

الجمهورية الجزائرية الديمقراطية الشعبية  
République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



جامعة الإخوة منتوري قسنطينة I  
Frères Mentouri Constantine I University  
Université Frères Mentouri Constantine I

Faculté des Sciences de la Nature et de la Vie  
Département de microbiologie

كلية علوم الطبيعة والحياة  
قسم الميكروبيولوجيا

Mémoire présenté en vue de l'obtention du diplôme de Master

Domaine : Sciences de la Nature et de la Vie

Filière : biotechnologie

Spécialité : *Mycologie et biotechnologie fongique*

N° d'ordre :

N° de série :

Intitulé :

---

**Développement d'un modèle Machine Learning pour la  
prédiction de la productivité d'un Bioréacteur**

---

Présenté par : Kabache Yasmine  
Bouzahzah Mohamed Lokmane

Le 22/06/2023

Jury d'évaluation :

**Président du jury :** Mme. Abdelaziz Ouidad (MCB - Université Frères Mentouri, Constantine 1).

**Rapporteur :** Mme. Djama Ouahiba (MCB - Université Frères Mentouri, Constantine 1).

**Examineur :** Mme. Meziani Meriem (MCB - Université Frères Mentouri, Constantine 1).

Année universitaire  
2022 - 2023

## **Remerciement**

Tout d'abord, nous remercions Dieu tout-Puissant de nous avoir accordé la grâce de la raison, la santé et la poursuite de la connaissance et du courage, pour nous avoir donnés la force, la volonté et la patience.

Nous profitons également de cette occasion pour exprimer notre gratitude à chacun d'entre nous.

Nous exprimons nos profonds remerciements à notre Encadreur et Directrice de notre mémoire Mme Djama Ouahiba. Nous la remercions pour sa patience, sa disponibilité et surtout ses sages conseils qui ont contribué à notre réflexion et sans qui ce travail n'aurait pas pu être arrivé à bon port. Nous la remercions d'avoir eu la patience de répondre à nos questions et pour le temps qu'elle nous a consacré tout au long de cette expérience combien enrichissante pour nous deux.

Nous tenons également à remercier chaleureusement tous les membres de notre jury de soutenance Mme Abdelaziz Ouidad et Mme Maziani Meriam.. Votre expertise, votre intérêt pour notre sujet et vos commentaires éclairés ont enrichi notre réflexion et ont été d'une valeur inestimable. Nous sommes honorés d'avoir pu bénéficier de vos connaissances lors de notre présentation.

Sans oublier de remercier Mlle Boubaker Nour El-Houda, doctorante à l'Université Abdelhamid Mehri Constantine 2, notre ami Chams Eddine Adjias, étudiant en informatique, pour leurs précieux conseils.

On remercie aussi Dr. Zamouche-Zerdazi Rania de la faculté Génie des procédés de l'UC3, pour sa serviabilité. Nous remercions Mme Abdelaziz Ouidad pour son aide lors de la collection de données.

Cordialement,

*Dédicace :*

*Allah tout puissant merci d'être toujours auprès de moi.*

*Je dédie ce travail à :*

*\_ Mon père Kabache Ahcene mon trésor et ma force dans ma vie.*

*\_ Ma mère Kabache Nora ma force, ma lumière de ma vie.*

*\_ Mes frères Houssame Addine, Mouhamed Amine, Imad et ma sœur*

*Hannane. Puisse-tu toujours être mon meilleur soutien*

*[Kabache Yasmine]*

*Chers Papa, Maman, Sœurs et Frère, ce dédicace de mémoire est un cadeau spécial pour vous exprimer tout mon amour et mon appréciation. Vous êtes des piliers dans ma vie, et chaque moment partagé ensemble est précieux. Ce dédicace est rempli de souvenirs soigneusement sélectionnés pour chaque membre de la famille. Je chéris notre amour, notre unité et notre connexion profonde. Merci d'être dans ma vie. Avec tout mon amour,*

*[Bouzahzah Mohamed Lokmane].*

## **Résumé :**

Notre travail appartient au domaine de biotechnologie qui sert à exploiter les techniques d'intelligence artificielle (IA) pour prédire la quantité d'une biomasse au niveau d'un bioréacteur en fonction de variance des certains paramètres. L'objectif de ce travail est de développer un modèle PPVC exploitant une des techniques de Machine Learning ou l'apprentissage automatique (ML) pour effectuer cette prédire. Sachant que, l'apprentissage automatique (ML) est une branche de l'intelligence artificielle (IA). Le modèle PPVC que nous avons créé est capable de prédire la quantité avec un taux de réussite très élevé (92.18%). Le champ applicatif de notre travail est la prédiction de la productivité du *Chlorella vulgaris* par un bioréacteur via lactosérum en variant les valeurs de : le rapport S/X, l'éclairage et le mode nutritionnel. Cela, permet de gagner de temps et le coût du matériel et matière car, nous avons la capacité de prédire la quantité sans faire des expériences.

## **Mots clé :**

Intelligence artificielle, Machine Learning (ML) ou l'apprentissage automatique, Prédiction, Bioréacteur.

**Abstract :**

Our work belongs to the field of biotechnology when we exploit the artificial intelligence techniques to predict the quantity of biomass in a bioreactor according to the variance of certain parameters. The objective of this work is to develop a PPVC model exploiting one of the techniques of Machine Learning (ML) to accomplish this prediction. Knowing that, ML is a branch of artificial intelligence (AI). The PPVC model that we created is able to predict the quantity with a very high success rate (92.18%). The field of application of our work is the prediction of the productivity of *Chlorella vulgaris* in a bioreactor via whey by varying the values of: the S/X ratio, the lighting and the nutritional mode. This saves time and the cost of material and material because we have the ability to predict the quantity without experimenting.

**Keywords:**

Artificial Intelligence, Machine Learning (ML), Prediction, Bioreactor.

## المخلص:

ينتمي عملنا إلى مجال التكنولوجيا الحيوية حيث قمنا باستغلال تقنيات الذكاء الاصطناعي للتنبؤ بكمية الكتلة الحيوية على مستوى مفاعل حيوي وفقاً لتباين معايير معينة. الهدف من هذا العمل هو تطوير نموذج PPVC باستغلال إحدى تقنيات التعلم الآلي أو التعلم الآلي (ML) لتنفيذ هذا التوقع. مع العلم أن ML هو فرع من فروع الذكاء الاصطناعي (AI). نموذج PPVC الذي أنشأناه قادر على التنبؤ بالكمية بمعدل نجاح مرتفع جداً (92.18%). مجال تطبيق عملنا هو التنبؤ بإنتاجية *Chlorella vulgaris* بواسطة مفاعل حيوي عبر مصل اللبن من خلال تغيير قيم: نسبة S / X والإضاءة والوضع الغذائي. هذا يوفر الوقت وتكلفة المواد والموارد لأن لدينا القدرة على التنبؤ بالكمية دون الحاجة إلى إجراء تجارب.

## كلمات مفتاحية:

الذكاء الاصطناعي، التعلم الآلي (ML)، التنبؤ، المفاعل الحيوي.

## Liste des figures :

<b>Figure 1</b> : les différents systèmes dans le bioréacteur .....	07
<b>Figure 2</b> : bioréacteur a agitation mécanique.....	09
<b>Figure 3</b> : bioréacteur à agitation pneumatique.....	10
<b>Figure 4</b> : Bioréacteur à colonne à bulles .....	11
<b>Figure 5</b> : bioréacteur à lit fluidisé .....	12
<b>Figure 6</b> : schéma d'un photobioréacteur .....	13
<b>Figure 7</b> : schéma d'un bioréacteur à membrane .....	14
<b>Figure 8</b> : La relation entre l'IA, ML, DL .....	21
<b>Figure 9</b> : Différents techniques pour le ML.....	22
<b>Figure 10</b> : Extrait du Dataset sur Excel.....	35
<b>Figure 11</b> : code python pour enrichie le Dataset.....	36
<b>Figure 12</b> : Extrait du Dataset enrichie sur Excel.....	36
<b>Figure 13</b> : Code python pour le chargement du mémoire et lecture du Dataset.....	37
<b>Figure 14</b> : Code python effectué la diffusion du Dataset (ensemble del'entraînement et l'ensemble pour le test).....	38
<b>Figure 15</b> : Code python effectué la construction le model PPCV.....	38
<b>Figure 16</b> : Extrait du Dataset issu de l'expérience réelle sur Excel. ....	39
<b>Figure 17</b> : code python pour le score d'entraînement du programme .....	40
<b>Figure 18</b> : code python pour le score de test du programme.....	40
<b>Figure 19</b> : code python pour le score d'entraînement du programme de Dataset ajusté.....	42
<b>Figure 20</b> : code python pour le score de test du programme de Dataset ajusté.....	42
<b>Figure 21</b> : code python pour la comparaison des valeurs réel avec les valeurs prédites.....	44
<b>Figure 22</b> : Evaluation des résultats de prédiction.....	46



## Liste des tableaux :

<b>Tableau 01</b> : les applications du bioréacteur.....	17
<b>Tableau 02</b> : Conditions d'incubation dans les bioréacteurs.....	28
<b>Tableau 03</b> : Quantité de la biomasse produite dans chaque expérience.....	29
<b>Tableau 04</b> : Caractéristiques de la machine utilisée pour le ML.....	31
<b>Tableau 05</b> : Principaux outils utilisés.....	33
<b>Tableau 06</b> : Différents bibliothèques python utilisées.....	34
<b>Tableau 07</b> : Les valeurs ajusté du Dataset.....	41
<b>Tableau 08</b> : comparaison de quantité réel avec quantité prédire du programme....	45

## **Les Abréviation**

- IA : Intelligence Artificielle
- ML : Machine Learning (Apprentissage Automatique)
- PBR : Photobioréacteur
- PPCV : Prédiction de la Production *Chlorella Vulgaris*
- Rapport (S/X) : Rapport (Substrat/ Biomasse)

# Sommaire

-Liste des figures

-Liste des tableaux

-Liste des abréviations

<b>Introduction.....</b>	<b>1</b>
<b>Chapitre I : Les bioréacteurs et leurs utilisations.....</b>	<b>3</b>
I.1 Définition.....	3
I.2 Description de bioréacteur.....	4
I.2.1 La cuve de fermentation.....	4
I.2.2 L'appareil de chauffage et de refroidissement (contrôle la température).....	4
I.2.3 Le système d'aération et d'agitation.....	5
I.2.4 Le système de contrôle de la mousse.....	6
I.2.5 Le système de contrôle du pH.....	6
I.3 Les différents types de bioréacteurs.....	8
I.3.1 Le bioréacteur à agitation mécanique.....	8
I.3.2 Le bioréacteur avec agitation pneumatique.....	9
I.3.3 Bioréacteur à colonne à bulles.....	10
I.3.4 Le bioréacteur à lit fluidisé.....	11
I.3.5 Les photobioréacteurs.....	12
I.3.6 Le bioréacteur à membrane.....	13
I.4 Les modes de fonctionnement.....	14
I.4.1 Le mode discontinu (ou batch).....	14
I.4.2 Le mode semi-continu (ou Fed- batch).....	15
I.4.3 Le mode continu (ou chemostat).....	16
I.5 Applications du bioréacteur.....	16
<b>Chapitre II : Concepts d'Intelligence Artificielle.....</b>	<b>18</b>
II.1 L'informatique.....	18
II.1.1 Définition.....	18
II.2 L'Intelligence Artificielle (IA).....	18
II.2.1 Définition.....	18

II.2.2 Historique .....	19
II.3 Apprentissage automatique ou Machine Learning (ML).....	22
II.3.1 Définition.....	22
II.3.2 Types de données.....	22
II.3.3 Types de techniques de ML.....	23
II.4 l'IA dans le domaine de la biologie.....	24
II.5 Processus de ML.....	25
II.5.1 Collection des données.....	25
II.5.2 Prétraitement des données.....	25
II.5.3 Choix du modèle.....	25
II.5.4 Entraînement du modèle.....	26
II.5.5 Évaluation.....	26
<b>Chapitre III : Matériel et méthode.....</b>	<b>27</b>
III.1 Matériel .....	27
III.1.1 données biologiques.....	27
III.1.2. Construction du Dataset.....	30
III.1.2.1. Les valeurs du rapport (s/x).....	30
III.1.2.2. Eclairage.....	30
III.1.2.3. Le mode nutritionnel.....	31
III.1.2.4. Quantité de la biomasse produite (g).....	31
III.1.3. Matériel informatique.....	31
III.1.4. Outils logiciels et bibliothèques.....	32
III.1.4.1. les Outils .....	32
III.1.4.2. Les bibliothèques.....	34
III.2 méthodes.....	35
III.2.1 Exploitation du Dataset.....	35
III.2.2 Apprentissage.....	38
III.2.2.1 Création du modèle.....	38
<b>Chapitre IV : Résultats et discussions.....</b>	<b>39</b>
IV.1 Résultats et discussions.....	39

IV.1.1 test d'overfitting.....	39
IV.1.2 Amélioration du modèle.....	40
IV.1.3 Test overfeeting avec Dataset ajusté.....	42
IV.1.4 Évaluation de la prédiction.....	43
<b>Conclusion.....</b>	<b>47</b>
<b>Références bibliographique.....</b>	<b>48</b>

# **Introduction**

## Introduction

Le lactosérum est défini comme la partie liquide résiduelle de la coagulation du lait (Bouznada *et al*, 2022). Il est utilisé comme source de substrat pour la production de la biomasse : microalgues *Chlorella vulgaris*. Pour cela, (Bouznada *et al*, 2022) ont exploité des photobioréacteurs (PBR) avec comme paramètres : le rapport substrat/biomasse (S/X), l'éclairage et le mode nutritionnel.

(Bouznada *et al*, 2022) ont varié les valeurs du rapport substrat/biomasse (S/X), l'éclairage et le mode nutritionnel afin de trouver à chaque fois la quantité finale de la biomasse *Chlorella vulgaris*. Pour trouver des quantités de *Chlorella vulgaris* pour des nouvelles valeurs de rapport substrat/biomasse (S/X), l'éclairage et le mode nutritionnel qui ne sont pas pris en charge dans les expériences de (Bouznada *et al*, 2022), il faut refaire ces expériences. La chose qui prend du temps et de matériel avec un coût élevé.

La solution de cette problématique est de construire un modèle de prédiction. La prédiction mathématique nécessite beaucoup de calculs. La chose qui prend du temps et augmente le risque des erreurs.

