



République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique



Université Larbi Tébessi - Tébessa

Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie

Année 2021-2022

Département : Mathématiques et Informatique

THESE

Présenté en vue de l'obtention du diplôme de MASTER

Domaine : Mathématiques et Informatique

Filière : Informatique

***Identification des traits de personnalité en
utilisant le deep learning***

Option

Systemes et multimédias

Par

M^r MOHAMMEDI Khaled

Devant le jury:

Dr. Nait-Hamoud Cherif Mohamed	MCB	Université Larbi Tébessi	Président
Dr. Djeddi Chawki	MCA	Université Larbi Tébessi	Encadrant
Mr. Gahmousse AbdEllatif	MAA	Université Larbi Tébessi	Co- Encadrant
Mr. Abbas Faycel	MCB	Université Larbi Tébessi	Examineur

الملخص

يمكن أن يتمتع البشر بالعديد من الخصائص الفريدة، كما أن الكتابة اليدوية تمتص معلومات قيمة وتقدم بشكل أفضل الحالات النفسية المختلفة للكاتب. يتم استخدام تقنية تحليل الكتابة اليدوية بهدف تحديد سمات شخصية الكاتب من خلال خط يده والذي يقيم الخصائص المختلفة للكتابة اليدوية لتحديد سمات الشخصية يسمى عالم الرسم البياني. هذه العملية تستغرق وقتًا طويلاً والدقة تعتمد على مهارة المحلل. لذلك يقدم هذا العمل نظاماً آلياً لتحديد سمات الشخصية من المستندات المكتوبة بخط اليد. يتضمن النظام المقترح خطوتين رئيسيتين، المعالجة المسبقة والتصنيف (تحديد سمات الشخصية). أولاً، قمنا باستخلاص عينات ذات أحجام مختلفة، ثم قمنا بإزالة العينات غير الضرورية، ثم تم تنظيم الفئات وفق سلم للنتائج وتم تسمية العينات وفقاً للفئات المنظمة، نتيجة البيانات تم زيادتها بتقنية بسيطة لزيادة البيانات، نتيجة البيانات النهائية قسمت إلى حجمين مختلفين أحدهما للتدريب والآخر للاختبار، وتم اختيار الشبكات العصبية التلافيفية كخوارزمية للتعلم العميق.

ركزت التجارب على مجموعة بيانات تتكون من 570 صورة للوثائق المكتوبة يدوياً. تم الحصول على أعلى معدل تعريف عن طريق التوزيع الذي يحتوي على فصول منظمة بترتيب ثنائي ويفصل بينها تدريب بنسبة 80% واختبار بنسبة 20%.

الكلمات المفتاحية: التعرف على ملامح الشخصية، وثنائق مكتوبة بخط اليد، علم الخط، الشبكة العصبية التلافيفية، زيادة البيانات، التعلم العميق.

Abstract

Humans can have many unique characteristics, and handwriting absorbs valuable information and better understands the different states of the writer. A technique called handwriting analysis used to find out more about personality of someone through their handwriting and the one who evaluates various characteristics of an individual's handwriting with the aim of identifying their personality traits is called a graphologist. This process is time consuming and the accuracy depends on the skill of the analyst.

Therefore, this work presents an automatic system for identifying personality traits from handwritten documents. The proposed system includes two main steps, pre-processing and classification (identification of personality traits). At first, we created patches of different sizes with the elimination of some unnecessary, then comes the phase of regularizing the classes and labeling the patches and according to these, data results was augment by a simple data augmentation technique, then the data set result has been split into two different sizes, one for training and the other for testing and algorithm choice for deep learning was the convolutional neural networks.

The experiments focused on a dataset comprising 570 images of handwritten documents. The highest result was obtained by a distribution that has classes regularized by a binary ranking and separated by 80% learning and 20% testing.

Keywords: Personality traits identification, graphology , handwritten documents ,data augmentation, deep learning, convolutional neural networks.

Résumé

Les humains peuvent avoir de nombreuses caractéristiques uniques, et l'écriture manuscrite absorbe des informations précieuses et comprend mieux les différents états de l'écrivain. Une technique appelée l'analyse de l'écriture manuscrite est utilisée pour en savoir plus sur une personne grâce à son écriture manuscrite et celui qui évalue les diverses caractéristiques de l'écriture manuscrite d'un individu dans le but d'identifier ses traits de personnalité s'appelle un graphologue. Ce processus prend du temps et la précision dépend de la compétence de l'analyste.

Par conséquent ce travail présente un système automatique d'identification des traits de personnalité à partir des documents manuscrits. Le système proposé comprend deux étapes principales, le prétraitement et la classification (identification des traits de personnalité). Dans un premier temps, nous avons créé des patchs de différentes tailles, puis certains patchs inutiles ont été éliminé, ensuite vient la phase de régularisation des classes et l'étiquetage des patchs et en fonction de celles-ci, le résultat des données a été augmenter par une simple technique d'augmentation des données, puis le résultat d'ensemble des données a été fractionner en deux tailles différentes, un pour l'apprentissage et l'autre pour le test et le choix d'algorithme pour l'apprentissage approfondie était les réseaux de neurones convolutifs.

Les expériences ont porté sur un ensemble de données comprenant 570 documents manuscrits. Le taux d'identification le plus élevé a été obtenu par une distribution qui a des classes régularisées par un classement binaire et séparées par 80% d'apprentissage et 20% de test.

Mots clés: Identification de traits de personnalité, graphologie, documents manuscrits, réseau de neurones convolutifs, augmentation des données, l'apprentissage en profondeur.

Remerciements

Tout d'abord, je voudrais remercier **Allah** le Tout-Puissant pour la volonté, la santé et la patience qu'il m'a accordées pendant toutes ma vie.

Steve Jobs :

« Ce qu'est un ordinateur pour moi, c'est l'outil le plus remarquable que nous n'ayons jamais conçu. C'est l'équivalent d'un vélo pour notre esprit. »

Je tiens à exprimer ma profonde gratitude et mes remerciements à mes encadrants de mon projet, Mr **GAHMOUSSE AbdEllatif** et Mr **DJEDDI Chawki**, qui m'ont accompagné, encouragé, poussé et dirigé tout au long de ma thèse, et je pense sincèrement qu'ils sont les meilleurs encadrants qu'un étudiant puisse rêver. S'ils me donnaient beaucoup de latitude dans mon travail, ils ont toujours fait preuve d'une grande disponibilité. Je n'oublierai jamais nos conversations, leurs conseils et surtout, leurs gentillesse.

Mes remerciements s'adressent à **tous les membres de jury** pour leur temps, pour le grand honneur qu'ils nous font en acceptant de juger ce travail.

Des milliers de remerciements que je tiens d'exprimer pour mes parents et frères, Je ne serais pas allé aussi loin sans l'équilibre, la chaleur, le soutien et le bonheur dans lesquels j'ai vécu, merci **ma Mère, mon Père, mes grands Frères**.

Mes vifs remerciements à **mes Professeurs** qui ont cru en moi, qui m'ont apporté tout le soutien nécessaire et n'ont pas cessé de saturé l'avidité technique par leurs connaissances en Informatique pendant mon parcours d'études, merci à tout le staff de notre faculté et spécialement à toute l'équipe de département de Mathématiques et Informatiques.

Dédicaces

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Voilà la partie qu'attendent avec impatience toutes personnes qui a eu où aura bientôt cette expérience. J'entame cette tâche non sans émotion, à la fois agréable car elle me rappelle toutes personnes qui m'ont tant donné et étaient à mes côtés, et aussi difficile en raison de leur nombre immense, alors...

Steve Jobs :

« Faire simple est plus difficile que faire compliqué,

Il faut travailler plus dur pour épurer votre pensée et la simplifier. »

Merci à tous pour votre soutien !

Liste des tableaux:

Tableau 1. 1 Indentification de personnalité à partir de la manière d'écriture des lettres.	5
Tableau 1. 2 Indentification de personnalité à partir des caractéristiques d'écriture. [2].....	6
Tableau 1. 3 Résultat des traits par questionnaire FFM.....	7
Tableau 1. 4 Différentes base de données liées au modèle FFM. [15]	10
Tableau 1. 5 L'apprentissage approfondie et ses caractéristiques.	13
Tableau 1. 6 Définitions des rôles des couches de CNN.	15
Tableau 2. 1 Différentes mesures des traits de personnalité.	20
Tableau 2. 2 Différents systèmes basé sur CNN de prédiction des traits de personnalité basés sur l'écriture manuscrite.	27
Tableau 3. 1 Nombre et taille des patches créer	34
Tableau 3. 2 Nombre des patches éliminer et à utiliser.	35
Tableau 3. 3 Classement par rapport au score.	36
Tableau 3. 4 Nombre des classes par classements obtenus pour chaque distribution.....	37
Tableau 3. 5 Nombre et pourcentage de chaque classe pour classement binaire.....	39
Tableau 3. 6 Nombre et pourcentage de chaque classe pour le classement ternaire.....	43
Tableau 3. 7 Nombre et pourcentage de chaque classe pour classement quinaire.....	47
Tableau 3. 8 Exemples des patches avant et après l'augmentation des données.....	49
Tableau 3. 9 Tableau comparatif des résultats obtenus à base des étiquettes binaires 80/20.	53
Tableau 3. 10 Tableau comparatif des résultats obtenus à base des étiquettes binaires 70/30.	53
Tableau 3. 11 Tableau comparatif des résultats obtenus à base des étiquettes binaires 60/40.	53
Tableau 3. 12 Tableau comparatif des résultats obtenus à base des étiquettes ternaires 80/20.	54
Tableau 3. 13 Tableau comparatif des résultats obtenus à base des étiquettes ternaires 70/30.	54
Tableau 3. 14 Tableau comparatif des résultats obtenus à base des étiquettes ternaires 60/40.	54
Tableau 3. 15 Tableau comparatif des résultats obtenus à base des étiquettes quinaires 80/20.	54
Tableau 3. 16 Tableau comparatif des résultats obtenus à base des étiquettes quinaires 70/30.	55
Tableau 3. 17 Tableau comparatif des résultats obtenus à base des étiquettes quinaires 60/40.	55
Tableau 3. 18 Tableau comparatif des résultats obtenus 80/20.	57
Tableau 3. 19 Tableau comparatif des résultats obtenus 70/30.	57
Tableau 3. 20 Tableau comparatif des résultats obtenus 60/40.	57
Tableau 3. 21 Comparaison des systèmes de prédiction des traits de personnalité à partir de l'écriture manuscrite. .59	

Liste des Figures:

Figure 1. 1 Questionnaire du modèle à cinq facteurs.....	7
Figure 1. 2 Processus du modèle à cinq facteurs.....	8
Figure 1. 3 Le processus général de l'apprentissage automatique.....	10
Figure 1. 4 l'intelligence artificielle et ses sous-domaines.	12
Figure 1. 5 Architecture générale d'un réseau de neurones.....	13
Figure 1. 6 L'architecture d'un réseau de neurone convolutif.	14
Figure 2. 1 La relation entre "ligne de base" et les traits de personnalité.[27].....	21
Figure 2. 2 La relation entre "Taille d'écriture" et les traits de personnalité.[27]	22
Figure 2. 3 La relation entre "Vitesse d'écriture " et les traits de personnalité.[27].....	23
Figure 3. 1 Approche proposée.	31
Figure 3. 2 Échantillons manuscrits.	33
Figure 3. 3 Création des patchs.	34
Figure 3. 4 Patchs inutiles.	35
Figure 3. 5 Processus d'élimination des patchs inutiles.	36
Figure 3. 6 Histogramme pour chaque distribution par classement binaire.	37
Figure 3. 7 Histogramme pour chaque distribution par classement de trois classes.	38
Figure 3. 8 Histogramme pour chaque distribution par classement de cinq classes.	38
Figure 3. 9 Pourcentage de chaque classe par classement binaire pour 64 patchs par échantillon.	40
Figure 3. 10 Pourcentage de chaque classe par classement binaire pour 100 patchs par échantillon.	40
Figure 3. 11 Pourcentage de chaque classe par classement binaire pour 256 patchs par échantillon.	40
Figure 3. 12 Pourcentage de chaque classe par classement binaire pour 361 patchs par échantillon.	41
Figure 3. 13 Pourcentage de chaque classe par classement binaire pour 400 patchs par échantillon.	41
Figure 3. 14 Pourcentage de chaque classe par classement ternaire pour 64 patchs par échantillon.	43
Figure 3. 15 Pourcentage de chaque classe par classement ternaire pour 100 patchs par échantillon.	43
Figure 3. 16 Pourcentage de chaque classe par classement ternaire pour 256 patchs par échantillon.	44
Figure 3. 17 Pourcentage de chaque classe par classement ternaire pour 361 patchs par échantillon.	44
Figure 3. 18 Pourcentage de chaque classe par classement ternaire pour 400 patchs par échantillon.	44
Figure 3. 19 Pourcentage de chaque classe par classement de cinq classes pou quinaire 64 patchs par échantillon.....	47
Figure 3. 20 Pourcentage de chaque classe par classement quinaire pour 100 patchs par échantillon.	48
Figure 3. 21 Pourcentage de chaque classe par classement quinaire pour 256 patchs par échantillon.	48

Liste des Figures

Figure 3. 22	Pourcentage de chaque classe par classement quinaire pour 361 patchs par échantillon.	48
Figure 3. 23	Pourcentage de chaque classe par classement quinaire pour 400 patchs par échantillon.	48
Figure 3. 24	architecture du réseau.	51
Figure 3. 25	Un exemple sur les taux d'exactitudes et les taux d'erreurs à base des étiquettes binaires. ...	53
Figure 3. 26	Un exemple sur les taux d'exactitudes et les taux d'erreurs à base des étiquettes ternaires. ...	54
Figure 3. 27	Un exemple sur les taux d'exactitudes et les taux d'erreurs à base des étiquettes quinaires. .	55
Figure 3. 28	Histogramme comparatif des résultats d'exactitude 80/20.	55
Figure 3. 29	Histogramme comparatif des résultats d'exactitude 70/30.	55
Figure 3. 30	Histogramme comparatif des résultats d'exactitude 60/40.	56
Figure 3. 31	Un patch et ses représentations d'étiquettes.	56
Figure 3. 32	Un exemple sur les taux d'exactitudes et les taux d'erreurs 80/20.	57
Figure 3. 33	Un exemple sur les taux d'exactitudes et les taux d'erreurs 70/30.	57
Figure 3. 34	Un exemple sur les taux d'exactitudes et les taux d'erreurs 60/40.	57
Figure 3. 35	Histogramme comparatif des résultats 80/20.	58
Figure 3. 36	Histogramme comparatif des résultats 70/30.	58
Figure 3. 37	Histogramme comparatif des résultats 60/40.	58

Table des matières :

Liste des tableaux.....	VI
Liste des Figures	VII
Table des matières.....	IX
Introduction Générale	1
Chapitre 01	
ANALYSE DE L'ECRITURE MANUSCRITE POUR LA CLASSIFICATION AUTOMATIQUE.	4
1. Introduction.....	5
2. Jeu de données	9
3. L'apprentissage automatique « Machine Learning »	10
3.1. L'apprentissage supervisé.....	10
4. L'apprentissage profondi « Deep Learning ».....	11
4.1. Les réseaux de Neurones (ANN)	12
4.1.1. Les réseaux de neurones convolutifs (CNN)	14
4.1.2. Evaluation du modèle CNN	15
4.2. Deep transfert learning.....	16
5. Conclusion	16
Chapitre 02	
ÉTAT DE L'ART	18
1. Introduction à l'identification des traits de personnalité.....	19
1.1. L'analyse d'écriture manuscrite.....	19
1.2. L'analyse graphologique.....	20
1.3. Modèle à cinq facteurs	23
1.3.1. Comment appliquer le modèle à cinq facteurs.....	24
2. Travaux connexes	25
3. Comparaison entre les travaux ultérieurs	27
4. Conclusion	28
Chapitre 03	
IMPLEMENTATION ET EXPERIMENTATION	29
1. Introduction.....	30
2. Approche proposée	30
3. Outils de développement.....	30
3.1. Environnement de développement et librairies.....	31

Table des matières

3.1.1.	Colab.....	32
3.1.2.	Tensorflow.....	32
3.1.3.	Keras.....	32
3.2.	Jeu de données.....	32
4.	Prétraitement.....	33
4.1.	Création des patches.....	33
4.2.	Elimination des patches inutiles.....	35
4.3.	Régularisation des classes et étiqueter les patches.....	36
4.4.	Analyse des étiquettes des patches.....	38
5.	Apprentissage et Classification (premières expériences).....	49
5.1.	Augmentation des données.....	49
6.	L'architecture CNN.....	51
7.	Expérimentations et résultats.....	52
7.1.	Scénario d'expérimentation 1.....	52
7.2.	Comparaison des résultats de 1 ^{er} Scénario.....	55
7.3.	Scénario d'expérimentation 2.....	56
7.1.	Comparaison des résultats de 2 ^{ème} Scénario.....	58
7.2.	Discussion des résultats.....	58
7.3.	Comparaison avec les travaux ultérieurs.....	59
8.	Conclusion.....	60
FUTURES PERSPECTIVES		
ET CONCLUSION GENERALE		
1.	Perspectives futures.....	62
2.	Conclusion générale.....	63
	Bibliographie.....	64

Introduction Générale

Introduction Générale

L'être humain a toujours développé ses outils de communication depuis son existence, mais l'écriture manuscrite est l'outil le plus ancien de partage de connaissances. Le style d'écriture diffère d'une personne à l'autre et c'est un facteur important dans la détermination de plusieurs caractéristiques de l'écrivain tel que l'âge, le sexe, la caractérisation morphologique et l'état de la santé physique, mentale et émotionnelle et aussi les traits de personnalité d'un écrivain.

Il existe de nombreuses méthodes pour déterminer la personnalité d'une personne, telles que les méthodes utilisées par les psychologues comme les interviews, les expressions faciales et les gestes. Cependant, la personnalité est une combinaison de diverses caractéristiques et qualités d'un individu, et d'un autre côté l'écriture manuscrite est un processus qui ne se produit pas du jour au lendemain, c'est les aspects conscients et inconscients de l'écrivain qui règlent les mouvements des doigts pendant l'écriture. Dans ce cas-là on trouve un lien entre le cerveau et le texte établi, c'est ce qui justifie que l'écriture manuscrite est une méthode puissante et efficace pour déduire la personnalité d'une personne. Et ce n'est pas facile d'identifier la personnalité d'une personne par son style d'écriture, car c'est un domaine qui nécessite un expert d'analyse de l'écriture manuscrite pour analyser plusieurs textes d'un seul écrivain et selon des règles spécifiques il pourra identifier la personnalité de cet écrivain, ce domaine s'appelle la graphologie et son expert on l'appelle un graphologue.

La graphologie à commencer à apparaître dès le début du dix-neuvième « **XIXe** » siècle et elle est considérée comme une nouvelle branche de la psychologie depuis le vingtième « **XXe** » siècle. Parmi les méthodes de la graphologie dans le but d'extraire des différents types de traits de personnalité c'est d'analyser diverses caractéristiques de l'écriture manuscrite comme la vitesse d'écriture, les lettre "A" ou "g", ligne de base, l'inclinaison, et d'une autre part il se trouve plusieurs type d'indicateurs de personnalité comme l'indicateur de types Myers-Briggs (MBTI), aussi le questionnaire à seize facteurs de personnalité (16PF) et le plus utiliser récemment le modèle à cinq facteurs (FFM) connu aussi par le nom (BIG FIVE) c'est un inventaire sous forme d'un ensemble des questions bien choisis et à partir de ce questionnaire un expert observateur fait extraire un ensemble des scores pour chaque trait de personnalité, les détails de cette méthode et ce formulaire et son utilisation sont décrites dans le 2^{ème} chapitre.

Introduction Générale

Mais tout cela a un coût, que ce soit en termes de temps, d'argent ou d'efforts, dès que c'est un travail très compliqué et on ne garantit pas toujours la disponibilité des graphologues dans les divers domaines. Et il faut bien noter qu'il ne s'agit pas de la science exacte et plus précisément c'est un domaine de la science humaine et sociale mais ce problème nécessite un moyen automatisé pour résoudre le problème d'identification des traits de personnalité à partir de l'écriture manuscrite. D'ici vient l'analyse informatisée de l'écriture manuscrite comme une solution pour analyser les différentes caractéristiques et les classifiées dans un temps idéal en évitant que les résultats d'analyse soient relativement subjectifs à cause des facteurs tels que des opinions et expériences différentes. Comme solutions il y en a beaucoup des travaux dans le domaine d'apprentissage supervisé du côté de l'apprentissage automatique « Machine Learning » tels que les classifieurs supervisée, SVM, Adaboost, KNN, les réseaux de neurones artificielles, et tout récemment quelques solutions en apprentissage approfondi.

