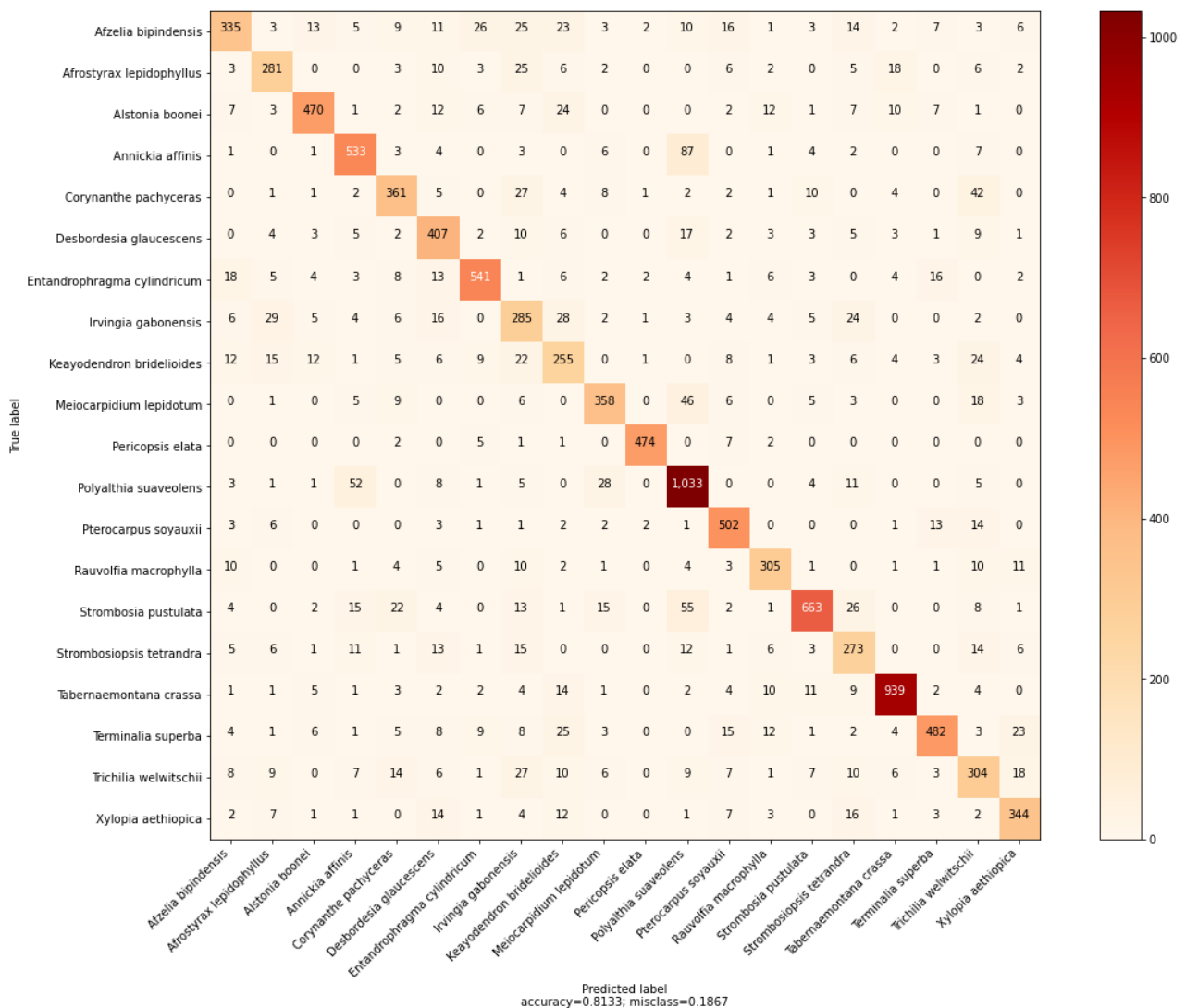


Figure 20 : Matrice de confusion des performances du modèle

Concernant les F-score, étant la performance globale du modèle, l'ensemble des espèces présentent des performances supérieures à 60% avec une moyenne de 80 % (Tableau 6).

L'intervalle du rappel varie de 65% à 96%. En effet, le nombre d'images correctement identifiées au sein des prédictions d'une même classe varie fortement.

Concernant la proportion d'images correctement classifiées au sein d'une classe théorique, elles suivent la même tendance que le rappel. En effet, la précision varie de 57% à 98%.

La fiabilité des résultats est un élément très important à soulever. Si le modèle donne un résultat, quelle en est la fiabilité ?

La fiabilité peut être « formalisée » à l'aide des rapports entre le rappel et la précision. En effet, plus le taux de faux positifs d'une espèce est élevé, plus l'utilisateur du modèle doit être critique envers ses résultats.

Pour *Polyalthia suaveolens*, *Pterocarpus soyauxii* et *Desbordesia glaucescens* on obtient des valeurs de rappel très élevées mais leurs valeurs de précision sont plus faibles. De ce fait, bien que le modèle n'identifie pas l'ensemble individus, le taux de faux positifs est très faible. La fiabilité sera donc effective (Tableau 6).

Pour les espèces *Keayodendron brideliodes*, *Irvingia gabonensis*, *Strombosiopsis tetrandra* et *Trichilia welwitschii*, la précision est faible et le taux de faux négatifs est donc élevé. Les résultats pour ces espèces sont les moins fiables du modèle car, en moyenne 6 résultats sur 10 sont corrects (F1-score de 60% à 70%) (Tableau 6).

De manière similaire, *Afzelia bipindensis* et *Meiocarpidium lepidotum* suivent la même tendance en ayant des valeurs de précision autour de 80% et des valeurs de rappel respectivement de 65% et 78%. De ce fait, malgré une reconnaissance élevée des individus, il y a plus de risques que ces espèces soient confondues avec d'autres (Tableau 6).

Finalement, certaines espèces présentent des performances équilibrées. C'est le cas d'*Afrostryax lepidophyllus*, *Annickia affinis*, *Corynanthe pachyceras*, *Rauwolfia macrophylla* et *Xylopia aethiopica*. Leur fiabilité suivra donc la même tendance que leur reconnaissance par le modèle : elles seront bien reconnues avec un risque minimum d'être confondues (F1-score entre 75 à 82%) (Tableau 6).

Tableau 6 : Résumé des performances du modèle pour chaque espèce étudiée

Classe	Précision	Rappel	F1-score
<i>Afzelia bipindensis</i>	79%	65%	71%
<i>Afrostryrax lepidophyllus</i>	75%	75,5%	75%
<i>Alstonia boonei</i>	89,5%	82%	86%
<i>Annickia affinis</i>	82%	82%	82%
<i>Corynanthe pachyceras</i>	79%	77%	78%
<i>Desbordesia glaucescens</i>	74%	84%	79%
<i>Entandrophragma cylindricum</i>	89%	85%	87%
<i>Irvingia gabonensis</i>	57%	67%	62%
<i>Keayodendron bridelioides</i>	61%	65%	63%
<i>Meiocarpidium lepidotum</i>	82%	78%	80%
<i>Pericopsis elata</i>	98%	96%	97, %
<i>Polyalthia suaveolens</i>	80%	89,5%	85%
<i>Pterocarpus soyauxii</i>	84%	91%	88%
<i>Rauwolfia macrophylla</i>	82%	83%	82%
<i>Strombosia pustulata</i>	91%	80%	85%
<i>Strombosiopsis tetrandra</i>	66%	74%	70%
<i>Tabernaemontana crassa</i>	94%	92,5%	93%
<i>Terminalia superba</i>	89,5%	79%	84%
<i>Trichilia welwitschii</i>	64%	67%	65%
<i>Xylopi aethiopica</i>	82%	82%	82%
<b>Exactitude</b>		<b>81,3%</b>	
Moyenne	80%	79,7%	79,7%
Moyenne pondérée	81,9%	81,3%	81,4%

### 3.3. Analyse des causes d'erreurs du modèle

#### 3.3.1. La présence d'images révélant des artefacts

À la suite de l'analyse à l'œil nu des 2104 images incorrectement identifiées par l'algorithme, 95% d'entre elles sont polluées par des artefacts. Pour 13 espèces, plus de 90% des arbres portent des artefacts (Figure 21).

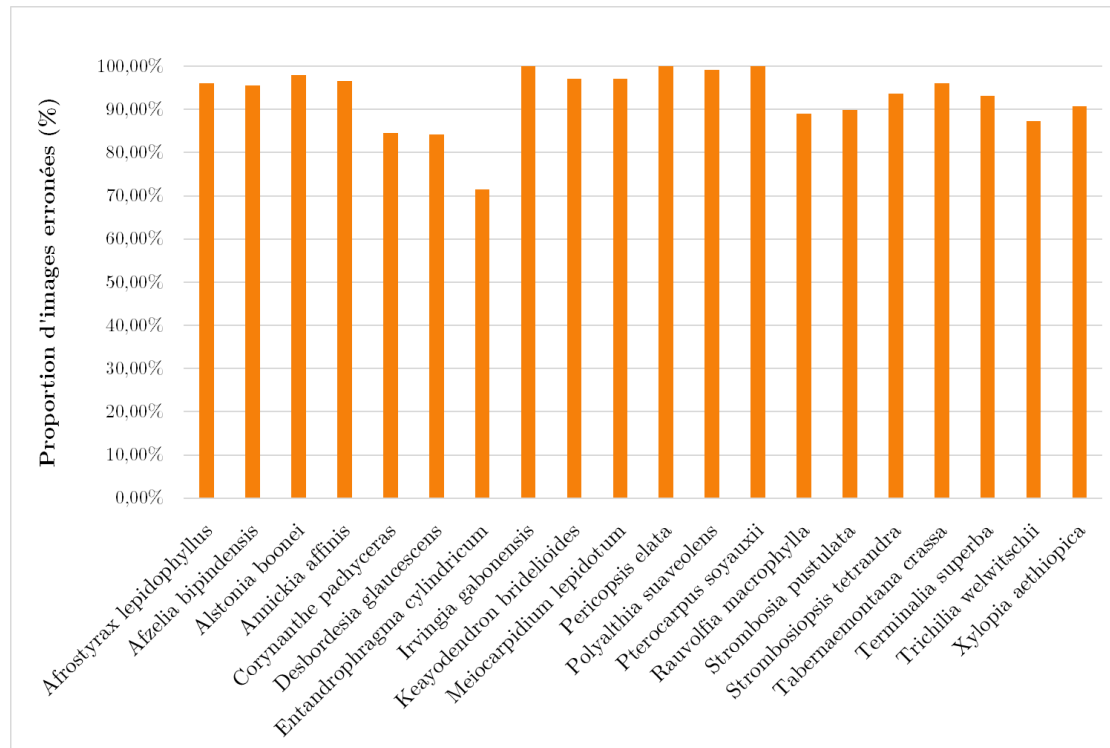


Figure 21: Proportions d'images relatives aux 20 classes choisies présentant des artefacts

L'observation des artefacts permet d'en définir deux types :

- Les artefacts provenant de l'opérateur : flou, présence d'arrière-plan et surexposition (Figure 22) ;
- Les artefacts liés aux caractéristiques du tronc: présence de mousse, lichens, termitières, tranche ou autre (branche, dégâts d'écorces, peinture, ...) (Figure 23).

### 3.3.1.1. Artéfacts dus à l'opérateur

Le flou se révèle être un artéfact qui se manifeste sur des images de l'intégralité des classes (Figure 22). Le pourcentage d'images floues par classe varie entre 9% et 42%. *Alstonia boonei*, *Entandrophragma cylindricum*, *Irvingia gabonensis*, *Pericopsis elata* et *Tabernaemontana crassa* ont la proportion d'images floues (plus de 35%) la plus élevée.

Concernant l'arrière-plan, 9% à 73% des images erronées par classe sont polluées par un arrière-plan apparent. *Strombosia pustulata* et *Trichilia welwitschii* ont les valeurs les plus élevées (plus de 58%) (Figure 22).

Le facteur de surexposition, défini par une luminosité trop élevée sur l'image, empêche de distinguer les caractéristiques de couleur et texture de l'écorce. Une surexposition est visible sur 5% des images et les valeurs d'intervalles sont entre 0 % et 20,5%. Le maxima est constaté pour l'espèce *Entandrophragma cylindricum* (Figure 22).

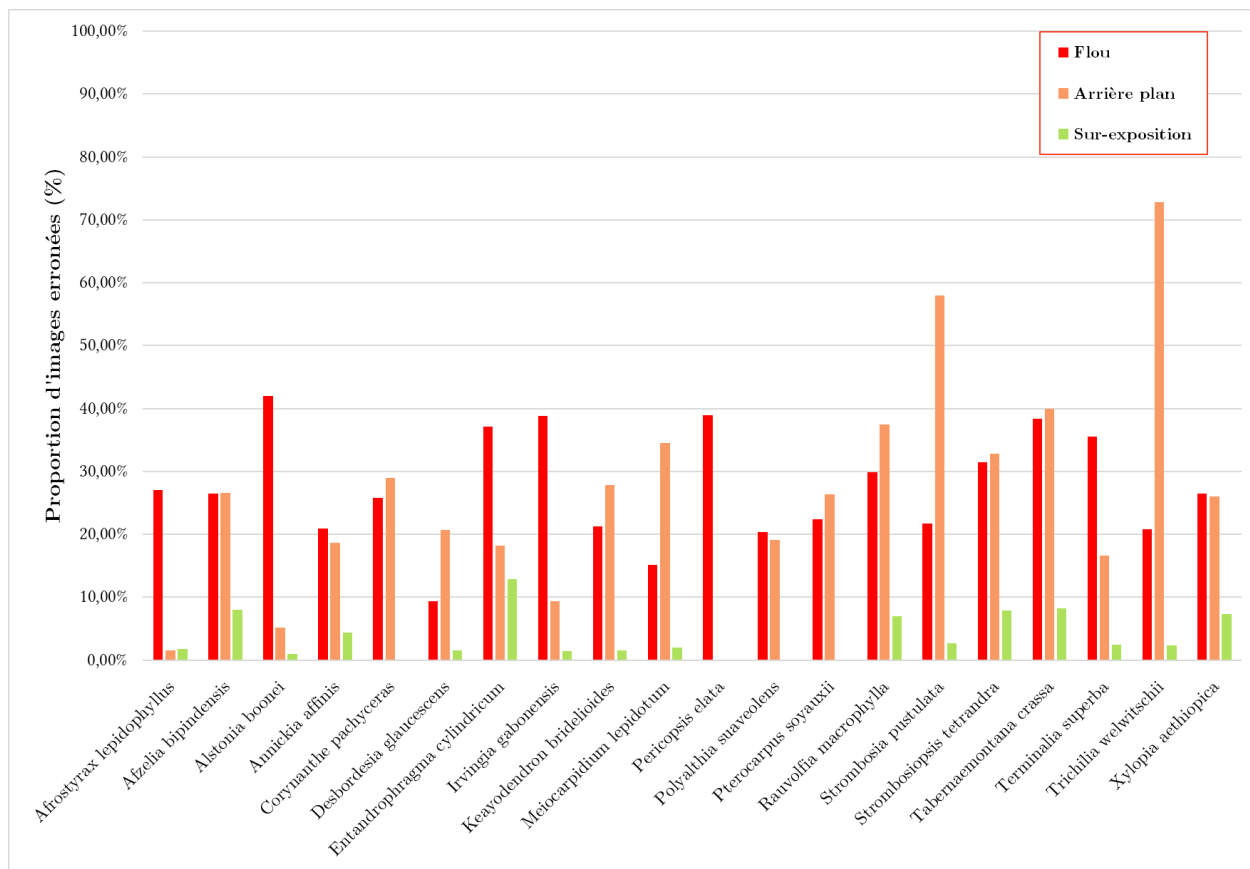


Figure 22 : Histogramme du pourcentage d'artéfacts lié à l'opérateur en fonction de la classe

### 3.3.1.2. Artefacts dus aux caractéristiques du tronc

En moyenne 68% des images par espèce recensent de la mousse, du lichen ou le mélange des deux. Selon les espèces, entre 33% et 87 % des images contiennent soit de la mousse, du lichen ou un mélange des deux. Le mélange des deux artefacts sur la même image est le plus courant (Figure 23).

Ensuite, le lichen est observé en moyenne sur 20% des images. Il est le plus fréquent (> 40%) pour les espèces suivantes : *Meiocarpidium lepidotum*, *Pterocarpus soyauxii*, *Strombosia pustulata*, *Trichilia welwitschii* et *Xylopiia aethiopica* (Figure 23).

La mousse est visible en plus faible proportion, en moyenne 18% des images par espèce (Figure 23).

Quant aux termitières, tranches et autres artefacts, ils n'apparaissent que sur moins de 5% des images (Figure 23).

