



RAPPORT DE L'APPLICATION DE LA DÉMARCHE SCIENTIFIQUE
STUDIES AND RESEARCH IN MACHINE LEARNING AND ELECTRONICS

**Modèle prédictif appliqué à un cadre de
dépollution métallique des sols et utilisant
des systèmes de détection inductifs**

AUTHOR : ADRIEN NICOLAS
PROMOTION : A4 COMPUTER
SCIENCE

Tutor : MARTINEZ Sinuhe

1 Remerciement

Je tiens à remercier mes tuteurs du CESI pour leur soutien et leur aide précieuse dans la validation de mon sujet de recherche. Leur expertise et leur conseil ont été précieux pour la direction que j'ai donnée à mon travail.

Je tiens également à remercier mes camarades de classe pour leur soutien moral au cours de ce projet de recherche. Leur motivation et leur enthousiasme ont été une source d'inspiration pour moi, et leur aide m'a été précieuse pour maintenir mon engagement envers ce projet.

Je suis reconnaissant envers tous ceux qui ont contribué de près ou de loin à la réussite de ce travail de recherche. Leur soutien a été inestimable et j'espère pouvoir leur rendre la pareille à l'avenir.

2 Résumé

Dans cette étude, nous avons utilisé un modèle prédictif pour améliorer les processus d'estimation de pollution d'un sol pour prioriser la dépollution métallique de celui-ci. Nous avons entraîné le modèle sur des données simulées, et réelles pour que le comparatif soit le plus représentatif possible. Les résultats ont montré que le modèle prédictif était efficace pour prédire les taux de contamination des sols en fonction de certaines catégories précises de métaux, et que l'utilisation de systèmes de détection inductifs permettait d'obtenir des prédictions aussi précises.

Nous avons conclu que notre approche pourrait être utilisée pour améliorer les processus de dépollution métallique des sols en fournissant une estimation précise des taux de contamination avant de commencer le processus de dépollution.

Nous avons ainsi exploré les moyens d'optimiser de l'étude d'un sol dans un contexte écologique afin de fournir une estimation précise de la pollution d'un terrain. Dans ce rapport, vous verrez les différentes techniques d'identification et de prédiction utilisées et celles privilégiées et adaptées pour notre étude.

In this study, we used a predictive model to improve the process of estimating soil pollution and prioritize metal remediation. We trained the model on simulated and real data to make the comparison as representative as possible. The results showed that the predictive model was effective at predicting contamination levels of soils based on specific categories of metals, and that the use of inductive detection systems allowed for similarly precise predictions.

We concluded that our approach could be used to improve metal remediation processes by providing a precise estimate of contamination levels before beginning the remediation process.

We have thus explored ways to optimize the study of soil in an ecological context to provide a precise estimate of soil pollution. In this report, you will see the different identification and prediction techniques used and the ones favored and adapted for our study.

Table des matières

1	Remerciement	1
2	Résumé	2
3	Introduction	6
4	Impact métallique	7
4.1	Dechets generaux	7
4.2	Dechets guerres mondiales	8
5	Systeme de detection inductif	10
5.1	Principe de fonctionnement	10
5.2	Details techniques	11
6	Etude du signal	12
6.1	Explication generale	12
6.2	Cas rapporté à l'étude	12
6.3	Concept algorithmique d'analyse du signal	13
6.3.1	Interpretation et prediction	13
7	Etat de l'art des moyens d'analyses des signaux	15
7.1	Introduction	15
7.2	Contexte historique	15
7.3	Etat actuel des connaissances	16
7.4	Themes émergents	17
7.5	Conclusion de l'etat de l'art	18
7.6	Etudes relatives	18
8	Composants types d'un SDI	19
8.1	Inductance	19
8.2	NE555 Timer	19
8.3	Circuit d'un Systeme de dectection inductif	20
8.4	Étude des signaux	21
8.4.1	Inductance de 150 mH	21
8.4.2	Inductance de 10 mH	22
8.4.3	Corrélation entre l'inductance et la fréquence	23

9	Données	24
9.1	Explication generale	24
9.1.1	Augmentation du dataset	25
9.2	Possibilités de Dataset	26
9.3	Preparation des données	27
9.3.1	Pourquoi traiter les données	27
9.3.2	Explication du fonctionnement de la pipeline	28
9.3.3	Résultat de la pipeline de préparation	29
10	Moyens de prédiction	30
10.1	Generalités des predictions	30
10.1.1	Algorithmes de regression	31
10.1.2	Algorithme de classification	36
10.1.3	Simulation d'experience	38
11	Conclusion	40
12	Glossaire	41

Table des figures

1	Tableau des elements les plus toxiques pour la santé	8
2	Schématique d'un composant NE555 Timer	19
3	Circuit électronique simulé d'un SDI	20
4	Signal généré par le circuit simulé avec une valeur d'inductance de 150 mH	21
5	Signal généré par le circuit simulé avec une valeur d'inductance de 10 mH	22
6	Graphique de corrélation entre la fréquence et l'inductance	23
7	Premier jeu de données	24
8	Nouveau jeu de données	24
9	Densité des jeux de données au fur et à mesure de la repopulation du dataset	25
10	Répartition du dataset	28
11	Dataset mis à l'échelle	29
12	Comparaison des modèles de regression sur un petit dataset	31
13	Comparaison des modèles de regression sur un gros dataset	32
14	dataset et cluster predit	33
15	Algorithme de clustering Kmeans sur un dataset d'inductances par rapport à des frequences	33
16	Algorithme de clustering Kmeans sur un dataset d'inductances par rapport à des frequences	34
17	Application de Random Forest afin de prédire un cluster spécifique en prenant en compte une inductance et une fréquence.	35
18	Matrice de confusion de la Logistic regression	36
19	Matrice de confusion de la Random Forest	36
20	Modèle de random forest permettant de prédire 4 groupes de permeabilité	37
21	Distribution des clusters de permeabilité à travers les datasets	38

3 Introduction

Aujourd'hui avec la crise climatique, la dépollution intervient sur de nombreux terrains comme les sols, nappes phréatiques, sédiments et eaux de surface. Ainsi, l'étude du sol pour un contexte d'écologie peut nous permettre une estimation de pollution sur une surface et un terrain donné. Une fois cette estimation faite, il serait donc possible de prioriser la dépollution d'endroits plus ou moins sensibles. La pollution du sol à caractère métallique pose un problème particulier, car les métaux ne sont aucunement biodégradables. Cela signifie que si aucune action n'est faite dans les sols, un phénomène appelé bio-accumulation pourrait apparaître dans nos champs agricoles, amenant dans le futur une forte toxicité sur les espèces récoltées.

L'enjeu de cette étude est de proposer une optimisation d'analyse de la composition d'un sol de ses objets métalliques pour découler d'une estimation d'un taux de pollution permettant de prioriser les endroits à dépolluer. En effet, la présence de polluants métalliques, notamment des métaux lourds dans le sol pose des problèmes de toxicité importants, tout en sachant que ces polluants peuvent migrer (sous l'effet de l'écoulement des eaux, de la manipulation de la terre, de plantations, de l'acidification du milieu...). Il est donc nécessaire de cibler la catégorie métallique la plus impactante pour prioriser les lieux de dépollution. En prenant l'exemple des métaux lourds, ceux-là peuvent être dangereux pour les sols en raison de leur toxicité. Ils peuvent contaminer les sols et affecter la croissance des plantes et la santé des organismes vivant dans le sol. Les métaux lourds les plus couramment associés à la contamination des sols sont le plomb, le mercure, le cadmium et le chrome. L'exposition à ces métaux peut entraîner des problèmes de santé pour les êtres humains et les animaux, tels que des troubles neurologiques, des dommages aux reins et au foie, et des problèmes de reproduction. Il est important de gérer correctement les déchets contenant des métaux lourds pour prévenir la contamination des sols.

Cette optimisation d'étude du sol donc, passe par l'acquisition des signaux de l'outil de détection des métaux (système de Détection Inductif), et l'étude des données de celui-ci. Un Système de Détection Inductif (appelé aussi détecteur de métaux) peut être utilisé pour participer à la dépollution d'un sol grâce à ses capacités de détections et discriminations de métaux. Les systèmes de détection inductifs sont des outils de mesure non destructifs qui permettent de détecter la présence de métaux enfouis dans le sol sans creuser. Ils fonctionnent en envoyant un champ électromagnétique dans le sol et en mesurant les réponses électromagnétiques pour détecter la présence de métaux. L'étude pourra se porter sur plusieurs paramètres influant, que ce soit à propos de l'objet, ou encore des paramètres du sol. Un enjeu à traiter en second plan est la possibilité d'identifier des objets métalliques à caractère explosifs enfouis, nocifs pour l'environnement et rentrant en compte dans l'étude d'une estimation de pollution d'un terrain.

4 Impact métallique

4.1 Dechets generaux

D'une manière générale, les métaux les plus connus de tous (cuivre, plomb, nickel, etc.) sont naturellement présents dans nos sols à plus ou moins fortes doses. En bref, les rejets de l'industrie, des ménages, des transports, de l'agriculture ou simplement de la pollution et de la présence de l'homme contribuent à la contamination diffuse de métaux dans nos sols. Chaque métal est souvent lié à des activités spécifiques, comme le plomb qui est principalement relié aux activités industrielles dans le nord et le nord-est autour de Paris, par exemple.

L'ensemble des métaux est évidemment dangereux pour l'environnement s'ils sont concentrés à forte dose. Mais certains le sont plus que d'autres. La famille de métaux la plus dangereuse est celle des métaux appelés "lourds", non pas pour leur poids, mais pour la masse volumique des matériaux.

La contamination des sites et des sols par les métaux lourds pose un risque important pour la santé publique car ils peuvent provoquer diverses affections médicales. L'exposition aux métaux lourds peut se produire de deux façons : inhalation de particules et ingestion directe (de la poussière) ou via les aliments contaminés. Dans les deux cas, le sol est un vecteur important d'infection. Pour cette raison, il est essentiel d'avoir des moyens de mesurer les niveaux de métaux lourds dans le sol.

La pollution des sols passe à travers la végétation, mais aussi à cause de l'infiltration des nappes phréatiques par lessivage. Cela a des impacts potentiels et relatifs sur la santé humaine et constitue donc un sujet de préoccupation majeur à des échelles de court et long terme. La véritable prise de conscience des risques écologiques liés aux sols pollués est apparue réellement tardivement, il y a de cela 40 ans. Durant cette période, le développement industriel fut important, et la production de masse cause des dommages importants sur le plan écologique. Aujourd'hui, l'environnement est susceptible d'être affecté par ce problème ; il est donc important d'identifier les sources de pollution, mais également de savoir évaluer les niveaux de pollution écologique d'un sol, servant ainsi de prévention.

La teneur en métaux du sol peut être d'origine naturelle ou anthropique, c'est-à-dire causée par les activités humaines. L'apport en métaux se fait par des voies différentes dans les deux cas. Sans prendre en compte l'intervention de l'homme, les sols contiennent déjà une concentration en métaux (certes faibles, mais à prendre en compte). Pour ce qui est de l'intervention humaine, les sols sont contaminés de plusieurs manières :

- **Les contaminations diffuses** : Par le biais de la voie atmosphérique qui peut se caractériser sous forme de poussière, se déposant sur les végétaux, sur les eaux de surface.
- **Les contaminations ponctuelles** : Par le biais des usines, d'une mines, d'engrais, traitements phytosanitaires etc.. Ce genre de contaminations fait intervenir un grand nombre de polluant concentrés sur une surface bien délimitée. Les métaux sont aussi caractérisés sous forme de poussières.
- **Les contaminations résiduelles** : Les métaux causant cette contamination sont sous une forme bien plus importantes que les contaminations précédentes. Cette contamination a été causée par l'humain en jetant, laissant tomber depuis des années des déchets métalliques. Ces déchets métalliques peuvent être le résumé des guerres passées, des déchets venant de toute part et ne se dégradant pas depuis des centaines d'années. Ce genre de contamination est d'autant plus inquiétant, car les déchets laissés par la guerre sont dangereux écologiquement avec les types de matériaux utilisés, et le nombre importants enfouis.

On compte en France une moyenne entre 3700 et 3900 sites pollués selon le ministère de l'Écologie et du Développement Durable. Parmi tous ces sites, la moitié est définie comme ne présentant pas de risque important pour l'environnement et ne nécessitant donc pas de travaux de dépollution

H																		He
Li	Be												B	C	N	O	F	Ne
Na	Mg												Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr	
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe	
Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn	

éléments majeurs

Mo

oligo-éléments

Na

macro-éléments

éléments toxiqueséléments peu toxiques

FIGURE 1 – Tableau des éléments les plus toxiques pour la santé

prioritaires. L'autre moitié est identifiée comme dangereuse pour l'environnement et nécessitant des travaux de dépollution prioritaires.

Concrètement, d'après le tableau périodique ci-dessous, on remarque que les éléments les plus toxiques et donc les plus polluants sont les métaux lourds tels que le plomb (Pb). Dans notre cadre d'étude, visant à démontrer qu'il est possible d'estimer la pollution d'un sol grâce à un système de détection inductif (que nous appellerons SDI[1]), ce type de matériau nous intéresse particulièrement, car il est plus susceptible d'être présent sous une forme détectable. Enfin, on remarque que le cuivre et le zinc sont moins toxiques et donc moins polluants. Nous les prendrons en compte dans notre étude car, tout comme le plomb, ils sont plus facilement détectables par le SDI en raison de leur concentration plus élevée.

4.2 Déchets guerres mondiales

Longtemps après la guerre, les zones qui ont été exposées aux bombardements continuent encore aujourd'hui de contenir des munitions, des obus et toutes sortes d'explosifs actifs. S'ils ne sont pas récupérés, ils finissent par se dégrader dans les sols et les contaminer par la suite. En France, les déchets de la Première et de la Seconde Guerre mondiale continuent encore de polluer, comme l'a rappelé l'association Robin des Bois à l'occasion de la célébration des deux armistices. Les obus et explosifs contenus dans le sol contiennent des produits toxiques pour l'environnement et pour l'être humain, tels que l'arsenic, le mercure et les sels de perchlorates. Rappelons que ces trois éléments sont répertoriés comme dangereux d'après le tableau périodique précédent. Les régions les plus touchées en France sont la Franche-Comté, l'Alsace, la Lorraine, la Champagne-Ardenne, l'Île-de-France, la Picardie et le Nord-Pas-de-Calais. Environ 500 à 800 tonnes de munitions sont collectées chaque année sur le territoire. Il est donc nécessaire de prioriser rapidement la dépollution des sols dans ces régions touchées par la pollution de la guerre.

À l'issue du conflit de la Première Guerre mondiale, l'urgence immédiate est la remise en état des sols agricoles afin de rétablir rapidement la production. La remise en état consiste souvent à simplement reboucher ce qui a été creusé, en oubliant ainsi les nombreux stocks et munitions souterrains. En effet, on compte 1,7 million de tonnes de projectiles totalement abandonnés, ainsi que 4 millions de projectiles chimiques. La grande problématique est donc de savoir comment localiser le plus rapidement possible ces mines environnementales afin de prioriser la dépollution de ces engins.