La solution idéale est de profiter des avantages de l'IA notamment sa branche apprentissage automatique ou Machine Learning (ML) pour créer un modèle de prédiction de la quantité de *Chlorella vulgaris* pour des nouvelles valeurs de rapport substrat/biomasse (S/X), l'éclairage et le mode nutritionnel sans refaire les expériences.

Donc, l'objectif visé dans ce travail est de développer un modèle de Machine Learning capable de réaliser cette prédiction avec un taux de réussite acceptable. De ce fait, nous allons introduire l'intelligence artificielle dans la Biotechnologie et notamment au niveau des bioréacteurs.

Ce mémoire est organisé selon le plan suivant :

Dans le premier chapitre, nous aborderons les bioréacteurs, en définissant et examinant en détail leur conception et en discutant de certains de leurs types, de leurs utilisations et de leur mode de fonctionnement. Dans le deuxième chapitre, nous aborderons la définition de l'informatique et de l'intelligence artificielle (IA), en mettant

en évidence quelques éléments historiques et l'usage de l'IA en biologie. Nous discuterons également de l'apprentissage automatique (Machine Learning - ML) et de ses techniques. Ensuite, nous terminons le chapitre par l'explication de processus de création des modèles dans le domaine du ML. Dans les troisième et quatrième chapitres, nous expliquerons notre travail en détaillant les matériaux utilisés, les techniques et les méthodes employées. Nous exprimerons également les résultats obtenus et les discussions en profondeur.



# **Chapitre I**

**Les bioréacteurs et leurs  
utilisations**

## **I.1 Définition**

Un bioréacteur est un type de cuve de fermentation utilisé pour produire divers produits chimiques et réactions biologiques. Il s'agit d'un récipient fermé avec un agencement approprié pour l'aération, l'agitation, le contrôle de la température et du pH et un drain de trop-plein ou un évent pour éliminer la biomasse résiduelle des micro-organismes cultivés avec leurs produits (Sanjogta, 2022).

Le bioréacteur est généralement un environnement biomécanique et un environnement biochimique, Il contrôle le transport des nutriments et de l'oxygène vers les cellules et les produits du métabolisme. Il peut également être défini comme un fichier dispositif conçu pour obtenir une croissance et une activité métabolique optimale d'un organisme grâce à l'action d'un biocatalyseur, enzyme ou micro-organismes et cellules d'animaux ou de plantes et la matière première peut être organique ou inorganique (Neha&Badri, 2022). Un composé chimique ou même une substance complexe (Neha&Badri, 2022).

Le produit de la transformation peut comprendre de la levure de boulanger, cellule protéines, ferments lactiques, aliments pour animaux, etc. ou métabolites primaires tels que les acides aminés, les acides organiques, vitamines, sucres, éthanol, etc. et métabolites secondaires par exemple : antibiotiques, etc. (Neha&Badri, 2022).

Les bioréacteurs peuvent être utilisés pour fabriquer des produits de biotransformation comme biotransformation de stéroïdes, L-sorbitol, etc. enzymes tel que l'amylase, lipase, cellulose, etc. et des produits recombinants par exemple : certains vaccins, des hormones comme l'insuline, des hormones de croissance, etc. (Neha&Badri, 2022).

## I.2 Description de bioréacteur

La conception et le mode de fonctionnement d'un bioréacteur reposent sur la production d'un organisme, les conditions optimales requises pour la formation du produit souhaité, la valeur du produit et son échelle de production. Une bonne conception de bioréacteur aidera à améliorer la productivité et à fournir des produits de qualité supérieure à des prix plus bas (Sanjogta, 2022).

Un bioréacteur est un dispositif qui comprend diverses fonctionnalités telles qu'un système d'agitateur, un système de distribution d'oxygène, un système de contrôle du mousse, ainsi que divers autres systèmes tels qu'un système de contrôle de température et de pH, des ports d'échantillonnage, un système de nettoyage et de stérilisation, des lignes de chargements et de vidange du réacteur. Les bioréacteurs varient considérablement en fonction de l'application (figure 01) (Sanjogta, 2022).

### I.2.1 La cuve de fermentation

La cuve de fermentation est un grand cylindre fermé en haut et en bas, relié à divers tuyaux et vannes. Elle est divisée en deux types : la cuve en verre et en acier inoxydable. La cuve en verre est généralement utilisée dans les industries à petite échelle. Elle est non toxique et résistante à la corrosion et la pression. La cuve en acier inoxydable est utilisée dans les industries à grande échelle (Sanjogta, 2022).

Elle comprend des ports de sondes de pH, de thermocouple et d'oxygène dissous, des ports de dégazage, d'acide et de base, un port d'inoculation, un tuyau pour la diffusion de l'air de processus, un arbre d'agitateur et des ports de recharge (Kana et al, 2003).

### I.2.2 L'appareil de chauffage et de refroidissement (contrôle la température)

Est un équipement utilisé dans les bioréacteurs pour contrôler la température de la culture en fermentation. Il est positionné à l'extérieur de la cuve de fermentation est équipé d'une veste de refroidissement qui ferme hermétiquement la cuve et fournit de l'eau de refroidissement (Sanjogta, 2022). Il peut se présenter sous forme de veste de refroidissement, de serpentins internes, de coussins thermiques ou de radiateurs infrarouges (Jagani et al 2010).

### I.2.3 Le système d'aération et d'agitation

Le procédé d'aération est l'une des parties très importantes d'un fermenteur qui fournissent une alimentation en oxygène et en nutriments à des cultures microbiennes. Ce système nécessite un bon brassage des bulles d'air et des cellules microbiennes pour un meilleur rendement (Sanjogta, 2022). Il se compose d'un compresseur, d'une entrée d'air, d'un système de stérilisation, d'un spargeur d'air et d'un système de stérilisation de sortie d'air pour éviter la contamination (Jagrit et *al*, 2014).

Trois types de systèmes sont employés pour réaliser l'aération. D'abord, des dispositifs mécaniques (turbines, agitateurs et pales ou à l'aide des pompes), ensuite, des insufflations d'air et enfin, par dispositif mixte par les deux systèmes (Toumi, 2009).

Le procédé d'agitation est composé d'un ou plusieurs éléments mobiles adaptés à leur taille pour créer un mouvement rotatif dans le liquide du bioréacteur pour homogénéiser la culture et effectuer un transfert de chaleur et de matière à l'intérieur de la cuve (Clément, 2017). Ce système est constitué d'un moteur électrique externe, d'une hélice et d'un ensemble de déflecteurs qui assurent un mélange homogène et une amélioration du transfert de masse à travers les différentes couches du liquide ainsi que la couche limite de bulles. Ce système fournit les conditions de cisaillement requises pour briser les bulles présentes dans la culture (Jagrit et *al* 2014). Ce système est reposé à quelque partie du fermenteur sont :

-l'impulseur (Agitateur) : Il est composées de lames d'impeller attachées à un moteur sur le couvercle. Ces lames jouent un rôle important dans la réduction de la taille des bulles d'air. Les lames distribuent les bulles d'air uniformément dans les supports de fermentation afin d'assurer une agitation continue du milieu de culture et empêcher les cellules de se déposer au fond (Sanjogta, 2022 ; Vishwajeet et *al*, 2018).

- Le spargeur (Aérateur) : est un dispositif utilisé pour injecter de l'air stérile dans une cuve de fermentation aérobie. Il permet d'assurer une aération adéquate du réacteur en libérant de l'air sous pression à travers de petits trous d'environ 5 à 10 mm présents dans

les tuyaux de sparger. Il existe trois types de sparger utilisés : Sparger poreux, Sparger à jet, Sparger-agitateur combiné (Sanjogta, 2022 ; Vishwajeet et *al*, 2018).

-Les baffles : sont des bandes métalliques fixées radialement sur la paroi du bioréacteur. Ces bandes sont utilisées pour améliorer la dispersion du gaz et le mélange et pour réduire la formation de vortex et augmenter l'efficacité de l'aération. Ils sont généralement inclus dans les cuves agitées de toutes tailles (Jagani et *al* 2010).

#### **I.2.4 Le système de contrôle de la mousse**

Le niveau de mousse dans le récipient doit être minimisé pour éviter tous types de contamination, c'est un aspect important du fermenteur, car une production excessive de mousse peut obstruer les filtres de sortie d'air et entraîner une accumulation de pression à l'intérieur du bioréacteur (Sanjogta, 2022 ; Vishwajeet et *al*, 2018 ; Jagrit et *al*, 2014). Ce système est travaillé par deux unités :

-Le détecteur de mousse : Détecter la présence de mousses.

-Une unité de contrôle : Éviter la formation et la dégradation de la mousse dans le milieu.

Ce système est positionné sur le dessus du fermenteur, avec une entrée dans le fermenteur (Sanjogta, 2022 ; Vishwajeet et *al*, 2018 ; Jagrit et *al* 2014).

#### **I.2.5 Le système de contrôle du pH**

Ce système permet d'évaluer et surveiller le changement de pH du milieu (Vishwajeet et *al*, 2018). Il repose sur des électrodes positionnées dans le bouillon pendant les fermentations associé à un écran d'affichage pour montrer les mesures de pH (Kana et *al*, 2003).

Le système utilise des agents neutralisants pour contrôler le pH. Ces agents doivent être non corrosifs et non toxiques pour les cellules lorsqu'ils sont dilués dans le milieu (Jagrit et *al* 2014).

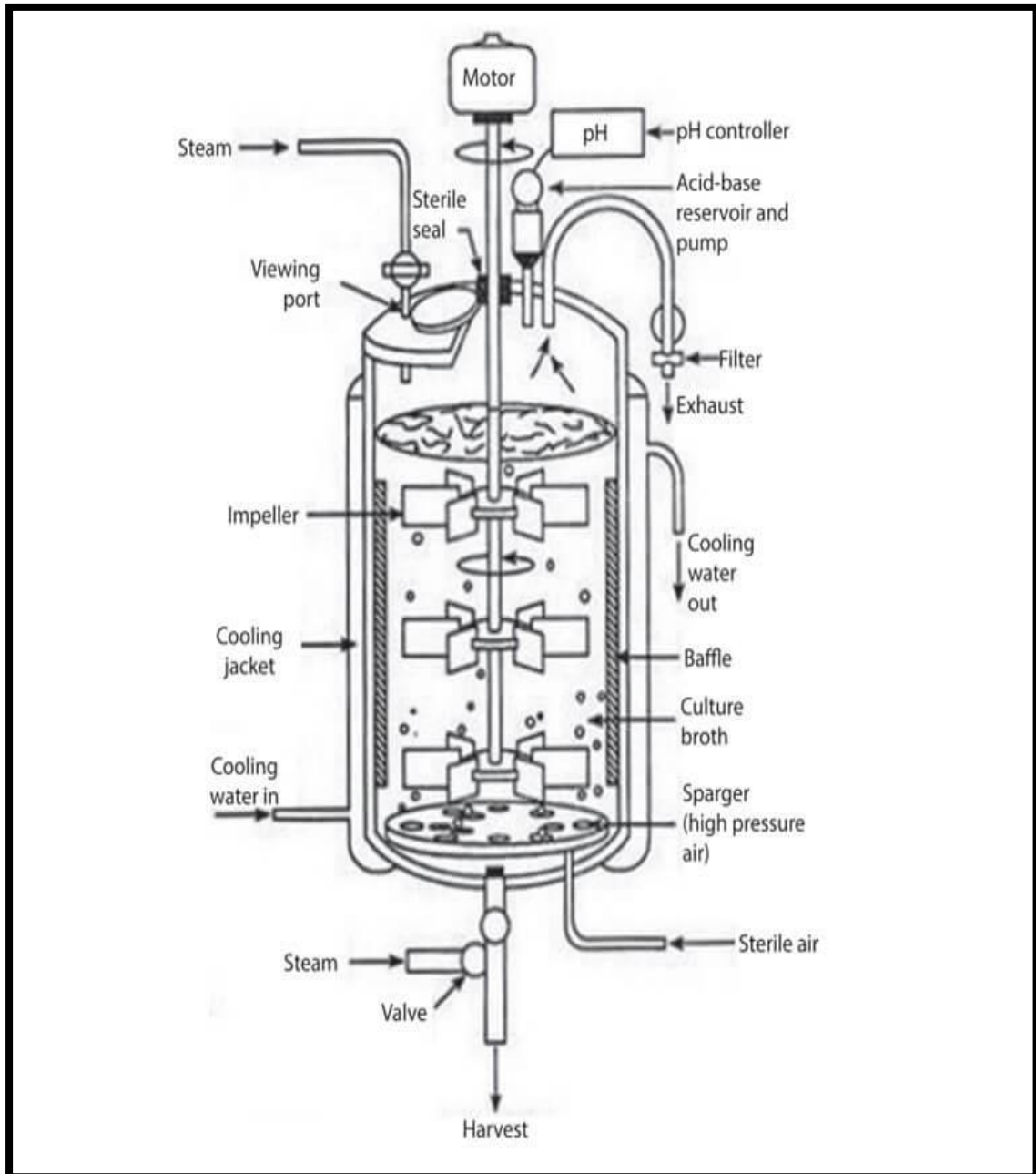


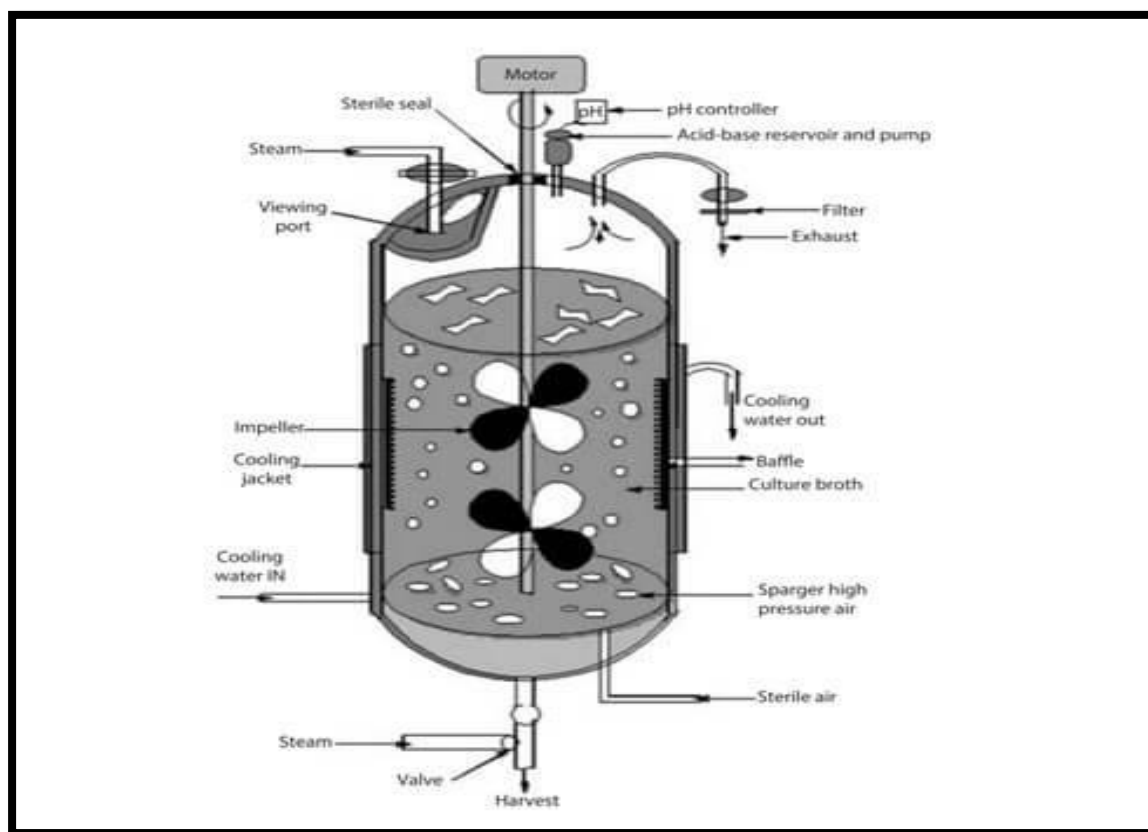
Figure 1 : les différents systèmes dans le bioréacteur (Sanjogta, 2022).

