# 

# 

# 

# 

# 

VOLTRON

**Section IA**

# 

*Proposition de valeur et*

*évaluation de la faisabilité*

## 

| **Version** | **Date** | **Auteur** | **Statut** |
| --- | --- | --- | --- |
| 1.0 | 27/05/2023 | CLAVIER Eliott, PIGNON Nathan | Première version du document |
| 1.1 | 29/05/2023 | BOUCHET Valentin | Ajout apprentissage continu |
| 1.2 | 30/05/2023 | PIGNON Nathan | Ajout des exigences et contraintes |
| 1.3 | 12/06/2023 | CLAVIER Eliott, PIGNON Nathan, RIPAULT Paul | Ajout de la description concrète de la proposition |
| 1.4 | 20/06/2023 | CLAVIER Eliott, RIPAULT Paul | Ajout documentation recherches vignes |
| 2.0 | 27/06/2023 | CLAVIER Eliott | Remise au propre du document |
| 2.1 | 02/07/2023 | CLAVIER Eliott | Ajout schéma architecture technique |
| 2.2 | 11/07/2023 | CLAVIER Eliott, PIGNON Nathan, RIPAULT Paul, BOUCHET Valentin | Complétion et ajustement spécifications techniques |

[**Exigences 5**](#_m7wx4eg8dhd1)

[Exigences fonctionnelles 5](#_otugtzmsrsoh)

[Analyse des données des vignobles 5](#_502fw7laduu)

[Suggestion sur la gestion des ressources humaines 5](#_s9p6ze3qj0wu)

[Prédiction des rendements de la vigne 5](#_xdorqar5bgm)

[Détection précoce des maladies de la vigne 5](#_nme33tgpo6g6)

[Gestion optimale de l'irrigation 6](#_x3eu64urfiu7)

[Surveillance des conditions environnementales 6](#_lollaj4bd7p7)

[Apprentissage continu 6](#_i7edgzw0ujhx)

[Exigences non-fonctionnelles 7](#_v0qca983hp0b)

[Compatibilité 7](#_tgzwar6okbjy)

[Confidentialité 7](#_yh4m1fhvn3z)

[Performance et disponibilité 7](#_oish9375gxxg)

[Scalabilité 7](#_m9zvsrxpzkt5)

[Fiabilité 8](#_idlrnangzzva)

[Facilité d’utilisation 8](#_48dy3py5vmmm)

[Durabilité 8](#_b4ovs7n9uqin)

[Risques et contraintes 9](#_keobtb1x3k5o)

[**Propositions 10**](#_laq68lb4mfy3)

[Proposition minimale 10](#_9nsw5l4a5txx)

[Proposition intermédiaire 10](#_2wn5g0vbvc7h)

[Proposition intégrale (choisie par le client) 11](#_a1op5y1wyv5a)

[Récapitulatif des propositions 12](#_9uiemdd30tj7)

[Architecture de la solution 13](#_a3ttn1vpmwgz)

[**Formatage et monitoring des données 14**](#_gydig6j2w7pd)

[Moyens 14](#_pam3flqblf88)

[Présentation 14](#_82aqv3oeog3k)

[**Apprentissage et prédictions 15**](#_pjbfolm307m6)

[Analyse des données météorologiques et images 15](#_jyckc9godu7x)

[Technologies 15](#_hzj793qb2gxe)

[Intégration de l’équation Penman-Monteith 15](#_q2jkxcelm7ay)

[Introduction aux réseaux de neurones artificiels (ANN) 16](#_kl0ig6sdxyqt)

[Réseaux de neurones convolutifs (CNN) et utilité 16](#_rt27qghztfh6)

[Solution de mise en place 17](#_9a3iz69lr3pz)

[Ordre de grandeur des coûts (monétaire et temps nécessaire) 18](#_gmqsnog70gs1)

[Jeux de données et modèles 18](#_w7etlhd1xau9)

[Environnement personnalisé 19](#_mlbcoc5susrb)

[Cloud computing 19](#_6mp964s5yjw3)

[Apprentissage continu 20](#_msajeur7ovdh)

[Visualisation des prédictions 22](#_kkxpzge95d0q)

[Feedback 23](#_ljka7qiomxa5)

[Mise en place 23](#_gkatuq4s7hho)

[Risques 23](#_39e9rjfspw36)

[Annotation 24](#_amrev3zz0vh)

[Interface et API 24](#_zayohgv5x6u)

[**Articles pertinents 25**](#_xt9nwshf085x)

[IA dans le domaine de la viticulture 25](#_y3v0dghxydx6)

[Maladies de la vigne 25](#_ya7ax25t3uop)

[Données et POC 25](#_gkw9xsi1b0at)

[**Tableau récapitulatif parasites / maladies de la vigne 27**](#_qzf0v4wvym68)

# 

# Exigences

## Exigences fonctionnelles

L'organisation Vignerons Indépendants cherche à améliorer la gestion de ses vignobles grâce à l'application de technologies innovantes. Leur objectif est de maximiser la productivité et la qualité de leurs vignes tout en minimisant les coûts et l'impact environnemental. Pour atteindre cet objectif, plusieurs domaines clés ont été identifiés où l'intelligence artificielle et la *big data* pourraient apporter une valeur significative:

### Analyse des données des vignobles

**Description :** Le système doit être capable d'analyser de grandes quantités de données de vignoble et de les transformer en informations exploitables.

**Détails :** Le système utilisera des algorithmes d'IA pour identifier les tendances et les modèles dans les données. Ces insights aideront les Vignerons Indépendants à prendre des décisions éclairées basées sur des données.

### Suggestion sur la gestion des ressources humaines

**Description :** Le système doit aider les vignerons dans la gestion des ressources humaines selon les besoins périodiques.

**Détails :** La *data visualisation* qui découle des prédictions du système doit permettre de mieux gérer les ressources humaines, notamment pour mieux diriger les équipes en temps-réel sur les zones sensibles sur lesquelles les vignerons doivent effectuer des actions correctives.

### Prédiction des rendements de la vigne

**Description :** Le système doit être capable de prédire les rendements de la vigne en fonction des conditions environnementales et de l'état de santé des vignes.

**Détails :** Le système utilisera l'apprentissage automatique pour analyser les données historiques et actuelles et faire des prédictions précises sur les rendements futurs. Cela aidera les Vignerons Indépendants à optimiser leur production.

### Détection précoce des maladies de la vigne

**Description :** Le système doit être capable de détecter les signes précoces de maladies de la vigne.

**Détails :** Le système utilisera l'analyse d'image et l'apprentissage automatique pour identifier les signes de maladies avant qu'elles ne se propagent. Cela aidera les Vignerons Indépendants à maintenir la santé de leurs vignes.

### Gestion optimale de l'irrigation

**Description :** Le système doit être capable de surveiller l'humidité du sol et de prédire les besoins en eau des vignes.

**Détails :** Le système utilisera des capteurs et des algorithmes d'apprentissage automatique pour fournir des recommandations d'irrigation précises et opportunes. Cela aidera les Vignerons Indépendants à économiser de l'eau et à améliorer la santé de leurs vignes.

### Surveillance des conditions environnementales

**Description :** Le système doit être capable de surveiller en temps réel les conditions météorologiques, la température, l'humidité et la qualité du sol.

**Détails :** Le système utilisera l'IA pour analyser ces données et fournir des alertes en cas de conditions qui pourraient affecter la santé ou la productivité des vignes. Cela aidera les Vignerons Indépendants à adapter rapidement leurs pratiques de gestion des vignobles.

### Apprentissage continu

**Description :** Le système doit être capable d’apprendre continuellement sur des nouvelles données pour ajuster au fur et à mesure ses recommandations ou ses alertes sur les vignobles et ses problématiques.

**Détails :** Le système utilisera des méthodologies d’apprentissage continu alimentées par l’arrivée de nouvelles données capturées et stockées. Les nouvelles pourront notamment permettre au fil du temps:

* de supporter la reconnaissance de nouvelles maladies qui n’auraient pas été découverte sur l’ensemble des vignobles sur une période de temps donnée
* de prédire des actions correctives plus justes selon les aléas des saisons et des années, où les conditions climatiques / météorologiques / foncières peuvent variées et être plus ou moins clémentes ou extrêmes pour l’activité des vignerons.

### 

## Exigences non-fonctionnelles

L’apport de réponses aux exigences fonctionnelles doit être accompagné de réponses à des exigences non fonctionnelles qui assurent la performance et de la pertinence des solutions proposées. Parmi ces exigences non-fonctionnelles, on peut relever:

### Compatibilité

**Description :** Le système doit être compatible à l’ensemble des solutions apportées par le projet

**Détails :** Pour minimiser les coûts et faciliter l'intégration du nouveau système, il est essentiel que celui-ci soit compatible avec les outils et les systèmes déjà en place dans l'organisation ou mis en place suite à la réponse de l’appel d'offres.

### Confidentialité

**Description :** Le système doit assurer la confidentialité des données à l’intérieur et l’extérieur du système.

**Détails :** Compte tenu de la sensibilité des données impliquées, il est essentiel que le système assure la sécurité et la confidentialité des données à tout moment. L’accès aux données est dissocié de leur exploitation et doit être suffisamment délimité pour éviter tout conflit par rapport à la concurrence entre vignerons de l’association Vignerons Indépendants et les vignerons externes.