Le bilan général de l'impact des métaux dans le sol est le suivant : les métaux caractérisés comme lourds sont ainsi les plus toxiques pour l'environnement. Il faut également prendre en compte que

les autres types de métaux sont écologiquement incompatibles, mais beaucoup moins impactants que les métaux lourds. De plus, en plus des déchets généraux précédemment cités, il faut prendre en compte, dans certaines régions de la France, la préoccupation des types de déchets dangereux tels que les explosifs, qui sont préoccupants tant du point de vue écologique que de la sécurité. La priorité ici est donc de détecter les traces de métaux lourds présentes en quantité suffisante pour le SDI, ainsi que de prioriser la localisation des engins explosifs parmi plusieurs paramètres à étudier.

5 Systeme de detection inductif

5.1 Principe de fonctionnement

Un détecteur de métaux (ou système de détection inductif) est un appareil permettant de localiser des objets métalliques en exploitant le phénomène physique de l'induction magnétique

Le domaine de l'électromagnétisme est une combinaison de l'électricité et du magnétisme. Dans notre vie de tous les jours, pour créer de l'électricité, un générateur fait tourner un tambour de fil de cuivre dans un champ appelé "magnétique". Cela résulte d'un courant électrique à l'intérieur du fil, on a donc créé de l'électricité grâce au champ magnétique mais la relation fonctionne également à l'inverse, et c'est comme cela que les moteurs électriques fonctionnent. Finalement, on remarque que l'électricité peut créer du magnétisme, et le magnétisme peut produire de l'électricité, d'où le terme électromagnétisme. Un champ magnétique variable entraîne un champ électrique variable. Il en va de même pour l'électricité, ceci est un concept important à comprendre.

un système de détection inductif (SDI) possède deux bobines, une bobine émettrice et une bobine de réception. Celles-ci se situent à l'avant du système. Le rôle de la bobine émettrice est de transmettre un champ magnétique autour du SDI. C'est en alimentant cette bobine par un courant électrique alternatif qu'elle va générer ce fameux champ magnétique.

La fréquence de fonctionnement d'un détecteur est un concept directement lié à ce champ magnétique. La fréquence de fonctionnement est variable selon les besoins de détection. Celle-ci permettra une discrimination différente et correspond à la fréquence à laquelle le sens du courant est alterné. Il y a ainsi 5 types de fréquence de fonctionnement connus chez les SDI que l'on précisera plus tard dans l'étude :

- Les détecteurs basse fréquence (4-8kHz)
 - Meilleur rejet du fer
 - Détecteurs plus stables en terrain pollué et plus sélectif
 - Profondeur accrue sur les cibles très conductrices
- Les détecteurs haute fréquence (+ de 18kHz)
 - Détecteurs plus performants en terrain minéralisé
 - Profondeur accrue sur les cibles faiblement conductrices
- Les détecteurs moyenne fréquence (10-14kHz)
 - Performances exceptionnelles sur les petites cibles peu conductrices
 - Rejetant le fer parfaitement
- Les détecteurs Multifréquence simultanée (MF)
 - Efficace sur tout type de terrain
 - Meilleure identification
 - Meilleure performance

Le gros point important des champs magnétiques est qu'ils ne sont pas affectés par les sols et obstacles à travers lesquels ils passent. Ainsi, si le SDI est suffisamment proche d'un objet métallique, le champ magnétique va agir sur les atomes de métal provoquant ainsi le déplacement électrique, et comme dit précédemment, cela va effectivement créer un champ électrique variable autour de cet objet en métal. En d'autres termes, cela va créer un nouveau champ magnétique autour de l'objet. C'est ici que la bobine réceptrice agit. Lorsque celle-ci va se déplacer dans le nouveau champ magnétique entourant l'objet, un courant électrique est généré dans la bobine, c'est donc ici que l'on va récupérer le signal entrant.

5.2 Détails techniques

Nous avons dit précédemment qu'un SDi était équipé de bobines émettrices et réceptrices. Voyons ainsi comment se comporte techniquement l'appareil : La présence d'un objet métallique à proximité d'une bobine modifiera son inductance. Selon le type de métal, l'inductance peut augmenter ou diminuer. Les métaux non magnétiques tels que le cuivre et l'aluminium à proximité d'une bobine réduisent l'inductance, car un champ magnétique changeant induit des courants de Foucault dans l'objet qui réduisent l'intensité du champ magnétique local. Les matériaux ferromagnétiques, tels que le fer, à proximité d'une bobine augmentent son inductance car les champs magnétiques induits s'alignent sur le champ magnétique externe. La mesure de l'inductance d'une bobine peut ainsi révéler la présence de métaux à proximité.

L'inductance d'une bobine dépend du champ magnétique que cette bobine génère autour d'elle (champ magnétique dont les variations vont ensuite créer la force électromotrice dans les spires de la bobine) mais ce champ magnétique peut justement varier en fonction des propriétés physiques de cet espace "autour". En l'occurrence, ce que l'on modifie, dans cet espace, c'est sa perméabilité μ qui a une valeur différente à l'endroit où on aura placé le "métal". Mais en alternatif il y a aussi les courants de Foucault induits dans le métal qui changent le champ magnétique et l'inductance. C'est comme cela que les portiques de sécurité détectent d'autres métaux que les ferromagnétiques. C'est aussi de cette façon que certains détecteurs de métaux fonctionnent. La self est le petit nom de la "self-inductance" ou inductance propre. Par opposition avec l'inductance mutuelle entre deux inductances (ne vous en préoccupez pas pour l'instant).

On mesure une inductance en mesurant son impédance ou le module de son impédance : on lui applique une tension alternative et on mesure le courant qui la traverse. Les amplitudes de la tension et du courant sont reliées par : $I = V/(L)$ où L est l'inductance (la self) et ω est la pulsation $2\pi f$. Il y a bien d'autres méthodes mais il faut connaître d'autres choses.

Lorsqu'un objet approche d'une self, beaucoup de paramètres de celle-ci varient : son inductance bien sûr, dans le cas d'un objet magnétique ou conducteur, mais également également ses pertes en valeur absolue, la façon dont elles varient en fonction de la fréquence, et dans le cas de corps diélectriques, leur influence sur l'inductance en fonction de la fréquence, plus les effets d'ordre supérieurs, si ce diélectrique a lui-même des pertes.

Et il est en effet tout à fait judicieux de considérer le couple self/cible (pour être général), comme des inductances couplées ayant généralement un k faible, et dont le secondaire est chargé par une impédance complexe.

6 Etude du signal

6.1 Explication generale

L'étude du signal est un domaine clé de la recherche en ingénierie et en physique. Elle vise à comprendre et à analyser les signaux électriques, électroniques et acoustiques qui sont produits par des systèmes physiques et électroniques. Les signaux peuvent être continus ou discrets, analogiques ou numériques, et peuvent être utilisés pour transmettre des informations dans de nombreux domaines, tels que la communication, la détection et la mesure.

La mesure et l'analyse d'un signal est un processus crucial pour comprendre les caractéristiques et les propriétés d'un signal. Il existe plusieurs étapes à suivre pour mesurer et analyser un signal de manière efficace.

- **Préparation de l'expérience** : Il est important de définir clairement les objectifs de l'analyse et de sélectionner les outils et les technologies appropriés pour réaliser les mesures. Il est également important de définir un protocole expérimental rigoureux pour garantir la qualité et la fiabilité des résultats obtenus.
- **Acquisition des données** : Il est important de réaliser des mesures précises et fiables du signal en utilisant des outils tels que des générateurs de signaux, des oscilloscopes et des analyseurs de spectre. Il est également important de considérer les effets de l'environnement sur le signal, tels que les perturbations électromagnétiques et les variations de température
- **Traitement des données** : Il est important de traiter les données obtenues pour en extraire les informations pertinentes. Cela peut inclure des techniques telles que la transformée de Fourier, la décomposition en ondelettes, la corrélation, etc.
- **Analyse des résultats** : Il est important d'interpréter les résultats obtenus en utilisant les concepts théoriques appropriés. Cela peut inclure des calculs tels que la fréquence, le rapport signal sur bruit, la distorsion, la stabilité, etc. Il est également important de tracer les résultats pour une meilleure visualisation.
- **Conclusion** : Il est important de déduire une conclusion des résultats obtenus en comparant les résultats avec les objectifs de l'analyse et les normes et standards pertinents. Il est également important de discuter des limites de l'analyse et de faire des recommandations pour des travaux futurs.

Il est important de noter que l'analyse d'un signal peut être très complexe et nécessiter des connaissances approfondies en mathématiques, en physique et en ingénierie électronique. La qualité de l'étude qui va suivre va donc dépendre d'un niveau d'analyse et d'interprétation d'un eleve étudiant de niveau 2ème année d'ecole d'ingenieurs.

6.2 Cas rapporté à l'étude

L'étude d'un signal dans le cadre d'un système de détection inductif est un domaine de recherche important en ingénierie électronique. L'analyse d'un signal dans ce contexte permet de comprendre les caractéristiques des signaux produits par le système et de déterminer les paramètres pertinents tel que la fréquence.

Il est important de prendre en compte les différents paramètres qui peuvent influencer la qualité et la fiabilité des résultats obtenus, tels que les propriétés électriques et les dimensions des conducteurs détectés, les conditions environnementales, les perturbations électromagnétiques et les variations de température. Il est également important de considérer les limites du système de détection inductif lui-même, telles que sa sensibilité, sa précision et sa résolution.

L'analyse d'un signal dans le cadre d'un système de détection inductif peut impliquer des méthodes d'analyse générales et basiques et de traitement du signal telles que la transformée de Fourier, la décomposition en ondelettes et la corrélation. Mais la possibilité d'étudier et de comparer un signal par le biais du machine learning est une piste et un travail à étudier pour extraire des résultats probants. Il est aussi important de disposer des outils et des technologies appropriés pour réaliser les mesures et les analyses, tels que des générateurs de signaux, des oscilloscopes et des analyseurs de spectre.

En somme, l'étude d'un signal dans le cadre d'un système de détection inductif est un domaine clé de la recherche en ingénierie électronique qui permet de comprendre les caractéristiques des signaux produits par le système et de déterminer les paramètres pertinents pour une détection efficace. Il est important de prendre en compte les différents paramètres influents et les limites du système pour obtenir des résultats précis et fiables.

6.3 Concept algorithmique d'analyse du signal

L'étude du signal en utilisant des concepts algorithmiques tels que l'apprentissage automatique (machine learning en anglais) est devenu un domaine de recherche de plus en plus populaire ces dernières années. Les techniques d'apprentissage automatique permettent d'extraire des informations cachées ou de découvrir des relations complexes dans les données de signal, permettant d'améliorer les performances des systèmes dans des domaines tels que la reconnaissance vocale, la vision par ordinateur, la détection de défauts, etc.

Il y a différentes techniques d'apprentissage automatique qui peuvent être utilisées pour cette étude. Les réseaux de neurones, par exemple, sont des modèles de traitement du signal qui peuvent être utilisés pour des tâches de classification, de reconnaissance de formes, de régression, etc. Les réseaux de neurones peuvent être entraînés à partir de données de signal pour apprendre à reconnaître des caractéristiques spécifiques dans le signal, comme les formes, les fréquences ou les modèles.

Les méthodes de classification non supervisée, comme l'analyse en composantes principales (PCA) ou l'analyse discriminante linéaire (LDA), peuvent également être utilisées pour étudier les signaux. Ces méthodes permettent de réduire la dimensionnalité des données de signal et de mettre en évidence les caractéristiques les plus pertinentes pour la classification.

6.3.1 Interpretation et prediction

L'interprétation et la prédiction de signaux en utilisant l'apprentissage automatique (machine learning) est un domaine de recherche en pleine croissance qui permet d'améliorer les performances des systèmes en utilisant des données de signal. Les techniques d'apprentissage automatique permettent d'extraire des informations cachées ou de découvrir des relations complexes dans les données de signal, permettant de prédire les comportements futurs et de prendre des décisions en conséquence.

Pour l'interprétation de signaux, les techniques d'analyse de données telles que l'analyse en composantes principales (PCA) ou l'analyse discriminante linéaire (LDA) peuvent être utilisées pour réduire la dimensionnalité des données de signal et mettre en évidence les caractéristiques les plus pertinentes pour l'interprétation. Les réseaux de neurones, tels que les réseaux de neurones profonds (DNN) peuvent également être utilisés pour interpréter les signaux en les entraînant à partir de données de signal pour reconnaître des caractéristiques spécifiques dans le signal.

Pour la prédiction de signaux, les techniques de régression, telles que les réseaux de neurones récurrents (RNN) ou les modèles de séries chronologiques, peuvent être utilisées pour prédire les valeurs futures d'un signal en fonction des valeurs passées. Les techniques de classification, telles que les

arbres de décision ou les forêts aléatoires, peuvent également être utilisées pour prédire la classe d'un signal en fonction de ses caractéristiques.

Il est important de noter que l'interprétation et la prédiction de signaux en utilisant l'apprentissage automatique nécessite une quantité importante de données de signal pour entraîner les modèles, ainsi qu'une bonne compréhension des techniques d'apprentissage automatique pour les utiliser de manière efficace. Il est donc important de disposer des connaissances et des outils appropriés pour utiliser ces techniques pour l'analyse de signaux.

7 Etat de l'art des moyens d'analyses des signaux

7.1 Introduction

Les capteurs inductifs sont largement utilisés dans l'industrie pour la détection de matériaux, la mesure de position et la détection de défauts. Cependant, l'analyse et l'identification des signaux provenant de ces capteurs peuvent être complexes en raison de la nature non linéaire des signaux et des interférences électromagnétiques. Pour résoudre ce problème, de nombreuses techniques de traitement de signal ont été proposées dans la littérature.

Dans cette étude, nous présentons un état de l'art sur les techniques d'analyse et d'identification des signaux provenant d'un système de détection inductif. Nous examinons les travaux les plus récents dans ce domaine, en nous concentrant sur les études qui ont proposé des méthodes novatrices pour améliorer la précision et la fiabilité des signaux inductifs.

Nous allons examiner les travaux de Hendrik Kruger et Hartmut Ewald, qui ont étudié la capacité de visualisation et classification en ligne pour le déminage humanitaire, et de Minh Dao-Johnson Tran; Canicious Abeynayake; Lakhmi C. Jain, qui ont utilisé des méthodologies de discrimination de cible utilisant l'extraction de caractéristiques morphologiques et basées sur les ondelettes avec des données de matrice de détecteurs de métaux

En examinant ces études, nous montrons comment les techniques de traitement de signaux peuvent être utilisées pour améliorer la qualité des signaux inductifs et pour extraire des informations précieuses sur les matériaux, les positions et les défauts. Enfin, nous soulignons les opportunités de recherche futures dans ce domaine, notamment en explorant les techniques d'apprentissage automatique et les méthodes de décomposition avancées pour une analyse plus précise et plus fiable des signaux inductifs.

7.2 Contexte historique

Les capteurs inductifs ont été utilisés pour la première fois dans les années 1940 pour détecter les métaux dans les mines et les carrières. Depuis lors, ces capteurs ont trouvé de nombreuses applications industrielles pour la détection de matériaux, la mesure de position et la détection de défauts. Cependant, l'analyse des signaux provenant de ces capteurs a été un défi constant en raison de la nature non linéaire des signaux et des interférences électromagnétiques.

Au fil des ans, de nombreuses techniques de traitement de signal ont été proposées pour améliorer la précision et la fiabilité des signaux inductifs. Dans les années 1990, la décomposition en ondelettes a été proposée comme une technique efficace pour extraire des caractéristiques discriminantes à partir de signaux inductifs (Daubechies, 1992). Cette technique a été largement utilisée pour la discrimination des matériaux à partir de signaux inductifs (Minh Dao-Johnson Tran; Canicious Abeynayake; Lakhmi C. Jain, 2007).