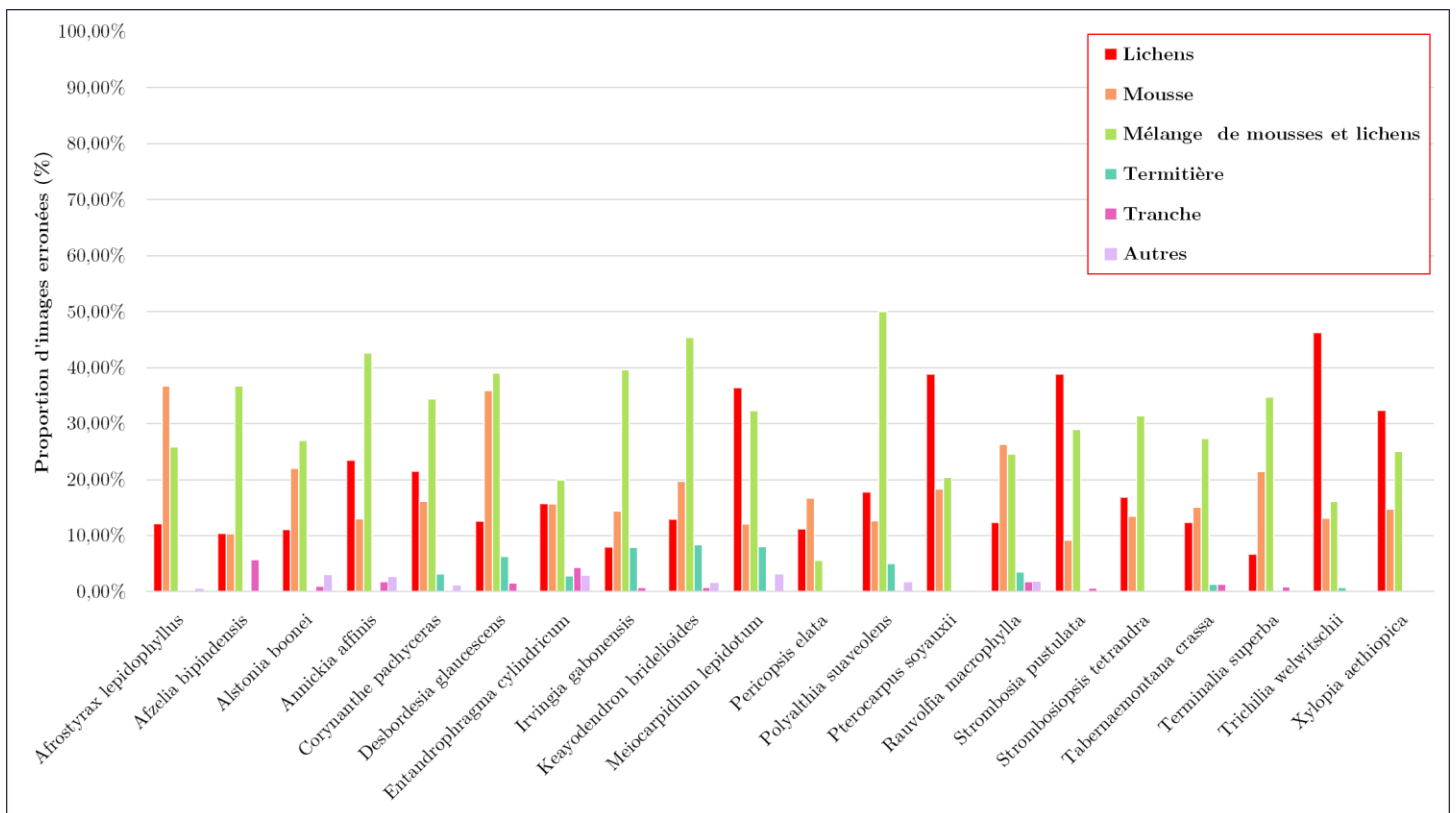


Figure 23 : Histogramme des artefacts présents sur le tronc en fonction de la classe



### 3.3.2. Etude de l'impact des artefacts : carte de chaleurs

Les artefacts présentent plusieurs risques pour le modèle. En théorie, certains de ces artefacts tels que la mousse, les termitières et les dégâts d'écorce peuvent créer un surapprentissage en devenant une caractéristique utilisée par le modèle pour reconnaître l'espèce. De plus, ils peuvent être source de faible performance car ils ne correspondent pas aux caractéristiques propres de l'espèce. Ceci constitue le risque le plus important révélé par notre étude. Ce propos est à nuancer concernant le lichen, pouvant être constant sur une espèce, étant donc une caractéristique pertinente des images de l'espèce (§2.4.1.1).

Comme défini au paragraphe 3.4.3.2., les informations analysées par le modèle ont été traduites en une carte des chaleurs. Cette carte représente les zones utilisées par le modèle pour la classification de l'image. Suite à la comparaison de celle-ci avec l'image initiale, les éléments concordant aux zones de haut intérêt de l'image peuvent être identifiés.

La Figure 24-A donne un exemple de classification d'espèce commune pour une écorce sans artefact. L'ensemble de l'écorce est prise en compte. De plus, deux zones mises en évidence (rouge foncé) sont les plus importantes pour la classification de l'image.

La mousse peut être reconnue comme élément d'importance pour l'identification de l'espèces (Figures 24-B et 24-C). Il en résulte alors une erreur d'identification.

Le flou tend à rendre le modèle confus sur les zones à identifier. En effet, la Figure 25-A considère l'ensemble de l'image comme importante pour la classification. Cela est problématique car le modèle ne différencie plus l'image même des zones neutres en bord d'image (pixels noirs).

Quant à l'arrière-plan, c'est une information inconnue pour le modèle et en cela, il risque de perturber la classification. Le modèle peut réagir de deux manières différentes : la première étant que le modèle focalise son identification sur la partie de l'image contenant le tronc (Figure 25-D) ; la seconde est plus problématique, à la place de se focaliser sur le tronc, le modèle définira la classe de l'image sur base de la zone d'arrière-plan (Figure 25-C).

Finalement, les termitières peuvent devenir la zone de l'image utilisée pour la classification. Cela est imagé au sein de la Figure 25-B où la zone ayant le plus d'importance pour la classification est celle de la termitière. L'image est donc incorrectement classifiée.

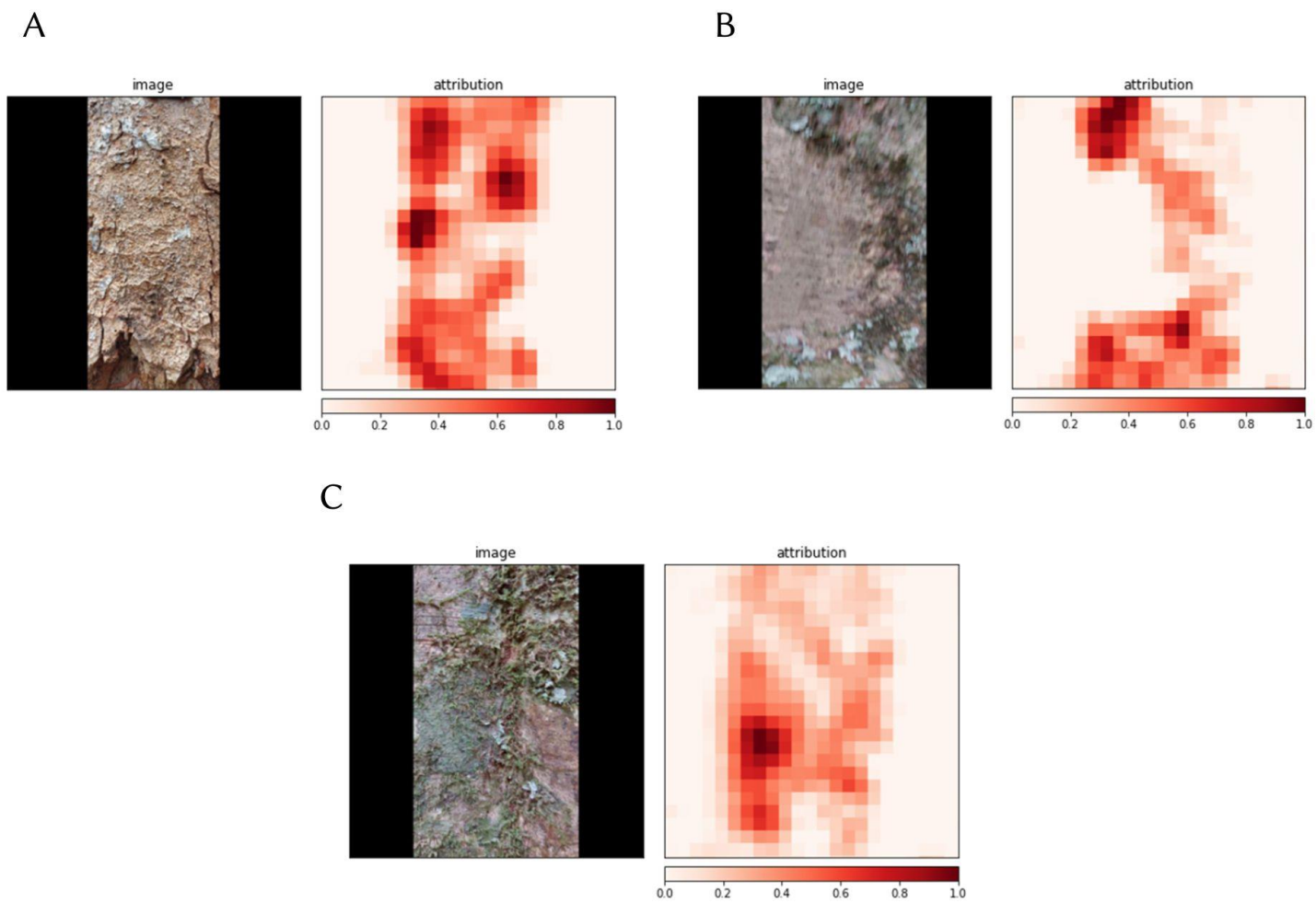


Figure 24 : Carte de chaleurs des zones d'intérêts du modèle pour la classification ( A : Exemple sans artefacts, B et C : Exemple avec mousse et lichens )

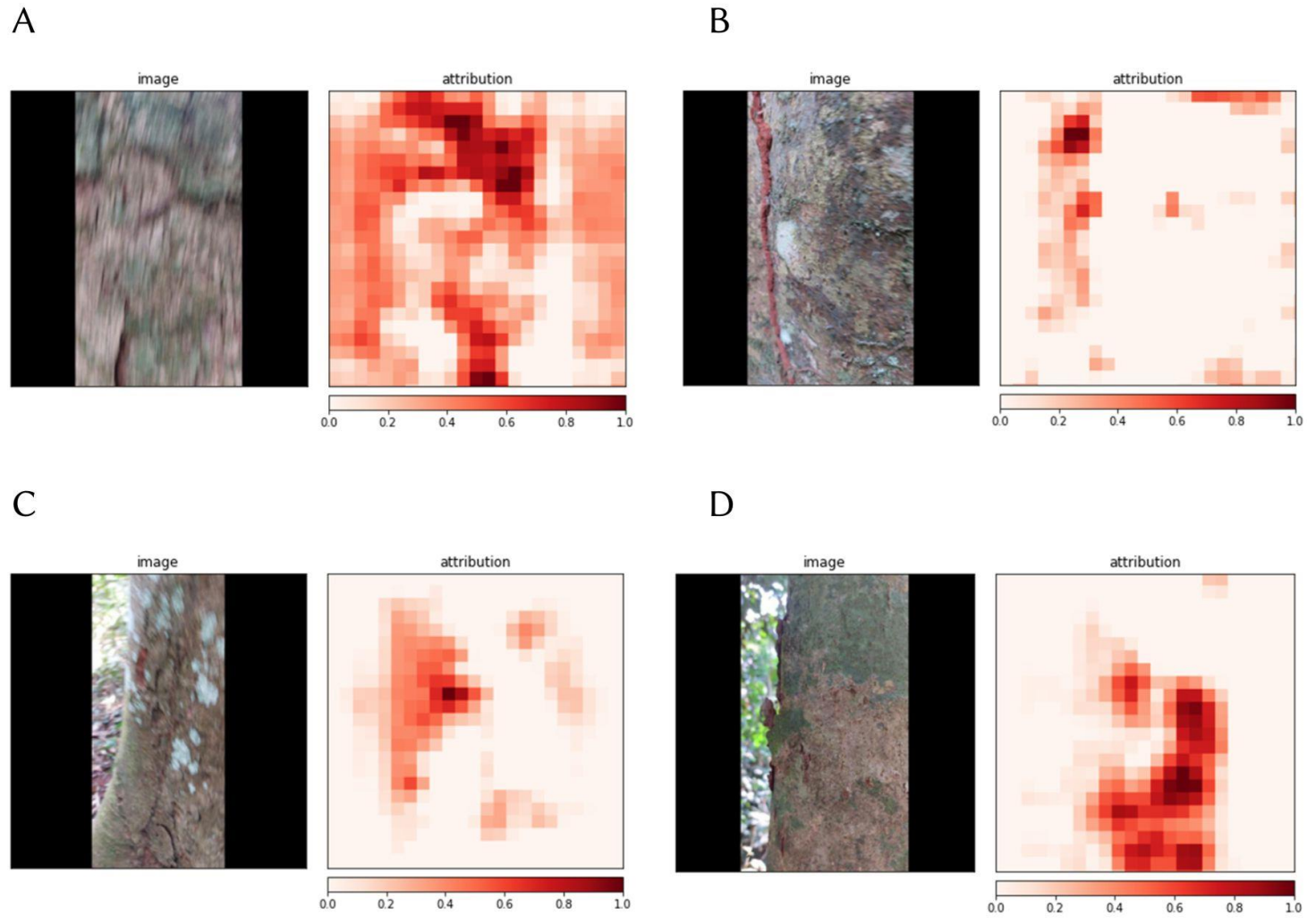


Figure 25 : Carte de chaleurs des zones d'intérêts du modèle pour la classification ( A : Image floue, B : Termitière, C et D : arrière-plan )

## 4. Discussion

### 4.1. Performance du modèle

La reconnaissance des arbres par l'écorce en forêt tropicale présente des contraintes évidentes car de nombreux facteurs confondants sont susceptibles de fausser l'interprétation. La présente étude a eu pour ambition d'utiliser les outils de l'intelligence artificielle, plus précisément le Deep learning, dans l'optique de permettre une reconnaissance automatique.

Le système développé a démontré une performance globale prometteuse. En effet, dans 8 cas sur 10, les espèces sont reconnues via des photographies d'écorce. Ce résultat est cependant à nuancer car toute moyenne cache une variabilité. Ainsi, *Pericopsis elata* est parfaitement reconnue 95 fois sur 100. A l'opposé, on se limite à 60 cas sur 100 pour *Irvingia gabonensis*. L'approche envisagée est donc intéressante mais encore perfectible.

#### 4.1.1. Comparaison des performances à la littérature disponible

A notre meilleure connaissance, la présente étude représente une première dans le domaine forestier tropical. Elle est néanmoins comparée à trois études récentes élaborées dans la même perspective.

La présente étude démontre qu'il est envisageable d'utiliser le Deep learning pour le développement d'outils de reconnaissance en milieu tropical.

L'étude de Carpentier et al. (2018) a démontré l'utilité du Deep learning dans le cadre de la reconnaissance d'espèces à l'aide d'un jeu de données de 23000 images pour 23 espèces ligneuses. Les performances moyennes annoncées étaient de 93,88 % d'exactitude en choix unique<sup>4</sup>. Lors de la réalisation de la base de données, la variabilité du diamètre, de la luminosité et de la distance de photographie au tronc ont été prises en compte.

Celle de Misra et al. (2020) quant à elle, présente des performances d'exactitude de 57,22% pour une base de données de 2587 images pour 101 espèces ligneuses. Le but de cette étude était de montrer que l'utilisation d'un outil d'intelligence artificielle offrait des performances plus élevées que les techniques de filtres anciennement utilisées.

Finalement, Kim et al. (2022) ont démontré que la reconnaissance automatique de 42 espèces d'arbres par l'écorce était efficace avec une exactitude de 90%. De plus, cette étude a permis de montrer que le modèle moins complexe utilisait des caractéristiques plus générales et mieux adaptées, tandis que le modèle le plus performant, doté d'un nombre de couches beaucoup plus important, indiquait des caractéristiques plus spécifiques et moins pertinentes pour les écorces.

---

<sup>4</sup> Seul le choix présentant le score de prédiction le plus élevé est considéré pour le calcul de l'exactitude.

Notre étude diffère en quatre points des trois études précédentes :

- Ses performances sont plus faibles que l'étude de Carpentier et al.(2018) de 10 % malgré un nombre d'images doublé pour un nombre de classes similaires.
- Ses performances sont plus élevées de 26% que celles de l'étude de Misra et al.(2020). Cependant, Misra et al.(2020) entraînent leur algorithme sur un jeu de données au nombre d'images extrêmement faible et avec un nombre de classes très élevé vis-à-vis de notre étude.
- Concernant l'étude de Kim et al.(2022), ses performances d'exactitude sont supérieures de 10% avec un nombre d'images plus faible (9375 images). Un lien intéressant entre nos deux études est la présence d'artefacts sur l'image comme source d'erreurs. De plus, certaines des erreurs restantes semblent également dues à des ressemblances interspécifiques.
- Finalement, notre étude diffère des trois précédentes par les conditions du milieu de recherche. Réalisée en forêt tropicale dense humide, les conditions de prises d'images différent et complexifient la reconnaissance d'images.

