

Handwritten Letter Recognition

ZRIRA Ilyass, BELLANCA Ugo et AHARRAR SOULALI Mohammed

15 janvier 2024

Remerciements

Nous tenons à exprimer notre profonde gratitude envers Guillaume Muller pour son soutien précieux et sa contribution significative à notre projet.

Directeur de stage

- École des Mines de Saint-Étienne : Pr. Mihaela JUGANARU

Mots-clés — Fouille de données (data mining), Apprentissage automatique, Analyse de données, Reconnaissance d'écriture manuscrite, CNN, LSTM

Contents

1	Introduction	4
1.1	Présentation du sujet	4
1.1.1	Projet Initial d'Aidar	4
1.1.2	Stage PROTECH Actuel	4
1.1.3	Objectifs du Stage PROTECH	4
1.2	Méthodologie de Collaboration et Premières Étapes d'Analyse	5
1.3	Présentation des données	5
2	Analyse exploratoire des données	8
2.1	Corrélation des variables	8
2.1.1	Corrélations Intra-Capteurs	8
2.1.2	Corrélations Inter-Capteurs	9
2.1.3	Indépendance des Mesures	9
2.1.4	Optimisation et Sélection des Caractéristiques	9
2.1.5	Exclusion des Données Magnétiques du Modèle d'Apprentissage	9
2.2	Analyse de la Réduction Dimensionnelle via PCA	9
2.2.1	Contribution des Composantes Principales	10
2.2.2	Composantes Principales avec Variance Nulle	10
2.2.3	Impossibilité de Réduire la Dimensionnalité par PCA	10
3	Apprentissage Machine	13
3.1	Stratégie de Modélisation Initiale	13
3.1.1	Choix du Modèle de Base	13
3.1.2	Extraction de Caractéristiques	13
3.1.3	Application de la Forêt Aléatoire	13
3.1.4	Analyse des Résultats du Modèle de Forêt Aléatoire	13
3.2	Évolution vers un Modèle de Réseau de Neurones Récurrents LSTM	15
3.2.1	Modélisation des Données Temporelles avec RNN LSTM	16
3.3	Conception et Mise en Œuvre du Modèle LSTM	16
3.3.1	Prétraitement des Données	16
3.3.2	Architecture du Modèle LSTM	16
3.3.3	Compilation et Entraînement	17
3.3.4	Évaluation du Modèle	17
3.4	Évaluation Détaillée des Performances du Modèle LSTM	17
3.4.1	Résultats du Rapport de Classification	17
3.4.2	Analyse de la Matrice de Confusion	17
3.4.3	Analyse de la Convergence du Modèle	19
3.4.4	Discussion des Performances	20
3.4.5	Conclusions et Perspectives Futures	20
4	Extension du modèle à la prédiction d'un nouvelle utilisateur	21
4.1	Méthodologie	21
4.1.1	Séparation des Données	21
4.1.2	Prétraitement et Architecture du Modèle	21
4.2	Entraînement et Évaluation du Modèle	21

4.3	Résultats et Discussion	21
4.3.1	Rapport de Classification	21
4.3.2	Matrice de Confusion	22
4.3.3	La problème du manque de données	22
5	Conclusion	24
5.1	Bilan du stage	24
5.2	Évaluation des résultats	24
5.2.1	Comparaison avec les Objectifs Initiaux	24
5.2.2	Manque de Données	24
5.3	Réflexions sur les Apprentissages	24
5.4	Défis Rencontrés	25
5.5	Perspectives d'Amélioration et Travaux Futurs	25
5.6	Conclusion Personnelle	25

Chapter 1

Introduction

1.1 Présentation du sujet

1.1.1 Projet Initial d'Aidar

Aidar Gaffarov a précédemment effectué un stage à l'École des Mines de Saint-Étienne en juin 2023, où il a développé un dispositif innovant nommé *Smart Glove*. Ce gant, équipé de divers capteurs, était conçu pour capturer des lettres minuscules écrites à la main. Un aspect crucial de ce projet était la capacité du gant intelligent à fonctionner sans connexion Internet, ce qui a nécessité l'adoption d'une approche d'apprentissage directement sur l'appareil.

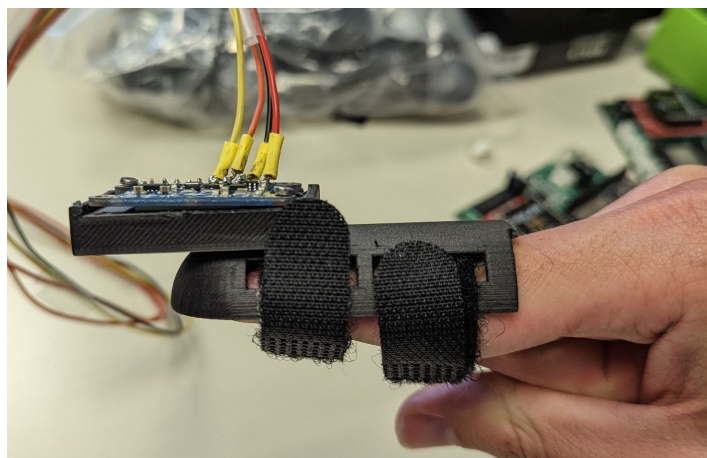


Figure 1.1: Smart Glove

1.1.2 Stage PROTECH Actuel

Dans la continuité de ce projet lors de notre stage PROTECH actuel, nous nous concentrons sur l'analyse exploratoire des données recueillies par les capteurs du *Smart Glove*. Cette phase d'analyse est essentielle pour comprendre en profondeur les caractéristiques des données, tant temporelles que spatiales. Par la suite, nous appliquerons diverses méthodes d'apprentissage automatique afin de prédire efficacement la lettre écrite. Ce stage PROTECH représente donc une évolution et une extension du travail initial d'Aidar, enrichissant le projet avec de nouvelles perspectives et compétences en matière d'analyse de données et d'apprentissage automatique.

1.1.3 Objectifs du Stage PROTECH

Les objectifs initiaux de ce stage sont donc :

- Analyse exploratoire des données fournies par les capteurs.
- Amélioration du modèle d'apprentissage automatique établi par Aidar, jugé insatisfaisant au niveau de son accuracy.

1.2 Méthodologie de Collaboration et Premières Étapes d'Analyse

Nous avons opté pour l'utilisation de GitLab, une plateforme qui a grandement facilité le partage et la gestion de notre code. Des réunions hebdomadaires étaient également organisées avec notre superviseur; elles nous permettaient de présenter l'avancement hebdomadaire de nos travaux et de planifier les étapes à venir.

Les premières phases de notre projet TECH ont été consacrées à une analyse exploratoire des données. Cette étape préliminaire s'est avérée cruciale pour nous familiariser avec les concepts théoriques de la science des données, domaine dans lequel nos connaissances étaient alors naissantes. Par la suite, nous avons développé une gamme de modèles de machine learning, des plus élémentaires aux plus avancés, qui ont finalement conduit à nos résultats les plus prometteurs.

Une comparaison approfondie de nos modèles avec ceux élaborés par Aidar a été effectuée. Forts de cette analyse comparative, nous avons proposé de relever un défi supplémentaire : la prédiction de la reconnaissance de lettres pour des utilisateurs auparavant inconnus du système.

1.3 Présentation des données

La procédure d'enregistrement des données a été rigoureusement définie pour garantir la cohérence et la précision des mesures. Chaque échantillon, correspondant au dessin d'une lettre, a une durée fixée à 1,2 secondes et se compose de 32 points de mesure distincts, permettant une analyse détaillée du mouvement en trois dimensions.

Au total, 16 participants ont activement contribué à l'élaboration de cette base de données. Pour chaque participant, 10 échantillons ont été enregistrés pour 10 lettres différentes, choisies de manière aléatoire. Ce protocole a abouti à la collecte de 100 échantillons par personne, constituant ainsi un ensemble de données global de 1600 échantillons.