### Performance et disponibilité

**Description :** Le système doit être suffisamment performant pour être utilisable dans un cas concret d’exploitation.

**Détails :** Le système se doit de répondre à deux axes liés à la performance:

* le système doit répondre rapidement lorsqu’il est sollicité
* le système doit apprendre et se rendre disponible à chaque apprentissage de manière continu dans le meilleur des cas

### Scalabilité

**Description :** Le système doit être capable de gérer une augmentation de la charge de travail sans affecter ses performances.

**Détails :** Comme les Vignerons Indépendants pourraient augmenter le nombre de vignobles utilisant le système à l'avenir, il est essentiel que le système puisse gérer une augmentation du volume de données et du nombre d'utilisateurs sans compromettre sa performance.

### 

### Fiabilité

**Description :** Le système doit être fiable et fournir des résultats précis et cohérents.

**Détails :** Les Vignerons Indépendants dépendent du système pour prendre des décisions importantes concernant la gestion de leurs vignobles. Par conséquent, il est essentiel que le système soit fiable et qu'il fournisse des résultats précis et cohérents à chaque fois.

### Facilité d’utilisation

**Description :** Le système doit être facile à utiliser pour les utilisateurs finaux.

**Détails :** Les Vignerons Indépendants auront besoin de comprendre et d'utiliser le système au quotidien. Par conséquent, il est essentiel que le système soit intuitif et facile à utiliser, même pour les utilisateurs qui ne sont pas techniquement avertis.

### Durabilité

**Description :** Le système doit être conçu de manière à minimiser son impact environnemental.

**Détails :** Compte tenu de l'engagement des Vignerons Indépendants envers la durabilité, il est essentiel que le système soit conçu de manière à minimiser son impact environnemental. Cela pourrait impliquer l'utilisation de serveurs éco-énergétiques ou l'optimisation du système pour réduire la consommation d'énergie.

## 

## Risques et contraintes

L'implémentation d'un système d'intelligence artificielle et de big data dans le domaine de la viticulture, bien que prometteuse, n'est pas sans défis. Il est crucial de prendre en compte les risques et contraintes potentiels pour assurer une transition réussie vers cette technologie innovante. Ces risques et contraintes peuvent être de nature technique, opérationnelle, financière ou réglementaire. Des risques et contraintes sont alors identifiés et associés à ce projet. Pour chacun, des stratégies d'atténuation efficaces pour chaque scénario sont indiquées. Les risques et contraintes soulevés sont les suivantes :

* L'intégration de nouvelles technologies dans les systèmes existants peut présenter des défis techniques. Il est essentiel de s'assurer que le nouveau système est compatible avec les outils et les systèmes déjà en place.
* La gestion de grandes quantités de données sensibles présente des risques en termes de sécurité et de confidentialité. Il est essentiel de mettre en place des mesures de sécurité robustes pour protéger les données contre les accès non autorisés et les violations de données.
* La précision des prédictions dépend de la qualité des données et des modèles d'apprentissage automatique utilisés. Il est essentiel de s'assurer que les données sont précises et à jour et que les modèles sont régulièrement mis à jour et validés.
* L'adoption de nouvelles technologies dépend de leur facilité d'utilisation et de leur valeur ajoutée perçue. Il est essentiel de fournir une formation adéquate aux utilisateurs et de communiquer clairement les avantages du nouveau système.
* Les performances du système peuvent être affectées par des conditions environnementales extrêmes, telles que des tempêtes ou des sécheresses. Cela pourrait impacter la qualité des données récoltées et biaiser ou impacter négativement la performance de la solution proposée au client. Il est essentiel de concevoir le système de manière à être résilient face à ces conditions, en adoptant un système discriminant permettant de faire le tri.
* L'implémentation et la maintenance du système peuvent entraîner des coûts significatifs. Il est essentiel de planifier et de gérer ces coûts de manière efficace.

# Propositions

Suite à l’appel d’offres et aux exigences fonctionnelles et non-fonctionnelles identifiées, plusieurs propositions de solution ont été apportées afin de répondre au mieux à la demande du client et ses exigences, ainsi que les exigences non fonctionnelles.

Chacune des propositions ont été budgétisée et leur réalisation estimée dans le temps. Le détail et la justification des coûts a été rédigée pour la solution choisie par le client **(proposition intégrale)**. L’ensemble des coûts et leurs justifications sont disponibles dans les deux documents suivants:

* [Tableau des coûts](https://docs.google.com/spreadsheets/d/1gzNDuPedpnHkAKogtQPsu95y6HtZlPELAK2M6g9MK6o/edit?usp=sharing)
* [Justification des coûts et explication de la procédure de réalisation](https://docs.google.com/document/d/1ynmv8OlxK-DKjkHKHvHVU8cKhN1hvMPyQcw0UrlY-5s/edit?usp=sharing)

## Proposition minimale

Enfin, la solution réfléchie par l’équipe d’experts en intelligence artificielle en ce qui concerne la proposition minimale est composée de:

* Une IA de reconnaissance des maladies sur les vignes, qui sont la cause principale d’une perte de production et qui est le plus pertinent à corriger en premier lieu. **Cette reconnaissance se fera par le biais d’images de vignes prises depuis le sol.**
* La mise en place de l’apprentissage continu du modèle sous forme de soumission de lot d’images périodiquement pour ajuster la performance de l’IA. Pour le cas des images, **une soumission chaque semaine** permet d’ajuster suffisamment les prédictions tout en limitant les coûts en faisant un entraînement plus conséquent mais moins sollicitant de manière ponctuelle.
* La mise à disposition d’une API permettant de requêter facilement le modèle entraîné pour obtenir des prédictions.
* Une interface de rapports permettant d’indiquer où les vignes présentent des potentiels cas de maladies sur un vignoble donné en temps-réel.

## Proposition intermédiaire

* L’ensemble des solutions composant la proposition minimale sont reprises et combinées aux propositions suivantes.
* Un script d’analyse de **données climatiques / météorologiques** qui aiderait les vignerons à optimiser leurs exploitations en apportant une meilleure gestion de l’irrigation et du vignoble grâce à l’application d’actions correctives. Ce script intégrerait **l’équation Penman-Monteith**. **Parmi les données à analyser, on peut relever la température, la pression de l’air, l’humidité dans l’air et la radiation solaire.**

En fonction de ces données, il sera possible de prédire plusieurs jours à l’avance les quantité d’eau à fournir pour irriguer correctement l’exploitation selon les données capturées par la station météo.

* La mise en place de l’apprentissage continu du modèle sous forme de soumission de lot de données périodiquement pour ajuster la performance de l’IA. Pour le cas des images, **une soumission une fois par semaine** semble adaptée pour pouvoir s’adapter aux problématiques saisonnières.
* La mise à jour de l’API et de l’interface de rapports en conséquence pour supporter l’analyse des **données climatiques et météorologiques**.

## Proposition intégrale (choisie par le client)

Enfin, la solution réfléchie par l’équipe d’experts en intelligence artificielle en ce qui concerne la proposition intégrale est composée de:

* Une IA de reconnaissance des maladies sur les vignes, qui sont la cause principale d’une perte de production et qui est le plus pertinent à corriger en premier lieu. **Cette reconnaissance se fera par le biais d’images de vignes prises depuis le sol.**
* La mise à disposition d’une API permettant de requêter facilement le modèle entraîné pour obtenir des prédictions.
* Un script d’analyse de **données climatiques / météorologiques** qui aiderait les vignerons à optimiser leurs exploitations en apportant une meilleure gestion de l’irrigation et du vignoble grâce à l’application d’actions correctives. Ce script intégrerait **l’équation Penman-Monteith**. **Parmi les données à analyser, on peut relever la température, la pression de l’air, l’humidité dans l’air et la radiation solaire.**

En fonction de ces données, il sera possible de prédire plusieurs jours à l’avance les quantité d’eau à fournir pour irriguer correctement l’exploitation selon les données capturées par la station météo.

