

APPRENTISSAGE CONTINU POUR LA CLASSIFICATION DES CONSOMMABLES DE SANTÉ

Mohamed Demes et Taher Lmouden

Encadré par: Massinissa Hamidi

ABSTRACT

Ces dernières années, répondre aux défis économiques liés au recyclage des dispositifs médicaux jetables est devenu de plus en plus crucial, si bien que le processus de recyclage a évolué. Pourtant, les solutions actuelles présentent de nombreux défis, le plus notable étant la forte implication de la main-d'œuvre, menant à des pertes de temps et d'argent, ce qui rend ces solutions inefficaces.

Parmi les solutions innovantes qu'on peut adopter, l'intégration de l'intelligence artificielle et l'Internet des objets (IoT) pour automatiser le processus du recyclage. Notre objectif est de créer un système qui utilise des cameras pour détecter les dechets medicaux recyclable. Cependant, des nouveaux types de dechets apparaissent continuellement, ce qui fait qu'on doit garantir que notre systeme évolue constamment, justifiant le recours à l'approche d'apprentissage continu* pour leur détection, en s'appuyant sur le 'few shot* learning', une méthode qui tire parti d'un petit nombre d'exemples pour entraîner notre modèle à généraliser vers de nouvelles données. D'où l'utilisation des modèles de fondation comme support; des modèles qui servent de base pour l'adaptation vers des tâches spécifiques avec un nombre limité d'exemples, offrant ainsi une méthode efficace pour l'apprentissage avec peu de données.

Mots clés: Foundation models, Internet des objets, Few shots, apprentissage continu.

TABLE DE MATIÈRES

1	Introduction	3
1.1	Objectif du Travail d'étude et de recherche	3
2	étude de l'existant	3
2.1	Modèles de fondation	3
2.2	Apprentissage Continu	4
3	Configuration expérimentale	5
3.1	Dataset description	5
3.2	Description des modèles	6
3.2.1	Réseau de neurones simple	6
3.2.2	ResNet50	6
3.2.3	ViT	6
4	Résultats expérimentaux	7
4.1	Évaluation du réseau de neurones simple	7
4.2	Évaluation des modèles de fondation sans fine-tuning	8
4.2.1	ResNet-50	8
4.2.2	ViT	9
4.3	Évaluation des modèles de fondation après fine-tuning	9
4.3.1	Paramètres figées	10
4.3.1.1	ResNet-50	10
4.3.1.2	ViT	11
4.3.2	Paramètres non-figées	11
4.3.2.1	ResNet-50	12
4.3.2.2	ViT	12
4.4	Exploration des modèles pour l'apprentissage en few-shot*	13
4.4.1	Paramètres figées	14
4.4.2	Paramètres non-figées	15
4.5	Evaluation des modèles dans le cadre d'apprentissage continu	16
4.5.1	Approche de régularisation	16
4.5.2	Approche architecturale	17
5	Conclusion	18
Glossaire		19
References		19

1 INTRODUCTION

En examinant l'évolution constante des quantités de déchets produits par le secteur de la santé, nous pouvons évaluer les impacts économiques des consommables médicaux et l'importance de leur gestion à l'avenir. Selon le Conseil de recyclage des plastiques de santé, les coûts associés aux déchets de soins de santé étaient estimés à 36 milliards de dollars en 2020 à l'échelle mondiale et devraient atteindre 55 milliards de dollars en 2025. La solution actuellement établie pour limiter ces dépenses considérables est le recyclage des consommables de santé. Toutefois, le processus de recyclage comporte plusieurs défis, principalement liés à la dépendance envers l'intervention humaine. Cette dépendance implique des risques d'erreurs émises par l'humain, et des risques de blessures puisqu'ils sont souvent en contact avec des déchets potentiellement dangereux, les exposant ainsi à des effets toxiques et à des maladies contagieuses.

La solution que nous proposons repose sur l'utilisation de l'intelligence artificielle et de l'Internet des Objets (IoT) pour automatiser le processus de recyclage. Cette approche réduit la dépendance humaine en privilégiant la technologie et l'intelligence artificielle, diminuant ainsi les risques pour les ouvriers. Pour atteindre cet objectif, nous envisageons de développer un modèle capable de détecter et de classifier les déchets, afin de prendre des décisions de recyclage adaptées à chaque catégorie. Toutefois, dans un déploiement réel, le système peut être confronté à des déchets qu'il n'a jamais rencontrés auparavant. Pour répondre à cette exigence et concevoir un modèle autonome et durable, nous envisageons d'adopter une approche d'apprentissage continu*. Cette approche vise à créer des modèles capables d'apprendre et de reconnaître de nouveaux types de déchets qui pourraient apparaître à l'avenir, en utilisant un minimum d'échantillons grâce à l'apprentissage "few-shot*". Ils luttent également contre l'oubli catastrophique (Catastrophic forgetting) tout en maintenant leur efficacité dans la détection des déchets déjà connus. Cette adaptation continue garantit que nos systèmes demeurent pertinents et performants. Dans notre solution, nous exploiterons les capacités des modèles de fondation en matière d'apprentissage en few-shot* et de leur adaptation pour contrer l'oubli catastrophique, afin de réaliser un apprentissage continu*.

1.1 OBJECTIF DU TRAVAIL D'ÉTUDE ET DE RECHERCHE

Dans ce travail, nous nous intéressons à la mise en place d'une solution pour la gestion des consommables médicaux basée sur l'intelligence artificielle et l'internet des objets. Nous nous focalisons sur le défi lié à la robustesse et aux capacités d'adaptation des modèles d'apprentissage lorsqu'ils sont déployés dans le monde réel et confrontés, par exemple, à l'apparition de nouvelles classes de consommables et à l'évolution des caractéristiques des classes déjà apprises.

2 ÉTUDE DE L'EXISTANT

2.1 MODÈLES DE FONDATION

Les modèles de fondation (Foundation Models) sont une classe de modèles d'apprentissage automatique qui sont pré-entraînés sur une vaste quantité de données générales avant d'être affinés sur des tâches spécifiques. Ces modèles sont reconnus par leur capacité à générer des performances impressionnantes sur une large gamme de tâches d'apprentissage profond, y compris la compréhension du langage naturel, la vision par ordinateur, et plus encore. Les modèles de fondation reposent sur trois principes fondamentaux. Premièrement, la flexibilité et la généralisation qui signifie que les modèles de fondations sont conçus pour fonctionner sur diverses tâches sans nécessiter de modifications architecturales majeures. Par exemple, le même modèle peut traiter à la fois le texte, l'image, les tâches multimédia, etc.

Deuxièmement, l'échelle (ou "scaling") implique l'utilisation de grandes quantités de données pour l'entraînement du modèle et de grandes ressources technologiques. Cela signifie que les modèles de fondation sont conçus pour fonctionner efficacement sur des ensembles de données massives et à s'adapter à des infrastructures informatiques de grande envergure pour traiter et analyser ces données volumineuses(1).