Pour notre objectif, dès qu'il y a beaucoup des travaux qu'ont été réalisé et proposé comme des solutions sont du côté d'apprentissage automatique mais les solutions les moins affrontées avec peu de résultats se trouvent dans le domaine d'apprentissage approfondi. À cet égard, dans ce travail on va d'offrir une solution d'identification des traits de personnalité en analysant l'écriture manuscrite basée sur le « Deep Learning ». C'est bien nécessaire de marquer qu'il y a un manque terrible des bases de données concernant ce domaine il existe presque deux Jeu de données la 1^{er} est HWxPI, des images des papiers surlignés alors on les considère comme des images bruyantes et difficile à gérer dans un tel travail, ainsi on a décidé de travailler sur l'autre jeu de données qui est un ensemble de 570 image d'écritures manuscrites liées au traits de personnalité extraites par le modèle à cinq facteurs et c'est une base de donnée qui a été créée par Umair & Siddiqi.

Cette thèse est composée de trois chapitres principaux :

- Le premier chapitre, intitulé de « ANALYSE DE L'ECRITURE MANUSCRITE POUR LA CLASSIFICATION AUTOMATIQUE », qui propose une présentation du domaine de L'analyse à partir de l'écriture manuscrite et l'identification des traits de personnalité, avec les jeux de données disponible et utilisées. Et aussi un revu sur l'apprentissage approfondi.

Introduction Générale

- Le 2ème chapitre, intitulé « ETAT DE L'ART » contient une description de la graphologie et les traits extraits depuis l'écritures manuscrites, ainsi une représentation du modèle à cinq facteurs (FFM). Avec une exposition des travaux connexes avec tableau comparatif des travaux les plus connexes dans la littérature.
- Le dernier chapitre, intitulé « IMPLEMENTATION ET EXPERIMENTATION », contient la modélisation de l'approche proposée, une présentation des implémentations réalisées avec les résultats obtenus, toutefois on termine avec quelques perspectives sur les travaux futurs.

Au terme de ce travail, nous concluons par une conclusion générale.

Chapitre 01:
ANALYSE DE L'ECRITURE
MANUSCRITE POUR LA
CLASSIFICATION
AUTOMATIQUE

1. Introduction

En prenant l'histoire des sciences humaines, de la sociologie ou de la psychologie moderne, celle de la graphologie est un concept assez ancien s'il est tenu compte des simples observations faites sur l'écriture et sur son écrivain, il s'avère que son objet et ses buts ont intéressé aussi bien les physiologistes, les philosophes, les artistes et aussi les psychologues et les sociologues, l'analyse de l'écriture manuscrite est une technique utilisée dans des divers domaines par exemple tels que la pédagogie, le recrutement, et le diagnostic de santé mentale et aussi pour identifier la personnalité d'une personne pour mieux comprendre ses émotions et ses éventuels comportements.[1]

L'identification de la personnalité à partir de l'écriture manuscrite peut fournir plusieurs examens de la personnalité, en fonction des caractéristiques utilisées. Si on fait appel à un graphologue pour examiner l'écriture d'un écrivain dans le but d'identifier ses traits de personnalité, il fera extraire les caractéristiques de son écriture manuscrite avec ses techniques spécialisées. Parmi ces techniques, on trouve l'analyse de la ligne de base, l'inclinaison, la pression, la taille, la vitesse d'écriture, et la manière d'écriture des lettres bien déterminées, comme "g", "A", "f", "s", etc. On peut voir quelques exemples dans les tableaux **Tableau 1. 1** et **Tableau 1. 2** et d'après ces dernières caractéristiques on peut déterminer son niveau émotionnel ou ses traits de personnalité. Cela aidera les personnes qui sont émotionnellement perturbées ou déprimées, les personnes qui ont besoin d'une aide psychologique pour surmonter leurs émotions négatives.

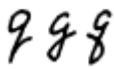
Lettre	Type	Emotion indiquée	Echantillon
Lettre « A »	Lettre spencérienne à l'ancienne	respect de la tradition, protection des autres.	
Lettre « f »	Retourner le trait à gauche	fluidité de la pensée, esprit rapide.	
Lettre « g »	G fait comme Q ou Q fait comme G	Peut signifier un complexe messianique ou de martyr.	
Lettre « s »	S Imprimé en script	lit beaucoup.	

Tableau 1. 3 Identification de personnalité à partir de la manière d'écriture des lettres.[2]

Type		Emotion indiquée	Echantillon
La ligne de base	Zone supérieure	Sphère de l'imagination, l'intellect, la créativité.	
	Zone médiane	Sphère de l'actualité, de la vie sociale.	
	Zone inférieure	Sphère de l'inconscient.	
L'inclinaison	Vertical	Ouvert à l'expérience	
	Incliné	Sensible, émotionnellement sain	
	Très Incliné	aimable, sensible	
	Extrêmement incliné	Extrêmement ardent, jaloux, facilement offensé	
	Incliné (de gauche)	Assez poli	
	Très Incliné (de gauche)	Indépendant, égoïste	
	Extrêmement incliné (de gauche)	Très évasif, la nature émotionnelle est en retrait	
La pression	Pression moyenne	Vitalité, volonté saines.	
	Pression forte	Beaucoup d'énergie, forte volonté, sévères, têtus .	

Tableau 1. 4 Identification de personnalité à partir des caractéristiques d'écriture. [3]

L'écriture manuscrite varie d'une personne à l'autre et aussi les méthodes et les techniques d'analyses. Parmi d'autres méthodes, et pour avoir certains traits de personnalité, on devrait remplir des questionnaires pour avoir un score (quantifié) précis comme résultat. Parmi ces techniques il existe une technique pour mesurer le concept de dogmatisme développé par Rokeach [4], où par la liste de préférences des traits pour évaluer le concept d'intérêt social qui a été proposé par Crandall [5], aussi on trouve l'indicateur de types Myers-

Briggs (MBTI) [6], et on trouve autre technique de la part de Allport et Odbert qui ont identifiés 4000 termes comme référence aux traits de personnalité [7], Et Cattell qui a créé le questionnaire à seize facteurs de personnalité (16PF) [8], et Goldberg qui a créé ses propres ensembles de synonymes et a reproduit, un exemple d'une partie de modèle connu FFM le modèle à cinq facteurs [9], qui est représenté dans **Figure 1. 1** et **Tableau 1. 5** et son processus dans la **Figure 1. 2**

1.	Worry about things.	Very Inaccurate <input type="radio"/>	Moderately Inaccurate <input type="radio"/>	Neither Accurate Nor Inaccurate <input type="radio"/>	Moderately Accurate <input type="radio"/>	Very Accurate <input type="radio"/>
2.	Make friends easily.	Very Inaccurate <input type="radio"/>	Moderately Inaccurate <input type="radio"/>	Neither Accurate Nor Inaccurate <input type="radio"/>	Moderately Accurate <input type="radio"/>	Very Accurate <input type="radio"/>
3.	Have a vivid imagination.	Very Inaccurate <input type="radio"/>	Moderately Inaccurate <input type="radio"/>	Neither Accurate Nor Inaccurate <input type="radio"/>	Moderately Accurate <input type="radio"/>	Very Accurate <input type="radio"/>
4.	Trust others.	Very Inaccurate <input type="radio"/>	Moderately Inaccurate <input type="radio"/>	Neither Accurate Nor Inaccurate <input type="radio"/>	Moderately Accurate <input type="radio"/>	Very Accurate <input type="radio"/>
5.	Complete tasks successfully.	Very Inaccurate <input type="radio"/>	Moderately Inaccurate <input type="radio"/>	Neither Accurate Nor Inaccurate <input type="radio"/>	Moderately Accurate <input type="radio"/>	Very Accurate <input type="radio"/>
6.	Get angry easily.	Very Inaccurate <input type="radio"/>	Moderately Inaccurate <input type="radio"/>	Neither Accurate Nor Inaccurate <input type="radio"/>	Moderately Accurate <input type="radio"/>	Very Accurate <input type="radio"/>
7.	Love large parties.	Very Inaccurate <input type="radio"/>	Moderately Inaccurate <input type="radio"/>	Neither Accurate Nor Inaccurate <input type="radio"/>	Moderately Accurate <input type="radio"/>	Very Accurate <input type="radio"/>
8.	Believe in the importance of art.	Very Inaccurate <input type="radio"/>	Moderately Inaccurate <input type="radio"/>	Neither Accurate Nor Inaccurate <input type="radio"/>	Moderately Accurate <input type="radio"/>	Very Accurate <input type="radio"/>
9.	Would never cheat on my taxes.	Very Inaccurate <input type="radio"/>	Moderately Inaccurate <input type="radio"/>	Neither Accurate Nor Inaccurate <input type="radio"/>	Moderately Accurate <input type="radio"/>	Very Accurate <input type="radio"/>
10.	Like order.	Very Inaccurate <input type="radio"/>	Moderately Inaccurate <input type="radio"/>	Neither Accurate Nor Inaccurate <input type="radio"/>	Moderately Accurate <input type="radio"/>	Very Accurate <input type="radio"/>
11.	Often feel blue.	Very Inaccurate <input type="radio"/>	Moderately Inaccurate <input type="radio"/>	Neither Accurate Nor Inaccurate <input type="radio"/>	Moderately Accurate <input type="radio"/>	Very Accurate <input type="radio"/>
12.	Take charge.	Very Inaccurate <input type="radio"/>	Moderately Inaccurate <input type="radio"/>	Neither Accurate Nor Inaccurate <input type="radio"/>	Moderately Accurate <input type="radio"/>	Very Accurate <input type="radio"/>
13.	Experience my emotions intensely.	Very Inaccurate <input type="radio"/>	Moderately Inaccurate <input type="radio"/>	Neither Accurate Nor Inaccurate <input type="radio"/>	Moderately Accurate <input type="radio"/>	Very Accurate <input type="radio"/>
14.	Make people feel welcome.	Very Inaccurate <input type="radio"/>	Moderately Inaccurate <input type="radio"/>	Neither Accurate Nor Inaccurate <input type="radio"/>	Moderately Accurate <input type="radio"/>	Very Accurate <input type="radio"/>
15.	Try to follow the rules.	Very Inaccurate <input type="radio"/>	Moderately Inaccurate <input type="radio"/>	Neither Accurate Nor Inaccurate <input type="radio"/>	Moderately Accurate <input type="radio"/>	Very Accurate <input type="radio"/>

Figure 1. 3 Questionnaire du modèle à cinq facteurs.

Trait	Extraversion	Névrosisme	Agréabilité	Conscience	Ouverture à l'expérience
Résultat	12	50	36	48	71

Tableau 1. 6 Résultat des traits par questionnaire FFM.

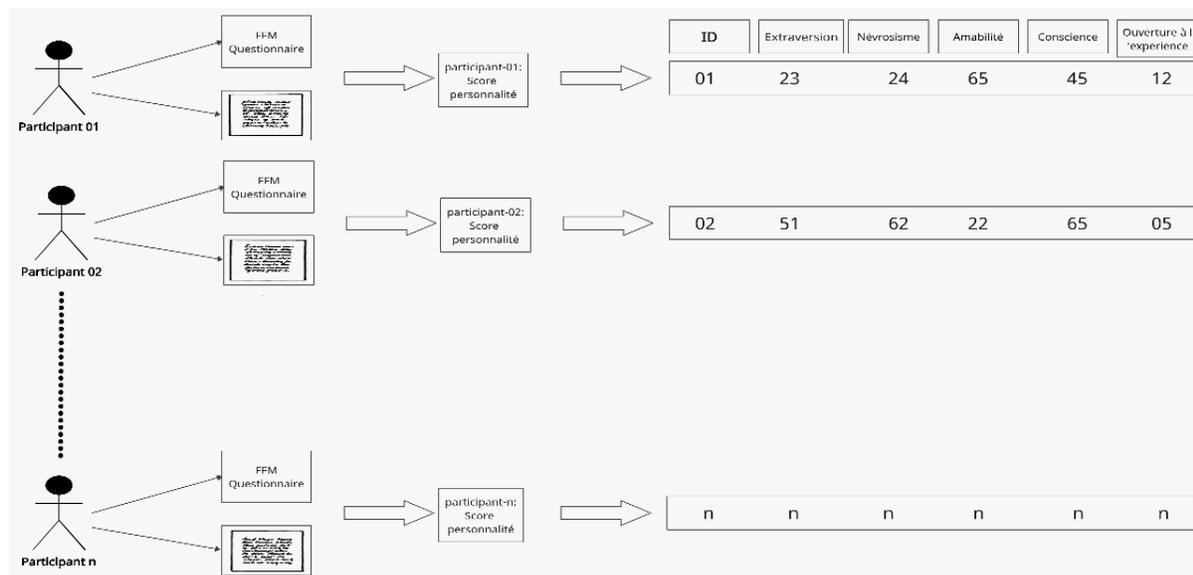


Figure 1. 4 Processus du modèle à cinq facteurs.

Les traits de personnalité peuvent être mesurés à l'aide de l'une des techniques les plus stables et couramment utilisées, c'est la méthode du modèle des cinq facteurs FFM. Son processus commence par remplir un questionnaire de test, et on peut y arriver au même résultat par l'analyse d'écriture d'un document. Comme c'est mentionné dans le **Tableau 1. 7** et **Figure 1. 5** c'est à base de questionnaire on s'attend à une distribution de scores (varie d'une personne à l'autre à base de leurs réponses au questionnaire) comme résultat des traits de personnalité « Extraversion, Névrosisme, Agréabilité, Conscience, Ouverture à l'expérience » voir **Tableau 1. 8**. Ces scores peuvent être obtenus par des évaluations des observateurs, ces derniers résultats de questionnaire peuvent être affectés en tant qu'étiquette du document d'écriture manuscrite de la même personne.

Auparavant, l'analyse manuelle de l'écriture manuscrite nécessitait la présence d'expertise et beaucoup de temps et aussi échantillons de l'écrivain pour prédire la personnalité de personne, la précision de l'analyse dépend de la compétence du spécialiste de l'écriture manuscrite. Même les spécialistes de l'écriture manuscrite se fatiguent lorsqu'ils doivent analyser beaucoup d'échantillons. Le coût d'avoir les services d'un graphologue est élevé, et sa disponibilité n'est pas toujours garantie.

En revanche, Les technologies actuelles ont donné la capacité aux machines pour apprendre et de reconnaître des images, des signaux et des données séquentielles. Dans ce cas-là, les techniques de graphologie sont pertinentes pour être quantifiables automatiquement

grâce à l'apprentissage automatique. Notant que certaines travaux ont utilisées l'écriture manuscrite avec plusieurs méthodes pour identifier les traits de personnalité à partir de l'écriture, on peut citer parmi eux tels que application du traitement d'image pour prédire la personnalité basé sur l'écriture manuscrite et de la signature [10], Identification des traits de personnalité basés sur l'écriture manuscrite à l'aide d'un boosting adaptatif et de caractéristiques texturales [11], Une enquête sur le profilage prédictif à partir de données de signatures manuscrites [12], Identification des traits de personnalité par l'analyse de l'écriture manuscrite [13],

Cependant, l'apprentissage profond nécessite un matériel puissant et des données d'apprentissage plus large. L'une des méthodes d'apprentissage profond les plus populaires pour traiter les images est celle des réseaux neuronaux convolutifs.

2. Jeu de données

Le Jeu de données est parmi les points d'évaluation de l'apprentissage automatique et des systèmes de classification et d'analyse de l'écriture manuscrite. C'est un moyen efficace pour comparer les travaux réalisés, mais dans le domaine d'identification des traits de personnalité par écriture manuscrite à base de modèle des cinq facteurs, on remarque qu'il existe un manque terrible des bases de données, deux Jeux de données ont été mises disponibles jusqu'à maintenant, le premier c'est le jeu de données **HWxPI** [14], qui est un ensemble des images d'écriture manuscrites en langue espagnole écrite sur des papiers surlignés, donc on les considère comme des images bruyantes et difficile à gérer et vraiment à éviter si on veut y arriver à un résultat remarquable, autrement on trouve « **Umair's Dataset** » créée par Umair & Siddiqi [15], cette dernière se compose de 570 images, 285 comme nombre de scripteurs dont chacun a participé avec un texte dépendant et un texte indépendant. Le questionnaire FFM a été utilisé pour évaluer chaque participant. Par conséquent, les résultats obtenus (entre 0 et 100) ont été attribués comme étiquettes aux images sous les classes FFM Extraversion, Névrosisme, Agréabilité, Conscience et Ouverture à l'expérience (E, N, A, C, O).

Base de données	Nombre d'échantillons	Nombre de scripteurs	Base de données Publique	Langue
TxPI-u Dataset	418	Non	Non	Espagnole
Umair's Dataset	570	285	Non	Anglais

Tableau 1. 9 Différentes base de données liées au modèle FFM. [16]

3. L'apprentissage automatique « Machine Learning »

L'apprentissage automatique est le processus de création d'un modèle à partir d'un ensemble de données d'apprentissage. Par exemple, classification ou déterminer de nouvelles données.

Les algorithmes d'apprentissage peuvent être classés en fonction du mode d'apprentissage qu'ils utilisent, la figure suivante présente un exemple du processus d'apprentissage automatique avec des différents modèles

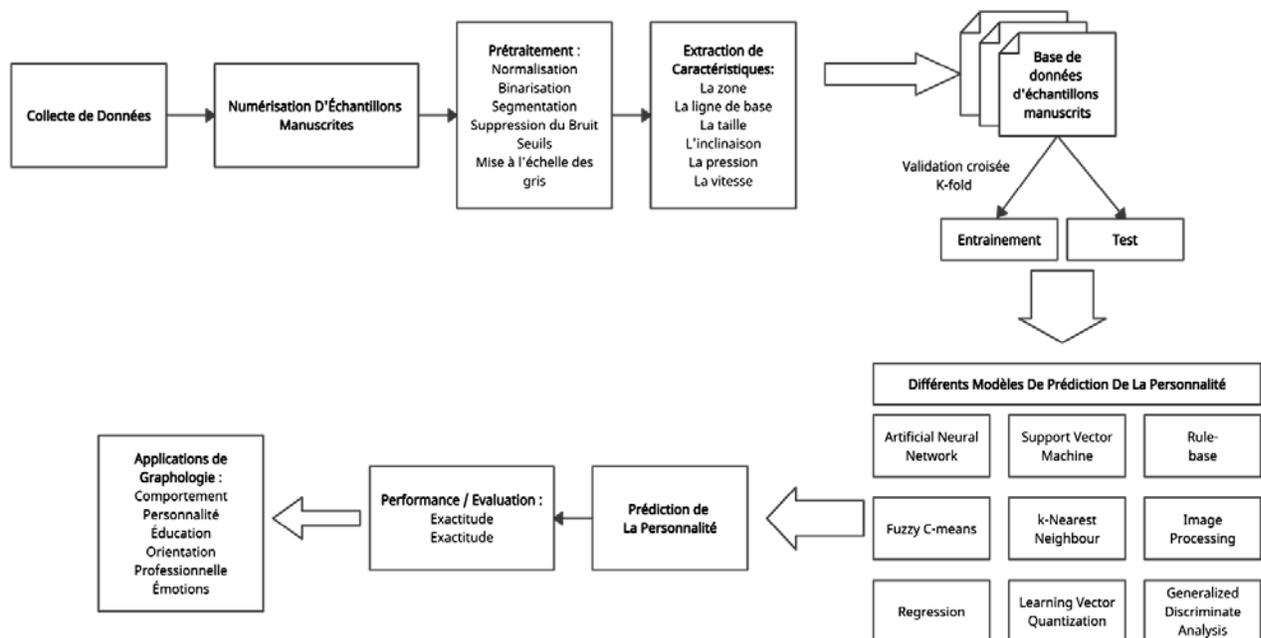


Figure 1. 6 Le processus général de l'apprentissage automatique.

3.1.L'apprentissage supervisé

On ne peut pas parler de l'apprentissage supervisé sans avoir passer par l'apprentissage non supervisé, ce dernier semble plus compliqué car il analyse et regroupe des ensembles de données non étiquetés à l'aide de méthodes d'apprentissage automatique pour former un algorithme, cela aide à détecter un motif lors de la description d'un modèle. Dans cette situation, les données d'entrée aident l'algorithme à extraire les règles et réglementations ainsi qu'à détecter les modèles significatifs.[17]

Les algorithmes d'apprentissage supervisé reposent sur un retour d'information direct pour la prédiction. Cette stratégie se concentre sur la formation d'un algorithme qui agit comme la meilleure fonction pour définir la sélection des données d'entrée en utilisant une collection de données étiquetées. C'est-à-dire qu'il est lié à une description ou à une information. L'algorithme apprend et recherche les relations qui lui permettent de lier les données à l'étiquette à partir de milliers ou de millions d'exemples étiquetés [18], dans le cas d'apprentissage automatique supervisé on trouve plusieurs algorithmes tels que les Réseaux Bayésiens (**BN**), Naïve Bayes (**NB**), Machine à Vecteur de Support (**SVM**), Arbres de décision et Forêts aléatoires (**RF**), K Voisin le plus proche (**KNN**).

4. L'apprentissage approfondi « Deep Learning »

Selon une étude de Deng L et al. [20] l'apprentissage en profondeur est une classe de techniques d'apprentissage automatique qui appartiennent au domaine de l'apprentissage automatique comme c'est démontré dans la **Figure1.7** La famille des techniques d'apprentissage en profondeur s'agrandit pour inclure les réseaux de neurones, les modèles probabilistes hiérarchiques et de nombreux algorithmes d'apprentissage de fonctionnalités supervisés et non supervisés.