### **I.3 Les différents types de bioréacteurs**

Pour exécuter avec succès la fermentation, il est essentiel de choisir le bon type de bioréacteur pour atteindre une productivité élevée des produits. Le transfert de masse gaz-liquide, le mélange liquide, la dynamique des fluides et les forces de cisaillement sont les critères fondamentaux les plus couramment utilisés pour évaluer et comparer les différents types de bioréacteurs (Toumi, 2009).

#### **I.3.1 Le bioréacteur à agitation mécanique**

Le bioréacteur à agitation mécanique est l'un des types les plus couramment utilisés en raison de sa flexibilité (Chisti, 2010). Il se compose d'un ou plusieurs agitateurs (hélices) qui sont soutenus par un axe central de forme cylindrique et alimentés par un moteur. Cette combinaison du diffuseur et des agitateurs (hélices) permet une amélioration de la distribution du gaz dans tout le récipient. Ce bioréacteur peut fonctionner en continu dans le fermenteur. Ces paramètres de contrôle sont la température qui est très simple. La construction est peu coûteuse. Ce type de bioréacteur est facile à utiliser, car l'exploitation nécessite peu de formation et le nettoyage est simple (Sanjogta, 2022). Le bioréacteur à agitation mécanique assure un transfert de masse considérable et un mélange avantageux mais il y a un problème d'attrition se produit dans le récipient avec les cellules immobilisées (Figure 02) (Vishwajeet et al, 2018).



**Figure 2 :** bioréacteur a agitation mécanique (Sanjogta, 2022).

### I.3.2 Le bioréacteur avec agitation pneumatique

Le bioréacteur avec agitation pneumatique est généralement utilisé pour les dispositifs de contact gaz-liquide ou gaz-liquide-solide. Il est également connu sous le nom de réacteur tour (Sanjogta, 2022 ; Vishwajeet et al, 2018).

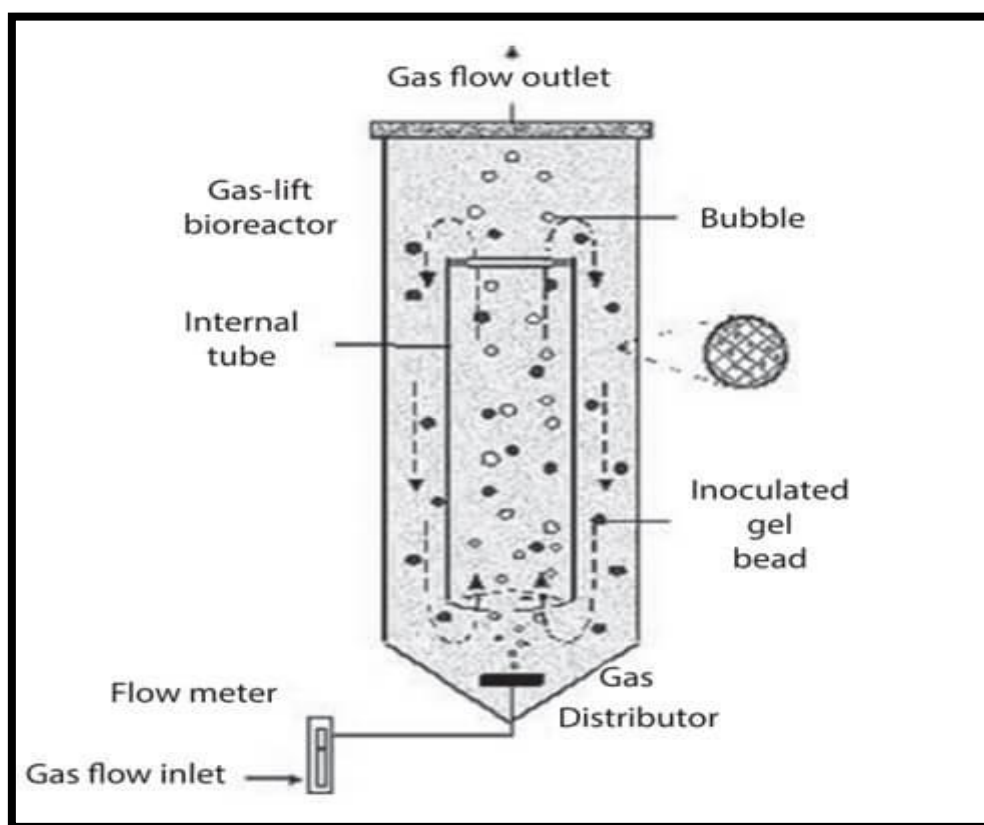
Ce réacteur peut être utilisé pour les cellules libres et immobilisées qui conviennent aux bactéries, levures, champignons, plantes et cellules animales. Ce type de fermenteur n'a pas de système d'agitation car il permet l'utilisation de l'expansion du gaz comprimé pour provoquer le mélange (Jagrit et al 2014).

Le système aérien divise le volume de fluide en deux zones pour améliorer la circulation, le transfert d'oxygène et égaliser les forces dans le réacteur (Sanjogta, 2022).



Le bioréacteur pneumatique possède quelques avantages qui sont : la simplicité de conception, sa stérilisation facile, son faible coût et sa faible consommation d'énergie (Sanjogta, 2022).

Cet équipement est souvent utilisé pour la culture de micro-organismes ou pour la production de substances d'intérêt industriel (Figure 03) (Vishwajeet et *al*, 2018).



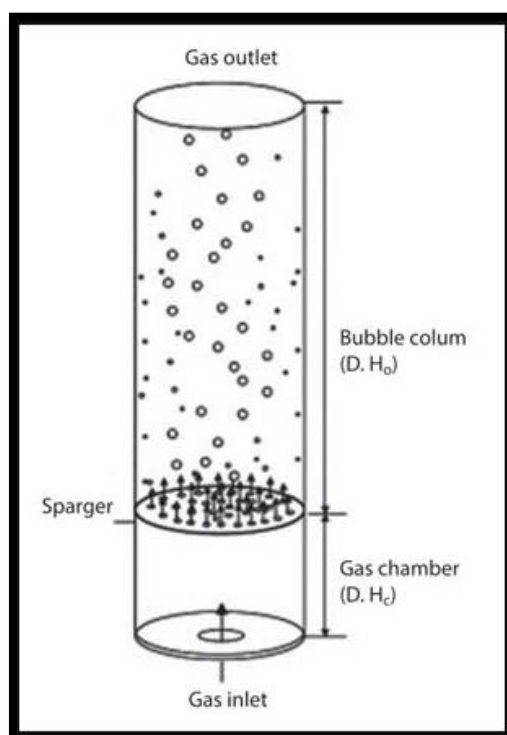
**Figure 3 :** bioréacteur à agitation pneumatique (Sanjogta, 2022).

### I.3.3 Bioréacteur à colonne à bulles

Le bioréacteur à colonne à bulles est constitué d'un récipient cylindrique équipé d'un diffuseur de gaz qui propulse des bulles de gaz dans une phase liquide ou une suspension liquide-solide. L'agitation est assurée par l'air ou le gaz qui est injecté à la base de la colonne à l'aide de tuyaux, de plaques perforées ou de diffuseurs métalliques à pores microscopiques (Sanjogta, 2022 ; Chisti, 2010). L'action de l'agitation et l'aération de bioréacteur à colonne à bulles est influencée par les changements des propriétés rhéologiques du fluide et le débit de gaz.

Par comparaison à d'autres réacteurs avec des agitateurs, l'utilisation de ce réacteur est simple, facile avec un faible coût d'exploitation.

Les réacteurs à colonne de bulles sont utilisés généralement dans des opérations biochimiques telles que la fermentation et le traitement biologique des eaux usées et dans la synthèse du méthanol et la fabrication d'autres carburants (Figure 04) (Sanjogta, 2022 ; Jagrit et al 2014).



**Figure 4 :** Bioréacteur à colonne à bulles (Sanjogta, 2022).

### I.3.4 Le bioréacteur à lit fluidisé

Les bioréacteurs à lit fluidisé sont constitués de lits de petites particules. Cela évite les problèmes tels que le colmatage, la chute de pression liquide élevée, la canalisation et le compactage de lit associés aux réacteurs à lit fixe (Sanjogta, 2022).

Il s'agit de dispositifs ressemblés à bioréacteur à colonnes à bulles avec une section transversale élargie près du haut. Le catalyseur est posé au fond du réacteur et les

réactifs sont pompés dans le bioréacteur à travers une pompe de distribution pour fluidiser le lit (Jagrit et *al*, 2014 ; Sanjogta, 2022).

Dans ces réacteurs, les cellules sont immobilisées dans de petites particules qui se déplacent avec le fluide. Ce qui améliore le transfert de masse, le transfert d'oxygène et la nutrition des cellules (Sanjogta, 2022).

Ces bioréacteurs peuvent être adaptés à la culture de micro-organismes à croissance lente, et au traitement des eaux usées (Figure 05) (Chisti, 2010).

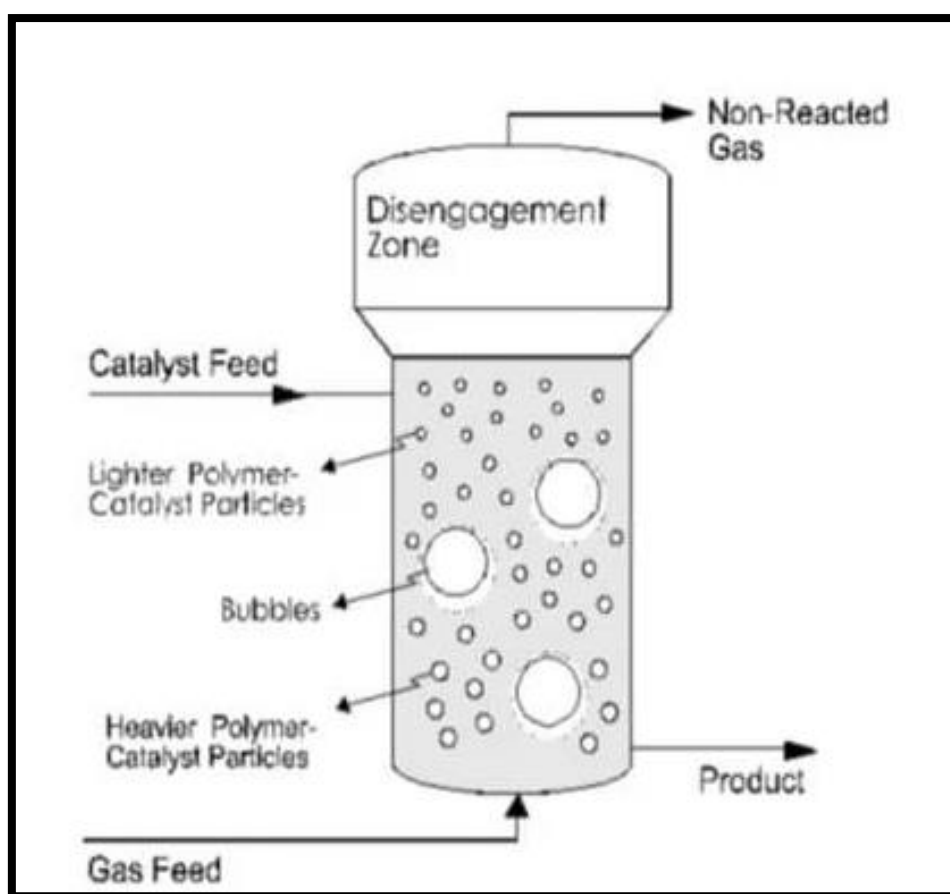
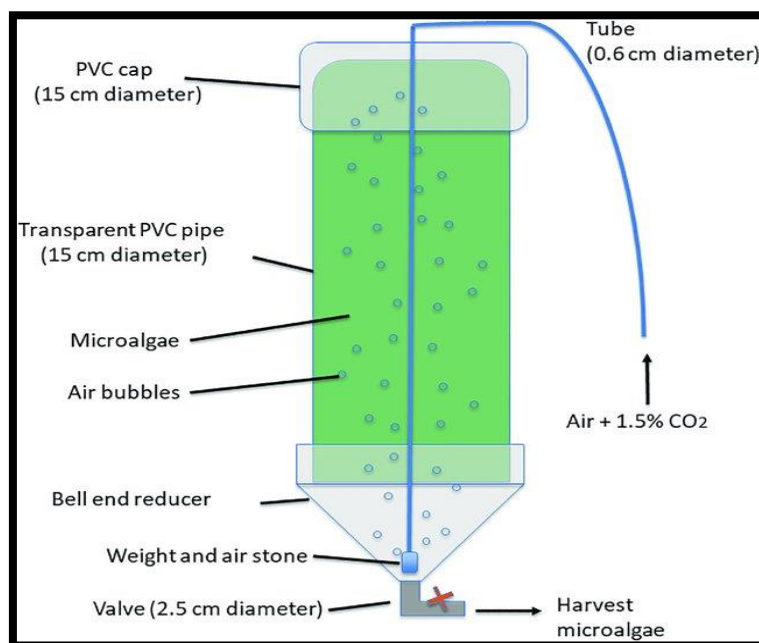


Figure 5 : bioréacteur à lit fluidisé (Sanjogta, 2022).