## 4.2. Etude du lien entre la qualité de la base de données et les performances du modèle

Au cours de la présente étude, une base de données présentant un grand nombre d'images pour 20 espèces ligneuses a été établie.

Dans le but d'analyser l'impact de la robustesse de la base de données sur les performances du modèle, 5 caractéristiques sont analysées une à une. Pour ce faire, les valeurs de F-1 score sont utilisées comme indicateur chiffré de la performance car il en constitue le paramètre le plus représentatif.

### 4.2.1.1. Nombre d'images

Le nombre d'images est un paramètre déterminant d'une base de données en Deep learning. En théorie, plus le modèle est alimenté par une quantité de données importantes plus il est performant. Dans le cas contraire, il sera sous-entraîné, il aura donc des faibles performances de reconnaissance pour des images qui sont en dehors de son propre jeu de données d'entraînement.

Si la base de données est déséquilibrée, c'est-à-dire si certaines espèces sont sur-représentées par rapport à d'autres, les performances intrinsèques du modèle risquent d'en être affectées. Dès lors, il est important de vérifier les proportions respectives des images relatives à chaque espèce (§ 4.4.2.4) et donc de minimiser les différences entre classes présentant de nombreuses données élevées et celles moins bien représentées. Il s'agit aussi « d'augmenter » les données par des prises de vue du même objet, sous différents angles artificiellement créés (§ 4.4.2.5).

Cette base de données surpasse, en termes de nombre d'images, l'ensemble des bases de données d'écorces préalablement créées (§1.1.3.4). Les bases de données de Carpentier et al. (2018) et Kim et al.(2018) ayant montré des performances de reconnaissance d'espèces de plus de 90% avec un nombre d'images plus faible, la base de données de la présente étude présente un nombre suffisant d'images pour 20 classes pour la création d'un outil d'intelligence artificielle. Elle surpasse de même largement la recommandation d'un minimum de 1000 images par classe.

De plus, la comparaison des deux espèces *Polyalthia suaveolens* et *Tabernaemontana crassa* possédant le nombre d'images le plus élevé à d'autres espèces moins bien représentées permet de voir qu'il n'y a pas de lien systématique entre performances et nombre d'images. Pour des nombres d'images différents, des F-scores proches sont observés (Tableau 7).

Tableau 7 : Comparaison du nombre d'images en fonction des performances ( F-1 score) par classe

Espèce	Nombre d'images	F-1 score
<i>Polyalthia suaveolens</i>	5754	84.71%
<i>Tabernaemontana crassa</i>	5025	93.25%
<i>Annickia affinis</i>	3314	82.00%
<i>Alstonia boonei</i>	2851	85.69%
<i>Rauvolfia macrophylla</i>	1834	82.43%
<i>Strombosiopsis tetrandra</i>	1830	69.91%

Il est donc possible de conclure qu'en termes de quantité de données, le jeu de données utilisé dans notre étude est suffisamment robuste.

#### 4.2.1.2. Nombre d'individus photographiés

Le nombre d'images est un facteur important dans la reconnaissance par IA, de même que la variabilité liée au nombre d'individus. A titre d'exemple, les valeurs ci-dessous pour deux espèces parmi les 20 investiguées, montrent que pour un nombre d'images et d'individus du même ordre de grandeur, le F-1 score peut être très différent (Tableau 8). Les autres espèces se situent entre ces deux extrêmes.

Dans le cadre de cette étude, le minimum de 61 individus par espèce est suffisant pour avoir des performances élevées. Cependant, l'augmentation du nombre d'individus tendra toujours à une nouvelle information de qualité pour le modèle. Cela restera donc toujours intéressant d'augmenter ce nombre pour perfectionner la robustesse de la base de données.

Tableau 8 : Comparaison des performances de 2 classes ayant le même nombre d'individus

Espèce	F1-score	Nombre d'individus	Unicité des caractéristiques morphologiques
<i>Irvingia gabonensis</i>	61.69%	68	-
<i>Pericopsis elata</i>	97.03%	68	++

#### 4.2.1.3. Variabilité de conditions écologiques

Diverses conditions écologiques se retrouvent au sein du site d'étude :

- Les conditions de canopée fermée qui se situent dans les parcours phénologiques ou layons de pré-exploitations.
- Les zones exploitées (Parcelles 1 et 2) où des zones ouvertes et des zones de couvert fermé se retrouvent.
- Les bords de routes à l'exposition lumineuse élevée.

Les conditions pluviométriques ont aussi impacté positivement la variabilité des conditions écologiques. Ainsi, suite à des épisodes pluvieux, des troncs ont été mouillés. Cela a modifié la texture et la couleur du tronc.

En figure 26, sont représentés l'impact d'un couvert forestier fermé vis-à-vis d'un couvert forestier ouvert et de la prise d'image après un épisode pluvieux.



Finalement, cette étude considère la variabilité géographique selon les limites de la concession forestière Pallisco. Celle-ci représente une étendue minimale vis-à-vis des surfaces forestières d'Afrique centrale. Sachant que le milieu dans lequel évoluent les espèces ligneuses influence leur morphologie, il semble pertinent de se questionner sur les performances du modèle sur des individus situés dans des zones phytogéographiques différentes d'Afrique centrale. Par exemple, ce modèle reconnaîtrait-il une même espèce située dans une région floristique différente ? Dans le but de tester les performances du modèle, il serait opportun d'élargir le champ d'investigation de l'étude à d'autres pays ou régions d'Afrique centrale.

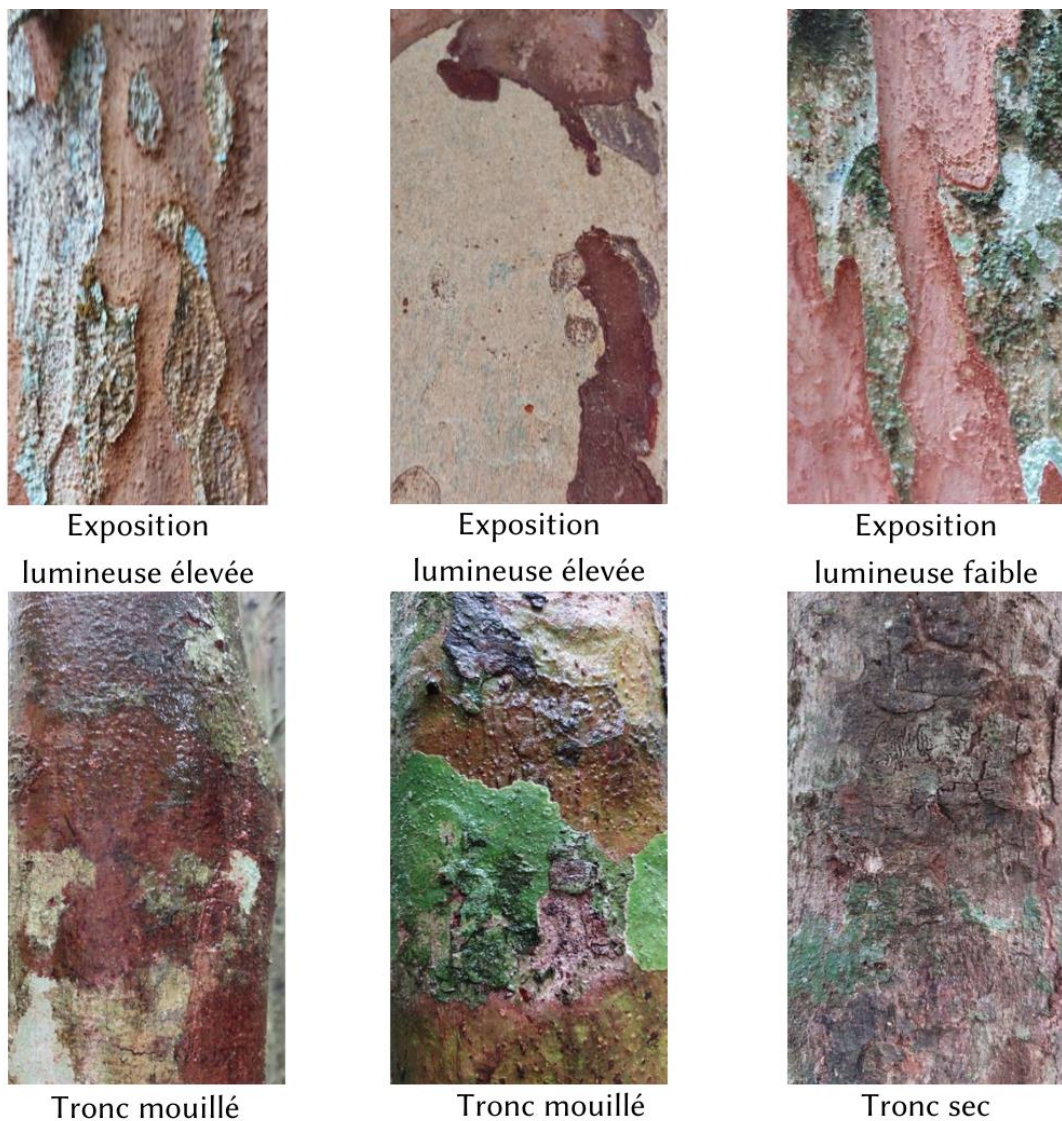


Figure 26 : Représentation d'images d'écorces impactées par les conditions climatiques. Les trois images du haut représentent *Periscopsis elata* exposé ou non à des conditions d'exposition élevées. Les trois images du bas correspondent à l'espèce *Trichilia welwitschii* avec le tronc mouillé suite à un épisode pluvieux ou un tronc sec.



#### 4.2.1.4. Diamètre

Une large gamme de diamètres est représentée dans cette étude. Dès lors, la gamme d'âges semble représentative des variations possibles en forêt tropicale.

La variabilité du diamètre mesuré pour les individus de chaque espèce influence peu les capacités de reconnaissance par le modèle. En effet, d'après le tableau 9, les diamètres répartis en 3 groupes selon leur variabilité ne montrent qu'une influence mineure sur les F-scores calculés.

Tableau 9 : Table des Intervalles de performance (F1-score) pour les classes de variabilités des diamètres

	Variabilité élevée	Variabilité moyenne	Variabilité faible
Intervalle de F1-score	71.35%-97.03%	75.44%-86.77%	61.69%-93.25%

#### 4.2.2. Impact des artefacts sur les performances du modèle

Dans cette étude, les artefacts ont été analysés à deux stades du travail. Premièrement, lors de la récolte des données, les artefacts ont été relevés sur l'ensemble des individus (§ 5.1.3) et deuxièmement à la suite de l'analyse des performances du modèle, ils ont été relevés sur les images erronées.

Deux types d'artefacts interviennent dans ces images :

- Les artefacts naturels,
- Les artefacts liés à l'opérateur.

##### 4.2.2.1. Artefacts naturels

Le lichen et la mousse sont présents et visibles sur de nombreuses images en proportion élevée. Seuls *Strombosia pustulata*, *Meiocarpidium lepidotum* et *Polyalthia suaveolens* possèdent ces artefacts sur moins de 50% de leurs individus.

L'impact de la mousse sur les performances du modèle est vérifié par les cartes de chaleur colligées à la Figure 24. Ces artefacts se retrouvant sur 68% des images erronées, la variation de performance du modèle en est potentiellement considérablement affectée. Toutefois, ces propos restent à nuancer car un grand nombre d'images ont été correctement identifiées malgré la présence d'artefacts. En effet, une majorité des individus de la base de données ont de la mousse ou un mélange de celle-ci et du lichen sur leur tronc.

En vue d'améliorer les performances, des modalités de prise d'images ou des prétraitements de la base de données devraient être envisagés pour permettre la suppression des images présentant plus de 50% de mousse sur l'image.

Le lichen pourrait être un potentiel impact mais celui-ci se retrouvant sur l'ensemble des individus d'une même espèce comme pour *Pterocarpus soyauxii*, *Keayodendron brideoides* et *Rauwolfia macrophylla*. Il est considéré comme une potentielle caractéristique de reconnaissance. Sur les espèces où il est présent en faible proportion, sa présence pourrait alors être un potentiel artefact et entraîner un risque de classification erronée.

Les termitières sont aussi fréquentes sur les troncs d'arbre. Cependant, dans la présente étude, leur occurrence est moindre sur les images car ce sont le plus fréquemment des lignes de terre sur le tronc alors que le lichen et la mousse ont tendance à recouvrir la surface du tronc. Cette hypothèse a été validée par l'examen des images erronées. En effet, bien que les termitières soient fort présentes sur les individus, leur présence sur les images erronément identifiées est minime. Cependant, comme démontré par les cartes de chaleurs, elles peuvent être une source de confusion pour le modèle.

En parallèle, deux artefacts liés à l'opérateur ont un impact élevé sur le caractère discriminant des reconnaissances d'espèces par IA. En effet, le flou et l'arrière-plan sont fréquemment sources d'erreurs et perturbent les critères de reconnaissance du modèle (Figure 25)

Le flou, décelé sur 25 % des images dont la classification est erronée, détériore la qualité des prises de vue empêchant le modèle de déterminer les caractéristiques importantes de l'écorce pour la classification de l'espèce. Cet artefact est donc considéré comme une problématique majeure. Lors d'une mission de terrain, éviter le flou peut être difficile. En effet, les conditions lumineuses pouvant être très faibles sous le couvert forestier, les performances d'un téléphone portable peuvent atteindre leurs limites en termes de mise au point et rendre difficile la réalisation d'une image nette.

La présence d'arrière-plan est de même problématique, elle amène une information nouvelle ne correspondant pas au schéma de reconnaissance créé par le modèle, cette tierce information peut être source d'erreur (Figure 25-D). Lors d'une mission de terrain, il peut aussi être complexe d'éviter de photographier l'arrière-plan vu le faible diamètre du tronc de certaines espèces. Cette observation est confirmée chez *Trichilia welwitschii* et *Strombosia pustulata* car une part importante de leurs individus ont un faible diamètre (entre 10 et 20 cm) et plus de 60% des images erronées présentent un arrière-plan.

En conclusion, l'impact de ces artefacts en forêt tropicale est déterminant. En effet, bien que le modèle ait été entraîné sur des images présentant des artefacts, la reconnaissance des espèces est cependant complexifiée par leur présence. La normalisation des prises de vues par un protocole de terrain plus stricte serait une perspective de résolution de cette problématique. Cette question sera discutée en §4.4.2.

### 4.2.3. Erreur humaine

L'erreur humaine est un facteur à considérer pour discuter les performances de ce modèle. En effet, 56 000 images ont été labellisées à la main dans un dossier à destination de la base de données, ce qui peut entraîner potentiellement des erreurs de labellisation.

Aussi, lors de la mission de terrain, certains individus restent complexes à identifier par une observation visuelle. *Trichilia welwitschii* peut-être difficilement différenciable d'autres *Trichilia spp.*, il est donc possible que des erreurs d'identification se soient glissées dans l'étude.

### 4.3. Le critère de l'écorce est-il suffisamment discriminant?

Au vu des performances du modèle, le critère de l'écorce semble suffisamment discriminant. Cependant, mettre en regard le lien entre les caractéristiques morphologiques et leur reconnaissance par le modèle permet de voir l'impact et les potentielles faiblesses de ce critère sur la reconnaissance.

Les résultats montrent une confusion entre espèces dont les écorces sont très semblables, par exemple *Annickia affinis* et *Polyalthia suaveolens*.

De ce fait, les espèces botaniques ayant les résultats les plus performants sont celles dont les caractéristiques botaniques sont distinctes et uniques vis-à-vis des autres, à l'instar de *Pericopsis elata* présentant des écailles laissant des marques rougeâtres caractéristiques et de *Tabernaemontana crassa* présentant un grand nombre de lenticelles. Toutefois, la présence de caractéristiques uniques n'assure pas pour autant une reconnaissance absolue de l'espèce par le modèle. A l'exemple de *Corynanthe pachyceras* ayant une couleur discriminante mais dont les performances de reconnaissance sont moyennes.

Le critère discriminant de l'écorce est aussi confirmé par l'étude de Misra et al.(2020). En effet, cette étude ayant un nombre élevé de classes pour un faible nombre d'images, a eu plus de 57 % d'exactitude pour 101 espèces. Cela offre donc des perspectives sur l'hypothèse qu'un grand nombre d'espèces d'arbres pourraient être discriminé par ce critère.

