



MEMOIRE

BOURAS SENIA





MEMOIRE

Pour l'obtention du diplôme de Master
Spécialité : Maintenance, fiabilité et qualité

Thème

***Système électronique liées aux erreurs
des médicaments look-alike***

Devant le jury composé de :

Nom et Prénom	Grade	Etablissement	Qualité
Aouimer Yamna	MCA	Université d'Oran 2	Présidente
MOUFOK Souad	MAA	Université d'Oran 2	Encadreur
HADJ JELLOUL Kheira	MAA	Université d'Oran 2	Examineur



Remercîment

Tout d'abord, je remercie Dieu de m'avoir donné la santé et la volonté de mener à bien ce travail de fin d'études.

Je tiens à remercier les membres du jury d'avoir accepté de participer au jury et de juger le travail de mon travail de fin d'étude.

Je tiens particulièrement à remercier mon superviseure, Mme MOUFOK Souad de m'avoir initié à la recherche et de m'avoir confié ce merveilleux sujet et pour son encadrement à mon égard, merci de poursuivre ce travail avec une grande disponibilité, de me faire profiter de son expérience et de me laisser une grande latitude pour mener à bien ce travail dans d'excellentes conditions.

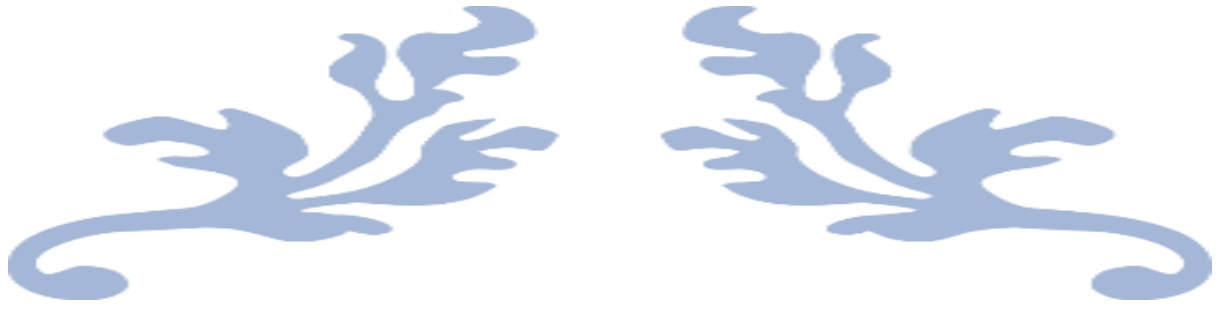
Mon profond remerciement à Monsieur «FEDALA Ismail», qui tous les mots de remerciements ne suffisent pas pour ses efforts consentis à mon égard, et aussi Monsieur «LARBI Djamel» et Monsieur «CHAREF Walid» pour m'avoir facilité la tâche, pendant la période de stage.

Ma profonde gratitude et mes vifs remerciements vont également aux personnels de la direction maintenance service compression/sud de l'entreprise "SONATRACH D" qui m'ont aidé avec leurs précieux conseils durant la période de stage.

Enfin, je remercie tous ceux qui ont contribué de près ou de loin à la réalisation de ce modeste travail.

BOURAS SENIA





Dédicace

Tout d'abord, Je tiens à remercier mon Dieu de m'avoir donné la force et le courage d mener à bien ce modeste travail. Je tien à dédier cette humble travail
à :

Ma tendre mère et mon très cher père ;

Mes précieuses sœurs : Hayat et Soria ;

Mes frères : Mohamed et Yassine ;

Mes meilleurs amis ;

Toutes mes amis d'enfance et du long parcours scolaire et universitaire ;

Toute ma famille

Tous ce qui m'aiment et que j'aime

BOURAS SENIA



Résumé

La similarité entre les noms de médicaments est une problématique très fréquente qui cause des erreurs médicamenteuses. Cette similarité peut conduire à des prises de mauvais médicaments, mauvaise dose, ...etc. Ces médicaments sont habituellement désignés selon leur dénomination anglophone par les termes de look-alike, Pour résoudre ce problème, il faut créer des stratégies permettant de regrouper les noms de médicaments similaires.

L'objectif de cette étude est d'évaluer les erreurs médicamenteuses liées aux noms de médicaments similaires en regroupant les noms de médicaments en fonction de leur similarité orthographique en utilisant deux mesures de similarités (Bi-Sim et Levenshtein). Afin de réduire la survenue d'erreurs de délivrance causées par la confusion entre les médicaments dont les noms se ressemblent visuellement.

Mots clés :

Look-Alike; LASA; erreurs de médicaments; similarité Bi-sim, distance Levenshtein.

الملخص

ُعدّ تشابه أسماء الأدوية مشكلة شائعة تُسبب أخطاءً دوائية. قد يؤدي هذا التشابه إلى تناول دواء خاطئ، أو جرعة خاطئة، وما إلى ذلك. عادةً ما يُشار إلى هذه الأدوية بأسمائها الإنجليزية باستخدام مصطلحاتٍ متشابهة. ولحلّ هذه المشكلة، يجب وضع استراتيجياتٍ لتجميع أسماء الأدوية المتشابهة معًا.

تهدف هذه الدراسة إلى تقييم أخطاء الأدوية المتعلقة بأسماء الأدوية المتشابهة، وذلك بتصنيف أسماء الأدوية بناءً على تشابهها الإملائي وذلك للحد من أخطاء صرف الأدوية الناتجة عن الخلط بين الأدوية ذات (Levenshtein و Bi-Sim) باستخدام مقياسي تشابه الأسماء المتشابهة بصريًا.

الكلمات المفتاحية:

LASA, تشابه bi-sim, مسافة Levenshtein, التشابه, أخطاء الدواء,

Summury

Similarity between drug names is a very common problem that causes medication errors. This similarity can lead to taking the wrong medication, wrong dose, etc. These medications are usually referred to according to their English names by the terms look-alike. To solve this problem, strategies must be created to group similar drug names.

The objective of this study is to assess medication errors related to similar drug names by grouping drug names based on their spelling similarity using two similarity measures (B₁-Sim and Levenshtein). This is to reduce the occurrence of dispensing errors caused by confusion between drugs with visually similar names

Key words:

Look-Alike, LASA, medication errors, Bi-sim similarity, Levenshtein distanc

Plan de travail

Remerciements.....	2
Dédicace.....	3
Résumé.....	4
Abstract	
Table des matières	
Liste des figures	
Liste des tableaux	
Introduction générale.....	11
<u>Chapitre 1 : Etat de l’art de la confusion des médicaments (LOOK ALIKE)</u>	13
1.1 Introduction	14
1.2 Définition et Classification des Médicaments Look-Alike	
1.2.1 Définition des erreurs médicamenteuses.....	14
1.2.2. Tableaux des erreurs (Tableau 01, Tableau 02).....	15
1.3 Définition des médicaments Look-alike Sound-alike (LASA)	16
1.3.1 Différences look-alike sound-alike.....	16
1.3.2 Exemple	18
1.4 Catégorisation des confusion possibles	19
1.5 Causes et facteurs favorisant la confusion.....	20
1.5.1 Facteurs liés à l’emballage.....	21
1.5.2 Facteurs liés à la nomenclature.....	22
1.6 Conséquences de confusion des médicaments.....	23
1.6.1 Risques pour les patients	23
1.6.2 Impacts économiques.....	23
1.6.3 Etudes de cas.....	24
1.7 Réglementation et Normes.....	25
1.7.1 Cadre réglementaire.....	25
1.8 Stratégies et solutions.....	28
1.9 conclusion	29
<u>Chapitre 2 : outils et méthodes</u>	31
2.1 Introduction.....	32
2.2 outils utilisés.....	33
2.3 Similarité textuelle.....	34
2.3.1 Similarité syntaxique.....	34
2.3.2 Similarité sémantique.....	34
2.4 Mesures de similarité.....	34
2.4.1 Matrice Dice (Bi-Sim).....	35
2.4.2 Distance de Levenshtein.....	42
2.5 Classification des matrices de similarité.....	46
2.6 Conclusion.....	50
<u>Chapitre 3 : Conception du travail</u>	51
3.1 Introduction.....	52
3.2 Architecture globale.....	52

3.2.1 FDA liste des médicaments.....	52
3.2.2 FDA liste.....	53
3.2.3 Méthodes de similarité (Bi-sim et Levenshtein).....	53
3.2.4 Fonction de pondération binaire.....	54
3.2.5 Score de silhouette.....	59
3.3 Méthodologie détaillée.....	59
3.3.1 Description des donnée.....	59
3.3.2 Prétraitement.....	59
3.3.3 Codage des données.....	59
3.4 Environnement de travail.....	61
3.4.1 Matlab.....	61
3.4.2 Présentation Matlab.....	61
<u>Chapitre 4 : Implémentation et résultats.....</u>	69
4.1 Introduction.....	70
4.2 Description des données.....	70
4.3 Modules développés.....	71
4.3.1 Module Bi-Sim.....	71
4.3.2 Module Levenshtein	71
4.4 Résultats.....	72
4.4.1 Prétraitement	72
4.4.2 Codage.....	72
4.4.3 Regroupement.....	72
4.4.4 Discussion.....	78
4.5 Conclusion.....	78
Conclusion générale et perspectives.....	79

Liste des figures :

Figure 1.1 Ressemblance entre des comprimés quadri sécables de LISINOPRIL EG et de PREVISCAN (fluidione)	21
Figure 1.2 Ressemblance entre Kétamine et Morphine	21
Figure 1.3 Couleur par classe pharmacologique pour un les gaz médians et anesthésie	
Figure1.4 La similarité visuelle (Sound-Alike) entre ondansetron et metoclopramide.....	22
Figure 1.5.la similarité visuelle (Sound-alike) entre cazaline et ceftriascon	23
Figure 1.6 la simulation suivant les recommandation ANSM.....	25
Figure 1.7 Siège de l'agence Européenne des médicament.....	27
Figure 1.8 Siège de la Food and Drug Administration	30
Figure 1.9 Siège de l'agence nationale de sécurité du médicament	30
Figure 2.1 Algorithme Bi-sim	40
Figure 2.2 Substitution (i remplacer par e)	47
Figure 2.3 Algorithme de distance Levenshtein.....	51
Figure 3.1 Méthodologie	57
Figure 3.2 L'interface générale du Matlab.....	63
Figure 3.3 Exemple de variable dans le Workspace.....	63
Figure 3.4Exemple de fichier.m enregistré dans le curent directory nommé TP MATLAB 2020	64
Figure 3.5 Exécution d'un fichier Matlab	66
Figure 3.6 Exemple d'exécution en mode interactif	67
Figure 3.7 Exemple d'exécution en mode exécutif (fichier.m)	68
Figure 3.8 Liste des noms de médicaments	70
Figure 3.9 Nombre de groupe en fonction des seuils	73

Liste des tableaux

Tableau 01 Erreurs au niveau de prescription	19
Tableau 02 Erreurs au niveau de la retranscription	20
Tableau 2.1 Exemple d'une matrice de similarité (Initialisation)	41
Tableau 2.2 Remplissage de la matrice de similarité, caractère par caractère	42
Tableau 2.3 Matrice de similarité entre deux noms(Doliprane et Doleprane)	43
Tableau 2.4 Initialisation de la matrice de similarité Bi-Sim	44
Tableau 2.5 Remplissage de la matrice de similarité	45
Tableau 2.6 Matrice de calcul de distance Levenshtein entre (Doliprane et Doleprane) ..	47
Tableau 2.7 Matrice de calcul de distance Levenshtein entre (Amoxicilline et Ampicilline).....	48
Tableau 2.8 Calcul de matrice de distance de Levenshtein entre un ensemble de médicament... ..	49
Tableau 2.9 Comparaison entre les deux méthodes Bi-Sim et Levenshtein	50
Tableau 2.10 Classification possible des matrices de similarité selon le type d'algorithme utilisé	51
Tableau 2.11 Classification possible des matrices de similarité selon le type de similarité détectée	52
Tableau 2.12 Classification possible des matrices de similarité selon l'usage en détection de confusion	52
Tableau 2.13 Exemple de score de confusion	53
Tableau 4.1 Seuils de similarité en fonction du nombre de groupes de médicaments	72
Tableau 4.2 Liste des noms de médicaments Seuil 0.2	73
Tableau 4.3 Liste des noms de médicaments Seuil 0.4	74
Tableau 4.4 Liste des noms de médicaments Seuil 0.5	74
Tableau 4.5 Liste des noms de médicaments Seuil 0.6	77