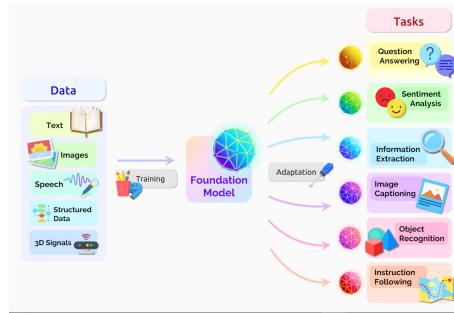


Figure 1: L'adaptation des modèles de fondation à différentes tâches.

Troisièmement, le transfert d'apprentissage (ou "transfer learning" en anglais) est une technique d'apprentissage automatique permettant d'initialiser le modèle avec des connaissances préalablement acquises au lieu de démarrer l'apprentissage à partir de zéro. En d'autres termes, le modèle est pré-entraîné sur une tâche source qui partage des caractéristiques ou des données similaires avec la tâche cible. Ensuite, les connaissances acquises par le modèle, sous forme de poids de réseau neuronal et de représentations apprises, sont transférées à un nouveau modèle destiné à la tâche cible, ce qui peut accélérer le processus d'apprentissage, améliorer les performances du modèle final, et adapter le modèle à des tâches spécifiques avec relativement peu de données supplémentaires pour le fine-tuning(1).

Grâce au transfert d'apprentissage, les modèles de fondation ont démontré une capacité d'apprentissage en few-shot* tout en conservant une précision élevée(6). L'apprentissage en few-shot*, qui signifie la capacité d'un modèle à apprendre à partir d'un minimum d'échantillons, est une technique qui a gagné en popularité ces dernières années, notamment grâce aux avancées dans le domaine de l'apprentissage automatique et de l'intelligence artificielle. Cette approche sera particulièrement pertinente pour notre tâche cible, car les nouvelles données à apprendre pour notre modèle auront des quantités de données limitées. Par conséquent, il est crucial de garantir une précision maximale avec un minimum d'échantillons.

2.2 APPRENTISSAGE CONTINU

L'apprentissage continu*, également appelé "continual learning", vise à développer des algorithmes capables d'acquérir de nouvelles connaissances de manière continue tout en préservant celles déjà acquises. C'est crucial dans les systèmes d'IA qui doivent s'adapter à un flux constant de nouvelles données ou de nouvelles tâches sans oublier les informations précédemment apprises. La formation incrémentielle est au cœur de cette approche, permettant aux modèles d'évoluer progressivement avec de nouvelles données. Cependant, la formation progressive des modèles pose un défi car ils ont tendance à ajuster excessivement les paramètres sur les données actuelles, ce qui peut entraîner l'oubli des informations précédemment apprises. Ce phénomène est appelé "oubli catastrophique" et demeure un problème de recherche ouvert(7). Les techniques d'apprentissage continu* sont utiles dans deux cas principaux :

- Adaptation rapide aux nouvelles données : Parfois, les modèles de machine learning doivent être régulièrement actualisés pour rester efficaces. Par exemple, un modèle détectant la fraude dans les transactions bancaires doit constamment apprendre de nouvelles techniques frauduleuses pour détecter rapidement les activités malveillantes.
- Personnalisation du modèle : Dans certains cas, les besoins de chaque utilisateur peuvent varier. Par exemple, dans un système de classification de documents, chaque utilisateur peut travailler avec des documents utilisant des termes et des styles différents. L'apprentissage continu* permet d'adapter automatiquement le modèle à chaque utilisateur en utilisant ses propres données, assurant ainsi des résultats précis et adaptés à ses besoins spécifiques.

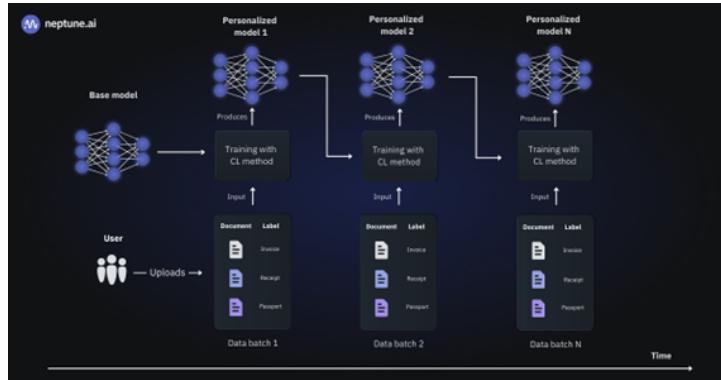


Figure 2: L'apprentissage continu pour la personnalisation du modèle.

On distingue trois principales approches dans l'apprentissage continu.

- Premièrement, les approches architecturales qui consistent à modifier la structure du modèle afin de lui permettre d'apprendre de nouvelles tâches sans oublier les précédentes. Par exemple, dans un scénario où un modèle doit être entraîné sur plusieurs tâches, différentes couches de classification peuvent être ajoutées au modèle de base pour chaque tâche spécifique. Cette approche permet de préserver les connaissances acquises tout en adaptant le modèle aux nouvelles données.
- Deuxièmement, la régularisation qui maintiennent la structure du modèle fixe pendant l'apprentissage, mais utilisent des techniques pour éviter l'oubli catastrophique. Par exemple, la distillation des connaissances consiste à transférer les connaissances d'un modèle pré-entraîné vers un modèle en cours d'apprentissage, afin de conserver les informations importantes, en incluant la modification des fonctions de perte et la sélection des paramètres à mettre à jour lors de l'apprentissage.
- Troisièmement, la méthode d'apprentissage continu basée sur la mémoire, implique de sauvegarder une partie des échantillons d'entrée dans une mémoire tampon pendant la formation, et à les réutiliser lors de l'apprentissage de nouvelles tâches. À titre d'illustration, le modèle sera entraîné sur un lot de données actuelles, ainsi que sur des échantillons sélectionnés de manière aléatoire dans la mémoire tampon. Cette approche permet de maintenir la diversité des données et d'éviter l'oubli catastrophique.

3 CONFIGURATION EXPÉRIMENTALE

3.1 DATASET DESCRIPTION

Le jeu de données utilisé pour les différentes expériences mises en œuvre est intitulé "Medical-Waste-4.0-Dataset", créé par la Région de Toscane. Il a été élaboré dans le but de constituer une ressource précieuse pour la conception et le test de méthodes de vision par ordinateur dédiées au tri primaire des déchets médicaux. Ce jeu de données contient 4245 images de résolution 1920 x 1080 réparties en 13 classes. Un échantillon représentatif de chaque classe est illustré dans la figure 3.



Figure 3: Echantillons de chaque classe du jeu de données.

Afin d'explorer les performances des modèles de base et les comparer à notre modèle de réseau de neurones simple, nous avons opté pour la division de notre jeu de données en trois ensembles distincts : l'ensemble d'entraînement, représentant 70% des échantillons (soit 2967 images), l'ensemble de validation, comprenant 15% des échantillons (soit 629 images), et enfin l'ensemble de test, comprenant 15% (soit 649 images).