Dans les années 2000, la transformation de Hilbert-Huang a été proposée comme une technique alternative pour l'analyse des signaux inductifs (Yu Zhang, Shen Wang, Songling Huang*, Wei Zhao, 2015). Cette technique a été utilisée avec succès pour de la reconnaissance des modes des signaux de détection d'onde dans une plaque métallique à l'aide de cette méthode, démontrant ainsi son efficacité.

Plus récemment, les techniques d'apprentissage automatique, telles que les réseaux de neurones artificiels et les machines à vecteurs de support, ont été proposées pour améliorer l'analyse des signaux

inductifs, plus précisément dans de la classification de mines anti personnels. (M. Baydoun, 2020). Ces techniques ont montré une grande promesse pour la détection de défauts et la caractérisation des matériaux à partir des signaux inductifs. Dans cet article, les chercheurs rapportent les tentatives d'utilisation de l'apprentissage automatique pour la prise de décision dans le processus de déminage. Ces chercheurs ont créé leur propre base de données, correspondant aux mines terrestres et/ou au fouillis (fausses alertes).

En examinant le contexte historique de la recherche sur l'analyse et l'identification des signaux provenant de capteurs inductifs, nous pouvons voir comment les techniques de traitement de signal ont évolué au fil du temps et comment elles ont été utilisées pour résoudre les problèmes associés aux signaux inductifs. Cette perspective peut aider à éclairer les défis actuels et les opportunités de recherche dans ce domaine en constante évolution.

7.3 Etat actuel des connaissances

L'état actuel des connaissances sur les techniques d'analyse et d'identification des signaux provenant d'un système de détection inductif montre que plusieurs approches ont été proposées dans la littérature scientifique.

Tout d'abord, l'article de Timur karimov. (2022) présente une approche basée sur l'analyse de Fourier pour extraire les caractéristiques du signal d'un capteur inductif et les utiliser pour la détection basé sur l'oscillateur chaotique Duffing. Cette méthode est basée sur l'observation que les signaux de capteurs inductifs peuvent être représentés comme une somme de sinusoides à différentes fréquences, qui peuvent être analysées par une transformée de Fourier.

Le troisième article, celui de Marko Šimić (2004), propose une approche basée sur les réseaux de neurones artificiels pour l'estimation des objets à partir des données d'un système de detection inductif. Dans cet article, les chercheurs présentent une nouvelle approche pour estimer la profondeur d'un objet métallique en utilisant un détecteur de métaux à induction pulsée combiné à un système de suivi électromagnétique. Ils utilisent un modèle d'approximation dipolaire pour représenter la réponse spatiale du détecteur de métal, et un réseau de neurones à convolution 1D pour estimer la profondeur. L'algorithme développé a été expérimentalement validé dans des conditions de laboratoire. En effectuant un seul passage horizontal sur un objet métallique situé dans une plage donnée de coordonnées, l'algorithme est capable d'estimer la profondeur de l'objet, indépendamment de sa forme, de sa taille et de ses propriétés matérielles, avec une erreur absolue moyenne inférieure à 4,5 mm.

Enfin, l'article de Yu Zhang et al. (2015) propose une approche basée sur la transformée de Hilbert-Huang pour l'identification des caractéristiques dynamiques des capteurs inductifs.

Cet article traite de la problématique liée à l'analyse et à la reconnaissance des signaux d'écho de détection des défauts dans les plaques métalliques, qui sont rendus complexes en raison des caractéristiques de dispersion et de modes multiples présentes dans la propagation des ondes de Lamb (LW).

Les chercheurs ont mis en place un système expérimental visant à détecter les signaux LW dans une plaque d'aluminium. Ils ont utilisé la transformée de Hilbert-Huang (HHT), une méthode récente d'analyse temps-fréquence, pour analyser ces signaux. Grâce à l'extraction du temps de vol des fonctions de mode intrinsèque (IMF) après la transformée de Hilbert (HT), les signaux LW superposés de différents modes ont pu être identifiés.

Les résultats expérimentaux ont confirmé la validité de l'analyse théorique et ont démontré que la méthode HHT permettait de reconnaître clairement les signaux LW superposés de différents modes dans les plaques métalliques. Cependant, la méthode rencontrait des difficultés pour distinguer les

signaux du même mode. Ainsi, la méthode HHT s'avère être une approche efficace pour la reconnaissance des signaux de détection LW de différents modes dans les plaques métalliques.

En conclusion, l'état actuel des connaissances montre que plusieurs approches ont été proposées pour l'analyse et l'identification des signaux provenant d'un système de détection inductif. Ces approches incluent l'analyse de Fourier, la transformée en ondelettes, la décomposition en composantes indépendantes, les réseaux de neurones artificiels et la transformée de Hilbert-Huang. Chaque méthode a ses avantages et ses limites, et le choix de la méthode appropriée dépend des caractéristiques spécifiques de l'application.

7.4 Themes émergents

Les thèmes émergents dans la recherche sur les systèmes de détection inductifs incluent l'utilisation de l'intelligence artificielle (IA) pour améliorer les performances de ces systèmes. Plusieurs études ont montré l'efficacité des réseaux de neurones artificiels (ANN) pour la détection de métaux dans les capteurs inductifs, tel que l'étude de Minh Joo, Jaehyun Yoon, Allah Rakhio Junejo Jaehyeok Doh (2022). Cet article présente une méthodologie qui utilise un drone équipé d'un détecteur de métaux et programmé avec des modèles d'apprentissage automatique (ML) pour la détection des métaux. Le processus de recherche proposé représente une technologie de détection de mines sans pilote sûre et efficace pour la localisation et l'élimination ultérieure des mines terrestres.

Les utilisateurs de cette méthodologie peuvent contrôler les drones à distance, ce qui leur permet de détecter les métaux enfouis sans entrer dans la zone minée, et de déterminer si le métal détecté est une mine ou non. Pour réaliser cette idée, les chercheurs ont tout d'abord amélioré le contrôle d'attitude du drone avec un détecteur de métaux en utilisant une optimisation basée sur un algorithme micro-génétique des gains de contrôle proportionnel-intégral-différentiel.

Ensuite, pour la détection des métaux, ils ont formé des modèles d'apprentissage automatique tels qu'une machine à vecteurs de support et un réseau de neurones à rétropropagation en utilisant un ensemble de données annotées. Enfin, ils ont construit un drone équipé d'un détecteur de métaux et des modèles ML entraînés, et ont expérimentalement validé leur méthodologie. Les résultats des expériences ont démontré la stabilité du vol des drones de détection de métaux sans pilote et un taux élevé de succès dans la détection des métaux.

Cependant, il reste encore des lacunes dans les connaissances actuelles, notamment en ce qui concerne l'optimisation des paramètres des capteurs inductifs et de l'amélioration de la précision des mesures. De plus, la plupart des études se concentrent sur l'analyse de signaux de capteurs inductifs dans des conditions statiques. Il serait intéressant de poursuivre les recherches dans des conditions dynamiques pour mieux comprendre les performances des capteurs inductifs dans des applications réelles.

En outre, il y a un manque de standardisation en termes de techniques de traitement des signaux pour les capteurs inductifs, ce qui rend difficile la comparaison des résultats entre différentes études. Il serait donc important de développer des normes de mesure et de traitement de signal pour les capteurs inductifs afin de faciliter la comparaison des résultats et d'assurer la qualité des mesures.

Enfin, l'intégration de capteurs inductifs avec d'autres technologies telles que l'imagerie par ordinateur ou la robotique pourrait ouvrir la voie à de nouvelles applications et améliorer l'efficacité de ces systèmes. Par exemple, les capteurs inductifs pourraient être utilisés pour la localisation de robots ou pour la navigation de véhicules autonomes. Ces pistes de recherche pourraient contribuer à l'avancement des connaissances sur les systèmes de détection inductifs et à leur utilisation dans un large éventail d'applications.

7.5 Conclusion de l'état de l'art

L'examen approfondi de la littérature existante met en évidence un certain nombre de tendances émergentes et l'état actuel des connaissances dans le domaine de la détection de métaux par le biais de l'apprentissage automatique. Il est important de souligner que cette recherche est relativement récente et que seules quelques études ont été publiées dans ce domaine. Cela démontre le caractère émergent et novateur de l'utilisation du machine learning[7] pour la reconnaissance des signaux de détection de métaux.

Les travaux de recherche disponibles se concentrent principalement sur l'application de techniques d'apprentissage automatique telles que les réseaux de neurones, les machines à vecteurs de support et les méthodes de classification[5] avancées pour analyser et interpréter les signaux de détection de métaux. Ces études mettent en évidence les avantages potentiels de l'utilisation de ces approches par rapport aux méthodes traditionnelles de détection de métaux.

Il convient également de noter que les études existantes se sont principalement concentrées sur des types spécifiques de détecteurs de métaux, tels que les détecteurs à induction pulsée et les ondes de Lamb. Cependant, il reste encore beaucoup à explorer en ce qui concerne l'application du machine learning à d'autres types de détecteurs de métaux et à d'autres aspects de la détection de métaux, tels que la localisation précise des objets métalliques et la distinction entre différents types de métaux.

En conclusion, l'état de l'art met en évidence le potentiel prometteur de l'apprentissage automatique pour la reconnaissance des signaux de détection de métaux. Bien que les recherches dans ce domaine soient encore limitées, les résultats obtenus jusqu'à présent soutiennent pleinement la pertinence et la justification de notre étude, qui se concentre sur l'application du machine learning à la détection de métaux.

7.6 Etudes relatives

Voici une liste des études relatives à l'état de l'art, avec les technologies citées :

"Handheld metal detector with online visualisation and classification for the humanitarian mine clearance", Hendrik Kruger ; Hartmut Ewald, 16 December 2008.

"A Target Discrimination Methodology Utilizing Wavelet[6]-Based and Morphological Feature Extraction With Metal Detector Array Data", Minh Dao-Johnson Tran ; Canicious Abeynayake ; Lakhmi C. Jain, 04 August 2011

"Mode Recognition of Lamb Wave Detecting Signals in Metal Plate Using the Hilbert-Huang Transform Method", Yu Zhang, Shen Wang, Songling Huang*, Wei Zhao, 2015

"Object Depth Estimation From Line-Scan EMI Data Using Machine Learning", Marko Šimić ; Davorin Ambruš ; Vedran Bilas, 08 December 2022

"Optimization : Drone-Operated Metal Detection Based on Machine Learning and PID Controller", Minhoo Joo, Jaehyun Yoon, Allah Rakhio Junejo Jaehyeok Doh, 21 March 2022

"Physical Contamination Detection in Food Industry Using Microwave and Machine Learning", Ali Darwish, 29 septembre 2022

"Identification des Objets Métalliques Enfouis par Réseaux de Neurones Artificiels dans un Système de Détection Inductif", Ahmed Chaouki Lahrech (1) , Abdelhalim Zaoui , Fethi Benyoubi, 03-07-2022

8 Composants types d'un SDI

Dans le cadre d'une simulation d'un système de détection inductif, il était nécessaire d'étudier plus précisément la partie électronique de l'appareil. On y remarque ainsi que la plupart des détecteurs de métaux sont équipés d'un composant nommé NE555 Timer. Comme son nom l'indique, ce composant est un circuit intégré utilisé pour la temporisation, et fait office de fréquenceur. Ainsi, les détecteurs de métaux NE555 sont aussi catégorisés en trois grandes familles :

- Détecteurs très basse fréquence.
- Détecteurs à induction d'impulsions.
- Détecteurs de métaux à usage spécifique.

Avant de parler du fonctionnement général du détecteur de métaux NE555, il est nécessaire de comprendre les concepts de :

- Inductance
- NE555 Timer operational system

8.1 Inductance

"L'inductance, en électromagnétisme et en électronique, est la capacité d'un conducteur d'électricité à annuler le changement du courant électrique qui le traverse. Ce flux de courant électrique produit un champ magnétique autour de ce conducteur électrique."

Dans le cadre d'un circuit du détecteur de métaux NE555 Timer, nous utilisons un inducteur qui détecte la présence du métal à proximité (une bobine). Plus un détecteur de métaux est proche de celui-ci, plus il y a de lignes de champ électrique produites et, par conséquent, la fréquence du signal en sortie de l'inducteur changera. Nous désignons l'inductance par "L". Par conséquent, la formule pour trouver l'inductance à travers n'importe quel conducteur peut être trouvée par : Inductance = flux magnétique de courant/courant. Dans la suite de l'étude, nous nous pencherons sur l'impact de l'inductance en fonction de la fréquence mesurée.

8.2 NE555 Timer

Le circuit intégré NE555 appartient à la famille des circuits intégrés 555 timer. Ce sont d'ailleurs les circuits intégrés les plus utilisés dans le monde de l'électronique. Ainsi, le Timer NE555 est un circuit intégré composé de 8 broches (descriptif ci-après), et notamment utilisé pour avoir à sa sortie des impulsions uniformes pouvant être réglées.

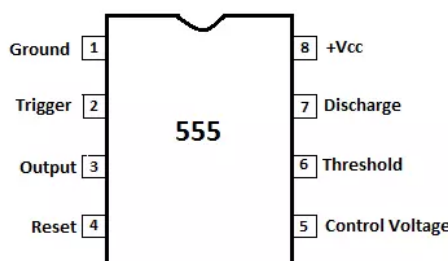


FIGURE 2 – Schématique d'un composant NE555 Timer

8.3 Circuit d'un Systeme de detection inductif

Pour simuler correctement le comportement des composants dans le circuit, le logiciel Proteus[4] a été privilégié pour sa vaste bibliothèque de composants et ses performances dans la simulation de circuits électroniques.

Ainsi, plusieurs points importants sont à prendre en compte dans l'étude de ce circuit :

- Tous les composants sont alimentés par une batterie délivrant 9V.
- Cette puissance entre ensuite dans le circuit NE555 qui délivre lui-même des impulsions uniformes basées sur le timer de sa broche de sortie (TH).
- Cette impulsion entre dans la résistance de 47kohms, permettant ainsi de contrôler le circuit principal.
- Ensuite, la résistance transmet ce courant à l'inducteur. L'inductance mesurée sur cet inducteur sera le critère privilégié pour déterminer le type de métal ainsi que sa distance par rapport à la bobine.
- Pour la simulation, dans Proteus, il n'est pas possible de déplacer un objet métallique devant l'inducteur. Il faut donc changer la valeur de l'inductance, simulant ainsi le type d'objet et sa distance par rapport à la bobine. Ainsi, plus il y a d'inductance, plus cela simule que l'objet métallique est proche du circuit.