Comme tous les réseaux de neurones, l'apprentissage approfondi nécessite des paramètres d'entrée X et Y (tel que X représente les données d'apprentissage et Y représente les labels). L'apprentissage automatique est une méthode d'apprentissage statistique dans laquelle chaque instance de l'ensemble de données est décrite par un ensemble de caractéristiques ou d'attributs. En revanche, l'apprentissage approfondi est une technique d'apprentissage statistique qui extrait des caractéristiques et des attributs à partir de données brutes. Cela se fait à l'aide d'un réseau de neurones avec plusieurs couches cachées, de grandes données et de puissantes ressources informatiques.

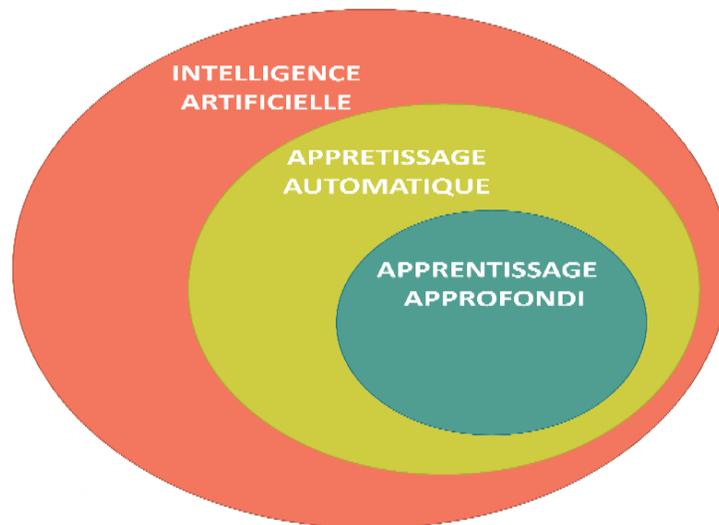


Figure 1. 8 l'intelligence artificielle et ses sous-domaines.

4.1. Les réseaux de Neurones (ANN)

Les réseaux de neurones artificiels tirent leurs concepts structurels et fonctionnels du système nerveux humain, mais à des échelles beaucoup plus petites. Cette technique est applicable aux problèmes où les relations peuvent être non linéaires ou assez dynamiques. Un NN peut capturer de nombreux types de relations, il permet à l'utilisateur de modéliser rapidement et relativement facilement des phénomènes qui autrement auraient pu être très difficiles ou impossibles à gérer. Diverses variantes de ANN existent comme par exemple les Réseaux de Neurones à Propagation Récurrentes (RNN), le NN Probabiliste (PNN), la Fonction de Base Radiale NN (RBNN) et le NN Complémentaire (CMPNN) classés sur la base de différentes méthodes mises en œuvre pour entraîner le réseau. Le premier réseau de neurones artificiels, appelés perceptrons, était basé sur des neurones biologiques avec une couche d'entrée et une couche de sortie entièrement connectées. Il peut être utilisé pour des tâches très simples telles que la classification de modèles séparables linéaires. Pour le cas du perceptron multicouche (MLP) c'est le classificateur ANN le plus utilisé, capable de modéliser des fonctions complexes et robuste aux entrées et bruits non pertinents. Il est difficile de s'entraîner, sauf si l'on choisit des poids plus proches des poids optimaux au départ pour résoudre des problèmes plus complexes.

La **Figure 1. 9** montre l'architecture d'un réseau de neurones. Chaque neurone du réseau reçoit d'abord une entrée, effectue une somme des entrées pondérées, applique une fonction d'activation, calcule une sortie et la transmet à la couche suivante.[21]

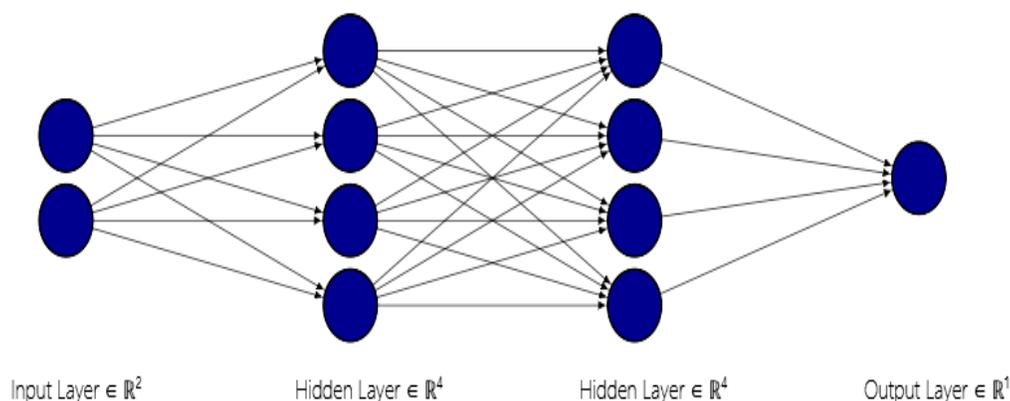


Figure 1. 10 Architecture générale d'un réseau de neurones

Ce tableau montre les caractéristiques de l'apprentissage approfondi :

Exigence de données	Besoin de données volumineuses
Résultat d'exactitude	Grande précision
Temps d'entraînement	plus de temps à s'entraîner.
Matérielle	GPU
Hyper-paramètres	Possibilité de réglage par plusieurs manières.

Tableau 1. 10 L'apprentissage approfondie et ses caractéristiques.

De manière générale, les réseaux de neurones artificiels sont des techniques développées pour simuler l'activité du cerveau humain et de ses neurones biologiques. Particulièrement utilisé pour les tâches de reconnaissance automatique de formes et les systèmes informatiques intelligents, et la classification automatique.

L'apprentissage en profondeur est une approche spécifique utilisée pour créer des topologies de réseaux de neurones qui traitent de grandes quantités de données pour des tâches d'apprentissage automatique. Chaque réseau de neurones donne de bons résultats pour des types de données spécifiques tels que (vidéo, audio, document, etc.). L'avantage de ses techniques d'apprentissage en profondeur est qu'il identifie manuellement et automatiquement les données caractéristiques des couches cachées du processus d'apprentissage. Cela rend la formation des réseaux de neurones plus facile et plus rapide et donne de meilleurs résultats.

Mais, dans le cas des grandes informations et données d'images un simple réseau ANN où ses variantes ne sont pas une solution et n'offre pas des bons résultats, d'ici on a les réseaux de neurones convolutifs qui offre une solution pour ce problème.

4.1.1. Les réseaux de neurones convolutifs (CNN)

Selon l'étude de [22], le réseau de neurones convolutifs "CNN" est un type spécifique de réseau de neurones artificiels qui utilise des perceptrons et des algorithmes d'unité d'apprentissage automatique pour l'apprentissage supervisé afin d'analyser des données. CNN peut être appliqué non seulement au traitement d'images, mais également au traitement du langage naturel "NLP" et à d'autres types de tâches cognitives. Comme d'autres types de réseaux de neurones artificiels, les réseaux de neurones convolutifs ont une couche d'entrée, une couche de sortie et une ou plusieurs couches cachées.

Certaines de ces couches sont des convolutions et utilisent des modèles mathématiques et algébriques pour propager les résultats aux couches suivantes.

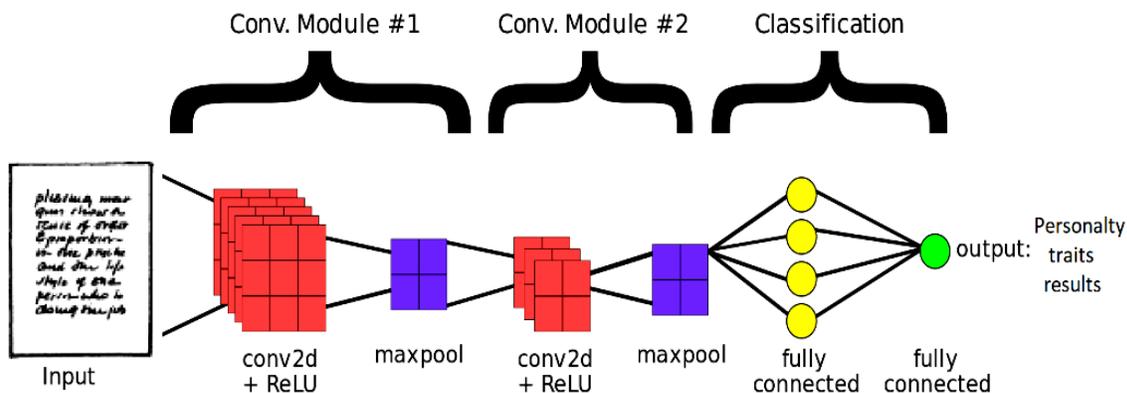


Figure 1. 11 L'architecture d'un réseau de neurone convolutif.

CNN est typique des techniques d'apprentissage en profondeur où des modèles plus sophistiqués, accélèrent le développement de l'intelligence artificielle en fournissant des systèmes qui simulent différents types d'activité biologique dans le cerveau humain.

Généralement, CNN se compose de quatre couches principales:

Couche «Layer »	Rôle
Convolution « Conv »	Traiter les données d'entrée ou le résultat de la couche médiane ou cachée.
Regroupement (Poolig) « Pool »	Compresser les informations en réduisant la taille de l'image intermédiaire.
Activation	Fonction de correction linéaire, est souvent appelée « relu » et met à jour les paramètres de la couche courante et de la couche précédente
Entièrement connectée « Fully connected »	Une couche de type perceptron

Tableau 1. 11 Définitions des rôles des couches de CNN.

4.1.2. Evaluation du modèle CNN

Pour évaluer les performances d'un modèle CNN, il existe un ensemble des métriques qui nous permet de calculer l'exactitude de model, La précision, Le rappel, F1-score, et voilà une brève explication de chacune des métriques [23]

Notez que :

- **TP** (True Positives) le nombre de vrais positifs.
- **FP** (False Positive) le nombre de faux positifs.
- **TN** (True Négatives) le nombre de vrais négatifs.
- **FN** (False Négatives) le nombre de faux négatifs.

L'exactitude de model (Accuracy) c'est la mesure de performance la plus intuitive. Il s'agit simplement du rapport entre le nombre d'observations correctement prédites et le nombre total d'observations.

La précision (Precision) c'est le rapport entre les observations positives correctement prédites et le nombre total d'observations positives prédites.

Le rappel (Recall) c'est le rapport entre les observations positives correctement prédites et toutes les observations de la classe réelle.

F1-score c'est la moyenne pondérée de la précision et du rappel.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$
$$\text{F1-score} = \frac{2 \times (\text{Recall} \times \text{Precision})}{(\text{Recall} + \text{Precision})}$$

Les réseaux de neurones convolutifs (CNN) ont remporté un énorme succès dans des nombreux domaines de la vision par ordinateur et l'analyse d'images. Cependant, il existe encore un grand potentiel pour améliorer les performances des systèmes de prédiction de la personnalité à partir de l'écriture manuscrite en intégrant toutes les informations utilisées par les experts spécialistes en graphologie, ce qui le rend notre choix comme un algorithme du Deep learning.

4.2. Deep transfert learning

D'autre part l'apprentissage en profondeur a récemment reçu de plus en plus d'attention de la part des chercheurs et a été appliqué avec succès dans de nombreux domaines.

Dans certains domaines, comme celui de notre étude c'est difficile de trouver des données volumineuse et le coût est assez élevé pour la collecte de données et les annotations ce qui rendent très difficile la création de grands ensembles de données annotées. Et malgré que l'apprentissage par transfert assouplit l'hypothèse selon laquelle les données d'entraînement doivent être distribuées indépendamment et identiques aux données de test, et motive l'utilisation de l'apprentissage par transfert pour résoudre le problème des données inadéquates [24], pour le cas de domaine de l'identification de personnalité à partir de l'écriture manuscrite, il n'existe pas encore un modèle de classification à la base du modèle à cinq facteurs FFM près à l'utilisation dans l'apprentissage par transfert.

5. Conclusion

Nous vivons à une époque où les possibilités sont infinies et la technologie Deep Learning peuvent nous aider à réaliser de nouvelles avancées technologiques avec ses outils et concepts existantes malgré que le manque des jeux des données qui sont liées au questionnaires comme le FFM l'algorithme de CNN semble une bonne solution pour notre cas, et avant passer à notre propre architecture pour notre étude d'identification de personnalité par écriture manuscrite on va faire une étude de l'existant dans ce domaine dans les sections suivantes.

Le chapitre suivant présente une étude détaillée sur les caractéristiques de l'écriture manuscrite ainsi que les cinq grands traits de personnalité et également un état de l'art dans le domaine de l'identification des traits de personnalité à partir de l'écriture manuscrite.

Chapitre 02:

ÉTAT DE L'ART

1. Introduction à l'identification des traits de personnalité

L'écriture manuscrite est l'un des attributs distinctifs des humains et pour l'écriture le cerveau humain consacre plus d'efforts qu'à la parole et à l'écoute [25], tout le monde a un style d'écriture [26], ce style n'est pas différent pour tous ceux qui partagent les mêmes traits de personnalité, certains appellent l'écriture manuscrite "l'écriture cérébrale" et la justifient parce que nous laissons une empreinte cérébrale de ce que nous pensons, ressentons et agissons sur le papier.[27]

On peut décrire Les traits de personnalité comme suit « des styles émotionnels, interpersonnels, expérientiels, comportementaux et motivationnels », et l'identification de ces traits peut être effectuée sur la base de différents aspects tels que la signature, la position de l'iris, les jeux vidéo en ligne, les hashtags des réseaux sociaux, le texte brut, les documents, l'écriture manuscrite. [28]

1.1. L'analyse d'écriture manuscrite

L'identification des traits de personnalité à partir de l'écriture manuscrite est prête à l'emploi dans divers domaines qu'on peut citer parmi eux la psychologie relationnelle, le recrutement et la gestion des ressources, les enquêtes criminelles, le diagnostic en santé mentale.[14]

L'étude et l'analyse de l'écriture manuscrite à l'aide de l'étude psychologique dans le but de déterminer les traits de personnalité s'appelle la graphologie, est décrite comme une étude scientifique et c'est une façon d'identifier, d'évaluer et d'interpréter le comportement d'un écrivain [26], et qui fait cette étude et analyse on l'appelle un graphologue, quand un graphologue est formé sera capable d'identifier avec précision la personnalité d'un écrivain individu.

Qu'on parmi eux peut citer :

- **Rokeach** a développé un questionnaire pour mesurer le concept de dogmatisme. [4]
- **Crandall** a fait une liste de préférences de traits pour évaluer le concept d'intérêt social.[5]
- **Jung** a proposé la théorie des types psychologique : les attitudes d'introversion et d'extraversion et de névrosisme et de psychoticisme.[29]
- L'indicateur de type **Myers – Briggs (MBTI)** a été développé sur la base de la psychologie jungienne.[6]
- **Allport et Odbert** ont trouvé 18 000 termes descriptifs de traits parmi lesquels 4 000 termes ont été identifiés qui faisaient clairement référence à des traits de personnalité.[7]
- **Cattell** a regroupé les synonymes et considéré 35 échelles de cluster; ils ont identifié 16 dimensions, il est devenu le questionnaire des seize facteurs de personnalité (16PF). [8]
- **Goldberg** a formé ses propres ensembles de synonymes et a reproduit le modèle à cinq facteurs (FFM). [9], [30]

Ce sont différentes mesures de traits de personnalité qui nécessite des techniques de collecte et de traitement des donnée et une évaluation, le trait de personnalité d'un individu peut être mesuré à l'aide de l'une des techniques les plus couramment utilisées présentées dans **Tableau 2. 1**

Mesures des Traits De Personnalité	
Identificateurs	Nombre des traits extraits
MBTI	4
16PF	5
FFM	5

Tableau 2. 2: Différentes mesures des traits de personnalité.

1.2. L'analyse graphologique

Les graphologues recherchent manuellement par des différentes méthodes en les appliquant dans le but d'extraire des résultats analytiques tout en identifiant les traits de personnalité ou de comportement d'un individu à partir de son écriture.

Il y a eu différentes théories pour des caractéristiques particulières de la personnalité. Quelques significations psychologiques des caractéristiques particulières de l'écriture manuscrite et de ses variantes sont mentionnées ci-dessous [2], [27], [28] :

- **Zone d'écriture** C'est la dimension verticale du mouvement d'écriture. Dans la séquence temporelle, les trois zones d'écriture, les zones supérieure, médiane et inférieure représentent respectivement le futur, le présent et le passé. Le développement de l'ego d'un individu remonte à sa sphère intellectuelle et spirituelle, son soi social quotidien et sa pulsion instinctive inconsciente. La proportion des trois zones détermine l'équilibre entre ces principaux domaines d'auto-développement.
- **La ligne de base (Baseline)** Le mouvement des mots écrits sur la ligne s'appelle la ligne de base, et il forme une ligne invisible dans les zones comme Il peut être considéré comme la ligne de modification de l'ego. Ils sont classés en trois types : ascendant se réfère aux personnes optimistes, tandis que les personnes pessimistes moyennes inférieures et la ligne de base se réfèrent aux personnes ayant la maîtrise de soi et [28] a donné les différentes lignes de base d'écriture et ses relations avec les caractéristiques de la personnalité comme est présenter dans la **Figure 2. 1** [2] :

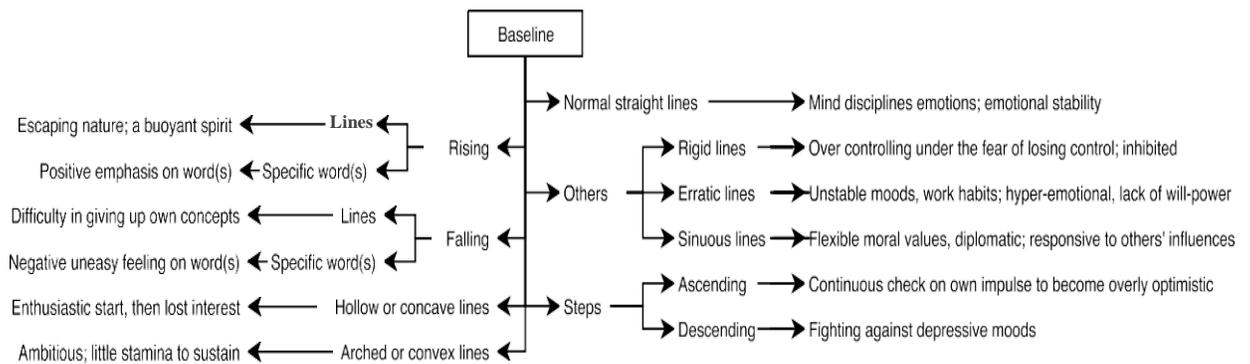


Figure 2. 2 La relation entre "ligne de base" et les traits de personnalité.[28]

- **Inclinaison des mots (Slant)** C'est la dimension horizontale du mouvement d'écriture et comment le mot est écrit dans le contexte de l'inclinaison ce qui permet de mesurer l'alignement social. Les dimensions zonales révèlent la perception, l'analyse, la pensée et le caractère individuel, et les dimensions obliques signifient (émotion, relation, communication et personnalité sociale).

- **Pression (Pressure)** La hauteur, la largeur et la profondeur sont considérées comme les trois dimensions de l'écriture manuscrite. La dimension de profondeur peut être remarquée lorsqu'une pression est appliquée contre la surface d'écriture avec un ombrage. La force lors de l'écriture indique le degré d'énergie disponible. L'écriture contient des coups vers le haut et vers le bas qui peuvent être considérés comme des coups de relâchement et des coups de contraction, respectivement.

D'autres aspects sont pris en compte dans l'analyse de l'écriture manuscrite. Parmi eux : la largeur du trait, fin ou épais, qui affiche l'énergie en action. La Pastiosité (Pastiosity) qui signifie un flux d'encre supplémentaire du stylo sur le papier. La netteté qui indique des traits d'écriture nets. La sensualité d'un individu peut être trouvée en examinant les motifs d'encrage.

- **Taille (Size)** La taille des mots montre l'importance d'un écrivain à ses propres actions, selon [2] La hauteur standard des lettres de la zone médiane est de 3 mm ou 1/8ème de pouce, les caractéristiques de personnalité associées à divers attributs de taille sont présentées par [28] dans la **Figure 2. 3**

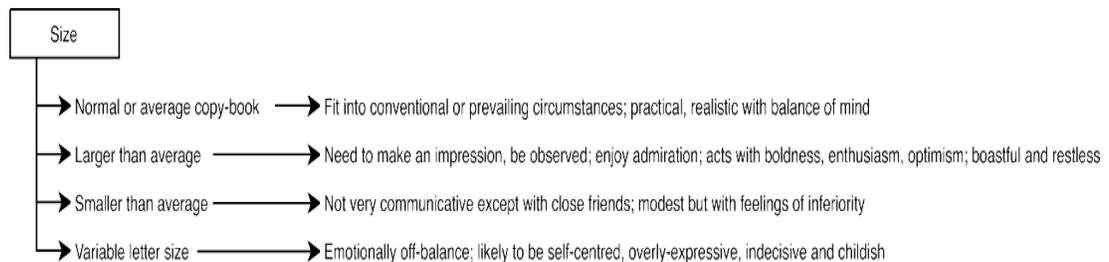


Figure 2. 4 La relation entre "Taille d'écriture" et les traits de personnalité.[28]

- **Espacement (Spacing)** L'espacement entre les lettres indique comment l'écrivain se rapporte à d'autres personnes à un niveau personnel. L'espacement entre les mots représente le besoin de confort émotionnel de l'écrivain et la distance qu'il souhaite maintenir entre lui et la société.