3.2 DESCRIPTION DES MODÈLES

3.2.1 RÉSEAU DE NEURONES SIMPLE

Le réseau de neurones convolutif présenté est un modèle simple conçu pour la classification d'images. Il se compose de blocs de convolution suivis de couches de pooling pour extraire des caractéristiques, puis ces caractéristiques sont aplatis et transmises à des couches entièrement connectées pour la classification. La fonction d'activation ReLU est utilisée pour introduire de la non-linéarité dans le modèle. Ce modèle a été utilisé pour comparer la robustesse à l'apprentissage continu d'un modèle simple à celle des modèles de fondations.

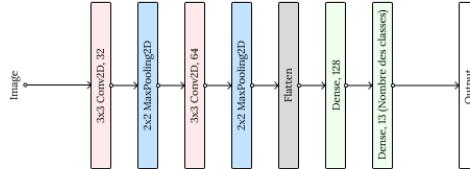


Figure 4: Architechture du modèle de réseau de neurones simple.

3.2.2 RESNET50

ResNet-50, un modèle de réseau neuronal convolutif (CNN) profond développé par Microsoft Research, fait partie de la famille des réseaux résiduels (ResNet), qui utilisent des connexions résiduelles pour faciliter l'entraînement de réseaux plus profonds. Avec ses 50 couches de convolution, ResNet-50 totalise 25,6 millions de paramètres. Ce modèle a été introduit dans le papier de recherche "Deep Residual Learning for Image Recognition"(4). Il est pré-entraîné sur le dataset ImageNet-1k, qui comprend 1 million d'images de résolution 224 x 224 réparties en 1000 classes. ResNet-50 est largement utilisé pour des tâches telles que la classification d'images et la détection d'objets, grâce à sa capacité à extraire des caractéristiques complexes à partir de données visuelles.

3.2.3 ViT

ViT, ou Vision Transformer, est un modèle révolutionnaire dans le domaine de la vision par ordinateur, développé par Google, qui applique les principes des transformateurs, à l'origine développés pour le traitement du langage naturel, à l'analyse d'images. Contrairement aux architectures CNN traditionnelles, ViT traite les images comme des séquences de jetons, ce qui lui permet de capturer des relations à longue portée entre les pixels. Avec ses 86,6 millions de paramètres, ViT a été introduit pour la première fois dans le papier de recherche intitulé "An Image is Worth 16x16

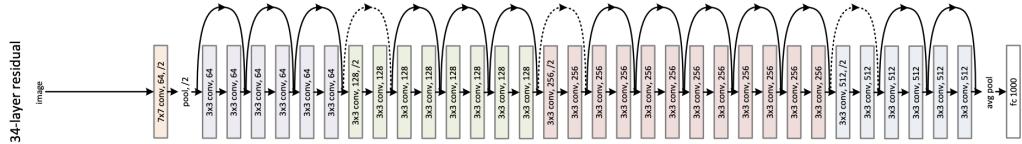


Figure 5: Architecture du modèle ResNet-50.

Words: Transformers for Image Recognition at Scale”(3). Il a été pré-entraîné sur une large collection d’images de manière supervisée, notamment ImageNet-21k(2), avec 14 millions d’images de résolution de 224x224 pixels réparties en 21,843 classes. Ensuite, le modèle a été fine-tuné sur ImageNet 2012, un ensemble de données comprenant 1 million d’images et 1000 classes, également à une résolution de 224x224 pixels.

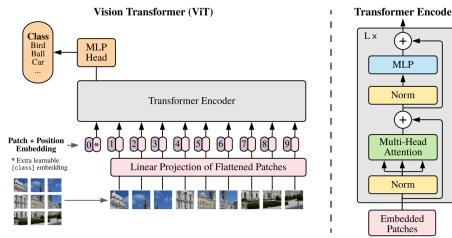


Figure 6: Architecture du modèle ViT.

4 RÉSULTATS ÉXPERIMENTAUX

4.1 ÉVALUATION DU RÉSEAU DE NEURONES SIMPLE

Dans un premier temps, nous avons évalué un modèle de réseau de neurones simple. Nous l’avons entraîné sur notre jeu de données d’entraînement pendant 10 époques, afin de l’utiliser comme référence de base pour la comparaison avec d’autres modèles fondamentaux. Ensuite, nous avons testé ce modèle sur notre ensemble de test pour évaluer ses performances. Les résultats obtenus ont été présentés dans les figures 7 et 8.

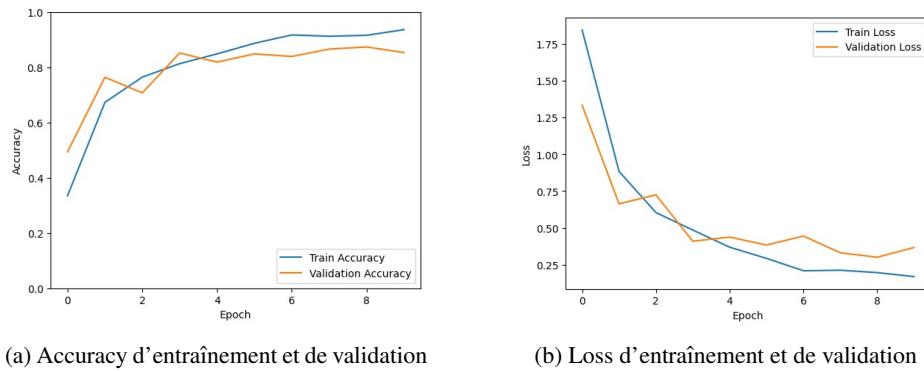


Figure 7: Accuracy et loss du modèle de réseau de neurones simple.

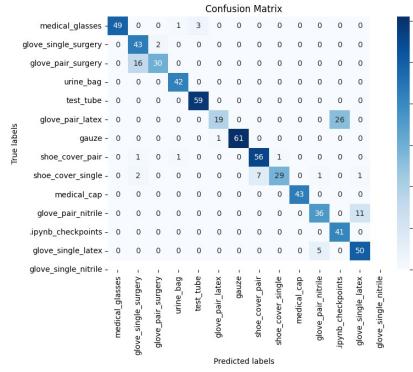


Figure 8: Matrice de confusion sur l'ensemble ‘Testing’ du modèle de réseau de neurones.

En examinant les graphiques d’accuracy et de perte sur les ensembles d’entraînement et de validation, nous avons constaté que le modèle atteignait de bonnes performances en dixième époque, avec une accuracy de 0.86 et une perte de 0.38 sur l’ensemble de test. Cette observation a été confirmée par la matrice de confusion, qui a témoigné de la capacité du modèle à prédire correctement les échantillons de test.

4.2 ÉVALUATION DES MODÈLES DE FONDATION SANS FINE-TUNING

Dans cette expérience, nous explorons la capacité de prédiction des modèles de fondation afin de déterminer s’ils peuvent correctement identifier nos classes. Pour obtenir des résultats plus tangibles, nous avons ajouté une catégorie nommée ’false prediction’ à la matrice de confusion. Cette catégorie indique que le modèle a attribué des classes qui ne correspondent pas à celles du dataset utilisé.