INTRODUCTION Général

De nombreux médicaments peuvent être confondus en raison des similarités de leurs noms ou de leurs présentations. Ces confusions sont fréquentes et conduisent souvent à des incidents graves ou même fatals. Ces médicaments sont habituellement désignés selon leur dénomination anglophone par les termes de look-alike, sound-alike (LASA). Tous les domaines d'activité médicale sont concernés et cette problématique touche aussi bien le secteur hospitalier que le secteur ambulatoire dans le monde entier. Chaque étape du « circuit du médicament » (depuis la prescription du médicament jusqu'à son administration au patient) peut être à l'origine de confusions entre des produits d'apparence ou de noms similaires. Depuis quelques années, la littérature médicale s'est enrichie d'articles traitant du sujet des LASA, mais peu de publications sont disponibles en français. Des démarches ont été proposées pour réduire le risque des LASA et plusieurs organisations ont émis des recommandations en ce sens. Les aspects d'étiquetage et d'apparence des produits mis sur le marché jouent un rôle prépondérant, quand bien même les exigences de fabrication sont rarement corrélées à des dispositions légales. De plus, des considérations économiques l'emportent souvent sur les préoccupations de sécurité. Le domaine de l'infectiologie est lui aussi concerné par la problématique des LASA, par exemple en raison de l'existence de nombreux médicaments administrés par voie parentérale, des dénominations très proches des céphalosporines ou encore du recours occasionnel à des médicaments peu connus de certains utilisateurs.

Les médicaments look-alike, sound-alike représentent un risque important pour les patients et des incidents liés à leur utilisation peuvent avoir des conséquences fatales. Au même titre que la plupart des activités médicales, le domaine de l'infectiologie est étroitement concerné par cette problématique. Les étiologies possibles des confusions de médicaments sont nombreuses. Des recommandations

visent toutefois à aider les professionnels de la santé à réduire ce risque et à les rendre attentifs aux dangers constitués par les médicaments

CHAPITRE 1

ETAT DE L'ART DE CONFUSION DES MEDICAMENTS (Look- Alike).

1.1. Introduction.

La sécurité médicamenteuse constitue une priorité majeure dans les systèmes de santé, tant en milieu hospitalier qu'en officine. Parmi les erreurs les plus fréquentes, celles liées à la confusion entre les médicaments dits "look-alike" (semblables d'apparence) représentent un risque important pour les patients.

Ces confusions peuvent survenir en raison de similitudes dans les noms commerciaux, les emballages, les formes galéniques ou encore l'étiquetage. De telles ressemblances peuvent induire des erreurs de prescription de dispensation ou l'administration, compromettant ainsi la qualité des soins.

Ce chapitre vise à faire un état de l'art sur le phénomène des médicaments look-alike. Il s'agira d'en définir les contours, d'examiner les facteurs contributifs identifiés dans la littérature scientifique, et de recenser les approches actuelles en matière de prévention et de gestion de ce type d'erreur médicamenteuse.

1.2. Définition et Classification des Médicaments Look-Alike.

Un évènement indésirable est une manifestation nocive et non recherchée subie par une personne participant à une thérapie. Un évènement indésirable grave peut contribuer à la survenue d'un décès, être susceptible de mettre en jeu le pronostic vital immédiat du sujet, entraîner une invalidité ou une incapacité ou encore provoquer ou prolonger une hospitalisation. Une erreur médicamenteuse est l'écart par rapport à ce qui aurait dû être fait au cours de la prise en charge thérapeutique médicamenteuse du patient ; il s'agit de l'omission ou de la réalisation non intentionnelle d'un acte relatif à un médicament, qui peut être à l'origine d'un risque ou d'un évènement indésirable pour le patient. Elle est par définition évitable car elle manifeste ce qui aurait dû être fait et qui ne l'a pas été au cours de la prise en charge thérapeutique médicamenteuse d'un patient.

Les causes d'erreur sont diverses, elles peuvent venir du conditionnement primaire (unitaire) ou secondaire (emballage), de la notice d'information (lisibilité (langue,

format), traduction), du produit (étiquetage), ou encore de l'information officielle (Compendium des médicaments).

La plus grande partie des erreurs vient d'un problème de confusion (soundalike, look-alike) ou de perception (utilisation de décimales ou d'abréviations). Il semblerait que 0.13 % des erreurs de prescriptions concernent le mauvais médicament, alors que les rapports volontaires rapportent un taux de 25 %. Au niveau de la prescription, les erreurs peuvent provenir du nom du médicament, de la quantité ou de la fréquence d'administration ; il en va de même au niveau de la retranscription (les taux d'erreurs se trouvent dans les tableaux n°1 et 2). Il peut y avoir des erreurs de spécialité (médicament, dose unitaire ou concentration) et des erreurs d'étiquetage (identité du patient, médicament ou concentration), mais également des erreurs de compte ou de répartition pour les comprimés et des erreurs de volume, de dilution (calcul ou manipulation) ou de contamination biologique pour les solutions. 3 % d'erreurs sont commises sur les comprimés, dont le 74 % sont des erreurs de sélection (20 % d'erreurs de répartition et

6 % de comptage), alors que sur les 6.5 % d'erreurs commises sur les seringues, seulement 28 % sont des erreurs de sélection (52 % pour la dilution et 20 % pour la quantité). Ces erreurs pourraient être éliminées en supprimant l'élément humain du processus. Toutes les mesures qui améliorent la visibilité, la lisibilité et permettent de différencier un objet vont permettre d'améliorer le processus de sélection.

Tableau 01: Erreurs au niveau de la prescription.

Nom du médicament	seule l'abréviation est présente	10.6 %
	nom illisible ou absent	0.5 %
Quantité	ambiguë	41.1 %
	absente	27.0 %
Fréquence	ambiguë	12.3 %
	absente	31.0 %

Tableau 02: Erreurs au niveau de la retranscription.

Nom du médicament	seule l'abréviation présente	est 10.3 %
Quantité	ambiguë	3.3 %
	absente	22.8 %
	ne correspond pas à la prescription	4.3 %
Fréquence	absente	0.6 %

Les confusions ou risque de confusion entre noms de médicaments, les compléments alimentaires ou dispositifs médicaux, ou entre dénominations communes internationales (DCI), font l'objet de signalements auprès de l'ANSM. Ces confusions peuvent s'accompagner

d'effets indésirables, parfois graves. L'ANSM souhaite attirer l'attention en particulier sur le risque de confusion entre Previscan® (fluindione, antithrombotique) et Permixon® (serenoa repens, indiqué dans les troubles mictionnels modérés liés à l'hypertrophie bénigne de la prostate). Plusieurs signalements d'erreur de délivrance entre ces deux spécialités ont été rapportés. Elles sont survenues dans un contexte de délivrance sur la base d'ordonnances manuscrite. Ces erreurs peuvent induire des manifestations hémorragiques graves dont certaines ont eu une issue fatale. [2]

1.3. Définition des médicaments look-alike et sound-alike (LASA).

Les deux problèmes principaux dans la ressemblance des médicaments sont celui du soundalike, qui concerne les médicaments ayant une phonétique similaire, et celui du lookalike, qui se situe à plusieurs niveaux (nom, conditionnement primaire, conditionnement secondaire et étiquette). Pour les noms, il a été proposé de séparer le problème en deux méthodes, une pour les médicaments qui ont des noms ayant une orthographe similaire (lookalike) et une pour ceux qui ont une phonétique similaire (sound-alike). Il est possible de confondre deux noms de médicaments, mais également

de mal comprendre une information qui a été transmise, par exemple en raison du bruit environnant ou parce que le port du masque étouffe les voix. Il ne faudrait jamais agir sans être certain de ce qui a été demandé. [1]



Figure 1.1. Ressemblance entre des comprimés quadri sécables de Lisinopril EG® et de PREVISCAN® (fluindione). [3]

Exemple :

- **Le Préviscan®** : confusion avec Lisinopril EG®
- Attention aux confusions de dosage :



Figure 1.2. Ressemblance entre Ketamine et Morphine.

Ketamine 50 et 250 mg

- Morphine 10 et 20 mg.

- Ne pas se fier qu'aux codes couleur : **Anesthésie** : 1 couleur par classe **Gaz médicaux** : 1 couleur par classe de gaz pharmacologique (ex: antalgiques) mais le corps de la bouteille est toujours blanc.

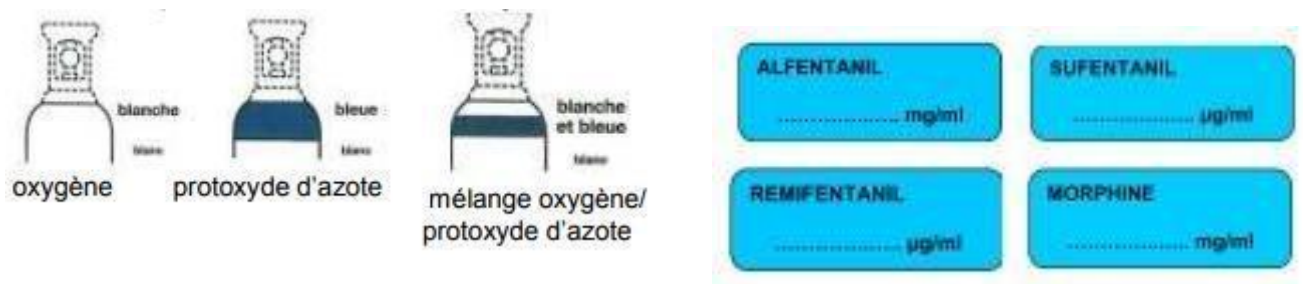


Figure 1.3. Couleur par classe pharmacologique.

1.4. Différences entre look-alike et sound-alike :

La confusion entre médicaments Look-Alike, Sound-Alike (LASA) est une source majeure d'erreurs médicamenteuses, et ces termes désignent deux types distincts de similarités susceptibles d'engendrer des erreurs :

a) Look-Alike (LA) : Similarité visuelle.

Cela concerne des médicaments ayant un aspect similaire, notamment au niveau de l'emballage, de la forme, de la couleur, ou du conditionnement.

Ces similarités peuvent entraîner des confusions lors de la dispensation, du stockage ou de l'administration des médicaments.

Exemples : Loxen® (nicardipine) et Loro xen® (piroxicam) peuvent être confondus en raison d'un packaging visuellement proche.



Figure 1.4. La similarité visuelle (sound-alike) entre Ondansetron et Metoclopramide [6].

b) Sound-Alike (SA) : Similarité phonétique.

Il s'agit de médicaments dont les noms se ressemblent phonétiquement, pouvant conduire à une confusion lors de la prescription orale ou de la transmission d'ordres verbaux.

Le risque est amplifié par des facteurs comme l'accent, la mauvaise articulation, ou les bruits ambiants.

Exemples : Celebrex® (célécoxib) et Cerebyx® (fosphénytoïne), qui ont une prononciation proche et peuvent être mal interprétés.