* La mise en place de l’apprentissage continu du modèle sous forme de soumission de lot de données périodiquement pour ajuster la performance de l'IA. Pour le cas des images, **une soumission tous les deux jours** semble adaptée pour pouvoir s’adapter rapidement aux problématiques saisonnières.
* Une interface de feedback où les vignerons pourraient indiquer si les prédictions sur les maladies des vignes potentiellement identifiées sont justes ou si les prédictions sur les quantités d’irrigation à apporter sont pertinentes, ce qui permet d’améliorer le processus d’apprentissage continu.
* Une visualisation selon une interface présentant une carte des vignes touchées par les maladies sur plusieurs vignobles donnés, pour pouvoir endiguer les problématiques climatiques / météorologiques ou liées à la maladie. **Cette visualisation pourrait également permettre de mieux répartir la charge de travail et diriger plus facilement les équipes vers les zones sensibles.**

## 

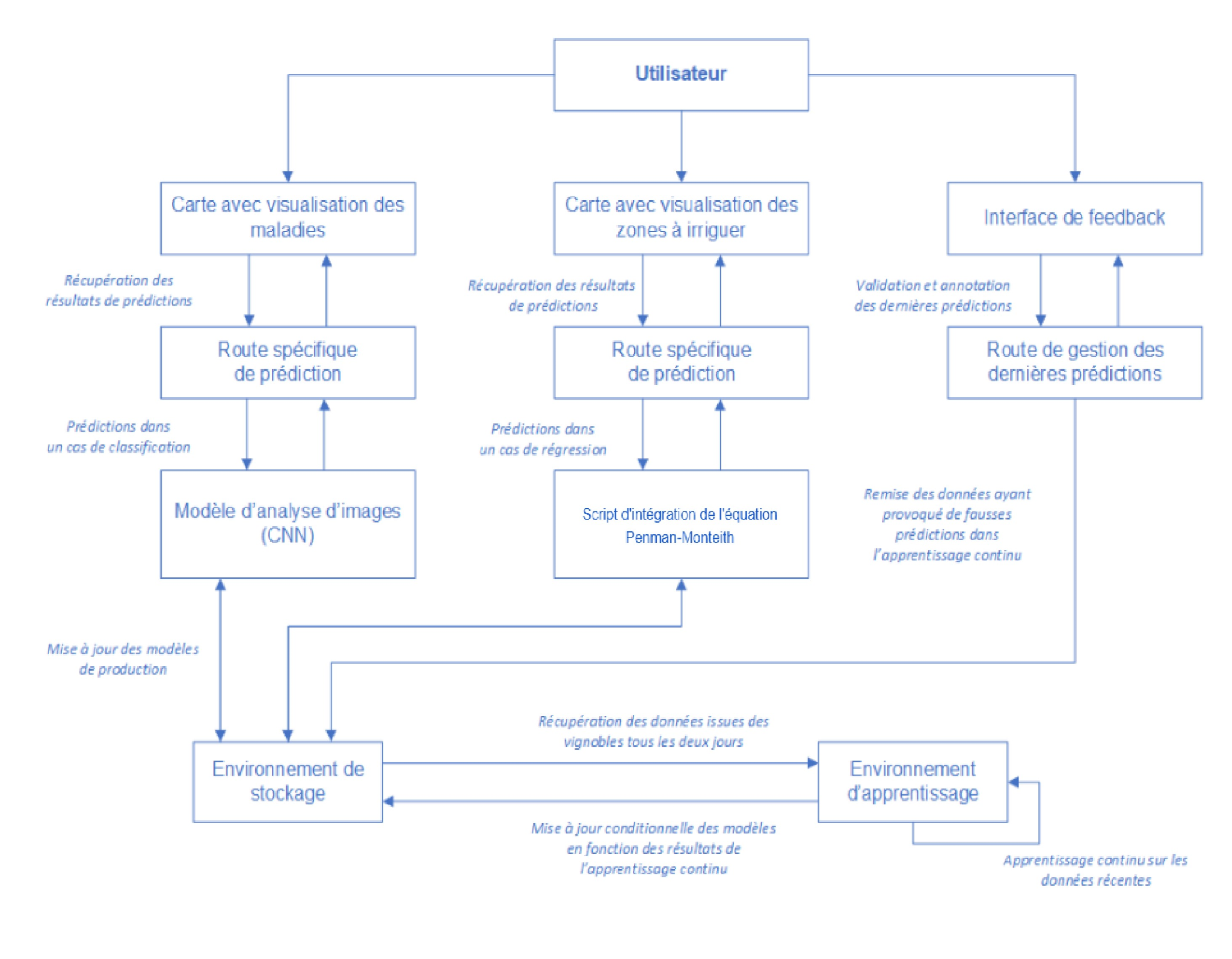
## Récapitulatif des propositions

|  | **Proposition minimale** | **Proposition intermédiaire** | **Proposition intégrale** |
| --- | --- | --- | --- |
| Analyse des données des vignobles | ✅ | ✅ | ✅ |
| Prédiction des rendements de la vigne | ✅ | ✅ | ✅ |
| Détection précoce des maladies de la vigne | ✅ | ✅ | ✅ |
| Gestion optimale de l’irrigation | ❌ | ✅ | ✅ |
| Surveillance des conditions environnementales | ❌ | ✅ | ✅ |
| Apprentissage continu amélioré avec interface de feedback | ❌ | ❌ | ✅ |
| Visualisation à grande échelle sur une carte des vignes touchées par des maladies | ❌ | ❌ | ✅ |

## 

## Architecture de la solution

Ci-dessous se trouve un schéma technique récapitulatif de la solution présentée dans la proposition intégrale, qui permet notamment de comprendre les tenants et aboutissants de chacune de ses composantes.



# Formatage et monitoring des données

## Moyens

Le formatage des données est une des étapes les plus importantes pour avoir une IA performante. Pour les **données caractérielles** (température, humidité …), il faudra les préparer et les formater. Cela peut impliquer différentes étapes, telles que la gestion des valeurs manquantes en les remplissant par des **techniques d'imputation**, la normalisation des valeurs numériques pour les mettre à la **même échelle** (par exemple, en utilisant la normalisation min-max ou la standardisation), l'encodage des variables catégorielles en utilisant des techniques telles que **l'encodage one-hot** ou **l'encodage des étiquettes**, et la **séparation des données** en ensembles d'entraînement, de validation et de test.

En ce qui concerne les **données de type image**, cela va être relativement similaire, cependant les images vont avoir quelques étapes supplémentaires. Parmi ces étapes, il y a le **redimensionnement de toutes les images** à une taille uniforme pour assurer la cohérence, en utilisant des techniques telles que **l'interpolation bilinéaire** ou la mise à l'échelle. Il est également possible d’appliquer des techniques d'augmentation des données, telles que la rotation, le retournement horizontal ou vertical, le recadrage aléatoire, pour augmenter la variabilité des données d'entraînement. De plus, les images doivent être converties en un format compatible avec votre modèle, tel qu'un tableau de pixels en niveaux de gris ou en couleurs RVB.

## Présentation

Concernant les images rassemblées, **un stockage par dossier qui identifient le type de maladie suffit.**

Pour le cas du stockage des données mesurées, on peut imaginer une feuille de type **Excel,** avec plusieurs colonnes notamment:

* Secteur
* Jour
* Heure
* Humidité de l’air
* Humidité du sol
* Température
* ***Autre données à l’instant T***

**La colonne *Secteur* peut permettre de réaliser des actions correctives plus précises**, notamment car les données mesurées vont constituer un suivi et vont être impactées par les actions correctives précédentes appliquées (ex: si on irrigue un jour J, l’humidité est susceptible d’augmenter pour le jour J+1).

Selon les activités du vignoble et le métier de vigneron, on peut aussi imaginer réaliser une deuxième feuille qui combine cette fois-ci les résultats d’une semaine en une ligne en faisant la moyenne des données mesurées. Cela peut permettre de mesurer l’évolution du vignoble sur l’année ou des périodes localisées.