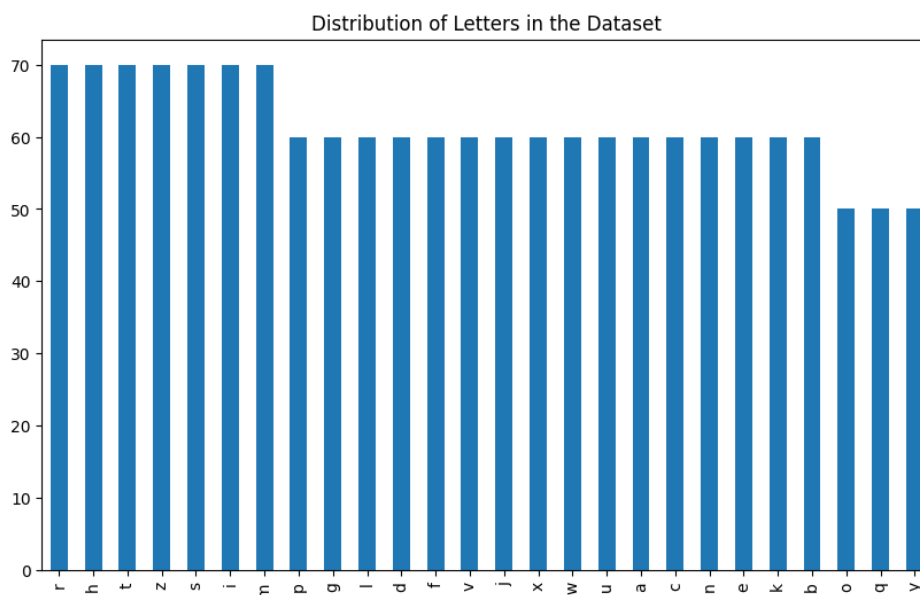


Figure 1.2: Distribution des lettres

Il convient de souligner que l'ensemble des données englobe toutes les lettres de l'alphabet anglais. Cette caractéristique est essentielle, car elle assure une représentation exhaustive et diversifiée des traits distinctifs de l'écriture manuscrite en trois dimensions.

Les données extraites du fichier `multiusers.csv` (fichier des données tirées du stage de Aidar) sont ainsi bien structurées en 1600 lignes et 4 colonnes clés, comme suit :

- **letter** : Cette colonne représente la lettre qui a été écrite en 3D à l'aide du gant intelligent. Chaque ligne correspond à une instance unique d'écriture de cette lettre.
- **features** : Cette colonne contient un ensemble de caractéristiques ou de mesures capturées par le gant intelligent. Les données sont enregistrées sous forme de dictionnaire JSON, incluant par exemple données d'accéléromètres (`acc_x`, `acc_y`, `acc_z`) ou de gyroscopes (`gyro_x`, `gyro_y`, `gyro_z`). Chaque entrée représente un ensemble de mesures pour un essai spécifique d'écriture de la lettre.
- **name** : Cette colonne indique le nom de l'utilisateur qui a réalisé l'essai d'écriture. Elle peut être utile pour identifier les variations ou les similitudes dans les styles d'écriture entre différents utilisateurs.
- **hand** : Cette colonne spécifie la main utilisée pour l'écriture, par exemple, droite ou gauche. Cette information est importante car la main utilisée peut influencer la façon dont la lettre est écrite et, par conséquent, les données capturées par le gant.

Les données exploitées sont issues de capteurs intégrés dans le gant intelligent, conçu pour enregistrer le mouvement de la main en trois dimensions lors de l'écriture de la lettre. Ces données multidimensionnelles se composent de mesures temporelles et spatiales précises, capturant chaque nuance du mouvement. Voici un résumé des types de données disponibles dans chaque enregistrement de la colonne **features** :

- **Accélération** (`acc_x`, `acc_y`, `acc_z`) : Mesure les accélérations linéaires le long des axes X, Y et Z.
- **Magnétisme** (`mag_x`, `mag_y`, `mag_z`) : Fournit les données magnétométriques le long des axes X, Y et Z.
- **Gyroscope** (`gyr_x`, `gyr_y`, `gyr_z`) : Enregistre les taux de rotation autour des axes X, Y et Z.
- **Orientation basée sur l'accéléromètre** (`acc_pitch`, `acc_heading`, `acc_roll`) : Ces mesures représentent l'orientation de l'objet (tangage, cap et roulis) déterminée par l'accéléromètre. L'accéléromètre, en mesurant l'accélération dans trois directions orthogonales, permet d'évaluer comment l'orientation de l'objet change par rapport à ces axes.
- **Orientation basée sur le gyroscope** (`gyr_pitch`, `gyr_heading`, `gyr_roll`) : Fournit des mesures d'orientation similaires, mais calculées à partir du gyroscope. Le gyroscope mesure la vitesse angulaire, offrant une perspective complémentaire à celle de l'accéléromètre, en particulier pour les mouvements rotationnels.
- **Temps** (`time`) : Marques temporelles pour chaque mesure, en millisecondes et depuis un certain point de départ.

L'intégration de ces différentes mesures offre une représentation exhaustive des mouvements manuscrits en trois dimensions, ouvrant la voie à des analyses détaillées et à la reconnaissance des lettres écrites à l'aide de techniques d'apprentissage automatique, c'est tout l'objet de notre PROTECH.

	acc_x	acc_y	acc_z	mag_x	mag_y	mag_z	gyr_x	gyr_y	gyr_z
mean	-4.2	0.1	8.2	2.0	25.1	-21.4	0.2	0.2	-0.1
std	4.2	4.3	4.4	19.5	11.6	10.0	1.2	2.7	1.6
min	-20.0	-20.0	-20.0	-40.6	-21.5	-61.6	-6.05	-17.4	-8.0
25%	-7.3	-2.5	5.7	-12.7	17.6	-27.5	-0.4	-1.0	-0.8
50%	-4.3	-0.2	8.3	0.1	26.3	-21.6	0.1	0.1	-0.1
75%	-1.0	2.3	10.4	15.6	33.1	-16.1	0.8	1.7	0.6
max	8.7	20.0	20.0	92.9	67.7	46.3	8.3	16.2	8.2

Figure 1.3: Summary() de **features**

Chapter 2

Analyse exploratoire des données

2.1 Corrélation des variables

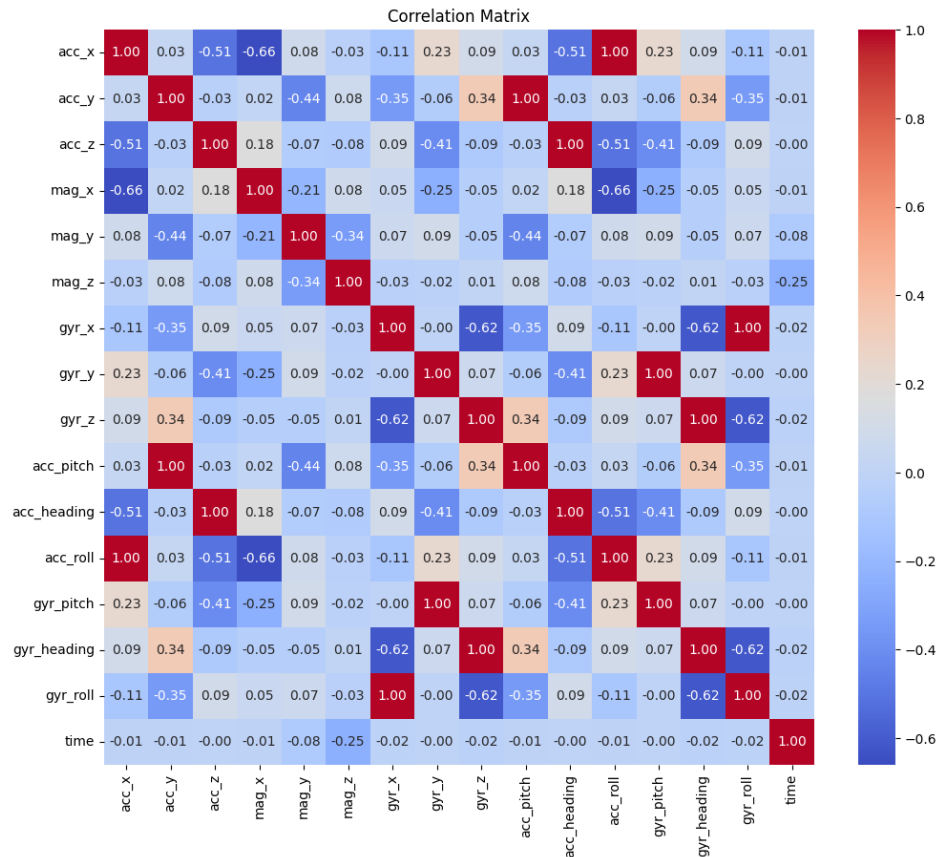


Figure 2.1: Corrélation des **features**

En examinant la matrice de corrélation pour les mesures sensorielles enregistrées par le gant intelligent, nous pouvons identifier les relations entre les différentes mesures qui sont cruciales pour la conception d'un algorithme d'apprentissage automatique efficace. Ci-dessous, nous discutons des implications de ces corrélations pour notre modèle d'apprentissage.