On peut expliquer cette différence par le suit :

- **Cognition et perception :** Les erreurs LASA sont liées aux biais cognitifs, notamment l'effet de similarité perceptive qui impacte la reconnaissance visuelle et auditive.
- **Facteurs humains :** Fatigue, stress ou surcharge cognitive augmentent le risque de confusion.
-

Conséquences cliniques : Une administration erronée peut entraîner une inefficacité thérapeutique, des effets indésirables graves, voire un danger vital.



Figure 1.5. La similarité visuelle (sound-alike) entre Cefazolin et Ceftriaxon.

1.5. Catégorisation des confusions possibles.

Les confusions entre médicaments, dues à des similitudes dans leurs noms ou apparences, constituent une source fréquente d'erreurs médicamenteuses. Ces erreurs peuvent survenir à différentes étapes du processus de soins, notamment lors de la prescription, de la dispensation ou de l'administration des médicaments.

L'Agence nationale de sécurité du médicament et des produits de santé (ANSM) distingue plusieurs types d'erreurs médicamenteuses :

- **Erreur avérée** : le patient a reçu un médicament incorrect, une dose inappropriée, par une mauvaise voie d'administration ou selon un schéma thérapeutique inadéquat.
- **Erreur interceptée** : l'erreur est détectée et corrigée avant l'administration du médicament au patient.
- **Risque d'erreur** : une situation où un danger potentiel pour le patient est identifié, par exemple en raison de conditionnements ou de dénominations similaires entre médicaments.

1.6. Causes et facteurs favorisant la confusions.

Les causes des erreurs médicamenteuses sont très variées : il s'agit majoritairement de confusions ou de risques de confusion entre des spécialités conditionnées sous forme d'ampoules de solution injectable, mais aussi entre des comprimés. Les autres causes sont principalement des défauts d'information (ex : informations manquantes, mentions erronées), des défauts de conditionnement (ex : inadaptation aux pratiques), d'étiquetage (ex : similitude d'étiquetage entre deux spécialités), et des similitudes de noms commerciaux (ex : risque de confusion par similitude entre deux noms de spécialités : des cas ont été rapportés concernant les spécialités Lamictal® et Lamisil® par exemple) ou de dénominations communes, et quelques cas de défauts de qualité. Près du tiers des cas s'avère être des erreurs de pratique impliquant le professionnel de santé, dont l'essentiel correspond à des erreurs d'inattention, parfois liées à des circonstances particulières (exemple : contexte d'urgence), ou des erreurs d'utilisation par le patient [6].

Suite aux différents signalements rapportant des erreurs de pratique, l'ANSM a identifié plusieurs causes possibles d'erreurs médicamenteuses (organisationnelles, humaines, techniques ou logistiques ...) sous-jacentes souvent associées :

1.6.1. Facteurs liés à l'emballage et à l'étiquetage.

- L'absence de lecture de la notice (en particulier de la posologie ou du mode d'administration) et de l'étiquetage,
- Similarité des emballages primaires ou secondaires (couleurs, forme, format).
- Police d'écriture trop petite ou difficile à lire, surtout pour les noms ou les dosages.
- Emballage générique non distinctif : plusieurs médicaments ont un design standard sans code couleur ou logo clair.
- Étiquetage peu clair ou surchargé : présence excessive d'informations peut masquer les éléments essentiels comme le nom ou le dosage.
- Mauvais positionnement des informations importantes : nom du médicament et dosage peu visibles.

Boîtes actuelles



Simulations suivant
les recommandations
ANSM



Figure 1.6. La simulation suivant les recommandations ANSM.

1.6.2. Facteurs liés à la nomenclature et la phonétique.

a. Similitude orthographique (look-alike)

- Nom commercial ou DCI visuellement proche : ex. Losec vs Lasix.
- Utilisation de préfixes/suffixes similaires : ex. Hydroxyzine vs Hydralazine.
- Mêmes lettres dominantes : quand les premières ou dernières lettres sont identiques ou très proches.
- Longueur et structure similaires : les noms longs avec le même nombre de syllabes ou lettres.

b. Similitude phonétique (sound-alike)

- Prononciation similaire : ex. Celebrex vs Celexa.
- Accent local ou différences linguistiques qui rendent les noms plus proches à l'oral.
- Vitesse ou clarté de la parole : dans un environnement bruyant, certains noms peuvent sembler identiques.

c. Nomenclature internationale ou commerciale

- Variations entre DCI et noms commerciaux qui créent de la confusion.
- Nom de marque différent d'un pays à l'autre pour le même médicament.
- Nom trop « marketing » sans lien logique avec la molécule ou l'indication.

d. Utilisation de chiffres ou lettres similaires

Par exemple : Zyrtec 10 et Zantac 150 peuvent prêter à confusion si mal lus ou mal stockés.

1.6.3. Facteurs humains et erreurs de perception.

- Erreurs en relation avec les logiciels de prescription informatisée
- Méconnaissance des médicaments
- Mauvaise lecture de l'ordonnance sans confirmation auprès du médecin
- Erreurs de substitution par méconnaissance de la liste des génériques

1.6.4. Facteurs liés à l'organisation et à la gestion des stocks.

- Conseil inadapté, inapproprié
- Erreurs de sélection, soit par confusion de nom, soit en raison du mode de rangement des médicaments dans les officines de ville

- Des problèmes de rangement des médicaments (médicament dangereux laissé à portée des enfants, etc.),

L'erreur peut donc trouver sa source :

- Dans une mauvaise conception du médicament et de l'information qui lui est relative (confusion de dénomination, conditionnement inadapté, problème d'étiquetage ou de notice d'information, etc.),
- Dans le dysfonctionnement de l'organisation systémique du processus de prise en charge thérapeutique du patient (organisation du circuit du médicament, facteurs humains, facteurs environnementaux, pratiques professionnelles, etc.).

Selon une étude réalisée par l'ANSM, la majorité des erreurs médicamenteuses est due aux similitudes de conditionnements et d'étiquetages (32,3%) ainsi qu'aux erreurs de pratique (29,6%).

1.7. Conséquences de la Confusion des Médicaments.

L'effet indésirable est défini dans le Code de la Santé Publique et répond à la définition suivante « Réaction nocive et non voulue à un médicament, se produisant aux posologies normalement utilisées chez l'homme pour la prophylaxie, le diagnostic ou le traitement d'une maladie ou pour la restauration, la correction ou la modification d'une fonction physiologique ou résultant d'un mésusage du médicament ou produit ».

L'effet indésirable est considéré comme grave selon le Code de la Santé Publique lorsqu'il répond à la définition suivante : « Effet indésirable létal, ou susceptible de mettre la vie en danger, ou entraînant une invalidité ou une incapacité importante ou durable, ou provoquant ou prolongeant une hospitalisation, ou se manifestant par une anomalie ou une malformation congénitale. »

L'analyse a posteriori de l'erreur permet de la caractériser et de la qualifier par sa nature, son type, la gravité de ses conséquences cliniques pour le patient et par l'étape de réalisation dans la chaîne de soins. L'analyse de l'ensemble des signalements impliquant clairement le médicament (sa présentation, sa dénomination...).

1.7.1. Risques pour la sécurité des patients.

- La confusion entre des médicaments similaires (look-alike, sound-alike) représente un danger majeur pour la sécurité des patients. Les erreurs médicamenteuses peuvent entraîner :
- Des effets indésirables graves, parfois mortels, lorsque le médicament administré n'est pas le bon.
- Un retard dans le traitement, surtout si le médicament nécessaire est remplacé par un autre non adapté.
- Des complications médicales, notamment chez les patients vulnérables (enfants, personnes âgées, patients polymédiqués).
- Une perte de confiance du patient envers le personnel soignant et le système de santé en général.

1.7.2. Impacts économiques et juridiques.

- Les erreurs dues à la confusion entre médicaments ont également des conséquences économiques et juridiques importantes :
- Coûts supplémentaires pour les établissements de santé (traitement des complications, prolongation de l'hospitalisation, réinterventions).
- Litiges et indemnisation : les erreurs peuvent conduire à des poursuites judiciaires, avec des compensations financières à verser aux patients ou à leurs familles.
- Perte de réputation pour les fabricants de médicaments et les établissements de santé.
- Rappel de lots de médicaments ou modification de l'emballage, entraînant des frais pour les laboratoires pharmaceutiques.

1.7.3. Études de cas et exemples concrets.

Plusieurs cas réels illustrent la gravité de ce problème :

- En France, une confusion entre Prozac (antidépresseur) et Prograf (immunosuppresseur) a provoqué des effets secondaires sévères chez un patient transplanté.
- À l'international, le cas bien documenté de la confusion entre Lamisil (antifongique) et

Lamictal (antiépileptique) a entraîné des crises d'épilepsie chez des patients ayant reçu le mauvais médicament.

- Dans certains hôpitaux, des erreurs entre insuline rapide (Actrapid) et insuline lente (Lantus) ont causé des hypoglycémies sévères.

1.8. Réglementation et Normes en Matière de Sécurisation des Médicaments.

La lutte contre les erreurs médicamenteuses liées à la confusion des médicaments lookalike (médicaments ayant un aspect similaire) repose sur un cadre réglementaire et normatif rigoureux, aussi bien au niveau national qu'international. Ce cadre vise à encadrer la fabrication, la distribution, l'étiquetage et la présentation des médicaments, afin de limiter les risques de confusion.

1.8.1. Réglementations nationales et internationales.

Au niveau international, de nombreuses agences de régulation ont mis en place des exigences spécifiques pour réduire les erreurs liées aux look-alike. Par exemple :

- a) **Union Européenne (EMA)** : L'agence européenne des médicaments impose des règles strictes concernant l'étiquetage, la couleur, la forme et le conditionnement des médicaments. Les autorités nationales, comme l'ANSM en France, veillent à l'application de ces règles.



Figure 1.7. Siège de l'Agence Européenne des Médicaments.

- b) **États-Unis (FDA)** : La Food and Drug Administration évalue les risques de confusion dès la phase de demande de mise sur le marché. Elle peut refuser un nom de médicament ou exiger des modifications visuelles si le produit est jugé trop semblable à un autre.



Figure 1.8. Siège de la Food and Drug Administration.

- c) **France (ANSM) :** L'Agence nationale de sécurité du médicament impose des contrôles sur les éléments visuels des médicaments (nom, emballage, forme galénique), et peut demander des tests de différenciation.



Figure 1.9 Siège de l'Agence Nationale de Sécurité du Médicament

1.9. Bonnes pratiques de fabrication et de distribution.

Les bonnes pratiques de fabrication (BPF) et de distribution (BPD) incluent des exigences pour éviter les erreurs dues aux médicaments look-alike :

- **BPF :** Les fabricants doivent intégrer des contrôles qualité sur les aspects visuels et l'identification claire des produits. L'utilisation de codes couleur, de formes distinctes, ou d'emballages spécifiques est encouragée.
- **BPD :** Les distributeurs sont tenus de stocker et de livrer les médicaments de manière à éviter les mélanges ou les erreurs lors de la préparation des commandes. L'étiquetage clair et le respect de la chaîne de traçabilité sont essentiels.

1.10. Recommandations des organismes de santé (OMS, FDA, ANSM, etc.).

Les grandes institutions de santé publient régulièrement des recommandations :

- **OMS** : Encourage la mise en place de systèmes de signalement des erreurs médicamenteuses et la standardisation de l'emballage et de l'étiquetage.
- **FDA** : Propose des lignes directrices pour la conception des étiquettes, l'utilisation de codes-barres, et l'évaluation des noms commerciaux via des tests de similitude phonétique et orthographique.
- **ISMP (Institute for Safe Medication Practices)** : Fournit des listes de médicaments look-alike et recommande l'utilisation de Tall Man lettering (ex. DOPamine vs DOBUTamine) pour améliorer la distinction visuelle.
- **ANSM** : Propose des recommandations sur le conditionnement et mène des campagnes de sensibilisation sur les risques liés aux confusions.