### I.3.5 Les photobioréacteurs

Un photobioréacteur est une unité spécialisée pour la fermentation qui est éclairée soit artificiellement, soit directement par la lumière du soleil. Ce bioréacteur est fabriqué généralement en verre ou en plastique transparent et les systèmes de réception de lumière sont intégrés aux tubes et panneaux plats. Les photobioréacteurs sont

généralement exploités en mode continu à une plage de température de 25 à 40 °C. L'utilisation de ce bioréacteur est pour les cultures photosynthétiques de micro-algues et de cyanobactéries pour produire des produits tels que l'astaxanthine et le bêta-carotène (Figure 06) (Sanjogta, 2022).



**Figure 6:** schéma d'un photobioréacteur (Jochum et al, 2022).

### I.3.6 Le bioréacteur à membrane

Ce système regroupe un réacteur biologique avec une séparation physique par des membranes poreuses (Grasmick et al, 2007). Ce qui permet l'élimination des matières organiques et des solides en suspension ainsi que l'élimination des niveaux élevés de nutriments. Les membranes poreuses sont trempées dans un réacteur biologique aéré. La taille des pores de la membrane varie de 0,035 microns à 0,4 microns. Ce système apporte un oxygène pur pour la suspension et un traitement biologique à taux encore plus élevé (Sanjogta, 2022).

Les bioréacteurs membranaires sont classés selon les types de membrane (organiques, inorganiques, polymériques, etc.) (Figure 07) (Toumi, 2009).

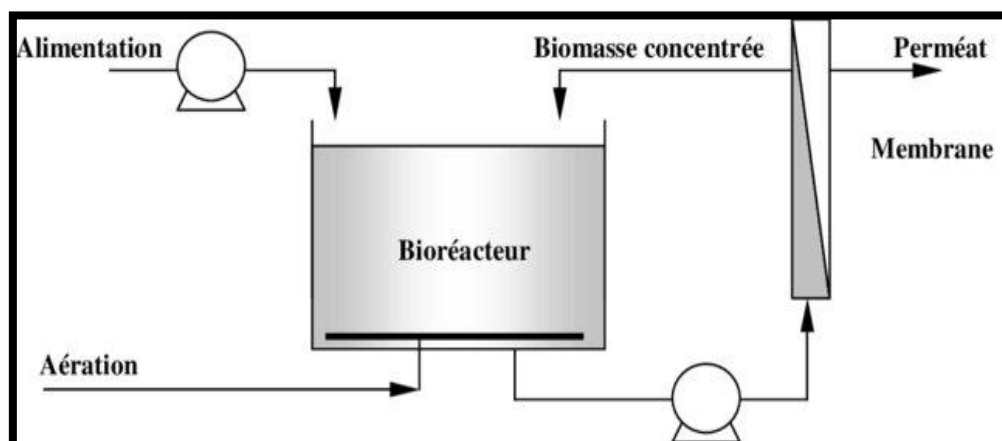


Figure 7 : schéma d'un bioréacteur à membrane (Nicolas, 1970)

## I.4 Les modes de fonctionnement

Dans la pratique, les modes opératoires se caractérisent par les échanges liquides, c'est-à-dire par la façon dont les réacteurs biologiques sont alimentés en substrat. Nous distinguons trois modes principaux (Grisales, 2007).

### I.4.1 Le mode discontinu (ou batch)

Dans ce mode de fonctionnement, la totalité des nutriments nécessaires à la croissance biologique est introduite dans le bioréacteur lors du démarrage de la réaction. Aucun apport ni prélèvement (excepté bien sûr pour quelques mesures hors lignes éventuellement) est par la suite réalisé et la réaction se poursuit à volume constant (Grisales, 2007).

Les seules actions possibles de l'opérateur ne concernent que les variables d'environnement par exemple : pH, température, vitesse d'agitation, aération, etc. Ce genre de mode ne nécessite pas beaucoup de moyens pour être mis en œuvre. Ce qui le rend attractif d'un point de vue industriel (Grisales, 2007).

Le deuxième avantage est de garantir l'axénicité des cultures, car les risques de contamination de la culture sont minimes. Le revers de la médaille est le peu des moyens qui permettent d'agir sur le fermenteur pour améliorer l'utilisation des micro-organismes (Grisales, 2007).

Il souffre cependant de deux inconvénients majeurs :

- sa durée requise, car à chaque nouveau remplissage du réacteur, avant que la réaction ne démarre, on observe une période de latence liée au temps d'adaptation des microorganismes, ce qui entraîne la génération de temps morts lors de l'opération (Grisales, 2007).

- son inadéquation lorsque, la réaction générale est sensible à une inhibition par le substrat (les quantités introduites sont en effet importantes, car devant couvrir l'ensemble des besoins des micro-organismes sur la totalité du fonctionnement) ou par un produit (toute production est en effet conservée dans le réacteur jusqu'à l'achèvement attendu de la réaction), ce qui se traduit par des durées de traitement allongées et des limitations de la charge initiale admissible (Grisales, 2007).

#### **I.4.2 Le mode semi-continu (ou Fed- batch)**

Tout en nécessitant un dispositif de stockage des affluents, ce mode de fonctionnement se distingue du précédent par un apport de nutriments différents tels que les besoins constatés des micro-organismes. La variation du volume du milieu réactionnel est donc une fonction directe de l'état d'avancement de la réaction (Grisales, 2007).

Ce mode de fonctionnement permet essentiellement de lever les problèmes d'inhibition associés au mode précédent et de fonctionner à des taux spécifiques de croissance proches de leur valeur maximale (Grisales, 2007).

A partir d'un volume initial préalablementensemencé, le réacteur est alimenté par un débit augmentant typiquement de façon exponentielle, nécessitant un contrôle en boucle fermée, cela réduit l'utilisation du Fed-batch en milieu industriel (Grisales, 2007).

Le mode semi-continu est principalement utilisé pour la production à partir de procédés anciens : production de vinaigre, de bière, de levure etc. Il est également préconisé lorsque la récupération des produits est réalisée en discontinu (accumulation intracellulaire par exemple) ou que l'on ne peut se permettre de relarguer des substances toxiques résiduelles (cas du fonctionnement en continu) (Grisales, 2007).

**I.4.3 Le mode continu (ou chemostat)**

C'est le mode le plus utilisé dans le domaine du traitement de l'eau (Queinnec, 2000). Il se caractérise par un volume de réaction constant, il est soumis à un soutirage de milieu réactionnel égal au flux d'alimentation en matière nutritive (en employant une régulation de niveau) (Demeyer et *al*, 1982).

Les procédés continus fonctionnent en régime permanent, en maintenant, pour des conditions d'alimentation fixées, le système dans un état stationnaire, tout en évitant tout phénomène inhibiteur grâce à l'effet de dilution dû à l'alimentation (Queinnec, 2000 ; Grisales, 2007). Ces modes de fonctionnement permettent également la production importante dans des réacteurs de taille réduite et ne nécessitent pas d'importants dispositifs de stockage en amont. (Queinnec, 2000).

**I.5 Applications du bioréacteur**

Selon son type, le bioréacteur est utilisé dans plusieurs domaines différents, notamment :

Tableau 01 : les applications du bioréacteur (Sanjogta, 2022).

type de bioréacteur	Application
Bioréacteur à agitation mécanique	Antibiotiques, acide citrique, exopolysaccharides, cellulose, enzymes chitinolytiques, laccase, xylanase, pectique et pectate lyase, culture de masse tissulaire, lipase, polygalacturonases, acide succinique.
Bioréacteur avec agitation pneumatique	Antibiotiques, Enzymes chitinolytiques, Exopolysaccharides, Acide gibbérélique, Laccase, Cellulase, Acide lactique, Polygalacturonases, Culture tissulaire de masse.
Bioréacteur à colonne à bulles	Culture d'algues, Enzymes chitinolytiques.
bioréacteur à lit fluidisé	Laccase.
Photobioréacteur	Traitement des eaux usées, gestion de la qualité de l'eau, assainissement des sols contaminés.
Bioréacteur à membrane	Alginate, Antibiotique, Hydrolyse de la cellulose, Production d'hydrogène, Traitement de l'eau, Traitement des composés organiques volatils.

Notre travail consiste à prédire la quantité de production de la biomasse en exploitant le *lactosérum* comme une source de substrat de culture des micro-algues (plus précisément la *Chlorella vulgaris*, une souche micro algale verte d'eau). La production a été réalisée par (Bouznada et al, 2022) à l'aide d'un photobioréacteur qui se base sur les variations des facteurs de biotransformation (c'est-à-dire la température, le pH, la lumière, la source de carbone, la salinité, les nutriments, etc.).

Afin de réaliser cette prédiction, nous exploitons les techniques de l'intelligence artificielle. En premier lieu, nous expliquons la notion de l'intelligence artificielle dans le chapitre suivant.



# **Chapitre II**

## **Concepts d'Intelligence Artificielle**

## II.1 L'informatique

### II.1.1 Définition

L'informatique désigne l'ensemble des sciences et des techniques qui permettent le traitement automatique de l'information comprenant à la fois la partie matérielle (électronique) et la partie immatérielle (logiciel) d'un ordinateur (Le Parc, 2017).

Il existe plusieurs branches de l'informatique. L'intelligence artificielle (IA) est une branche de l'informatique la plus adaptée pour résoudre des problèmes qui nécessitent le raisonnement.

## II.2 L'Intelligence Artificielle (IA)

### II.2.1 Définition

Il n'y a pas de définition universelle de l'IA et chaque professionnel la définit en fonction de l'école qu'il suit (Belbekri, 2022).

L'IA est un ensemble de technologies informatiques et de théories dans plusieurs domaines concerné par la pensée, la logique et l'intelligence. Elle vise à concevoir des programmes capables de résoudre de problèmes, de traiter le langage naturel et de réaliser d'autres tâches en rapport avec l'intelligence (Lucas, 1992).

L'IA est définie comme un domaine de l'informatique qui vise à concevoir des systèmes capables de donner des outputs similaires à ceux qu'il produits par un système cognitif humain. De plus, L'IA est une discipline dans le domaine de l'informatique dont l'objectif principal est bien représenté en transformant en programmes informatiques l'expertise humaine dans plusieurs domaines (Lucas, 1992).

L'IA consiste en un ensemble de connaissances et de technologies, lesquelles ont leur unité en fonction d'un objectif commun lié à l'activité de programmation, expliquer certains processus cognitifs liés à l'intelligence humaine, ou résoudre des problèmes complexes (par des moyens informatiques) dans plusieurs domaines de la connaissance (Lucas, 1992).

On peut mentionner trois types de l'IA (Belbekri, 2022):

- IA faible qui peut être inférieure ou équivalente à l'humain.

- IA moyenne qui peut être supérieure à celle de la plupart des hommes.
- IA forte qui peut être supérieure à tout humain.

## II.2.2 Historique

L'histoire de l'IA a commencé en 1943 avec la publication de l'article de Warren McCullough et Walter Pitts intitulé "Un Calcul Logique des Idées Immanentes dans l'Activité Nerveuse". Dans cet article, les scientifiques présentent le premier modèle mathématique pour la construction d'un réseau de neurones. Après cette date, l'IA est entrée dans une phase de développement rapide et varié :

Entre [1950 ; 1960], SNARC est apparu comme un premier ordinateur avec un réseau de neurones qui a été développé en 1950 par Dean Edmonds et Marvin Minsky, deux étudiants de Harvard. La même année, Alan Turing a publié le Test de Turing, qui est encore utilisé aujourd'hui pour évaluer l'IA<sup>1</sup>.

En 1952, Arthur Samuel a développé un logiciel qui pouvait apprendre à jouer aux échecs. C'est lors du discours de John McCarthy en 1956, dans la conférence « Dartmouth Summer Research Project on Artificial Intelligence », le terme "intelligence artificielle" a été utilisé pour la première fois, en relation avec Arthur Samuel. Les chercheurs ont présenté les objectifs et la vision de l'IA lors de cet événement. Beaucoup de gens pensent que cette conférence a marqué le véritable début de l'IA<sup>1</sup>.

En 1959 encore Arthur Samuel a inventé le terme "apprentissage automatique".

En revanche, John McCarthy et Marvin Minsky ont fondé le projet de IA du MIT<sup>1</sup>.

Entre [1960 ; 1980], John McCarthy a également créé le "AI Lab" de l'université de Stanford en 1963. Après cette période le doute a jeté un froid sur le domaine de l'IA au cours des années suivantes. En 1966, le rapport américain ALPAC souligne le manque de progrès dans la recherche en traduction automatique visant à traduire instantanément le langage russe dans un contexte de guerre froide. De nombreux projets financés par le gouvernement des États-Unis seront abandonnés.

De même, en 1973, le gouvernement britannique a publié son rapport "Lighthill", qui a mis en lumière les déceptions de la recherche en IA. Une fois de plus, les coupes

budgetaires ont entraîné le ralentissement des projets de recherche. Cette période d'incertitude a perduré jusqu'en 1980 et elle a été nommée « premier hiver de l'IA »<sup>1</sup>.

Entre [1980 ; 2000], avec la création de R1 (XCON) par Digital Equipment Corporations (DEC), le premier hiver de l'IA prendra fin. Ce système expert commercial est destiné à établir des commandes pour de nouveaux systèmes informatiques, ce qui entraîne un véritable boom des investissements qui dure plus d'une décennie.

Quelque pays comme les États-Unis et le Japon ont réalisé des investissements importants dans la recherche en IA. Plus d'un milliard de dollars sont investis chaque année par les entreprises dans les systèmes experts, et le secteur est en croissance.

Malheureusement, le marché des dispositifs "Lisp" s'effondre en 1987 lorsque des substituts moins coûteux apparaissent. C'est « le deuxième hiver de l'IA ». Les entreprises perdent de l'intérêt pour les fournisseurs de systèmes spécialisés. De nombreux milliards de dollars ont été gaspillés dans des projets de recherche infructueux par le gouvernement américain et japonais<sup>1</sup>.