2.1.1 Corrélations Intra-Capteurs

Les valeurs de corrélation parfaites le long de la diagonale principale sont attendues puisqu'elles représentent la corrélation d'une mesure avec elle-même. Notamment, des corrélations négatives

modérées sont observées entre `acc_x` et `acc_z`, suggérant une relation inverse entre ces axes pendant les mouvements d'écriture.

2.1.2 Corrélations Inter-Capteurs

Une forte corrélation négative entre `acc_x` et `mag_x` indique que les mouvements capturés par l'accéléromètre le long de l'axe X sont inversement liés aux mesures magnétiques sur le même axe. Cette relation peut être exploitée pour mieux comprendre comment l'orientation du gant influence les champs magnétiques détectés, mais ce n'est pas l'objet du stage PROTECH.

2.1.3 Indépendance des Mesures

Des corrélations très faibles entre certaines mesures, comme `gyr_y` par rapport à `mag_y` et `mag_z`, indiquent que ces capteurs fournissent des informations distinctes. Cela est essentiel pour un modèle d'apprentissage, car cela signifie que ces caractéristiques peuvent contribuer à une diversité d'informations.

2.1.4 Optimisation et Sélection des Caractéristiques

Les corrélations observées suggèrent des redondances potentielles, notamment entre les mesures d'accélération et magnétiques, qui offrent une opportunité de réduire la dimensionnalité des données. Cette simplification est avantageuse pour l'entraînement de modèles d'apprentissage automatique, comme les réseaux de neurones, où la sélection de caractéristiques avec des corrélations faibles à modérées est essentielle pour éviter la multicollinéarité et préserver la diversité des informations.

De plus, la nécessité d'optimiser les modèles en fonction des ressources computationnelles disponibles commande de combiner ou de transformer les caractéristiques corrélées pour alléger la complexité computationnelle. Cela est particulièrement pertinent pour les applications embarquées telles que le gant intelligent, où les ressources de calcul sont souvent limitées.

2.1.5 Exclusion des Données Magnétiques du Modèle d'Apprentissage

Sur la base de l'analyse de la matrice de corrélation et des considérations d'optimisation des ressources computationnelles, une décision stratégique a été prise d'exclure les données magnétiques de notre modèle d'apprentissage. Cette décision est justifiée par plusieurs facteurs clés identifiés lors de l'évaluation préliminaire des caractéristiques :

- Les mesures magnétiques, bien que fournissant une perspective distincte sur l'orientation spatiale, montrent des corrélations élevées avec les mesures d'accélération. Cette redondance suggère que les informations essentielles capturées par les magnétomètres sont déjà représentées dans d'autres caractéristiques sensorielles.
- La sensibilité des magnétomètres aux interférences externes peut introduire du bruit dans les données, ce qui pourrait potentiellement dégrader la performance du modèle. (C'est effectivement le cas d'ailleurs)
- Les contraintes de calcul imposées par le matériel disponible nécessitent une sélection rigoureuse des caractéristiques qui maximise l'information tout en minimisant la complexité.

En conséquence, l'accent est mis sur l'exploitation des données d'accélération et de gyroscope qui offrent des informations complémentaires et robustes pour l'apprentissage de la dynamique de l'écriture en 3D. Cette approche ciblée permet de construire un modèle plus efficace et adapté aux capacités de traitement de notre système embarqué.

2.2 Analyse de la Réduction Dimensionnelle via PCA

L'analyse des composantes principales (PCA) effectuée sur les données issues du gant intelligent révèle des aspects importants pour la réduction de dimensionnalité en lien avec l'apprentissage automatique.

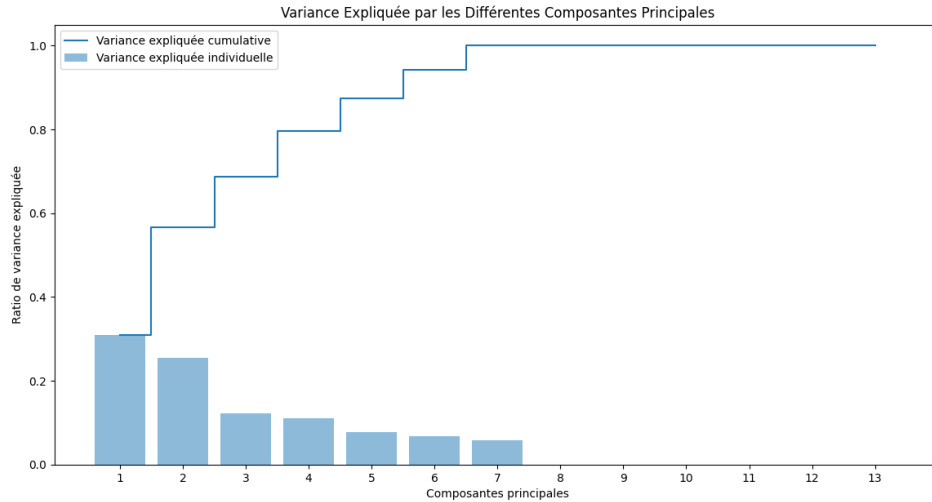


Figure 2.2: PCA des 16 **features**

2.2.1 Contribution des Composantes Principales

Les premières composantes du PCA capturent une part peu importante de la variance des données. Cette observation suggère que ces composantes ne contiennent pas la majorité des informations pertinentes et qu'une réduction dimensionnelle, en se concentrant sur ces premières composantes, ne conserverait pas les caractéristiques essentielles des données bien que simplifiant le modèle.

2.2.2 Composantes Principales avec Variance Nulle

Il est notable que six composantes principales présentent une variance expliquée nulle. Ces composantes correspondent vraisemblablement aux six variables d'orientation d'Euler, calculées à partir des mesures de l'accéléromètre et du gyroscope. Ces variables n'apportent pas de variance additionnelle puisqu'elles sont dérivées directement des données d'accélération et de rotation (Les effets d'échelles ne jouent pas puisque les données ont été standardisées). Par conséquent, elles ne contribuent pas d'informations supplémentaires au-delà de ce qui est déjà capturé par les variables d'origine.

2.2.3 Impossibilité de Réduire la Dimensionnalité par PCA

Comme illustré à la Figure 2.3, une tentative de clustering des données en deux dimensions à l'aide de PCA ne permettrait pas une séparation claire des 26 lettres de l'alphabet.

Insuffisance de la Réduction à Deux Dimensions

Il est évident que la réduction à deux composantes principales ne parvient pas à capturer la complexité inhérente aux données d'écriture manuscrite. Les mouvements de la main et les caractéristiques stylistiques de l'écriture sont de nature multidimensionnelle et ne peuvent être fidèlement représentés dans un espace bidimensionnel sans une perte d'information notable.

Complexité des Données et Modélisation

Les données issues du gant intelligent renferment des attributs subtils qui sont essentiels à la reconnaissance des lettres et qui sont malheureusement omis lors d'une réduction dimensionnelle excessive. Des modèles qui prennent en compte la temporalité et la séquence des données, tels que les réseaux de neurones récurrents, seront nécessaires pour une modélisation adéquate.