# Apprentissage et prédictions

## Analyse des données météorologiques et images

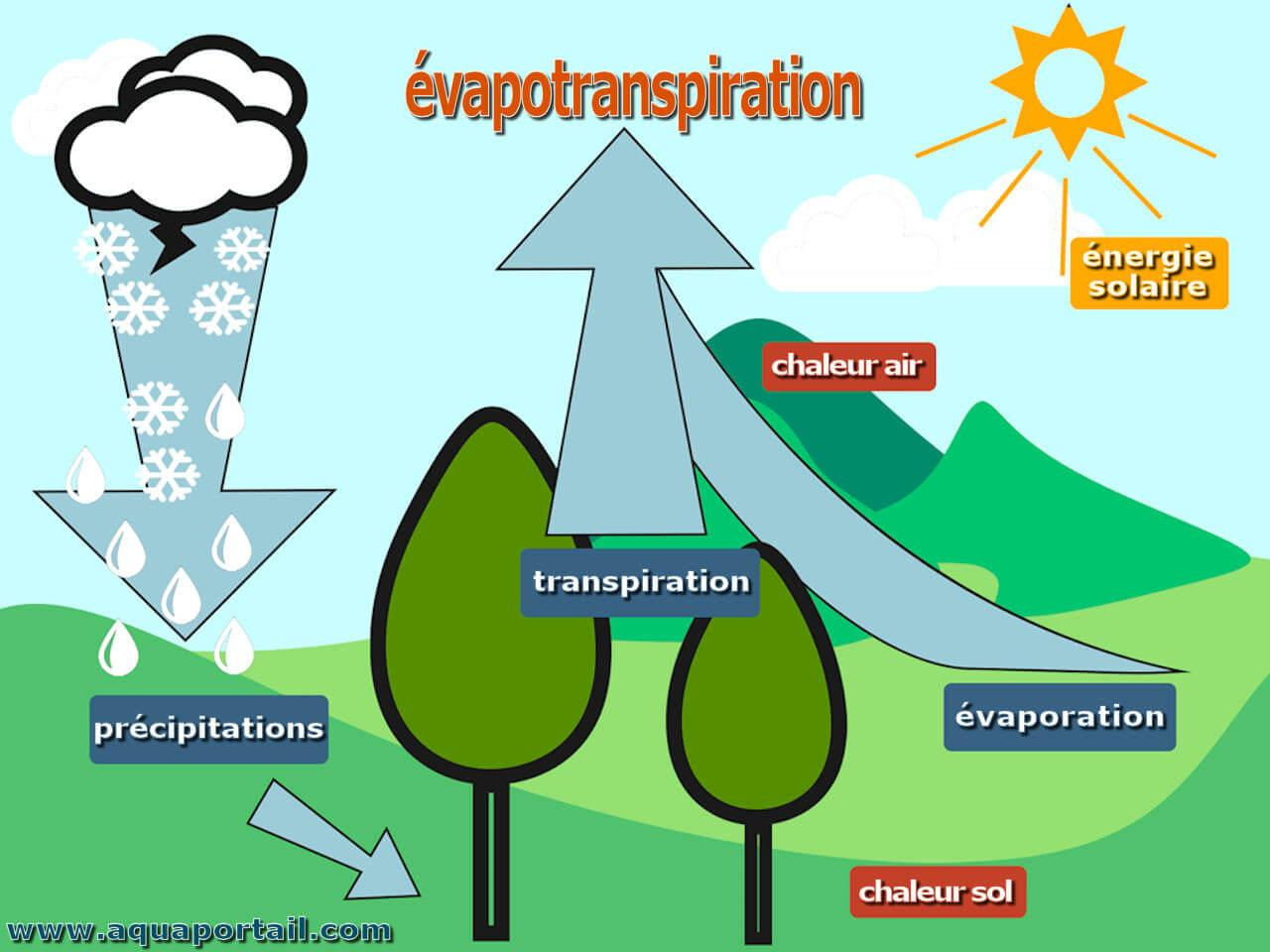
### Technologies

En exploitant les données stockées et formatées concernant les mesures faites (humidité de l’air, température, etc.) ainsi que les images produites sur les vignes des vignobles du client, il est possible de prédire certains scénarios quant à l’impact de ces facteurs sur le vignoble et d’anticiper des actions correctives à apporter pour maximiser la production et éventuellement faire des économies. Ces prédictions peuvent se faire par le biais de l’utilisation **de réseaux de neurones artificiels (*deep learning*);** c’est aujourd’hui le moyen le plus poussé et performant pour obtenir des prédictions précises de manière automatisée, ce qui limite l’action de l’humain et donc les coûts et les éventuelles erreurs liées à l’homme.

Il est également possible de réaliser des calculs mathématiques selon certaines données afin d’estimer exactement les ressources nécessaires à une tâche donnée et donc optimiser les coûts de l’exploitation et la qualité finale du produit.

Selon les données produites par les différents matériels mis en place et selon les besoins exprimés dans l’appel d’offre, **deux solutions complémentaires** peuvent être conçues pour maintenir au mieux l’activité du vignoble.

#### Intégration de l’équation Penman-Monteith



Afin d’estimer les quantités d’eau nécessaire à la bonne irrigation des vignes, il est possible d’intégrer l’équation de **Penman-Monteith**. Cette équation permet de calculer **l'évapotranspiration potentielle (notée ETo)** de l’environnement étudié et d’en déduire sa perte en eau et le manque à compenser.

Le calcul de l’évapotranspiration potentielle prend en compte l’ensemble des caractéristiques météorologiques **journalières** pouvant avoir un impact sur le phénomène. Parmi ces caractéristiques, on retrouve notamment l’humidité, la radiation solaire, la température moyenne, les précipitations ou encore la vitesse du vent.

L'évapotranspiration potentielle couplée aux “crop coefficients” de l’espèce de vigne permet ensuite de calculer l’irrigation nécessaire en mm par plant. Le **“Crop Coefficient” (Kc)** est une valeur située **autour de 1** qui dépend du type de vigne étudiée et de la période de la saison.

**L’irrigation nécessaire en mm est alors le résultat du produit de l'évapotranspiration potentielle par le “crop coefficient” correspondant.**

**L’intégration de l’équation et son calcul seront réalisés par un script (Python).**

[Plus de détails concernant l’implémentation de cette équation dans le cas d’étude actuel](https://docs.google.com/document/u/0/d/1nYMldgadF0Y2ZlABp8OWHRxzAd5tPPrst_9fCgox9Mg/edit)

#### Introduction aux réseaux de neurones artificiels (ANN)

Pour expliquer simplement, un ANN (Artificial Neural Network), ou réseau neuronal artificiel en français, est un modèle mathématique inspiré du fonctionnement des neurones dans le cerveau humain. C'est une technique utilisée en intelligence artificielle pour permettre à un ordinateur d'apprendre à partir de données et d'effectuer des tâches spécifiques.

Un ANN est composé de plusieurs couches de neurones artificiels interconnectés. Chaque neurone reçoit des entrées, effectue des calculs sur ces entrées, puis transmet le résultat à d'autres neurones. Ces calculs sont basés sur des poids et des seuils ajustables, qui sont modifiés lors de l'apprentissage du réseau.

Avant d’être utilisé, l’ANN doit être entraîné en exposant le réseau à un ensemble de données d'entraînement, comprenant des exemples d'entrées avec les résultats attendus en face. Le réseau ajuste alors les poids et les seuils de ses neurones afin de minimiser l'écart entre les résultats obtenus et les résultats attendus au fur et à mesure de l'entraînement.

**Un ANN peut ensuite être mis à disposition pour réaliser des prédictions sur des données qui ne sont pas intervenus dans l'entraînement de ce dernier.**

#### Réseaux de neurones convolutifs (CNN) et utilité

En ce qui concerne les images capturées sur les vignes des vignobles du client, il est possible de notifier les viticulteurs de la présence de certaines anomalies grâce à un **réseau de neurones convolutifs dit CNN.**

Un CNN (Convolutional Neural Network), est un type de réseau neuronal utilisé principalement pour l'analyse et le traitement des données visuelles, telles que des images. Il est largement utilisé dans des domaines tels que la vision par ordinateur, la reconnaissance d'images et la classification d'images. **C’est une extension de l’ANN.**

La particularité d'un CNN réside dans son architecture qui exploite les opérations de convolution. La convolution consiste à appliquer un ensemble de filtres (matrices) sur l'image d'entrée pour extraire des caractéristiques spécifiques. Ces filtres identifient des motifs visuels tels que des bords, des textures ou des formes présentes dans l'image.

De la même manière qu’un simple ANN, un CNN doit suivre une phase d'entraînement. Les données fournies en entrée sont des images qui permettent au CNN de reconnaître et de généraliser les caractéristiques visuelles, ce qui lui permettra par la suite de classifier de nouvelles images lors de ses prédictions. **Ces prédictions vont pouvoir porter sur plusieurs domaines:**

* Détecter les maladies des vignes, cette fois-ci selon des images; un CNN peut être entraîné à détecter les signes de maladies sur les feuilles ou les vignes à partir des images capturées par la caméra. En utilisant un ensemble d'images étiquetées avec des exemples de feuilles saines et infectées, le CNN peut apprendre à reconnaître les motifs visuels associés à différentes maladies.
* Suivant la croissance des vignes, les images capturées par la caméra peuvent également être utilisées pour surveiller la croissance des vignes au fil du temps. Un CNN peut être utilisé pour extraire des caractéristiques telles que la densité foliaire, la hauteur des vignes, ou l'étendue de la canopée. On peut alors identifier les zones à faible rendement et prendre des décisions de gestion appropriées.
* Evaluer la maturité des raisins en parallèle du suivi de la croissance des vignes. Un CNN peut être formé pour reconnaître les caractéristiques visuelles des raisins mûrs, tels que la couleur ou la taille, à partir d’images aériennes, ce qui aide à déterminer le moment optimal pour la récolte, en maximisant la qualité des raisins et la production de vin.

**Dans le cas actuel d’étude, l’utilisation d’un CNN permettra de détecter les maladies des vignes sur des images de feuilles de vignes fournies pour permettre au client de l’aider dans la reconnaissance des maladies et les anticiper.**

**Le CNN sera mis en place à l’aide des outils Tensorflow et Keras, spécialisés respectivement dans le *Machine Learning* et le *Deep Learning*.**

### Solution de mise en place

Concernant l’aspect **Deep Learning**, on a donc une possibilité de réseaux neuronaux pouvant permettre d’améliorer la performance du vignoble, à savoir un CNN de classification. Complémentairement, l’intégration de l’équation Penman-Monteith permet également de répondre à des exigences fonctionnelles du client.

Sachant cela, il est possible de dimensionner un environnement type pouvant pour l’algorithme de **Machine Learning** proposée. Ci-dessous, est proposé les solutions techniques optimales pour la **proposition intégrale** afin de démontrer les avantages et risques de cette dernière. Parmi ces solutions, on trouve une solution à faible coût, une solution sans contrainte budgétaire et humaine établie et une solution qui semble la plus optimale.