1.11. Stratégies et Solutions pour Réduire la Confusion.

Pour faire face aux risques liés aux médicaments look-alike, plusieurs stratégies ont été développées dans les domaines du design, de la technologie, de la formation, et de la vérification. Ces approches visent à renforcer la sécurité tout au long du circuit du médicament.

1.11.1. Amélioration du design des emballages et étiquetages.

L'un des moyens les plus efficaces pour limiter les confusions est l'optimisation de l'apparence visuelle des médicaments :

- **Utilisation du Tall Man Lettering** : Cette technique consiste à mettre en majuscules une partie des noms de médicaments pour les différencier (ex. hydrOXYzine vs hydrALAZINE).
- **Codes couleur** : L'utilisation de couleurs spécifiques pour certaines classes thérapeutiques permet de distinguer les produits visuellement.
- **Typographie claire et lisible** : Éviter les polices similaires ou trop petites, ainsi que les étiquettes surchargées d'informations.
- **Différenciation des formes et conditionnements** : Des formes galéniques variées (capsules, comprimés, ampoules) et des emballages bien distincts réduisent les risques de confusion.

1.11.2. Rôle des technologies (codes-barres, RFID, IA).

La technologie joue un rôle croissant dans la sécurisation des circuits de médicaments :

- **Codes-barres et Datamatrix** : Permettent de vérifier automatiquement l'identité du médicament au moment de la dispensation ou de l'administration.
- **Technologie RFID** : Offre une traçabilité en temps réel des médicaments, utile notamment en milieu hospitalier.
- **Systèmes d'aide à la décision clinique (CDSS) et intelligence artificielle** : Aident à détecter les erreurs potentielles lors de la prescription ou de la préparation.

1.11.3. Formation et sensibilisation des professionnels de santé.

La vigilance des professionnels reste un élément clé :

- **Sessions de formation continue** : Ciblées sur les risques liés aux look-alike et les bonnes pratiques à adopter.
- **Guides internes et alertes** : Fournis par les pharmacies hospitalières ou les autorités sanitaires.
- **Culture de sécurité** : Encourager le signalement des erreurs et quasi-accidents sans sanction favorise l'apprentissage organisationnel.

1.11.4. Procédures de vérification et double contrôle.

Mettre en place des processus de validation supplémentaires permet de limiter les erreurs humaines :

- **Double vérification** : Surtout lors de la préparation ou de l'administration de médicaments à risque élevé.
- **Check-lists et protocoles standardisés** : Pour renforcer la rigueur dans les étapes sensibles.
- **Implication du patient** : Dans certains contextes (ambulatoire, domicile), la vérification active par le patient ou un aidant peut contribuer à éviter des erreurs.

1.12. Conclusion.

La confusion entre médicaments look-alike représente un enjeu majeur de sécurité dans les systèmes de santé du monde entier. À travers cet état de l'art, il apparaît clairement que ces erreurs peuvent survenir à toutes les étapes du circuit du médicament, de la prescription à l'administration, et qu'elles sont souvent liées à des similarités visuelles ou phonétiques entre produits.

Les études, rapports d'incidents et retours d'expérience soulignent l'importance d'une approche globale, intégrant à la fois la réglementation, l'amélioration du design, l'innovation technologique et la formation des professionnels. Les recommandations d'organismes tels que l'OMS, la FDA ou l'ANSM montrent également une volonté commune de standardiser les pratiques et de renforcer la prévention.

Cette revue met en lumière la complexité du problème, mais aussi les nombreuses solutions existantes ou en développement. Elle constitue une base solide pour orienter les stratégies d'amélioration continue en matière de sécurisation des médicaments look-alike, en particulier dans le cadre hospitalier et industriel.

CHAPITRE 2.

MATERIELS ET METHODES.

2.1. Introduction

Lorsque l'on entend parler de calcul de similarité en informatique, c'est souvent dans le cadre de données textuelles : comment classer, regrouper des documents ? Comment, pour une requête, retourner la liste des documents les plus pertinents (dans le cas d'un moteur de recherche par exemple) ? La notion de similarité sémantique (distance sémantique) est utilisée pour exprimer la ressemblance entre des concepts. En recherche d'information conceptuelle, les mesures de similarité jouent un rôle important, en particulier dans le processus de désambiguïsation des termes, la pondération des concepts et l'évaluation de la pertinence.

L'objectif des mesures de similarité sémantique est d'estimer la ressemblance entre les documents.

Ce chapitre vise à présenter les approches méthodologiques utilisées pour identifier les risques de confusion entre médicaments dits look-alike, c'est-à-dire ceux dont le nom ou l'apparence est visuellement similaire. L'objectif principal est d'expliquer les fondements et la pertinence des méthodes de similarité dans le contexte de la détection préventive des erreurs médicamenteuses.

La ressemblance entre les noms commerciaux ou l'apparence des conditionnements peut engendrer des erreurs lors de la prescription, de la dispensation ou de l'administration des médicaments. Pour faire face à ce risque, l'utilisation d'outils de comparaison basés sur la similarité (lexicale, visuelle) permet une évaluation systématique et objective de la proximité entre deux médicaments. Ces méthodes sont ainsi essentielles pour anticiper les confusions potentielles, renforcer la sécurité des patients, et orienter les décisions réglementaires ou les actions correctives en pharmacie hospitalière ou industrielle.

2.2. Outils utilisés pour détecter la confusion des médicaments

L'analyse des risques de confusion entre médicaments repose sur la collecte et l'exploitation de bases de données fiables contenant des informations sur les noms commerciaux, les principes actifs, et parfois les caractéristiques visuelles des médicaments. Plusieurs sources ont été mobilisées dans ce travail :

- **Base hospitalière :** Données internes issues du logiciel de gestion pharmaceutique de l'hôpital, incluant les spécialités utilisées en routine, leurs noms commerciaux, dosages, formes galéniques et conditionnements.

- **ANSM (Agence Nationale de Sécurité du Médicament et des produits de santé) :** Accès à des listes de spécialités pharmaceutiques autorisées sur le marché français, utilisées comme référence réglementaire.
- **Vidal :** Base de données professionnelle largement utilisée en milieu hospitalier et officinal, permettant de croiser les noms commerciaux avec les données thérapeutiques et galéniques.

Au total, environ (le nombre des médicaments utilisés dans cette étude) ont été analysés. Les données ont été extraites et consolidées sous format Excel pour un traitement initial, puis organisées dans des tableaux SQL pour faciliter les requêtes, les croisements et l'automatisation des calculs de similarité.

Cet outil constitue le socle des analyses réalisées dans les sections suivantes, permettant l'application des méthodes de détection des similarités entre noms de médicaments à des fins de prévention des risques.

2.3. Similarité textuelle

Évaluer la similarité entre documents textuels est une des problématiques importantes de plusieurs disciplines comme l'analyse de données textuelles, la recherche d'information ou l'extraction de connaissances à partir de données textuelles (Text Mining). Dans chacun de ces domaines, les similarités sont utilisées pour différents traitements :

- En analyse de données textuelles, les similarités sont utilisées pour la description et l'exploration de données ;
- En recherche d'information, l'évaluation des similarités entre documents et requêtes est utilisée pour identifier les documents pertinents par rapport à des besoins d'information exprimés par les utilisateurs ;
- En Text Mining, les similarités sont utilisées pour produire des représentations synthétiques de vastes collections de documents.

Les techniques mises en œuvre pour calculer les similarités varient bien évidemment selon les disciplines, mais elles s'intègrent cependant le plus souvent dans une même approche générale en deux temps :

- a. Les documents textuels sont d'abord associés à des représentations spécifiques qui vont servir de base au calcul des similarités. Bien que la nature précise des représentations utilisées dépende fortement du domaine d'application, il faut noter que, presque dans tous

- b. les cas, les documents sont représentés sous la forme d'éléments d'un espace vectoriel de grande dimension.
- c. Un modèle mathématique est choisi pour mesurer les similarités.
 - **Similarité syntaxique** : En mathématiques et en informatique, une mesure permettant de comparer des documents textuels, consiste à comparer des chaînes de caractères. C'est une métrique qui mesure la similarité ou la dissimilarité entre deux chaînes de caractères. Par exemple, les chaînes de caractères "Sam" et "Samuel" peuvent être considérées comme similaires. Une telle mesure sur les chaînes de caractères fournit une valeur obtenue algorithmiquement. Parmi de telles mesures de similarité, citons par exemple, la distance de Levenshtein (ou distance d'édition), le coefficient de Dice, l'indice de Jaccard, la distance euclidienne, le cosinus, ...
 - **Similarité sémantique** : La similarité sémantique est un concept selon lequel un ensemble de documents ou de termes se voient attribuer une métrique basée sur la ressemblance de leur signification / contenu sémantique. Concrètement, cela peut être réalisé en définissant une similitude topologique, par exemple, en utilisant des ontologies pour définir une distance entre les mots, ou en définissant une similitude statistique, par exemple en utilisant un modèle d'espace vectoriel pour corréliser les termes et les contextes à partir d'un corpus de texte approprié (co-occurrence). Parmi de telles mesures de similarité, citons par exemple, Resnik, LSA (Analyse sémantique latente), ESA (Analyse sémantique explicite), ...

2.4. Différentes mesures de similarité

Nous calculons la similarité entre deux médicaments $S_{ab}=S_{ba}$ et chaque nom de médicament est comparé aux autres noms, mais pas à lui-même. Ainsi, la matrice obtenue est une matrice carrée de $n \times n$ où n est le nombre de médicaments.

2.4.1. Matrice de similarité Dice (Bi-Sim)

2.4.1.1. Définition

L'algorithme de similarité Bi-Sim est un algorithme de similarité appliqué aux noms de médicaments en comparant leurs caractères. Il suit une approche similaire aux algorithmes d'édition de distance comme Levenshtein, mais en utilisant une fonction de similarité spécifique $Nsim(i,j)$ entre caractères. [3]

2.4.1.2. Principe de la méthode Bi-Sim Similarity

L'algorithme de similarité Bi-Sim [7] est utilisé pour calculer une matrice de similarité entre deux noms de médicaments en comparant leurs caractères un par un.

Il fonctionne de manière similaire aux algorithmes de distance d'édition (comme Levenshtein) mais utilise une fonction de similarité personnalisée appelée $Nsim(i,j)$.

L'objectif est de construire une matrice M où chaque élément $M(i,j)$ indique la similarité entre le i -ème caractère du premier mot et le j -ème caractère du second mot.

2.4.1.3. Application à la Similarité des Médicaments

L'algorithme donné dans la figure 2.1., peut être utilisé pour :

- Détecter des noms de médicaments similaires (éviter les erreurs de prescription).
- Regrouper des médicaments en fonction de leur nom pour analyser des confusions potentielles.
- Créer une base de données où les erreurs de saisie sont corrigées automatiquement.
- Il aide à détecter les erreurs de saisie.
- Il permet de regrouper des médicaments similaires.