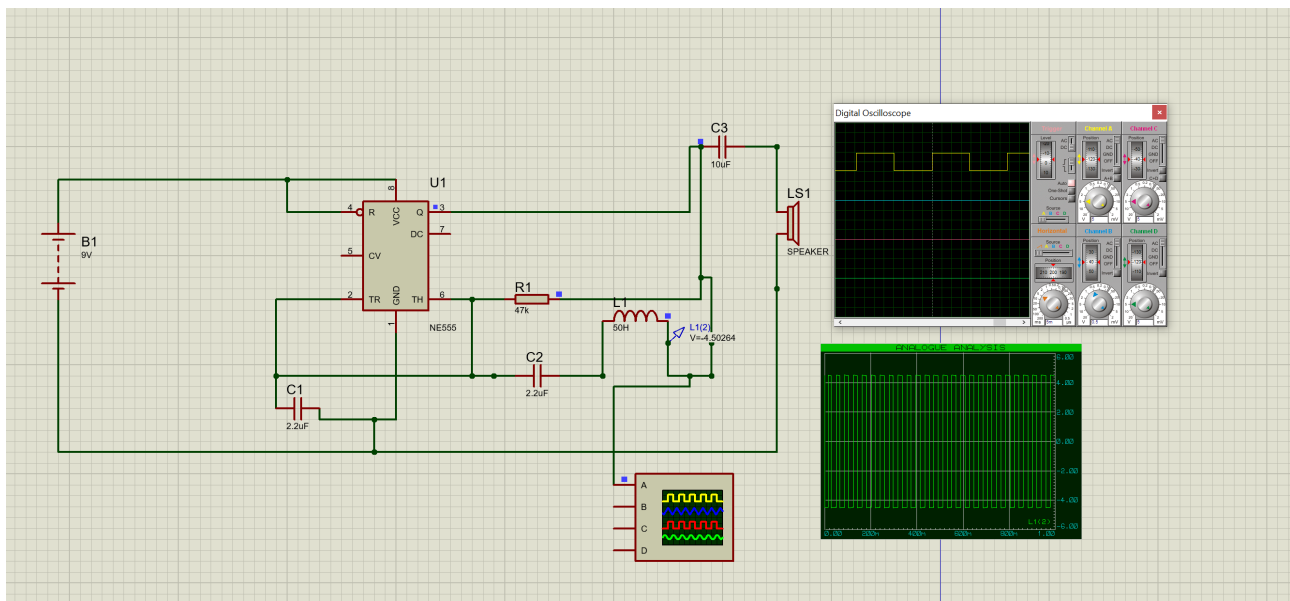


FIGURE 3 – Circuit électronique simulé d'un SDI

Le défi à présent est de mesurer le signal en sortie de l'inductance. On remarque ainsi un signal périodique, comme le montrent l'oscilloscope et le graphique en vert. Pour mieux catégoriser ce signal et comparer différentes valeurs d'inductance et de signal, exportons les valeurs de ce graphique au format CSV pour les exploiter dans un script Python.

8.4 Étude des signaux

8.4.1 Inductance de 150 mH

Voyons ainsi comment catégoriser un signal et trouver la corrélation entre l'inductance de la bobine et la fréquence du signal renvoyé en sortie. Pour cela, l'idée est de sélectionner une très courte portion du signal (quelques millisecondes) afin d'en extraire la période et de calculer la fréquence. Voyons un signal mesuré pour une inductance de 150 mH :

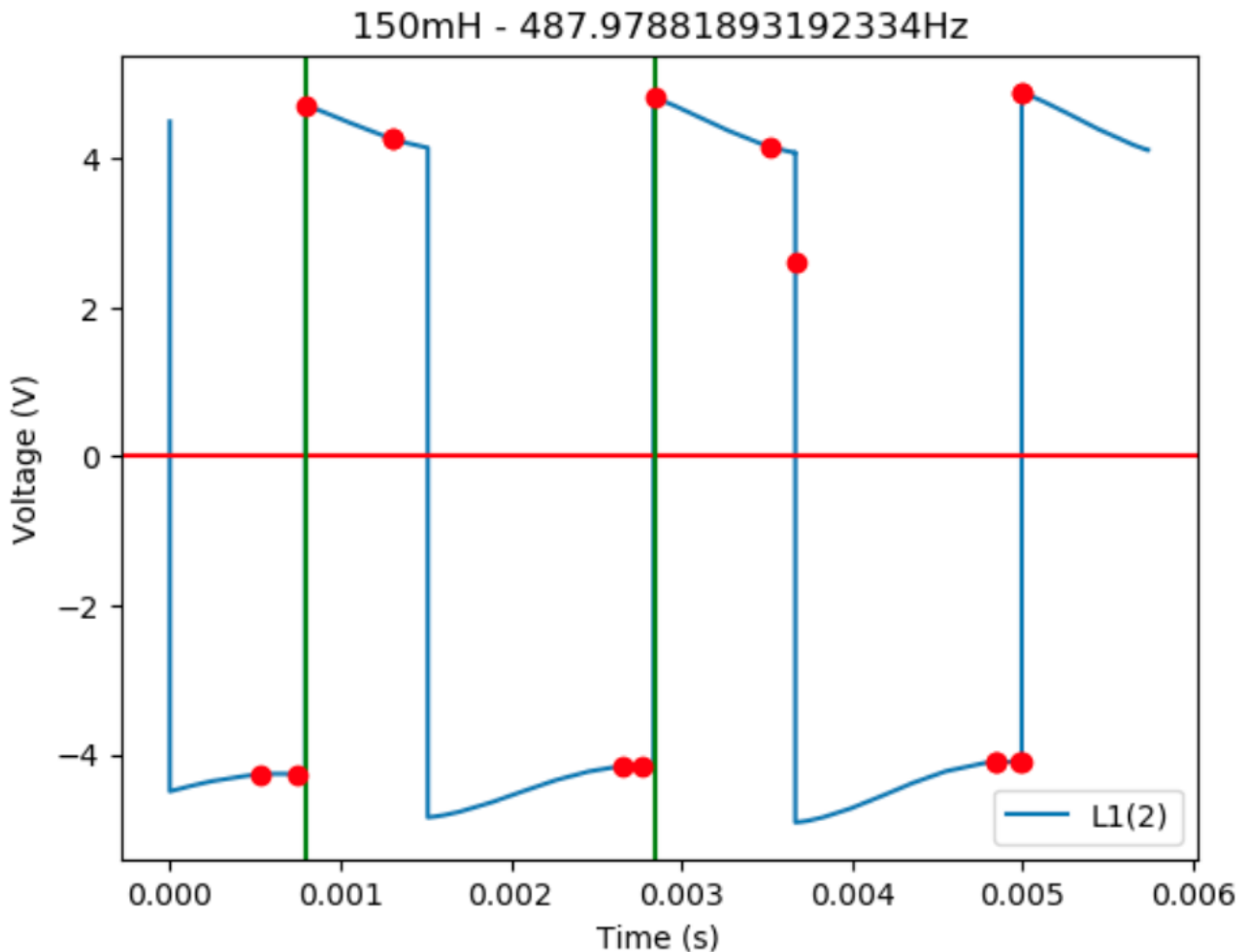


FIGURE 4 – Signal généré par le circuit simulé avec une valeur d'inductance de 150 mH

Appliquons un algorithme nous permettant de trouver les maximums locaux sur la courbe. L'idée est d'avoir deux points en symétrie pour délimiter clairement un cycle complet d'impulsion (et déterminer la période). Ci-dessus, en rouge, sont représentés les points des minimums locaux. Ceux sélectionnés pour délimiter les deux bornes d'un cycle sont représentés par les courbes verticales vertes. Ainsi, la période est définie par la différence entre les deux bornes par rapport à l'axe du temps. Une fois la période trouvée, et la fréquence se calculant par l'inverse de la période, on trouve une valeur d'environ 488 Hz pour une inductance de 150 mH d'après les signaux de la simulation dans Proteus. Cela nous donne une base pour pouvoir comparer cette valeur d'inductance avec les autres. On remarque également que le pic de tension maximum dépasse les 4V, tout comme le pic de tension minimum dépasse lui les -4V.

8.4.2 Inductance de 10 mH

Comparons maintenant avec une valeur d'inductance plus faible et voyons comment la fréquence et la courbe réagissent.

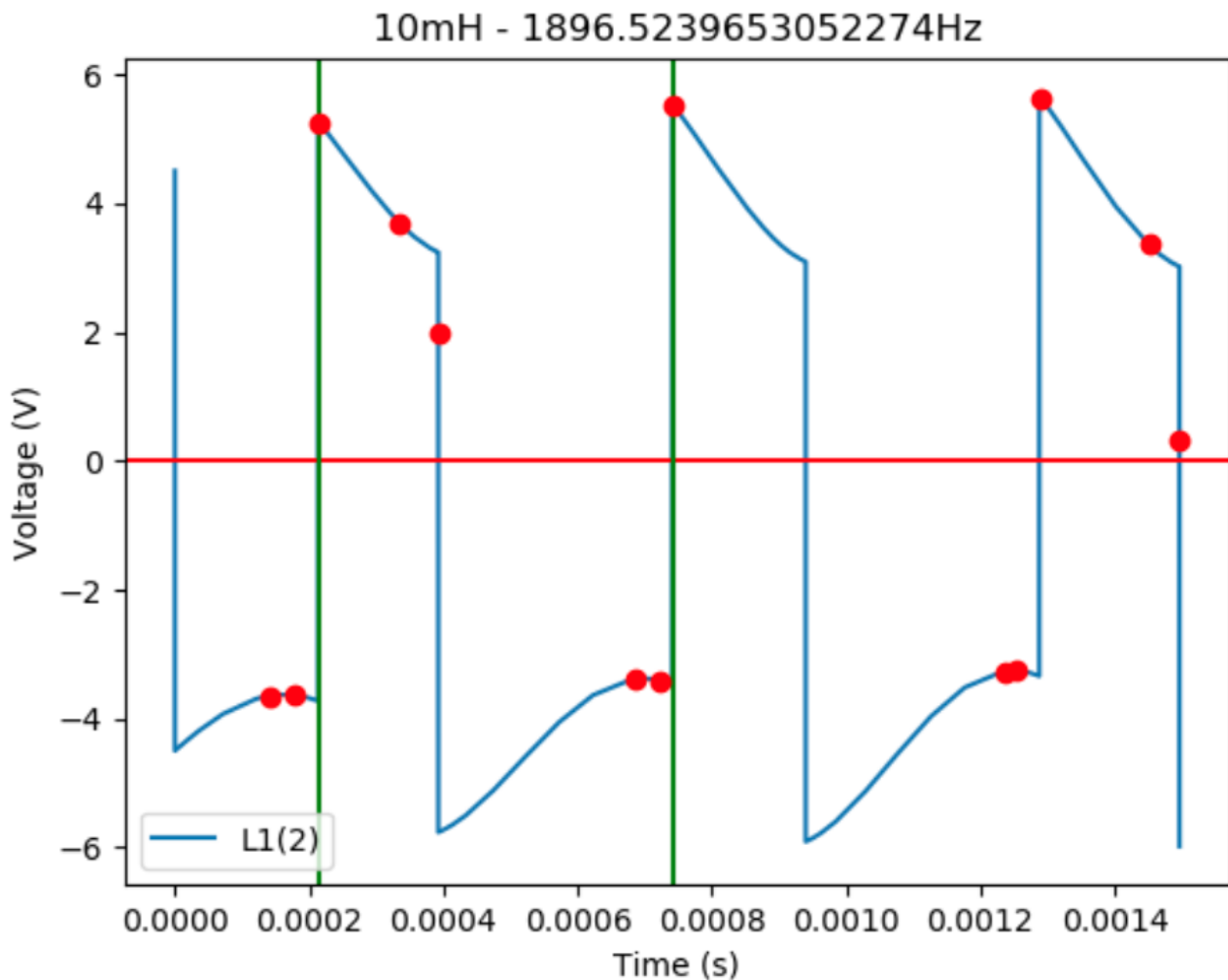


FIGURE 5 – Signal généré par le circuit simulé avec une valeur d'inductance de 10 mH

Comme le montre la courbe ci-dessus, et après avoir calculé la période ainsi que la fréquence, on remarque qu'une inductance plus faible augmente considérablement la fréquence. On remarque également que le pic de tension dépasse ici les 5V en positif et en négatif, et que la forme du signal varie. Les minimums locaux sont plus marqués, montrant ainsi une réelle évolution avec le changement de l'inductance par rapport au signal.

La période du signal est également beaucoup plus courte que celle du signal correspondant à l'inductance de 150 mH. Cela s'explique par l'évolution de la fréquence et de la relation entre l'inductance et cette fréquence. On peut également se demander si le changement de valeur d'inductance fait évoluer la fréquence de manière linéaire ou si cette relation suit une fonction ou une loi particulière.

8.4.3 Corrélation entre l'inductance et la fréquence

Pour vérifier cela, toujours à partir des valeurs de la simulation et à l'aide de notre algorithme, les fréquences des signaux provenant de différentes valeurs d'inductance ont été calculées et stockées pour en faire un graphique représentatif de ces valeurs :

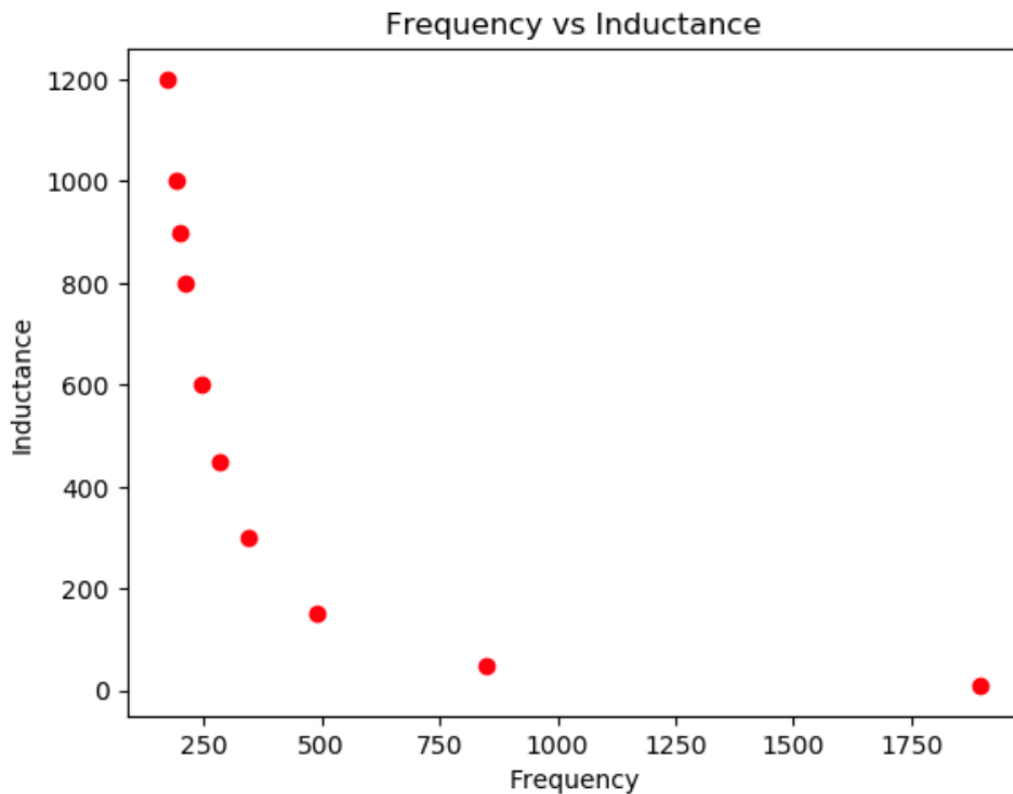


FIGURE 6 – Graphique de corrélation entre la fréquence et l'inductance

Comme on peut le remarquer, cela confirme la première affirmation : "si on diminue l'inductance, on augmente considérablement la fréquence". La courbe prend une allure exponentielle. Ici, peu de données sont représentées compte tenu de la quantité d'informations à traiter à partir des signaux récupérés depuis le simulateur, mais cela suffit pour voir l'aspect général de la courbe et commencer éventuellement à ajouter des concepts de prédiction de fréquence par rapport à l'inductance, et vice versa.