## 4.4. Limites de l'étude

### 4.4.1. Méthodologie de création du modèle

#### 4.4.1.1. Algorithme sélectionné

Comme présenté, le domaine de l'intelligence artificielle est un domaine vaste presque illimité en termes de possibilité de création. Lors de cette étude, seul l'algorithme Resnet-18 a été testé. Il s'est révélé performant pour répondre aux objectifs fixés en début de recherche (§ 1.1).

Au-delà de ses performances, il présente un second atout : ses performances sont associées à un nombre de couches faible lors de la phase de reconnaissance des « features ». Ce nombre de couches faible entraînera un modèle légèrement moins performant en terme de classification mais présentant une complexité plus faible. Cela représente un gain de temps d'exécution et de technicité indéniables. La complexité a un impact sur les paramètres de temps, de puissance de calcul et permet d'éviter le surapprentissage. Plus un modèle est complexe, plus le temps d'exécution et les capacités informatiques nécessaires augmentent.

Sachant que le contexte de la recherche se déroule dans des zones forestières éloignées, cette information prend toute son importance. En effet, cette étude vise à créer un outil exploitable facilement par les acteurs de la foresterie tropicale. Offrir un outil qui allie performances élevées et besoins technologiques faibles est donc primordial.

Dans le cadre de cette étude, utiliser des modèles présentant un nombre de couches élevé (Resnet 32 ou 64) pour une augmentation de l'exactitude pourrait ne présenter qu'un faible intérêt car comme dit précédemment, la complexification est associée à une augmentation des besoins technologiques.

Certes, la comparaison de différents algorithmes reste une possibilité d'optimisation de cet outil.

#### 4.4.1.2. Hyperparamètres du modèle

Une comparaison des hyperparamètres permet d'étudier des perspectives précises d'amélioration de l'outil de reconnaissance automatique actuel. L'étude de Carpentier et al.(2018) et de Kim et al.(2022) ont utilisé l'optimiseur « Adam »<sup>5</sup>, plus complexe que la descente de gradient stochastique.

Les performances de ces deux études étant supérieures à celle-ci, la différence d'optimiseurs pourrait être une première cause de cette différence de performance. Une première piste de recherche serait d'appliquer les paramètres et modèles sélectionnés de ces études à notre base de données pour voir l'impact de ceux-ci sur les performances de l'outil (Tableau 10).

Tableau 10 : Comparaison des hyperparamètres de cette étude à ceux utilisés par Carpentier et al.(2018) et Kim et al.(2022). Seuls les hyperparamètres disponibles en littérature ont été comparés.

Etude	Algorithme	Préentraîné Sur ImageNet	Optimisation	Taille du lot	Fonction d'optimisation
Carpentier et al.(2018)	Resnet18 et Resnet32	Oui	Adam	Non disponible	0.0001 avec une diminution à l'époque 16 et 33
Kim et al.(2022)	VGG-16 et EfficientNet	Oui	Adam	8	Non disponible
La présente étude	Resnet18	Oui	SGD	8	0.001 avec une diminution de 10-1 après 4 valeurs d'exactitude non maximale

#### 4.4.2. Limite du matériel de terrain

Un des choix méthodologiques de cette étude est la décision de réaliser la base de données uniquement sur base d'un smartphone avec un appareil photo de 64MP.

Lors de la mission de terrain, cet appareil a montré plusieurs limites :

- Le temps de terrain ayant été limité à 8 semaines de travail, 1400 images étaient prises en moyenne par jour. Ce nombre pouvait monter jusqu'à 3000 images/jour. Les performances du téléphone ont été impactées par cette prise d'images à répétition pendant une durée de temps élevée. Suite à plusieurs heures de terrain, la mise au point du téléphone devenait plus instable et l'image devenait plus sujette au flou.
- Finalement, le téléphone ne présentant pas les performances d'un appareil photographique et donc des réglages professionnels, les paramètres étaient complexes à optimiser pour faire ressortir au mieux la texture et la couleur de l'écorce. Lorsque les conditions lumineuses étaient faibles, sous un couvert forestier

---

<sup>5</sup> Pour plus d'informations sur ce procédé mathématique, l'article de Kingman et Ba (2017) est disponible.

dense par exemple, il pouvait être complexe d'avoir une image aussi discriminante qu'elle l'était à l'œil nu.

En résumé, le téléphone portable a permis de créer les conditions réelles de prise de vue d'un utilisateur mais lors de l'utilisation intensive de l'appareil et dans des conditions de terrain complexes, la qualité de ses images s'en est vue altérée.

Pour répondre à cette problématique, de nombreuses pistes sont envisagées :

- Un cadre d'utilisation du matériel sur terrain pourrait être envisagé tel qu'un nombre d'heures de prise d'images limité,
- L'utilisation d'une application de prise d'images permettant le réglage précis des paramètres du capteur
- L'utilisation de deux smartphones par jour pour limiter l'altération de la qualité de prise d'images
- Le backup des données régulier des données et leur suppression de la mémoire interne afin d'empêcher le ralentissement du processeur.

De mon point de vue, l'utilisation d'un appareil photo de qualité n'est pas une solution car :

- Premièrement, cela ne correspondrait plus à l'objectif de la présente étude. En effet, l'outil de reconnaissance artificiel utilisable doit répondre à l'objectif d'être accessible et utilisable dans le domaine de gestion forestière tropicale.
- Deuxièmement, le traitement des images ayant nécessité une diminution de la qualité de celle-ci, un outil plus performant n'aurait pas d'intérêt. Le seul avantage réel étant une mise au point plus adaptée aux conditions du milieu.

#### 4.5. Identification botanique, lien entre l'humain et la technologie

L'écorce n'est habituellement pas un critère considéré seul en botanique tropicale. L'espèce est habituellement identifiée à l'aide des caractéristiques de tranches, feuilles, formes du tronc ou houppier.

Il est donc intéressant de se questionner sur le fait que si l'homme est capable de reconnaître une espèce sur base de la combinaison de plusieurs de ces critères, la machine ne serait-elle pas de même plus performante en combinant ces caractéristiques ?

L'application Plant@net utilisant la fusion multimodale et donc plusieurs caractéristiques morphologiques pour la reconnaissance de l'espèce est un exemple d'outil de reconnaissance automatique performant. En effet, ses capacités de reconnaissance automatique ont déjà été prouvées sur des espèces et écosystèmes divers. Avec un jeu de données suffisamment conséquent sur la thématique des arbres tropicaux, l'hypothèse que cette technologie serait plus performante qu'un algorithme CNN simple semble prometteuse.

## 5. Conclusion et perspectives

Cette étude conclut que l'utilisation d'une intelligence artificielle de type Resnet18 est réalisable et performante pour la reconnaissance d'espèces ligneuses en utilisant des images d'écorces en milieu tropical.

Cette étude a également permis la création d'une base de données d'images d'écorces originale qui englobe 56 000 prises de vues pour 20 espèces ligneuses. Cet aspect montre aussi qu'il est possible de créer une base de données robuste et d'ampleur suffisante dans un laps de temps court et avec des outils technologiques relativement simples, en l'occurrence : un téléphone portable, un ordinateur au processeur performant avec des logiciels en libre accès dédiés.

A contrario, cette étude met en évidence l'impact des artefacts sur les performances de classification ainsi qu'un risque de confusion par le modèle des espèces ayant des caractéristiques morphologiques similaires.

La base de données s'est montrée suffisamment représentative en termes de nombre d'individus, nombre d'images, de variabilité de conditions écologiques et de gamme d'âges. Cependant, elle demeure perfectible, notamment en termes d'exploration géographique et d'espèces. Le nombre d'espèces d'arbres en Afrique tropicale représentant plus de 4000 espèces, les champs de complexification de ce type d'outil sont infinis.

Lors de cette étude le critère de l'écorce s'est montré suffisamment discriminant pour la différenciation d'une vingtaine de classes. Le critère « aspect de l'écorce » serait-il encore discriminant pour 500 voire plus de 4000 espèces?

En conséquence, dans la continuité de ce travail, plusieurs pistes de réflexions doivent être envisagées.

Premièrement, le programme pourrait être complexifié en ajoutant un autre critère morphologique, à l'exemple des caractéristiques de la tranche ou de la feuille. Cet ajout pourrait potentiellement être envisagé selon les deux optimisations suivantes :

- Une description morphologique du second critère par l'utilisateur sous la forme d'un traitement préalable à l'envoi de l'image au modèle. Par la suite, l'image serait traitée par le modèle de reconnaissance automatique associé aux caractéristiques décrites. Le nombre d'espèces, objet de l'étude, pourrait donc être augmenté drastiquement tout en réduisant les risques d'erreur.
- Un algorithme plus complexe en utilisant la technique de fusion multimodale comme dans l'application Pl@ntnet (§ 2.3.4). Cette technique permettrait de combiner l'image d'une écorce avec une image décrivant une autre caractéristique morphologique de l'individu. Cela permettrait d'augmenter les probabilités de reconnaissance de chaque espèce en combinant plusieurs informations sur celle-ci, tel que le ferait un botaniste en forêt tropicale.

En second lieu, la complexification pourrait être l'ajout d'une carte de chaleur au préalable à la classification ( § 1.3.4) comme dans les articles de Kim et al. ( 2022) et Zhu et al. (2019). Elle permettrait de se focaliser sur les zones réellement importantes pour la classification. Par exemple, elle pourrait permettre la suppression automatique des arrière-plans, diminuant ainsi les biais qu'ils induisent.

Finalement, une seule architecture de modèle, le Resnet-18 a été testée. De ce fait, une comparaison de différentes architectures de CNN pourrait permettre d'augmenter et d'améliorer les performances de reconnaissance de cette même base de données. Des architectures plus profondes, à l'exemple du Resnet-32, pourraient être envisagées dans des recherches complémentaires.

En conclusion, les perspectives d'amélioration et d'utilisation à court terme de cette étude sont les suivantes :

- L'augmentation de la variabilité de la base de données en termes de conditions écologiques ;
- L'augmentation du nombre d'espèces ligneuses associée à la complexification de la technologie utilisée ;
- La création d'une technologie d'intelligence artificielle exploitable de manière simple en vue de tendre vers une application à grande échelle.

A long terme, des perspectives d'utilisation pourraient être envisagées de manière exponentielle car toute personne détentrice d'un smartphone et d'un accès à internet aurait accès à un outil de reconnaissance d'arbres tropicaux.

Cette possibilité pourrait faciliter les inventaires forestiers et avoir un impact économique (gestion optimisée), écologique (gestion plus durable) et touristique (gestion plus ouverte) car l'accessibilité à l'information favoriserait l'intérêt général pour la forêt tropicale et ses nombreuses richesses si favorables à l'homme.

L'intelligence artificielle au service d'une gestion intelligente du poumon vert mondial était il y a encore peu de temps un rêve, il peut devenir réalité.



## 6. Bibliographie

Abdullahi H., Sheriff R. & Mahieddine F., 2017. Convolution neural network in precision agriculture for plant image recognition and classification. In : *Proceedings of the 7th International Conference on Innovative Computing Technology (INTECH)*, Luton, UK, 16–18.

Al Dahdah, M., 2019. Les géants du numérique au chevet de l’Afrique. Le téléphone portable comme nouvel outil de santé globale. *Politique africaine*, 156, 101-119. <https://doi.org/10.3917/polaf.156.0101>

Barré P., Stöver B.C., Müller K.F. & Steinhage V., 2017. LeafNet : A computer vision system for automatic plant species identification. *Ecological Informatics*, 40, 50–56, DOI:[10.1016/j.ecoinf.2017.05.005](https://doi.org/10.1016/j.ecoinf.2017.05.005).

Ben Ameer R., 2018. Fusion multimodale pour la reconnaissance d’espèces d’arbres. Thèse de doctorat : Université Grenoble Alpes (France). <https://tel.archives-ouvertes.fr/tel-01921590>

Benítez Á., Aragón G. & Prieto M., 2019. Lichen diversity on tree trunks in tropical dry forests is highly influenced by host tree traits. *Biodivers Conserv* **28**(11), 2909–2929, DOI:[10.1007/s10531-019-01805-9](https://doi.org/10.1007/s10531-019-01805-9).

Blaanco L. J., Travieso C. M., Quinteiro J. M., Hernandez P. V., Dutta M. K. & Singh A., 2016. A bark recognition algorithm for plant classification using a least square support vector machine. in : *2016 Ninth International Conference on Contemporary Computing (IC3)*, 1–5.

Bressane A., Roveda J.A., & Martins A.C., 2015. Statistical analysis of texture in trunk images for biometric identification of tree species. *Environmental Monitoring and Assessment*, 187, 1-9.

Carpentier M., Giguère P. & Gaudreault J., 2018. Tree Species Identification from Bark Images Using Convolutional Neural Networks. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1075–1081.

Chen L., Li S., Bai Q., Yang J., Jiang S. & Miao Y., 2021. Review of image classification algorithms based on Convolutional Neural Networks. *Remote Sensing* **13**(22), 4712, DOI:[10.3390/rs13224712](https://doi.org/10.3390/rs13224712).

Chollet F., 2018. *Deep Learning with Python*. États-Unis d’Amérique : Manning Publications.

Dainou K., Dubart N., Bracke C. & Doucet J.-L., 2017. Evaluation de la présence de hautes valeurs de conservation (hvc) dans les concessions gérées par la société Pallisco, Cameroun. Massif forestier des UFA 10.030, 10.031, 10.039, 10.042, 10.044 et 10.047.

Delplanque A., 2019. Travail de fin d’études : Identifying african mammal species in aerial images with object detection algorithms. Gembloux Agro-Bio Tech

Deng J., Dong W., Socher R., Li L.-J., Li K. & Fei-Fei L., 2009. ImageNet: A large-scale hierarchical image database. *In: 2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255.

Dhruv P. & Naskar S., 2020. Image Classification Using Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN): A Review. *In: Learning and Information Processing, Advances in Intelligent Systems and Computing*. Singapore: Springer, 367–381.

Doumenge C., Mboma R., Sist P., Costa Procopio L. Bouka G., GourletFleury S., Lowry P. & Chevillotte H., 2016. La connaissance des espèces: une quête du Graal pour la gestion durable des forêts tropicales, *Quels botanistes pour le 21e siècle ? Métiers, enjeux et opportunités*, 40-49.

Doucet J.-L., 2006. Cours d'essences forestières des régions chaudes. *Laboratoire de foresterie des régions tropicales et subtropicales, Gembloux Agro-Bio Tech - Uliège*

Dupuis C., 2019. Mission de collecte des données sur le site de la société forestière Pallisco dans le cadre du projet IDROC et de la thèse de doctorat de Chloé Dupuis.

Durand T. & Durand H., 1909. Sylloge florae congolanae [Phanerogamae], Bruxelles : Maison A. de Boeck, 738.

Elgendy, 2020. Deep learning for vision systems.

Fayolle A., Swaine M.D., Bastin J.-F., Bourland N., Comiskey J.A., Dauby G., Doucet J.-L., Gillet J.-F., Gourlet-Fleury S., Hardy O.J., Kirunda B., Kouamé F.N. & Plumptre A.J., 2014. Patterns of tree species composition across tropical African forests. *Journal of Biogeography* 41(12), 2320–2331, DOI:[10.1111/jbi.12382](https://doi.org/10.1111/jbi.12382).

Feteke F., Fayolle A., Dainou K., Bourland N., Dié A., Lejeune P., Doucet J.-L. & Beeckman H., 2017. Variations saisonnières de la croissance diamétrique et des phénologies foliaire et reproductive de trois espèces ligneuses commerciales d'Afrique centrale. *Bois & forêts des tropiques* 330, 3, DOI:[10.19182/bft2016.330.a31315](https://doi.org/10.19182/bft2016.330.a31315).

Fiel S. & Sablatnig R., 2010. Automated identification of tree species from images of the bark, leaves or needles 63. *Thèse de doctorat : Institute of Computer Aided Automation Vienna University of Technology*

Figuroa-Mata G. & Mata-Montero E., 2020. Using a Convolutional Siamese Network for Image-Based Plant Species Identification with Small Datasets. *Biomimetics* DOI:[10.3390/biomimetics5010008](https://doi.org/10.3390/biomimetics5010008).

Goëau H., Bonnet P. & Joly A., 2019. Overview of LifeCLEF Plant Identification task 2019: diving into data deficient tropical countries .

Gogul I. & Kumar V.S., 2017. Flower species recognition system using convolution neural networks and transfer learning. *In: 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN)*. 1–6.

Grinblat G.L., Uzal L.C., Larese M.G. & Granitto P.M., 2016. Deep learning for plant identification using vein morphological patterns. *Computers and Electronics in Agriculture* 127, 418–424, DOI:[10.1016/j.compag.2016.07.003](https://doi.org/10.1016/j.compag.2016.07.003).

Hallé N., 1966. Flore du Gabon (12) : Rubiaceae (1). *Muséum national d'Histoire naturelle*, Paris, 277p.

He J., Gu H. & Liu W., 2012. Imbalanced Multi-Modal Multi-Label Learning for Subcellular Localization Prediction of Human Proteins with Both Single and Multiple Sites. *PLOS ONE* 7(6), e37155, DOI:[10.1371/journal.pone.0037155](https://doi.org/10.1371/journal.pone.0037155).