2.4.1.4. Algorithme Bi-Sim

Entrée : Deux chaînes X (taille n) et Y (taille m)

Sortie : Matrice M et Score de similarité de S

Début

$n \leftarrow \text{Longueur}(X)$

$m \leftarrow \text{Longueur}(Y)$

//Etape 1 : Initialisation de la matrice

Crée une matrice M de taille $(n+1)(m+1)$

Pour i de 0 à n faire

$M[i,0] \leftarrow 0$

Fin pour

Pour j de 0 à m faire

$M[0,j] \leftarrow 0$

Fin pour

//Etape 2 Remplissage de la matrice

Pour i de 1 à n faire

```

    Pour j de 1 à m faire
        Si  $X[i,1] = Y[j,1]$ 
             $M[i,j] \leftarrow 1$ 
        Sinon
             $M[i,j] \leftarrow 0$ 
        Fin si
    Fin pour
Fin pour
//Etape 3 calcule du score de similarité
Somme  $\leftarrow 0$ 
Pour i de 1 à n faire
    Pour j de 1 à m faire
        Somme  $\leftarrow$  somme +  $M[i,j]$ 
    Fin pour
Fin pour
 $S \leftarrow$  Somme /  $ma(n,m)$ 
Retourner M ,S
Fin

```

Figure 2.1. Algorithme Bi-Sim.

a. Explication de l'algorithme de similarité Bi-Sim.

L'objectif est de construire une matrice de similarité M qui représente la similarité entre deux noms de médicaments caractère par caractère.

- Entrées :

- **D** : Ensemble de noms de médicaments $D=\{T1,T2,...,TL\}$.
- Chaque nom de médicament T_i est une séquence de caractères $T_i=\{U1,U2,...,UR\}$.

- Sortie :

- Une matrice de similarité M où $M(i,j)$ représente la similarité entre le i -ème caractère de x et le j -ème caractère de y .

b. Étapes détaillées.

1. Créer une matrice M de dimensions $(n+1) \times (m+1)$, où :

- n est la longueur du premier mot x .
- m est la longueur du second mot y .
- On ajoute une ligne et une colonne pour tenir compte des positions vides.

2. Initialiser la première colonne et la première ligne de la matrice à zéro pour gérer l'absence de caractères.
3. Remplir la matrice M en comparant chaque caractère de x avec chaque caractère de y à l'aide de $Nsim(i,j)$.
4. Lire la matrice M pour analyser la similarité globale entre les deux mots.

c. **Exemples d'application.**

Exemple 1:

Prenons deux noms de médicaments à comparer :

- $x = \text{"Doliprane"}$
- $y = \text{"Doleprane"}$

Nous allons utiliser une fonction $Nsim(i,j)$ simple qui donne :

- $Nsim(a,b)=1$ si les caractères a et b sont identiques.
- $Nsim(a,b)=0$ sinon.

➤ **Étape 1 : Création de la matrice M, et Initialisation de la première ligne et colonne à zéro.**

Les mots ont :

- $n=9$ caractères ("Doliprane").
- $m=9$ caractères ("Doleprane").

On crée une matrice M de (10×10) (avec une ligne et colonne supplémentaires pour gérer les positions vides).

Tableau 2.1. Exemple de d'une matrice de similarité (Initialisation).

	0	D	O	L	I	P	R	A	N	E
0	0	0	0	0	0	0	0	0	0	0
D	0									
O	0									
L	0									
E	0									
P	0									
R	0									
A	0									

N	0									
E	0									

➤ **Étape 2 : Calcul des similarités caractère par caractère (remplissage de la matrice de similarité).**

On remplit la matrice $M(i,j)$ avec $Nsim(i,j)$:

- $M(1,1) = Nsim(D, D) = 1$
- $M(2,2) = Nsim(o, o) = 1$
- $M(3,3) = Nsim(l, l) = 1$
- $M(4,4) = Nsim(i, i) = 1$ □ $M(5,5) = Nsim(p, p) = 1$
- $M(6,6) = Nsim(r, r) = 1$
- $M(7,7) = Nsim(a, a) = 1$
- $M(8,8) = Nsim(n, n) = 1$
- $M(9,9) = Nsim(e, e) = 1$

Cela donne les résultats données dans le tableau 2.2.

Tableau 2.2. Remplissage de la matrice de similarité, caractère par caractère.

	0	D	O	L	I	P	R	A	N	E
0	0	0	0	0	0	0	0	0	0	0
D	0	1	0	0	0	0	0	0	0	0
O	0	0	1	0	0	0	0	0	0	0
L	0	0	0	1	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	1	0	0	0	0
R	0	0	0	0	0	0	1	0	0	0
A	0	0	0	0	0	0	0	1	0	0
N	0	0	0	0	0	0	0	0	1	0
E	0	0	0	0	0	0	0	0	0	1

L'algorithme de similarité Bi-Sim permet de calculer une matrice de similarité entre les noms de médicaments caractère par caractère.

Le résultat final de l'algorithme Bi-Sim est la somme des valeurs dans la matrice M, qui représente la similarité globale entre les deux noms de médicaments. Récapitulatif du calcul pour "Doliprane" et "Doleprane".

La matrice obtenue est :

Tableau 2.3. Matrice de similarité entre deux noms (Doliprane et Doleprane).

	o	D	o	l	i	p	r	a	n	e
o	0	0	0	0	0	0	0	0	0	0
D	0	1	0	0	0	0	0	0	0	0
o	0	0	1	0	0	0	0	0	0	0
l	0	0	0	1	0	0	0	0	0	0
i	0	0	0	0	0	0	0	0	0	0
p	0	0	0	0	0	1	0	0	0	0
r	0	0	0	0	0	0	1	0	0	0
a	0	0	0	0	0	0	0	1	0	0
n	0	0	0	0	0	0	0	0	1	0
e	0	0	0	0	0	0	0	0	0	1

➤ Étape 3 : Calcul de la Similarité Totale.

On somme tous les "1" dans la matrice :

$$1+1+1+1+1+1+1+1=8$$

Donc, la similarité totale entre "Doliprane" et "Doleprane" est 8.

❖ Indice de Similarité Normalisé.

Si on veut un score de similarité entre 0 et 1, on peut normaliser selon la formule suivante :

$$Similarité = \frac{\sum M(i,j)}{\max(n,m)} \quad (1)$$

Ici, n=9 et m=9, donc :

$$Similarité = \frac{8}{9} = 0.8889$$

Interprétation du Résultat :

- Une valeur de **0.8889** signifie une forte similarité entre les deux noms.
- La seule différence est le caractère "i" dans "Doliprane" remplacé par "e" dans "Doleprane".

- **Exemple 2 :**

Si les deux noms de médicaments n'ont pas la même longueur, l'algorithme de similarité BiSim fonctionne toujours, mais il faut ajuster la taille de la matrice et traiter les différences de longueur.

Cas d'exemple : "Doliprane" (9 caractères) et "Dolipran" (8 caractères) :

- **Étape 1 : Création de la matrice M, et Initialisation de la première ligne et colonne à zéro.**

On crée une matrice M de taille 10×9 (ajout d'une ligne et d'une colonne supplémentaires pour les décalages).

Tableau 2.4. Initialisation de la matrice de similarité Bi-Sim

	0	D	O	L	I	P	R	A	N
0	0	0	0	0	0	0	0	0	0
D	0								
O	0								
L	0								
I	0								
P	0								
R	0								
A	0								
N	0								
E	0								

- **Étape 2 : Calcul des similarités caractère par caractère (remplissage de la matrice de similarité).**

On compare les caractères entre eux selon la suite suivante :

$$\begin{cases} \text{si } x[i] = y[j], & \text{alors } M(i, j) = 1 \\ \text{sinon,} & M(i, j) = 0 \end{cases} \quad (2)$$

Tableau 2.5. Remplissage de la matrice de similarité

	0	D	O	L	I	P	R	A	N
0	0	0	0	0	0	0	0	0	0
D	0	1	0	0	0	0	0	0	0
O	0	0	1	0	0	0	0	0	0
L	0	0	0	1	0	0	0	0	0
I	0	0	0	0	1	0	0	0	0
P	0	0	0	0	0	1	0	0	0
R	0	0	0	0	0	0	1	0	0
A	0	0	0	0	0	0	0	1	0
N	0	0	0	0	0	0	0	0	1
E	0	0	0	0	0	0	0	0	0

➤ **Étape 3 : Calcul de la Similarité Totale** On somme tous les 1 :

$$1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 = 8$$

Normalisation par la plus grande longueur (9) :

$$\text{Similarité} = \frac{8}{9} = 0.8889$$

Normalisation par la plus grande longueur (9) :

Interprétation du Résultat

- La valeur est proche de 1, donc ces deux noms sont très similaires.
- L'algorithme gère bien les noms de longueurs différentes.
- S'il y a plus de différences, la valeur de similarité diminue.

2.4.2. Distance de Levenshtein.

2.4.2.1. Définition.

La distance de Levenshtein a été proposée pour la première fois par Levenshtein et al. [4]. Elle est destinée à comparer la similarité entre deux chaînes de caractères en tenant compte de l'ordre des caractères. En utilisant deux chaînes comme paramètres d'entrée d_1 et d_2 , l'algorithme de Levenshtein calcule le nombre minimum d'opérations nécessaires pour transformer une chaîne de caractères en une autre. Les opérations autorisées sont :

- **Insertion** d'un caractère
- **Suppression** d'un caractère
- **Substitution** d'un caractère par un autre.

Le pseudocode de l'algorithme de la distance de Levenshtein est présenté dans la figure 2.3.

Pour obtenir la distance de Levenshtein $\text{Sim}_{\text{Levenshtein}}(d_1, d_2)$ entre les documents d_1 et d_2 , il s'agit d'associer à chacune de ces opérations un coût. Le coût des opérations est toujours égal à 1, sauf dans le cas d'une substitution de caractères identiques. Notons que cette distance a été étendue pour prendre en compte la grammaire, la phonétique, ...

2.4.2.2. Exemple simple.

Mot 1 (horizontal) : Doliprane

Mot 2 (vertical) : Doleprane

On construit une matrice où, chaque cellule (i, j) représente le coût minimum pour transformer les i premiers caractères de "Doliprane" en les j premiers caractères de "Doleprane".

Initialisation de la matrice :

On remplit la première ligne et la première colonne :

- 1ère ligne : 0, 1, 2, 3, ..., n
- 1ère colonne : 0, 1, 2, 3, ..., m

Ici les deux mots ont 9 lettres.

Tableau 2.6. Matrice de calcul de distance Levenshtein entre (Doliprane et Doleprane).

	0	D	O	L	I	P	R	A	N	E
	0	1	2	3	4	5	6	7	8	9
D	1	0	1	2	3	4	5	6	7	8
O	2	1	0	1	2	3	4	5	6	7
L	3	2	1	0	1	2	3	4	5	6
I	4	3	2	1	1	2	3	4	5	6
P	5	4	3	2	2	1	2	3	4	5
R	6	5	4	3	3	2	1	2	3	4
A	7	6	5	4	4	3	2	1	2	3
N	8	7	6	5	5	4	3	2	1	2
E	9	8	7	6	6	5	4	3	2	1

Explication :

- Jusqu'à la lettre "i" dans Doliprane et "e" dans Doleprane, il faut remplacer "i" par "e", ce qui coûte 1 opération.
- Tout le reste est identique entre les deux mots.
- Ainsi, distance finale = 1.

En résumé :

- Une seule substitution ("i" → "e").
- Distance Levenshtein = 1.



Figure 2.2. Substitution (i remplacer par e)

2.4.2.3. Exemple complexe :

Médicament 1 : *Amoxicilline*

Médicament 2 : *Ampicilline*

On veut calculer la distance de Levenshtein entre ces deux mots :

Tableau 2.7. Matrice de calcul de distance Levenshtein entre (*Amoxicilline* et *Ampicilline*).