L'espacement entre les lignes fait référence à l'ordre de la philosophie et du raisonnement de l'écrivain. Il fournit des indices sur le degré d'interaction avec sa société.

- **Vitesse (Speed)** La vitesse d'écriture représente le rythme des gestes, des actions et des réactions de l'écrivain, selon [2] il est difficile de dessiner des points à grande vitesse et ils peuvent se transformer en symboles tels que des virgules ou des tirets.

Lorsque le texte est écrit lentement les traits extraits sont : contrôlés, hésitants, instables, voire retouchés. D'autre part, une vitesse d'écriture plus rapide signifie : spontanéité et naturel.

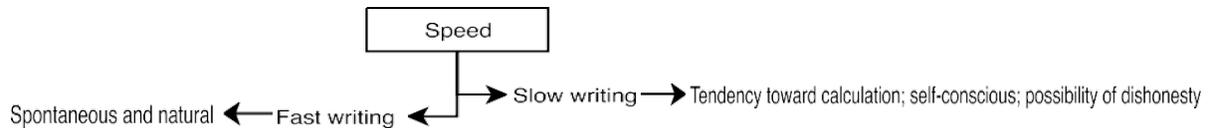


Figure 2. 5 La relation entre "Vitesse d'écriture " et les traits de personnalité.[28]

- **Marge (Margin)** Le script anglais ou français est écrit de manière horizontale de gauche à droite; un écrivain généralement commence son texte du côté gauche du papier qui dépeint le passé, tandis que le côté droit symbolise ses objectifs futurs, la paresse et l'indécision de l'écrivain sont montrées par moins d'espace en bas alors que la marge supérieure montre la convention. La façon dont une écriture manuscrite est disposée sur une page peut représenter une qualité de l'étiquette de l'écrivain, une propension socioculturelle, un penchant pour l'art ou l'absence de ces qualités.
- **Traits (Strokes)** Selon [2], la ligne droite et la forme angulaire symbolisent l'aspiration humaine et l'esprit abstrait ; deux types de traits existent, connectés révèlent les personnes capables de s'adapter rapidement au changement, et non connectés révèlent signifient que ces personnes n'acceptent pas le changement d'une manière simple.

Traditionnellement, extraire les traits de personnalité et l'état mental, le comportement et les sentiments d'un individu à partir de son écriture manuscrite est fait manuellement par des différentes méthodes soit par l'analyse du style d'écriture et d'autres caractéristiques particulières ou bien par des questionnaires tels que le fameux questionnaire du modèle à cinq facteur qu'on va le détailler dans la section suivante.

1.3. Modèle à cinq facteurs

Le modèle à cinq facteurs (FFM) continue de montrer son efficacité en tant qu'outil d'évaluation de la personnalité, la pertinence de ce modèle est considérée comme un effet de sa base en langage naturel, il comprend cinq traits de personnalité quantifiables selon [2] sont présenté comme suit :

- **Extraversion**

Les personnes ayant un score élevé d'extraversion sont toujours entourées d'autres personnes communicatives, faciles à approcher et affirmées, et les personnes ayant de faibles score n'aiment pas les bavardages et être dans le centre d'attention.

- **Névrosisme**

Réfère au réaction de quelqu'un face à une expérience négative, ces personnes sont instable émotionnellement et vulnérables, contrairement aux personnes qui ont un score faible sur ce niveau ce qui indique qu'ils peuvent bien gérer le stress.

- **Agréabilité**

Indique la foi de quelqu'un envers les autres, ce facteur identifie les personnes qui se soucient des autres et qui sont confiantes, honnêtes et plus modestes. Un score faible indique qu'une personne s'intéresse peu aux autres.

- **Conscience**

Se rapporte à la maîtrise de soi contre les actions impulsives. Ce facteur réfère les personnes disciplinées, prudentes et organisées et qui prêtent attention aux détails. Tandis que les personnes ayant de faibles scores sur ce trait n'aiment pas la structure et le calendrier.

- **Ouverture à l'expérience**

Les personnes qui ont ce trait dans leur personnalité s'acceptent les comportements inhabituels ou les expériences uniques. Les personnes ayant un score élevé d'ouverture à l'expérience sont créatives, abstraites et imaginatives, tandis que les personnes avec un score faible sont plus traditionnelles, ils n'apprécient pas les nouveautés.

1.3.1. Comment appliquer le modèle à cinq facteurs

Les traits de personnalité peuvent être mesurés à l'aide de la méthode du modèle des cinq facteurs FFM. C'est simple il suffit de remplir un questionnaire de test du modèle des cinq facteurs, et répondre à 60 questions par cocher la case qui vous convient.

À base de questionnaire on s'attend à une distribution de scores comme résultat des traits de personnalité « Extraversion, Névrosisme, Agréabilité, Conscience, Ouverture à l'expérience » et ces scores se varient d'une personne à l'autre à base de leurs réponses. Ces scores peuvent être obtenus par des évaluations des observateurs, mentionnant que peut-on avoir les mêmes résultats des scores de ce modèle par l'analyse graphologique.

Auparavant, l'analyse manuelle de l'écriture manuscrite nécessite la présence d'expertise et beaucoup de temps et aussi échantillons de l'écrivain pour prédire la personnalité de personne, la précision de l'analyse dépend de la compétence du spécialiste de

l'écriture manuscrite. Le coût d'avoir les services d'un graphologue est élevé. Dans la plupart des cas, les résultats d'analyse des experts sont relativement subjectifs, causés par des facteurs tels que des opinions et expériences différentes, dans ce cas l'analyse informatisée de l'écriture manuscrite offre une meilleure solution pour analyser les différentes caractéristiques et les classifiées en utilisant les méthodes informatisées de classification à l'aide des techniques d'analyse d'images. [14].

2. Travaux connexes

Après une recherche dans le domaine d'identification des traits de personnalité à partir de l'écriture manuscrite on a remarqué que la littérature sur ce domaine comprend plusieurs travaux. Par exemples des modèles de Feed forward neural networks [31], réseaux de neurones [15], Rank-SVM [14] et autre modèles d'apprentissage automatique. Cette section, pour but d'arborer les différents travaux consacrés à l'identification des traits de personnalité à partir de l'écriture manuscrite basé sur l'apprentissage automatique et spécialement l'apprentissage approfondi et en fonction de l'approche utilisée et les résultats obtenus.

Par commencement on va citer les travaux connexes basé sur l'apprentissage automatique

Dans l'an 2018 **Ramirez et al** [32], [33] ont présenté un système de prédiction des traits de personnalité basé sur le modèle des cinq facteurs FFM, le travail était sur la base de données TxPIu créés par eux même dans un cadre d'une compétition. On a proposé plusieurs systèmes, parmi lesquels un système qui s'est basé sur les HoG (Histograms of Oriented Gradient), a utilisé SVM pour la classification. Les résultats était inférieur par rapport à aux études menées par **Gahmousse et al** [11], [14] qui ont présenté deux travaux sur les deux jeux de données pour identifier les traits « FFM » par écritures manuscrite, le 1^{er} travail était sur le jeu de données nommé HWxPI [33] et l'autre sur le même jeu de données de Umair [15]. Les systèmes proposés étaient basés sur des techniques d'extraction des caractéristiques, tel que l'OBIFC « the oriented basic image features » et « Edge Hinge » respectivement. Et pour le 1^{er} travail on a utilisé une classification basée sur le « Rank-SVM », le résultat était d'une exactitude de prédiction globale de 58,6% et pour le 2^{ème} était une classification combinée basé sur AdaBoost, le résultat obtenu était d'une exactitude de prédiction globale de 83,02%.

Et pour les études basées sur l'apprentissage approfondi

En 2018 une recherche de **Gavrilescu et al** [31] qui ont proposé une méthode qui consiste à prédire les traits « FFM » basée sur deux outils psychologiques le modèle des cinq facteurs FFM

et l'analyse graphologique, le travail était sur une base de données collectée auprès de 128 sujets contenant à la fois des textes prédéfinis et aléatoires, ils ont appliqués une nouvel architecture surnommé de Feed-forward neural networks. Et le résultat obtenu était exactitude globale de 84.4 %.

De sa part, **Lemos et al** [26] ont présenté un système de prédiction de différents traits de personnalité comme « Optimiste, Pessimiste, Équilibré, Pragmatique » par écritures manuscrite, le travail était sur une base de données d'un nombre non mentionné, les image d'écriture manuscrite prise de l'utilisateur via un site Web puis converti en noir et blanc ensuite améliorer le contraste ou la netteté de l'image pour rendre les courbes d'écriture nettes. Dans la dernière étape, un réseau de neurones convolutifs (CNN) était introduit pour extraire les fonctionnalités à partir des images d'écriture manuscrite et prédire les différents traits d'un utilisateur.

Il faut aussi noter que dans la même compétition effectué par **Ramirez et al** [33] sur la base de données HWxPI, le système proposé a due divisé les échantillons de le jeu de données en patchs, pour obtenir 216 patchs de chaque images résultant en environ 90 000 images en niveaux de gris, et en appliquant un réseau de neurones convolutifs (CNN) comme classificateur constitué de cinq couches convolutives pour extraire les caractéristiques des patchs et de 3 couches entièrement connectées pour effectuer la classification. Le résultat obtenu c'était une exactitude globale de 50.22 %.

Ainsi de suite, vient **Valdez-Rodríguez et al** [34] qui a proposé une méthode qui consiste à prédire les traits « FFM » par écritures manuscrite avec une technique de recadrage des images d'écriture manuscrite de la base de données nominé HWxPI (Handwritten texts for Personality Identification) [33], pour obtenir un certain nombre des patches par chaque image et puis les convertir en niveaux de gris.

Et en utilisant un réseau de neurones convolutifs (CNN) qui est basée sur l'architecture de AlexNet [35], le réseau se compose de cinq couches convolutives et c'est pour extraire les caractéristiques des patches et aussi trois couches entièrement connectées pour effectuer la classification des traits de personnalité, finalement la méthode proposée a atteint une exactitude de 50.23 %.

Et dernièrement on trouve **Umair.S et al** [15] a fait un travail similaire de ce dernier mais sur une différente base de données qui se compose de 285 échantillons écrits qui a était créer par lui-

même en reliant les échantillons manuscrits par ses traits « FFM » après les avoirs eux par un formulaire d'enquête (questionnaire FFM).

Et en utilisant un réseau de neurones convolutifs (CNN), le réseau se compose de deux couches convolutives et c'est pour extraire les caractéristiques des patches et aussi deux couches entièrement connectées pour effectuer la classification des traits de personnalité et son résultat a obtenu une exactitude d'exactitude globale de 59,73 %.

3. Comparaison entre les travaux ultérieurs

Nous citons les différentes techniques qui ont été utilisées dernièrement pour prédire les traits de personnalité, l'année et l'ensemble de données et technique utilisé et le nombre d'images et le résultat obtenus sont présenté dans le **Tableau 2. 3**

Etude par :	Année :	Jeu de données :	Classifieur utiliser :	Nombre d'images :	Résultat :
Lemos et al. [26]	2018	Non mentionné	CNN	Non mentionné	Non mentionné
Ramirez et al [33]	2019	HWxPI	CNN	418 images	50.22 %
Valdez-Rodríguez et al. [34]	2019	HWxPI	CNN	418 images	50.23 %.
UMAIR.S et al. [15]	2020	Propre jeu de données	CNN	285 images	59.73%

Tableau 2. 4: Différents systèmes basé sur CNN de prédiction des traits de personnalité basés sur l'écriture manuscrite.

4. Conclusion

Après avoir présenté des notions de généralité en rapport avec le lien entre la personnalité et l'écriture manuscrites d'une façon concise l'analyse d'écritures manuscrites, l'identifications des traits de personnalité, en passant par l'analyse graphologique, ainsi que quelques bases d'identificateurs de personnalité et on mettez en face le modèle à cinq facteurs qui est le modèle lié avec la base de données qu'on va travailler sur elle, et aussi on a mentionné des différents travaux connexes dans le domaine de l'identification de la personnalité à partir de l'écriture manuscrites on remarquant que dans la littérature ils ont mets d'accord que les résultats extraites par des questionnaires peuvent être extraites en appliquant des méthodes de graphologie aussi et l'objectif de la présentation de ces éléments, c'est de mettre en face les difficultés rencontrées par les graphologues dans l'analyse des manuscrites et la nécessité d'une solution informatisée, et pour cela notre travail va s'orienter vers nos expérimentations concernant l'identification des traits de personnalité à partir de l'écriture manuscrite en utilisant le Deep Learning dans le chapitre suivant.

Chapitre 03:
IMPLEMENTATION ET
EXPERIMENTATION

1. Introduction

Dans ce chapitre, nous présenterons la base de données utilisées, l'approche proposée, la préparation de données et l'étiquetage, puis nous discuterons l'architecture CNN proposée pour identifier les traits de personnalité à partir de l'écriture manuscrite que nous avons réalisée afin d'évaluer l'efficacité de l'architecture.

2. Approche proposée

Notre approche et ses étapes sont décrites de début par obtenir un jeu de données des images d'écritures manuscrites liées directement au résultats du modèle FFM et ce jeu de données c'est Umair's Dataset [15], ensuite on a effectué un prétraitement constituer d'une étape de création des patches de la même taille (fragment d'un échantillon) dans le but d'avoir des nouvelles variantes de ce jeu de données ensuite une purification des données a été appliquée à base de l'étape de conversion en noir et blanc du jeu de données qui a été déjà patchés et aussi la détermination d'un seuil (cette conversion aide seulement dans l'étape de purification des données), en résultat tous les patches inutiles ont été éliminés, subséquemment vient la partie de régularisation des classes sous différents classements en basant sur le score FFM, à partir de cette régularisation le jeu de données patchés purifié a été étiqueté et en résultat on a plusieurs variantes d'un jeu de données patchés purifié et étiqueté.

Après la préparation des données, on a réalisé un nombre des expérimentations avec des différents paramètres et aussi différents tailles des patches du jeu de données.

Malheureusement, on s'est heurté à une stagnation de résultats et ceci nous a poussés à appliquer une simple technique d'augmentation des données sur le jeu de données pour avoir un jeu de données patchés purifié étiqueté et augmenté, sur ce dernier on a appliqué un certain nombre des expérimentations sous différents scénarios, et on a sorti avec notre architecture de CNN proposé afin de l'évaluer. La figure **Figure 3. 1** représente l'approche proposée.

3. Outils de développement

Le matériel utilisé est un ordinateur personnel avec une configuration de :

- Processeur : Intel® Core™ i5-5200U CPU @ 2.20GHz 2.20 GHz
- Mémoire installée (RAM) : 8.00 Go
- Type de système : Système d'exploitation Windows 10 professionnel, processeur 64 bits x64.

3.1. Environnement de développement et librairies

Notre étude a été appliquée sur un environnement de développement afin d'avoir des résultats avec des librairies qui sont mentionnées dans la section suivante.

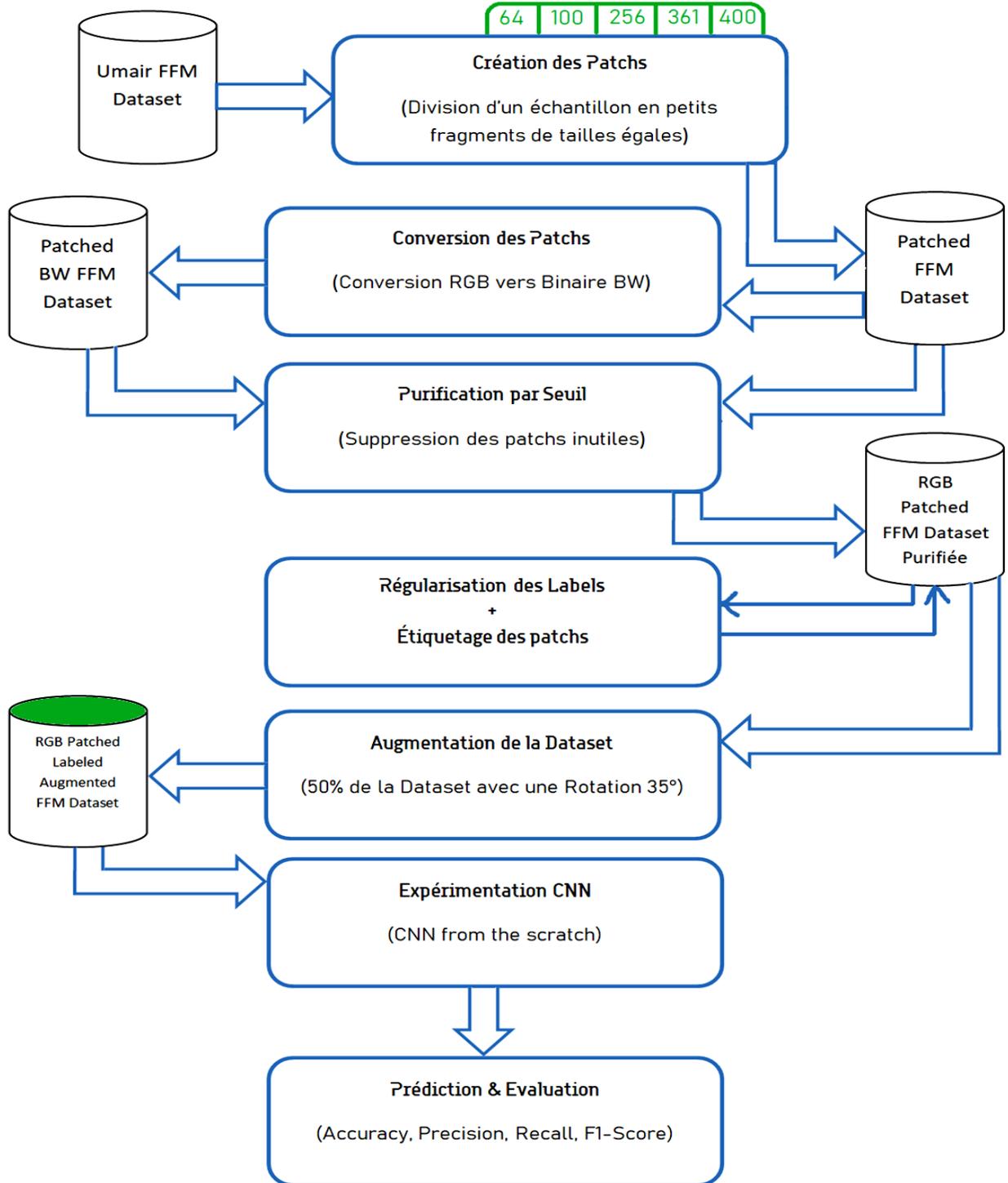


Figure 3. 2 Approche proposée.

3.1.1. Colab

Colab (ou "Colaboratory") C'est un environnement interactif appelé **notebook Colab basé sur Python**, qui vous permet d'écrire et d'exécuter du code Python dans votre navigateur. Les notebooks Colab sont enregistrés dans votre compte Google Drive. Vous pouvez facilement les partager avec vos collaborateurs.

Colab vous permet d'importer un ensemble de données d'images, d'entraîner un classificateur d'images sur cet ensemble et d'évaluer le modèle. Les notebooks Colab exécutent ce code sur les serveurs cloud de Google. Vous avez donc à votre disposition toute la puissance du matériel Google, y compris les GPU et TPU, quelle que soit la puissance de votre ordinateur. Vous n'avez besoin que d'un navigateur.[36]

3.1.2. Tensorflow

TensorFlow est une bibliothèque logicielle open source pour le calcul numérique de haute performance, dédiée au machine learning. Son architecture flexible permet un déploiement facile du calcul sur diverses plates-formes (CPUs, GPUs, TPUs), Elle propose un écosystème complet et flexible d'outils, de bibliothèques et de ressources communautaires permettant aux chercheurs d'avancer dans le domaine du machine learning.

TensorFlow offre davantage de flexibilité et de contrôle grâce à des fonctionnalités telles que l'API fonctionnelle de Keras et l'API de sous-classification de modèle pour la création de topologies complexes. Initialement développé par des chercheurs et des ingénieurs de l'équipe de Google Brain au sein de l'organisation de l'IA de Google, il s'appuie sur l'apprentissage automatique et l'apprentissage en profondeur. [37].

3.1.3. Keras

tf.keras est l'API de haut niveau de TensorFlow permettant de créer et d'entraîner des modèles d'apprentissage en profondeur. Elle est utilisée dans le cadre du prototypage rapide, de la recherche de pointe et du passage en production.[38]

3.2. Jeu de données

Le jeu de données utiliser dans note travail Umair's Dataset a été créé par [5], afin de pouvoir répondre aux problèmes d'identification des traits de personnalité à partir d'écriture manuscrite. Ce jeu de données se compose de 570 échantillons d'écriture manuscrite sont écrites par 285 individus (hommes et femmes), âgés de 20 à 40 ans. Chaque personne s'est

portée volontaire pour écrire deux échantillons et à répondre aux questionnaires FMM par lequel il sera évalué. Le texte du premier paragraphe était le même pour tous les participants (Dépendant) tandis que le second était différent (Indépendant).