Conclusion

La réduction dimensionnelle à deux ou trois dimensions principales est insuffisante pour le problème de classification des lettres. Cela met en évidence la nécessité d'une approche plus nuancée pour la

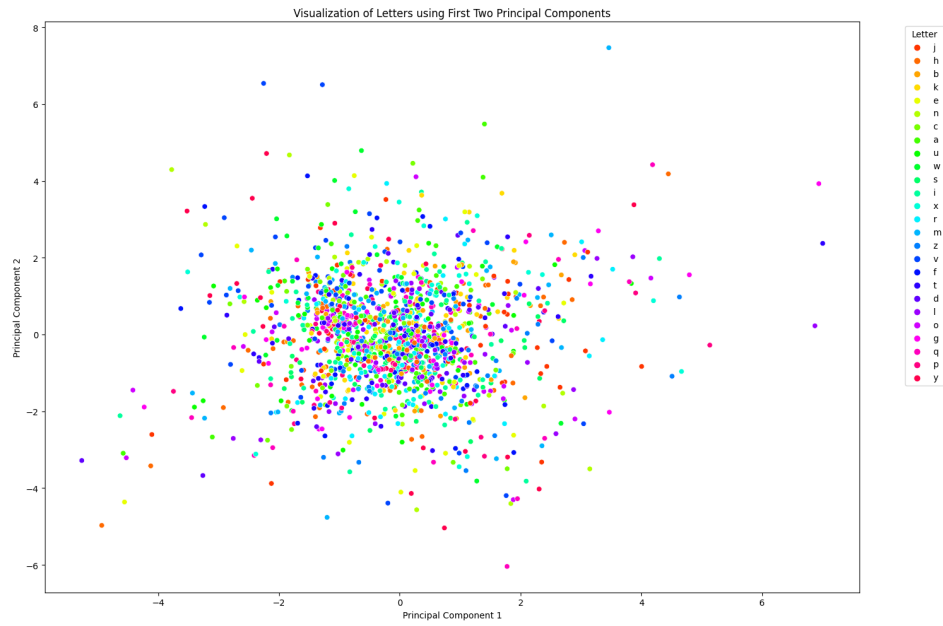


Figure 2.3: Classification des 26 lettres dans les 2 premières composantes principales de la PCA sur **features**

réduction de la dimensionnalité et la sélection de caractéristiques, en tenant compte de la complexité et de la richesse des données de mouvement manuscrit.

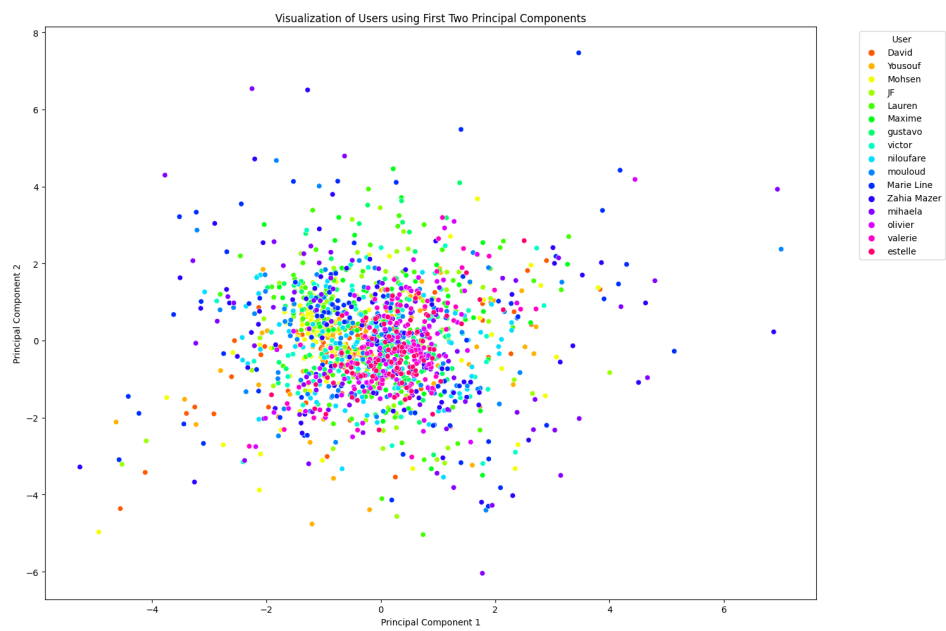


Figure 2.4: Classification des 16 utilisateurs dans les 2 premières composantes principales de la PCA sur **features**

Chapter 3

Apprentissage Machine

3.1 Stratégie de Modélisation Initiale

Dans le cadre initial de notre recherche, nous adoptons une approche pragmatique visant à établir une base de référence pour la précision de la classification que nous pouvons atteindre avec un modèle relativement simple et efficace. Cette étape initiale est cruciale pour évaluer la faisabilité du projet et pour établir un point de comparaison pour des modèles plus complexes qui seront développés ultérieurement.

3.1.1 Choix du Modèle de Base

Le modèle de base sera conçu pour être rapide et efficace, offrant ainsi une estimation initiale de l'exactitude (*accuracy*) que nous pouvons espérer obtenir avec les données du gant intelligent. Cette démarche nous permettra de comprendre rapidement les défis inhérents à la tâche de classification et d'identifier les axes d'amélioration potentiels pour des modèles plus avancés.

3.1.2 Extraction de Caractéristiques

Pour chaque enregistrement de lettre, nous extrairons les moments statistiques clés à partir des trois axes des capteurs d'accéléromètre, de gyroscope, et des mesures d'orientation d'Euler. Ces caractéristiques comprennent la moyenne, la médiane, l'écart-type, le minimum, le maximum, la range (max - min), l'asymétrie et la kurtosis pour chaque axe X, Y et Z.

3.1.3 Application de la Forêt Aléatoire

Afin d'évaluer la performance de notre modèle de base, nous implémenterons une forêt aléatoire. Les données seront divisées en deux ensembles : 80% pour l'entraînement et 20% pour les tests. Cette répartition permettra d'entraîner le modèle tout en conservant une portion de données non observées pour évaluer de manière fiable sa performance.

3.1.4 Analyse des Résultats du Modèle de Forêt Aléatoire

Suite à l'entraînement du modèle de forêt aléatoire sur l'ensemble de données constitué de 80% des enregistrements et l'utilisation des 20% restants pour le test, nous avons obtenu des résultats significatifs qui méritent une analyse détaillée.

Matrice de Confusion

La matrice de confusion, comme illustré dans la Figure X, fournit un aperçu de la performance du modèle en termes de prédictions correctes et incorrectes. Chaque colonne représente les instances dans un ensemble de prédiction, tandis que chaque ligne représente les instances dans un ensemble réel. Les valeurs sur la diagonale de la matrice reflètent le nombre de prédictions correctes pour chaque lettre, et les autres cellules montrent les cas où les lettres ont été confondues avec d'autres par le modèle.

Il est à noter que certaines lettres montrent une haute précision de classification, tandis que d'autres présentent une confusion significative, ce qui suggère que le modèle peut avoir des difficultés avec des lettres ayant des caractéristiques de mouvement similaires, ainsi la lettre la plus confondu est m avec n comme on pouvait s'en douter.