Pour cette solution, trois grands axes sont abordés: la **récupération de données**, la **construction et l’expérimentation du modèle** et **l'entraînement et le maintien du modèle**. Pour les estimations de temps, le matériel jouera forcément sur le temps que prendra les différentes étapes, c’est pourquoi ce dernier est évalué en prenant une base de **16 vCPUs avec 64go de RAM**. Même si ce n’est pas la solution choisie, il est facile de convertir le temps que cela prendra en fonction de la configuration souhaitée.

|  | **Solution proposée** | **Coût (humain et budgétaire)** | **Avantages** | **Risques** |
| --- | --- | --- | --- | --- |
| **Récupération des données** | Jeux de données existants combinés avec plusieurs vignobles voisins et des jeux de données génériques | Aucun coût monétaire pour les jeux de données génériques  Aucun coût monétaire pour les jeux de données déjà existants  Peut nécessiter de collaborer avec d’autres vignobles et de devoir investir | Beaucoup de données potentielles, avec des données génériques de base qui assure de pouvoir entraîner à minima | Peu confidentiel dans le cas d’une collaboration, potentiellement trop générique et peu adapté aux vignobles du client |
| **Construction et expérimentation** | Modèle optimal expérimenté sur un environnement personnalisé basé sur des architectures déjà reconnues | Investissement pour mettre en place un environnement (GPU) + temps de recherche | Résultats très précis et générique sur un type de vignoble | Demande du temps de recherche pour obtenir les meilleurs résultats et modèles les plus optimaux |
| **Entraînement et maintien** | Environnement dédié au vignoble avec matériel adapté et performant | Temps d'entraînement conséquent et environnement continuellement sollicité pour apprendre en continu | Un environnement en accès continu permettant de pouvoir entraîner librement | Peut nécessiter des périodes de maintenance; des procédures en cas d’incident matériel doivent être anticipées |
|  | Plateformes de cloud computing avec une offre avancée fortement disponible et performante | Moins coûteux sur le court et moyen terme, plus coûteux sur le long terme  Nécessite une expertise sur la plateforme choisie | Redimensionnement de la solution plus simple par rapport à l’investissement d’un environnement dédié fixe | Peu confidentiel, nécessité d’adapter la solution de stockage à une solution faisant partie de la même suite de cloud computing choisie, disponibilité pouvant fluctuer |

### Ordre de grandeur des coûts (monétaire et temps nécessaire)

#### Jeux de données et modèles

Comme explicité, il est possible d’avoir des jeux de données gratuits concernant les feuilles de vignes ou dans photos de vignoble vu du ciel. En ce qui concerne l’option pour construire un jeu de données de zéro, **le coût monétaire dépendra d’abord de l’ensemble du matériel physique à acheter pour capturer différentes images et capturer des données selon certains paramètres** *(voir section IoT).* Il faudra également tenir compte du coût de stockage de ces données *(voir section Big Data).* Il est aussi possible d’obtenir des modèles pré entraînés sur ces jeux de données gratuitement, ou plus simplement des architectures déjà implémentés de modèles complexes et potentiellement adaptés à au cas d’usage.

Le temps nécessaire pour faire un jeu de données à partir de zéro, prend forcément du temps pour avoir une base de travail adéquate. Il faut avoir des données dans le plus de conditions climatiques différentes ce qui nécessite d’avoir des relevés pour chacune des saisons dans l’année pour avoir une base solide. La solution la plus optimale serait donc de coupler des données déjà existantes avec les données qui vont être accumulées au fur et à mesure.

[*Exemples de jeux de données et de modèles à libre disposition*](#_gkw9xsi1b0at)

#### Environnement personnalisé

Lors de la construction d’un environnement d’expérimentation et d'entraînement personnalisé, le principal coût réside dans l’investissement du matériel de calcul permettant d’effectuer les opérations d'entraînement. Dans le cadre d’un environnement personnalisé, plusieurs options plus ou moins performantes permettent de réaliser ce genre de tâches:

* un CPU (avec RAM), qui constitue la base de la ressource matérielle pour l'entraînement. Les coûts peuvent varier entre **200 et 1000 euros (d’un i5 à un i9 par exemple)** entre de l’entrée de gamme et de la haute gamme concernant le CPU. On peut y ajouter la RAM, qui doit être dimensionnée correctement avec le CPU, qui peut varier **entre 100 et 600 euros (16go à 64go avec variation de fréquence plus ou moins haute et technologie DDR3, DDR4 voire DDR5)**.
* un GPU en plus du CPU, essentiel pour booster l'entraînement d’un CNN en particulier et plus rapide qu’un CPU en général. Le prix d’un GPU peut varier entre **300 et 1500 euros** **pour ceux accessible au grand public et qui conviendrait au client**.

L’avantage de ces solutions est qu'elles permettent d’avoir un coût moins important sur le long terme et de centraliser l’activité d'entraînement chez le client. Elle peut être néanmoins moins performante que certaines solutions proposées par les offres de **cloud computing** et nécessite aussi de la **maintenance** en cas de problème si le client n’est pas expert dans le domaine (en cas de casse matérielle par exemple).

#### Cloud computing

Il est aussi possible de souscrire des offres de cloud computing souvent plus performantes que de construire un environnement personnalisé, avec une garantie de disponibilité sur des horaires précises. L’avantage de ces solutions est aussi qu’il est possible de dimensionner correctement l’usage du matériel sur des plages dédiées et d’optimiser les coûts de location. Parmi les offres de cloud computing, on peut noter AWS, Google Cloud, Azure ou Scaleway.

Google Cloud permet à titre d’exemple d’avoir **16 vCPUs avec 64go de RAM pour 400$ par mois dédié au calcul**. Google Cloud possède aussi une offre permettant la location d’un TPU. Un TPU est beaucoup plus performant qu’un GPU mais surtout plus coûteux et pas spécialement adapté au client. Pour donner une idée du coût, il est possible d’avoir **un TPU entrée de gamme à disposition 24h/24h pour 3600$ par mois.**

Pour de la location de GPU (type RTX 3070), Scaleway propose un tarif de **0.98€/h soit environ 720€ par mois à temps plein.**

En comparaison avec Google Cloud, **Azure** permet d’avoir **16 vCPUs avec 64go de RAM pour environ 560$ par mois dédié au calcul.**

Le choix de la solution de cloud computing dépend du prix mais aussi des l’ensemble des solutions complémentaires proposées (comme les solutions de stockage et potentiellement la personnalisation plus ou moins poussée de l’offre ainsi que l’engagement; pour l’engagement certaines offres se positionne sur une année, d’autres sur plusieurs années comme 3 ans).

L’intérêt d’utiliser du cloud computing réside dans le fait qu’il est obligatoire de stocker et traiter de grandes quantités de données générées par le projet, le cloud computing répond à ce besoin. Il a également d’autres avantages comme sa scalabilité et sa performance, ce qui est essentiel pour le bon fonctionnement du système. Enfin, les offres de cloud computing assurent généralement une disponibilité continue permettant de ne pas subir des périodes de creux.

## Apprentissage continu

L'apprentissage continu peut être réalisé en utilisant différentes techniques, voici une explication de ces approches :

* **Transfer Learning** (Transfert d'apprentissage) qui vise à utiliser un modèle pré-entraîné sur des données similaires, puis à le réutiliser pour résoudre une tâche spécifique dans un domaine différent. Le modèle pré-entraîné est utilisé comme point de départ, et seules les couches finales du modèle sont ré-entraînées sur les nouvelles données spécifiques au vignoble. Cela permet de bénéficier des connaissances déjà acquises par le modèle pré-entraîné et d'accélérer le processus d'apprentissage sur les nouvelles données.
* **Online Training** (Apprentissage en ligne) qui consiste à mettre à jour le modèle en temps réel à mesure que de nouvelles données arrivent. Le modèle initial est entraîné sur un ensemble de données initiales, puis il est mis en production et utilisé pour prendre des décisions. À mesure que de nouvelles données sont collectées, elles sont utilisées pour mettre à jour le modèle existant, en utilisant des techniques d'apprentissage en ligne telles que la rétropropagation du gradient ou les algorithmes d'apprentissage incrémental. Cela permet d'adapter le modèle en continu aux nouvelles observations et d'améliorer ses performances au fil du temps.
* **Incremental Training** (Apprentissage incrémental) qui consiste à ajouter de nouvelles données au modèle existant et à le réentraîner périodiquement sur l'ensemble étendu des données accumulées. Le modèle initial est entraîné sur les données initiales, puis de nouvelles données sont collectées et ajoutées au jeu de données existant. Le modèle est ensuite ré-entraîné sur l'ensemble étendu des données, en utilisant des techniques telles que la rétropropagation du gradient ou les algorithmes d'apprentissage en ligne. Cela permet d'adapter le modèle aux nouvelles données tout en préservant les connaissances apprises précédemment.
* **Le fenêtrage,** autre technique, qui permet de repartir de zéro en entraînant un nouveau modèle en se basant uniquement sur les données les plus récentes. Dans le cas du l’activité des vignobles, cela pourrait permettre d’obtenir un modèle plus spécifique à une période en cours, surtout pour ce qui concerne la proposition d’actions correctives et d’optimisation.
* **Les feedbacks** peuvent participer à l’apprentissage continu car ils permettent directement d’ajuster un modèle d’intelligence artificielle si ses prédictions sont fausses ou trop aléatoires.