L'inductance est une caractéristique des circuits électriques qui décrit leur tendance à emmagasiner de l'énergie sous forme de champ magnétique lorsqu'un courant électrique les traverse. Plus un circuit a d'inductance, plus il résiste au changement de courant. Cela signifie que si on applique une tension alternative à un circuit inductif, il y aura une différence entre la tension appliquée et la tension effectivement mesurée dans le circuit.

La fréquence de la tension alternative a un effet sur cette différence de tension. Plus la fréquence est élevée, plus la différence de tension sera faible. Cela signifie que si on applique une tension alternative à haute fréquence à un circuit inductif, le circuit ne résistera pas autant au changement de courant et la différence de tension sera faible. En revanche, si on applique une tension alternative à basse fréquence à un circuit inductif, le circuit résistera davantage au changement de courant et la différence de tension sera plus importante.

En résumé, l'inductance d'un circuit est directement liée à sa fréquence : plus la fréquence est élevée, moins le circuit a d'inductance.

9 Données

9.1 Explication generale

Le jeu de données actuel se compose de 10 éléments. Il s'agit d'une première étape dans le projet et il n'y a pas beaucoup de données pour prédire beaucoup de choses. Ainsi, le prochain objectif est d'obtenir un jeu de données plus important.

	Frequency	Inductance
0	1896.52	10.00
1	345.84	300.00
2	283.24	450.00
3	243.95	600.00
4	173.74	1200.00
5	487.98	150.00
6	211.91	800.00
7	200.19	900.00
8	189.95	1000.00
9	847.79	50.00

FIGURE 7 – Premier jeu de données

Une nouvelle caractéristique que nous allons introduire dans ce nouveau jeu de données est la perméabilité. Celle-ci est un paramètre physique qui définit une catégorie métallique spécifique. Il y a quatre catégories métalliques : ferromagnétique, paramagnétique, diamagnétique et non magnétique. Les coefficients de perméabilité sont définis par μ_m , avec μ_m appartenant à $[0,6,1]$ pour le diamagnétique, $[1,100]$ pour le paramagnétique, $[101,5000]$ pour le ferromagnétique et $[1]$ pour le non-perméable.

En générant une perméabilité aléatoire, nous pouvons également calculer l'inductance en utilisant la formule : $L = \mu_m \cdot N^2 \cdot A / (2\pi f)$, où μ_m est la perméabilité, N est le nombre de tours, A est la surface de la bobine et f est la fréquence.

En calculant l'inductance, nous pouvons également calculer la fréquence en utilisant la formule : $f = \frac{1}{2\pi\sqrt{LC}}$, où L est l'inductance et C est la capacité (constante pour tout le jeu de données).

Nous utiliserons en premier lieu une fonction qui effectuera toutes ces étapes, et nous l'appellerons un certain nombre de fois pour obtenir des données dans notre jeu de données.

	Frequency	Inductance	Permeability
0	43873.359488	0.001316	Non permeable
1	43873.359488	0.001316	Non permeable
2	48880.540689	0.001060	diamagnetique
3	878.522050	3.281972	ferromagnetique
4	809.421983	3.866253	ferromagnetique
5	55327.651893	0.000827	diamagnetique
6	43873.359488	0.001316	Non permeable
7	2075.138697	0.588228	ferromagnetique
8	43873.359488	0.001316	Non permeable
9	5711.824893	0.077641	paramagnetique

FIGURE 8 – Nouveau jeu de données

9.1.1 Augmentation du dataset

Dans le domaine de la science des données, l'augmentation de la taille d'un dataset revêt une importance cruciale. Un dataset plus volumineux permet d'améliorer la précision des modèles prédictifs et d'obtenir des résultats plus fiables. Pour atteindre cet objectif, différentes méthodes peuvent être utilisées, notamment la duplication de données avec l'ajout de bruit.

Lorsque nous parlons de duplication de données, il s'agit de créer de nouvelles instances en copiant des observations existantes dans le dataset. Cette approche peut sembler contre-intuitive, car elle peut conduire à un dataset redondant et biaisé. Cependant, lorsque le bruit est ajouté aux données dupliquées, cela peut réellement apporter des avantages significatifs.

Le bruit, dans ce contexte, fait référence à l'introduction de variations aléatoires dans les caractéristiques ou les étiquettes des données dupliquées. Cela peut être réalisé en appliquant des transformations aléatoires aux données existantes ou en ajoutant des valeurs aléatoires à certaines caractéristiques. Par exemple, si nous dupliquons une observation d'un objet métallique dans notre étude, nous pouvons ajouter du bruit en modifiant légèrement son inductance ou sa fréquence.

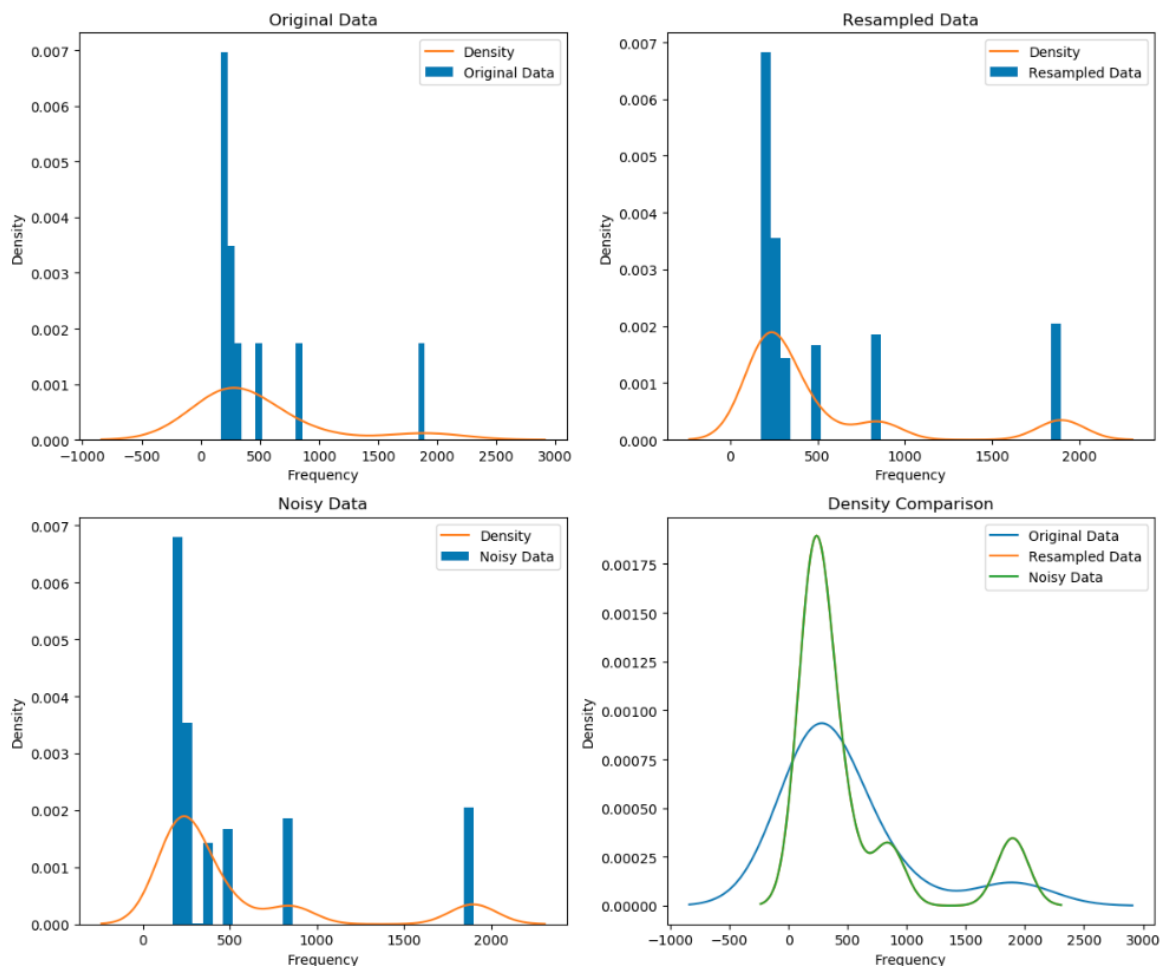


FIGURE 9 – Densité des jeux de données au fur et à mesure de la repopulation du dataset

Les 4 graphiques obtenus présentent les distributions de fréquence pour différentes étapes de la population de données.

Le premier graphique représente le jeu de données initial (Original Data). Il montre la distribution de fréquence du dataset initial avec son histogramme et sa courbe de densité. Ce graphique est utilisé comme point de référence pour comparer les autres graphiques.

Le deuxième graphique correspond au jeu de données après avoir été resamplé (Resampled Data). On peut observer une augmentation du nombre de données par rapport au jeu de données initial. Cela se traduit par une distribution de fréquence plus étalée et moins concentrée dans certaines plages de valeurs.

Le troisième graphique représente le jeu de données après avoir été soumis à un bruit (Noisy Data). On peut constater une amélioration de la répartition des données par rapport au jeu de données resamplé. Le bruit ajouté a permis de répartir les valeurs de fréquence de manière plus uniforme, réduisant ainsi les concentrations dans certaines plages de valeurs.

Le dernier graphique présente une comparaison des courbes de densité des trois jeux de données : Original Data, Resampled Data et Noisy Data. On peut observer que la courbe verte représentant la densité des données du jeu de données Noisy Data présente des variations significatives par rapport aux autres courbes. Cela indique qu'en ajoutant du bruit, nous avons introduit de nouvelles données qui n'étaient pas nécessairement présentes auparavant. Ces nouvelles données ont renforcé les densités déjà présentes et ont conduit à une modification importante de la distribution globale du dataset.

L'avantage de cette méthode réside dans sa capacité à renforcer la robustesse du modèle en exposant celui-ci à une plus grande variété de cas. En ajoutant du bruit, nous introduisons des scénarios plus réalistes et diversifiés dans le dataset, permettant ainsi au modèle de mieux généraliser et de gérer des situations inconnues lorsqu'il est confronté à de nouvelles données.

De plus, l'ajout de bruit peut contribuer à réduire le surapprentissage (overfitting) en introduisant une certaine régularisation. Lorsque les données sont trop spécifiques et sans variation, le modèle peut apprendre des relations trop complexes qui ne se généralisent pas bien aux nouvelles données. L'ajout de bruit permet d'atténuer cette tendance en introduisant une dose de variabilité dans le dataset.

Il convient de souligner que l'ajout de bruit doit être effectué de manière judicieuse et contrôlée. Il est important de comprendre le domaine d'application spécifique et de veiller à ce que le bruit ajouté reste cohérent et réaliste par rapport au problème étudié. Une analyse et une évaluation appropriées sont nécessaires pour mesurer l'impact de l'ajout de bruit sur les performances du modèle.

9.2 Possibilités de Dataset

L'ouverture de notre étude à de nouvelles données est essentielle pour améliorer la précision de la détection métallique. En plus des paramètres existants tels que la perméabilité, la fréquence et l'inductance, il est pertinent d'envisager l'ajout de nouvelles caractéristiques pour une reconnaissance plus précise.

1. Humidité des sols : L'humidité des sols peut avoir un impact significatif sur la détection métallique. En ajoutant cette donnée à notre dataset, nous serons en mesure d'évaluer comment la conductivité électrique varie en fonction du degré d'humidité du sol. Cela nous permettra de mieux comprendre l'interaction entre l'eau présente dans le sol et les métaux enfouis.

2. Surface de l'objet : La surface de l'objet métallique peut influencer les propriétés électromagnétiques détectées. En considérant cette caractéristique, nous pourrions examiner comment la rugosité, la forme ou la texture de l'objet affectent les mesures de détection. Cela nous aidera à développer des modèles plus précis pour distinguer différents types d'objets métalliques.

3. Composition chimique : En intégrant des données sur la composition chimique des métaux, nous pourrions mieux différencier les métaux les uns des autres. Chaque type de métal a des propriétés spécifiques en termes de conductivité électrique et de réponse aux champs magnétiques. En

incluant ces informations dans notre dataset, nous pourrions affiner nos modèles pour une identification plus précise des métaux.

4. Profondeur d'enfouissement : La profondeur à laquelle les métaux sont enfouis peut également avoir un impact sur leur détection. En ajoutant cette donnée, nous pourrions étudier comment les propriétés électromagnétiques varient en fonction de la profondeur. Cela nous permettra de mieux cibler les zones où se trouvent les objets métalliques à des profondeurs spécifiques.

5. Conditions environnementales : Les conditions environnementales telles que la température, l'humidité de l'air et les conditions météorologiques peuvent influencer les mesures de détection. En incorporant ces facteurs dans notre dataset, nous serons en mesure de comprendre comment ils affectent les performances de nos modèles et d'ajuster nos méthodes en conséquence.

L'ajout de nouvelles données à notre dataset ouvre de nombreuses possibilités pour améliorer la précision de la détection métallique. L'inclusion de paramètres tels que l'humidité des sols, la surface de l'objet, la composition chimique, la profondeur d'enfouissement et les conditions environnementales permettra une meilleure compréhension des interactions entre les métaux et leur environnement. Ces données supplémentaires nous permettront de développer des modèles plus précis et de prendre des décisions éclairées pour la dépollution des sols et d'autres applications liées à la détection métallique.

9.3 Préparation des données

La préparation des données est une étape essentielle dans tout projet d'apprentissage automatique. Elle vise à transformer les données brutes en un format approprié pour l'entraînement d'un modèle. La pipeline de préparation est une séquence d'étapes de traitement appliquées aux données, permettant de les nettoyer, les transformer et les mettre à l'échelle de manière systématique. Cette pipeline garantit que les données d'entrée sont cohérentes, complètes et adaptées aux algorithmes d'apprentissage automatique.

9.3.1 Pourquoi traiter les données

La préparation des données est cruciale pour plusieurs raisons :

Nettoyage des données : Les données brutes peuvent contenir des valeurs manquantes, des valeurs aberrantes ou des erreurs. Le traitement des données permet de gérer ces problèmes en les éliminant ou en les remplaçant par des valeurs appropriées.

Transformation des données : Les modèles d'apprentissage automatique peuvent nécessiter des données dans un format spécifique. Par conséquent, il est souvent nécessaire de transformer les variables pour satisfaire les exigences du modèle. Cela peut inclure la conversion de variables catégorielles en variables numériques, l'encodage des variables catégorielles, la normalisation des variables numériques, etc.

Réduction du bruit : Les données brutes peuvent contenir des variations non pertinentes ou du bruit. La pipeline de préparation permet de réduire le bruit en filtrant les informations non pertinentes et en ne conservant que les caractéristiques importantes pour le modèle.

Mise à l'échelle des données : Certains algorithmes d'apprentissage automatique, tels que les modèles basés sur les distances, nécessitent que les données soient mises à l'échelle sur une même plage. La mise à l'échelle des données permet d'assurer une comparaison équitable entre les différentes caractéristiques.

9.3.2 Explication du fonctionnement de la pipeline

La pipeline de préparation des données se compose de plusieurs étapes, qui sont exécutées séquentiellement sur les données d'entrée. Chaque étape effectue une opération spécifique sur les données et passe les données transformées à l'étape suivante. Voici les principales étapes de la pipeline de préparation présentée :

StratifiedShuffleSplit : Cette étape utilise la méthode de découpage stratifié pour diviser les données en ensembles d'entraînement et de test tout en maintenant la répartition des catégories dans la variable cible (Permeability).