4.2.1 RESNET-50

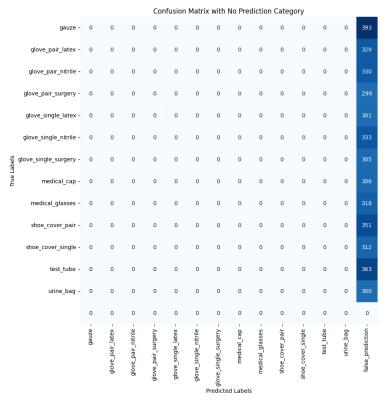


Figure 9: Matrice de confusion pour le jeu de données utilisant le modèle ResNet-50 non fine-tuné.

D’après les résultats de la matrice de confusion, nous avons remarqué que le modèle prédit l’intégralité de notre ensemble de données en assignant des classes qui ne correspondent pas aux classes réelles du dataset. Ces prédictions erronées sont toutes répertoriées comme des prédictions de la classe ”false predictions”. Cela explique les résultats obtenus en termes de précision de 0.0 et de perte de 15.57.

4.2.2 ViT

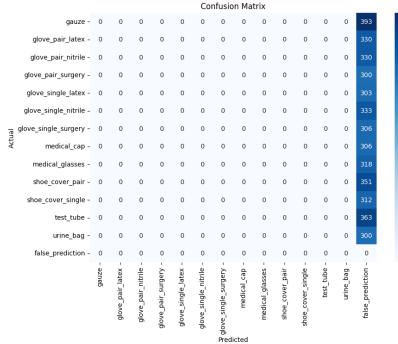


Figure 10: Matrice de confusion pour le jeu de données utilisant le modèle ViT non fine-tuné.

Similairement au ResNet-50, le modèle ViT sans affinement n'a réussi à prédire correctement aucune des images de notre dataset. En conséquence, nous avons obtenu une accuracy de 0.0 et un taux de perte de 2.56.

Ces résultats presque nuls étaient attendus, étant donné que ni ResNet-50 ni ViT ne reconnaissent les classes spécifiques de notre dataset. Dans le but d'obtenir des résultats plus représentatifs de la réalité, nous envisageons de finetuner ces deux modèles. Cette étape d'affinement (finetuning) est cruciale pour évaluer de manière plus précise et concrète leurs performances sur notre dataset. En ajustant les modèles pour mieux s'adapter aux spécificités de nos données, nous espérons améliorer significativement leur capacité de prédiction et obtenir des indicateurs de performance plus favorables.

4.3 ÉVALUATION DES MODÈLES DE FONDATION APRÈS FINE-TUNING

Dans cette expérience, nous allons effectuer un ajustement fin (fine-tuning) de modèles pré-entraînés avec 10 époques, en testant à la fois avec des paramètres figés et non figés, pour mieux adapter leurs poids aux caractéristiques spécifiques de notre jeu de données. Cette approche vise à optimiser les performances des modèles pour la détection de consommables médicaux.

Pour le modèle ResNet-50, nous avons retiré la dernière couche entièrement connectée (FC) du modèle de base et l'avons remplacée par une nouvelle couche 'Sequential' dotée de 13 sorties, correspondant au nombre de classes dans notre dataset. De plus, nous avons ajouté une couche Conv2D au début pour un prétraitement léger des données. Pour le modèle Vision Transformer (ViT), nous avons enrichi le modèle de base avec quatre couches supplémentaires : une couche de Dropout, qui désactive 50% des neurones de la couche précédente pour prévenir le surapprentissage ; une couche linéaire (Linear) pour transformer les caractéristiques ; une fonction d'activation ReLU pour introduire de la non-linéarité ; et une dernière couche de classification (Classifier), qui assure que la sortie correspond au nombre de classes dans le dataset. Les architectures modifiées adoptées pour ces expérimentations sont illustrées dans les Figures 11 et 12.

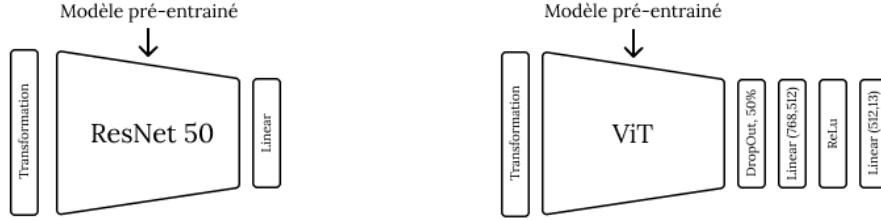


Figure 11: Architecture de ResNet-50 pour le fine-tuning

Figure 12: Architecture de ViT pour le fine-tuning

4.3.1 PARAMÈTRES FIGÉES

Pour ces expériences, nous avons gelé les paramètres internes des modèles, ne laissant que les paramètres des dernières couches ajoutées ajustables lors de l'entraînement. L'ajustement se fait par optimisation basée sur le calcul de la perte, ce qui signifie que les modèles sont entraînés pour minimiser la différence entre leurs prédictions et les vraies étiquettes des données, permettant ainsi une amélioration progressive de leurs performances dans la tâche spécifique.

4.3.1.1 ResNet-50

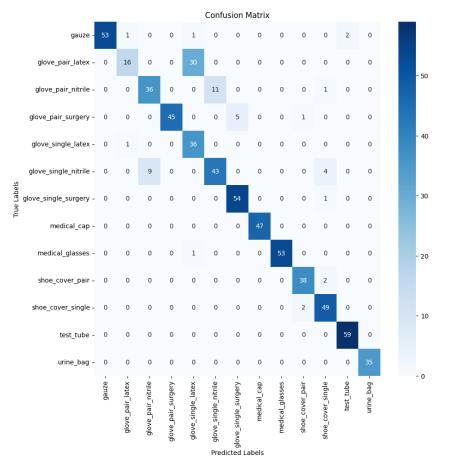
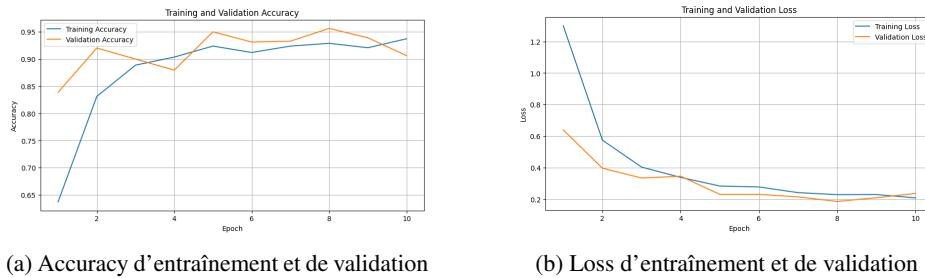


Figure 14: Matrice de confusion sur l'ensemble ‘Testing’ du modèle ResNet-50 finetuné avec paramètres figées.