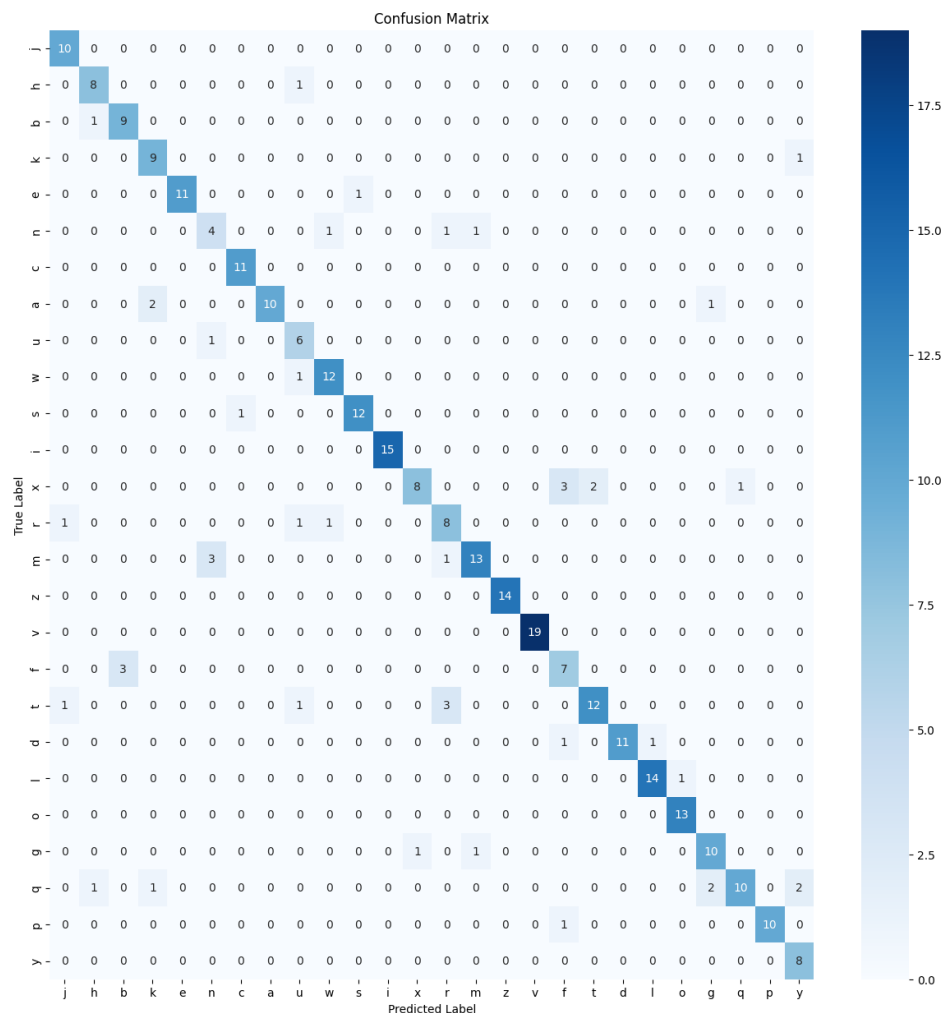


Figure 3.1: Matrice de Confusion du modèle de Forêt Aléatoire sur les moments statistiques des features

Rapport de Classification

Le rapport de classification, présenté à la Figure 3.1, récapitule la précision, le rappel et le score F1 pour chaque lettre, ainsi que les moyennes pour l'ensemble du modèle. Ces métriques fournissent une compréhension plus fine de la performance du modèle pour chaque classe. Par exemple, les lettres avec un score F1 élevé sont bien distinguées par le modèle, tandis que celles avec des scores plus bas peuvent nécessiter une attention supplémentaire ou indiquer des zones où les données d'entraînement sont insuffisantes ou trop similaires à d'autres classes.

Interprétation Globale

L'accuracy globale du modèle s'élève à 0.85625, ce qui est un résultat prometteur pour une première itération. Cependant, l'objectif de cette étape de modélisation n'était pas uniquement d'atteindre une haute précision, mais également de comprendre les limites du modèle de base et d'identifier les opportunités pour des améliorations. Les résultats indiquent une distinction efficace pour certaines lettres, tandis que d'autres semblent être sources de confusion pour le modèle. Ces résultats serviront

Letter	Precision	Recall	F1-Score	Support
a	1.00	0.77	0.87	13
b	0.75	0.90	0.82	10
c	0.92	1.00	0.96	11
d	1.00	0.85	0.92	13
e	1.00	0.92	0.96	12
f	0.58	0.70	0.64	10
g	0.77	0.83	0.80	12
h	0.80	0.89	0.84	9
i	1.00	1.00	1.00	15
j	0.83	1.00	0.91	10
k	0.75	0.90	0.82	10
l	0.93	0.93	0.93	15
m	0.87	0.76	0.81	17
n	0.50	0.57	0.53	7
o	0.93	1.00	0.96	13
p	1.00	0.91	0.95	11
q	0.91	0.62	0.74	16
r	0.62	0.73	0.67	11
s	0.92	0.92	0.92	13
t	0.86	0.71	0.77	17
u	0.60	0.86	0.71	7
v	1.00	1.00	1.00	19
w	0.86	0.92	0.89	13
x	0.89	0.57	0.70	14
y	0.73	1.00	0.84	8
z	1.00	1.00	1.00	14
Accuracy				
0.85625				
Macro Avg				
	0.85	0.86	0.84	320

Table 3.1: Rapport de Classification du modèle de Forêt Aléatoire

de base pour l’exploration de modèles plus complexes qui pourraient mieux capturer la dynamique de l’écriture manuscrite tridimensionnelle et améliorer la précision de la classification.

Perspectives

Des techniques de modélisation plus avancées, telles que les réseaux de neurones profonds, pourraient être envisagées pour capter les subtilités des mouvements de la main que la forêt aléatoire pourrait ne pas saisir.

3.2 Évolution vers un Modèle de Réseau de Neurones Récurrents LSTM

Après avoir établi une base de référence avec notre modèle de forêt aléatoire, nous orientons notre recherche vers une approche qui considère la nature temporelle des données issues du gant intelligent. Reconnaissant que chaque geste d’écriture constitue une séquence d’événements dynamiques dans le temps, nous avons l’idée d’adopter un modèle qui peut saisir ces dépendances temporelles. Dans cette optique, nous proposons l’utilisation d’un modèle RNN LSTM pour notre analyse.

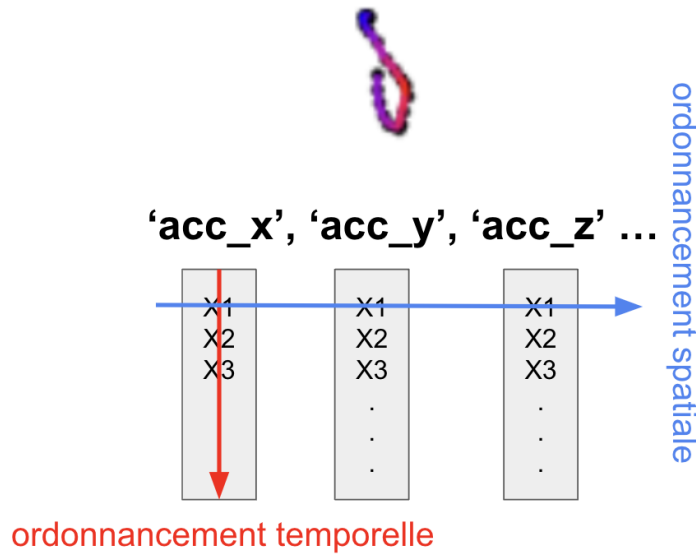


Figure 3.2: Nature spatiale (Aidar) VS temporelle (Ce PROTECH) des données

3.2.1 Modélisation des Données Temporelles avec RNN LSTM

Les réseaux de neurones récurrents, en particulier ceux utilisant des cellules LSTM, sont particulièrement bien adaptés pour modéliser des séquences de données. Les LSTM sont conçues pour surmonter le problème de disparition du gradient, permettant au réseau de se souvenir des informations sur de longues périodes, ce qui est essentiel pour capturer la dynamique complexe de l'écriture manuscrite.

Avantages des LSTM pour les Séries Temporelles

Les LSTM sont capables de reconnaître des motifs dans les séries temporelles, ce qui les rend idéales pour des tâches telles que la reconnaissance de l'écriture, où l'ordre et le contexte des mouvements précédents sont cruciaux pour prédire la suite du geste. Contrairement au modèle de forêt aléatoire, qui traitait les données de manière isolée, le modèle RNN LSTM prendra en entrée l'ensemble des séries temporelles pour chaque geste d'écriture. En traitant les données non pas comme un ensemble statique de points, mais comme une série d'actions interconnectées, les LSTM offrent un cadre robuste pour apprendre des caractéristiques significatives à partir des données temporelles.