He K., Zhang X., Ren S. & Sun J., 2015. Deep Residual Learning for Image Recognition.

Jendoubi S., Coquin D. & Boukezzoula R., 2018. K-NN crédibilise pour la reconnaissance automatique des espèces d'arbres. *LFA 2018 - Rencontres Francophones Sur La Logique Floue et ses Applications*, Arras, France.

JMN, 2015. Etude d'impact environnemental détaillée de l'UFA 10 047b. Société Pallisco sarl. Cameroun.

Joly A., Bonnet P., Goëau H., Barbe J., Selmi S., Champ J., Dufour-Kowalski S., Affouard A., Carré J., Molino J.-F., Boujemaa N. & Barthélémy D., 2016. A look inside the Pl@ntNet experience: The good, the bias and the hope. *Multimedia Systems* 22(6), 751–766, DOI:[10.1007/s00530-015-0462-9](https://doi.org/10.1007/s00530-015-0462-9).

Kim T.K., Hong J., Ryu D., Kim S., Byeon S.Y., Huh W., Kim K., Baek G.H. & Kim H.S., 2022. Identifying and extracting bark key features of 42 tree species using convolutional neural networks and class activation mapping. *Sci Rep* 12(1), 4772, DOI:[10.1038/s41598-022-08571-9](https://doi.org/10.1038/s41598-022-08571-9).

Kingma D.P. & Ba J., 2017. Adam: A Method for Stochastic Optimization. *ICLR 2015*.

Lee S.H., Chan C.S., Mayo S.J. & Remagnino P., 2017. How deep learning extracts and learns leaf features for plant classification. *Pattern Recognition* 71, 1–13, DOI:[10.1016/j.patcog.2017.05.015](https://doi.org/10.1016/j.patcog.2017.05.015).

Letouzey R., 1972. Manuel de botanique forestière – Afrique tropicale. Tome 1 : Botanique générale. CTFT, Nogent-sur-Marne, 189.

Letouzey R., 1986. Notice de la carte phytogéographique du Cameroun au 1/ 500,000. Institut de la carte internationale de la végétation, Toulouse, France.

Loukou A.F., 2012. Les TIC au service du développement en Afrique. *Tic & société* 5 (2), DOI:[10.4000/ticetsociete.1047](https://doi.org/10.4000/ticetsociete.1047).

Meunier Q., Carl M. & Doucet J.-L., 2015. Les arbres utiles du Gabon, Presses Agronomiques de Gembloux, 172.

Misra D., Crispim-Junior C. & Tougne L., 2020. Patch-Based CNN Evaluation for Bark Classification. *Computer Vision. ECCV 2020 Workshops*, 12540, 197-212.

Mukherjee S., Kumar P., Saini R., Roy P.P., Dogra D.P. & Kim B.-G., 2017. Plant Disease Identification using Deep Neural Networks. *Journal of Multimedia Information System* 4(4), 233–238, [DOI:10.9717/JMIS.2017.4.4.233](https://doi.org/10.9717/JMIS.2017.4.4.233).

Nilsback M-E., Zisserman A., 2008. Automated flower classification over a large number of classes. Visual Geometry Group, Oxford.

Pallisco-CIFM & Nature+, 2015. Résumé des plans d'aménagement et des directives FSC pour les UFA 10-030, 10-031, 10039, 10-041, 10-042 et 10-044 exploitées par la société PALLISCO et ses partenaires.

Ratajczak, R. et al., 2019. Efficient Bark Recognition in the Wild. Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 4, 240-248.

Slik et al., 2015. An estimate of the number of tropical tree species. *PNAS Early Edition* 1-6.

Sun Y., Xin Q., Huang J., Huang B. & Zhang H., 2019. Characterizing Tree Species of a Tropical Wetland in Southern China at the Individual Tree Level Based on Convolutional Neural Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12(11), 4415–4425, [DOI:10.1109/JSTARS.2019.2950721](https://doi.org/10.1109/JSTARS.2019.2950721)

Svab M., 2014. Computer-vision-based tree trunk recognition.

Vivien J. & Faure J.J., 2011. Arbres des forêts denses d'Afrique centrale. Nguila Kerou, Clohars Carnoet, France, 945 p.

Wang Z., Li H., Zhu Y. & Xu T., 2017. Review of Plant Identification Based on Image Processing. *Arch Computat Methods Eng* 24(3), 637–654, DOI:[10.1007/s11831-016-9181-4](https://doi.org/10.1007/s11831-016-9181-4).

Wilks C. et Issembé Y., 2000. Guide pratique d'identification: les arbres de la Guinée Equatoriale: Région continentale. Projet Curet Bata, Guinée Equatoriale, Prépresse Communications, Paris

Zhu Y., Sun W., Cao X., Wang C., Wu D., Yang Y. & Ye N., 2019. TA-CNN: Two-way attention models in deep convolutional neural network for plant recognition. *Neurocomputing* 365, 191–200, DOI:[10.1016/j.neucom.2019.07.016](https://doi.org/10.1016/j.neucom.2019.07.016).

## Lien web

Albumentation. Demo · Streamlit, 2022. . <https://albumentations-demo.herokuapp.com/>, (14/07/2022).

Jardin botanique de Meise, 2022 . <https://www.floredafriquecentrale.be/#/en/home>, (05/08/2022).

Jardin botanique de Meise, 2022.La Flore du Gabon, August-5-2022. . <https://www.plantentuinmeise.be/fr/pQErwSd/flore-du-gabon>, (05/08/2022).

Pallisco-Cifm, 2022. Présentation.August-5-2022. . <http://www.pallisco-cifm.com/index.php/fr/chiffres-cles>, (05/08/2022).

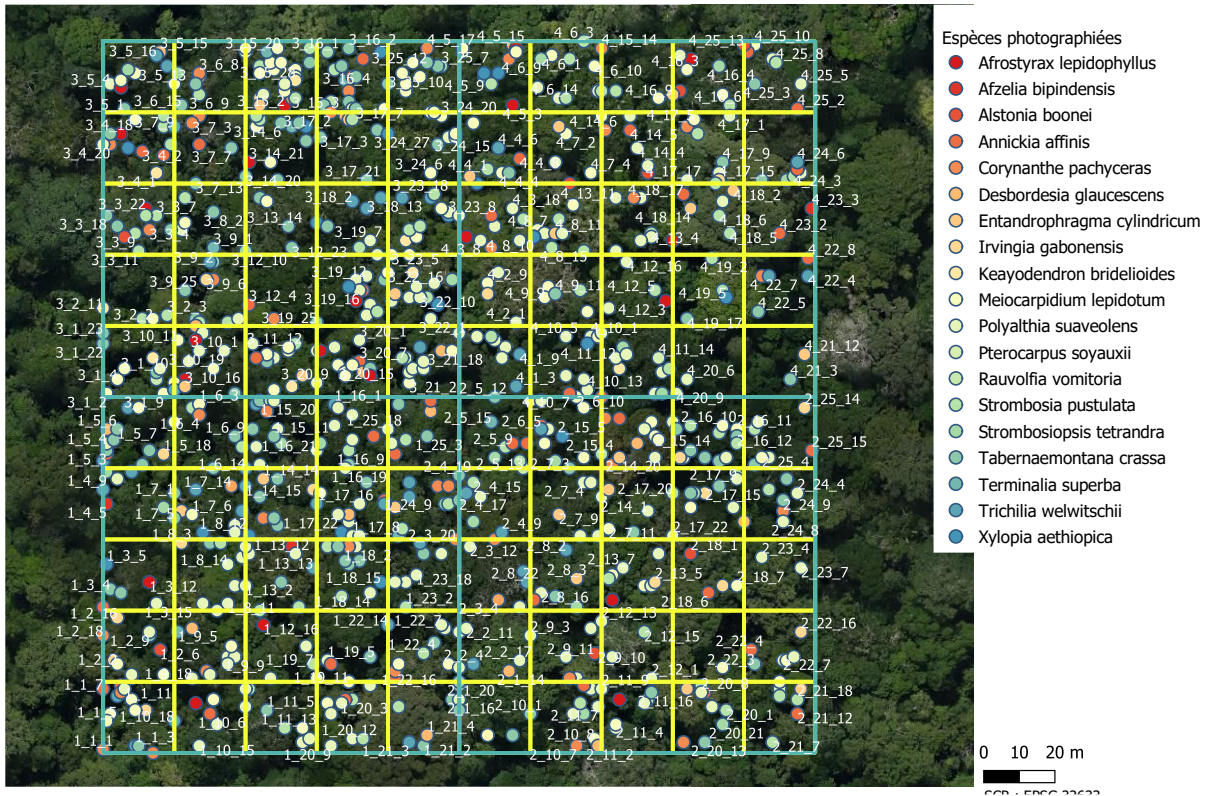
Pl@ntNet, 2022. Accueil, August-15-2022. . *Pl@ntNet*. <https://plantnet.org/>, (2/08/2022).

PROTA4U, 2022. . <https://www.prota4u.org/database/searchresults.asp>, (17/06/2022).

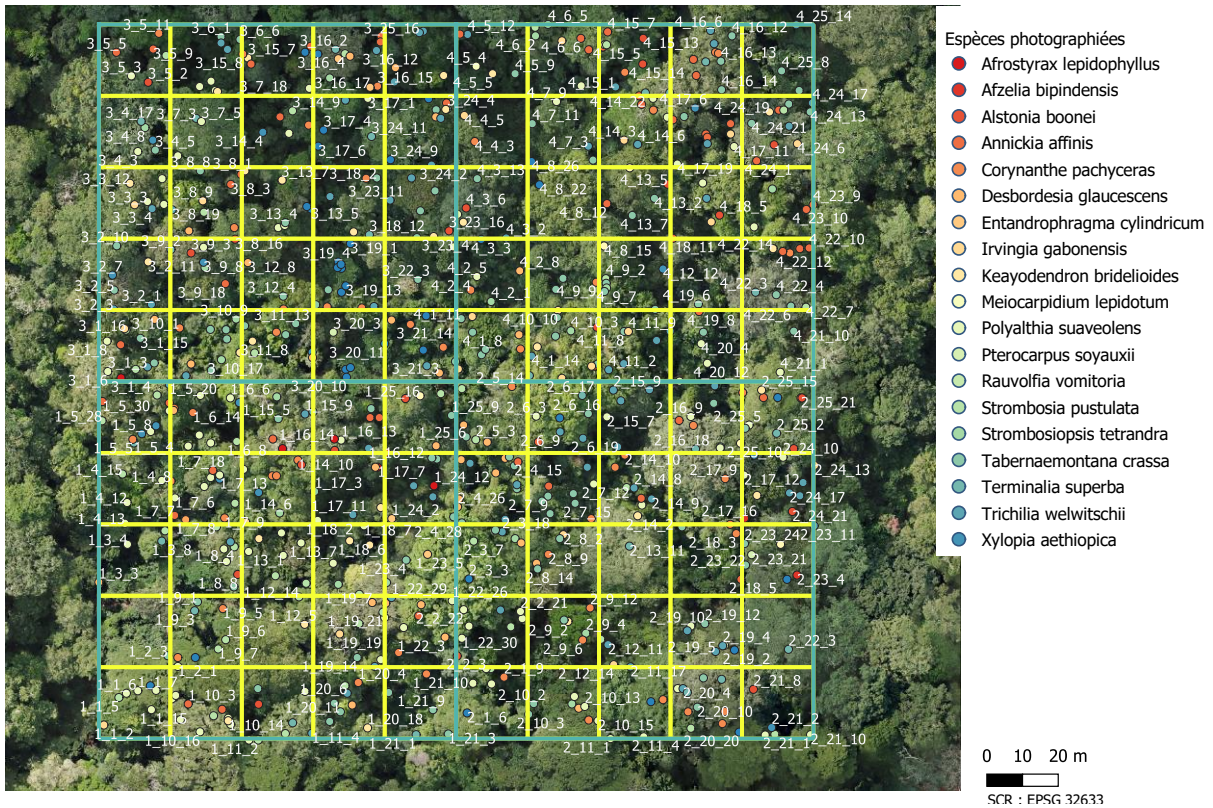


# 7. Annexes

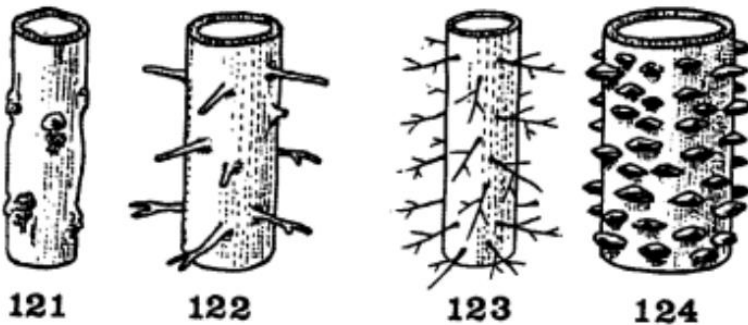
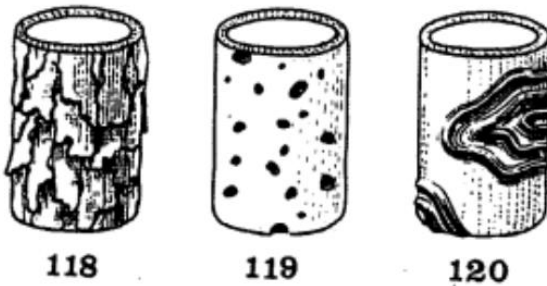
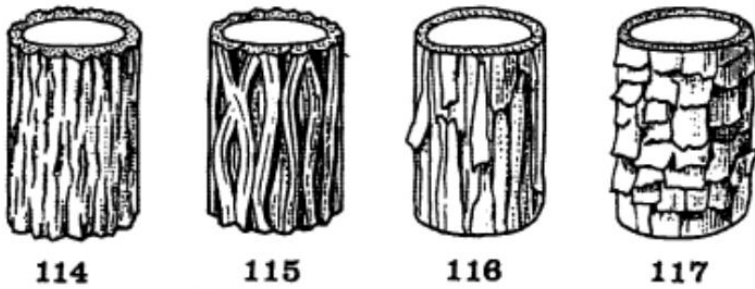
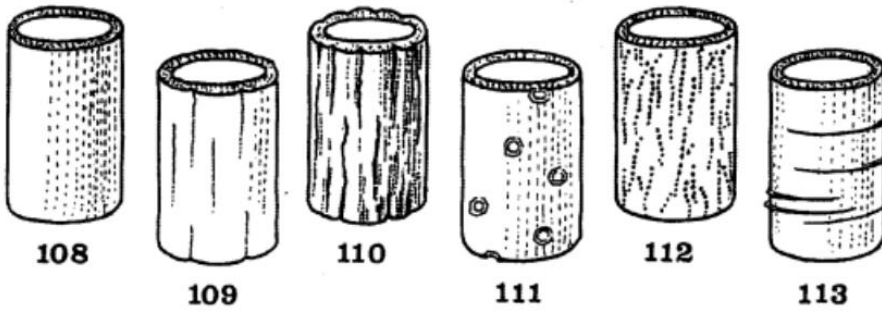
Annexe 1 : Cartographie des arbres d'intérêts répertoriées dans la Parcelle 1



Annexe 2 : Cartographie des arbres d'intérêts répertoriées dans la Parcelle 2



Annexe 3 : Illustration des différentes morphologies d'écorces (Letouzey, 1972)



- 108. Lisse
- 109. Fissurée longitudinalement
- 110. Légèrement rugueuse
- 111. Tavelée
- 112. Lenticellée
- 113. Bourrelets horizontaux
- 114. Crevassées longitudinalement
- 115. Dessin analogue à un filet étiré
- 116. Se détachant en lanière
- 117. A écailles ou plaques rectangulaires
- 118. A écailles ou plaques irrégulières
- 119. A écailles ou plaques circulaires
- 120. Crêtes sinueuses parallèles
- 121. Excroissances bosselées
- 122. Excroissances pointues
- 123. Epines ramifiées
- 124. Epines coniques



Annexe 4 : Capture de l'interface Albumentations permettant le test des différentes augmentations de données

The screenshot displays the Albumentations demo interface in a web browser. The browser's address bar shows the URL `https://albumentations-demo.herokuapp.com`. The interface is divided into a control panel on the left and a main display area on the right.

**Control Panel (Left):**

- Upload my image:** A dropdown menu.
- Upload your image (jpg, jpeg, or png):** A text input field containing `AFBI_39_1_36.jpg` and a `browse files` link.
- Select transformation NP1:** A dropdown menu set to `Rotate`.
- Select transformation NP2:** A dropdown menu set to `None`.
- Params of the Rotate:**
  - limit:** A slider ranging from `-360` to `360`, with markers at `-90` and `90`.
  - interpolation:** Radio buttons for `0`, `1`, `2`, `3`, and `4`. `4` is selected.
  - border\_mode:** Radio buttons for `0`, `1`, `2`, and `3`. `0` is selected.