Par conséquent, les résultats obtenus (valeurs entières entre 0 et 100) ont été assignés comme étiquettes aux images de texte sous les classes FFM « Extraversion, Névrosisme, Agréabilité, Conscience, et Ouverture à l'expérience » (E,N,A,C,O).[11]

La figure 3.2 représente quelques échantillons manuscrits du jeu de données

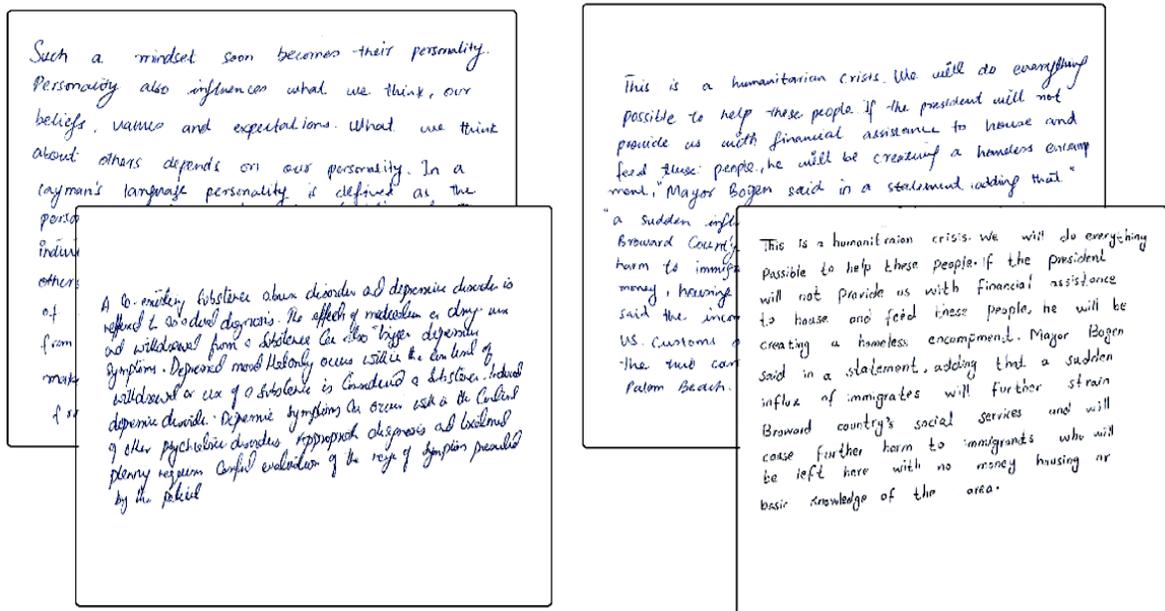


Figure 3.3 Échantillons manuscrits.

4. Prétraitement

L'étape de prétraitement est subdivisée en plusieurs étapes avant d'effectuer des expérimentations, comme un début c'est la partie de création des patches suivi par une élimination des patches inutiles ensuite la régularisation des classes et l'étiquetages et après une analyse des résultats des étiquettes des patches, on va expliquer et détailler tous ces étapes dans ce suit :

4.1. Création des patches

À partir des échantillons d'écriture manuscrits de la taille k hauteur par k largeur où $k= 1900$ nous avons fixé les tailles des patches (comme c'est indiqué avec la couleur de surbrillance du texte dans le **Tableau 3. 1**) qu'on doit créer pour nos tests, ces patches sont créé en fonction d'un nombre positive z , et ils sont de la taille $(N \times N)$, et $N = (k \div \sqrt{z})$, où z

est le nombre des patches obtenus par chaque images, et il est défini par un nombre > 1 et multiplier par lui-même, par ici nous avons créé des petits patches carrés de la mêmes taille en couleurs et en noir et blanc, qui veut dire une division entière de l'image de l'écriture manuscrite et cette technique de division assure que les patches soient égaux en termes de taille sans avoir des pertes de données, les patches en résultats auront automatiquement le même indice d'échantillons suivi par numéro de la ligne et de la colonne de l'emplacement du patch dans l'image originale par exemples pour extraire 256 patches par image de taille 118 par 118 les deux premiers patches extrait prennent un nom comme suit: 001-01_01_01, 001-01_01_02.

Le **Tableau 3. 2** présente les 5 variantes du jeu de données chacune d'eux et en version RGB (originale) et en version binaire (noir et blanc) et la **Figure 3.3** représente une démonstration de processus de création des patches.

Patches par image	Nombre des patches	Taille du patch
64	36480	237 × 237
100	57000	190 × 190
256	145920	118 × 118
361	205770	100 × 100
400	228000	95 × 95

Tableau 3. 3: Nombre et taille des patches créer

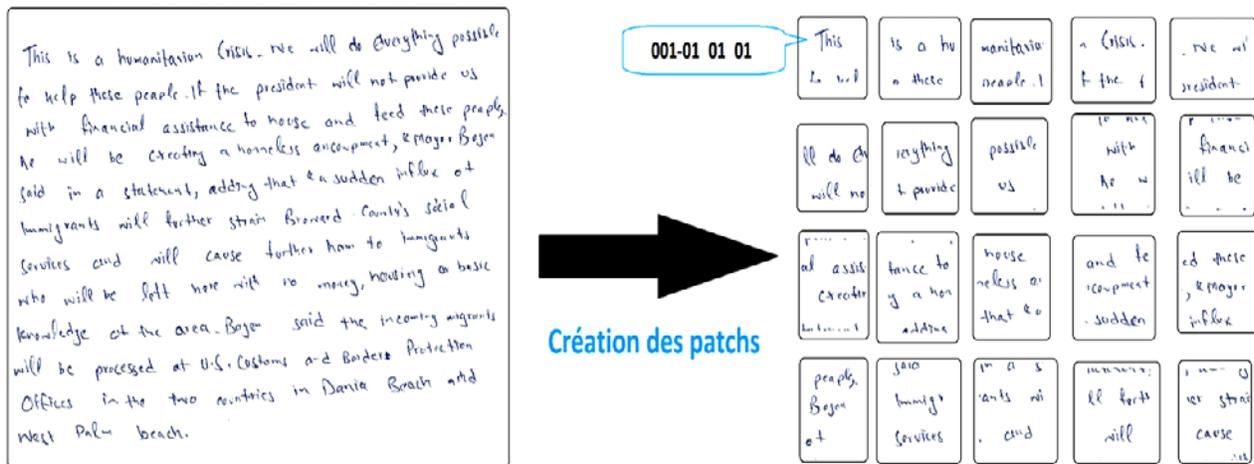


Figure 3. 4 Création des patches.

Concernant cette étape de création des patches il y a plusieurs techniques pour créer plusieurs patches à partir d'une seule image d'écriture manuscrite, telles que le patching orienté zone d'intérêt, et aussi la technique du patching aléatoire et c'est une technique avec perte de données qui a été utilisée par [15]...etc.

4.2. Elimination des patches inutiles

En résultat des patches on retrouve que certains des patches créés ne contenaient qu'un fond totalement blanc ou il occupe la majorité du patch avec un minimum de données par exemple (un point, une tache, un trait, ...).

Pour avoir seulement les patches convenables, on a mené un nombre des tests pour déterminer un seuil basé sur le rapport minimum entre le pourcentage des pixels noirs par rapport au pourcentage des pixels blancs dans un patch le seuil choisi été un pourcentage de **0.85%** du pourcentage des pixels noirs.

Ce seuil a été utilisé pour éliminer les patches dispensables qui contenaient une quantité insuffisante de données, et la **Figure 3. 5** est un exemple des patches inutiles.

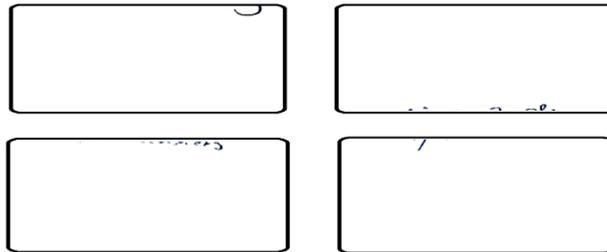


Figure 3. 6 Patches inutiles.

Et le résultat est dans le **Tableau 3. 4** qui représente le nombre des patches qui nous reste après le processus d'élimination des patches inutiles.

Patches par image	Nombre des patches inutiles	Nombre des patches à utiliser	Taille du patch
64	18 179	18 301	237 × 237
100	30 299	26 701	190 × 190
256	82 849	63 071	118 × 118
361	121 951	83 819	100 × 100
400	136 500	91 500	95 × 95

Tableau 3. 5: Nombre des patches éliminer et à utiliser.

Donc cette technique de patching choisi nous assure qu'il n'aura pas de perte de données et assure aussi l'extraction de la zone d'intérêt d'une manière indirecte et voilà une illustration du processus de la purification des patches dans la figure suivante

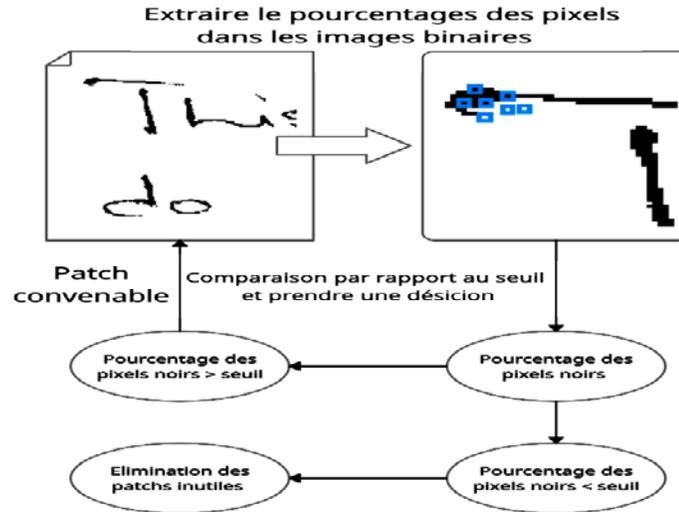


Figure 3. 7 Processus d'élimination des patchs inutiles.

4.3.Régularisation des classes et étiqueter les patches

Le jeu de données utiliser contient des échantillons d'écriture manuscrits liées aux résultats FFM et chaque trait est dans une plage d'échelle de 0 à 100.

On va classés les résultats en :

- Classement basée sue 2 classes (**binaire**) : Un score bas = 0, Un score haut = 1.
- Classement basée sur 3 classes (**ternaire**) : Un score bas=0, Un score moyen=1, Un score haut=2.
- Classement basée sur 5 classes (**quinaire**) : Un score bas=0, Un score moyen bas=1, Un score moyen=2, Un score moyen haut=3, Un score haut=4.

Par exemple pour le classement binaire, si la valeur du trait est totalement inférieure à 51, on lui assigne une étiquette (0) et si le trait a une valeur supérieure ou égale à 51 donc on lui assigne l'étiquette (1).

Le **Tableau 3. 6** présente une illustration de classement par rapport aux scores

Nombre de Classe	Plage de score et sa représentation de 0 → 100				
Classement binaire	0 → 50 : en 0		51 → 100 : 1		
Classement de 3 classes	0 → 33 : 0	31 → 67 : 1		68 → 100 : 2	
Classement de 5 classes	0 → 20 : 0	21 → 40 : 1	41 → 60 : 2	61 → 80 : 3	81 → 100 : 4

Tableau 3. 7: Classement par rapport au score.

Après avoir bien préparé les patches et classer les résultats des traits, tous les patches extraits de même échantillon on va les attribuer les mêmes classements d'échantillons originaux par rapport au classes (Extraversion, Névrosisme, Agréabilité, Conscience, Ouverture à l'expérience), on peut remarquer les classes dominantes dans tous les classes dans le tableau et les figures suivantes

Nombre patches	2 classements		3 classements			5 classements				
	0	1	0	1	2	0	1	2	3	4
64	61327	30178	37346	43839	10320	18669	30303	23764	14948	3821
100	89426	44079	54505	63899	15101	27242	44173	34648	21855	5587
256	211305	104050	128655	151010	35690	64237	104567	81821	51541	13189
361	281057	138038	170909	200949	47237	85462	138836	109115	68240	17442
400	307052	150448	186981	219183	51336	93543	151507	119008	74525	18917

Tableau 3. 8: Nombre des classes par classements obtenus pour chaque distribution.

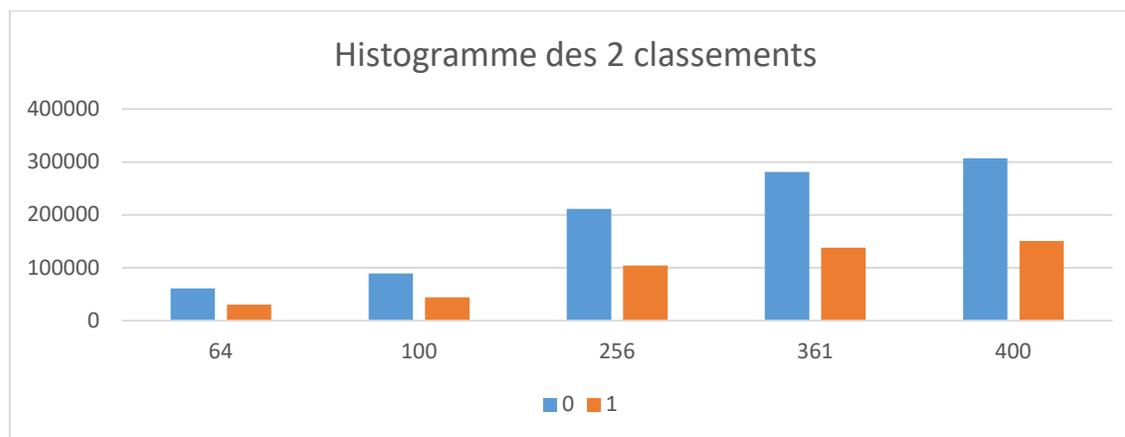


Figure 3. 8 Histogramme pour chaque distribution par classement binaire.

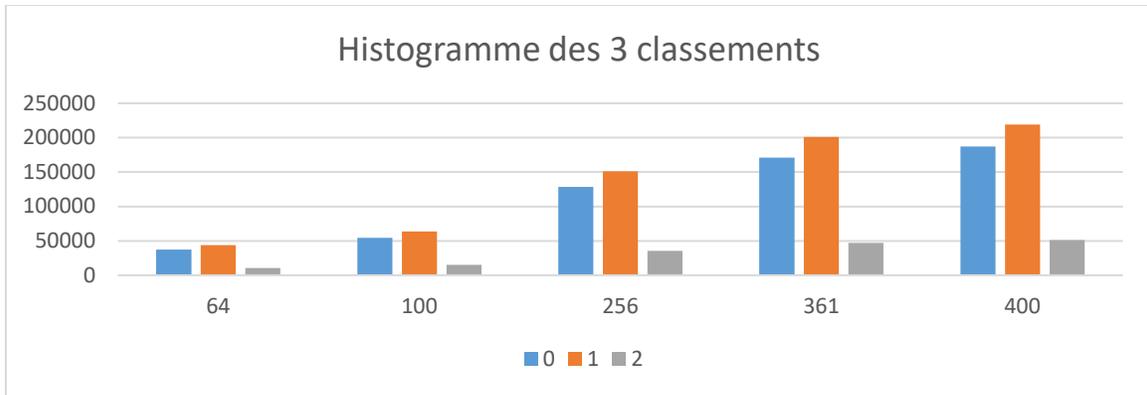


Figure 3. 9 Histogramme pour chaque distribution par classement de trois classes.

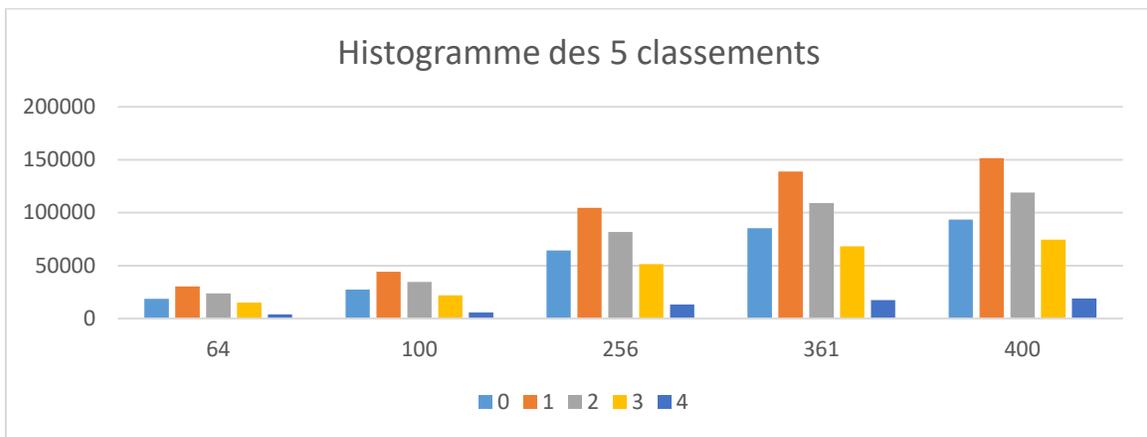


Figure 3. 10 Histogramme pour chaque distribution par classement de cinq classes.

4.4. Analyse des étiquettes des patches

Après avoir préparé les patches et leurs étiquettes des traits (Extroversion, Neruoticism, Agreeableness, Conscientiousness, Openness to Experience), on les analysé par chaque classe et par chaque distribution pour voir les classements dominants dans chaque expérience:

- Classement binaire

Patches par image	Classe	Classement	Nombre	Pourcentage
64	Extroversion	0	10432	57 %
		1	7869	43 %
	Neruoticism	0	11450	62.5 %
		1	6851	37.5 %
	Agreeableness	0	12206	66.7 %
		1	6095	33.4 %

	Conscientiousness	0	11870	64.8 %
		1	6431	35.2 %
	Openness to Experience	0	15369	84 %
		1	2932	16 %
100	Extroversion	0	15157	57 %
		1	11544	43 %
	Neruoticism	0	16742	63 %
		1	9959	37 %
	Agreeableness	0	17830	67 %
		1	8871	33 %
	Conscientiousness	0	17247	64.5 %
		1	9454	35.5 %
Openness to Experience	0	22450	84 %	
	1	4251	16 %	
256	Extroversion	0	35871	57 %
		1	27200	43 %
	Neruoticism	0	39528	63 %
		1	23543	37 %
	Agreeableness	0	42119	67 %
		1	20952	33 %
	Conscientiousness	0	40859	65 %
		1	22212	35 %
Openness to Experience	0	52928	84 %	
	1	10143	16 %	
361	Extroversion	0	47759	57 %
		1	36060	43 %
	Neruoticism	0	52813	63 %
		1	31006	37 %
	Agreeableness	0	56026	66.84 %
		1	27793	33.15 %
	Conscientiousness	0	54182	64.64 %
		1	29637	35.35 %
Openness to Experience	0	70277	83.84 %	
	1	13542	16.15 %	
400	Extroversion	0	52210	57 %
		1	39290	43 %
	Neruoticism	0	57595	63 %
		1	33905	37 %
	Agreeableness	0	61198	67 %
		1	30302	33 %
	Conscientiousness	0	59233	65 %
		1	32267	35 %
Openness to Experience	0	76816	84 %	
	1	14684	16 %	

Tableau 3. 9: Nombre et pourcentage de chaque classe pour classement binaire.

D'après **Tableau 3. 10** on remarque que les pourcentages des classements par rapport au nombre des patchs par échantillon sont presque les même pour toutes les distributions des patchs par images

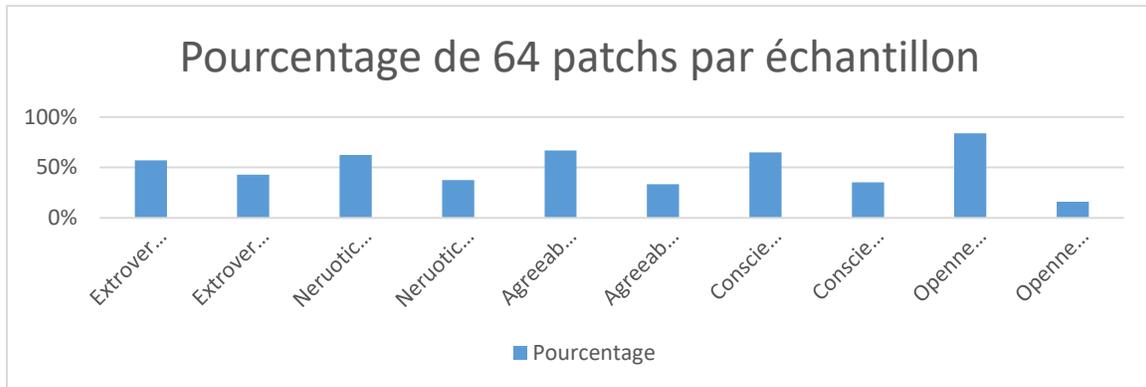


Figure 3. 11 Pourcentage de chaque classe par classement binaire pour 64 patchs par échantillon.

