

IA4DIVE : l'IA pour le développement de bioplastiques à base de racine d'enDIVE

1 Le sujet de recherche choisi et le contexte scientifique et économique

Cette thèse se déroulera dans le cadre du projet **DIVE** (Développement Innovant pour la Valorisation de la racine d'Endive) qui s'inscrit dans une dynamique régionale, européenne et mondiale qui a pour objectif la diminution de la pollution plastique et la réduction de ses usages, notamment dans le secteur agro-alimentaire. **DIVE** souhaite donc développer la valorisation d'un co-produit de la filière endive : la racine d'endive après forçage afin de (i) créer des alternatives sûres, saines, efficaces et écologiques pour la conservation alimentaire ; (ii) créer ainsi une nouvelle sous-filière structurante pour une filière d'importance régionale dans les Hauts-de-France et actuellement en crise.

Le projet **DIVE** est un projet interdisciplinaire qui réunit autour d'une thématique d'économie circulaire des experts des domaines suivant : la biochimie, le génie des procédés, les matériaux, l'informatique et les sciences humaines et sociales, le marketing et l'économie. Il découle d'une partie des résultats obtenus dans le projet région Hauts-de-France Economie Circulaire et Nouveaux Modèles de Développement – FermEndive qui a permis d'établir la preuve de concept d'un emballage alimentaire réalisé à partir de racines d'endive.

Le projet **DIVE** s'appuiera sur des travaux qui porte sur l'utilisation de la catalyse enzymatique pour améliorer les propriétés mécaniques, structurelles et biologiques de matériaux de type bioplastiques pour des alternatives d'emballages alimentaires biosourcés et compostables conçus à partir de racines d'Endive afin d'améliorer les propositions d'emballage développés dans FermEndive et de proposer de nouvelles alternatives d'emballage alimentaire (voir figure 1).

Cette thèse s'articule autour de l'axe informatique du projet **DIVE**, qui poursuit trois objectifs. Il s'agira dans un premier temps d'obtenir et de traiter les données sur les propriétés des bioplastiques afin de nourrir plusieurs méthodes d'apprentissage, telles que les réseaux de neurones profond (DNN) ou des forêts d'arbres aléatoires (RF). Le rôle de ces méthodes sera de prédire les propriétés mécaniques, structurelles, biologiques et antimicrobienne des emballages développés en fonction d'une part de la composition de la matière étudiées et d'autre part des procédés de transformation utilisés. De nombreuses expérimentations seront réalisées pour déterminer les approches les plus efficaces sur cette problématique. Ce modèle prédictif sera ensuite utilisé conjointement avec une métaheuristique pour réaliser le processus inverse : prédire les procédés et la composition de la biomasse nécessaires pour obtenir des propriétés biomécaniques désirées pour les emballages alimentaires.

2 L'état du sujet dans le laboratoire d'accueil

Le laboratoire d'accueil est le laboratoire Modélisation, Information & Systèmes (MIS UR 4290) de l'Université de Picardie Jules Verne. Le sujet se déroulera au sein du domaine Optimisation, Cryp-