**Main Display Area (Right):**

- Demo of Albumentations:** The main heading.
- Original image:** A photograph of a rock surface.
- Transformed image:** The same rock surface, rotated counter-clockwise.
- Press 'R' to refresh:** A text prompt.
- Random params used:** A JSON object showing the parameters used for the transformation:

```
{
  "Rotate": {
    "angle": -15.464944957387395
  }
}
```
- Docstring for Rotate:** A text block stating: `Rotate the input by an angle selected randomly from the uniform distribution.`

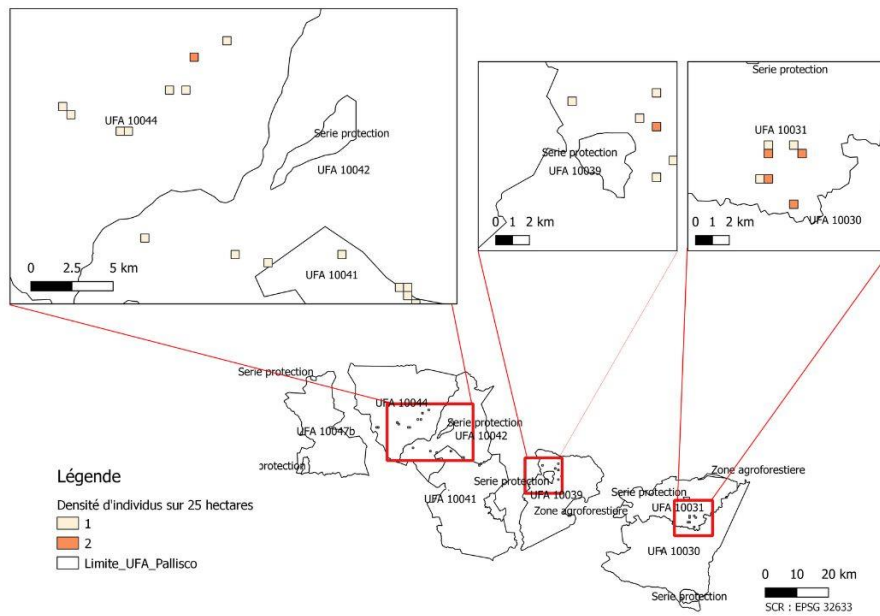


Annexe 5 : Nombre d'individus par classe et catégorie de division du jeu de données

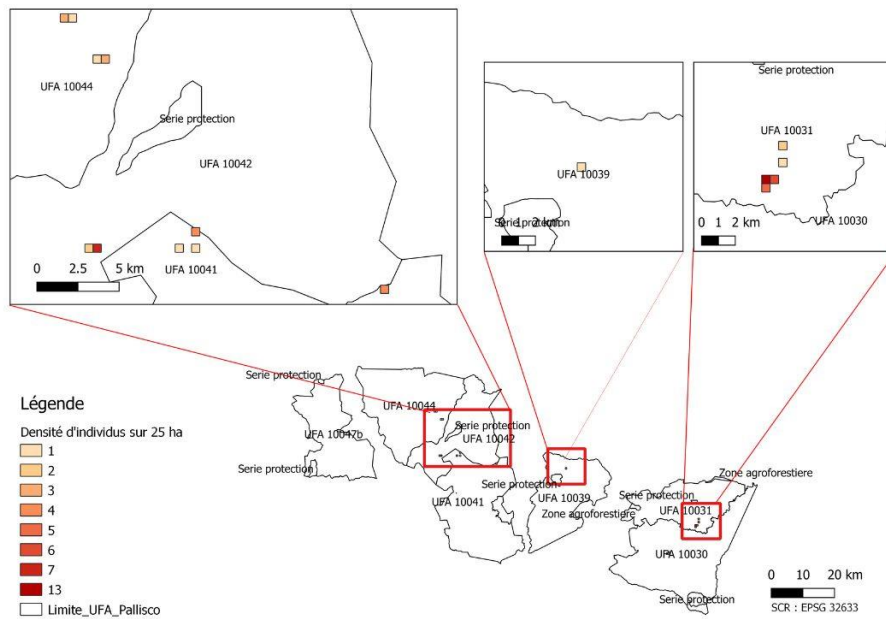
<b>Espèce</b>	<b>Entraînement</b>	<b>Validation</b>	<b>Test</b>	<b>Total</b>
0	1833	258	517	2608
1	1301	185	372	1858
2	1996	283	572	2851
3	2336	326	652	3314
4	1658	235	471	2364
5	1737	249	483	2469
6	2234	318	639	3191
7	1508	208	424	2140
8	1357	197	391	1945
9	1609	229	460	2298
10	1769	236	494	2499
11	4031	570	1153	5754
12	1875	274	551	2700
13	1284	181	369	1834
14	2909	418	832	4159
15	1282	180	368	1830
16	3557	453	1015	5025
17	2158	381	612	3151
18	1595	231	453	2279
19	1437	199	419	2055
Total général	39466	5611	11247	56324

Annexe 6 : Cartographie de la distribution des individus de chaque espèce par une cartographie de la densité ( chaque pixel représente 25 hectares)