L'année de 1997, a marqué l'histoire de l'IA. Gary Kasparov, le champion du monde d'échecs, a été battu par Deep Blue, l'IA d'IBM. C'est la première fois que la machine l'emporte sur l'humanité<sup>1</sup>.

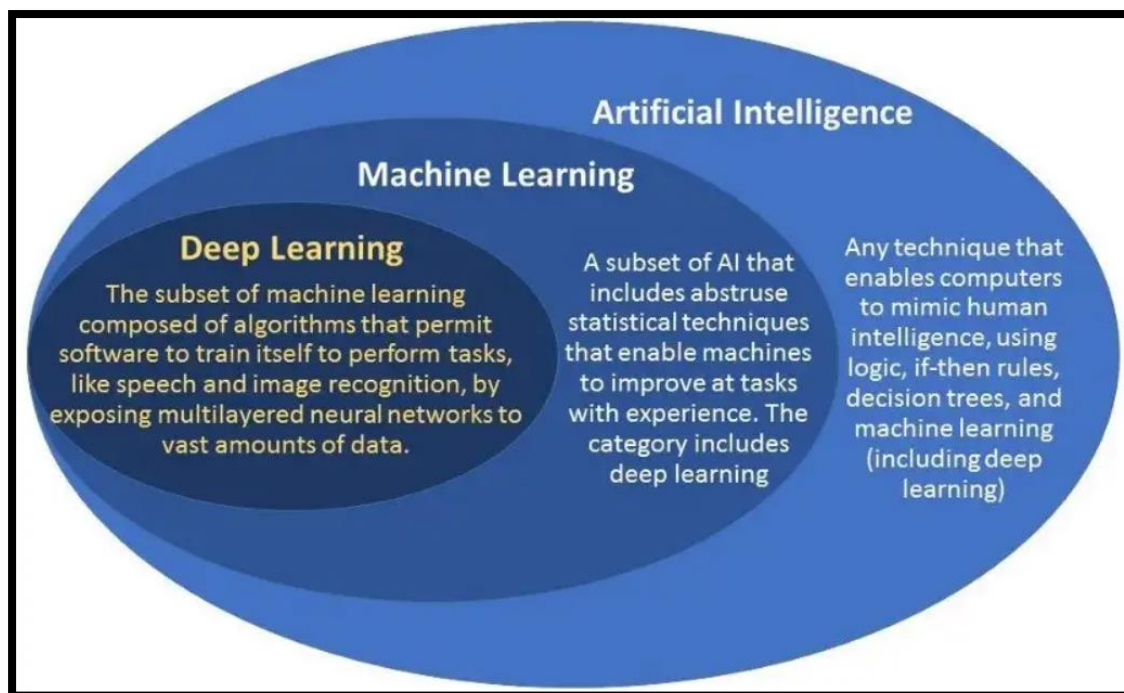
Entre [2000 ; 2022], l'intelligence artificielle a connu une évolution rapide et atteint son point culminant pendant cette période. Ces avancées technologiques permettent un renouveau de l'intelligence artificielle en 2008. Google réalise des progrès significatifs dans le domaine de la reconnaissance vocale et intègre cette fonctionnalité dans ses applications pour smartphones. Et en 2012, Andrew Ng a utilisé 10 millions de vidéos YouTube comme données d'entraînement pour alimenter un réseau de neurones. Grâce à l'apprentissage profond, ce réseau de neurones peut reconnaître une conversation sans avoir préalablement appris ce qu'est un chat. C'est le début d'une nouvelle ère pour l'apprentissage profond<sup>1</sup>.

Lee Sedol, le champion de Go en titre, a été vaincu par le système « AlphaGo » de Google DeepMind en 2016, marquant une nouvelle victoire de l'IA sur l'humanité.

L' IA artificielle a également dominé l'industrie du jeu vidéo, comme en témoignent AlphaStar de DeepMind sur Starcraft et OpenAI Five sur Dota 2.

Maintenant, L'utilisation de l'apprentissage profond et de l'apprentissage automatique est désormais répandue dans les entreprises de tous les secteurs, avec de nombreuses applications. L'IA ne cesse de se développer et surprend par ses performances. Le rêve d'une IA générale se rapproche de plus en plus de la réalité<sup>1</sup>.

Il existe plusieurs sous branches d'IA, par exemple machine learning et deep learning (figure 08). Cependant, la machine learning est plus adaptée pour résoudre des problèmes moins complexes.



**Figure 8 :** La relation entre l'IA, ML, DL (Moore et Raghavachari, 2019).

## II.3 Apprentissage automatique ou Machine Learning (ML)

### II.3.1 Définition

L'apprentissage automatique ou Machine Learning (ML) est une composante de l'IA bien qu'il cherche à résoudre des problèmes basés sur des exemples historiques ou passés (Berry et *al*, 2020).

C'est un domaine intéressant, issu de nombreuses disciplines telles que l'optimisation, algorithmes, les statistiques ou traitement du signal (Azencott, 2022).

Elle est basée sur le développement d'algorithmes qui améliore ses performances en apprenant des données (Zhang et Lu, 2021). Elle est déjà utilisée depuis des décennies dans la reconnaissance automatique de caractères. Elle est déjà utilisée maintenant pour se protéger contre la fraude bancaire, identifier les visages dans le viseur de notre appareil photo, ou traduire automatiquement des textes d'une langue vers une autre recommander des films, des livres, ou autres produits adaptés à nos goûts (Azencott, 2022).

### II.3.2 Types de données

Généralement, la disponibilité des données est la clé de la création d'un modèle d'apprentissage automatique. (Sarker, 2021).

Il existe plusieurs types de données. Les plus utilisés sont les données structurées et les données non structurées (figure 09) (Abdelaziz & Noui, 2022).

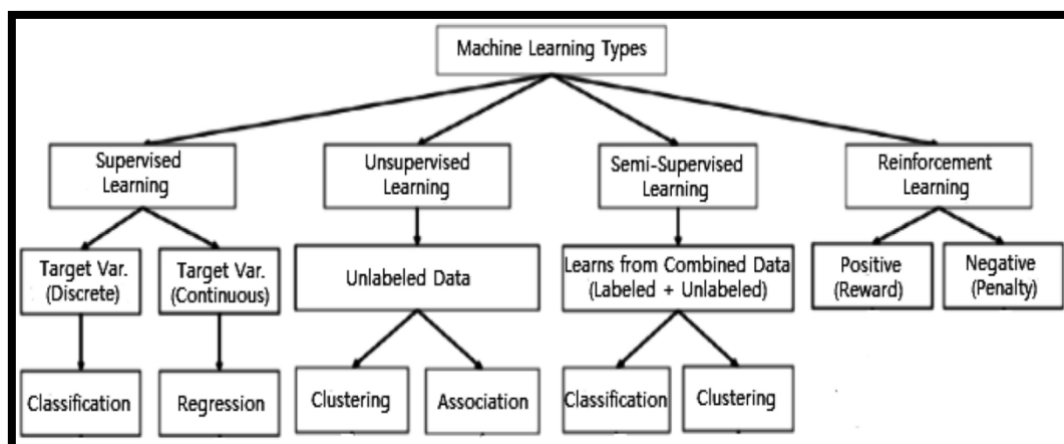


Figure 9 : Différents techniques pour le ML (Sarker, 2021).

**Données structurée** : Elles ont une structure bien définie, conforme à un modèle de données suivant un ordre standard. Ces données sont facilement accessibles et bien organisées. Par exemple : les bases de données relationnelles, les données structurées comme : noms, dates, adresses, numéros de carte de crédit, etc. (Sarker, 2021).

**Données non structurée** : D'un autre côté, il n'y a pas de format ou d'organisation prédéfini pour les données non structurées. Ce qui les rend beaucoup plus difficiles à capturer, traiter et analyser, contenant principalement du matériel multimédia et du texte. Par exemple : les documents de traitement de texte, les fichiers PDF, les fichiers audio, les vidéos, les images, les wikis, les images, les présentations, les pages Web, les données de capteurs électroniques et de nombreux autres types de documents commerciaux peuvent être considérés comme des données non structurées (Sarker, 2021).

### II.3.3 Types de techniques du l'apprentissage automatique

Les algorithmes d'apprentissage automatique sont principalement divisés en quatre catégories : apprentissage supervisé, apprentissage non supervisé, apprentissage semi-supervisé et apprentissage par renforcement (Sarker, 2021)

**Apprentissage supervise** : L'apprentissage supervisé est généralement la tâche de l'apprentissage automatique pour apprendre une fonction qui mappe une entrée à une sortie en fonction d'échantillons de paires entrée-sortie. Il utilise des données de formation étiquetées et une collection d'exemples de formation pour déduire une fonction. Cette technique est effectuée lorsque certains objectifs sont identifiés à accomplir à partir d'un certain ensemble d'entrées, c'est-à-dire une approche axée sur les tâches. Les supervisées les plus courantes sont la "classification" qui sépare les données et la "régression" qui ajuste les données. Par exemple : prédire l'étiquette de classe ou le sentiment d'un morceau de texte, comme une critique de produit, c'est-à-dire la classification de texte.

-La classification : La classification est définie comme le processus d'attribuer une ou plusieurs classes prédéfinies à chaque élément où la valeur de sortie est discrète. La classification binaire est mentionnée comme le type de classification le plus simple, où chaque élément est classé dans l'une des deux catégories. Peut être étendu à une classification multiple en définissant plus de catégories. (Jo, 2021). Ces méthodes sont

utilisées par exemple : pour identifier des textes ou des objets, faire des prédictions au niveau de l'image, comme la classification des maladies (Srindi et *al*, 2021).

-La régression : la régression est définie comme le processus d'estimation la valeur de sortie dépend de plusieurs facteurs. En régression, la valeur de sortie est continue (Jo, 2021).

**Apprentissage non supervisé :** L'apprentissage non supervisé analyse des ensembles de données non étiquetés sans intervention humaine, c'est-à-dire un processus piloté par les données. Il s'agit d'une large utilisation pour l'extraction de caractéristiques génératives, identifiant des tendances et des structures significatives, des regroupements dans les résultats et à des fins exploratoires. Les tâches d'apprentissage non supervisé les plus courantes sont l'estimation de la densité, l'apprentissage des caractéristiques, le regroupement, recherche de règles d'association et la détection d'anomalies, etc. (Sarker, 2021).

- Clusterisation : définie comme le processus de division d'un groupe d'éléments en sous-groupes dont chacun contient des éléments similaires les uns aux autres que celles des autres groupes (Jo, 2021). Il est utilisé dans le ML, recherche médicale, recherche d'informations, modèles de reconnaissance, bio-informatique et l'analyse d'image (Dike et *al*, 2018).

## II.4 L'Intelligence Artificielle dans le domaine de la biologie

IA est un domaine en pleine expansion, qui connaît de nombreuses applications dans divers secteurs, parmi les domaines émergents où l'IA montre un potentiel fascinant se trouve la biologie. L'application de l'IA dans la biologie offre de nouvelles perspectives qui utilisations de l'IA dans le domaine de la biologie en mettant l'accent sur ses applications prometteuses exemple :

- Découverte de médicaments : L'IA peut être utilisée pour analyser de vastes bases de données chimiques et identifier les composés prometteurs pour le développement de médicaments. Des algorithmes d'apprentissage automatique peuvent être utilisés pour prédire l'activité des molécules, accélérant ainsi le processus de découverte de médicaments (Ramsundar et *al*, 2019).



- Prédiction de la structure des protéines : La prédiction de la structure des protéines est une tâche complexe mais cruciale pour comprendre leur fonction. Des approches basées sur l'IA, telles que l'apprentissage profond, ont montré des résultats prometteurs dans la prédiction de la structure des protéines à partir de leur séquence d'acides aminés (Senior *et al*, 2020).
- Modélisation des réseaux biologiques : Les réseaux biologiques, tels que les réseaux de régulation génétique, sont extrêmement complexes. L'IA peut être utilisée pour modéliser ces réseaux et comprendre comment les différentes molécules interagissent entre elles (Maetschke *et al*, 2013).

L'application de l'IA dans la biologie offre un potentiel immense pour accélérer la recherche, la découverte de médicaments, la compréhension des systèmes biologiques complexes et la médecine personnalisée.

## II.5 Processus de l'apprentissage automatique

Il s'agit de l'ensemble des étapes à suivre pour créer un modèle pour résoudre un problème donné :

### II.5.1 Collection des données

Il s'agit de construire Datas qui permet de regrouper d'une manière structurée les données d'un problème à résoudre (Yang *et al*, 2019).

### II.5.2 Prétraitement des données

Il s'agit de nettoyer le Dataset par :

- ✓ suppression de données inutiles, répétées, incomplètes et manquantes
- ✓ enrichissement de Dataset par d'autres données ;
- ✓ décomposition des données (Yang *et al*, 2019) ;

### II.5.3 Choix du modèle

Selon le problème traité, on peut choisir :

- la régression : pour résoudre un problème de prédiction.

- le Clustering (K-means ou le voisin K) : pour résoudre les problèmes tels que la détection d'anomalies, la classification d'images, etc.
- naïve bayes : pour résoudre un problème de classification (Yang *et al*, 2019).

#### II.5.4 Entraînement du modèle

Les données de Dataset sont séparées en :

- 80 % pour entraîner l'algorithme choisi.
- 20 % pour tester et vérifier la performance du résultat (Yang *et al*, 2019).

#### II.5.5 Évaluation

Une fois que le modèle de ML a été construit, il va falloir l'évaluer puis l'améliorer si ses performances ne sont pas bonnes. Une baseline permettra de démarrer cette analyse.

Une baseline est un élément qui permet de comparer le modèle construit par rapport à autre chose<sup>1</sup> :

- avec les performances d'un modèle existe, ou
- avec les connaissances métier d'un expert

Une deuxième étape à étudier dans l'évaluation du modèle créé est de garantir l'absence de l'overfitting (sur apprentissage)<sup>1</sup>. Overfitting signifie que le modèle est très bon dans la prédiction de jeux de données d'entraînement, mais il est possible que les performances soient moins grandes dans le jeux de données de test. Prenant l'exemple<sup>1</sup> d'un élève qui révise ses leçons, et qui les apprend par cœur – lors de son entraînement –. Le jour de l'évaluation – du test – celui a 2/20. Cela signifie qu'il bien trop appris la leçon, mais qu'il n'a pas réussi à généraliser sur d'autres exercices.