En analysant les graphiques d'accuracy et de perte pour un modèle ResNet avec des paramètres figés, nous observons que le modèle atteint une performance optimale à la dixième époque. L'accuracy sur l'ensemble de test est de 88.67% et la perte est de 0.28. Ces résultats sont corroborés par la matrice de confusion, qui montre que le modèle prédit efficacement les échantillons de l'ensemble de test.

4.3.1.2 ViT

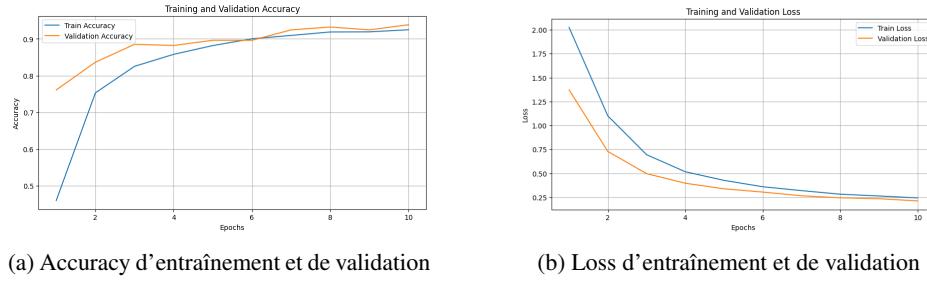


Figure 15: Accuracy et loss du modèle ViT finetuné avec paramètres figées.

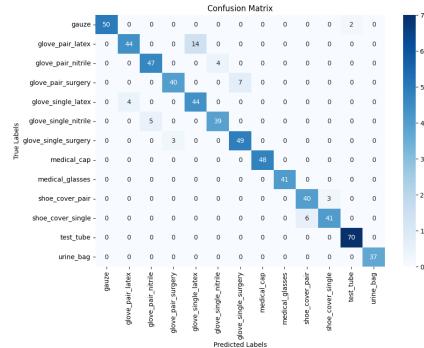
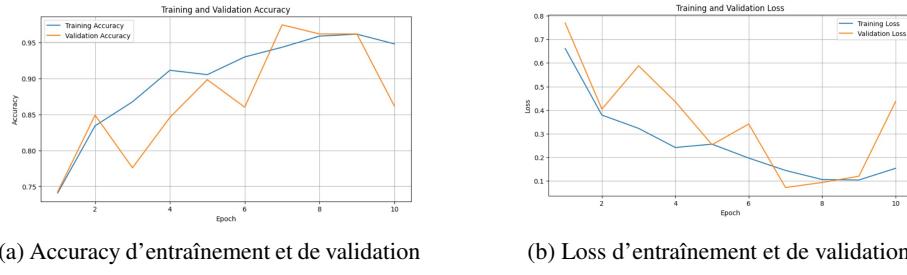


Figure 16: Matrice de confusion sur l'ensemble ‘Testing’ du modèle ViT finetuné avec paramètres figées.

Le modèle ViT finetuné avec paramètres figées a également atteint son apogée à la dixième époque. À ce stade, il a démontré une précision de 92.47% et une perte de 0,23 sur l'ensemble de test, marquant ainsi une performance significative. Ce que nous pouvons voir d'après la matrice de confusion.

4.3.2 PARAMÈTRES NON-FIGÉES

Pour les expériences avec les paramètres non figés, nous avons choisi d'ajuster l'ensemble des paramètres du modèle pendant l'entraînement. Contrairement au fine-tuning où seules les dernières couches sont ajustées, cette approche permet une adaptation complète du modèle à notre jeu de données spécifique. Ainsi, lors de l'optimisation basée sur la perte, tous les paramètres du modèle sont modifiés afin de minimiser la divergence entre les prédictions du modèle et les étiquettes réelles des données.



(a) Accuracy d’entraînement et de validation (b) Loss d’entraînement et de validation

Figure 17: Accuracy et loss du modèle ResNet-50 finetuné avec paramètres non-fixés.

4.3.2.1 ResNet-50

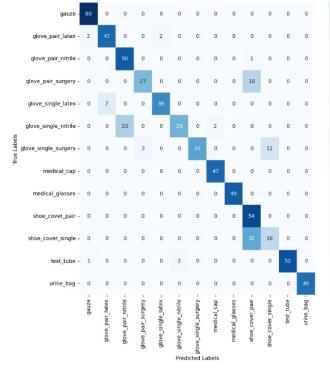
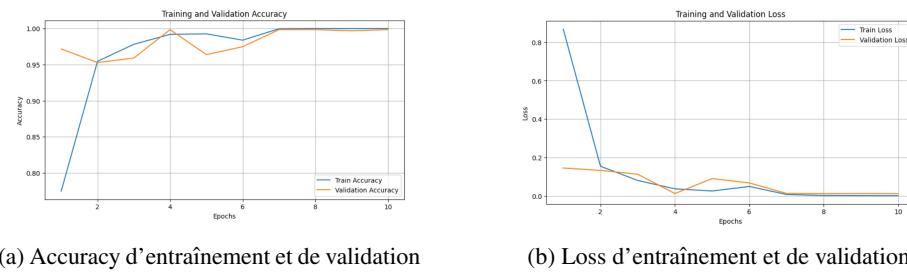


Figure 18: Matrice de confusion sur l’ensemble ‘Testing’ du modèle ResNet-50 finetuné avec paramètres non-fixés.

Après un ajustement fin (fine-tuning) avec des paramètres non figés, le modèle ResNet-50 a montré des fluctuations non monotones dans les graphiques d’accuracy et de perte. Contrairement à d’autres modèles, celui-ci a atteint le maximum de ses performances durant la septième époque. Après avoir évalué le modèle de la septième époque, il a enregistré une précision de 97.15% et une perte de 0.11 sur l’ensemble de test, démontrant ainsi sa capacité à extraire efficacement les caractéristiques des échantillons.

4.3.2.2 ViT



(a) Accuracy d’entraînement et de validation (b) Loss d’entraînement et de validation

Figure 19: Accuracy et loss du modèle ViT finetuné avec paramètres non-fixés.

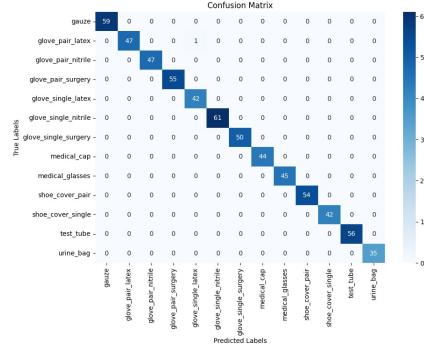


Figure 20: Matrice de confusion sur l’ensemble ‘Testing’ du modèle ViT finetuné avec paramètres non-figées.

Dans cette étude, nous avons employé un modèle Vision Transformer (ViT) avec des paramètres non figés. Nos résultats démontrent que ce modèle prédit avec une efficacité exceptionnelle les échantillons de l’ensemble de test. Nous avons notamment observé une accuracy presque parfaite de 99.84% et une perte extrêmement faible de 0.008. Ces performances remarquables sont également reflétées dans la matrice de confusion, qui indique que le modèle a correctement prédit presque tous les échantillons testés. Cette excellente capacité de généralisation du modèle ViT dans cette configuration dynamique est ainsi clairement mise en évidence.