3.3 Conception et Mise en Œuvre du Modèle LSTM

Pour exploiter pleinement la nature séquentielle de nos données, nous avons conçu un modèle basé sur un réseau de neurones récurrents à longue mémoire à court terme (LSTM).

3.3.1 Prétraitement des Données

Les données brutes, initialement sous forme de chaînes JSON, sont converties en dictionnaires et structurées en séries temporelles. Chaque série est normalisée en utilisant un `StandardScaler` pour mettre à l'échelle les caractéristiques numériques, essentiel pour une performance optimale de l'apprentissage profond.

3.3.2 Architecture du Modèle LSTM

Le modèle est construit avec l'API `Sequential` de Keras, en commençant par une couche LSTM avec 50 unités et retour de séquences pour permettre la connexion à des couches subséquentes. Une couche de `Dropout` est ajoutée pour réduire le surajustement en omettant aléatoirement des caractéristiques

lors de l'entraînement. La deuxième couche LSTM, sans retour de séquences, condense l'information en préparation pour la classification. La dernière couche, dense, utilise l'activation *softmax* pour une classification multiclasse.

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 32, 50)	12600
dropout (Dropout)	(None, 32, 50)	0
lstm_1 (LSTM)	(None, 50)	20200
dropout_1 (Dropout)	(None, 50)	0
dense (Dense)	(None, 26)	1326

=====
Total params: 34126 (133.30 KB)
Trainable params: 34126 (133.30 KB)
Non-trainable params: 0 (0.00 Byte)

Figure 3.3: Architecture du RNN LSTM

3.3.3 Compilation et Entraînement

Le modèle est compilé avec la fonction de perte *categorical_crossentropy*, adaptée aux problèmes de classification multiclasse, et l'optimiseur *adam* pour l'ajustement des poids. L'entraînement est effectué sur les données d'apprentissage avec des labels encodés en One-Hot, pendant 100 époques avec une taille de lot (ou batch) de 4.

3.3.4 Évaluation du Modèle

Après l'entraînement, le modèle est évalué sur l'ensemble de test. Un rapport de classification détaillé est généré, fournissant la précision, le rappel et le score F1 pour chaque lettre, permettant une analyse approfondie des performances du modèle. De plus, une matrice de confusion est affichée, offrant une visualisation claire des prédictions du modèle par rapport aux vraies étiquettes.

3.4 Évaluation Détaillée des Performances du Modèle LSTM

3.4.1 Résultats du Rapport de Classification

Le tableau 3.2 présente le rapport de classification détaillé obtenu après l'entraînement du modèle LSTM. Chaque lettre de l'alphabet, représentant une classe distincte dans notre ensemble de données, est évaluée en termes de précision, rappel et score F1. Les résultats montrent une performance globale élevée avec une accuracy macro et pondérée de 0.93, ce qui indique une capacité considérable du modèle à généraliser sur des données non vues lors de l'entraînement. Nous avons donc un modèle largement supérieur à celui de Aidar, dont l'accuracy est de 78%. Ainsi, le point de vue temporelle s'avère meilleur que celui spatiale.

3.4.2 Analyse de la Matrice de Confusion

La Figure 3.4 illustre la matrice de confusion générée à partir des prédictions du modèle sur l'ensemble de test. La matrice révèle un nombre élevé de prédictions correctes le long de la diagonale, indiquant une forte concordance entre les labels prédits et les vrais labels. Cependant, certaines confusions sont observées pour des lettres avec des mouvements manuscrits similaires, suggérant des pistes pour une analyse plus approfondie et un raffinement du modèle. Notamment, on remarque que les lettres qui

Lettre	Précision	Rappel	F1-Score	Support
a	0.90	0.75	0.82	12
b	1.00	1.00	1.00	12
c	1.00	0.92	0.96	12
d	0.85	0.92	0.88	12
e	0.86	1.00	0.92	12
f	1.00	0.92	0.96	12
g	1.00	0.92	0.96	12
h	0.93	0.93	0.93	14
i	0.73	0.79	0.76	14
j	0.92	0.92	0.92	12
k	1.00	1.00	1.00	12
l	1.00	1.00	1.00	12
m	0.93	1.00	0.97	14
n	1.00	0.83	0.91	12
o	1.00	0.90	0.95	10
p	1.00	1.00	1.00	12
q	1.00	1.00	1.00	10
r	0.86	0.86	0.86	14
s	1.00	1.00	1.00	14
t	0.93	0.93	0.93	14
u	0.91	0.83	0.87	12
v	0.80	1.00	0.89	12
w	0.85	0.92	0.88	12
x	1.00	0.92	0.96	12
y	1.00	1.00	1.00	10
Accuracy				
0.93				
Macro Avg				
	0.94	0.93	0.93	320

Table 3.2: Rapport de Classification du modèle RNN LSTM

posent problème sont celles auxquelles nous penserions : les "n" confondus avec des "m" ou encore "d" avec "a".

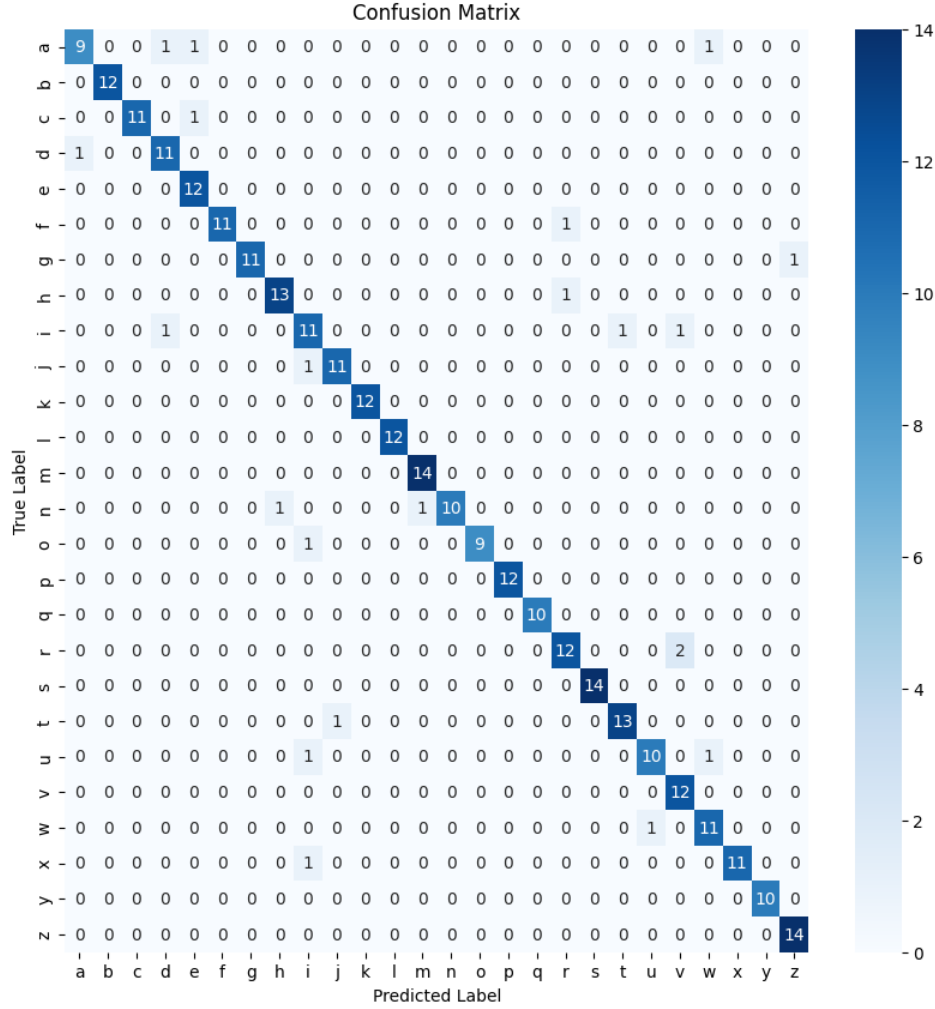


Figure 3.4: Matrice de Confusion du modèle RNN LSTM des **features**

3.4.3 Analyse de la Convergence du Modèle

La performance du modèle LSTM au cours des itérations d'entraînement est cruciale pour évaluer sa capacité à apprendre et à généraliser à partir des données. La Figure 3.5 illustre la courbe de perte du modèle sur l'ensemble d'apprentissage (train loss) et l'ensemble de validation (validation loss) au fil des époques.