Nous allons appliquer ce processus pour créer un modèle de prédiction de la quantité de production de la *Chlorella vulgaris* en exploitant le lactosérum comme une source de substrat de culture.

---

<sup>1</sup> <https://www.jedha.co/formation-ia/evaluer-ameliorer-modele-ml>

# **Chapitre III**

## **Matériel et méthode**

## III.1 Matériel

### III.1.1 données biologiques

Notre étude consiste à prédire la valeur de la quantité de *Chlorella Vulgaris* produite dans l'étude qui a été réalisée par (Bouznada et al, 2022). Cette dernière permet d'étudier l'utilisation du lactoserum récupéré des effluents de l'industrie laitière pour cultiver des microalgues *Chlorella Vulgaris* dans des photobioréacteurs et revalorisation des sous-produits. Dans la mémoire de (Bouznada et al, 2022), les microalgues *Chlorella Vulgaris* ont été cultivés (industrielle et naturelle) en mode batch dans des photobioréacteurs (PBR). Certains PBR sont équipés dans des conditions aseptiques avec un aérateur pour assurer l'évacuation du CO<sub>2</sub> à travers un tube d'échappement vers l'extérieur et un tube stérile pour les prélèvements. Certains PBR sont soumis à un éclairage artificiel.

Pour ce système de culture (Bouznada et al, 2022), les réacteurs ont été placés dans un incubateur à 30 °C. L'incubateur est équipé d'un éclairage à l'aide de la lumière LED à une intensité de 3,3  $\mu\text{watt}/\text{cm}^2$ . L'intérieur de la porte a été couvert d'aluminium pour que la lumière soit homogène dans tous les PBR. L'éclairage a été réglé à l'aide d'un programmateur pour assurer une illumination de 18h et 6h d'obscurité. Une agitation douce a été assurée par une plaque à 6 postes en utilisant des barreaux magnétiques identiques pour une bonne reproductibilité des résultats.

D'autres PBR sont directement exposés à la lumière naturelle suivie d'une mesure de l'absorbance du matin (8 h) et du soir (13 h) à une longueur d'onde de 673 nm. Les mesures du pH, d'Oxygène dissous, température ont été réalisés dans ces PBR. Une visualisation microscopie à l'aide d'un microscope Optika utilisant un grossissement x10 et un décalage du rapport S/X a été réalisée également.

Toutes les conditions d'incubation dans les bioréacteurs sont résumées dans les tableaux suivants.

Deux méthodes d'incubation ont été testées, un mode chimiotrophe et un mode mixotrophe, car la souche *Chlorella vulgaris* peut utiliser le carbone organique ou le carbone métallique en tant que source de carbone et source d'énergie est ce qui est

récupéré des réactions d'oxydation ou source de lumière (rayonnement solaire ou éclairage artificiel).

**Tableau 02** : Conditions d'incubation dans les bioréacteurs (Bouznada et *al*, 2022).

<b>Expérience (R)</b>	<b>Eclairage (lumière)</b>	<b>Rapport (S/X)</b>	<b>Mode de cultivation</b>
R2 ; R3	Obscurité	0.2 0.4 1 2 6 12	Chimiotrophe
R5	Eclairage naturel avec une aération externe VR 250 mL.	0.2 0.4 1 2 6 12	Mixotrophe
R6	Eclairage naturel avec une aération externe VR 500 mL	0.2 0.4 1 2 6 12	Mixotrophe
R7	Eclairage artificiel chronométré	0.2 0.4 1 2 6 12	Mixotrophe

**Tableau 03 :** Quantité de la biomasse produite dans chaque expérience (Bouznada et *al*, 2022).

Expérience	La rapport S/X	Quantité de la biomasse produite (g)
R2, R3	0.2	0,7714
	0.4	0,5677
	1	0,5629
	2	1,3381
	6	0,5677
	12	26,9156
R5	0.2	0,5476
	0.4	1,9182
	1	0, 3202
	2	1,9234
	6	5,7774
	12	0,1296
R6	0.2	0,2399
	0.4	1.1637
	1	2.0526
	2	4.3018
	6	4.0262
	12	3.9608
R7	0.2	0,2080
	0.4	0.1780
	2	0.0840
	6	0.0360

### III.1.2. Construction du Dataset

Nous avons créé une base des données (Dataset) sur Excel. Cette base est constituée de quatre colonnes dont trois colonnes représentent les attributs du Dataset qui sont les rapports : S/X (substrat sur la biomasse), la lumière et le mode de nutrition. Tandis que la quatrième colonne représente la sortie qui est la quantité de la biomasse (voir la figure 10).

#### III.1.2.1. Les valeurs du rapport (s/x)

Les valeurs originales du rapport (S/X) sont : "0,2" ; "0,4" ; "1" ; "2" ; "6" ; "12.

#### III.1.2.2. Eclairage

Dans l'étude originale la lumière n'est pas mesurée par des valeurs quantitatives. Puisque ML ne peut pas traiter des valeurs quantitatives, nous proposons le codage suivant des valeurs de la lumière :

- Les valeurs avant l'enrichissement (valeurs initiales mesurées durant les expériences)
  - "-1" : Obscurité.
  - "0" : Eclairage naturel avec une aération externe VR 250 mL.
  - "1" : Eclairage naturel avec une aération externe VR 500 mL.
  - "2" : Eclairage artificiel chronométré.
  
- Nouvelles valeurs après l'enrichissement du Dataset

Ce sont des valeurs qui ne sont pas mesurées dans les expériences. Elles sont ajoutées juste après les premières itérations de l'entraînement de modèle ML. Ces valeurs sont des valeurs intermédiaires entre les valeurs originales comme : "0.2" ; "0.5" ; "0.8" ; "-0.2" ; "-0.9" ; "1.2" ...

### III.1.2.3. Le mode nutritionnel

Dans l'étude originale, le mode nutritionnel n'est pas mesuré par des valeurs quantitatives. Puisque ML ne peut pas traiter des valeurs quantitatives, nous proposons le codage suivant des valeurs de ce mode :

- Les valeurs avant l'enrichissement
  - "0" : Chimiotrophe.
  - "1" : Mixotrophe.
  
- Nouvelles valeurs après l'enrichissement du Dataset :
  - "0.1" ; "0.4" ; "0.5" ; "0.8"...

### III.1.2.4. Quantité de la biomasse produite (g)

Dans l'étude originale, la quantité a été mesurée par des valeurs quantitatives. C'est pour ça nous avons gardé les valeurs de l'expérience.

### III.1.3. Matériel informatique

Le matériel informatique que nous avons utilisé pour réaliser notre travail est représenté dans le tableau suivant :

**Tableau 04 : Caractéristiques de la machine utilisée pour le ML.**

Ordinateur	Caractéristiques
Processeur	Intel(R) Core (TM) i3-3217U CPU @ 1.80GHz 1.80 GHz
Mémoire installée RAM	4,00 Go
Stockage	Disque dur 77.7 GO
Système d'exploitation	Windows 10 Professionnel
Type de système	Système d'exploitation 64 bits, processeur x64



### **III.1.4. Outils logiciels et bibliothèques**

Nous décrivons brièvement les outils et les bibliothèques utilisés pour réaliser le travail dans cette partie.

#### **III.1.4.1. les Outils**

Le travail a été réalisé en utilisant le langage de programmation Python, via Jupyter notebook de l'anaconda.

Tableau 05 : Principaux outils utilisés

Outil	Description
<b>Python 3.10</b>	Python est le langage de programmation le plus utilisé dans le domaine du Machine Learning, du Data science et de Big Data. Python est un langage de programmation de haut niveau, qui permet aux programmeurs de se concentrer sur ce qu'ils font plutôt que sur la façon dont ils le font. Ainsi, il permet d'écrire des programmes prend moins de temps qu'un autre langage. C'est un langage idéal pour les débutants (Bastien, 2023).
<b>Anaconda 3<sup>2</sup></b>	Anaconda est un outil dont la distribution est libre et open source. Il est destiné à la programmation dans un environnement Python et R. Anaconda est largement utilisé en IA ou ML, en sciences de données. Cette distribution scientifique de Python contient de nombreux packages nécessaires à l'analyse de données. Anaconda est un gestionnaire d'environnement open source.
<b>Jupyter notebook 6.5.2<sup>3</sup></b>	Jupyter Notebook est l'application Web originale pour créer et partager des documents informatiques. JupyterLab est le dernier environnement de développement interactif basé sur le Web pour les blocs-notes, le code et les données. Son interface flexible permet aux utilisateurs de configurer et d'organiser des flux de travail en science des données, en informatique scientifique, en journalisme informatique et en ML.
<b>Excel 2013<sup>4</sup></b>	Un programme développé par Microsoft qui utilise des formules et des fonctions pour organiser des données et des nombres à l'aide de feuilles de calcul.

---

<sup>2</sup> <https://www.jedha.co/>

<sup>3</sup> <https://jupyter.org/>

<sup>4</sup> <https://corporatefinanceinstitute.com/>

### III.1.4.2. Les bibliothèques

Les fonctions python utilisées proviennent de ces bibliothèques principales :

**Tableau 06 :** Différents bibliothèques python utilisées

Bibliothèque	Description
<b>Pandas</b> <sup>5</sup>	Pandas est une bibliothèque logicielle open source du langage de programmation Python. Il est spécialement conçu pour le traitement et l'analyse de données en Python. Il est puissant, flexible et facile à utiliser. Grâce à Pandas, le langage Python permet enfin de charger, d'aligner, de manipuler ou encore de fusionner des données.
<b>Numpy</b> <sup>6</sup>	Numpy est une bibliothèque Python utilisée pour travailler avec des tableaux. Numpy est un projet open source et vous pouvez l'utiliser librement. Il s'agit également des fonctions pour travailler dans le domaine de l'algèbre linéaire, de la transformée de Fourier et des matrices
<b>Matplotlib</b> <sup>7</sup>	Matplotlib est une bibliothèque de dessin open source qui prend en charge différents types de dessin. Il est souvent utilisé dans les serveurs d'applications Web, les shells et les scripts Python. Ils vous permettent également de créer des tableaux, des graphiques, des diagrammes à barres et d'autres types de graphiques avec seulement quelques lignes de code.
<b>Sklearn</b> <sup>8</sup>	C'est une bibliothèque Python qui donne accès à des versions efficaces d'un grand nombre d'algorithmes courants. Elle offre également une API uniformisée et propre. Une fois que vous avez compris l'utilisation et la syntaxe de base de Scikit-Learn pour un type de modèle, le passage à un nouveau modèle ou algorithme est très simple. La bibliothèque permet non seulement de faire de la modélisation, mais elle peut également fournir des étapes de pré-traitement, que nous verrons dans la suite de l'article.

<sup>5</sup> <https://datascientest.com/>

<sup>6</sup> <https://www.w3schools.com/>

<sup>7</sup> <https://www.simplilearn.com/>

<sup>8</sup> <https://datascientest.com/>

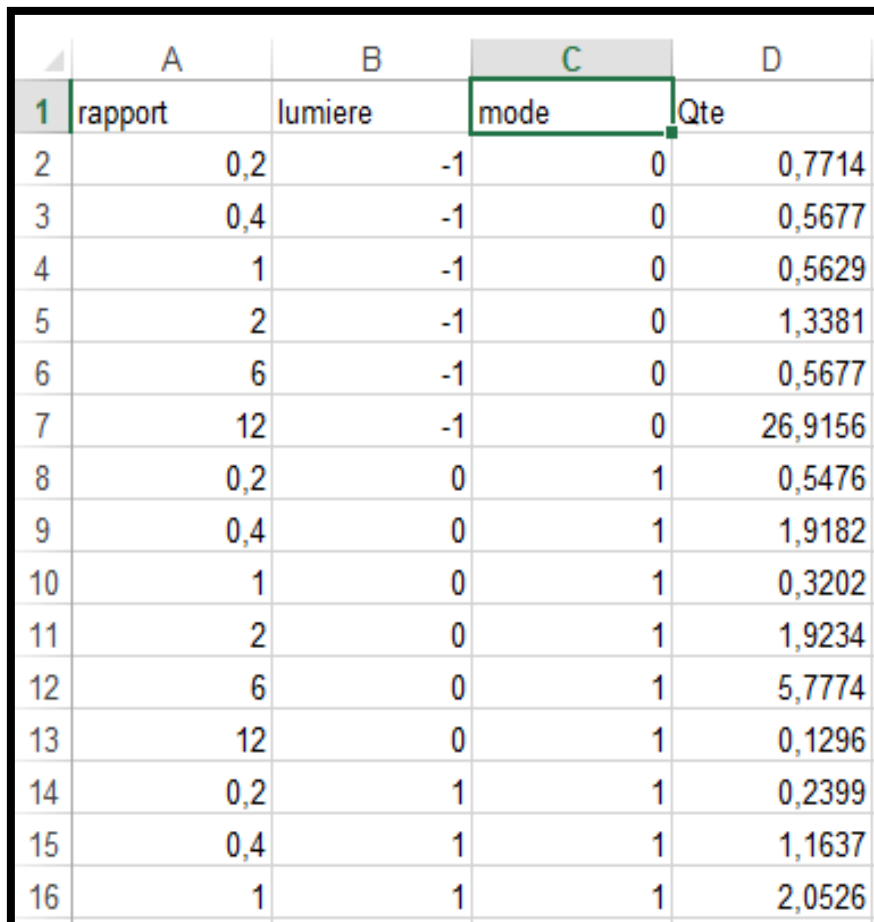
## III.2 méthodes

La section d'exploitation du Dataset et la section de l'apprentissage constituent les deux principales sections de la procédure de notre travail complet.

### III.2.1 Exploitation du Dataset

#### - Création du Dataset sur Excel :

Un extrait du Dataset sous format Excel est représenté dans (Figure 10).