		A	M	O	X	I	C	I	L	L	I	N	E
	0	1	2	3	4	5	6	7	8	9	10	11	12
A	1	0	1	2	3	4	5	6	7	8	9	10	11
M	2	1	0	1	2	3	4	5	6	7	8	9	10
P	3	2	1	1	2	3	4	5	6	7	8	9	10
I	4	3	2	2	2	2	3	4	5	6	7	8	9
c	5	4	3	3	3	3	2	3	4	5	6	7	8
I	6	5	4	4	4	4	3	2	3	4	5	6	7
L	7	6	5	5	5	5	4	3	2	3	4	5	6
L	8	7	6	6	6	6	5	4	3	2	3	4	5
I	9	8	7	7	7	7	6	5	4	3	2	3	4
N	10	9	8	8	8	8	7	6	5	4	3	2	3
E	11	10	9	9	9	9	8	7	6	5	4	3	2

La distance de Levenshtein entre "Amoxicilline" et "Ampicilline" est 2. Cela signifie qu'il faudrait 2 opérations (une suppression et une substitution) pour transformer l'un en l'autre :

- Remplacer "o" par "p"
- Supprimer "x"

2.4.2.4. Exemple sur un ensemble de médicaments

Supposons l'ensemble suivant de médicaments :

- Dafalgan

- Doliprane
- Doleprane
- Advil

On construit une matrice de distance Levenshtein où chaque case (i ,j) représente la distance entre les médicaments i et j.

Tableau 2.8. Calcule de matrice de distance de Levenshtein entre un ensemble de médicaments.

	Dafalgan	Doliprane	Doleprane	Advil
Dafalgan	0	7	7	6
Doliprane	7	0	1	7
Doleprane	7	1	0	7
Advil	6	7	7	0

Explications :

- Distance (Doliprane, Doleprane) = 1 → (seulement une substitution : "i" → "e")
- Distance (Dafalgan, Doliprane) = 7 → (plusieurs substitutions nécessaires)
- Distance (Dafalgan, Advil) = 6 → (noms très différents).

2.4.2.5. Comparaison entre les résultats de bi-sim et Algorithme Levenshtein

Les deux tableaux montrent que "Doliprane" et "Doleprane" sont très similaires : La matrice de similarité indique 8/9 caractères identiques. La distance de Levenshtein = 1, ce qui signifie qu'il ne faut qu'un seul changement pour passer de l'un à l'autre.

Ces deux méthodes sont complémentaires :
 La première est plus stricte (utile pour détection de plagiat, etc.)
 La seconde est plus souple (utile pour recherche approximative, correcteurs orthographiques, etc.)
 On peut résumer les différences dans ce tableau :

	Matrice de similarité	Distance de levenshtein
Méthode	Comparaison strict caractère par caractère	Alignement global avec tolérance aux erreurs
Résultat obtenu	8 caractères identiques	1 opération de transformation
Sensibilité aux petites erreurs	Faille	Elevée

2.4.2.5. Algorithme Levenshtein

L'algorithme de calcul de la distance Levenshtein présentée dans la figure 2.3. Est présentée dans la figure suivante, où S1 et S2 représentent deux chaînes de caractères, m et n représentent la taille des deux chaînes de caractères S1 et S2 respectivement, d représente la distance [8].

```

Fonction Levenshtein(s1,
s2) m ← longueur(s1) n
← longueur(s2) créer
matrice d[0..m, 0..n]

```

```

    Pour i de 0 à m
    d[i][0] ← i    pour j
    de 0 à n      d[0][j]
    ← j

```

```

        Pour i de 1 à m      pour j de 1 à n      si
        s1[i-1] = s2[j-1] alors      coût ← 0
        sinon      coût ← 1      d[i][j] ←
        min(      d[i-1][j] + 1, //
        suppression      d[i][j-1] + 1, //

```

```

insertion          d[i-1][j-1] + coût //
substitution

Retourner d[m][n]

```

Figure 2.3. Algorithme de distance Levenshtein.

2.5. Classification des matrices de similarité.

Pour détecter la confusion des médicaments look-alike (look-alike drus ou LASA), on peut utiliser des matrices de similarité basées sur des algorithmes comme Levenshtein et DICE pour mesurer la ressemblance entre les noms des médicaments. Voici une classification possible de ces matrices de similarité selon leurs caractéristiques et usages :

2.5.1. Selon le type d'algorithme utilisé

Tableau 2.10. Classification possible des matrices de similarité selon le type D'algorithme utilisé.

Type	Exemple	Description
Edition de chaînes	Levenshtein	Mesure le nombre minimal d'opérations (insertion, suppression, substitution) nécessaires pour transformer un mot en un autre.
Phonétique	DICE, Soundex, Metaphone	Analyse la similarité des sons pour détecter des confusions orales.
Basé sur la structure	N-gram, Qgram	Compare des sous-chaînes communes pour évaluer la similarité.

Hybride	Combinaison Levenshtein + BICE	Combine des aspects phonétiques et orthographiques pour une meilleure détection des confusions.
----------------	--------------------------------------	---

2.5.2. Selon le type de similarité détectée

Tableau 2.11. Classification possible des matrices de similarité selon le type de similarité détectée.

Type de similarité	Algorithmes adaptés	Exemple
Orthographique	Levenshtein, N-gram	<u>Celebrex</u> vs <u>Celexa</u>
Phonétique	DICE, Soundex	<u>Hydroxyzine</u> vs <u>Hydralazine</u>
Contextuelle / sémantique	Word embeddings (avancé)	Pas directement détectable par Levenshtein ou BICE, mais utile dans NLP

2.5.3. Selon l'usage en détection de confusion

Tableau 2.12. Classification possible des matrices de similarité selon l'usage en détection de confusion.

Usage	Algorithmes	Remarques

Pré-filtrage de grandes bases	N-gram, Levenshtein	Permet de réduire les candidats avant des analyses plus poussées.
Analyse fine de paires	Levenshtein, BICE	Pour comparer des noms spécifiques.
Classement par risque de confusion	Score combiné (pondéré)	Par exemple, 70% Levenshtein + 30% BICE.

Tableau 2.13. Exemple de score de confusion

Médicament A	Médicament B	Score Levenshtein	Score DICE	Score combiné	Risque
Celebrex	Celexa	0.77	0.85	0.81	Élevé
Zantac	Xanax	0.66	0.88	0.76	Moyen

2.6. Conclusion

Dans ce chapitre, nous avons présenté les outils méthodologiques adoptés pour identifier les risques de confusion entre les médicaments look-alike. L'approche repose principalement sur l'utilisation de la matrice BICE, permettant une évaluation structurée de la similarité orthographique entre les noms de médicaments selon plusieurs critères, ainsi que sur le calcul de la distance de Levenshtein qui quantifie les différences entre deux chaînes de caractères.

L'association de ces deux méthodes permet de croiser une évaluation qualitative (basée sur les types de confusion potentielle) et quantitative (mesure numérique de la similarité), offrant ainsi une détection plus robuste et objective des paires de noms à risque. Cette méthodologie constitue une base solide pour l'analyse des médicaments look-alike et la mise en place de mesures préventives adaptées dans le cadre de la sécurisation du circuit du médicament.

CHAPITRE 3 :

CONCEPTION DU TRAVAIL

3.1. Introduction

Après avoir effectué un regroupement des médicaments présentant des similitudes visuelles (look-alike), il devient essentiel de structurer les données recueillies de manière exploitable en vue d'une analyse approfondie. Cette étape marque le début de la phase de conception du travail, qui consiste à coder les données de manière systématique afin de les analyser, de modéliser les risques de confusion, et de proposer des solutions adaptées.

Dans cette optique, le logiciel Matlab a été choisi pour son efficacité en matière de traitement numérique, de visualisation graphique et d'analyse statistique. Le codage des données permet de transformer les caractéristiques qualitatives des médicaments (telles que le nom, la forme, la couleur, l'emballage) en variables quantitatives pouvant être traitées par Matlab. Ce processus facilite l'identification de schémas, la classification des produits à risque, et la mise en place d'outils d'aide à la décision pour prévenir les erreurs médicamenteuses.

Ce chapitre décrit donc en détail la méthode de codage utilisée, la structure des données, les choix techniques effectués pour leur traitement, ainsi que les différentes étapes d'analyse réalisées à l'aide de Matlab. L'objectif est de fournir une base solide pour l'interprétation des résultats et la formulation de recommandations concrètes pour améliorer la sécurité médicamenteuse.

3.2.1. Architecture globale :

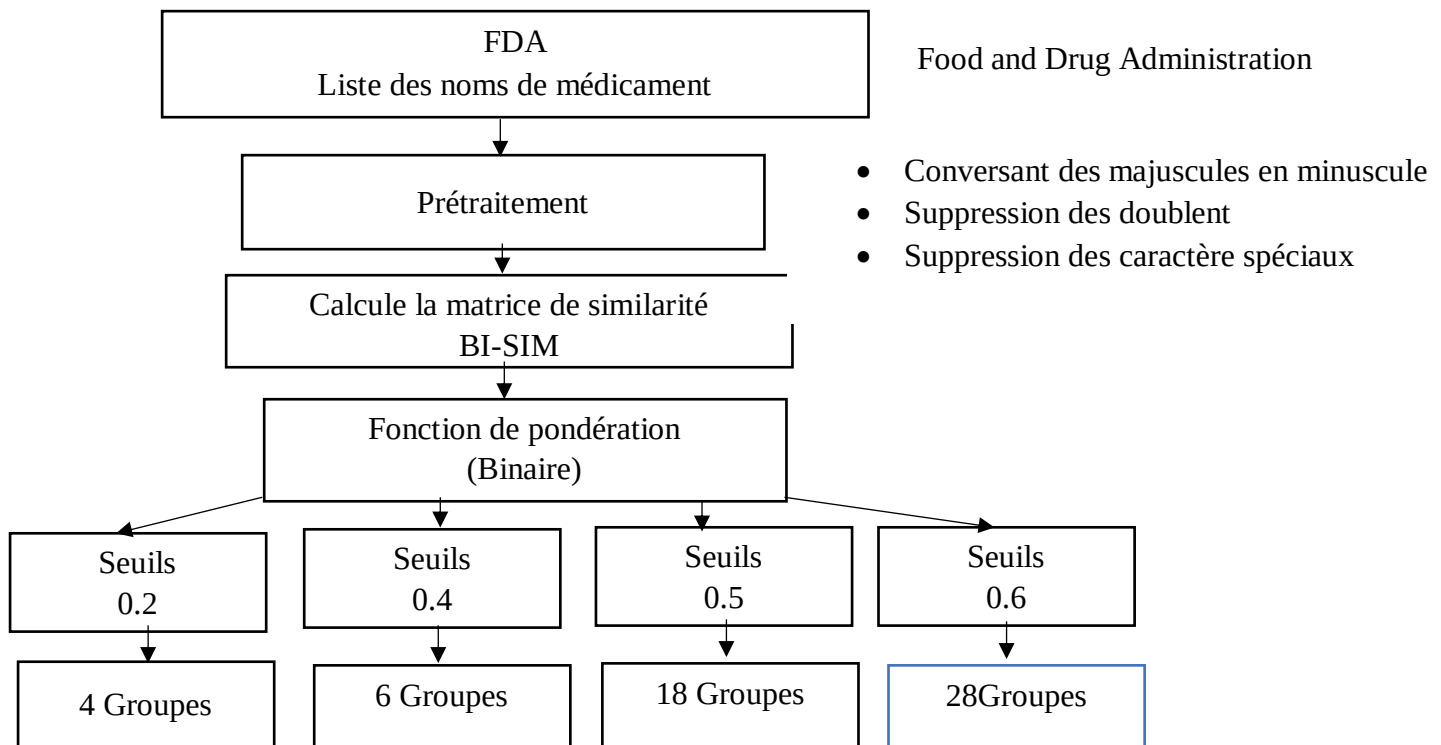


Figure3.1. Méthodologie

Notre système classe les médicaments d'une manière spécifique, en prenant les noms de médicaments comme entrées, une base initiale de médicaments et de leurs propriétés et en les transformant en une base organisée qui peuvent être utilisées dans les phases suivante :

3.2.1.1. FDA – Liste des noms de médicaments

- **Définition** : La FDA (Food and Drug Administration) fournit une base de données contenant les noms des médicaments autorisés aux États-Unis.
- **Objectif** : Cette liste constitue la base de travail pour détecter les similarités entre noms de médicaments et identifier les risques de confusion.