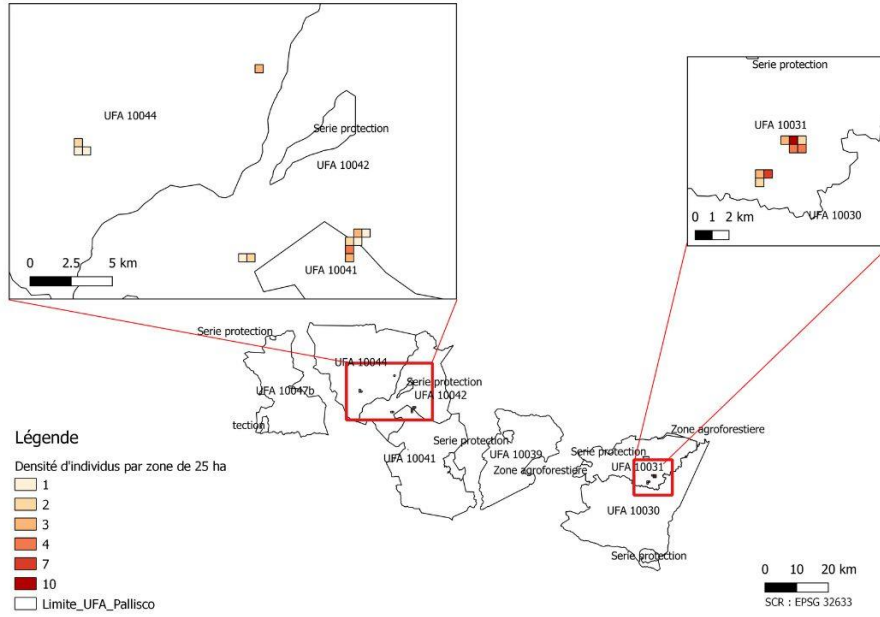
## Afzelia bipindensis



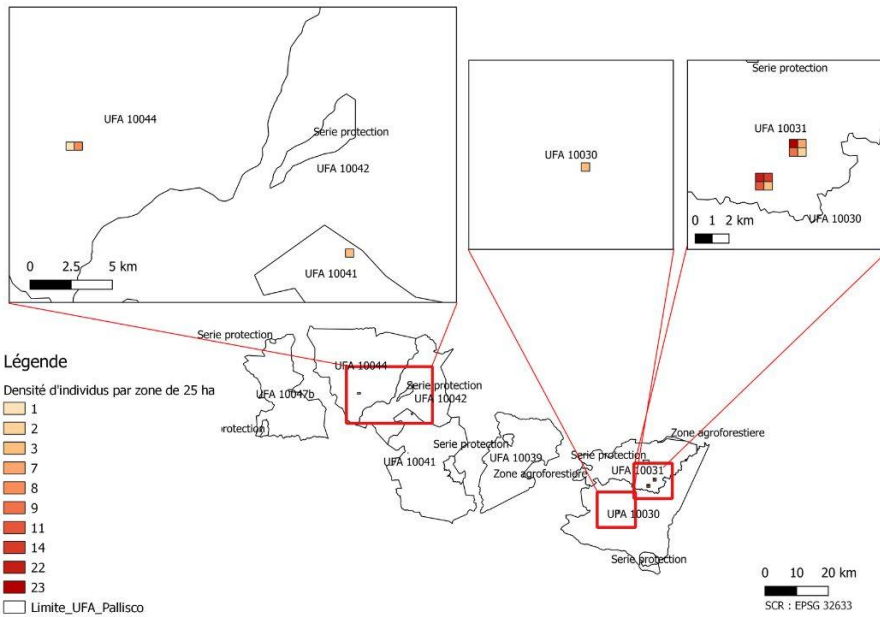
## Afrostyrax lepidophyllus



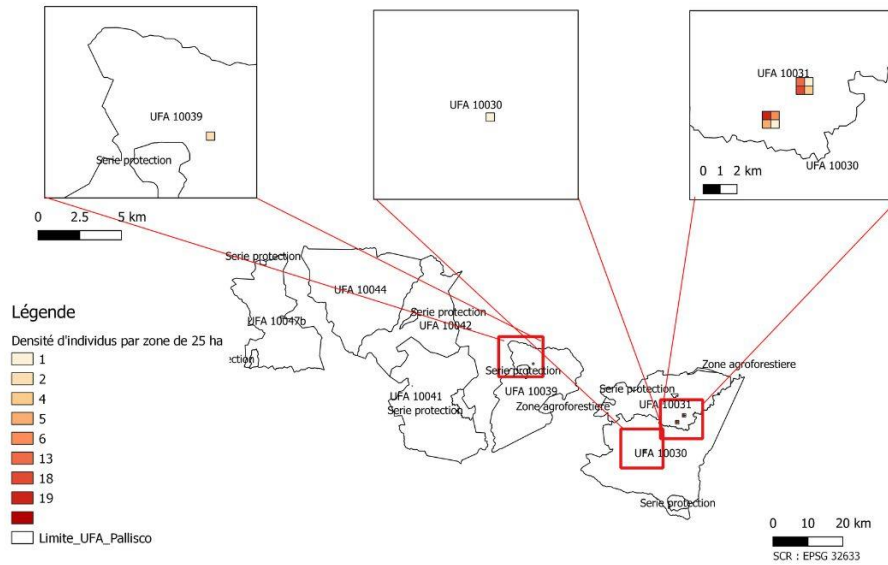
# Alstonia boonei



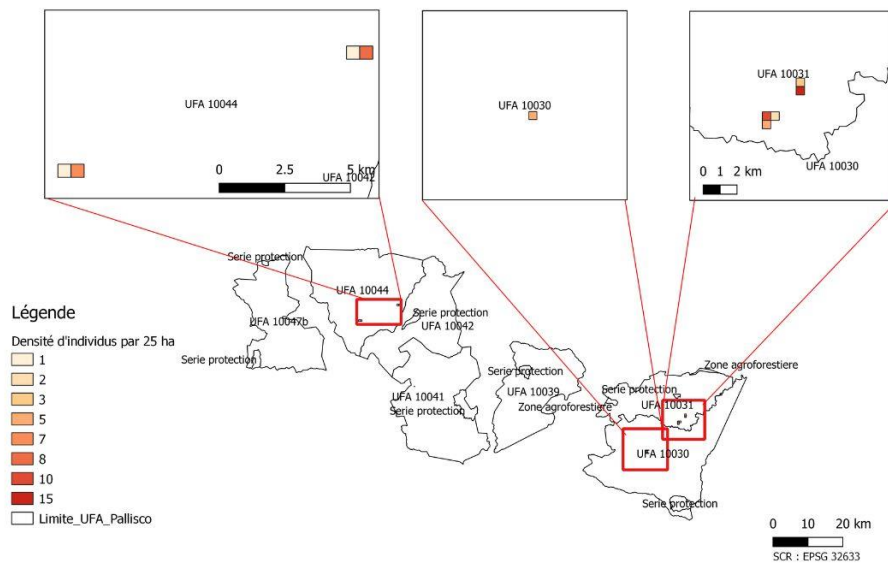
# Annickia affinis



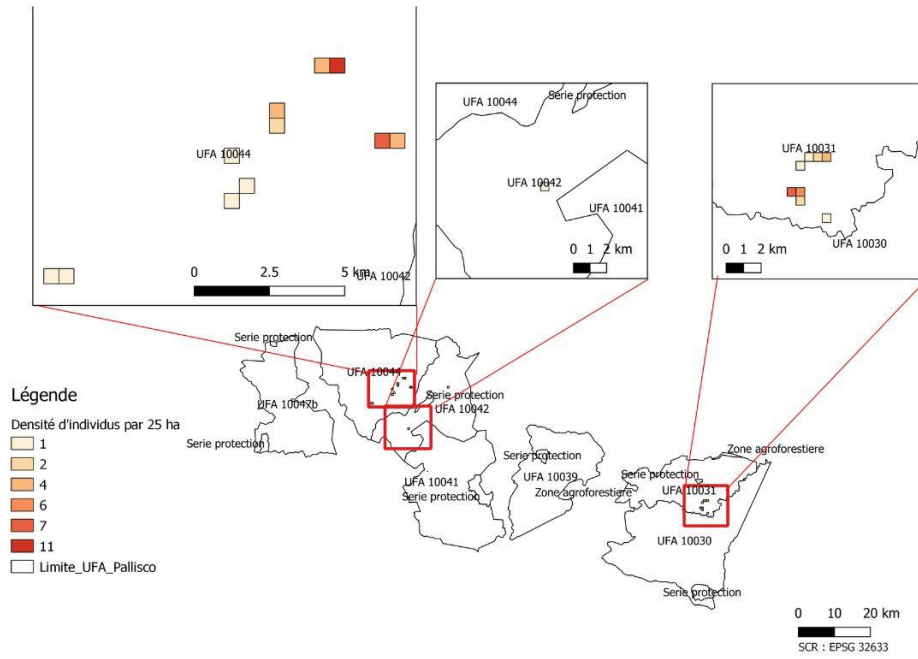
# Corynanthe pachyceras



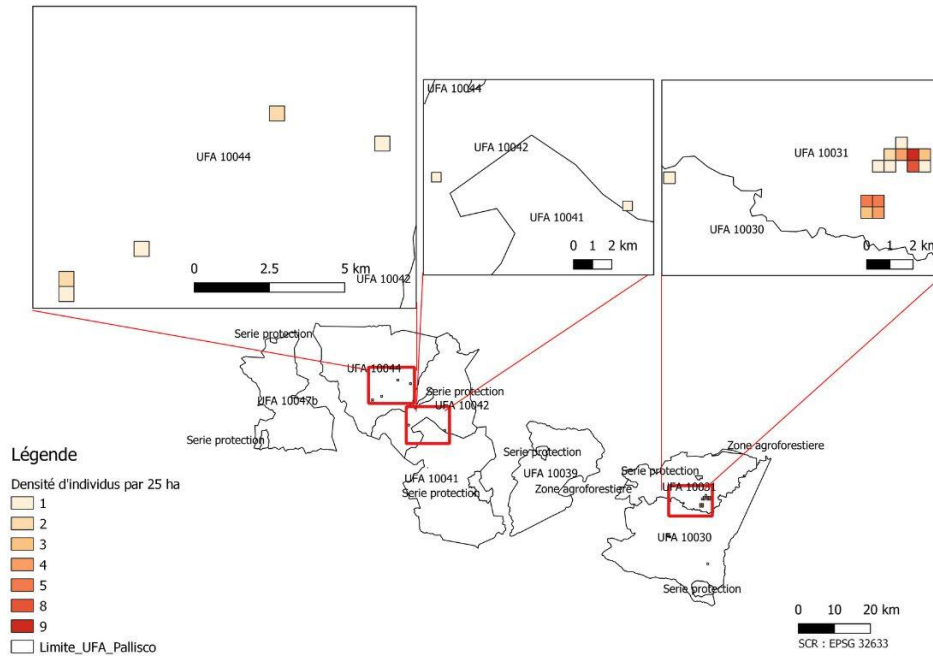
# Desbordesia glaucescens



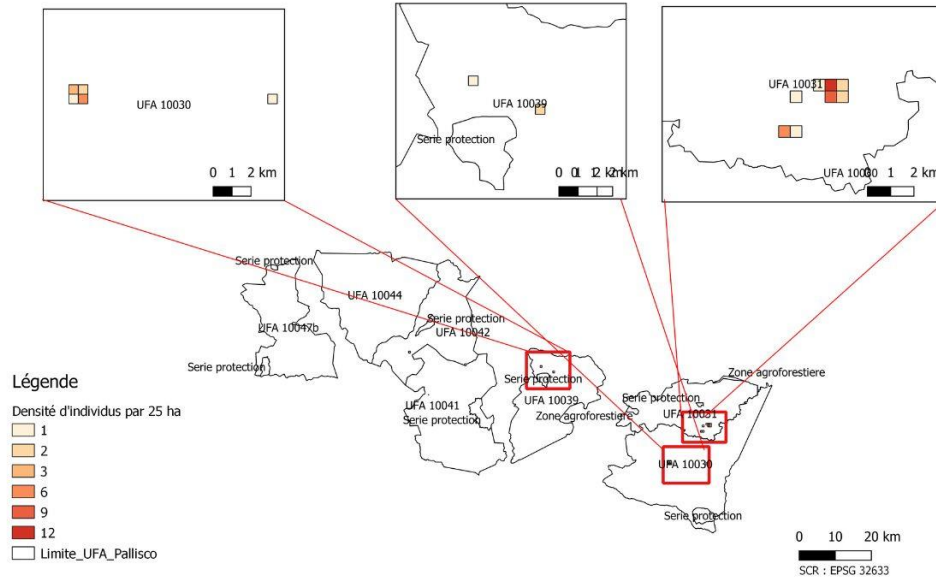
# Entandrophragma cylindricum



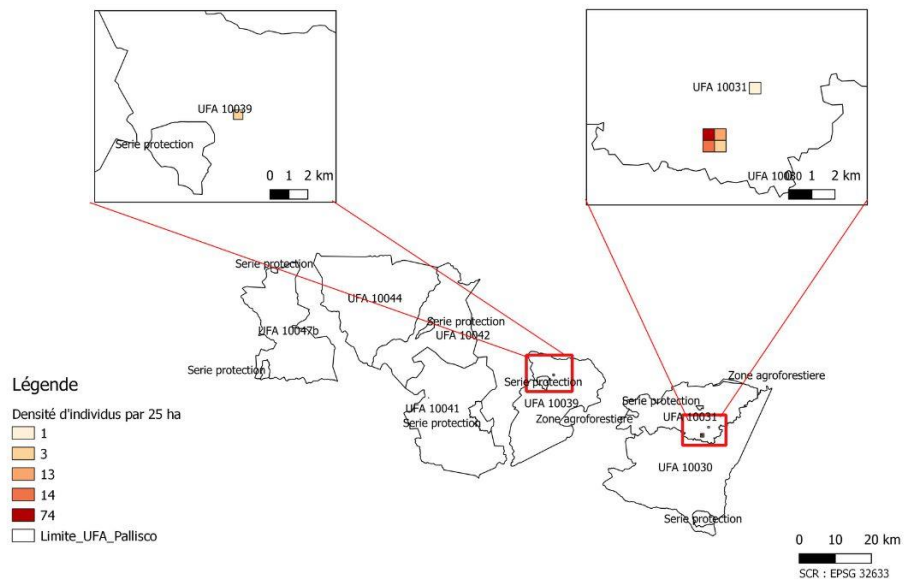
# Irvingia gabonensis



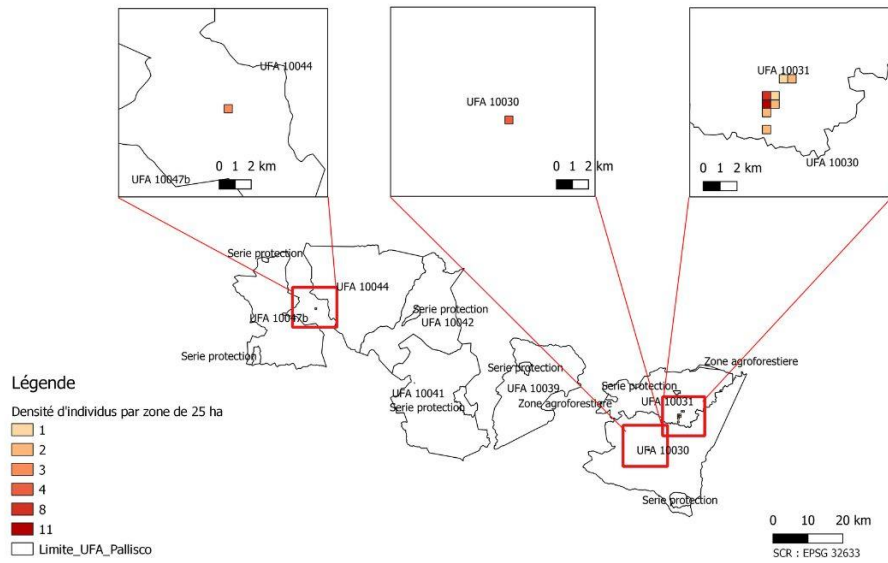
# Keayodendron bridelioides



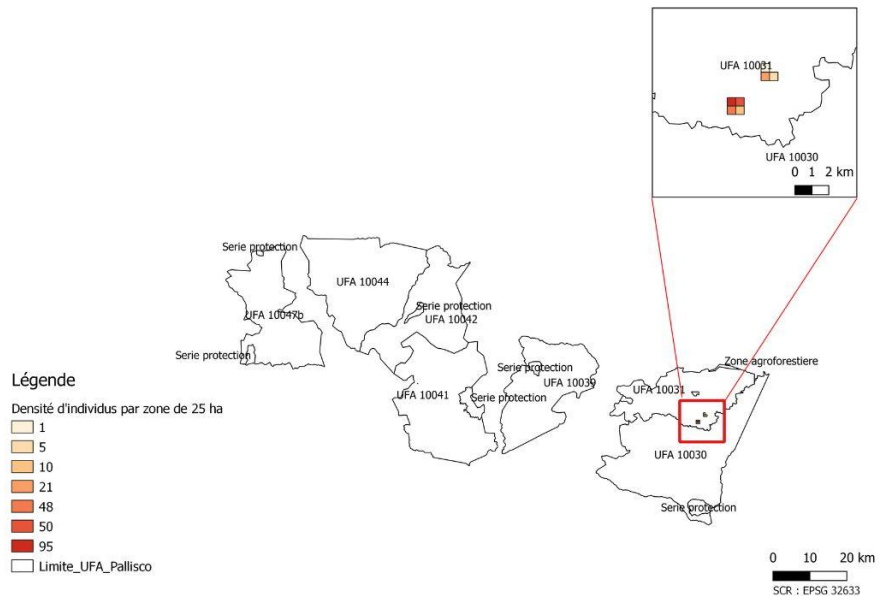
# Meiocarpidium lepidotum



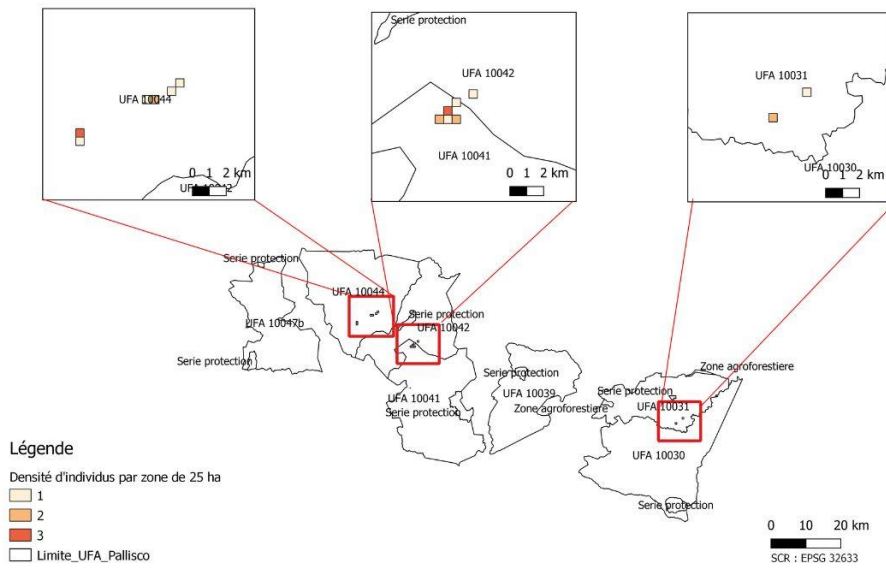
# Pericopsis elata



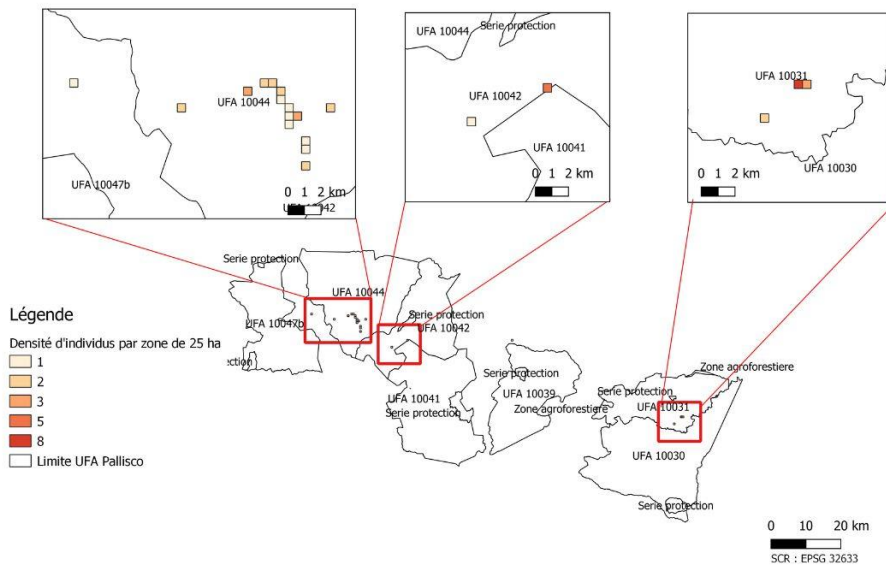
# Polyalthia suaveolens



# Pterocarpus soyauxii

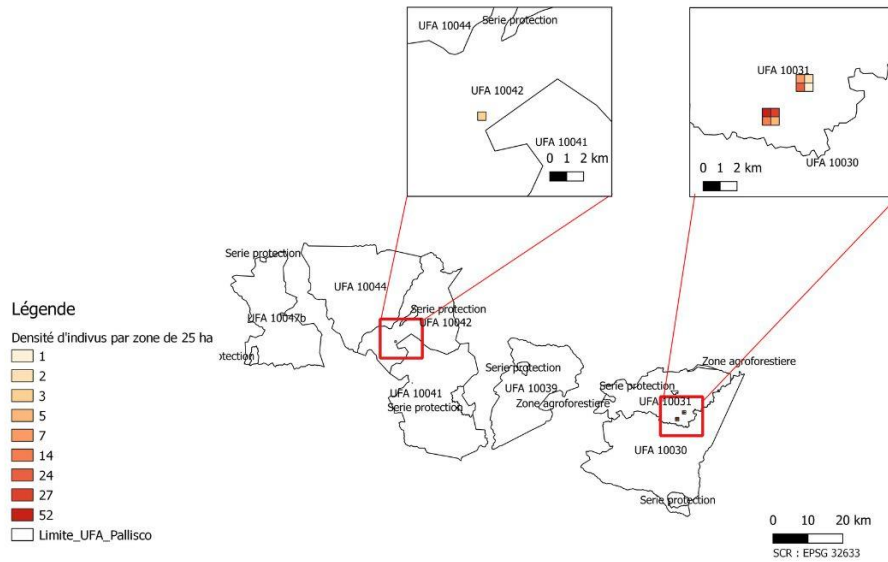


# Rauvolfia macrophylla

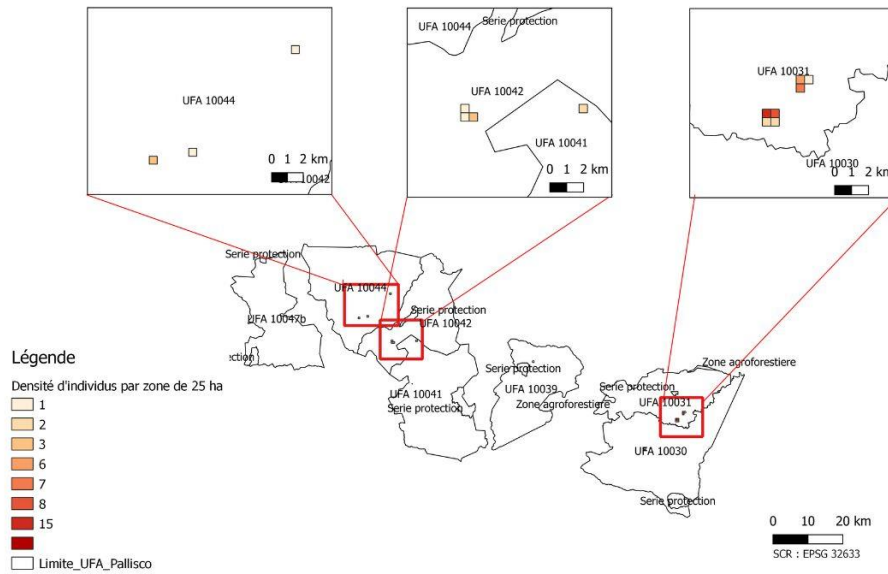




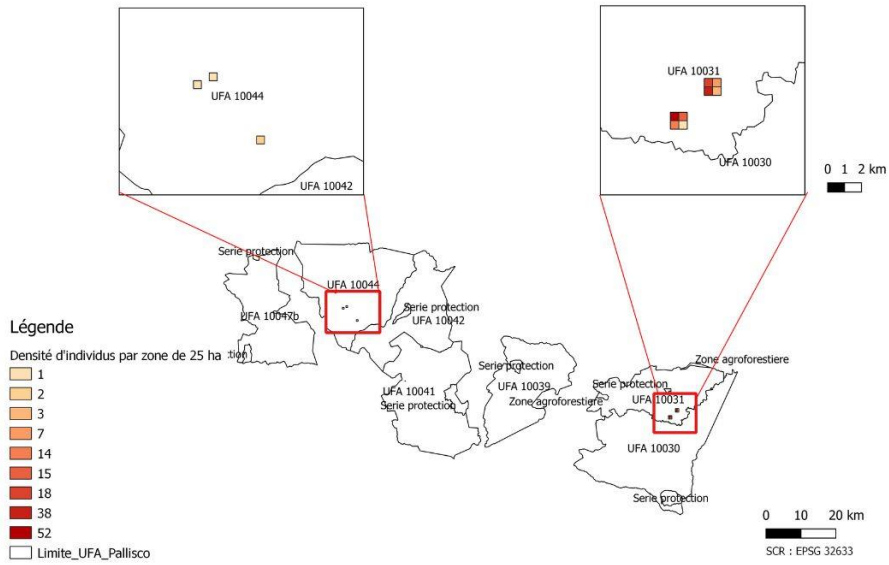
# Strombosia pustulata



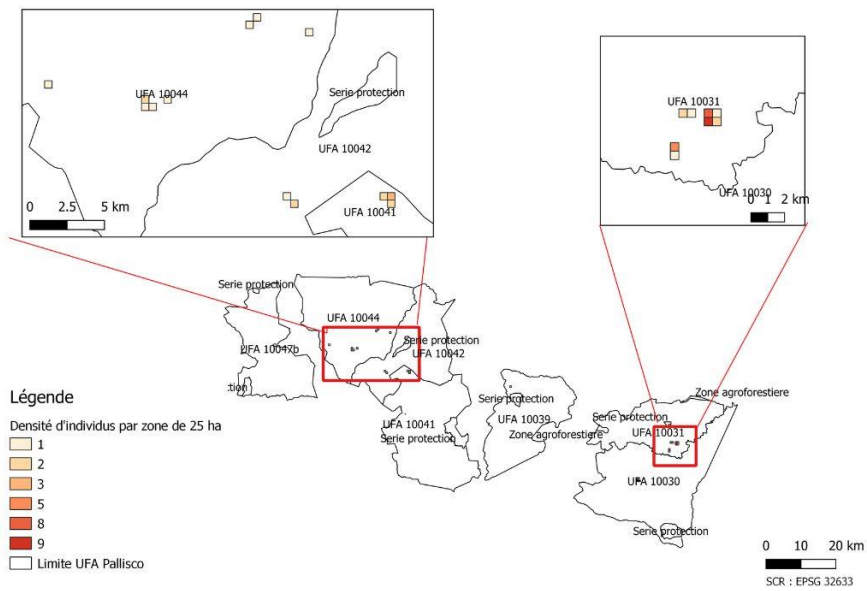
# Strombosiosis tetrandra



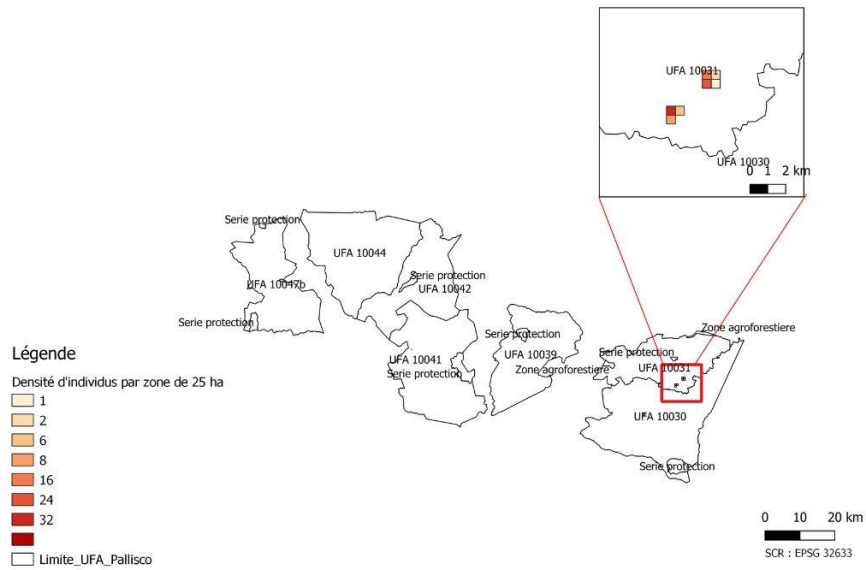