Chacune de ces approches présente des avantages et des inconvénients. Le **Transfer Learning** est **rapide et efficace** lorsqu'il existe des **modèles pré-entraînés** disponibles, mais il **peut** **ne pas être aussi adapté aux** **spécificités** du vignoble. L'**Online Training** permet d'adapter le modèle en temps réel, mais nécessite une **gestion continue des nouvelles données** et une **infrastructure** pour mettre à jour le modèle en production. L'**Incremental Training** permet **d'ajouter** de nouvelles **données** de manière **périodique**, mais nécessite de **conserver** **l'ensemble** **des données** et de **réentraîner le modèle régulièrement.**

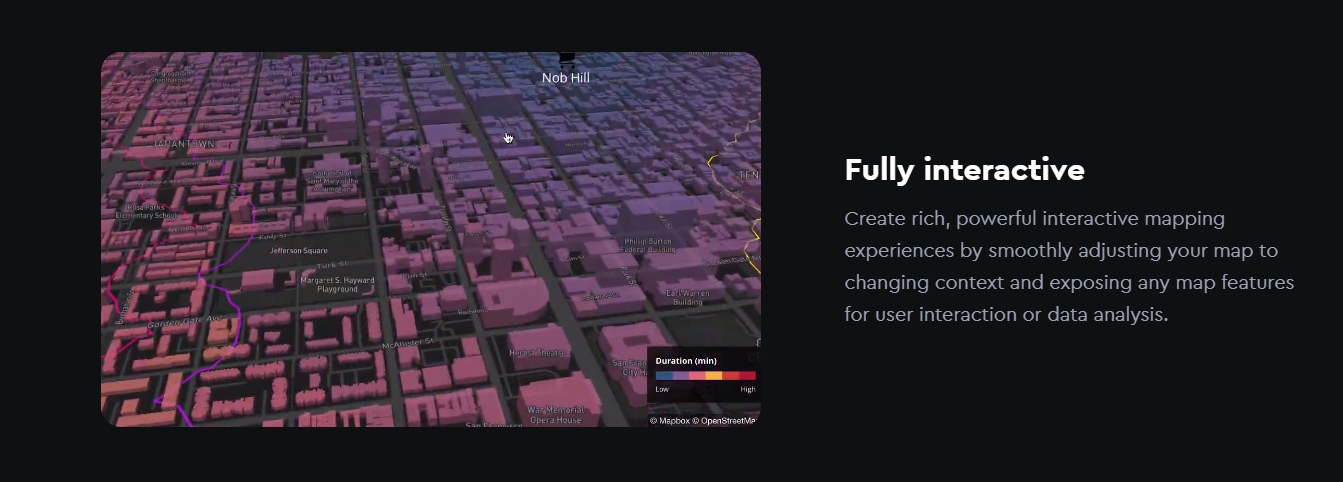
Le choix de l'approche dépendra des besoins spécifiques du client, de la disponibilité des données et des ressources informatiques, ainsi que des performances souhaitées du modèle. Une combinaison de ces approches peut également être envisagée en fonction des exigences du système.

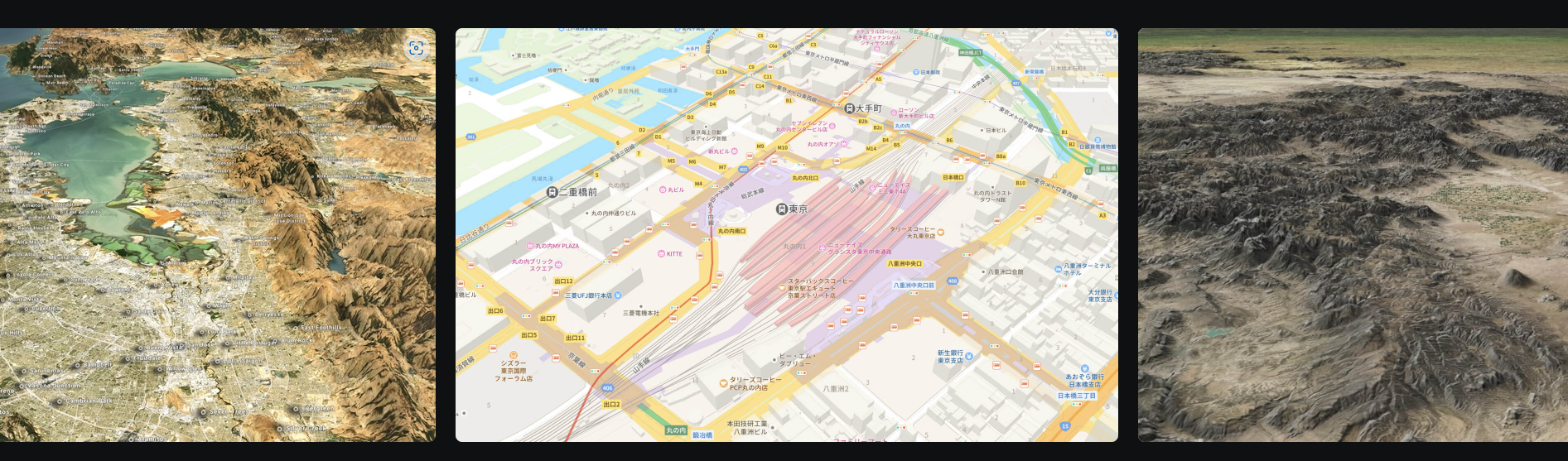
L’intérêt de mettre en place de l’apprentissage continu dans ce cas d’étude est que le client a besoin que le système conçu puisse apprendre continuellement à partir de nouvelles données. Cela permet d'ajuster les recommandations ou les alertes sur les vignobles et ses problématiques au fur et à mesure. L’apprentissage continu permet également d’intégrer de nouvelles classes dans la classification de nos modèles.

Cette mécanique est intéressante car elle permet de compléter le périmètre fonctionnel d’un modèle tout en conservant sa performance. Pour le cas des vignobles par exemple, il est possible que certaines maladies ne soient pas représentées par les données capturées sur une saison. **Il sera possible d’inclure la reconnaissance de cette nouvelle maladie la saison suivante.**

## Visualisation des prédictions

Afin d’implémenter une visualisation fiable, rapide et agréable au client lors de la détection d’une maladie dans la vigne, l’utilisation de la librairie **MapBox** est mise en avant afin d’obtenir efficacement et rapidement une carte du secteur touché par la maladie avec un trajet pour permettre à l’utilisateur de s’y rendre.