4.4 EXPLORATION DES MODÈLES POUR L’APPRENTISSAGE EN FEW-SHOT*

Après avoir évalué les performances de nos modèles suite à un fine-tuning et constaté leur efficacité sur notre jeu de données, nous souhaitons désormais explorer leur capacité d’apprentissage en few-shot*. Cette approche consiste à utiliser le minimum d’échantillons nécessaires pour apprendre, ce qui est crucial pour l’intégration de nouveaux types de déchets pour lesquels les données disponibles sont limitées. Pour ce faire, nous envisageons d’entraîner nos modèles avec différentes fractions des données d’entraînement (10%, 20%, 30%, etc.) et d’évaluer leurs performances après chaque session d’entraînement sur l’ensemble de test. Les détails de cette répartition, incluant le nombre d’échantillons par classe et le total d’échantillons, en fonction du pourcentage de données utilisé, sont présentés dans le tableau 2. Cela nous permettra de comparer l’efficacité des modèles de fondation par rapport à un modèle de réseau de neurones simple, ainsi que de déterminer la quantité de données nécessaire pour atteindre leur performance optimale.

Afin de garantir une distribution équilibrée des classes dans notre jeu de données durant l’apprentissage, nous avons choisi de maintenir un nombre égal d’échantillons par classe pour chaque fraction de données (de 10% à 90%). Comme illustré dans la figure 21, où nous avons représenté la quantité d’échantillons par classe en fonction de la quantité totale de données d’entraînement, cette stratégie assure une représentation équilibrée. Pour la dernière fraction, nous avons utilisé tous les échantillons disponibles de chaque classe, compte tenu de la distribution initialement inégale des échantillons par classe. Cette méthode prévient le risque que notre modèle favorise l’apprentissage d’une classe au détriment des autres, un risque particulièrement élevé avec de petites portions de données où un ou deux échantillons peuvent significativement influencer les résultats.

Table 1: Nombre d’images d’entraînement en fonction de la quantité de données.

Quantité de données	10%	20%	30%	40%	50%	60%	70%	80%	90%
Nbr d’images par classe	21	42	63	84	105	126	147	168	189
Nbr total d’images	273	546	819	1092	1365	1638	1911	2184	2457

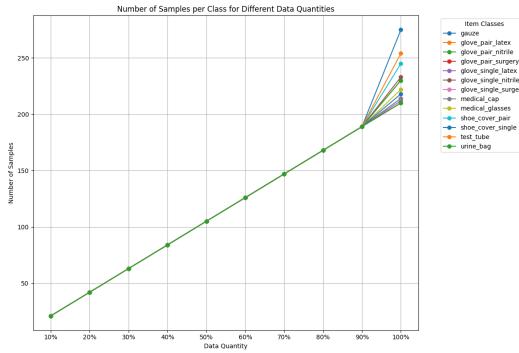


Figure 21: Répartition des échantillons par classe en fonction de la quantité de données

4.4.1 PARAMÈTRES FIGÉS

Les modèles de fondation avec paramètres figés sont entraînés avec des échantillons de données augmentant progressivement en taille, dans le but de comparer leur capacité d'apprentissage avec celle de réseau de neurones simple en utilisant de petites quantités de données.

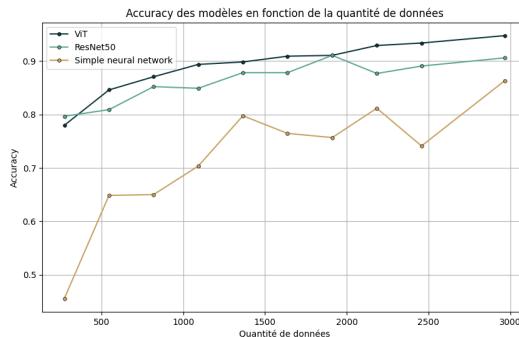
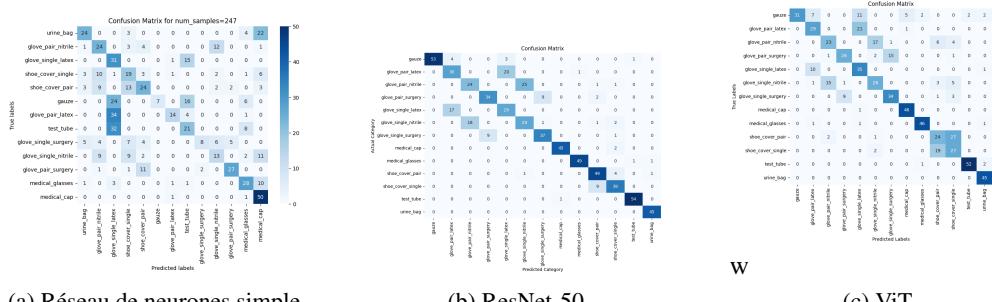


Figure 22: Accuracy des modèles en fonction du quantité de données.



(a) Réseau de neurones simple

(b) ResNet-50

(c) ViT

Figure 23: Matrices de confusion des modèles entraînés sur 273 échantillons et testés sur l'ensemble de test.

Les résultats présentés dans la figure 22 illustrent une comparaison des précisions entre les trois modèles. Nous observons une supériorité notable des modèles de fondation par rapport au modèle simple : en effet, alors que le modèle simple, entraîné sur 273 images, a atteint une précision de 45.52% sur l'ensemble de test, le ResNet-50 a obtenu 82.58% et le ViT a réalisé des performances exceptionnelles avec une précision de 90.29%. Il apparaît que ResNet-50 et ViT requièrent seulement 273 images pour parvenir à de bonnes performances, tandis que le modèle simple n'a surpassé les 80% de précision qu'avec 80% des données, représentant 2184 images. Cette disparité confirme l'efficacité supérieure des modèles de fondation pour l'apprentissage en few-shots* comparativement aux modèles plus simples.

4.4.2 PARAMÈTRES NON-FIGÉES

La même procédure a été effectuée avec les modèles de fondation dont les paramètres ne sont pas figés, afin de comparer leur capacité d'apprentissage à celle d'un réseau de neurones simple en s'entraînant avec un faible volume de données.

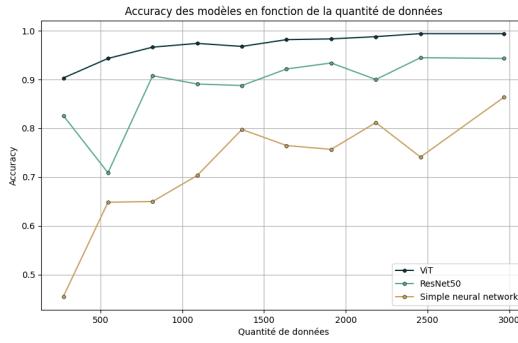


Figure 24: Accuracy des modèles en fonction du quantité de données.