La première étape consiste à récupérer la base de données officielle des noms de médicaments publiée par la FDA (Food and Drug Administration).

Cette base constitue le jeu de données de référence sur lequel les méthodes de comparaison seront appliquées.

3.2.1.2. FDA.Liste

- **Définition :** C'est l'extraction ou la récupération structurée de cette liste depuis la source FDA.
- **Objectif :** Préparer la liste pour un traitement automatique (comparaison de similarité entre les noms).

Il s'agit ici d'une étape de prétraitement et de structuration de la liste de médicaments obtenue, enfin de la rendre exploitable par des algorithmes de calcul de similarité.

3.2.1.3. Méthodes de calcul de similarité (Bi-Sim et Levenshtein)

Deux approches différentes sont utilisées pour comparer les noms de médicaments :

► Bi-Sim (Bi grams Similarity) :

Cette méthode divise chaque nom de médicament en paires de lettres successives (bigrams).

Elle mesure la proportion de bigrammes communs entre deux noms. Elle est utile pour détecter des similarités visuelles ou structurelles.

Distance de Levenshtein :

- Il s'agit d'un algorithme d'édition qui calcule le nombre minimal d'opérations (insertions, suppressions, substitutions) nécessaires pour transformer un mot en un autre.
- Il est pertinent pour identifier des confusions phonétiques ou orthographiques mineures.

3.2.1.4. Fonction de pondération binaire

- **Définition :** Cette fonction attribue une **valeur binaire (0 ou 1)** selon qu'un seuil de similarité est atteint ou non.
- **Objectif :** Filtrer les paires de médicaments potentiellement confondus :
- 1 → similarité significative détectée
- 0 → pas de similarité

Pour chaque méthode, on applique une fonction de pondération binaire qui décide si deux noms sont considérés comme similaires ou non :

- Si la similarité dépasse un certain seuil, on assigne la valeur 1.

- Sinon, la valeur 0 est attribuée.

Cette simplification permet un traitement plus clair pour l'analyse suivante.

3.2.1.5. Calcul de performance – Score de silhouette

- **Définition** : Le score de silhouette est une métrique d'évaluation de la qualité de regroupement (clustering). Il mesure si un élément est bien attribué à son groupe.
- **Application ici** : Permet d'évaluer la performance des méthodes de similarité (Bi-Sim et Levenshtein) dans la détection des noms similaires.
- **Interprétation** :
 - Score proche de 1 → bon regroupement
 - Score proche de 0 → confusion entre les groupes
 - Score négatif → mauvais regroupement

Après avoir identifié les paires similaires, une analyse de la qualité des regroupements est réalisée à l'aide du score de silhouette:

- Ce score évalue si les noms similaires sont bien regroupés entre eux.
- Un score faible ou négatif suggère que des noms différents sont mal classés ensemble.

3.3. Méthodologie détaillée

Ci-dessous, nous détaillons chacune des étapes que nous utilisons dans notre système.

3.3.1- Description des données

3.3.2- Prétraitement

L'objectif du prétraitement est de transformer les données de format brut au format éparé, pour pouvoir utiliser les données dans les phases suivantes, à savoir, le calcul de la similarité et le regroupement.

Dans notre travail, le prétraitement utilisé comporte juste un nettoyage des données, le nettoyage consiste à supprimer tous les doublant (les noms de médicaments qui se répète plusieurs fois).

3.3.3- Codage des données

Processus qui consiste en général à assigner des valeurs numériques aux réponses exprimées au départ sous forme numérique, textuel, sonore ou visuelle. Toutefois, les logiciels d'analyse qualitative utilisent aussi des codes sous forme de texte pour faire des analyses qualitatives.

Le principal objectif est de faciliter le traitement automatique des données dans un but analytique.

Dans ce travail, nous avons besoin de deux méthodes pour le traitement et le codage des données, qui se présentent comme suit :

3.3.3.1- Calcul de similarité Bi-Sim

Le calcul de similarité est une tâche essentielle du traitement automatique de données, textuelles ou non, et a toujours reçu une attention particulière de la communauté scientifique.

3.3.3.2- Distance de Levenshtein

La distance de Levenshtein mesure la similarité entre deux chaînes de caractères. Elle est égale au nombre minimal de caractères qu'il faut supprimer, insérer ou remplacer pour passer d'une chaîne à l'autre.

Son nom provient de Vladimir Levenshtein qui l'a définie en 1965 [6]. Elle est aussi connue sous le nom de « distance d'édition » ou encore de « déformation dynamique temporelle », notamment en reconnaissance vocale.

Cette distance est d'autant plus grande que le nombre de différences entre les deux chaînes est grand. La distance de Levenshtein peut être considérée comme une généralisation de la Hamming. On peut montrer en particulier que la distance de Hamming est un majorant de la distance de Levenshtein.

Le coût minimal pour aller de M à P en utilisant la distance de Levenshtein consiste à effectuer les opérations élémentaires suivantes :

- Substitution d'un caractère de M en un caractère de P.
- Ajout dans M d'un caractère de P.
- Suppression d'un caractère de M.

On associe ainsi à chacune de ces opérations un coût. Par exemple, dans les exemples suivants, le coût est toujours égal à 1, sauf dans le cas d'une substitution de caractères identiques.

3.3.4. Regroupement

Le regroupement de données la FDA, en particulier dans le domaine pharmaceutique, concerne l'organisation et la représentation des données aux fins de soumission et d'évaluation réglementaire. Cela inclut l'utilisation de formats de données standardisés comme le SEND, ainsi que la compilation de données provenant d'études cliniques et d'autres sources pour soutenir les demandes d'autorisation de nouveaux médicaments ou dispositifs médicaux.

But du regroupement de données :

- **Évaluer la sécurité :** En regroupant les données sur les effets secondaires, les réactions indésirables et les événements adverses, la FDA peut identifier les risques potentiels liés à l'utilisation d'un médicament ou d'un dispositif médical.
- **Évaluer l'efficacité :** Le regroupement des données cliniques, telles que les résultats des essais cliniques et les données de suivi après commercialisation, permet d'évaluer l'efficacité d'un produit.

3.4. Environnement de travail

3.4.1. Langage MATLAB

MATLAB est présent dans des systèmes automobiles de sécurité active, des véhicules spatiaux, des appareils de surveillance médicale, des réseaux électriques intelligents et des réseaux mobiles LTE. Il est utilisé dans les domaines de l'apprentissage automatique, le traitement du signal, la vision par ordinateur, les communications, la finance computationnelle, la conception de contrôleurs, la robotique et bien plus.

3.4.2. Présentation MATLAB

La plate-forme MATLAB est optimisée pour résoudre les problèmes scientifiques et techniques. Le langage MATLAB, basé sur les matrices, est le moyen le plus naturel au monde pour exprimer les mathématiques computationnelles. Les graphiques intégrés permettent de visualiser facilement les données afin d'en dégager des informations. Grâce à la vaste

bibliothèque de boîtes à outils prédéfinies, vous pouvez commencer directement par les algorithmes essentiels à votre domaine. L'environnement bureau encourage l'expérimentation, l'exploration et la découverte. Les outils et les fonctionnalités MATLAB sont tous testés rigoureusement. Ils sont conçus pour fonctionner conjointement.

Selon la version utilisée, l'interface peut changer légèrement mais les points centraux resteront identiques. La figure 3.1 illustre l'interface générale du Matlab.

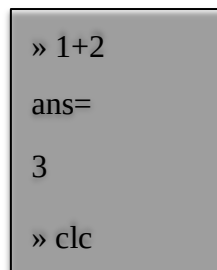
Par défaut, on retrouve quatre fenêtres :

1. Command Window
2. Command History
3. Current Directory
4. Workspace.

3.4.3. Description des fenêtres.

3.4.3.1. Command Window

L'une des plus importantes fenêtres de Matlab, la Command Window traite des instructions données. C'est après l'invite "prompt" qu'il faut entrer les instructions demandées. Les résultats s'afficheront dès le retour de ligne. Le ligne commence toujours par ». Essayez la commande suivante :



```
» 1+2
ans=
3
» clc
```

La commande `clc` permet de vider l'écran de Command Window pour répartir sur une fenêtre propre.

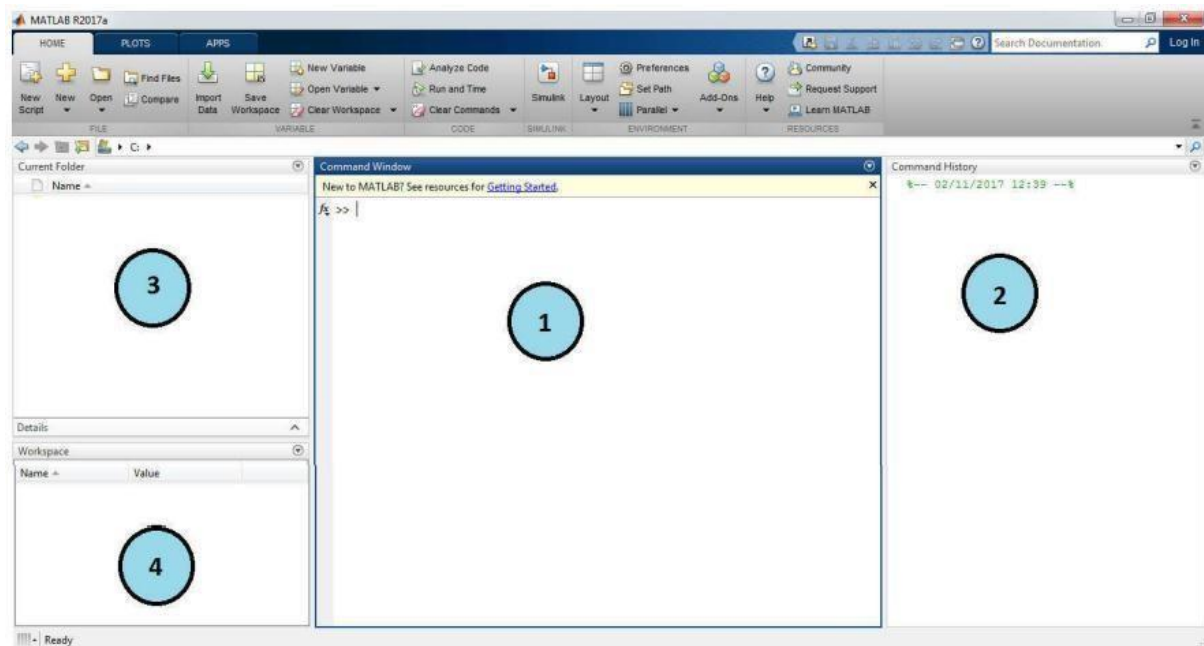


Figure 3.2. L'interface générale du Matalab.

3.4.3.2. Workspace

La fenêtre nommée Workspace permet de visualiser les variables mises en mémoire. On y retrouve leur nom, leur dimension ainsi que le type de variable. Matlab étant basé sur les matrices, toutes les variables sont constituées de plusieurs dimensions :

- Un scalaire est une matrice 1 x 1.
- Un vecteur est une matrice 1 x n, n x 1.
- Une matrice de taille n x m.

Il est possible de double-cliquer sur une variable pour l'afficher. Un clic-droit sur les variables offre de nombreuses options telles que: Copiez, Collez, Supprimez etc.