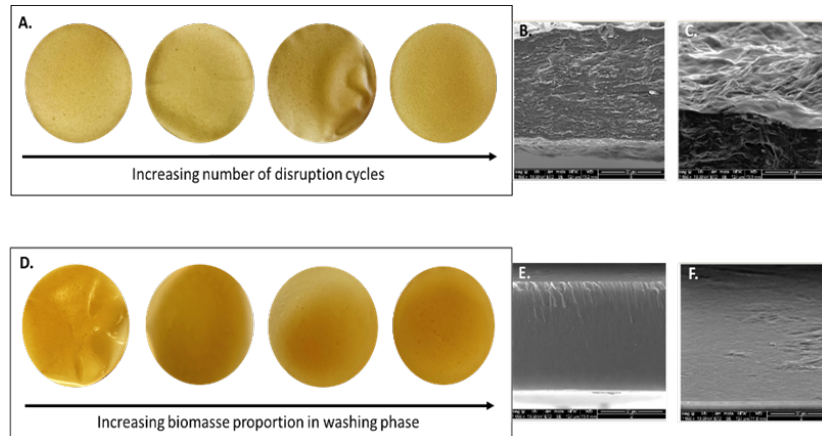


FIGURE 1 – Illustrations de bioplastiques

tographie, Intelligence Artificielle (OCIA) qui possède une expertise en matière d'IA et d'optimisation. Les principaux acteurs de ce projet ont encadré de nombreuses thèses sur l'optimisation de la localisation de ressources, sur la construction de tournées de véhicules ou la planification de tâches et de personnels. Les techniques de la recherche opérationnelle sont bien connues et maîtrisées par l'équipe [2, 10, 7]. Les outils de l'Intelligence artificielle sont également utilisés et combinés aux algorithmes d'optimisation [5].

3 Les objectifs visés et les résultats escomptés

La thèse consistera à partir des données récoltées dans le projet mais également à partir des données de la littérature, d'étudier la ou les corrélation(s) existante(s) entre la composition de la matière et les propriétés mécaniques, structurales, biologiques et antimicrobiennes des solutions d'emballages développées. L'étude de ces corrélations permettra de développer un outil de modélisation prévisionnelle permettant de faire le lien entre la composition de la matière brute et les propriétés des films obtenus à l'aide de la mise en place d'un outil de machine learning basé à la fois sur l'acquisition de données du projet **DIVE** et l'intégration de données bibliographiques déjà existantes. Trois types de données pourront être saisies dans l'outil conçu : des données de compositions de matières, des données concernant les propriétés des emballages ainsi que les informations en lien avec les différentes étapes du procédé de formulation des emballages. Les objectifs de cet outil seront d'une part de prédire quels types d'emballage il est possible d'obtenir en fonction de la composition de la matière première et d'autre part de tenir compte de la variabilité de composition de la matière première, liée aux conditions environnementales, pour adapter le procédé de formulation afin de préserver les propriétés souhaitées pour l'emballage (voir [15, 17, 19]).

Pour parvenir à ce résultat nous utiliserons des méthodes de régression issues du machine learning, incluant certaines méthodes à base d'arbres : arbres de décision [13], forêt d'arbres aléatoires [1] ou l'amplification de gradient [6], qui se sont montrées efficaces sur des problèmes de prédiction similaires. Ces méthodes seront également comparées à des approches basées sur les réseaux de neurones : perceptron multicouche [14], réseaux de neurones artificiels [11], réseaux de neurones profonds [8] ou encore réseaux de neurones convolutifs [9]. Des méthodes à noyaux, telles que les machines à vecteur de support [3] et la régression ridge à noyau [18], seront aussi envisagées.

L'efficacité de l'outil développé sera testée en laboratoire afin de comparer les modèles développés par l'outil et les résultats expérimentaux obtenus à l'échelle laboratoire. À partir de l'outil de modélisation prévisionnel, un mécanisme inverse sera proposé [12, 4]. Celui-ci, à partir de propriétés ciblées du polymère identifiera les caractéristiques de la biomasse et procédés d'obtention.

4 Le programme de travail et l'échéancier prévisionnel

Les objectifs se décomposent en 3 axes :

1. Traitement des données brutes. Le premier objectif est d'obtenir un dataset exploitable et fiable pour pouvoir construire un modèle prédictif pertinent répondant au problème. Cela passera par plusieurs opérations sur les valeurs associées : Nettoyage, Normalisation, Imputation intelligente et Étude de la corrélation des variables.
2. Étude et conception du modèle prédictif. Il existe plusieurs travaux de la littérature présentant différents modèles d'apprentissage tels que les réseaux de neurones convolutionnels (CNN), les réseaux de neurones profonds (DNN), les machines à vecteur de support (SVM), les forêts aléatoires (RF), la régression linéaire (LR), etc. Notre objectif est de trouver le meilleur type de modèle adapté à notre problème et pour cela définir un grand nombre d'expérimentations. Il faudra établir un protocole d'entraînement efficace et tester la précision du modèle obtenu. Une attention particulière sera portée à l'explicabilité de celui-ci.
3. Processus inverse. À l'issue de ce premier modèle, nous souhaitons établir un processus inverse au modèle prédictif. Il aura pour but de déterminer les meilleures caractéristiques de la biomasse et les procédés de transformation à employer afin d'obtenir des propriétés ciblées du polymère. Les méthodes issues de la recherche opérationnelle comme les algorithmes évolutionnaires seront utilisées pour la mise en œuvre de cette dernière phase.