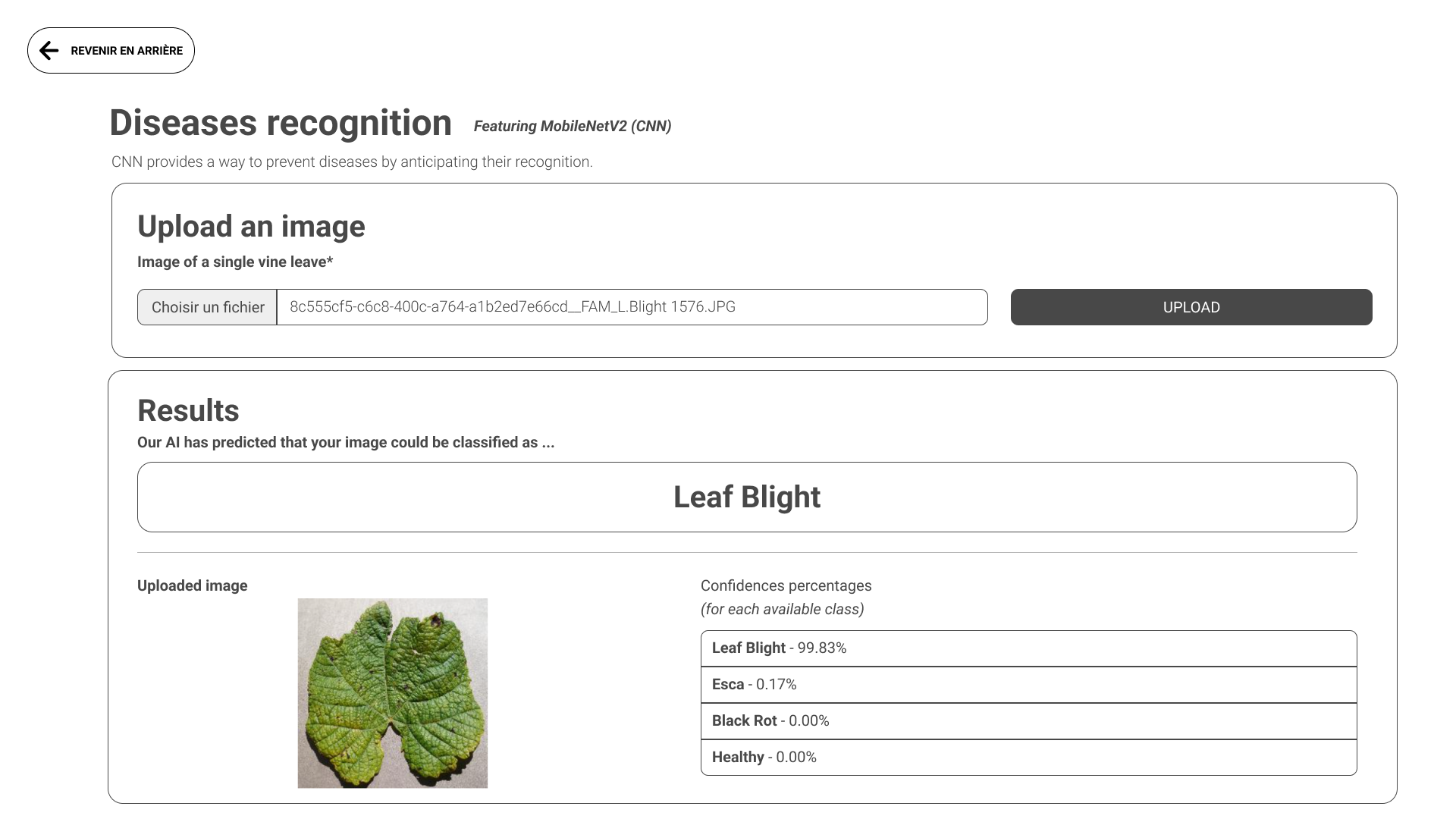
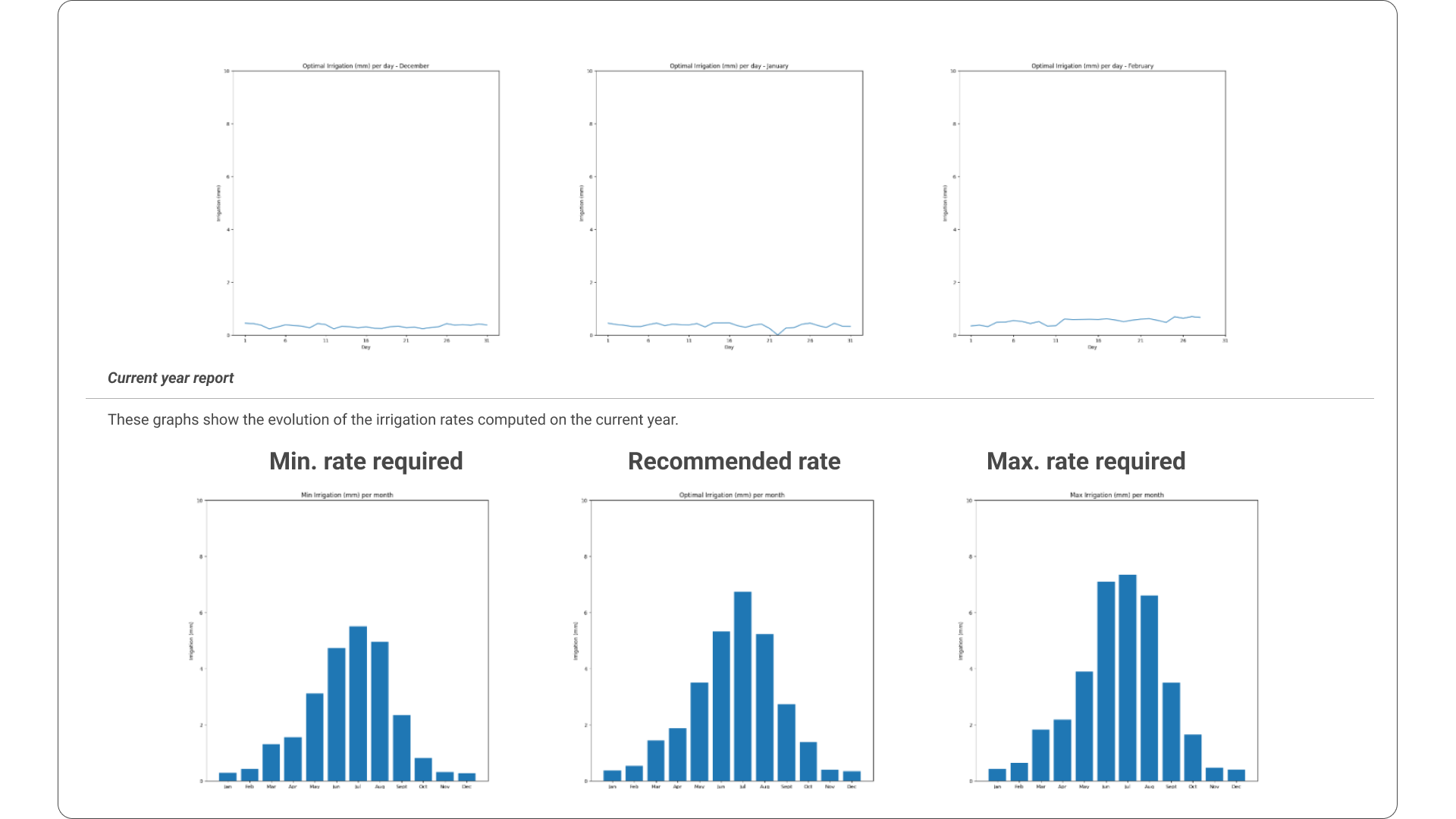
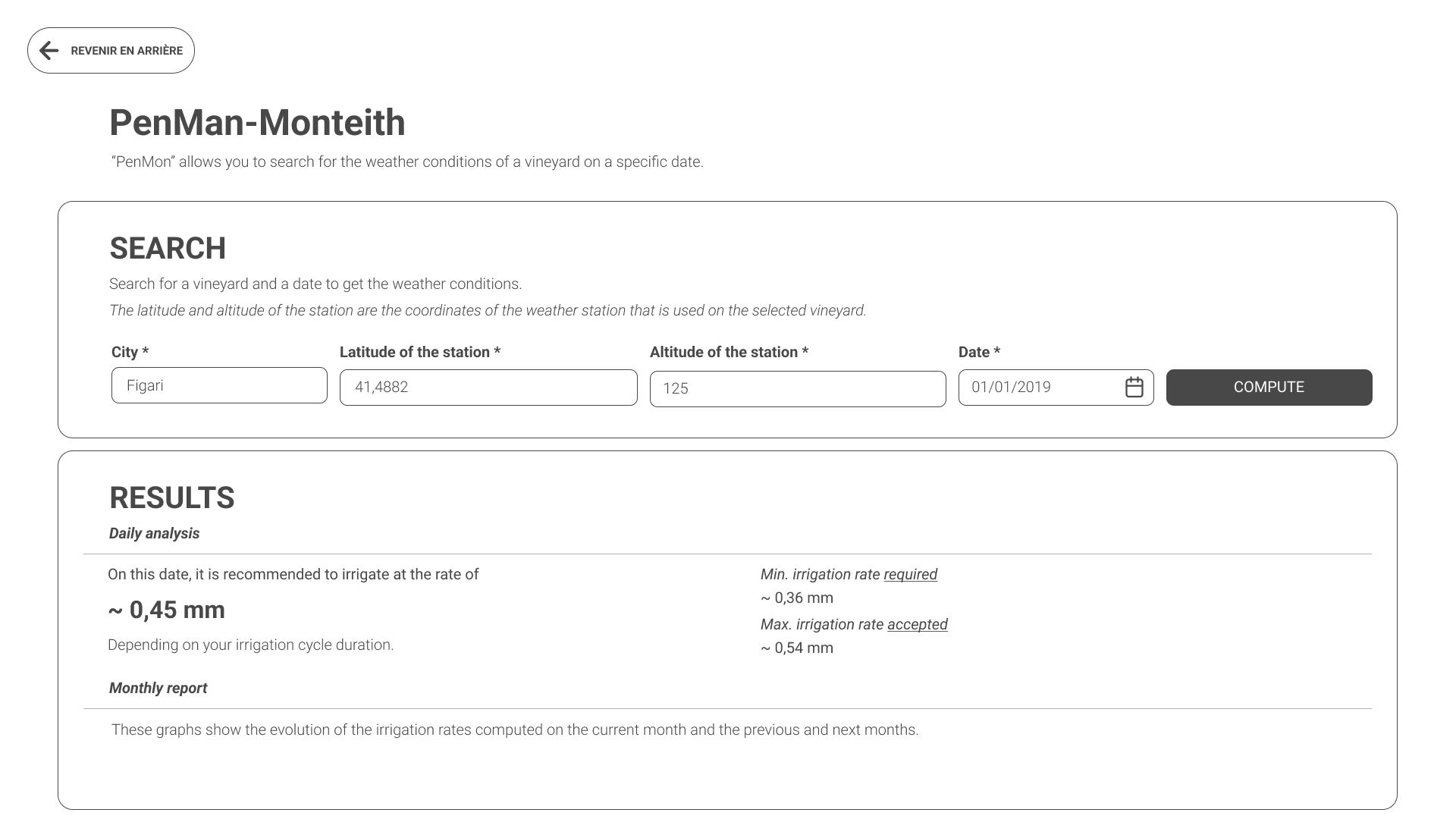
Cette interface sera accessible depuis un poste du client et indiquera à l’aide d’une carte interactive la zone de détection de la maladie / parasite avec une image de la zone concernée pour que le client puisse vérifier que l’IA ne fasse pas d’erreur.

L’interface est divisée en plusieurs onglets de façon à administrer au maximum les différentes fonctionnalités.