Interprétation de la Courbe de Perte

La courbe de perte de l'entraînement montre une chute rapide lors des premières époques, indiquant que le modèle apprend efficacement les caractéristiques pertinentes des données. Par la suite, la perte se stabilise, ce qui suggère que le modèle commence à converger et à atteindre un point d'équilibre entre l'apprentissage et la complexité du modèle.

La courbe de perte de validation présente des fluctuations plus importantes, mais suit généralement une tendance descendante, ce qui est un signe positif indiquant que le modèle ne surajuste pas significativement.

Des recherches futures pourraient explorer des techniques de régularisation adaptées pour minimiser l'écart entre la perte d'entraînement et de validation, afin d'améliorer la performance du modèle sur

des données inédites.

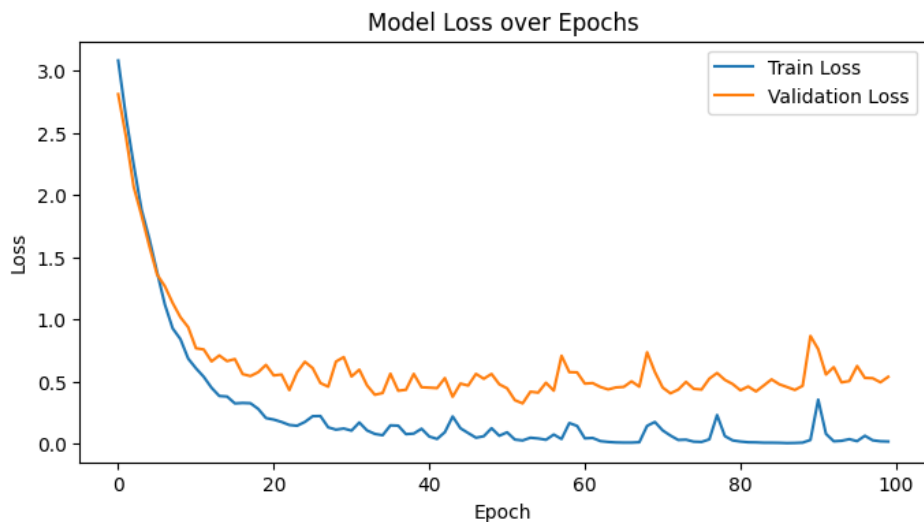


Figure 3.5: Évolution de la perte du modèle LSTM au cours des époques

3.4.4 Discussion des Performances

Les métriques élevées de précision et de rappel pour la majorité des lettres démontrent la capacité du modèle LSTM à apprendre les caractéristiques distinctives de chaque lettre dans l'espace tridimensionnel des mouvements manuscrits. Les lettres avec des scores F1 inférieurs seront sujettes à des investigations supplémentaires pour déterminer si des caractéristiques supplémentaires ou des ajustements du modèle pourraient améliorer la reconnaissance.

3.4.5 Conclusions et Perspectives Futures

L'adoption d'un modèle LSTM pour l'analyse des données temporelles capturées par le gant intelligent s'est avérée être une approche prometteuse, surpassant le modèle de forêt aléatoire précédent. Cette avancée démontre le potentiel des RNN LSTM dans la reconnaissance de l'écriture manuscrite et ouvre des avenues pour de futures améliorations, notamment en explorant des architectures de modèles plus complexes et en intégrant des techniques d'augmentation des données.

Chapter 4

Extension du modèle à la prédiction d'un nouvelle utilisateur

Dans ce chapitre, nous évaluons l'extension de notre modèle LSTM en testant sa capacité à généraliser sur un utilisateur inconnu. L'objectif est de déterminer si le modèle formé sur un ensemble d'utilisateurs peut prédire avec précision les lettres manuscrites par un nouvel utilisateur qui n'a pas contribué aux données d'apprentissage.

4.1 Méthodologie

4.1.1 Séparation des Données

Les données d'un utilisateur unique sont isolées du reste du jeu de données pour servir d'ensemble de test. Cette démarche simule l'application réelle du modèle à un utilisateur dont les données n'ont pas été vues pendant la phase d'entraînement, mettant ainsi à l'épreuve la capacité de généralisation du modèle.

4.1.2 Prétraitement et Architecture du Modèle

Le prétraitement des données et l'architecture du modèle LSTM restent les mêmes que ceux utilisés pour les utilisateurs connus, permettant une comparaison directe des performances.

4.2 Entraînement et Évaluation du Modèle

Le modèle est entraîné sur les données des utilisateurs connus et ensuite évalué sur l'ensemble de test de l'utilisateur inconnu. L'entraînement vise à minimiser la perte catégorique croisée, tandis que l'évaluation se concentre sur la précision, le rappel et le score F1 pour chaque lettre prédite.

4.3 Résultats et Discussion

Les résultats de l'expérimentation sur un utilisateur inconnu du modèle ont été compilés et sont présentés ci-dessous. Ces résultats permettent d'évaluer la capacité du modèle à généraliser à de nouvelles données et à identifier les zones nécessitant des améliorations.

4.3.1 Rapport de Classification

Le tableau suivant présente le rapport de classification détaillé pour chaque lettre prédite par le modèle pour l'utilisateur inconnu. Il fournit les mesures de précision, de rappel, de score F1 et le support pour chaque classe.

Les résultats reflètent des défis notables dans la classification de certaines lettres, indiquant une performance variable du modèle qui pourrait être attribuée à la complexité intrinsèque des données

Table 4.1: Rapport de classification pour les prédictions sur l'utilisateur inconnu

Letter	Precision	Recall	F1-Score	Support
d	0.25	0.10	0.14	10
g	0.57	0.40	0.47	10
h	1.00	0.20	0.33	10
l	0.00	0.00	0.00	10
m	1.00	0.40	0.57	10
p	0.00	0.00	0.00	10
r	0.67	0.20	0.31	10
s	0.75	0.30	0.43	10
t	0.67	0.20	0.31	10
z	1.00	0.30	0.46	10
Accuracy				
0.55				
Macro Avg				
	0.59	0.21	0.30	100

Table 4.2: Rapport de Classification du modèle RNN LSTM

d'écriture manuscrite. On remarque notamment un score de Recall très bas, ce qui traduit une capacité limitée du modèle à détecter toutes les instances réelles de chaque lettre (Comme "l" et "p", où le modèle n'a correctement prédit aucune instance)

4.3.2 Matrice de Confusion

La matrice de confusion, illustrée à la Figure 4.1, offre une visualisation détaillée des performances du modèle en révélant les erreurs de classification spécifiques.

L'analyse de la matrice de confusion montre des erreurs de classification significatives pour certaines lettres comme "b" avec "l", tandis que d'autres sont correctement reconnues. Ces résultats suggèrent que le modèle peut être sujet à un surajustement aux caractéristiques des utilisateurs d'entraînement et pourrait bénéficier d'une généralisation améliorée. Ceci nous vient d'un manque crucial de donnée.

4.3.3 La problème du manque de données

La performance du modèle sur des données d'un utilisateur inconnu met en lumière un défi significatif : la quantité de données disponible pour l'entraînement. Avec seulement 1600 observations en trois dimensions, notre jeu de données pâlit en comparaison avec des benchmarks établis comme le dataset MNIST, qui comprend 60.000 images (et en 2D qui plus est). Il est évident que pour une tâche complexe de reconnaissance de l'écriture en trois dimensions, une augmentation substantielle du volume de données est impérative pour améliorer la généralisation du modèle.

La recherche de méthodes pour enrichir notre jeu de données, que ce soit par des techniques d'augmentation ou par la collecte de nouvelles données, devient alors une priorité. Des datasets plus vastes et diversifiés permettront non seulement d'affiner la précision de notre modèle, mais aussi de mieux représenter la variabilité naturelle des mouvements d'écriture manuscrite.