# Tabernaemontana crassa



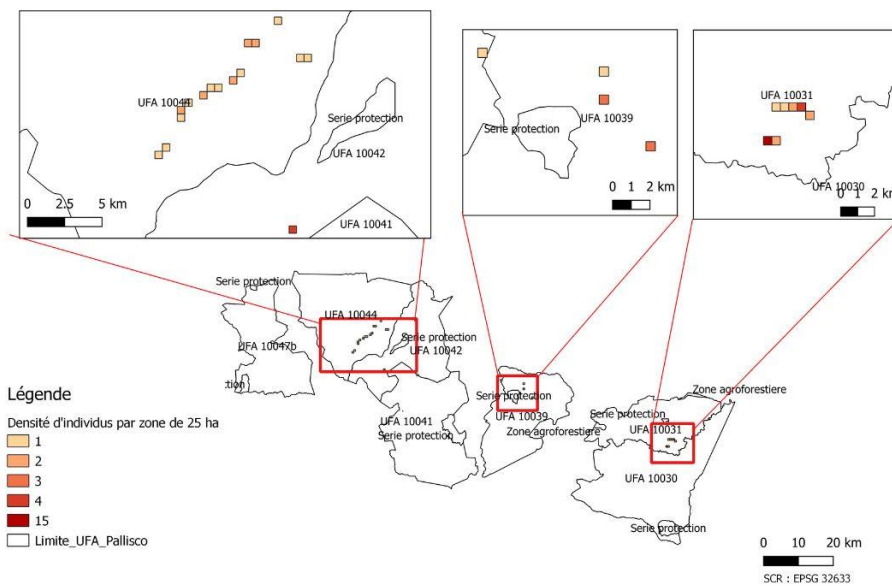
# Terminalia superba



# Trichilia welwitschii

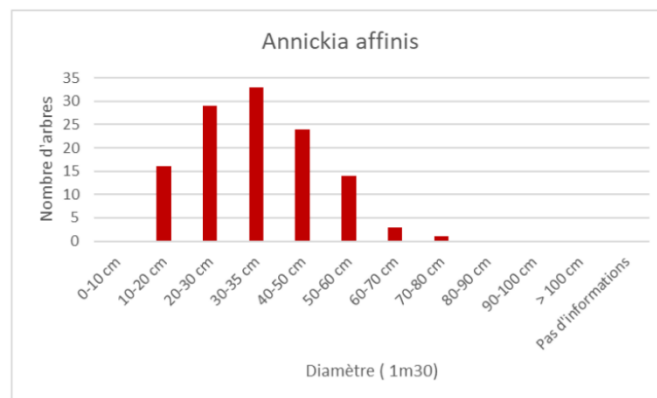
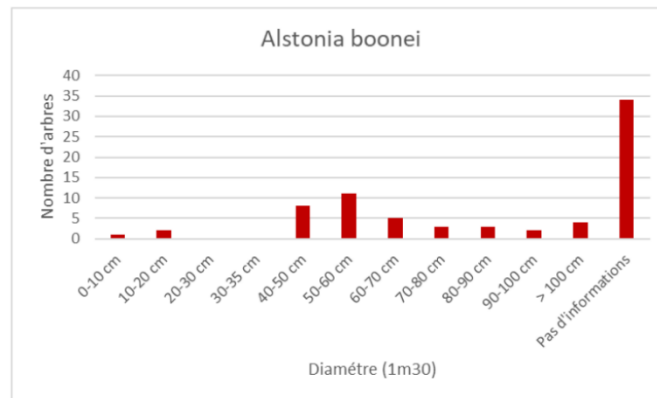
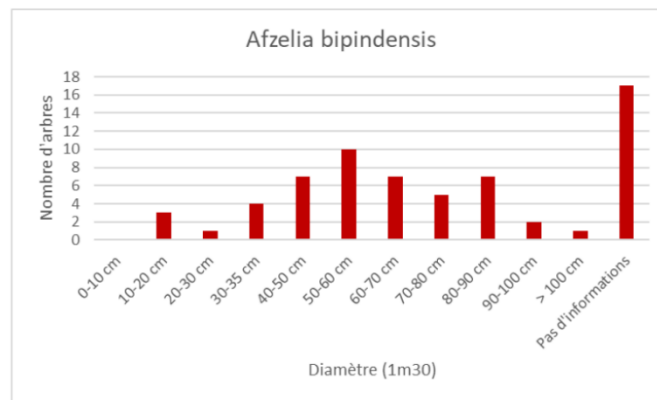
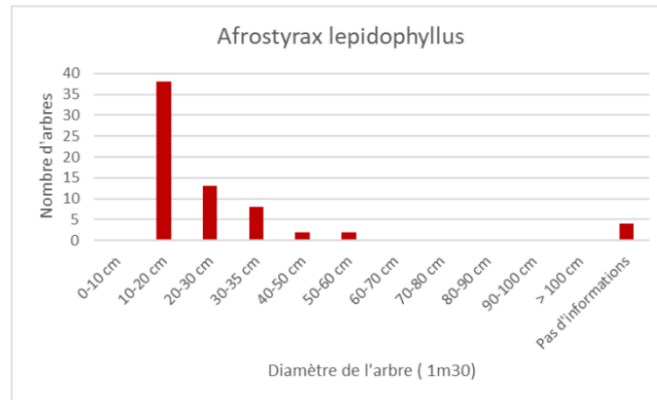


# Xylopiya aethiopica

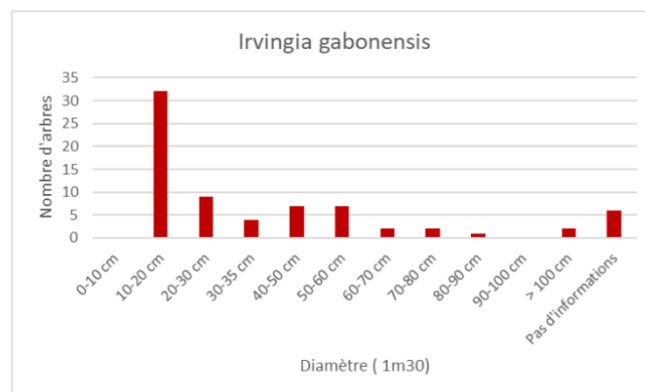
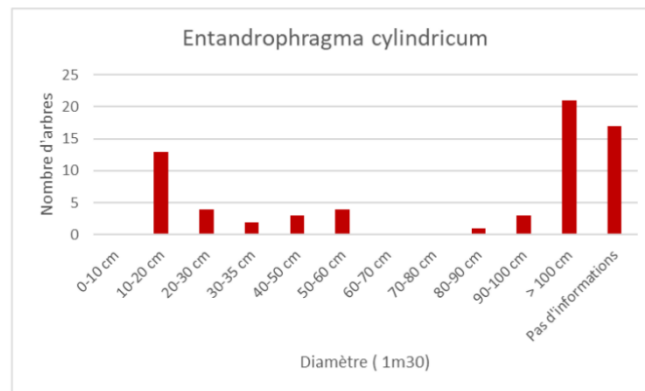
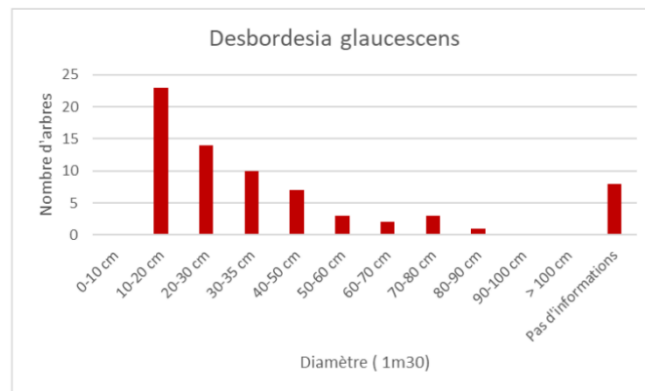
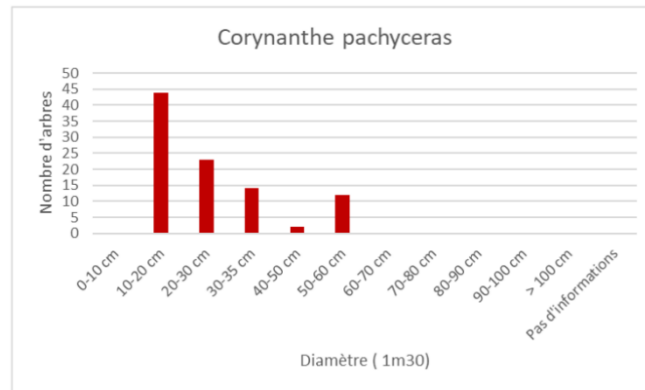


Annexe 7 : Histogrammes du nombre d'arbres par classe de diamètres pour l'ensemble des espèces ( Page A :  
 Espèces en A, Page B : Espèces de C à I , Page C : Espèces de K à Po, Page D : Espèces de Pt à S, Page E :  
 Espèces de T à X)

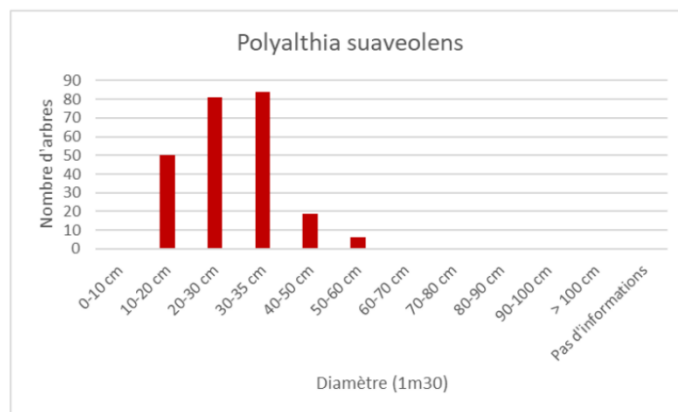
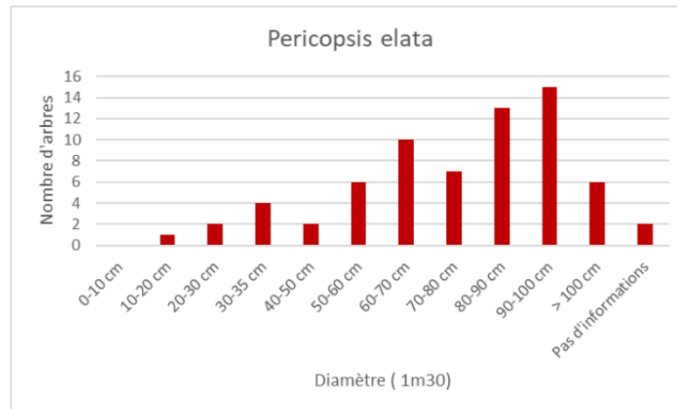
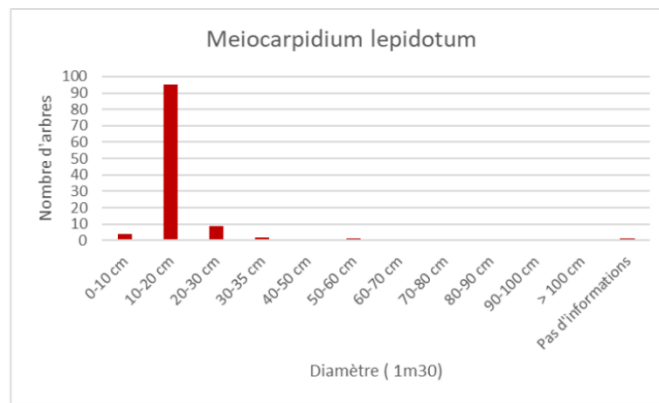
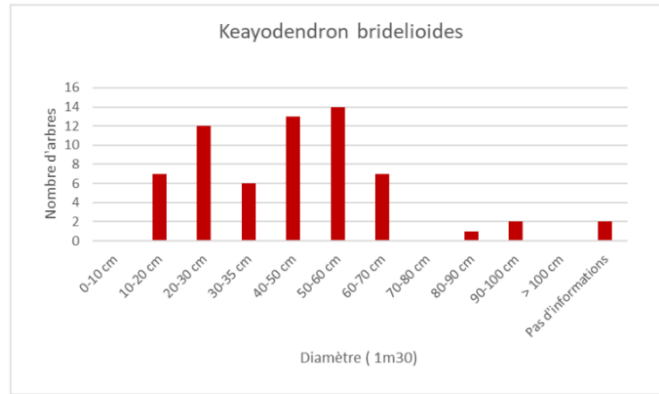
A



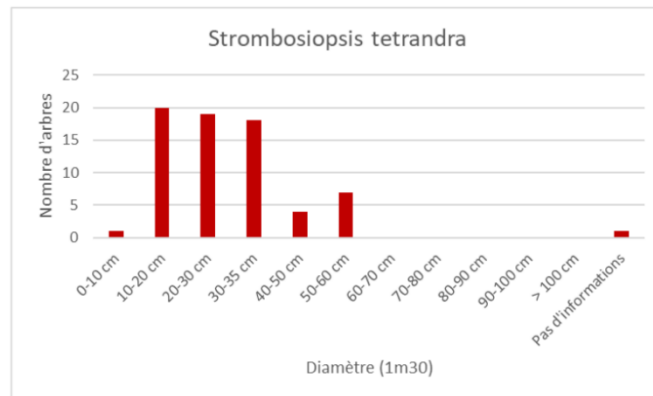
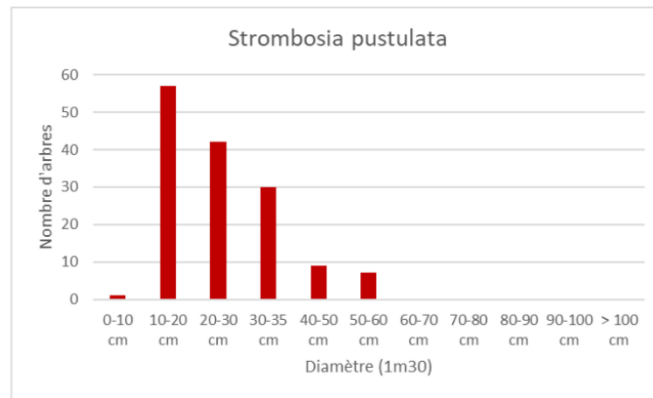
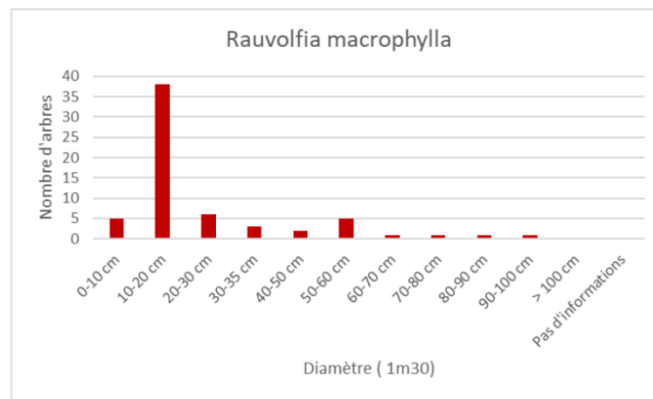
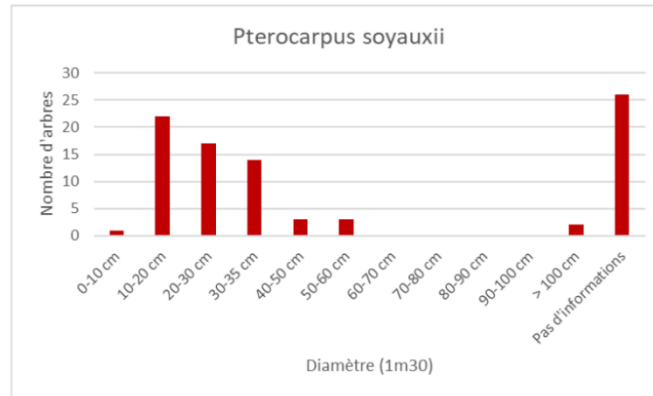
B



C



D



E