L'échéancier associé est présenté figure 2.



FIGURE 2 – Diagramme de Gantt du projet

5 Les collaborations prévues

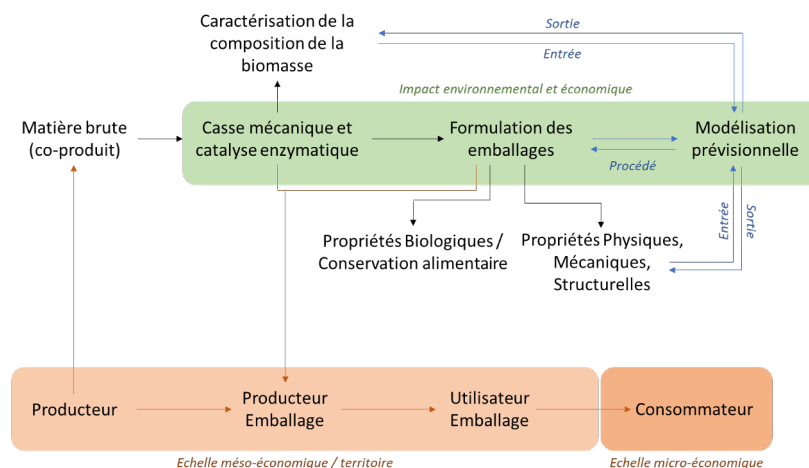


FIGURE 3 – Interactions dans le projet

Les collaborations interdisciplinaires sont indispensables à ce projet (voir figure 3). Tant en amont pour récolter un maximum de données sur des bioplastiques qui ont été conçus dans le cadre du projet **DIVE**, qu'en aval où les résultats des modèles prédictifs pourront être confirmés en produisant les bioplastiques à partir des caractéristiques d'entrée. De la même manière, le modèle prédisant le processus inverse devra faire appel aux compétences interdisciplinaires pour confirmer les propriétés souhaitées. Les partenaires sont (entres autres) :

- BioEcoAgro (UPJV, U-Lille) : l'Unité Mixte de Recherche Transfrontalière BioEcoAgro, fondée en janvier 2020, rassemble des chercheurs de l'INRAE, de l'Université de Liège, de U-Lille, de Junia, de l'UPJV, de l'Université d'Artois et de l'ULCO ;
- LG2A (UPJV) : Le Laboratoire de Glycochimie et des Agroressources d'Amiens (LG2A) dédie ses travaux de recherche majoritairement aux glycosciences ;
- UMET (U-Lille) : L'UMR Unité Matériaux Et Transformations (UMET) regroupe aujourd'hui une partie des activités de science des matériaux sur le site du campus scientifique de l'Université de Lille, et des activités de recherche sur la caractérisation des interactions entre bactéries et matériaux.

Côté informatique, l'implication de Sara Tari (ULCO, LISIC) sur ce projet permettra de renforcer l'étude et la conception du processus inverse en examinant les paysages adaptatifs induits par les caractéristiques de la biomasse [16].

6 Informations Pratiques

La thèse débutera en octobre 2026, pour une durée de 3 ans. Elle se fera au sein du laboratoire de recherche Modélisation, Information & Systèmes (MIS), de l'Université de Picardie Jules Verne (UPJV) situé à Amiens. L'équipe d'accueil au sein du laboratoire est l'équipe OCIA (Optimisation, Cryptographie, Intelligence Artificielle).