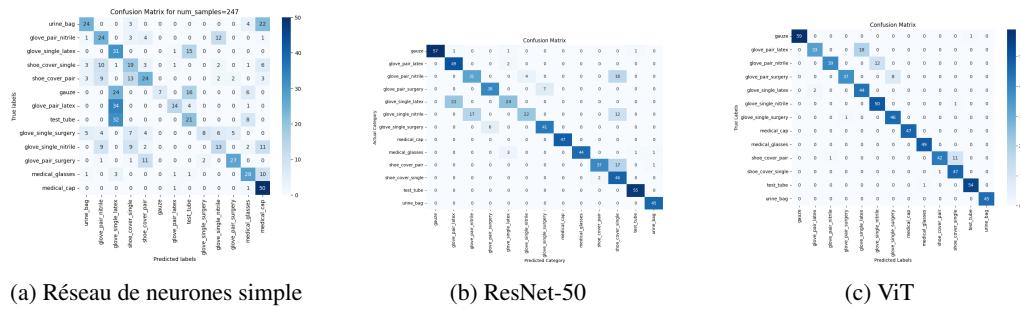


Figure 25: Matrices de confusion des modèles entraînés sur 273 échantillons et testés sur l'ensemble de test.

Tout comme les modèles de fondation avec des paramètres figés, ceux avec des paramètres non figés démontrent une supériorité par rapport au modèle simple. Par exemple, le ResNet-50 atteint une précision de 82,58% avec seulement 273 échantillons, tandis que le ViT affiche des performances exceptionnelles avec une précision de 90,29%, également avec 273 échantillons. Cette supériorité confirme l'efficacité des modèles de fondation lorsqu'ils sont utilisés avec un nombre minimal d'échantillons et valide leur pertinence pour notre tâche cible.

Table 2: Conclusion: accuracy de toutes les modèles en fonction de la quantité de données.

Quantité de données	10%	20%	40%	50%	70%	80%	90%	100%
Réseau de neurones	45.52%	64.83%	70.32%	79.74%	76.45%	81.16%	74.09%	86.34%
ResNet-50 non figée	82.58%	70.87%	89.06%	88.75%	93.37%	89.98%	94.45%	94.29%
ViT non figée	90.29%	94.29%	97.38%	96.76%	98.30%	98.76%	99.38%	99.38%
ResNet-50 figée	90.29%	94.29%	97.38%	96.76%	98.30%	98.76%	99.38%	99.38%
ViT figée	90.29%	94.29%	97.38%	96.76%	98.30%	98.76%	99.38%	99.38%

4.5 EVALUATION DES MODÈLES DANS LE CADRE D'APPRENTISSAGE CONTINU

Après avoir confirmé la capacité des modèles de fondation en apprentissage Few-Shots*, nous avons évalué leur capacité à apprendre de manière continue. Pour ce faire, nous avons entraîné nos modèles de manière incrémentale : d'abord sur un ensemble de 7 classes du jeu de données, puis ajouté progressivement 2 classes supplémentaires à chaque itération jusqu'à atteindre un total de 13 classes. Conformément aux recommandations de l'article 'Continual Learning: Methods and Application' (7), nous avons débuté par l'approche de régularisation, que nous avons implémentée en utilisant la méthode de distillation. Ensuite, nous avons opté pour l'approche architecturale, en explorant notamment les architectures dynamiques.

4.5.1 APPROCHE DE RÉGULARISATION

Dans le cadre de l'approche de régularisation, nous avons implémenté une approche connue sous le nom de "Learning Without Forgetting" (LWF)(5), pour transférer les connaissances des modèles précédents vers les modèles actuels tout en continuant à apprendre de nouvelles données. Pour ce faire, nous avons utilisé une fonction de perte de distillation qui calcule la divergence Kullback-Leibler (KL) entre les probabilités softmax des sorties du modèle actuel et celles des sorties du modèle précédent. Cette perte est pondérée avec la perte de la cross-entropy traditionnelle pour former une perte totale. L'objectif est de permettre au modèle actuel d'imiter les prédictions du modèle précédent tout en adaptant ses poids pour mieux s'adapter aux nouvelles données, comme décrit dans l'approche LWF. Cette approche de distillation est essentielle dans le cadre de l'apprentissage continu, car elle permet aux modèles de conserver les connaissances précédemment acquises tout en s'adaptant à de nouveaux exemples de manière progressive et efficace. Les准确ies présentées dans la figure 27 sont calculées à chaque étape en utilisant un ensemble de test qui contient l'ensemble des classes que le modèle a apprises jusqu'à cette étape.

Dans notre approche, nous avons réservé l'application de la distillation aux modèles dont les paramètres étaient figés, puisqu'ils avaient présenté les meilleures performances en termes de prédiction dans les étapes précédentes.

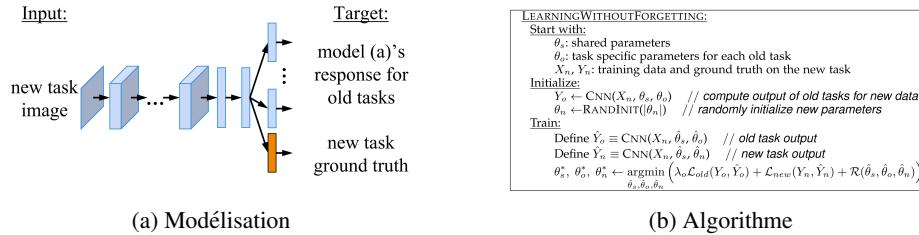


Figure 26: Architecture du 'Learning without forgetting'.

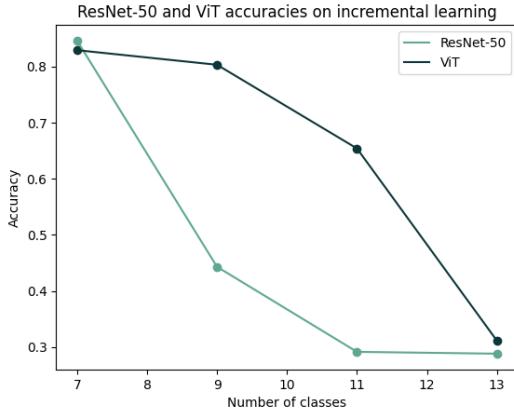


Figure 27: Accuracy des modèles en apprentissage incrémental (Approche de régularisation).

Lors de l'utilisation de la régularisation par distillation, nous avons constaté une différence notable entre le modèle Vision Transformer (ViT) et le modèle ResNet50. Contrairement à ResNet50, où l'ajout de nouvelles classes entraîne souvent une chute de précision dès les premières classes supplémentaires, la précision de ViT a montré une dégradation plus graduelle au fur et à mesure de l'ajout de classes. Cependant, malgré cette tendance, les performances globales de ViT n'ont pas été aussi élevées que celles de ResNet50. Par exemple, en arrivant à 13 classes, la précision justifiée par la méthode de distillation pour ViT était de 31.12%, tandis que ResNet50 atteignait une précision de 28.81% dans des conditions similaires.