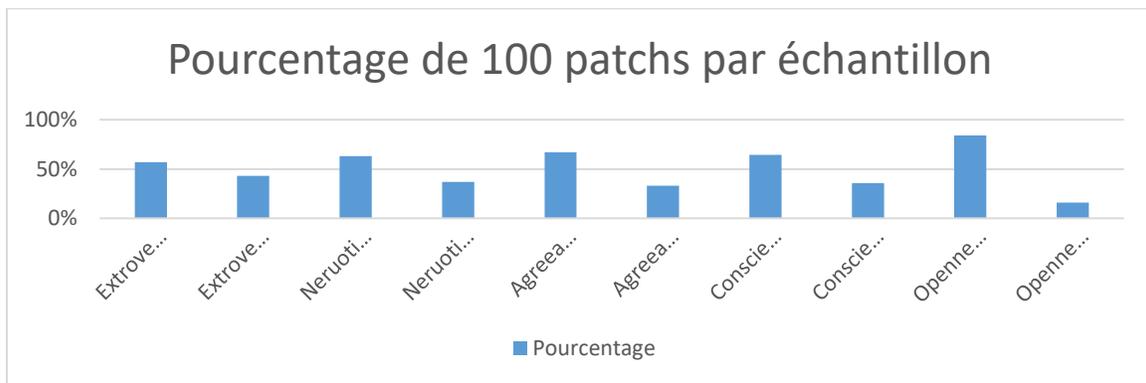


Figure 3. 12 Pourcentage de chaque classe par classement binaire pour 100 patchs par échantillon.

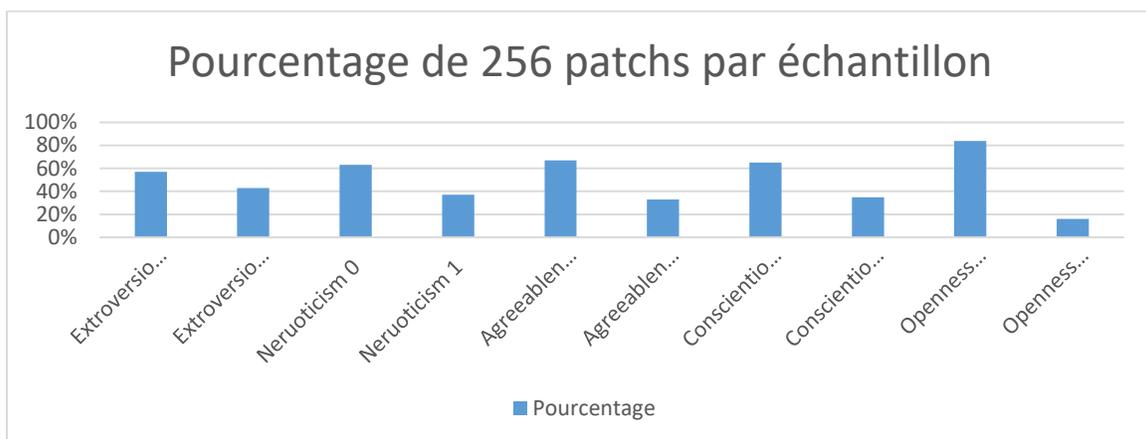


Figure 3. 13 Pourcentage de chaque classe par classement binaire pour 256 patchs par échantillon.

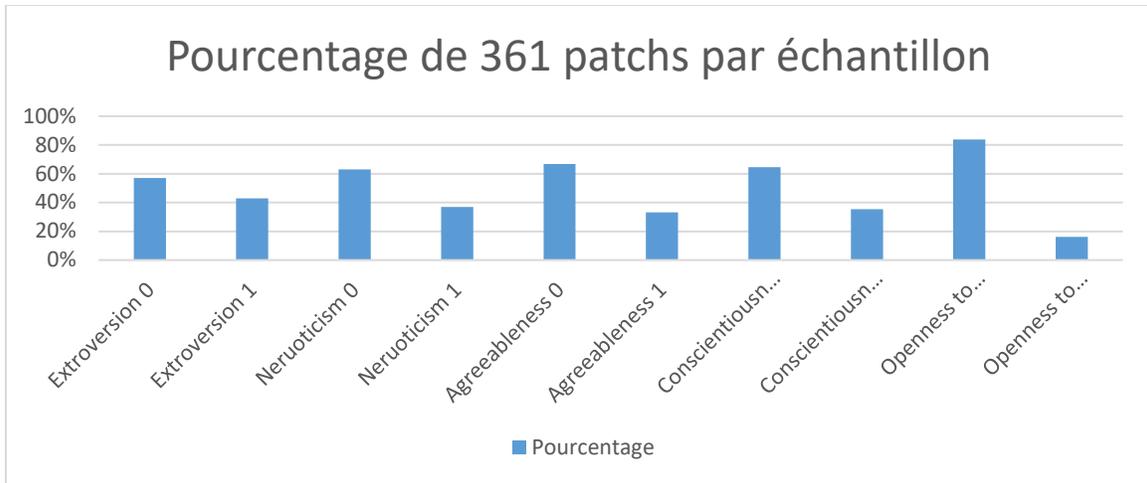


Figure 3. 14 Pourcentage de chaque classe par classement binaire pour 361 patchs par échantillon.

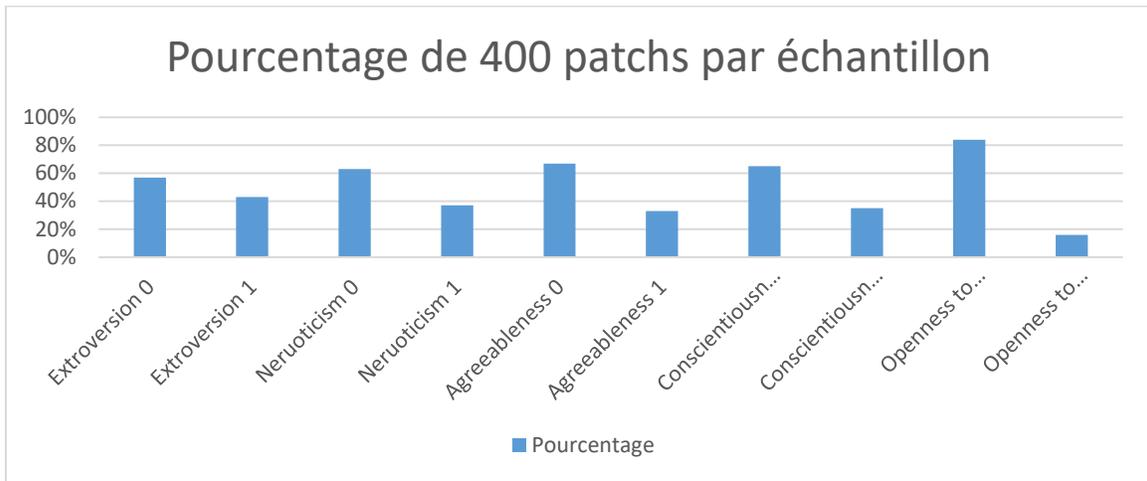


Figure 3. 15 Pourcentage de chaque classe par classement binaire pour 400 patchs par échantillon.

D’après les figures précédentes on remarque que les pourcentages pour le niveau bas (classement : 0) sont dominant pour le classement binaire par score.

- Classement ternaire

Patchs par image	Classe	Classement	Nombre	Pourcentage
64	Extroversion	0	4905	26.80 %
		1	10756	14.42 %
		2	2640	58.77 %
	Neruoticism	0	5495	30 %
		1	9724	53.13 %
		2	3082	16.84 %
	Agreeableness	0	6898	37.69 %
		1	10182	55.63 %
		2	1221	6.67 %

	Conscientiousness	0	7169	39.17 %	
		1	8988	49.11 %	
		2	2144	11.71 %	
	Openness to Experience	0	12879	70.37 %	
		1	4189	22.88 %	
		2	1233	6.37 %	
100	Extroversion	0	7129	26.69 %	
		1	15650	58.61 %	
		2	3922	14.68 %	
	Neruoticism	0	8061	30.18 %	
		1	14163	53.04 %	
		2	4477	16.67 %	
	Agreeableness	0	10116	37.88 %	
		1	14825	55.52 %	
		2	1760	6.59 %	
	Conscientiousness	0	10369	38.83 %	
		1	13191	49.40 %	
		2	3141	11.67 %	
	Openness to Experience	0	18830	70.52 %	
		1	6070	22.73 %	
		2	1801	6.74 %	
	256	Extroversion	0	16878	26.76 %
			1	36999	58.66 %
			2	9194	14.57 %
Neruoticism		0	19026	30.16 %	
		1	33415	52.97 %	
		2	10630	16.85 %	
Agreeableness		0	23764	37.67 %	
		1	35079	55.61 %	
		2	4228	6.70 %	
Conscientiousness		0	31093	39.03 %	
		1	24261	49.29 %	
		2	7357	11.66 %	
Openness to Experience		0	44366	70.34 %	
		1	14424	22.86 %	
		2	4281	6.78 %	
361		Extroversion	0	22471	26.80 %
			1	49122	58.60 %
			2	12226	14.58 %
	Neruoticism	0	25380	30.27 %	
		1	44516	53.10 %	
		2	13923	16.61 %	
	Agreeableness	0	31534	37.62 %	
		1	46679	55.69 %	
		2	5606	6.68 %	
	Conscientiousness	0	32634	38.93 %	
		1	41423	49.41 %	
		2	9762	11.64 %	

400	Openness to Experience	0	58890	70.25 %
		1	19209	22.91 %
		2	5720	6.82 %
	Extroversion	0	24623	26.91 %
		1	53582	58.55 %
		2	13295	14.53 %
	Neruoicism	0	27779	30.35 %
		1	48544	53.05 %
		2	15177	16.58 %
	Agreeableness	0	34498	37.7 %
		1	50967	55.7 %
		2	6035	6.59 %
	Conscientiousness	0	35648	38.95 %
		1	45187	49.38 %
		2	10665	11.65 %
	Openness to Experience	0	64433	70.41 %
		1	20903	22.84 %
		2	6164	6.73 %

Tableau 3. 11: Nombre et pourcentage de chaque classe pour le classement ternaire.

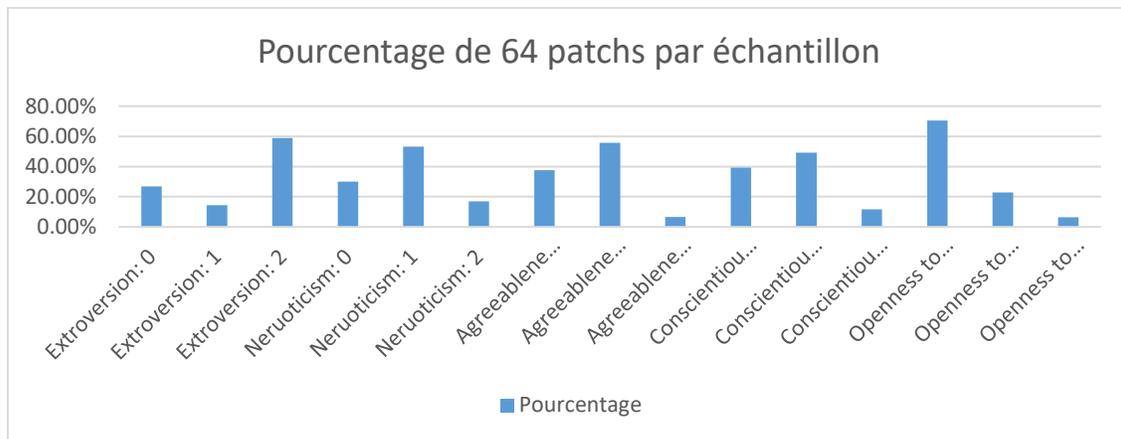


Figure 3. 16 Pourcentage de chaque classe par classement ternaire pour 64 patchs par échantillon.

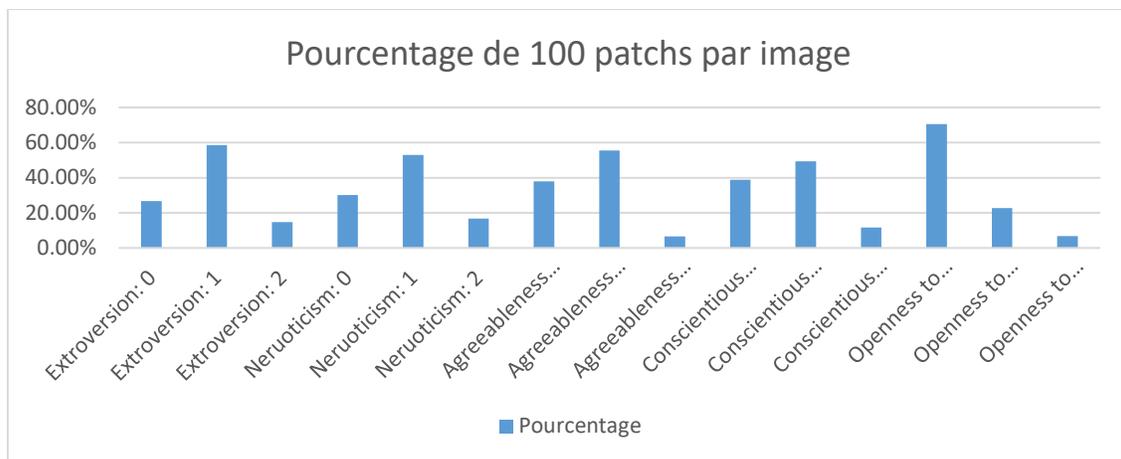


Figure 3. 17 Pourcentage de chaque classe par classement ternaire pour 100 patchs par échantillon.

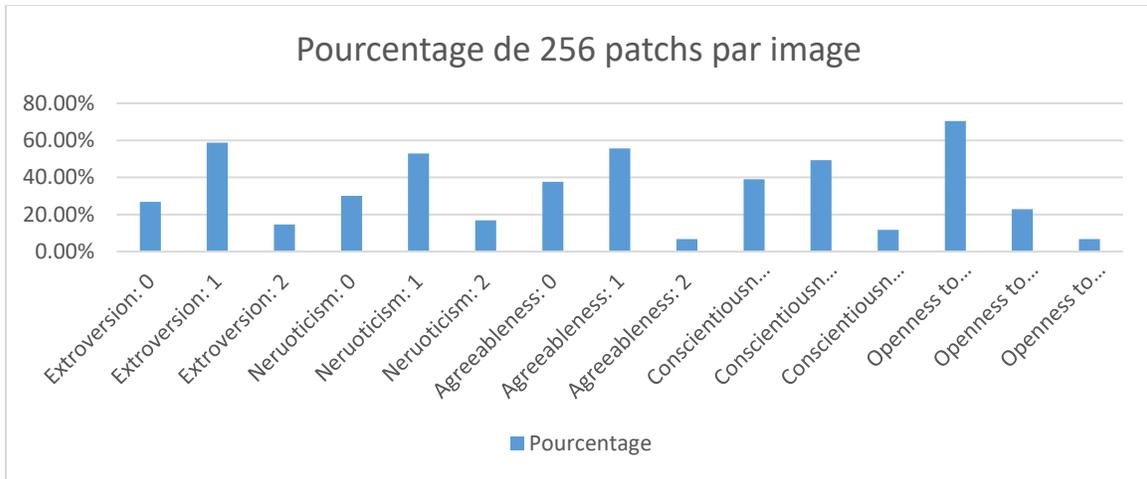


Figure 3.18 Pourcentage de chaque classe par classement ternaire pour 256 patchs par échantillon.

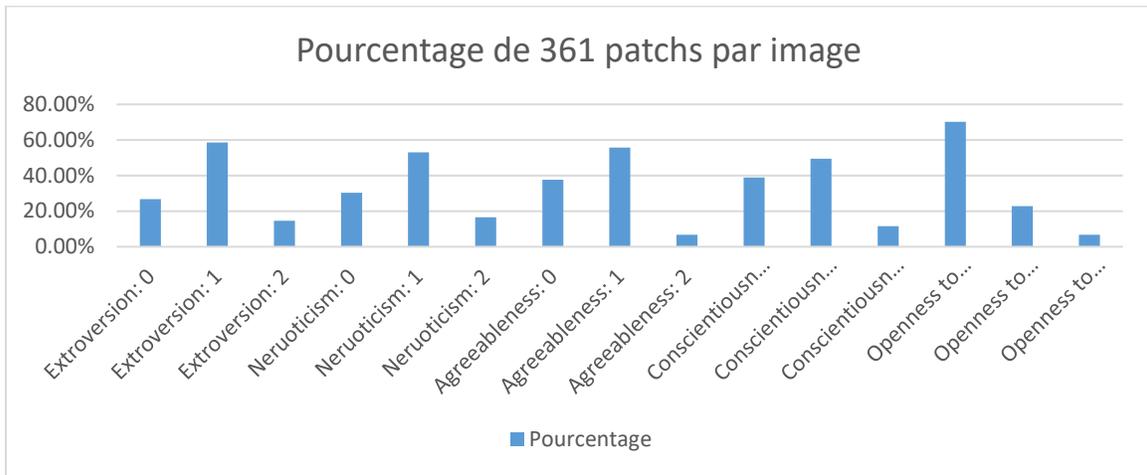


Figure 3.19 Pourcentage de chaque classe par classement ternaire pour 361 patchs par échantillon.

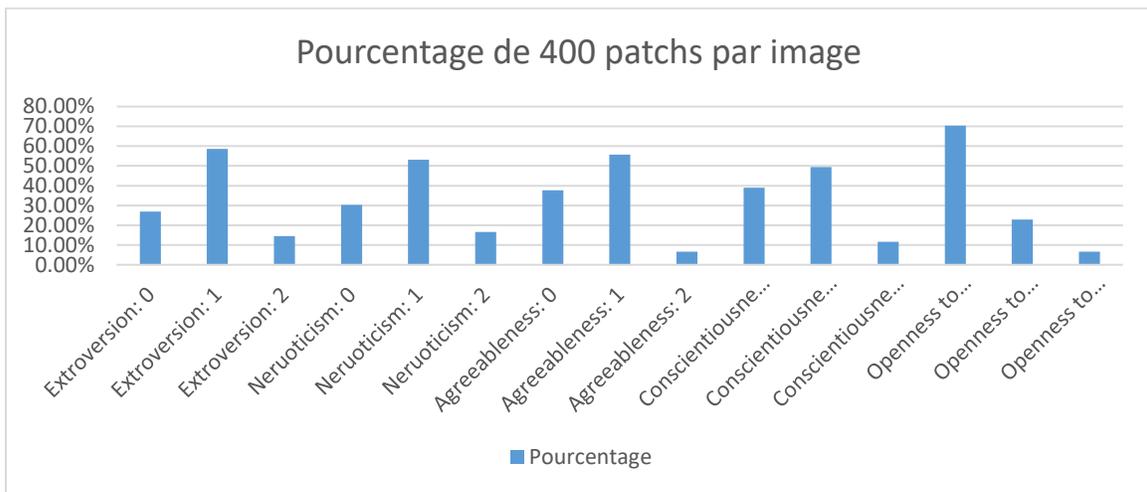


Figure 3.20 Pourcentage de chaque classe par classement ternaire pour 400 patchs par échantillon.