Séparation des données : Les ensembles d'entraînement et de test sont extraits à partir de l'ensemble de données généré. Les données cibles (Permeability) sont également séparées des caractéristiques d'entraînement.

Pipeline numérique : Cette étape définit une pipeline spécifique pour les caractéristiques numériques. Dans notre cas, la seule transformation appliquée est la mise à l'échelle standard (StandardScaler), qui permet de centrer les variables autour de zéro et d'ajuster leur écart type.

ColumnTransformer : Cette étape combine les différentes pipelines (dans notre cas, uniquement la pipeline numérique) avec les colonnes correspondantes du jeu de données. Elle permet d'appliquer les transformations spécifiques à chaque type de variable (numérique, catégorielle, etc.) de manière simultanée.

Fit et transform : Une fois la pipeline définie, elle est appliquée à l'ensemble d'entraînement. Une méthode est utilisée pour ajuster les transformations sur les données d'entraînement et les appliquer simultanément. Cela permet de calculer les paramètres de transformation (comme la moyenne et l'écart type pour la mise à l'échelle standard) à partir des données d'entraînement et d'appliquer ces transformations aux données d'entraînement et de test.

Résultats de la pipeline : Une fois la pipeline exécutée, les données d'entraînement sont transformées selon les étapes spécifiées. Le résultat est un ensemble de données d'entraînement préparé, où les caractéristiques numériques ont été mises à l'échelle.

Création du DataFrame préparé : Les données transformées sont converties en un DataFrame pandas, où les colonnes correspondent aux caractéristiques originales. Cela permet de conserver la structure du jeu de données et facilite l'analyse ultérieure.

Histogramme des données préparées : Enfin, un histogramme des données préparées est tracé pour visualiser la distribution des valeurs après la préparation. Cela peut aider à détecter d'éventuels problèmes de normalité, de valeurs aberrantes ou de transformations non souhaitées.

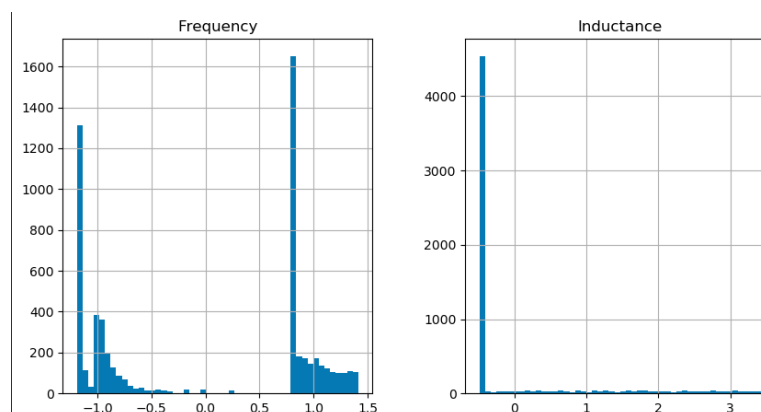


FIGURE 10 – Répartition du dataset

Dans notre étude, nous constatons que la répartition des inductances est assez complexe. La plupart des valeurs se situent dans un intervalle restreint, en relation avec la valeur de fréquence utilisée pour le calcul. En ce qui concerne la fréquence dans notre jeu de données d'exemple, elle se divise clairement en deux clusters, représentant deux groupes distincts avec des valeurs faibles et élevées, respectivement dans l'intervalle $[-1;1]$.

9.3.3 Résultat de la pipeline de préparation

La conclusion tirée des résultats de la pipeline de préparation, en mettant à l'échelle un dataset, est que cette étape est essentielle pour améliorer la performance des modèles d'apprentissage automatique. En redimensionnant les données, nous avons réussi à uniformiser l'échelle de toutes les caractéristiques, ce qui facilite la comparaison et l'interprétation des valeurs.

	Frequency	Inductance
0	0.90	-0.49
1	0.82	-0.49
2	0.82	-0.49
3	1.06	-0.49
4	-1.14	0.02

FIGURE 11 – Dataset mis à l'échelle

En mettant à l'échelle le dataset, nous avons également réduit les effets de variables indépendantes ayant des amplitudes très différentes, ce qui peut influencer négativement les performances du modèle.

De plus, en utilisant une pipeline de préparation, nous avons pu automatiser le processus de mise à l'échelle, ce qui facilite la réutilisation et l'adaptation de cette étape sur de nouveaux ensembles de données.

En conclusion, la mise à l'échelle du dataset à l'aide de la pipeline de préparation est une étape cruciale pour obtenir des résultats plus précis et cohérents lors de l'entraînement des modèles d'apprentissage automatique. Cela permet d'améliorer la performance du modèle en uniformisant l'échelle des caractéristiques et en réduisant les effets des variables indépendantes.

10 Moyens de prédiction

10.1 Generalités des predictions

Dans cette partie, nous aborderons les moyens de prédiction pour estimer la perméabilité d'un objet en fonction de sa fréquence et de son inductance perçue. L'objectif général est de développer des méthodes et des outils qui permettent de prédire la perméabilité d'un matériau à partir de ces caractéristiques physiques.

Nous allons explorer deux approches principales pour la prédiction de la perméabilité : les modèles de régression et les modèles de classification. Ces deux approches offrent des perspectives différentes pour aborder le problème de prédiction et peuvent être adaptées en fonction des besoins et des contraintes spécifiques de notre étude.

L'utilisation de modèles de régression nous permettra d'estimer la perméabilité en tant que valeur continue, ce qui nous donnera une indication précise de la magnétisation du matériau. Nous explorerons différentes techniques de régression, telles que la régression linéaire, la régression polynomiale ou encore les modèles d'apprentissage automatique plus avancés comme les réseaux de neurones.

D'autre part, nous considérerons également l'approche des modèles de classification. Cette approche vise à classer les matériaux en catégories prédéfinies de perméabilité, telles que ferromagnétique, paramagnétique, diamagnétique et non magnétique. Les modèles de classification nous permettront de prédire la catégorie à laquelle appartient un matériau en fonction de ses caractéristiques électromagnétiques.

En utilisant ces deux approches, nous pourrions évaluer la performance de différents modèles et sélectionner celui qui offre les meilleures prédictions pour notre problème spécifique. Nous explorerons également les techniques d'évaluation des modèles, telles que la validation croisée, pour garantir des résultats fiables et généralisables.

En résumé, cette partie sur les moyens de prédiction vise à explorer différentes méthodes, notamment les modèles de régression et de classification, pour estimer la perméabilité des matériaux à partir des caractéristiques de fréquence et d'inductance.

10.1.1 Algorithmes de regression

Nous utiliserons 3 algorithmes de prédiction comme la régression linéaire, la régression polynomiale et la Random Forest[2]. Le résultat était bon mais pas tout à fait représentatif de la situation. C'est pourquoi, avec un jeu de données plus important, le résultat des 3 algorithmes sera plus représentatif et moins faux.

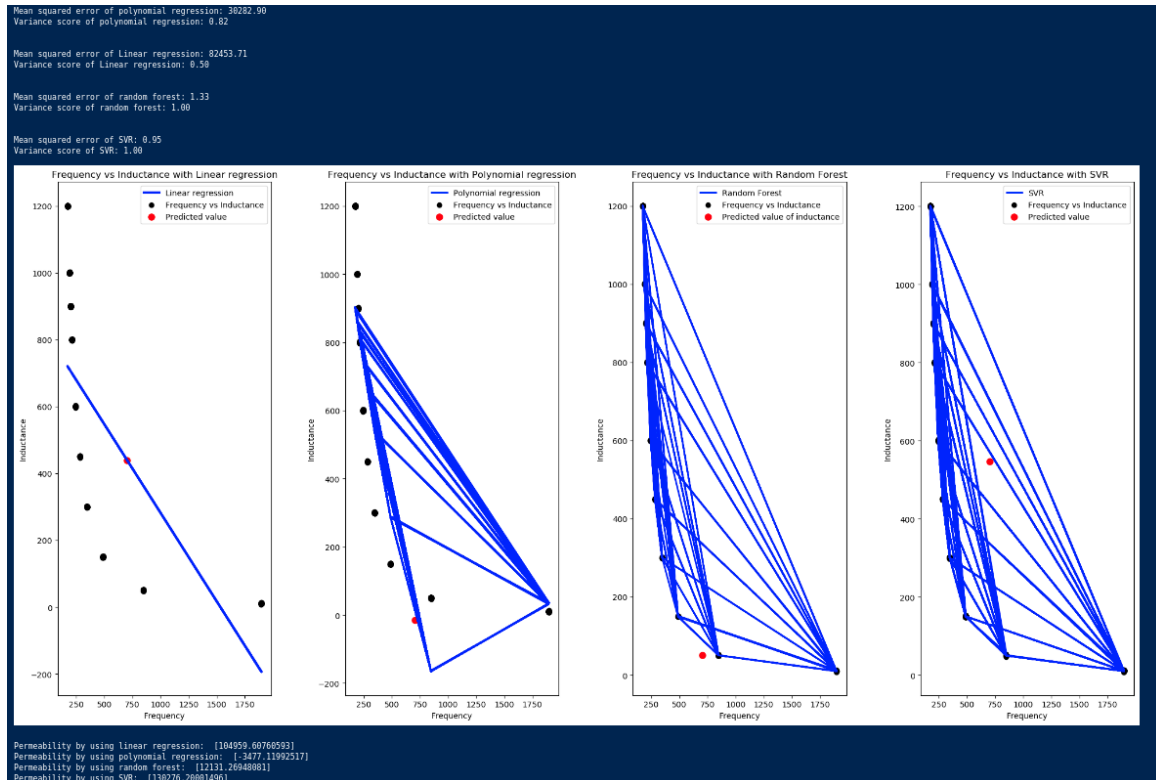


FIGURE 12 – Comparaison des modèles de regression sur un petit dataset

Pour comparer chaque algorithme, l'idée est de mettre en entrée une valeur de fréquence. Divisé en 3 fonctions différentes, il nous permettra d'obtenir une valeur d'inductance. Cette valeur d'inductance sera représentée dans le graphique par un point rouge. Après avoir obtenu l'inductance, un simple calcul nous permettra de déterminer la perméabilité de la fréquence spécifique.

Ainsi, 3 paramètres permettront de comparer la précision des 3 algorithmes : - Erreur quadratique moyenne : L'erreur quadratique moyenne (MSE) est une mesure de la différence entre les valeurs observées d'une variable dépendante et les valeurs prédites de la variable dépendante sur la base d'un modèle statistique. - Score de variance : Le score de variance est une mesure de la précision d'un modèle de régression. Il est calculé comme le rapport de la variance expliquée à la variance totale, et il varie de 0 à 1. Un score de variance de 1 indique que le modèle explique parfaitement la variance dans les données, tandis qu'un score de variance de 0 indique que le modèle n'explique aucunement la variance des données. - Valeur de perméabilité.

Resultats Avec un ensemble de données plus grand, nous pouvons voir beaucoup de différence entre chaque algorithme. Premièrement, la régression linéaire n'est pas vraiment précise. En fait, nous pouvons voir que l'erreur quadratique moyenne est très élevée et que le score de variance est très faible. Donc, la prédiction n'est pas vraiment bonne.

Le deuxième algorithme est la régression polynomiale et son erreur quadratique moyenne est un peu inférieure à la régression linéaire, et sa précision est totalement meilleure. Cet algorithme est plus précis, mais comme on peut le voir sur la figure, la prédiction ne correspond pas avec l'aspect de la courbe generale

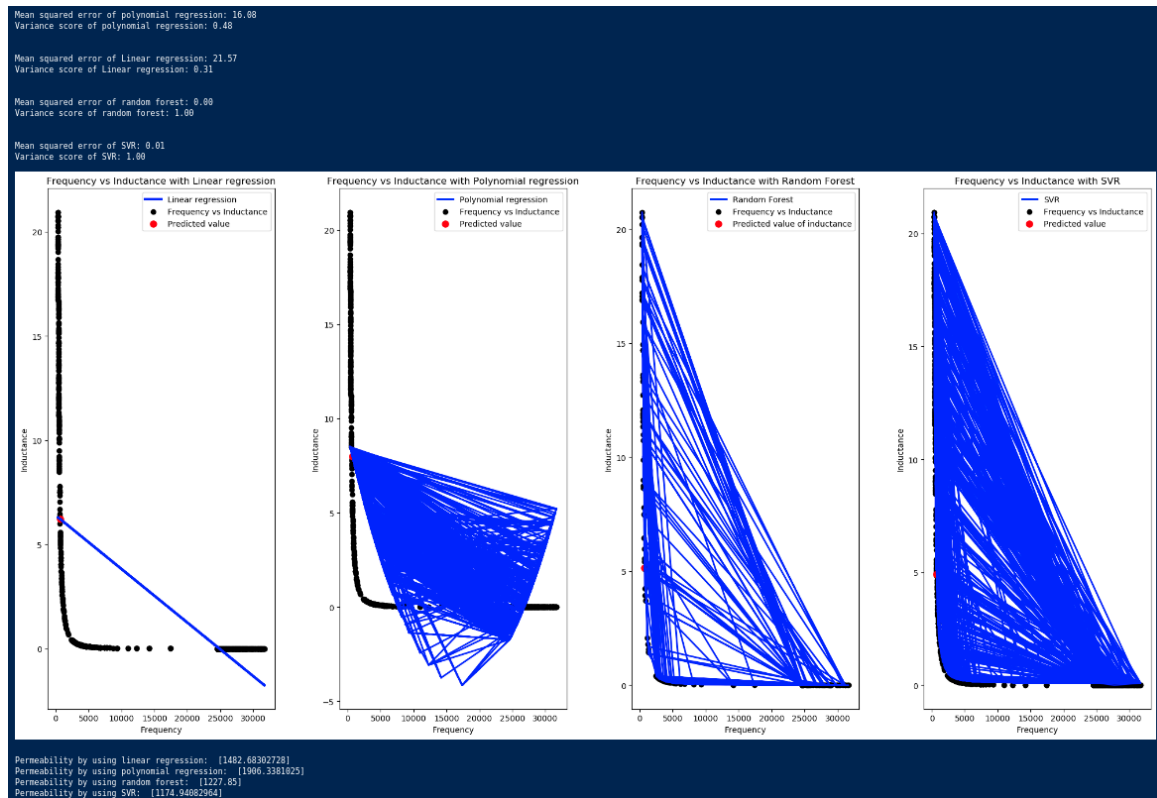


FIGURE 13 – Comparaison des modèles de regression sur un gros dataset

Enfin, les deux derniers algorithmes sont **Random Forest** et **SVR algorithm** (Support Vector Regression). L'erreur quadratique moyenne est la plus faible des 3 algorithmes et le **score de variance est le plus élevé**. Ainsi, la prédiction est la **plus précise**. La prédiction correspond vraiment au graphique réel des données. Entre les deux algorithmes, leurs résultats sont **vraiment proches et aussi leur précision**. Mais, en comparaison entre les résultats de perméabilité et les données réelles à l'intérieur de l'ensemble de données, **L'algorithme de forêt aléatoire est plus proche des données réelles, et donc, le plus précis**.

Prédiction des catégories de regroupement et de perméabilité En utilisant le cluster Kmeans[3] généré ci-dessous, et le meilleur algorithme de prédiction dans notre cas, il est maintenant possible de prédire la catégorie de perméabilité en fonction de la fréquence grâce au Cluster.