	A	B	C	D
1	rapport	lumiere	mode	Qte
2	0,2	-1	0	0,7714
3	0,4	-1	0	0,5677
4	1	-1	0	0,5629
5	2	-1	0	1,3381
6	6	-1	0	0,5677
7	12	-1	0	26,9156
8	0,2	0	1	0,5476
9	0,4	0	1	1,9182
10	1	0	1	0,3202
11	2	0	1	1,9234
12	6	0	1	5,7774
13	12	0	1	0,1296
14	0,2	1	1	0,2399
15	0,4	1	1	1,1637
16	1	1	1	2,0526

Figure 10 : Extrait du Dataset sur Excel

#### - Enrichissement :

Dans le but d'améliorer les résultats nous avons ajouté aux valeurs originales de l'expérience des valeurs à l'aide de la fonction Rolling (Figure 11) appliqué sur python. Puisque le volume de Dataset est insuffisant, on a augmenté le nombre des données de 22 donnés (lignes dans le tableau Excel) à 160 donnés (lignes dans le tableau Excel).

Le rôle principal de la fonction Rolling est de calculer les médianes des valeurs précédentes et suivantes pour chaque ligne du Dataset et créer un nouveau Dataset avec ces nouvelles lignes (Figure 12).

```
df= pd.read_csv("/content/df1df1.csv")
df = pd.DataFrame(df)
num_rows = df.shape[0]
df_before = df.shift(1).rolling(window=2).median().fillna(method="bfill").iloc[1:]
df_after = df.shift(-1).rolling(window=2).median().fillna(method="ffill").iloc[:-1]
df1= pd.concat([df,df_before, df_after], ignore_index=True)
print(df1)
```

Figure 11 : Code python pour enrichie le Dataset

	A	B	C	D
17	2	1	1	4,3018
18	6	1	1	4,0262
19	12	1	1	3,9608
20	0,2	2	1	2,5757
21	0,4	2	1	2,3458
22	2	2	1	2,8105
23	6	2	1	5,9671
24	0,3	-1	0	0,6695
25	0,3	-1	0	0,6695
26	0,7	-1	0	0,5653
27	1,5	-1	0	0,9505
28	4	-1	0	0,9529
29	9	-1	0	13,7416
30	6,1	-0,5	0,5	13,7316
31	0,3	0	1	1,2329
32	0,7	0	1	1,1192
33	1,5	0	1	1,1218
34	4	0	1	3,8504
35	9	0	1	2,9535
36	6,1	0,5	1	0,1847
37	0,3	1	1	0,7018
38	0,7	1	1	1,6081
39	1,5	1	1	3,1772
40	4	1	1	4,1640
41	9	1	1	3,9935
42	6,1	1,5	1	3,2683

Figure 12 : Extrait du Dataset enrichie sur Excel.

- **Chargement sur mémoire et lecture du Dataset :**

Nous avons exploité la fonction `read_excel` pour charger sur mémoire et lire le Dataset (Figure 13).

```
In [46]: # Importing the Libraries
import numpy as np # for array operations
import pandas as pd # for working with DataFrames
import requests, io # for HTTP requests and I/O commands
import matplotlib.pyplot as plt # for data visualization
%matplotlib inline

# scikit-Learn modules
from sklearn.model_selection import train_test_split # for splitting the data
from sklearn.metrics import mean_squared_error # for calculating the cost function
from sklearn.ensemble import RandomForestRegressor # for building the model

In [100]: dataset = pd.read_excel("C:/Users/pc/Documents/az2.xlsx")
dataset.head()

Out[100]:
   rapport  lumiere  mode  Qte
0         0.2     -1.0   0.0  0.7714
1         0.4     -1.0   0.0  0.5677
2         1.0     -1.0   0.0  0.5629
3         2.0     -1.0   0.0  1.3381
4         6.0     -1.0   0.0  0.5677

In [101]: x = dataset.drop('Qte', axis = 1) # Features
y = dataset['Qte'] # Target
```

**Figure 13 :** Code python pour le chargement sur mémoire et lecture du Dataset.

### III.2.2 Apprentissage

Les données du Dataset ont été divisées en deux ensembles. Le premier représenté 80% pour entraîner le modèle et le 20% resté pour tester le modèle (figure 14).

```
In [102]: # Splitting the dataset into training and testing set (80/20)
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 28)
```

**Figure 14 :** Code python effectué la diffusion du Dataset (ensemble pour l'entraînement et l'ensemble pour le test).

#### III.2.2.1 Création du modèle

Le modèle créé est le Random Forest Regressor, car le problème traité est un problème de régression qui permet de prédire la quantité de la biomasse produite (la sortie), qui est une variable numérique, à partir des autres variables numériques (Eclairage, Mode de cultivation, le rapport S/X) représentant les variantes des variables d'entrée.

Le Random Forest Regressor : Un type de méthodes de ML, connu sous le nom de "Random Forest Regression" ou forêt aléatoire, combine plusieurs arbres de décision aléatoires. Chacun de ces arbres étant entraîné sur un sous-ensemble de données. L'algorithme est plus stable et moins erratique lorsque de nombreux arbres sont utilisés. L'approche de régression de la forêt aléatoire, qui fonctionne bien avec des données volumineuses et la plupart des types de données, est un modèle fréquemment utilisé (Figure 15). Nous appelons le modèle ML que nous avons créé par PPCV (Prédiction de la Production *Chlorella Vulgaris*).

```
In [103]: # Initializing the Random Forest Regression model with 10 decision trees
model = RandomForestRegressor(n_estimators = 100, random_state = 0)
|
# Fitting the Random Forest Regression model to the data
model.fit(x_train, y_train)

Out[103]:
RandomForestRegressor
RandomForestRegressor(random_state=0)
```

**Figure 15 :** Code python effectué la construction le model PPCV.

# **Chapitre**

## **IV :**

### **Résultats et discussions**



## IV.1 Résultats et discussions

Dans ce qui suit, nous évaluons le modèle **PPCV** que nous avons créé dans le chapitre 3. La fonction prédéfinie dans python *sklearn* est également utilisée pour effectuer le test du modèle.

### IV.1.1 Test d'Overfitting (sur\_apprentissage)

Nous avons utilisé Dataset issu de l'expérience réelle (Figure 16) :

1	rapport	lumiere	mode	Qte
2	0.2		-1,0	0,0 0.7714
3	0.4		-1,0	0,0 0.5677
4		1,0	-1,0	0,0 0.5629
5		2,0	-1,0	0,0 1.3381
6		6,0	-1,0	0,0 0.5677
7		12,0	-1,0	0,0 26.9156
8	0.2		0,0	1,0 0.5476
9	0.4		0,0	1,0 1.9182
10		1,0	0,0	1,0 0.3202
11		2,0	0,0	1,0 1.9234
12		6,0	0,0	1,0 5.7774
13		12,0	0,0	1,0 0.1296
14	0.2		1,0	1,0 0.2399
15	0.4		1,0	1,0 1.1637
16		1,0	1,0	1,0 2.0526
17		2,0	1,0	1,0 4.3018
18		6,0	1,0	1,0 4.0262
19		12,0	1,0	1,0 3.9608
20	0.2		2,0	1,0 2.5757
21	0.4		2,0	1,0 2.3458
22		2,0	2,0	1,0 2.8105
23		6,0	2,0	1,0 5.9671

**Figure 16** : Extrait du Dataset issu de l'expérience réelle sur Excel.

Nous avons trouvé les résultats suivants après la première itération d'entraînement du modèle **PPCV** :

- Le score du test : -11.17 (Figure 17).

```
In [14]: from sklearn.metrics import r2_score
         r2_score(y_test, y_pred)

Out[14]: -11.179739079433906
```

**Figure 17 :** Code python pour le score d'entraînement du programme.

- Le score de l'entraînement : 0.8020 (Figure 18).

```
In [15]: y_pred_train = model.predict(x_train)
         r2_score(y_train, y_pred_train)

Out[15]: 0.8020807330057215
```

**Figure 18 :** Code python pour le score de test du programme.

## Discussion

Selon ses résultats, nous avons trouvé que le score du test est très inférieur au score d'entraînement. Cela, signifie que ce modèle souffre d'overfitting (voir chapitre 2). Ce dernier est produit à cause de l'insuffisance de la taille du Dataset qui est très petit. Nous estimons que PPCV nécessite une amélioration.

### IV.1.2 Amélioration du modèle PPCV

Pour améliorer le modèle, il faut ajuster les données au niveau du Dataset. De ce fait, nous avons exploité la fonction Rolling (Figure 11). Cette opération est appelée Feature Engineering<sup>9</sup>. La fonction Rolling a apporté les modifications suivantes au niveau du Dataset.

<sup>9</sup> <https://www.jedha.co/formation-ia/evaluer-ameliorer-modele-ml>

Tableau 07 : Les valeurs ajusté du Dataset.

Type des valeurs ajusté	Les valeurs ajustées
<b>Rapport S/X</b>	0.1 ; 0.3 ; 0.5 ; 0.7 ; 0.9 ; 1.1 ; 1.2 ; 1.3 ; 1.5 ; 1.6 ; 1.8 ; 2.1 ; 2.3 ; 2.4 ; 2.5 ; 2.7 ; 3.4 ; 3.6 ; 3.8 ; 4 ; 4,7 ; 4.8 ; 4.9 ; 5 ; 5.1 ; 5,2 ; 6.1 ; 6.5 ; 7.5 ; 9
<b>Eclairage</b>	-0.5 ; 0.5 ; 1.5 ; -0.8 ; -0.2 ; 0.2 ; 0.8 ; 1.2 ; 1.8 ; 1.2 ; 1.4 ; 1.1 ; -0.7 ; -0.3 ;

	-0.6 ; -0.1.
<b>Mode</b>	0.5 ; 0.2 ; 0.8 ; 0.1 ; 0.9 ; 0.3

### Discussion

Nous avons ajouté des nouvelles valeurs pour l'amélioration de model par exemple, la nouvelle valeur ajoutée dans l'attribut mode "0.1 " signifie que, le mode est mixte entre le mode Chimiotrophe et le mode Mixotrophe, mais il est proche au Chimiotrophe. Tandis que, la nouvelle valeur ajoutée dans l'attribut mode "0.9 " signifie aussi que, le mode est mixte entre le mode Chimiotrophe et le mode Mixotrophe, mais il est proche au Mixotrophe.

### IV.1.3 Test de l'Overfeeting avec Dataset ajusté

Après plusieurs itérations d'entraînement, nous avons trouvé les résultats suivants :

- Le pourcentage du test : 92.18% (Figure 19).

```
In [36]: from sklearn.metrics import r2_score
         r2_score(y_test, y_pred)

Out[36]: 0.9218468773252795
```

**Figure 19** : Code python pour le score d'entraînement du programme de Dataset ajusté.

- Le pourcentage d'entraînement : 89.62% (Figure 20).

```
In [37]: y_pred_train = model.predict(x_train)
         r2_score(y_train, y_pred_train)

Out[37]: 0.8962420944314644
```

**Figure 20** : Code python pour le score de test du programme de Dataset ajusté.

**Discussion :**

Nous avons trouvé que le pourcentage du test est plus proche au pourcentage d'entraînement. Nous constatons que le modèle créé **PPCV** a débarrassé le problème d'overfitting.

De plus le pourcentage d'entraînement montre que **PPCV** a réussi à apprendre 89.62% des données du Dataset. De plus, le pourcentage du test montre que **PPCV** a réussi à donner des prédictions correctes avec un taux de 92.18%.

**IV.1.4 Évaluation de la prédiction**

Pour évaluer la prédiction, nous avons comparé les résultats de la prédiction obtenues après l'exécution du modèle crée PPVC avec les résultats du Dataset.

La comparaison entre les valeurs réelles et les valeurs de test sert à trouver :

- Le nombre des prédictions qui sont proches aux valeurs réelles.
- Le nombre des prédictions qui sont exactes et égales aux valeurs réelles.
- Le nombre des prédictions qui sont éloignées aux valeurs réelles.

Afin d'estimer si la valeur d'une prédiction est proche ou éloignée, nous proposons de calculer la distance entre les deux valeurs. Tous simplement, cette distance désigne la différence entre les deux valeurs. Nous proposons le seuil 1.31 (g).

- Si la différence est égale ou inférieure à 1.31 alors la prédiction est proche.
- Si la différence est strictement supérieure à 1.31 alors la prédiction est éloignée.
- Si les deux valeurs sont égales, la différence est égale à 0 alors la prédiction est exacte.

La fonction `pd.DataFrame` permet d'effectuer cette comparaison (figure 21).

-

```
In [54]: df = pd.DataFrame({'real_value': y_test, 'Predicted_value': y_pred})  
  
# Display the DataFrame  
print(df)
```

	real_value	Predicted_value
127	3.1865	2.907162
24	0.9505	0.831456
35	1.6081	1.438757
61	2.9232	3.040150
99	7.3472	8.009683
106	1.5691	1.897424
149	1.9321	2.574556
85	1.5945	1.352529
49	3.4254	5.295138
70	7.4147	7.387174
146	3.5612	3.633233
139	2.1444	2.079667
157	5.2856	5.494613
65	1.9678	1.961878
136	5.8788	5.650055
137	3.0663	2.837280
98	0.9517	2.269145
14	2.0526	1.568317

**Figure 21 :** Code python pour la comparaison des valeurs réelles avec les valeurs prédites.

Le Dataset contient 160 données (lignes). 20% de ces données sont exploitées pour le test soit 32/160.