- Classement par quinaire

Patchs par image	Classe	Classement	Nombre	Pourcentage
64	Extroversion	0	1670	9.2 %
		1	4561	24.92 %
		2	6540	35.73 %
		3	4100	22.40 %
		4	1430	7.81 %
	Neruoticism	0	2644	14.44 %
		1	5887	32.16 %
		2	4922	26.89 %
		3	4025	22 %
		4	823	4.49 %
	Agreeableness	0	2533	13.84 %
		1	8001	43.71 %
		2	4603	25.15 %
		3	2818	15.39 %
		4	346	1.89 %
	Conscientiousness	0	2432	13.28 %
		1	6358	34.74 %
		2	5858	32 %
		3	2785	15.21 %
		4	868	4.74 %
Openness to Experience	0	9390	51.30 %	
	1	5496	30 %	
	2	1541	10 %	
	3	1220	6.66 %	
	4	354	1.93 %	
100	Extroversion	0	2424	9.07 %
		1	6671	24.98 %
		2	9497	35.56 %
		3	5972	22.36 %
		4	2137	8 %
	Neruoticism	0	3885	14.55 %
		1	8577	32.12 %
		2	7160	26.81 %
		3	5908	22.12 %
		4	1171	4.38 %
	Agreeableness	0	3731	13.97 %
		1	11656	43.65 %
		2	6707	25.11 %
		3	4123	15.44 %
		4	484	1.81 %
	Conscientiousness	0	3490	13.07 %
		1	9203	34.46 %
		2	8657	32.42 %
		%3	%4072	15.25 %
		4	1279	4.79 %

	Openness to Experience	0	13712	51.35 %
		1	8066	30.20 %
		2	2627	9.83 %
		3	1780	6.66 %
		4	516	1.93 %
256	Extroversion	0	5730	9.08 %
		1	15741	24.95 %
		2	22419	35.54 %
		3	14203	22.51 %
		4	4978	7.89 %
	Neruoticism	0	9133	14.48 %
		1	20384	32.31 %
		2	16865	26.73 %
		3	13876	22 %
		4	2813	4.46 %
	Agreeableness	0	8770	13.9 %
		1	27524	43.63 %
		2	15912	25.22 %
		3	9678	15.34 %
		4	1181	1.88 %
	Conscientiousness	0	8300	13.15 %
		1	21912	34.74 %
		2	20312	32.20 %
		3	9570	15.17 %
		4	2977	4.72 %
Openness to Experience	0	32304	51.21 %	
	1	19006	30.13 %	
	2	6313	10 %	
	3	4214	6.68 %	
	4	1234	1.95 %	
361	Extroversion	0	7660	9.13 %
		1	20916	24.95 %
		2	29881	35.64 %
		3	18732	22.34 %
		4	6630	7.90 %
	Neruoticism	0	12260	14.62 %
		1	27079	32.30 %
		2	22531	26.88 %
		3	18267	21.79 %
		4	3682	4.39 %
	Agreeableness	0	11655	13.90 %
		1	36607	43.67 %
		2	21175	25.26 %
		3	12816	15.29 %
		4	1566	1.86 %
	Conscientiousness	0	11019	13.14 %
		1	29002	34.60 %
		2	27086	32.31 %
		3	12769	15.23 %

	Openness to Experience	4	3943	4.70 %
		0	42868	51.14 %
		1	25232	30.10 %
		2	8442	10 %
		3	5656	6.74 %
	Extroversion	0	8432	9.21 %
		1	22889	25.01 %
		2	32554	35.57 %
		3	20446	22.34 %
		4	7179	7.84 %
	Neruoicism	0	13457	14.70 %
		1	29483	32.22 %
		2	24535	26.81 %
		3	20031	21.89 %
		4	3994	4.36 %
400	Agreeableness	0	12756	13.94 %
		1	39965	43.67 %
		2	23051	25.19 %
		3	14038	15.34 %
		4	1690	1.84 %
	Conscientiousness	0	12034	13.15 %
		1	31610	34.54 %
		2	29665	32.42 %
		3	13882	15.17 %
		4	4309	4.70 %
	Openness to Experience	0	46864	51.21 %
		1	27560	30.12 %
		2	9203	10.05 %
		3	6128	6.69 %
		4	1745	1.90 %

Tableau 3. 12: Nombre et pourcentage de chaque classe pour classement quinaire.

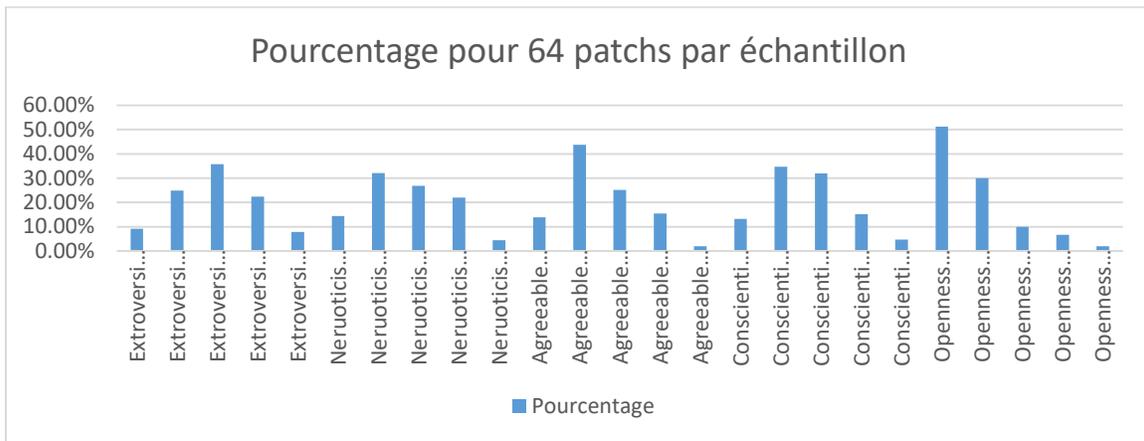


Figure 3. 21 Pourcentage de chaque classe par classement de cinq classes pou quinaire 64 patchs par échantillon.

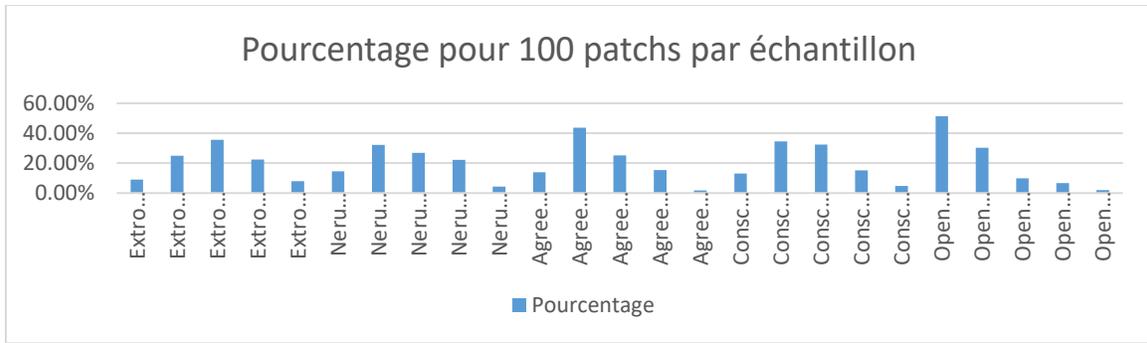


Figure 3. 22 Pourcentage de chaque classe par classement quinaire pour 100 patchs par échantillon.

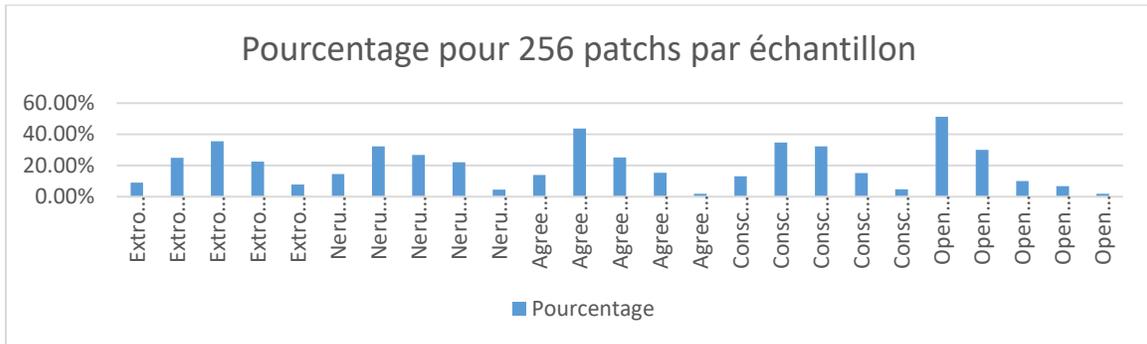


Figure 3. 23 Pourcentage de chaque classe par classement quinaire pour 256 patchs par échantillon.

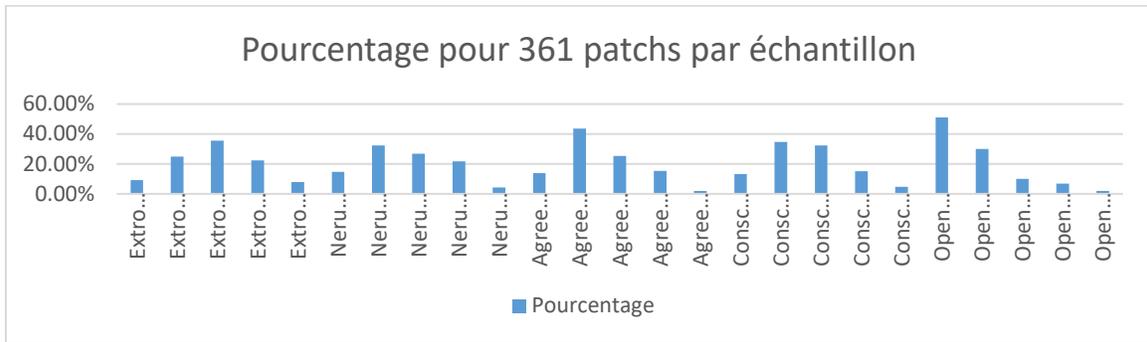


Figure 3. 24 Pourcentage de chaque classe par classement quinaire pour 361 patchs par échantillon.

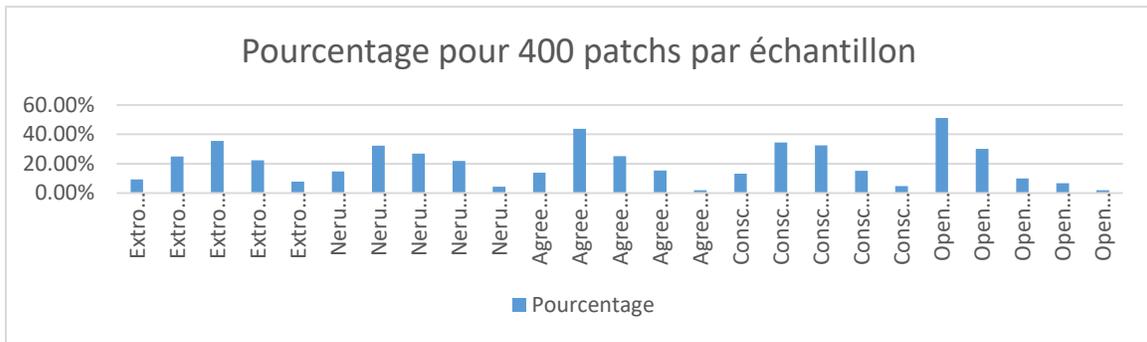


Figure 3. 25 Pourcentage de chaque classe par classement quinaire pour 400 patchs par échantillon.

5. Apprentissage et Classification (premières expériences)

D'après certain nombre de tests appliquer sur plusieurs architectures en démarrant de simple architecture CNN d'une seule couche cachée et en changeant les paramètres globaux à chaque fois passant par ajouter d'autres couches et aussi s'inspirant de plusieurs travaux et architectures vu à partir des recherches faites en 2^{ème} chapitre dans la littérature [15], [33], [34] et d'autres architectures connus dans le domaine de deep learning, en parcourant plusieurs scénarios, en testent plusieurs inputs des patches de différentes tailles...etc. Au final, à chaque test les résultats étaient presque stagnés des fois dans des taux inférieurs et autre fois égaux aux taux des travaux de recherche similaires. Et à partir de là, nous avons décidé d'appliquer l'augmentation des données.

À propos d'appliquer l'augmentation des données pour les images il existe plusieurs techniques, parmi eux on trouve le retournement (Flip), la translation (Shift), la rotation en degré, mise à l'échelle, la luminosité... etc. mais dans le cas de l'écriture manuscrite il y a des techniques qui ne convient pas comme le retournement et la translation et aussi la mise à l'échelle et pour ne pas trop compliqué le processus de on a dû choisir une simple technique de rotation pour l'augmentation.

5.1. Augmentation des données

Après plusieurs exécutions et tests, ces derniers ont menée pour déterminer qu'il faut améliorer les performances et les résultats de notre modèle.

On a choisi de travailler seulement sur la distribution de 256 patches par échantillon de la taille 118×118 et d'introduire une technique d'augmentation simple où l'augmentation est effectuée sur la moitié des données en appliquant une rotation de 35 degré pour ajouter nouveaux échantillons pour avoir en résultat 94 606 patches dans l'ensemble des données alors qu'on avait que 63 071 patches avec les expérimentations précédentes.

Le tableau suivant prend quelques exemples des patches avant et après l'application de l'augmentation des données

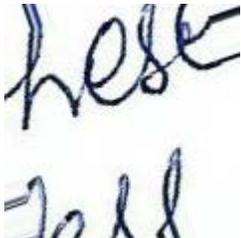
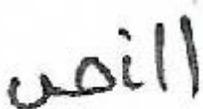
Avant l'augmentation des données	Après l'augmentation des données
	
	
	
	
	

Tableau 3. 8: Exemples des patchs avant et après l'augmentation des données.

6. L'architecture CNN

Cette architecture est un résultat de, comme c'est mentionner dans les deux sections précédentes, l'application de l'augmentation des données avec une série d'expérimentations sur plusieurs semaines.

On va définir notre architecture du réseau CNN utilisé et ses différentes couches sont montré dans la **Figure 3. 26**

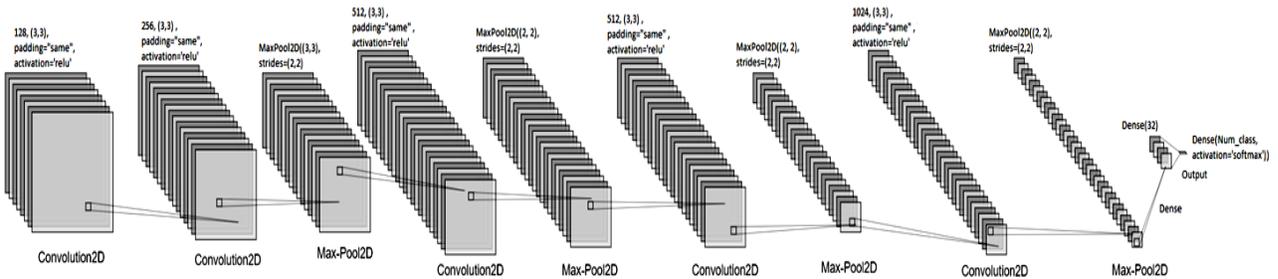


Figure 3. 27 architecture du réseau.

L'architecture proposé sur la Table, consiste à avoir l'ensemble des données des patches originaux de taille 180×180 pixels en entrée. Les images d'entrée sont déjà normalisées et sont transmises au modèle CNN. Le CNN se compose de 11 couches : cinq couches convolutives, quatre couches de max-pooling et une couche entièrement connecté et une couche de sortie.

Structure

La partie d'extraction des caractéristiques :

- Couche en entrée prend en taille de $(180 * 180 * 3)$
- Couches convolutives :
 - Première couche est composée de 128 filtres.
 - Deuxième couche est composée de 256 filtres.
 - Troisième couche est composée de 512 filtres.
 - Quatrième couche est composée de 512 filtres.
 - Cinquième couche est composée de 1024 filtres.
 - La taille de chaque filtre est de $3*3$ pour tous les couches.
 - Fonction d'activation pour tous les couches sont "Relu" pour forcer les neurones à retourner des valeurs positives.

- Couches de max pooling ; A partir de la deuxième couche convolutive, nous plaçons une couche de maxpooling après chaque couche convolutive, la taille de ses filtres suit celle de la couche convolutive qui la précède.

La partie de la classification

Elle est constituée de deux couches entièrement connectées. Cependant, ces couches ne peuvent accepter que des données unidimensionnelles. Pour convertir nos données 3D en 1D, nous allons utiliser la fonction « Flatten », qui va essentiellement remodeler les données de 3D en un vecteur 1D.

Vu que les dernières couches sont des couches entièrement connectées. Elles ont des connexions complètes avec toutes les activations de la couche précédente et la dernière couche c'est la couche de sortie, elle prend cependant la valeur du nombre des classes régulariser.

7. Expérimentations et résultats

Pour l'expérimentation on a dû suivre plusieurs scénarios d'expérimentations, et notre expérience va tester deux scénario :

- Scénario d'expérimentation 1 vise plusieurs scénarios d'inputs des patches avec chaque trait de personnalité tout seul sur plusieurs classements différents.
- Scénario d'expérimentation 2 vise plusieurs tailles d'inputs des patches avec tous les traits de personnalités ensemble sur plusieurs classements différents.

7.1. Scénario d'expérimentation 1

Dans la première expérimentation on va travailler sur tous les classements vus dans la partie de régularisation des classes. Et à chaque fois on test le modèle avec un seul trait de personnalité séparer d'autres traits, et pour chaque apprentissage on effectue le changement seulement sur nos étiquettes, et ce qui rend le résultat de prédiction un vecteur, c'est un apprentissage d'un modèle multi-classe, ensuite on va appliquer un changement sur le taux d'input d'apprentissage et du test avec la fonction **Train_test_split** et c'est une fonction utilisée dans les algorithmes d'apprentissage approfondi pour avoir à partir d'un ensemble globale deux ensemble de données l'un pour l'apprentissage du modèle et l'autre pour le test.

Paramètres Fixés :

- **Batch_Size** =
- **Epoques** =
- **Learning_rate** =

Apprentissage avec les étiquettes binaires (2)

Voilà les résultats des tests d'apprentissage avec les étiquettes binaires sur les différents inputs.

Avec input de 80% pour apprentissage et 20% pour le test

Résultat	Extroversion	Névrosisme	Agréabilité	Conscience	Ouverture à l'expérience
ACCURACY					
LOSS					

Tableau 3. 9: Tableau comparatif des résultats obtenus à base des étiquettes binaires 80/20.

Avec input de 70% pour apprentissage et 30% pour le test

Résultat	Extroversion	Névrosisme	Agréabilité	Conscience	Ouverture à l'expérience
ACCURACY					
LOSS					

Tableau 3. 10: Tableau comparatif des résultats obtenus à base des étiquettes binaires 70/30.

Avec input de 60% pour apprentissage et 40% pour le test

Résultat	Extroversion	Névrosisme	Agréabilité	Conscience	Ouverture à l'expérience
ACCURACY					
LOSS					

Tableau 3. 13: Tableau comparatif des résultats obtenus à base des étiquettes binaires 60/40.

Un exemple sur les taux d'exactitudes et les taux d'erreurs pour les étiquettes binaires

Figure 3. 28 Un exemple sur les taux d'exactitudes et les taux d'erreurs à base des étiquettes binaires.

Apprentissage avec les étiquettes ternaires (3)

Voilà les résultats des tests d'apprentissage avec les étiquettes binaires sur les différents inputs

Avec input de 80% pour apprentissage et 20% pour le test

Résultat	Extroversion	Névrosisme	Agréabilité	Conscience	Ouverture à l'expérience
ACCURACY					
LOSS					

Tableau 3. 14: Tableau comparatif des résultats obtenus à base des étiquettes ternaires 80/20.

Avec input de 70% pour apprentissage et 30% pour le test

Résultat	Extroversion	Névrosisme	Agréabilité	Conscience	Ouverture à l'expérience
ACCURACY					
LOSS					

Tableau 3. 15: Tableau comparatif des résultats obtenus à base des étiquettes ternaires 70/30.

Avec input de 60% pour apprentissage et 40% pour le test

Résultat	Extroversion	Névrosisme	Agréabilité	Conscience	Ouverture à l'expérience
ACCURACY					
LOSS					

Tableau 3. 16: Tableau comparatif des résultats obtenus à base des étiquettes ternaires 60/40.

Un exemple sur les taux d'exactitudes et les taux d'erreurs pour les étiquettes ternaires

Figure 3. 29 Un exemple sur les taux d'exactitudes et les taux d'erreurs à base des étiquettes ternaires.

Apprentissage avec les étiquettes quinaires (5)

Voilà les résultats des tests d'apprentissage avec les étiquettes quinaires sur les différents inputs

Avec input de 80% pour apprentissage et 20% pour le test

Résultat	Extroversion	Névrosisme	Agréabilité	Conscience	Ouverture à l'expérience
ACUURACY					
LOSS					

Tableau 3. 17: Tableau comparatif des résultats obtenus à base des étiquettes quinaires 80/20.

Avec input de 70% pour apprentissage et 30% pour le test

Résultat	Extroversion	Névrosisme	Agréabilité	Conscience	Ouverture à l'expérience
ACCURACY					
LOSS					

Tableau 3. 18: Tableau comparatif des résultats obtenus à base des étiquettes quinaires 70/30.

Avec input de 60% pour apprentissage et 40% pour le test

Résultat	Extroversion	Névrosisme	Agréabilité	Conscience	Ouverture à l'expérience
ACCURACY					
LOSS					

Tableau 3. 19: Tableau comparatif des résultats obtenus à base des étiquettes quinaires 60/40.

Un exemple sur les taux d'exactitudes et les taux d'erreurs pour les étiquettes quinaires

Figure 3. 30 Un exemple sur les taux d'exactitudes et les taux d'erreurs à base des étiquettes quinaires.

7.2. Comparaison des résultats de 1^{er} Scénario

On va présenter des histogrammes comparatifs pour chaque expérimentation tester de différents classements des étiquettes

Comparaison des résultats sur les étiquettes expérimenter avec input de 80% et 20%

Voilà l'histogramme comparatif pour les résultats d'exactitude obtenus par chaque ensembles d'étiquettes

Figure 3. 31 Histogramme comparatif des résultats d'exactitude 80/20.

Comparaison des résultats sur les étiquettes expérimenter avec input de 70% et 30%

Voilà l'histogramme comparatif pour les résultats d'exactitude obtenus par chaque ensembles d'étiquettes

Figure 3. 32 Histogramme comparatif des résultats d'exactitude 70/30.

Comparaison des résultats sur les étiquettes expérimentées avec input de 60% et 40%

Voilà l'histogramme comparatif pour les résultats d'exactitude obtenus par chaque ensemble d'étiquettes

Figure 3. 33 Histogramme comparatif des résultats d'exactitude 60/40.