L'objectif ultime est de bâtir un modèle aussi robuste et fiable que ceux développés pour des datasets comme MNIST, mais adapté à la complexité accrue des données en trois dimensions. Cela ouvrirait la voie à des applications pratiques plus étendues du gant intelligent dans le domaine de la reconnaissance de l'écriture manuscrite.

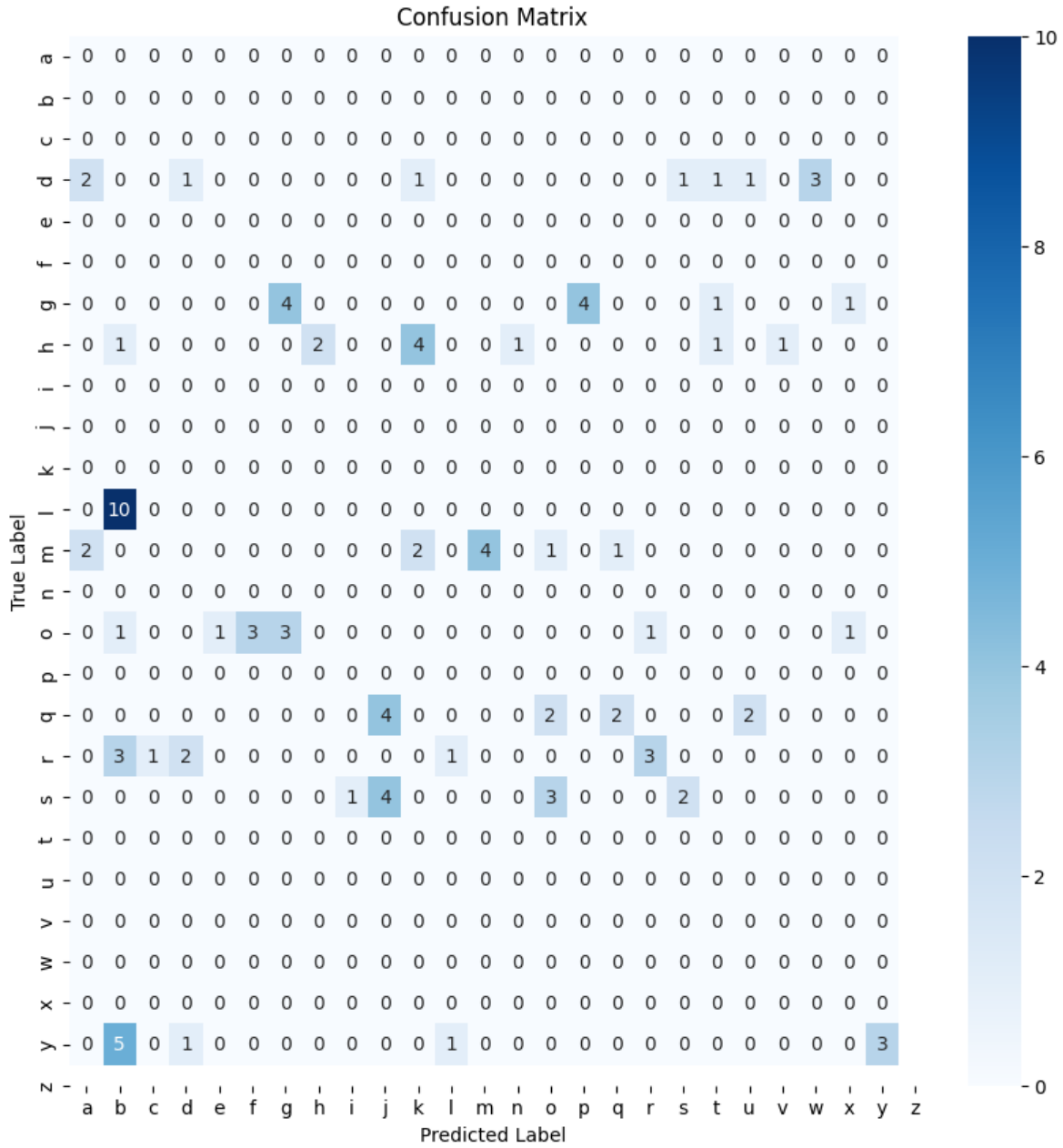


Figure 4.1: Matrice de confusion du modèle LSTM pour l'utilisateur inconnu.

Chapter 5

Conclusion

5.1 Bilan du stage

Le stage PROTECH a été une opportunité précieuse d’appliquer des concepts théoriques à des problèmes pratiques de reconnaissance de l’écriture manuscrite. Ce chapitre résume les résultats obtenus, évalue la réalisation des objectifs initiaux, et reflète sur les apprentissages et les défis rencontrés.

5.2 Évaluation des résultats

Les objectifs du stage étaient ambitieux, visant à améliorer significativement l’exactitude des prédictions de lettres manuscrites grâce au gant intelligent du stage d’Aidar. Les modèles développés au cours du stage, notamment la forêt aléatoire et le réseau de neurones récurrents LSTM, ont démontré des performances prometteuses. Le modèle LSTM, en particulier, a affiché une capacité notable à modéliser la séquence temporelle complexe des mouvements d’écriture, passant d’une accuracy de 78% à 93%.

5.2.1 Comparaison avec les Objectifs Initiaux

Par rapport aux objectifs fixés au début du stage, les résultats montrent une amélioration notable de la précision des prédictions. Cependant, le défi de généraliser le modèle à un utilisateur inconnu reste partiellement résolu, indiquant la nécessité d’une collecte de données plus vaste et d’une exploration continue des méthodes de traitement et de modélisation des données.

5.2.2 Manque de Données

Un constat majeur est le manque relatif de données par rapport à la complexité de la tâche de reconnaissance en 3D. Le stage a mis en évidence que, bien que la technologie du gant intelligent soit avancée, la quantité de données disponibles est un facteur limitant crucial. À titre comparatif, le dataset MNIST offre une quantité substantielle de données qui permettent des avancées significatives dans la reconnaissance d’images 2D. Pour atteindre une telle efficacité, il est impératif d’augmenter le volume et la diversité des données collectées avec le gant intelligent.

5.3 Réflexions sur les Apprentissages

Ce stage a été riche en apprentissages, offrant une immersion dans le domaine de l’apprentissage automatique appliqué à la reconnaissance de l’écriture manuscrite. Les expériences pratiques acquises à travers la mise en œuvre de différents modèles ont renforcé notre compréhension des principes sous-jacents à la fouille de données et à l’intelligence artificielle.

5.4 Défis Rencontrés

Les défis rencontrés durant le stage ont été variés, allant de la compréhension des données complexes en trois dimensions à la mise en œuvre de modèles d'apprentissage profond adaptés à ces données. L'un des défis les plus marquants a été l'adaptation des modèles pour généraliser à un utilisateur inconnu, un problème qui reste ouvert et nécessite une attention particulière pour les développements futurs.

5.5 Perspectives d'Amélioration et Travaux Futurs

Le stage a posé les bases pour de futures recherches et développements dans le domaine. Il est clair que l'amélioration de la précision de la reconnaissance des lettres manuscrites nécessitera une collecte de données plus importante et peut-être l'exploration de nouvelles architectures de réseau neuronal ou de techniques d'apprentissage semi-supervisé.

5.6 Conclusion Personnelle

Personnellement, ce stage a été une opportunité inestimable pour nous tous en termes d'apprentissage et de développement professionnel. Les compétences que nous avons acquises en matière d'analyse de données et de modélisation constitueront un atout précieux pour nos futures carrières dans les domaines de l'ingénierie des données et de l'intelligence artificielle. Chacun de nous peut attester de l'impact significatif que cette expérience a eu sur notre compréhension technique et notre vision de l'avenir de la technologie. Nous sommes reconnaissants pour les défis que nous avons rencontrés et surmontés ensemble, car ils ont renforcé notre collaboration et notre esprit d'équipe, tout en nous préparant à devenir des professionnels compétents et innovants dans nos domaines respectifs.