	Frequency	Inductance	Permeability	Cluster
0	55280.602547	0.000829	0.629878	3
1	696.841598	5.216415	3964.0	2
2	1154.562092	1.900228	1444.0	2
3	5134.988326	0.096064	73.0	1
4	56500.410081	0.000793	0.602974	3

FIGURE 14 – dataset et cluster prédit

La forêt aléatoire nous permettra de prédire le nombre de cluster lié à la fréquence. En fait, la corrélation entre une fréquence, une inductance, une perméabilité et donc le cluster a été établie récemment. Il suffit de faire la même chose ci-dessus avec l'induction de prédiction, mais maintenant avec le cluster.

Optimisation des paramètres des modèles Dans le cadre de l'optimisation des modèles de régression, nous utilisons différentes techniques pour améliorer leurs performances. Cela implique d'explorer l'espace des hyperparamètres du modèle et de trouver la combinaison optimale qui permettra d'obtenir de meilleures prédictions.

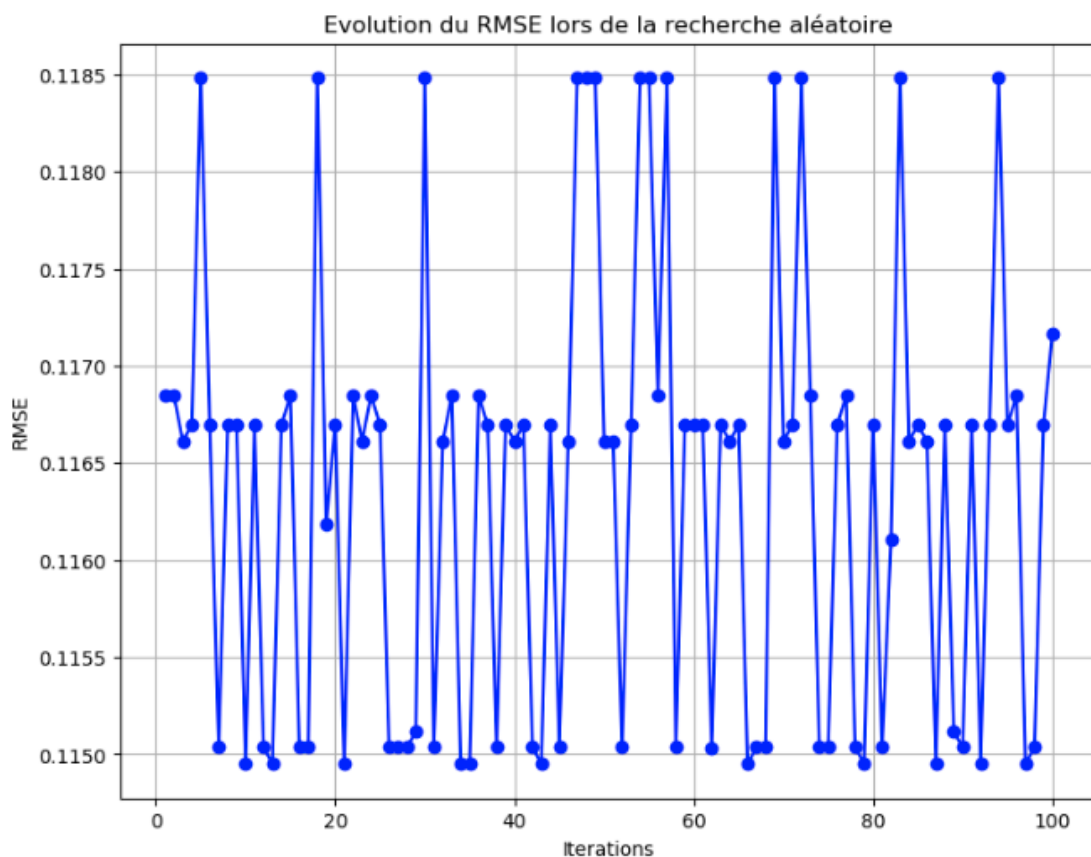


FIGURE 15 – Algorithme de clustering Kmeans sur un dataset d'inductances par rapport à des fréquences

Nous allons utiliser une recherche aléatoire (RandomizedSearchCV) pour explorer différents hyperparamètres du modèle RandomForestRegressor. Nous définissons une distribution de valeurs possibles pour chaque hyperparamètre, puis nous sélectionnons aléatoirement des combinaisons de ces

valeurs et évaluons leurs performances en utilisant la validation croisée avec la métrique de l'erreur quadratique moyenne négative (negative mean squared error). L'objectif est de trouver les hyperparamètres qui minimisent l'erreur quadratique moyenne négative, ce qui correspond à maximiser la performance du modèle.

Par le biais du graphique ci-dessus, celui-ci va représenter l'évolution de ces scores en fonctions des paramètres utilisés. On prendra ainsi la combinaison d'hyperparamètre le plus optimal en prenant le score RMSE le plus faible.

En résumé, l'optimisation d'un modèle de régression implique l'exploration des hyperparamètres pour trouver la meilleure configuration possible. Cela nous permet d'améliorer les performances du modèle et d'obtenir des prédictions plus précises.

Prediction du clustering en fonction d'une fréquence Comme vous pouvez le voir ci-dessous, il y a 2 graphiques représentant la fréquence et l'inductance mais avec un paramètre en plus : la perméabilité. La perméabilité est la cible de l'algorithme de clustering, et la fréquence et l'inductance sont les caractéristiques de l'algorithme de clustering. Concrètement, on peut voir que la perméabilité diminue lorsque la fréquence augmente, et l'inductance diminue. C'est un résultat important car le cluster généré sera précis s'il existe une corrélation entre la perméabilité et la fréquence et l'inductance.

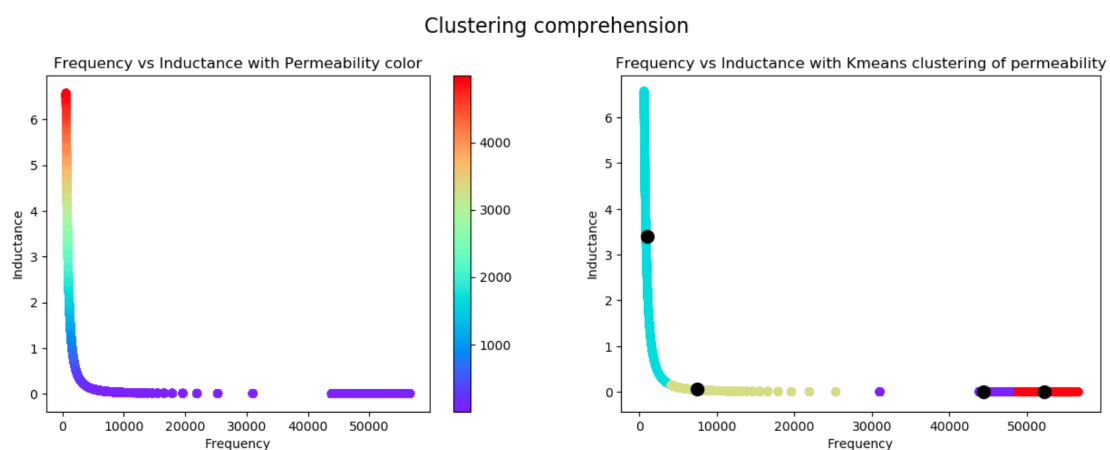


FIGURE 16 – Algorithme de clustering Kmeans sur un dataset d'inductances par rapport à des fréquences

Sur le deuxième graphique, on peut voir les 4 clusters représentant les 4 principales catégories de perméabilité. Grâce à cela, nous pouvons ajouter une colonne supplémentaire dans le jeu de données nommé "Cluster", représentant par 0,1,2 et 3 (non perméable, ferromagnétique, paramagnétique et diamagnétique). Cela sera utile pour la prochaine étape du projet, lorsque nous voudrions prédire la catégorie principale en fonction de la fréquence et de l'inductance.

Prediction du clustering en fonction d'une fréquence donnée Comme vous pouvez le voir ci-dessous, la prédiction est à nouveau très précise, et nous pouvons également trouver tous les différents clusters en fonction cette fois-ci de la fréquence. En point rouge, nous pouvons voir la valeur du cluster pour la valeur de fréquence définie de 45000Hz. On remarque ainsi que la valeur se placera dans le cluster 0, qui regroupe un certain nombre de valeur entre les valeurs de 40000 et 50000 Hz, ce qui reste ainsi entièrement cohérent. D'après les valeurs du dataset, et les cluster générés récemment, ce resultat reste extremement precis. Le problème avec cette methode de prediction et de clustering, est que la precision est plutot variable dependant des cluster et des valeurs de permeabilité et d'autres parametres.

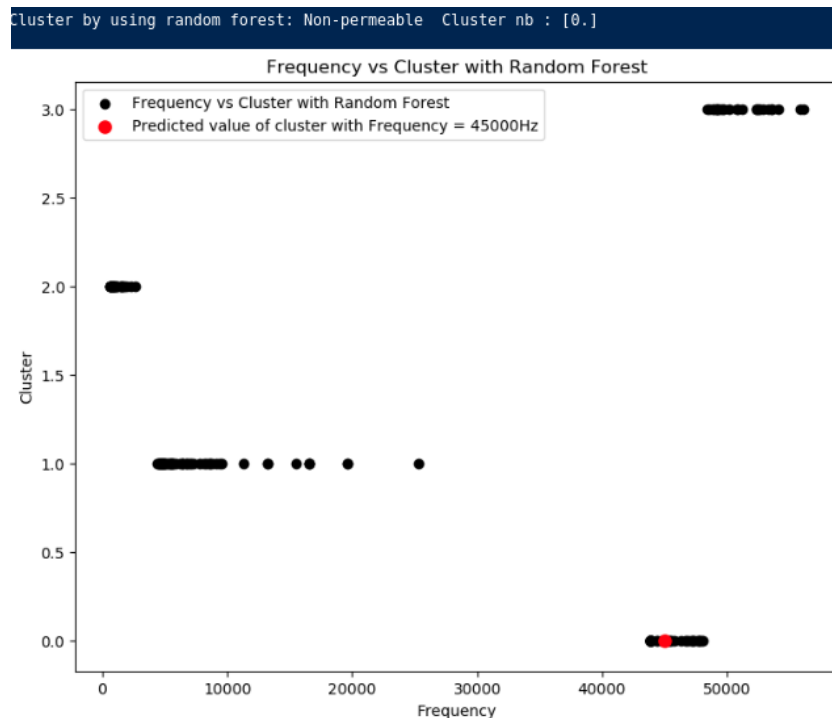


FIGURE 17 – Application de Random Forest afin de prédire un cluster spécifique en prenant en compte une inductance et une fréquence.

Au niveau du cluster 1, peu de valeurs se trouvent entre 10000 et 30000 Hz, ainsi, si la prediction se base sur une valeur de frequence entre 10000 et 30000Hz, la precision sera faible, mais categorisera la valeur de cette frequence en tant que Cluster 1 (soit la categorie Diamagnetique).

Pour conclure, cette methode, grace au resultat obtenu nous permet de prédire avec une valeur de fréquence, la catégorie de perméabilité spécifique du métal.

Dans ce projet, nous avons vu comment générer un jeu de données avec une simulation de détecteur de métaux, et comment utiliser ce jeu de données pour prédire la catégorie de perméabilité du métal en fonction de la fréquence, et cela avec des methodes de regression. Nous avons également vu comment regrouper les données en 4 groupes, et comment prédire le cluster en fonction de la fréquence.

10.1.2 Algorithme de classification

Indicateurs de performance Dans le cadre de la prédiction avec la classification, nous explorons différentes méthodes de classification pour déterminer la perméabilité d'un objet en fonction de sa fréquence et de son inductance perçue. Nous comparons les performances de deux méthodes de classification, à savoir la régression logistique (Logistic Regression) et la forêt aléatoire (Random Forest Classifier).

Dans la première étape, nous entraînons un modèle de régression logistique (LogisticRegression) sur les données de test et effectuons des prédictions. Nous calculons ensuite l'exactitude (accuracy) du modèle en comparant les prédictions avec les valeurs réelles. Nous utilisons également une matrice de confusion (confusion matrix) pour évaluer les performances du modèle, qui montre le nombre de vrais positifs, de faux positifs, de vrais négatifs et de faux négatifs. Enfin, nous générons un rapport de classification (classification report) qui fournit des mesures telles que la précision, le rappel et le score F1 pour chaque classe.

```
Accuracy_LR: 90.75
Confusion Matrix_LR:
[[100   0   0   0]
 [ 26  75   0   0]
 [  3   0 118   0]
 [  0   0   8  70]]
```

FIGURE 18 – Matrice de confusion de la Logistic regression

De manière similaire, nous utilisons un modèle de classification par forêt aléatoire (RandomForestClassifier) et effectuons les mêmes évaluations de performances. Nous analyserons ainsi la méthode la plus optimale comprenant la précision la plus précise. Dans notre cas, on remarque que la matrice de confusion, et surtout la précision du modèle de Random Forest, est largement supérieure à la régression logistique, pour cela, on continuera la suite de l'étude avec ce modèle.

```
Accuracy_RF: 99.75
Confusion Matrix_RF:
[[100   0   0   0]
 [  0 101   0   0]
 [  1   0 120   0]
 [  0   0   0  78]]
```

FIGURE 19 – Matrice de confusion de la Random Forest

La performance presque parfaite de la matrice de confusion de la forêt aléatoire peut être attribuée à plusieurs facteurs.

Premièrement, la forêt aléatoire est capable de modéliser des relations non linéaires et de capturer des interactions complexes entre les variables. Cela lui permet de mieux s'adapter aux données et de générer des prédictions plus précises. En revanche, la régression logistique est limitée par sa nature linéaire et peut avoir du mal à modéliser des relations non linéaires.

Deuxièmement, la forêt aléatoire bénéficie de la combinaison des prédictions de multiples arbres de décision. Chaque arbre peut apporter des informations différentes et contribuer à une meilleure performance globale. La régression logistique, quant à elle, repose sur un seul modèle et peut être plus sensible aux erreurs et aux biais inhérents à ce modèle unique.

Enfin, la répartition et la séparation des données dans le jeu de données peuvent également jouer un rôle dans la performance de la forêt aléatoire. Si les différentes classes de perméabilité sont bien séparées et distinctes dans l'espace des variables, la forêt aléatoire peut mieux les distinguer et générer des prédictions plus précises. Cela peut expliquer pourquoi la matrice de confusion de la forêt aléatoire montre des résultats presque parfaits, avec un nombre élevé de vrais positifs et de vrais négatifs.

Resultats des predictions Dans la deuxième étape, nous affichons trois graphiques qui représentent les clusters de classification en fonction de la fréquence et de l'inductance. Le premier graphique montre les clusters basés sur les classes réelles du jeu de données. Le deuxième graphique représente les clusters prédits par la régression logistique, en prenant en compte les limites perçues par l'algorithme pour la définition des groupes. Enfin, le troisième graphique illustre les clusters prédits par la forêt aléatoire.