Tout d’abord, lorsque l’utilisateur accède à l'application, il arrive sur un menu qui lui permet de sélectionner les infos qu’il souhaite voir. Ensuite, 3 choix lui sont disponibles, le premier permet à l’utilisateur d’obtenir, en fonction du vignoble sélectionné, des suggestions liées à l’irrigation dans la zone.

Le second, conduit à une carte utilisant la géolocalisation pour diriger l’utilisateur à l’endroit où une potentielle vigne malade ou parasitée pourrait se trouver ainsi que des boutons pour valider ou non si la vigne est bien malade pour améliorer le modèle de l’IA.

Enfin, le 3ème onglet contient la gestion des images, le client ajoute manuellement une image qui va être traitée par l’IA et, cette même IA, détectera si il y a maladie ou non.



## 

## 

## 

## 

## 

## 

## Feedback

Bien qu’une intelligence artificielle soit fiable, elle ne l’est jamais à 100%. Afin de tendre au maximum vers la fiabilité maximale, une interface de feedback est un élément incontournable. Grâce à cette interface, il est possible de mettre en place de l’apprentissage continu et le modèle sera constamment ré-entraîné sur de nouvelles données, ce qui va augmenter considérablement sa précision au cours des années.

#### Mise en place

Lorsque de nouvelles données seront ajoutées, elles seront dans un premier temps prédits par le modèle puis soumises à vérification par un vigneron.

Une interface sera disponible pour que les vignerons puissent vérifier les nouvelles données. L’interface sera faite pour que le vigneron puisse choisir les prédictions qu’il souhaite vérifier, comme par exemple le *black-rot* (une maladie). Lorsqu’il va choisir de vérifier cette classe, les nouvelles données qui ont obtenu cette prédiction vont alors s’afficher et il devra les valider si les prédictions sont vraies, ou dire à quelle classe elles appartiennent si la prédiction est fausse.

Une fois les vérifications faites, le modèle est ré-entraîné avec les nouvelles données et les données corrigées puis mis à jour.

#### Risques

Un risque auquel il est possible d’être confronté avec une interface de feedback de ce genre est une erreur de classification de la part de l’utilisateur. **Pour cela, l’interface sera conçue pour être la plus intuitive et la plus épurée possible afin de réduire le risque d’erreur, même si le risque n’est pas complètement effaçable.**

## Annotation

En plus de la fonctionnalité de feedback dans l'interface, un système d'annotation permettra aux vignerons de classer les photos en fonction de leur état (malade ou sain) en amont de leur utilisation dans le processus d’apprentissage continu. Cela fournira un jeu de données prétraitées pour l'IA, qui aura déjà les informations de classe associées aux images.

Pour réaliser cette fonctionnalité, l’outil **OpenSeadragon**, une bibliothèque JavaScript open-source et polyvalente conçue spécifiquement pour la visualisation et la navigation d'images gigapixels, sera intégrée au projet. OpenSeadragon offre une solution flexible et multiplateforme pour zoomer et naviguer en douceur à travers des images à très haute résolution. Grâce à ses fonctionnalités avancées, telles que la compression, la mise en cache et le rendu progressif, OpenSeadragon permet une expérience de visualisation fluide même avec des images extrêmement détaillées. Cette bibliothèque offre également une personnalisation étendue de l'interface utilisateur et la possibilité d'ajouter des fonctionnalités interactives supplémentaires, ce qui en fait un choix polyvalent pour notre projet.

*Exemple de visualisation d’image:* [*https://openseadragon.github.io/#examples-and-features*](https://openseadragon.github.io/#examples-and-features)

L'interface d'annotation sera soigneusement conçue et optimisée pour être ergonomique et accessible, afin de faciliter le processus d'annotation pour les vignerons.

## Interface et API

Enfin, dans le but d’assembler l’ensemble des composantes de la solution une API ainsi qu’une interface seront développées.

L’interface présentera les résultats des prédictions effectuées sur les nouvelles données capturées ainsi que les calculs effectuées. Elle présentera aussi une IHM permettant d’annoter ou d’effectuer un retour sur l’annotation apportée par l’IA afin d’améliorer le processus d’apprentissage continu de l’IA d’analyse d’images de feuilles de vignes.

L’API servira à faire le lien entre le modèle d’analyse d’images de feuilles de vignes, le script d’intégration de l’équation Penman-Monteith et l’interface.

L’interface sera réalisée à l’aide d’un *framework web* **(Angular)**et fera le lien avec **OpenSeadragon** et **Mapbox.** Quant à elle, l’API sera conçue à l’aide de **Flask,** un *framework web* d’application serveur léger et adapté au besoin de notre système.

# 

# Articles pertinents

## IA dans le domaine de la viticulture

* **Utilité des images prises dans le ciel par drone pour détecter des maladies sur un vignoble :**

<https://www.vitisphere.com/actualite-86425-la-flavescence-doree-detectee-par-drone-avant-2021-.html>

* **Drones utilisés pour surveiller un vignoble et détecter les maladies :**

<https://www.rtl.fr/actu/sciences-tech/chouette-le-drone-qui-permet-aux-viticulteurs-de-surveiller-leurs-vignes-7793482289>

<https://www.chouette.vision>

* **Application des techniques d’imagerie pour protéger les vignes :**

<https://youtu.be/ixiEJEsrIaU>

## Maladies de la vigne

* **Guide des maladies de la vigne :**

[Guide d'identification des principales maladies de la vigne - agriculture.canada.ca](https://agriculture.canada.ca/fr/production-agricole/lutte-antiparasitaire-agriculture/ressources-matiere-lutte-antiparasitaire-agriculture/guide-didentification-principales-maladies-vigne#blanc)

[Maladies de la vigne : mildiou, oïdium, black-rot, excoriose, botrytis et maladies du bois (basf.fr)](https://www.agro.basf.fr/fr/cultures/vigne/maladies_de_la_vigne/)

* **Guide des Ravageurs de la vigne :**

[Ravageurs de la vigne : cochylis, cicadelle, eudemis - BASF](https://www.agro.basf.fr/fr/cultures/vigne/ravageurs_de_la_vigne/)

[Ravageurs de la vigne : Bayer-Agri, conseils pour la protection de vos cultures](https://www.bayer-agri.fr/cultures/ravageurs-de-la-vigne/)

* **Cycle de la vigne :**

[Calendrier et cycle de la vigne (viticabrol.fr)](https://viticabrol.fr/content/18-calendrier-cycle-de-la-vigne)

## Données et POC

* **Jeu de données de feuille de vignes :**

<https://www.kaggle.com/datasets/muratkokludataset/grapevine-leaves-image-dataset>

<https://www.sciencedirect.com/science/article/abs/pii/S0263224121013142?via%3Dihub>

* **Notebook sur ce jeu de données :**

<https://www.kaggle.com/code/andreamontemurro/gravepine-disease-classification>

* **Jeu de données de feuille de vignes malades :**

<https://www.kaggle.com/datasets/piyushmishra1999/plantvillage-grape>

* **Jeu de données météorologiques :**

https://www.kaggle.com/datasets/juldel/meteo-data-france-parsing-python

Visualisation et interface

* **Utilisation de MapBox GL JS :**

[Create web maps with GL JS | Mapbox](https://www.mapbox.com/mapbox-gljs)

# Tableau récapitulatif parasites / maladies de la vigne

| **Maladie** | **Quand** | **Conditions propices** | **Photo** |
| --- | --- | --- | --- |
| Mildiou | début mai à mi-septembre avec une forte croissance de juin à fin août | vignobles humide et  ombragés |  |
| période de pluie,  forte rosé et brouillard |
| 22° à 28° |
| Blanc ou Oïdium | début mai à mi-septembre avec une forte croissance de mi-juin à fin août | tous les vignobles |  |
| période nuageuse |
| 20° à 28° |
| Anthracnose ou anthracnose maculée | début mai à fin septembre avec une forte croissance de début juin à mi-septembre | tous les vignobles |  |
| période de pluie, forte pluie, orages et vent violents |
| 20° à 26° |
| Pourriture grise | fin-mai à mi-juillet et mi-août à fin septembre avec une forte croissance de juin à mi-juillet et en septembre | tous les vignobles |  |
| période de pluie, forte pluie et vent violents |
| 16° à 26° |
| Pourriture Noire | mi-juin à fin août avec une forte croissance de fin juin à fin juillet | tous les vignobles |  |
| période de pluie, forte pluie et vent violents |
| 20° à 28° |

| **Parasite** | **Quand** | **Maturité de la vigne** | **Photo** |
| --- | --- | --- | --- |
| Cicadelle de la flavescence | début juillet à mi-septembre | Toute la période de maturité à partir des Boutons floraux |  |
| Cochylis | mi-avril à mi-mai et fin juin à fin juillet | Boutons floraux et Fermeture de grappe |  |
| Eudemis | mi-avril à mi-mai et fin juin à fin juillet | Boutons Floraux et Nouaison petit pois à Maturité |  |
| Thrips de la vigne | avril à octobre |  |  |
| Pyrale de la vigne | mars à juillet |  |  |
| Metcalfa de la vigne | mi-avril à fin septembre |  |  |
| Mange-bourgeon de la vigne | avril à début novembre |  |  |