7.3. Scénario d'expérimentation 2

Dans la deuxième expérimentation on va aussi travailler sur tous les classements vus dans la partie de régularisation des classes. Mais cette fois-ci on entraîne le modèle sur tous les traits de personnalité pendant 80 époques, et ce qui rend le résultat de prédiction sous forme matricielle chaque ligne donne un résultat d'un trait de personnalité et chaque colonne définit le niveau de ce dernier, c'est un entraînement d'un modèle multi-label multi-classe et pour chaque entraînement on effectue un changement sur le taux d'ensembles des données d'entrée d'entraînement et du test pour voir comment ce taux s'influe sur les performances du modèle.

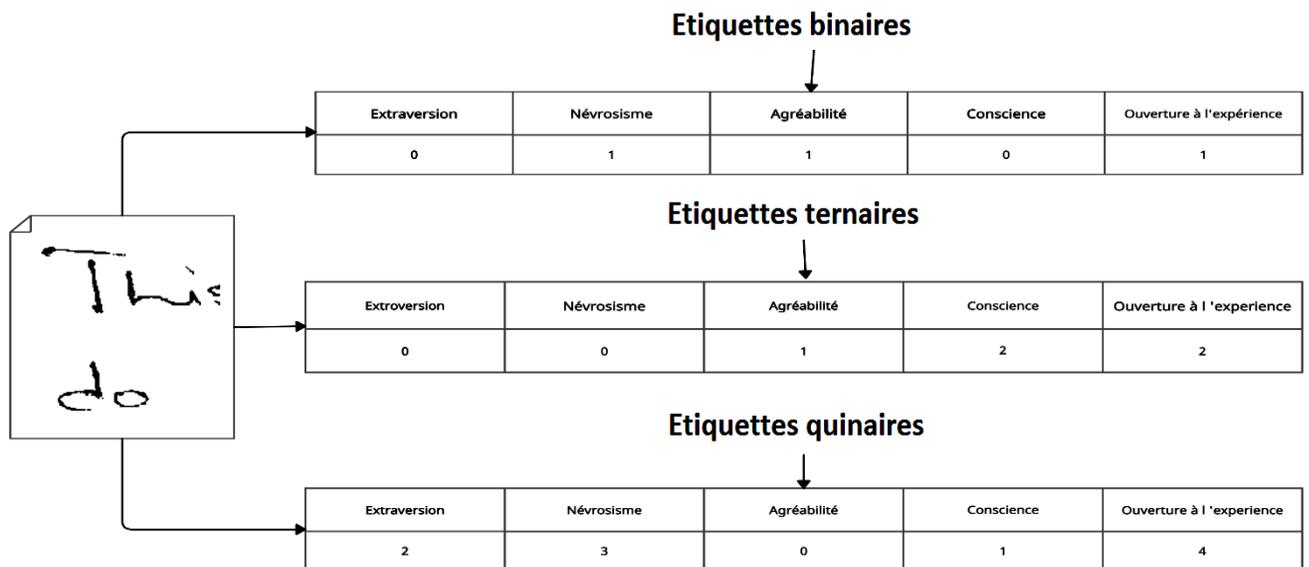


Figure 3. 34 Un patch et ses représentations d'étiquettes.

- **Paramètres Fixés**

Batch_Size =, Epoques =, Learning_rate =

Apprentissage avec les différentes étiquettes

Voilà les résultats des tests d'apprentissage avec les étiquettes binaires et ternaires et quinaires sur les différents inputs

Avec input de 80% pour apprentissage et 20% pour le test

	2 classes	3 classes	5 classes
ACCURACY			
LOSS			
F1-SCORE			
PRECISION			
RECALL			

Tableau 3. 20: Tableau comparatif des résultats obtenus 80/20.

Un exemple sur les taux d'exactitudes et les taux d'erreurs pour le scénario de 80/20

Figure 3. 35 Un exemple sur les taux d'exactitudes et les taux d'erreurs 80/20.

Avec input de 70% pour apprentissage et 30% pour le test

	2 classes	3 classes	5 classes
ACCURACY			
LOSS			
F1-SCORE			
PRECISION			
RECALL			

Tableau 3. 21: Tableau comparatif des résultats obtenus 70/30.

Un exemple sur les taux d'exactitudes et les taux d'erreurs pour le scénario de 70/30

Figure 3. 36 Un exemple sur les taux d'exactitudes et les taux d'erreurs 70/30.

Avec input de 60% pour apprentissage et 40% pour le test

	2 classes	3 classes	5 classes
ACCURACY			
LOSS			
F1-SCORE			
PRECISION			
RECALL			

Tableau 3. 20: Tableau comparatif des résultats obtenus 60/40.

Un exemple sur les taux d'exactitudes et les taux d'erreurs pour le scénario de 60/40

Figure 3. 37 Un exemple sur les taux d'exactitudes et les taux d'erreurs 60/40.

7.1. Comparaison des résultats de 2^{ème} Scénario

On va présenter des histogrammes comparatifs pour chaque expérimentation tester par scénario des différents inputs

Avec un ensemble de données 80% pour apprentissage et 20% pour tester

Figure 3. 38 Histogramme comparatif des résultats 80/20.

Avec un ensemble de données 70% pour apprentissage et 30% pour tester

Figure 3. 39 Histogramme comparatif des résultats 70/30

Avec un ensemble de données 60% pour apprentissage et 40% pour tester

Figure 3. 40 Histogramme comparatif des résultats 60/40.

7.2. Discussion des résultats

Pour le 1^{er} scénario et d'après les résultats obtenus

On remarque que l'ensemble des données... pour l'apprentissage du modèle et... pour le test... Et aussi pour classification par étiquettes Ce qui signifie que la régularisation des classes à un effet sur le taux d'exactitude.

Et pour le 2^{ème} scénario et d'après les résultats obtenus

On remarque que l'ensemble des données... pour l'apprentissage du modèle et... pour le test donnent les meilleurs résultats des taux globaux et aussi pour classification par étiquettes... on remarque que ça donné les meilleurs résultats en termes de taux... par rapport aux autres étiquettes tandis que la classification par étiquettes... a eu le moindre taux d'erreurs ce qui signifie que la régularisation des classes à un effet sur le taux...

On peut conclure que l'augmentation des données a donné un effet... sur le taux... et aussi la régularisation des classes pour l'étiquetage a démontré qu'elle a un effet... sur le taux... et... pour le taux...

7.3. Comparaison avec les travaux ultérieurs

Nous présentons une comparaison de notre système avec les travaux du différentes techniques qui ont été utilisées dernièrement pour prédire les traits de personnalité, en mentionnent l'année et le nombre d'images et le résultat obtenus dans le

Etude par :	Année :	Jeu de données :	Classifieur utiliser :	Nombre d'images :	Résultat :
Lemos et al. [26]	2018	Non mentionné	CNN	Non mentionné	Non mentionné
Ramirez et al [33]	2019	HWxPI	CNN	418 images	50.22 %
Valdez-Rodríguez et al. [34]	2019	HWxPI	CNN	418 images	50.23 %.
UMAIR.S et al. [15]	2020	Propre jeu de données	CNN	285 images	59.73%
Notre approche	2022	Umair's Dataset [15]	CNN	285 images + Data Augmentaion	

Tableau 3. 22: Comparaison des systèmes de prédiction des traits de personnalité à partir de l'écriture manuscrite.

8. Conclusion

Ce travail a présenté une approche d'identification des traits de personnalité à partir de documents manuscrit en appliquant l'apprentissage automatique en profondeur (deep learning), spécifiquement les réseaux de neurones convolutifs, le CNN compare les images tranche par tranche et trouve des caractéristiques approximatives qui sont à peu près similaires dans deux images différentes, le CNN est bien meilleur pour détecter la similitude uniquement par comparaison plein cadre.

Les évaluations ont été effectuées sur un jeu de données contenant des échantillons de texte manuscrit recueillis auprès de 285 sujets différents et en appliquant sur elle un patching étudié suivi d'une augmentation des données par la rotation.

Pour le cas de l'architecture CNN, avec une brève comparaison avec le dernier travail dans le même contexte de recherche, pour notre cas c'était l'utilisation de quatre couches cachées par contre c'était deux couche pour son travail, on remarque bien le point de la différence ce qui signifie que c'est les couches cachées d'un nombre supérieur apprennent des fonctionnalités plus complexes.

Un autre aspect intéressant de cette recherche, l'un est d'appliquer deux méthodes l'une sur chaque trait de personnalité séparément des autres et l'autre sur l'ensemble des traits.

De deux c'est d'extraire certains ensembles de données de ce jeu de données et de l'évaluer sur la même architecture réseau en changeant la couche de sortie.

Si on avait l'accès aux machines plus performant et un jeu des données plus équilibré et plus volumineuse, on aurait fait des tests de plus afin d'obtenir des résultats de classification beaucoup mieux. En fait, pour notre cas des résultats de certains ensembles de données ont surpassé d'autres.

Nos contributions présentées dans ce travail nous permettent d'atteindre des résultats prometteurs, mais ouvrent également plusieurs pistes qui pourront être exploitées dans le futur.

**FUTURES PERSPECTIVES
ET CONCLUSION
GENERALE**

1. Perspectives futures

Cette étude est une recherche scientifique sur le thème de la classification automatique des traits de la personnalité à partir de l'écriture ; néanmoins, il reste du travail à faire pour améliorer l'efficacité et la précision. Le déploiement du système d'identification était prévu, mais il en est encore au stade de la planification. Pour la suite de ce modeste travail, nous espérons étendre le système d'identification des traits de la personnalité en créant un jeu de données plus volumineuse arabe afin de pouvoir l'introduire dans de nombreux domaines tels que les services de recrutement, santé, l'éducation, ... etc.

D'autre part, il est intéressant d'utiliser d'autres technique d'augmentation de données supplémentaires mais il faut garder à l'esprit que les performances du système proposé dépendent non seulement de la technique d'augmentation des données, mais aussi de la taille des patchs choisis et d'autres aspects.

Dans ce cas, ce serait fascinant de combiner les techniques proposées avec d'autres existantes pour améliorer les performances du système proposé.

2. Conclusion générale

L'écriture manuscrite c'est un moyen de communication et d'autre part c'est un outil d'identification des traits de personnalité.

L'identification des traits de personnalités par la machine minimise les couts et facilite plusieurs taches dans plus d'un domaine comme les stratégies du recrutement, les domaines sanitaires où d'autres domaines.

Pour réaliser notre projet on a choisi d'utiliser l'apprentissage approfondi et les réseaux de neurones convolutionnels (CNN) comme un algorithme et technique d'apprentissage pour identifier les traits de personnalité à partir de documents manuscrits.

Des exemples de documents manuscrits de 258 auteurs liées directement aux trait des personnalités extraites à partir les formulaires de modèle à cinq facteurs « FFM / BIG FIVE » ont été utilisés avec une simple technique d'augmentation pour évaluer les performances du système proposé.

On a atteint un taux d'exactitude après l'étude effectuée, donc comme conclusion nous pouvons dire que l'apprentissage approfondi et l'augmentation des données ont donné un résultat notable et pour l'obtention d'une meilleure exactitude de classification des données nécessitent un jeu de données beaucoup plus grande.

Note :

La version complète de cette thèse a été discutée avec les membres des jurys et la raison du manque de résultats dans cette version est que ce travail a été nominé pour publication dans une revue scientifique. A tous ceux qui s'intéressent vraiment à ce travail et à ses résultats veuillez nous contacter sur :

- Dr. DJEDDI Chawki : c.djeddi@univ-tebessa.dz
- Mr. GAHMOUSSE AbdEllatif : abdellatif.gahmousse@univ-tebessa.dz
- Mr. MOHAMMEDI Khaled : mohammedi.khaled02@gmail.com

Bibliographie:

- [1] S. Yoon, S. Choi, S. H. Cha, and C. C. Tappert, "Writer profiling using handwriting copybook styles," *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, vol. 2005, pp. 600–604, 2005, doi: 10.1109/ICDAR.2005.256.
- [2] K. Amend and M. Ruiz, "Handwriting analysis: The complete basic book," 1980, Accessed: Apr. 23, 2022. [Online]. Available: <https://books.google.com/books?hl=fr&lr=&id=UTR8DwAAQBAJ&oi=fnd&pg=PT15&dq=Handwriting+Analysis+-+Karen+Amend&ots=CC5ojQf8QV&sig=R1Z0XG6xVTbx8kl1HJ1rokTC5EY>
- [3] S. v. Kedar, D. S. Bormane, A. Dhadwal, S. Alone, and R. Agarwal, "Automatic emotion recognition through handwriting analysis: A review," *Proceedings - 1st International Conference on Computing, Communication, Control and Automation, ICCUBEA 2015*, pp. 811–816, Jul. 2015, doi: 10.1109/ICCUBEA.2015.162.
- [4] M. Rokeach, "Open and closed mind," 1960, Accessed: Apr. 23, 2022. [Online]. Available: <http://ir.lucknowdigitalibrary.com:8080/xmlui/bitstream/handle/123456789/10411/94052.pdf?sequence=1>
- [5] J. E. Crandall, "Adler's concept of social interest: Theory, measurement, and implications for adjustment," *Journal of Personality and Social Psychology*, vol. 39, no. 3, pp. 481–495, 1980, doi: 10.1037/0022-3514.39.3.481.
- [6] I. B. Myers, M. H. McCaulley, and R. Most, *Manual, a guide to the development and use of the Myers-Briggs type indicator*. consulting psychologists press, 1985.
- [7] G. W. Allport and H. S. Odbert, "Trait-names: A psycho-lexical study.," *Psychological Monographs*, vol. 47, no. 1, pp. i–171, 1936, doi: 10.1037/H0093360.
- [8] A. D. Cattell, H. E., & Mead, "Cattell, H. E., & Mead, A. D. (2008). The sixteen personality factor questionnaire (16PF). The SAGE handbook of personality theory and assessment, 2, 135-178.," *The Sage Handbook of Personality Theory and Assessment*, vol. 2, pp. 135–159, 2008, Accessed: Apr. 23, 2022. [Online]. Available: [http://people.wku.edu/richard.miller/520 16PF Cattell and Mead.pdf](http://people.wku.edu/richard.miller/520%2016PF%20Cattell%20and%20Mead.pdf)
- [9] L. R. Goldberg, "The Structure of Phenotypic Personality Traits," *American Psychologist*, vol. 48, no. 1, pp. 26–34, 1993, doi: 10.1037/0003-066X.48.1.26.
- [10] E. C. Djamal, R. Darmawati, and S. N. Ramdhan, "Application image processing to predict personality based on structure of handwriting and signature," *Proceeding - 2013 International Conference on Computer, Control, Informatics and Its Applications: "Recent Challenges in Computer, Control and Informatics"*, IC3INA 2013, pp. 163–168, 2013, doi: 10.1109/IC3INA.2013.6819167.
- [11] A. Gahmousse, R. Yousfi, and C. Djeddi, "Handwriting Based Personality Traits Identification Using Adaptive Boosting and Textural Features," *Pattern Recognition and Artificial Intelligence*.

Bibliographie

- MedPRAI 2021. Communications in Computer and Information Science, vol 1543. Springer, Cham., pp. 216–227, 2022, doi: 10.1007/978-3-031-04112-9_16.*
- [12] M. Fairhurst and M. Abreu, “An investigation of predictive profiling from handwritten signature data,” *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, pp. 1305–1309, 2009, doi: 10.1109/ICDAR.2009.94.
- [13] T. Mekhaznia, C. Djeddi, and S. Sarkar, “Personality Traits Identification Through Handwriting Analysis,” *Communications in Computer and Information Science*, vol. 1322 CCIS, pp. 155–169, 2021, doi: 10.1007/978-3-030-71804-6_12/FIGURES/9.
- [14] A. Gahmousse, A. Gattal, C. Djeddi, and I. Siddiqi, “Handwriting based Personality Identification using Textural Features,” *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy, ICDABI 2020*, Oct. 2020, doi: 10.1109/ICDABI51230.2020.9325664.
- [15] S. 'Umair, “Identification of personality traits using computerized analysis of handwriting,” Master thesis, Bahria University, 2020.
- [16] C. DJEDDI, “Contribution à l’analyse et la caractérisation de l’écriture manuscrite,” BADJI MOKHTAR UNIVERSITY, ANNABA, 2014.
- [17] L. Thomas, “Sur Quelques Problèmes d’Apprentissage Supervisé et Non Supervisé,” *Mathématiques appliquées*, UNIVERSITE MONTPELLIER II, MONTPELLIER, 2009.
- [18] K. K. Jha, R. Jha, A. K. Jha, M. A. M. Hassan, S. K. Yadav, and T. Mahesh, “A Brief Comparison on Machine Learning Algorithms Based on Various Applications: A Comprehensive Survey,” *CSITSS 2021 - 2021 5th International Conference on Computational Systems and Information Technology for Sustainable Solutions, Proceedings*, 2021, doi: 10.1109/CSITSS54238.2021.9683524.
- [19] “A review of supervised machine learning algorithms | IEEE Conference Publication | IEEE Xplore.” <https://ieeexplore-ieee-org.sndl1.arn.dz/document/7724478> (accessed Jun. 06, 2022).
- [20] L. Deng and D. Yu, “Deep Learning: Methods and Applications,” *Foundations and Trends® in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014, doi: 10.1561/20000000039.
- [21] “Introduction to artificial neural networks,” *Proceedings - Electronic Technology Directions to the Year 2000, ETD 1995*, pp. 36–62, 1995, doi: 10.1109/ETD.1995.403491.
- [22] M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional Networks arXiv:1311.2901v3 [cs.CV] 28 Nov 2013,” *Computer Vision–ECCV 2014*, vol. 8689, no. PART 1, pp. 818–833, 2014, doi: 10.1007/978-3-319-10590-1_53.
- [23] M. A. el Mrabet, K. el Makkaoui, and A. Faize, “Supervised Machine Learning: A Survey,” *Proceedings - 4th International Conference on Advanced Communication Technologies and Networking, CommNet 2021*, 2021, doi: 10.1109/COMMNET52204.2021.9641998.
- [24] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, “A Survey on Deep Transfer Learning: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4–7, 2018, Proceedings, Part III,” 2018, pp. 270–279. doi: 10.1007/978-3-030-01424-7_27.

Bibliographie

- [25] S. Thomas, M. Goel, A. Agarwal, and A. A. Hazariwala, "Application of Machine Learning to Detect Neuroticism in Individuals Using Handwriting Analysis," *Lecture Notes in Electrical Engineering*, vol. 783, pp. 521–531, 2022, doi: 10.1007/978-981-16-3690-5_46/TABLES/3.
- [26] N. Lemos, K. Shah, R. Rade, and D. Shah, "Personality Prediction based on Handwriting using Machine Learning," *Proceedings of the International Conference on Computational Techniques, Electronics and Mechanical Systems, CTEMS 2018*, pp. 110–113, Dec. 2018, doi: 10.1109/CTEMS.2018.8769221.
- [27] D. A.-A. (Tamil N. C. I. for and undefined 2008, "Personality Profile Through Handwriting Analysis," *academia.edu*, Accessed: Apr. 23, 2022. [Online]. Available: https://www.academia.edu/download/35958256/20111213045530Graphology-12_4.pdf
- [28] K. Chaudhari and A. Thakkar, "Survey on handwriting-based personality trait identification," *Expert Systems with Applications*, vol. 124, pp. 282–308, Jun. 2019, doi: 10.1016/J.ESWA.2019.01.028.
- [29] C. Jung, *Psychological types*. Taylor and Francis, 2016. doi: 10.4324/9781315512334/PSYCHOLOGICAL-TYPES-CARL-JUNG.
- [30] L. R. Goldberg, "An Alternative 'Description of Personality': The Big-Five Factor Structure," *Journal of Personality and Social Psychology*, vol. 59, no. 6, pp. 1216–1229, 1990, doi: 10.1037/0022-3514.59.6.1216.
- [31] M. Gavrilescu and N. Vizireanu, "Predicting the Big Five personality traits from handwriting," *Eurasip Journal on Image and Video Processing*, vol. 2018, no. 1, pp. 1–17, Dec. 2018, doi: 10.1186/S13640-018-0297-3/TABLES/5.
- [32] G. Ramírez-De-La-Rosa, E. Villatoro-Tello, and H. Jiménez-Salazar, "TxPI-u: A resource for Personality Identification of undergraduates," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 5, pp. 2991–3001, Jan. 2018, doi: 10.3233/JIFS-169484.
- [33] G. Ramírez *et al.*, "Overview of the Multimedia Information Processing for Personality & Social Networks Analysis Contest," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11188 LNCS, pp. 127–139, 2019, doi: 10.1007/978-3-030-05792-3_12/FIGURES/3.
- [34] J. E. Valdez-Rodríguez, H. Calvo, and E. M. Felipe-Riverón, "Handwritten Texts for Personality Identification Using Convolutional Neural Networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11188 LNCS, pp. 140–145, 2019, doi: 10.1007/978-3-030-05792-3_13/TABLES/1.
- [35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017, doi: 10.1145/3065386.
- [36] Colab, "Bienvenue dans Colab !" <https://colab.research.google.com/>
- [37] Tensorflow, "Why TensorFlow." <https://www.tensorflow.org/?hl=fr>
- [38] Tensorflow, "Tf.KERAS." <https://www.tensorflow.org/guide/keras>