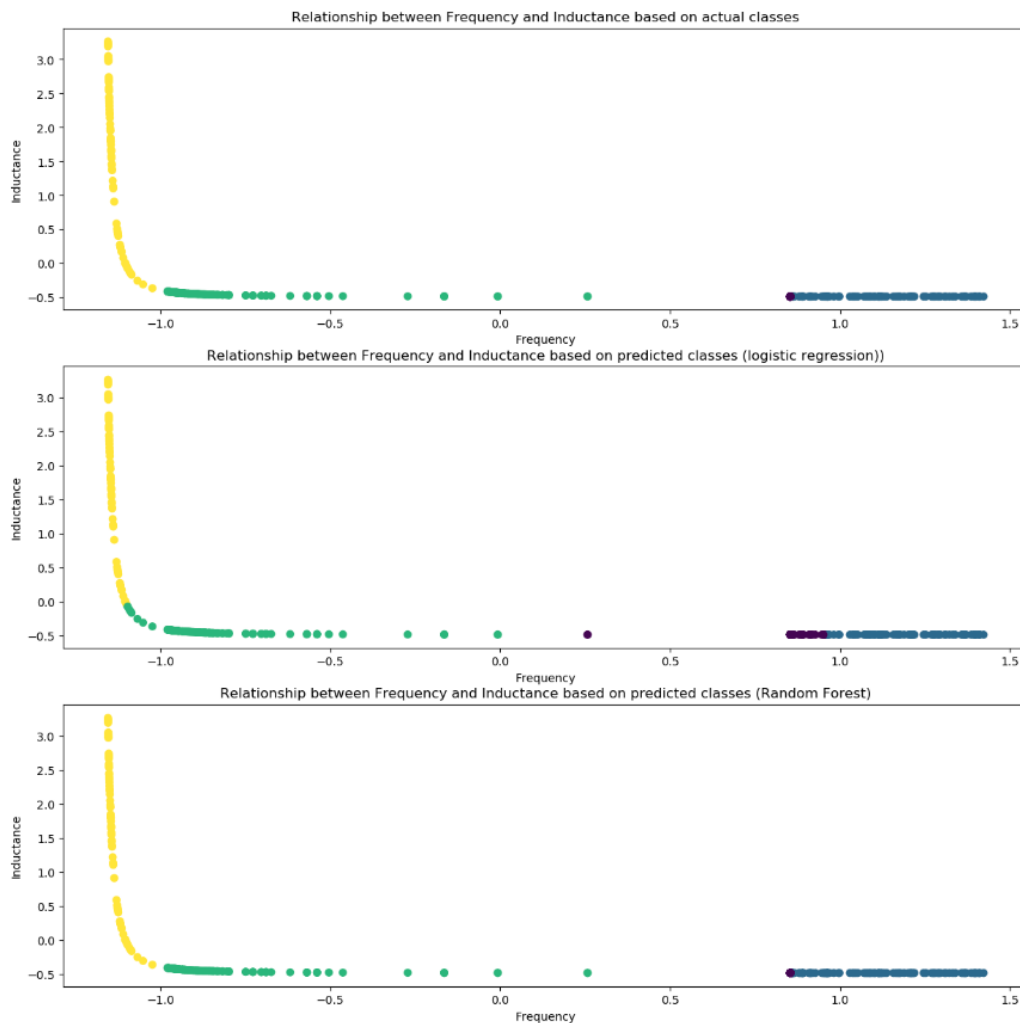


FIGURE 20 – Modèle de random forest permettant de prédire 4 groupes de permeabilité

Ces graphiques permettent de visualiser la performance des modèles de classification et de comparer les résultats. En observant les graphiques, on peut remarquer la proximité entre les clusters réels et les clusters prédits par le modèle de régression logistique. Cependant, le modèle de forêt aléatoire

montre une prédiction presque parfaite, avec une séparation nette des clusters correspondant aux différentes classes de perméabilité.

En résumé, l'utilisation de modèles de classification tels que la régression logistique et la forêt aléatoire permet de prédire la perméabilité en fonction de la fréquence et de l'inductance perçue. Les évaluations de performances, telles que l'exactitude, la matrice de confusion et le rapport de classification, nous aident à comparer les performances des modèles. De plus, les graphiques des clusters permettent de visualiser les résultats des prédictions et de mettre en évidence les performances des modèles de classification.

10.1.3 Simulation d'expérience

Imaginons que chaque dataset de 1000 enregistrements représente une surface donnée, sur laquelle des relevés ont été effectués à l'aide d'un système de détection inductif.

Rappelons que ces 1000 enregistrements, dans le cadre de chaque dataset, et pour chaque dataset ont été générés aléatoirement avec une seed différente à chaque fois.

Chaque surface correspond à un champ ou une prairie d'une même étendue. Les relevés permettent de collecter des paramètres qui sont utilisés pour prédire la perméabilité métallique de chaque dataset.

Dans cette expérience, nous avons comparé 30 terrains, ce qui équivaut à 30 datasets. L'objectif du graphique que nous avons utilisé est de rapidement identifier quel terrain est majoritairement composé d'un cluster spécifique de perméabilité (et donc d'une catégorie métallique). Cette approche permet ainsi de prioriser la dépollution des terrains en fonction de la nature de leur perméabilité.

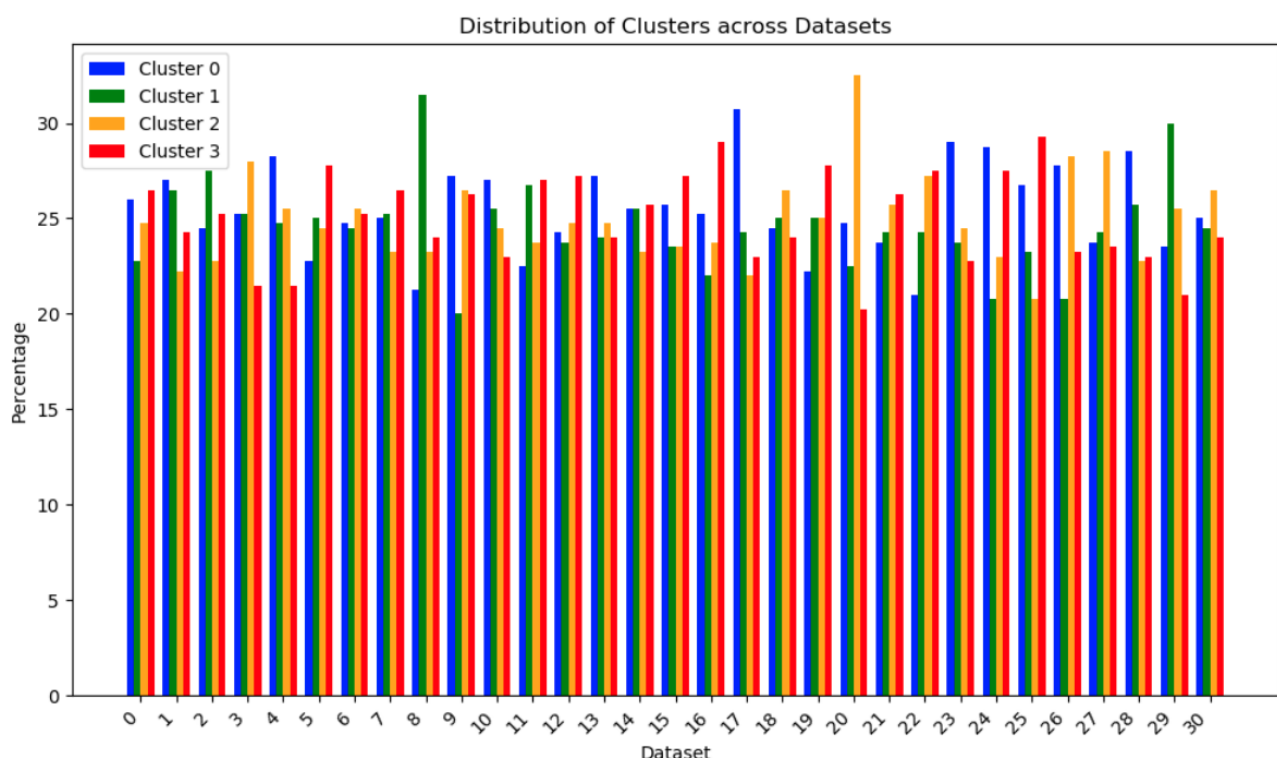


FIGURE 21 – Distribution des clusters de perméabilité à travers les datasets

En effectuant cette étude approfondie, vous pourrez définir rapidement quels sont les terrains présentant les perméabilités les plus problématiques, ce qui facilitera la décision de prioriser leur dépollution. Cette méthode permet ainsi d'optimiser les ressources et d'agir de manière ciblée pour remédier aux problèmes de pollution.

Dans le cadre de cette simulation, on remarque rapidement que le terrain concentrant le plus haut taux de concentration d'une catégorie métallique égale au cluster est le dataset 20. Si on observe que le cluster 2 représente un danger pour l'environnement, on priorisera la dépollution du terrain numéro 20.

Conclusion de l'expérience En utilisant des techniques de classification et de prédiction, notamment l'algorithme de Random Forest, nous avons pu analyser les caractéristiques des terrains et prédire la perméabilité métallique correspondante. Grâce à des graphiques et des statistiques, nous avons pu visualiser la répartition des différents clusters de perméabilité dans chaque terrain.

Les résultats ont montré que certaines surfaces étaient majoritairement composées d'un cluster spécifique de perméabilité métallique, indiquant ainsi une concentration plus élevée de matériaux métalliques dans ces terrains. Cette information est cruciale pour définir les zones les plus problématiques en termes de pollution métallique et pour établir des priorités en matière de dépollution.

En conclusion, cette expérience démontre qu'il est possible de cibler efficacement des clusters spécifiques de perméabilité métallique dans des terrains en utilisant des techniques de prédiction. Cette approche permet une identification rapide des zones présentant les problèmes de pollution les plus importants, facilitant ainsi la prise de décision pour la dépollution et l'allocation des ressources. Elle offre une méthode efficace pour prioriser les actions de dépollution et contribuer à la protection de l'environnement

11 Conclusion

En conclusion, cette étude a exploré l'optimisation de l'analyse de la composition des sols pour estimer le taux de pollution métallique et ainsi prioriser les zones à dépolluer. Dans un contexte de crise climatique, la dépollution des sols revêt une importance capitale pour prévenir les effets néfastes de la pollution métallique sur l'environnement et la santé humaine.

Nous avons mis en œuvre une approche basée sur l'utilisation d'un Système de Détection Inductif, un outil de mesure non destructif permettant de détecter la présence de métaux enfouis dans le sol. En analysant les données acquises par cet outil, nous avons appliqué des techniques de classification et de prédiction pour estimer le taux de pollution et identifier les clusters de perméabilité métallique dans les terrains étudiés.

L'étude a permis de démontrer que la prédiction des clusters de perméabilité métallique est possible en utilisant des méthodes telles que l'algorithme de Random Forest. En visualisant les résultats à l'aide de graphiques et de statistiques, il est devenu possible d'identifier les terrains présentant une concentration élevée de matériaux métalliques, ce qui les rend plus sensibles à la pollution et nécessitant une dépollution prioritaire.

La détection et la gestion des métaux lourds dans les sols revêtent une grande importance, car ces polluants peuvent entraîner des problèmes de santé pour les êtres humains, les animaux et affecter la biodiversité. Les métaux lourds tels que le plomb, le mercure, le cadmium et le chrome sont particulièrement préoccupants en raison de leur toxicité et de leur impact sur la croissance des plantes. Ainsi, cette étude démontre que l'optimisation de l'étude des sols à travers l'acquisition des signaux du système de Détection Inductif et l'analyse des données permet une estimation précise du taux de pollution métallique. Cette approche offre une méthode efficace pour identifier les zones les plus contaminées et prioriser les actions de dépollution, contribuant ainsi à la protection de l'environnement et à la préservation de la santé publique. Il est essentiel de continuer à développer et à mettre en œuvre de telles approches afin de lutter contre la pollution métallique et de garantir un avenir durable pour notre planète.

12 Glossaire

SDI : Système de détection inductif, représentant un détecteur de métaux

Capteurs inductifs : Des dispositifs utilisés dans l'industrie pour détecter des matériaux, mesurer des positions et détecter des défauts. Les signaux provenant de ces capteurs peuvent être non linéaires et soumis à des interférences électromagnétiques.

Techniques de traitement de signal : Méthodes utilisées pour analyser et extraire des informations à partir des signaux inductifs, tels que l'analyse de Fourier, la transformée en ondelettes, la décomposition en composantes indépendantes, etc.

Décomposition en ondelettes : Une technique de traitement de signal qui permet d'extraire des caractéristiques discriminantes à partir de signaux inductifs en analysant leur contenu fréquentiel.

Transformation de Hilbert-Huang : Une méthode alternative d'analyse des signaux inductifs qui a été utilisée avec succès pour la reconnaissance des modes des signaux de détection d'onde dans une plaque métallique.

Apprentissage automatique : Des techniques telles que les réseaux de neurones artificiels et les machines à vecteurs de support qui sont utilisées pour améliorer l'analyse des signaux inductifs et la classification des mines anti-personnel.

IA : L'utilisation de l'intelligence artificielle pour améliorer les performances des systèmes de détection inductifs, par exemple en utilisant des réseaux de neurones artificiels pour la détection des métaux ou des drones équipés de modèles d'apprentissage automatique pour la détection des mines terrestres.

Réseaux de neurones artificiels (ANN) : Des modèles d'apprentissage automatique utilisés pour la détection de métaux dans les capteurs inductifs, permettant une détection précise et une localisation efficace des métaux.

NE555 : Le NE555, plus couramment nommé 555, est un circuit intégré utilisé pour la temporisation

Proteus : Proteus est une suite logicielle destinée à l'électronique.

Inductance : Quotient du flux d'induction créé par un courant (dans un circuit) par l'intensité de ce courant.

Fréquence : Nombre de périodes ou de cycles complets de variations qui se succèdent en une seconde.

Perméabilité : La perméabilité est une aptitude d'un milieu à se laisser traverser par un fluide de forme liquide, gazeuse, ou des ondes.

Champ magnétique : Le terme de champ magnétique désigne une région de l'espace soumise à l'action d'une force provenant d'un aimantaimant.

Bobine : cylindre sur lequel s'enroule un fil conducteur isolé qu'un courant électrique peut parcourir.

Électromagnétisme : Partie de la physique qui traite des relations entre l'électricité et le magnétisme.

Paramagnétique : matériaux paramagnétiques (air, aluminium, magnésium, platine...) et

diamagnétique : matériaux diamagnétiques (argent, cuivre, eau, or, plomb, zinc...),

non magnétique : matériaux ferromagnétiques (cobalt, fer, mu-métal, nickel...).

Dataset : nt couramment utilisés en machine learning. Ils regroupent un ensemble de données cohérents qui peuvent se présenter sous différents formats

Pipeline de préparation : outil permettant de préparer des données et entraîner un modèle Machine

Random Forest : Technique de Machine Learning très populaire auprès des Data Scientists

MSE : C'est une mesure de la qualité d'un estimateur dans le cadre d'une regression

Matrice de confusion : C'est une mesure de la qualité d'un estimateur dans le cadre d'une classification permettant de finir les vrais positifs etc.

Kmeans : K-means (ou K-moyennes) : C'est l'un des algorithmes de clustering les plus répandus.

Clusters : Groupement (d'un petit nombre d'objets).

Références

- [1] **SDI** : Ressources d'un système de détection inductif
- [2] Steven J. Rigatti, MD, DBIM, DABFM **Random Forest**. : RandomForest
- [3] by Mohiuddin Ahmed 1, ORCID,Raihan Seraj **The k-means Algorithm : A Comprehensive Survey and Performance Evaluation** : Kmeans article
- [4] **Proteus** : Proteus logiciel
- [5] **Modèle de classification** Classifier ressources
- [6] **Wavelet ressources** Wavelet
- [7] **Machine Learning** Review of machine learning ressources