Nous regroupons les résultats de comparaisons dans le tableau suivant :

**Tableau 08** : comparaison de quantité réel avec quantité prédire du programme.

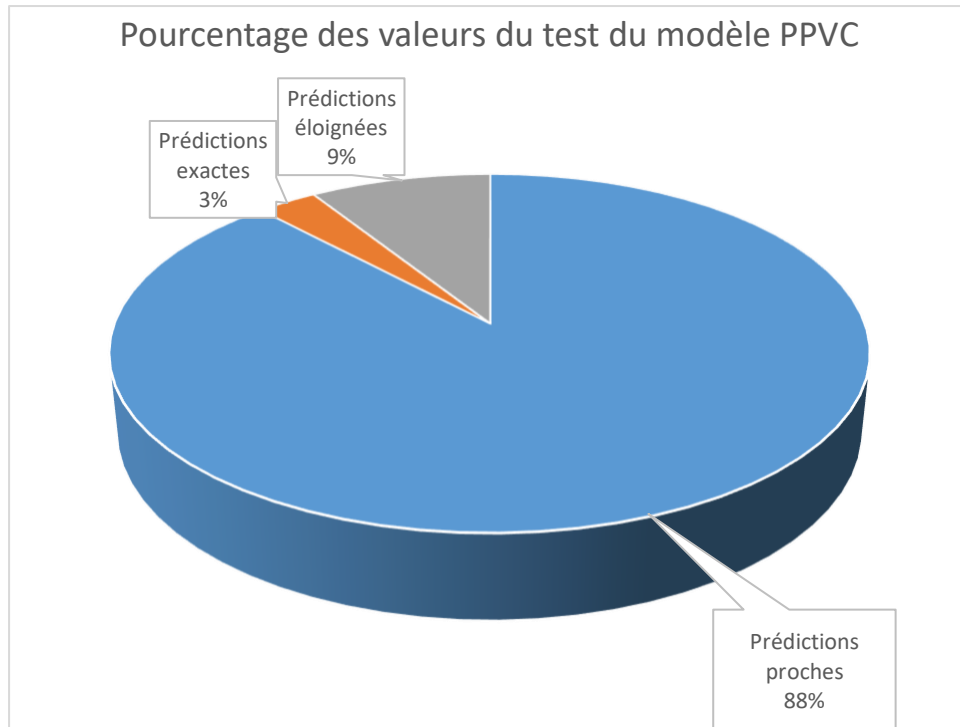
Quantité réelle	Quantité prédite	Distance Entre quantité réelle et prédite	Rapport X/S	Eclairage	Mode de cultivation
3.1865	2.907162	0,2793	2,2	1,6	0.9
0.9505	0.831456	0,1190	1,5	-1.0	0.0
1.6081	1.438757	0,1693	0.7	1.0	1.0
2.9232	3.040150	0,1170	3,6	1,8	1.0
7.3472	8.009683	0,6625	6,5	-1.0	0.0
1.5691	1.897424	0,3283	7,5	0.2	1.0
1.9321	2.574556	0,6425	1,2	0.8	0.7
1.5945	1.352529	0,2420	0.5	0.5	0.5
3.4254	5.295138	1,8697	3,4	-0.2	0.8
7.4147	7.387174	0,0275	4,8	-0.6	0.4
3.5612	3.633233	0,0720	5.0	1,2	1.0
2.1444	2.079667	0,0647	4,5	0.1	1.0
5.2856	5.494613	0,2090	4,9	0.3	0.6
1.9678	1.961878	0,0059	1,8	-0.2	0.2
5.8788	5.650055	0,2287	3.0	-0.2	0.8
3.0663	2.837280	0,2290	1,9	-0.1	0.9
0.9517	2.269145	1,3174	2,8	-1.0	0.0
2.0526	1.568317	0,4843	1.0	1.0	1.0
7.4873	7.240968	0,2463	4,6	-0.5	0.5
1.3900	1.745026	0,3550	3.0	0.8	1.0
0.6174	0.601625	0,0158	0.5	-1.0	0.0
1.4179	1.970867	0,5530	2,1	0.9	1.0
3.1967	2.585145	0,6116	1,2	1.0	1.0
3.4716	3.236503	0,2351	4,8	1,4	1.0
1.9153	2.167039	0,2517	1,9	0.9	1.0
7.3441	6.836477	0,5076	5,1	-0.9	0.1
1.8311	1.729383	0,1017	1,6	0.0	1.0
4.3018	2.635843	1,6660	2.0	1.0	1.0
2.5291	2.449914	0,0792	2,1	0.5	0.5
7.4355	6.962268	0,4732	4,4	-0.5	0.5
13.7366	10.960546	2,7761	7,5	-0.8	0.2
0.4432	1.585491	1,1423	3,2	0.8	1.0

**Discussion**

- Le nombre des prédictions qui sont proches aux valeurs réelles : 27/32.
- Le nombre des prédictions qui sont exactes et égales aux valeurs réelles : 1/32 (la ligne 65 sur dataset).
- Le nombre des prédictions qui sont éloignées aux valeurs réelles : 3/32.

Le calcul des pourcentages pour la confirmation des résultats du programme :

- Le pourcentage des prédictions proches :  $(28*100)/32 = 87.5\%$ .
- Le pourcentage des prédictions exactes :  $(1*100)/32 = 3.12\%$ .
- Le pourcentage des prédictions éloignées :  $(3*100)/32 = 9.37\%$ .



**Figure 22 :** Evaluation des résultats de prédiction.

Le pourcentage de réussite de prédiction est : le pourcentage des prédictions proches combiné avec le pourcentage des prédictions exactes ( $87.5\% + 3.12\% = 90.62\%$ ).



# Conclusion

**Conclusion :**

Le présent travail entre dans le domaine de la biotechnologie qui permet d'exploiter les nouvelles techniques notamment, de l'intelligence artificielle pour améliorer les résultats des systèmes biologiques.

Le but de ce travail est de prédire la quantité de la biomasse produite en fonction de certains paramètres. Le champ applicatif de notre travail est la productivité du *Chlorella vulgaris* par un bioréacteur via lactosérum en variant les valeurs de : le rapport S/X, l'éclairage et le mode nutritionnel.

Nous avons créé un modèle PPVC qui permet de réaliser cette prédiction. Ce modèle a été testé et il a donné un taux très élevé de réussite (92.18%). Ce taux signifie que le modèle est capable de prédire la quantité de biomasse.

L'apport de ce modèle est de prédire les résultats d'un bioréacteur pour des nouvelles valeurs des paramètres sans faire les expériences réelles. Cela nous permet de gagner de temps et le coût du matériel et matière.

Comme future travail, nous visons à exploiter des techniques de l'intelligence artificielle les plus avancées pour trouver les paramètres idéals (le rapport X/S, l'éclairage et le mode nutritionnel) pour une meilleure qualité de biomasse.

# **Références bibliographique**

**Abdelaziz, A. and Noui, M. G. E. B. (2022).** « Annotation des SNPs génomiques pour la prédiction de la maladie génétique/héréditaire d'Alzheimer par Deep Learning. Mémoire Master recherche : Bioinformatique ». Constantine : Université Frère Mentouri Constantine 1, 64 p.

**Azencott, C.-A. (2022).** « Introduction au machine learning ». Malakoff : Dunod.

**Belbekri, A. (2022).** « Classification des hydrolases à partir des données protéomiques par approche Deep Learning ». Mémoire Master Recherche : BioInformatique. Constantine : Université Frère Mentouri Constantine 1, 35 p.

**Berry, Michael W., Azlinah Mohamed, et Bee Wah Yap, éd. (2020).** Supervised and Unsupervised Learning for Data Science. Cham : Springer International Publishing.

**Bouznada, M. O., Demmane, D et Messaci Safa. (2022).** « utilisation du lactoserum récupéré des effluents de l'industrie laitière pour cultiver des microalgues dans des photobioréacteurs et revalorisation des sous-produits ». Mémoire Master recherche : Génie de l'Environnement. Constantine : Université Salah Boubnider, Constantine 03. 119p.

**Chisti Y. (2010).** « Fermentation technology ». *Industrial Biotechnology, Sustainable Growth and Economic Success*. 149–171.

**Clément, H. (2017).** « Instrumentation, modélisation et automatisation de fermenteurs Levuriers À destination œnologique ». Thèse de doctorat : Génie des procédés : Université Paris-Saclay.

**Demeyer A., Jacob F., Jay M., Menguy G., Perrier J. (1982).** « La conversion bioénergétique du rayonnement solaire et les biotechnologies », *Edition Technique & Documentation*, Paris, pp.187-212.

**Dike H. U., Zhou Y., Deveerasetty K. K. and Qingtian Wu. (2018).** « Unsupervised Learning Based On Artificial Neural Network : A Review ». *2018 IEEE International Conference on Cyborg and Bionic Systems (CBS)* [en ligne], (page consultée 20/05/2023). [doi: 10.1109/CBS.2018.8612259](https://doi.org/10.1109/CBS.2018.8612259).

**Grasmick A., Cabassud C., Spérandio M. and Christelle Wisniewsk. (2007).** Bioréacteurs à membranes et traitement des eaux usées. *Chimie verte*. 4, 127–130.

**Grisales Palacio V.H. (2007).** « Modélisation et commande floues de type Takagi-Sugeno appliquées à un bioprocédé de traitement des eaux usées ». Thèse de Doctorat : Systèmes Automatiques. Université Paul Sabatier-Toulouse III, 230p.

**Gueguim Kana, E. B., Oloke, J. K., Lateef, A. and Zebaze Kana Martiale gaetan. (2003).** « Constructional features of a 15-litre home-made bioreactor for fed-batch fermentations ». *African Journal of Biotechnology*, 2(8), 233-236

**Jagani, H., Hebbar, K., Gang, S., Raj, V., Chandrashekhar, H. and J.Venkata Rao (2010).** « An Overview of Fermenter and the Design Considerations to Enhance Its Productivity ». *Pharmacology*, 1, 261-301

**Jagriti, S., Nirmala, K. and Soumitra Biswas (2014).** « Bioreactors – Technology & Design ». *The Scitech Journal*, 1(6), 28-36

**Jo, Taeho. (2021).** Machine Learning Foundations : Supervised, Unsupervised, and Advanced Learning. Cham : Springer International Publishing.

**Jochum, M., Moncayo, L.P. and Jo, Y.-K. (2018).** « Microalgal cultivation for biofertilization in rice plants using a vertical semi-closed airlift photobioreactor ». *PLOS ONE*, 13(9).

**Le parc, P. (2017).** « Cours de l'informatique », support de cours pour les étudiants de 1<sup>ère</sup> année LMD Sciences et Techniques. Université de Bretagne Occidentale

**Lucas, M. A. (1992).** « INTELLIGENCE ARTIFICIELLE ET PHILOSOPHIE : LES CRITIQUES DE H. L DREYFUS ET J. SEARLE À L'INTELLIGENCE ARTIFICIELLE ». Exigence Partielle De La Maîtrise En Philosophie : Québec : l'Université Du Québec À Trois Rivères, 217 p.

**Maetschke, S. R., Madhamshettiwar, P. B., Davis, M. J., & Ragan, M. A. (2013).** « Supervised, semi-supervised and unsupervised inference of gene regulatory networks ». *Briefings in Bioinformatics*, 14(4), 326-336.

**Moore, J.H. and Raghavachari, N. (2019)** « Artificial intelligence based approaches to identify molecular determinants of exceptional health and life span-an interdisciplinary workshop at the National Institute on Aging », *Frontiers in Artificial Intelligence*, 2.

**Neha Rakesh Dhor and Dr. Badri N. Mohapatra (2022).** « Bioreactors – Technology and Design Analysis ». *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*, 2(2).

**Queinnec, I. (2000).** « Contribution à la commande de procédés biotechnologiques : application au traitement biologique de la pollution », Thèse de Doctorat, Université Paul Sabatier-Toulouse III, 96 p.

**Ramsundar, B., Liu, B., Wu, Z., Verras, A., Tudor, M. and Sheridan, Robert. (2019).** « Is There Hope for Computer-Aided Drug Discovery? » *Future Medicinal Chemistry*, 11(5), 429-431.

**Sanjogta Thapa Magar.** Microbe notes [en ligne]. (page consultée le 24/02/2023). <https://www.microbenotes.com/bioreactor/?fbclid=IwAR2Nz10KKWN66sjXVbna9ft4k-GoFzopIfQEa53iMBeZTP1mngiwcFjYFEw#3-aeration->

**Sarker I.H. (2021).** « Machine Learning: Algorithms, Real-World Applications and Research Directions ». *SN Computer Science* [en ligne], 2 (page consultée 20/05/2023). [doi: 10.20944/preprints202103.0216.v1](https://doi.org/10.20944/preprints202103.0216.v1)

**Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., ... & Kohli, P. (2020).** Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792), 706-710.

**Srinidhi C.L., Ciga O. and Martel Anne. L. (2021).** « Deep neural network models for computational histopathology : A survey ». *Medical Image Analysis*, 67, p101813.

**Toumi, H. (2009).** « Modélisation et identification paramétrique des processus de fermentation lactique ». Mémoire de Magister : Génie chimique. Université Ferhat Abbas, Sétif, Algérie, 100 p.

**Vishwajeet, G., Anil, P., Shubham, J., Poorva, S., Akshay, B., Akshay, J., Navnidhi, C. (2018).** « Designing of Fermenter and its utilization in food industries », Department of Food Technology and Nutrition, Lovely Professional University, Punjab, [en ligne] [doi:10.20944](https://doi.org/10.20944)

**Yang, J., Xu, J., Zhang, X., Wu, C., Lin, Tao et Yibin Ying. (2019).** « Deep learning for vibrational spectral analysis : Recent progress and a practical guide», *Analytica Chimica Acta*, 1081, pp. 6–17.

**Zhang C. et Lu Y. (2021).** « Study on Artificial Intelligence : The State of the Art and Future. Prospects ». *Journal of Industrial Information Integration*, 23

**Mémoire présenté en vue de l'obtention du Diplôme de Master**

**Domaine :** Sciences de la Nature et de la Vie

**Filière :** biotechnologie

**Spécialité :** *Mycologie et biotechnologie fongique*

**Titre**

**Développement d'un modèle Machine Learning pour la prédiction de la productivité d'un Bioréacteur**

**Résumé**

Notre travail appartient au domaine de biotechnologie qui sert à exploiter les techniques d'intelligence artificielle (IA) pour prédire la quantité d'une biomasse au niveau d'un bioréacteur en fonction de variance des certains paramètres. L'objectif de ce travail est de développer un modèle PPVC exploitant une des techniques de Machine Learning ou l'apprentissage automatique (ML) pour effectuer cette prédire. Sachant que, l'apprentissage automatique (ML) est une branche de l'intelligence artificielle (IA). Le modèle PPVC que nous avons créé est capable de prédire la quantité avec un taux de réussite très élevé (92.18%). Le champ applicatif de notre travail est la prédiction de la productivité du *Chlorella vulgaris* par un bioréacteur via lactosérum en variant les valeurs de : le rapport S/X, l'éclairage et le mode nutritionnel. Cela, permet de gagner de temps et le coût du matériel et matière car, nous avons la capacité de prédire la quantité sans faire des expériences.

**Mot clés :** Intelligence artificielle, Machine Learning (ML) ou l'apprentissage automatique, Prédiction, Bioréacteur.

**Membre du jury :**

**Président du jury :** Mme. Abdelaziz Ouidad (MCB - Université Frères Mentouri, Constantine 1).

**Rapporteur :** Mme. Djama ouahiba (MCB - Université Frères Mentouri, Constantine 1).

**Examineur :** Mme. Maziani Meriam (MCB - Université Frères Mentouri, Constantine 1).

**Présentée par :** Kabache Yasmine  
Bouzahzah Mohamed Lokmane

**Année universitaire :** 2022 -2023