Encadrants :

- Olivier Gérard (olivier.gerard@u-picardie.fr), Enseignant/Chercheur au laboratoire MIS, équipe OCIA, UPJV ;
- Laure Brisoux Devendeville (laure.devendeville@u-picardie.fr), Professeur des Universités au laboratoire MIS, équipe OCIA, UPJV.

References

- [1] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [2] S. Caillard, L. Brisoux Devendeville, and C. Lucet. SimU-TACS: Ant Colony System for a Planning Problem in Health Simulation Training. *Applied Soft Computing Journal*, page 110848, 2023. doi : 10.1016/j.asoc.2023.110848.
- [3] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- [4] S. Degaugue, O. Gérard, J. Scouarnec, C. Lucet, S. Tari, and L. Brisoux Devendeville. Generation of instances with estimated landscape features for the two level p-median location problem. In M. S. Krejca and N. Pillay, editors, *Evolutionary Computation in Combinatorial Optimization*, pages 52–67, Cham, 2026. Springer Nature Switzerland.
- [5] M. Fagot, L. Brisoux Devendeville, and C. Lucet. Adaptive local search for a pickup and delivery problem applied to large parcel distribution. In B. Dorronsoro, F. Chicano, G. Danoy, and E.-G. Talbi, editors, *Optimization and Learning*, pages 186–199, Malaga, Espagne, 3–5 May 2023. Springer Nature Switzerland. doi : 10.1007/978-3-031-34020-8_14.
- [6] J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
- [7] O. Gérard, C. Lucet, L. Brisoux Devendeville, and S. Darras. An Adaptive Large Neighborhood Search Method to Plan Patient’s Journey in Healthcare. In *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems - IFIP WG 5.7 International Conference, APMS 2021*, 631, pages 289–287, Nantes, France, 5–9 Sept. 2021. doi : 10.1007/978-3-030-85902-2_31.
- [8] G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.
- [9] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.
- [10] C. Lucet, S. Caillard, and L. Brisoux Devendeville. Variable neighborhood search for a planning problem with resource constraints in a health simulation center. *Applied Intelligence*, 52:6245–6261, 2021. doi : 10.1007/s10489-021-02730-7).
- [11] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- [12] G. Paliana, C. N. Iverson, T. Lookman, and B. L. Marrone. Machine-learning-based predictive modeling of glass transition temperatures: a case of polyhydroxyalkanoate homopolymers and copolymers. *Journal of Chemical Information and Modeling*, 59(12):5013–5025, 2019.
- [13] J. R. Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

- [14] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.
- [15] H. Salma, Y. M. Melha, L. Sonia, H. Hamza, and N. Salim. Efficient prediction of in vitro piroxicam release and diffusion from topical films based on biopolymers using deep learning models and generative adversarial networks. *Journal of Pharmaceutical Sciences*, 110(6):2531–2543, 2021.
- [16] J. Scouarnec, C. Lucet, S. Tari, and L. Brisoux Devendeville. Healthcare facility location problem and fitness landscape analysis. In *European Conference on Evolutionary Computation in Combinatorial Optimization (Part of EvoStar)*, pages 217–231. Springer, 2025.
- [17] H. D. Tran, C. Kim, L. Chen, A. Chandrasekaran, R. Batra, S. Venkatram, D. Kamal, J. P. Lightstone, R. G. P. Shetty, M. R. J. Laws, M. Shelton, and R. Ramprasad. Machine-learning predictions of polymer properties with polymer genome. *Journal of Applied Physics*, 128(17), Novembre 2020.
- [18] G. Wahba. *Spline models for observational data*. SIAM, 1990.
- [19] S.-S. Wang, P. Lin, C.-C. Wang, Y.-C. Lin, and C.-W. Tung. Machine learning for predicting chemical migration from food packaging materials to foods. *Food and Chemical Toxicology*, 178:113942, 2023.