4.5.2 APPROCHE ARCHITECTURALE

Dans cette expérience, nous avons utilisé une approche architecturale visant à adapter le modèle ResNet50 pour l'apprentissage continu avec un nombre variable de classes, en ajoutant des modules dédiés pour chaque ensemble supplémentaire de classes. Chaque fois que nous avons eu besoin d'ajouter de nouvelles classes, nous avons créé un module supplémentaire, parallèle aux modules déjà existants, avec une sortie correspondant au nombre de classes ajoutées. Ces nouveaux modules comportent des couches linéaires adaptées pour intégrer les caractéristiques spécifiques des nouvelles classes. En reliant ces modules directement au ResNet fine-tuner sur l'ensemble des classes initiales, à chaque étape, nous figeons tous les paramètres du modèle, à l'exception de ceux du dernier module ajouté, afin d'optimiser seulement les paramètres responsables des nouvelles classes.

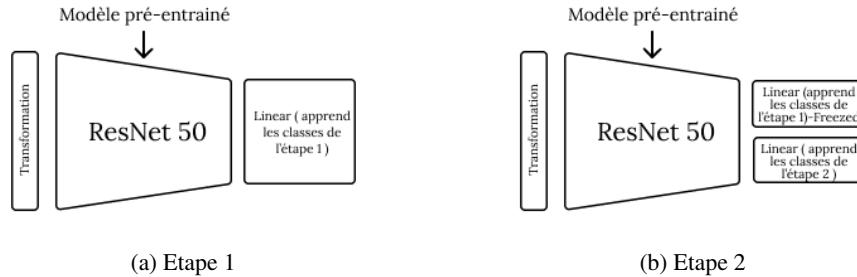


Figure 28: Architecture du modèle resNet-50 au différentes étapes de l'apprentissage.

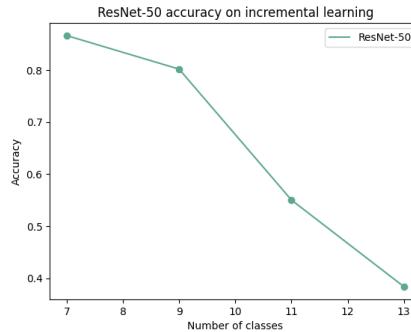


Figure 29: Accuracy du ResNet-50 en apprentissage incrémental (Approche architecturale).

Lorsque nous avons appliqué une approche architecturale à ResNet50 pour l'apprentissage continu, nous avons constaté que cette méthode n'a pas conduit à des performances satisfaisantes sur le long terme. Malgré des résultats initiaux encourageants, la précision du modèle a rapidement diminué au fur et à mesure de l'ajout de nouvelles classes. Notre adaptation architecturale n'a pas réussi à maintenir la précision du modèle de manière durable lors de l'apprentissage de nouvelles tâches.

Les résultats obtenus jusqu'à présent pour les deux approches ne sont pas encore satisfaisants pour notre tâche cible dans le monde réel. En d'autres termes, nous n'avons pas réussi à maintenir une précision stable, et la diminution de la précision après chaque ajout de classes suggère que nos modèles risquent d'obtenir des résultats de plus en plus médiocres à terme. Cependant, il nous reste à explorer la troisième approche, celle de l'apprentissage continu basée sur la mémoire, que nous n'avons pas encore essayée. Cette approche, qui promet des résultats meilleurs que les deux précédentes, implique l'intégration des images des anciennes classes, ce que nous avons réservé pour la phase ultérieure. En conservant des échantillons des anciennes classes, cette approche offre la possibilité au modèle de mieux généraliser et de maintenir des performances plus stables au fur et à mesure de l'ajout de nouvelles classes.

5 CONCLUSION

Dans ce travail, nous nous sommes intéressés à la mise en place d'une solution pour la gestion des consommables médicaux basée sur l'intelligence artificielle et l'internet des objets. Nous nous sommes focalisé sur le défi lié à la robustesse et aux capacités d'adaptation des modèles d'apprentissage lorsqu'ils sont déployés dans le monde réel et confrontés, par exemple, à l'apparition de nouvelles classes de consommables et à l'évolution des caractéristiques des classes déjà apprises. Pour cela, nous avons évalué les capacités des modèles de fondation (notamment ResNet-50 et ViT) pour la détection des consommables de santé dans des cadres (1) fine-tuning, (2) fine-tuning en few-shot* (contraintes sur les quantités de données disponibles) et (3) class-incremental (apparition de nouvelles classes de consommables de manière continue). Les résultats empiriques que nous avons obtenus montrent notamment que le modèle de fondation ViT permet d'atteindre de très bonnes performances dans les cadres (1) et (2) que nous avons exploré. Dans le cadre (3) class-incremental, ce même modèle est capable de reconnaître les nouvelles classes tout en maintenant de bonnes performances sur les anciennes classes. L'utilisation de ce type de modèles en conjonction avec des algorithmes d'apprentissage continu offre des perspectives prometteuses pour le développement de systèmes IA robustes et adaptables aux évolutions des environnements dans lesquels ils seront déployés. Alors que nous avons commencé à gratter la surface de ses possibilités, les prochaines étapes consisteront à étendre son application à de nouveaux domaines et à résoudre des problèmes plus complexes. En adoptant une approche à long terme, nous pourrions envisager d'intégrer des mécanismes d'adaptation encore plus sophistiqués, inspirés par la neuroplasticité, permettant aux modèles d'apprentissage automatique de s'ajuster en permanence à un monde en constante évolution. En poursuivant ces recherches et en collaborant avec diverses disciplines, nous pourrions déverrouiller tout le potentiel de l'apprentissage continu pour créer des systèmes d'IA plus robustes, adaptables et bénéfiques pour l'ensemble de la société.

GLOSSAIRE

Apprentissage en few-shot Une branche de l'apprentissage machine qui vise à entraîner des modèles capables de généraliser à de nouvelles tâches ou de nouvelles classes avec seulement quelques exemples d'entraînement. Cela implique souvent l'utilisation de techniques spéciales pour l'apprentissage à partir de données rares ou peu nombreuses.. 19

apprentissage continue Un paradigme d'apprentissage machine où un modèle est entraîné progressivement sur de nouvelles données au fil du temps, permettant une adaptation continue aux changements et une amélioration des performances au fur et à mesure de l'acquisition de nouvelles connaissances.. 19

Modèles de fondation Des modèles d'apprentissage machine pré-entraînés sur de grandes quantités de données pour apprendre des représentations générales des données. Ces modèles servent souvent de point de départ pour des tâches spécifiques, offrant une initialisation efficace et des performances améliorées.. 19

REFERENCES

- [1] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray O gut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. On the opportunities and risks of foundation models. pages 3–4, 2022.
- [2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021.
- [4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [5] Zhizhong Li and Derek Hoiem. Learning without forgetting, 2017.
- [6] Fan Liu, Tianshu Zhang, Wenwen Dai, Wenwen Cai, Xiaocong Zhou, and Delong Chen. Few-shot adaptation of multi-modal foundation models: A survey. pages 5–6, 2024.
- [7] Mateusz Wojcik. Continual learning: Methods and application. 2024.