Il est possible d'effacer certaines variables ainsi que de les éditer. Pour toutes les effacer, utilisez la commande Clear Workspace dans le menu Edit.

La figure 3.3 illustre un exemple de variables dans le Workspace.

Workspace		
Name	Value	Size
A	5	1x1
B	[1 2 3 4]	1x4
C	[1 2 3; 4 5 6; 7 8 9]	3x3

Scalaire
 Vecteur
 Matrice

Figure 3.3. Exemple de variable dans le Workspace.

3.4.3.3. Current Directory

Le Current Directory est le répertoire courant où sont enregistrés les fichiers-M. Il est fortement conseillé de se créer un répertoire autre que celui fournit par Matlab afin de mieux gérer les fichiers contenus dedans (Voir Figure 3.3).

- ✧ Pour compiler un fichier-M, il doit être enregistré dans le répertoire courant.
- ✧ Si un fichier-M appelle une fonction autre qu'une fonction de Matlab (i.e. une fonction créée par l'utilisateur), le fichier-M appelant la fonction ainsi que le fichier-M définissant la fonction doivent être dans le même répertoire courant.

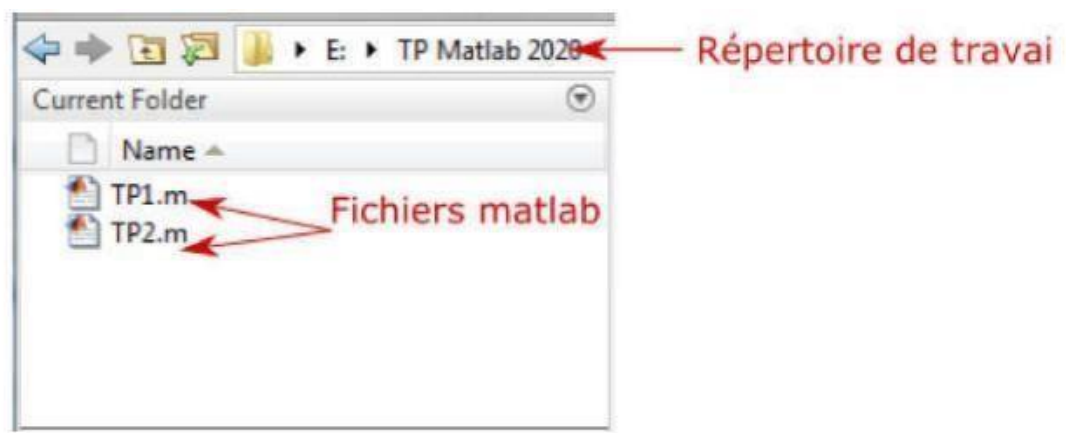


Figure 3.4. Exemple de fichier.m enregistré dans le Current Directory nommé TP Matlab 2020.

3.4.3.4. Command History

Lorsque l'on effectue une analyse de données sur les résultats d'une expérience il est essentiel de conserver une trace de toutes les opérations qui ont été réalisées. C'est la base de la traçabilité et de la reproductibilité des résultats scientifiques.

Il est possible de cliquer sur une commande dans cette fenêtre pour l'exécuter à nouveau. On peut également remonter dans la liste de commandes en se plaçant dans la Command Window et en pressant les flèches de direction.

Matlab est également capable de remonter dans la liste de commandes en ne prenant en compte que les commandes qui commencent par certains caractères. Si l'on recherche une commande que l'on a entrées précédemment, on peut taper le début de cette commande puis la flèche du haut et Matlab cherchera uniquement parmi les commandes ayant le même début.

Par ailleurs, il est parfois important d'enregistrer dans un fichier indépendant la liste des commandes utilisées. Pour ce faire on utilise :

» `diary filename`

Cette commande va créer un fichier `filename` qui enregistrera toutes les commandes entrées jusqu'à ce que l'on utilise `diary off`.

Editor/Debugger

Pour traiter des fichiers-M, il faut utiliser l'éditeur/débogueur de Matlab. Cette fenêtre ne fait pas partie de l'interface de base de Matlab et s'ouvre lorsqu'on ouvre un fichier-M ou lorsqu'on crée un nouveau fichier-M.

Sur la barre des boutons, un bouton est essentiel : le bouton RUN compile le programme, i.e. exécute les commandes du programme. Il peut aussi être exécuté avec F5 (Voir Figure 3.5).

Un autre élément intéressant de l'éditeur consiste à laisser le curseur sur une variable, et sa valeur apparaît (si le programme a été compilé au moins une fois).

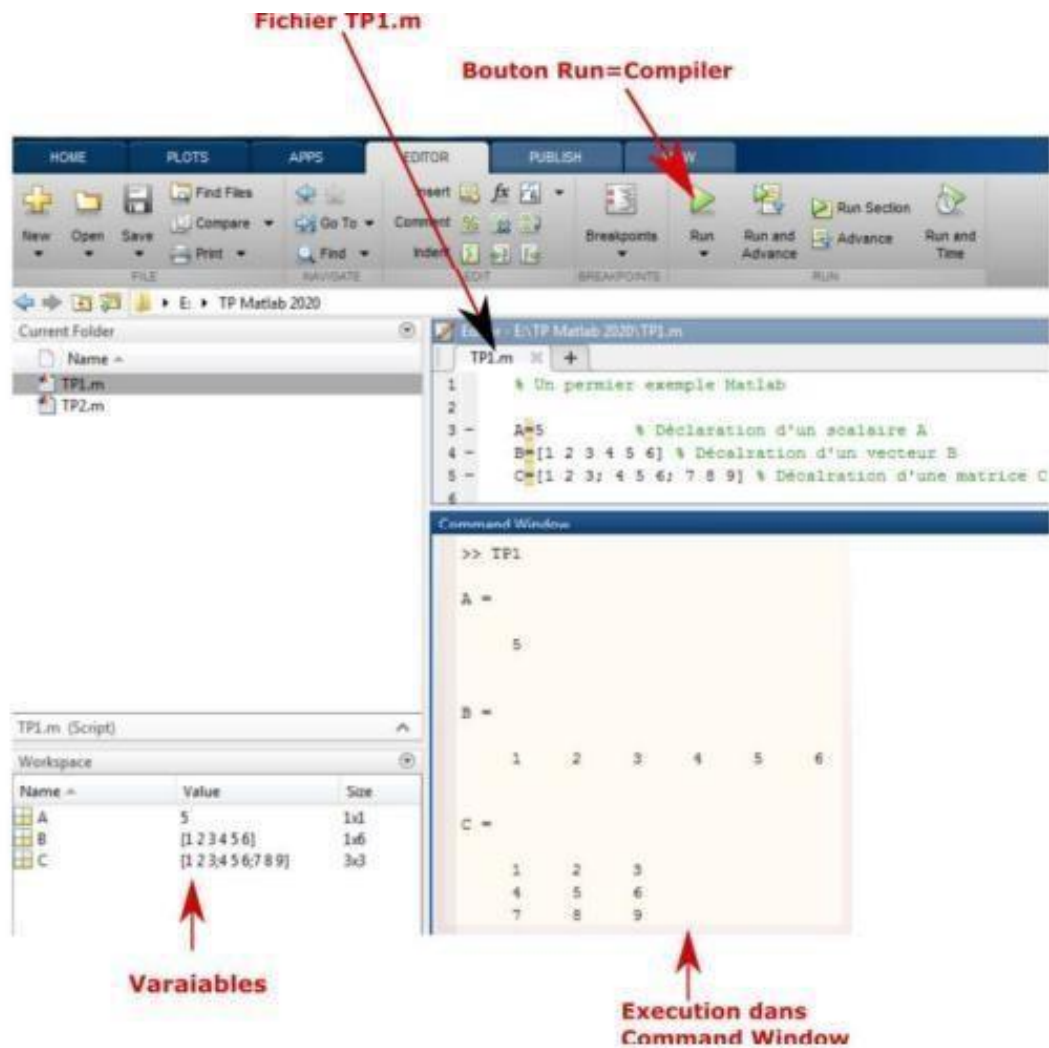


Figure 3.5. Exécution d'un fichier Matlab.

Première interaction avec MATLAB

Il existe deux modes de fonctionnement Matlab :

Mode interactif: Matlab exécute les instructions au fur et à mesure qu'elles sont données par l'utilisateur.

Le moyen le plus simple d'utiliser MATLAB est d'écrire directement dans la fenêtre de commande (Command Window) juste après le curseur (prompt) `>>`.

Pour calculer une expression mathématique il suffit de l'écrire comme ceci (voir figure 1.5):

- Le ";" et le "...": Le point-virgule à la fin d'une ligne signale à MATLAB de ne pas retourner le résultat de l'opération à l'écran.

» `A= 1+3`

» A=4

» B=2*3;

»

Dans l'éditeur/débogueur tout comme dans la fenêtre de commande, l'utilisation de «<...>» est utile pour continuer sur la ligne d'en dessous l'instruction en cours.

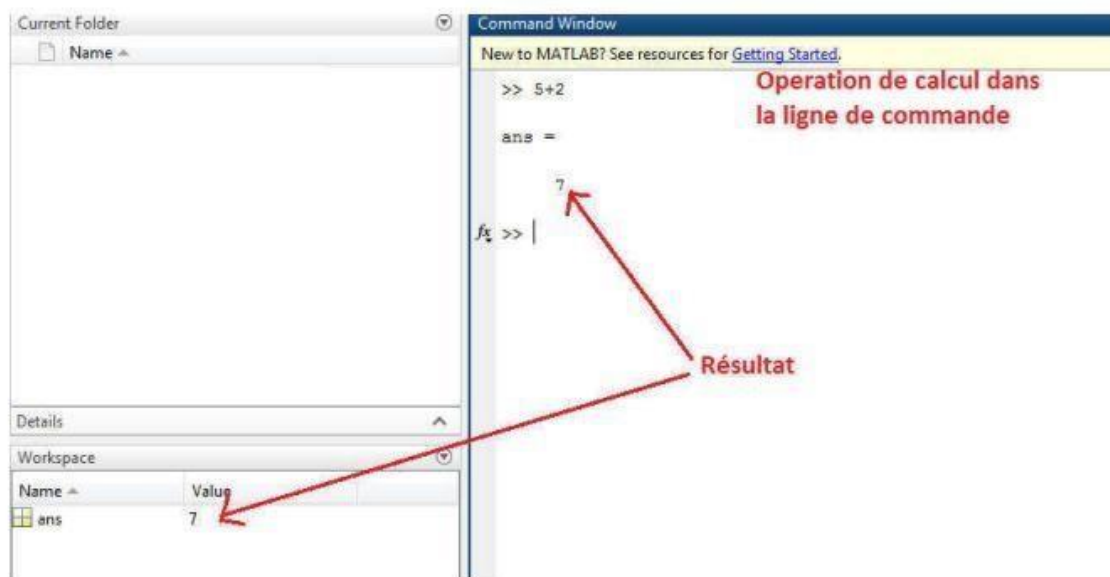


Figure 3.6. Exemple d'exécution en mode interactif.

↳ **Mode exécutif** : Matlab exécute ligne par ligne un programme Matlab (d'extension .m). Un programme Matlab (ou 'm-file' en anglais) est une suite d'instructions Matlab écrites dans un éditeur de texte et sauveées dans un fichier avec une extension .m .



Figure 3.7. Exemple d'exécution en mode exécutif (fichier.m).

Le tracé de courbes dans le plan : plot

La fonction plot permet de placer des points de coordonnées (x; y) sur un graphique. Nous illustrons diverses possibilités à travers quelques exemples.

Exemple:

```

x=[1 2 3 4]; % définit le vecteur ligne de coordonnées (1,2,3,4) et
le nomme x
y=[5 1 2 1]; % définit le vecteur ligne de coordonnées (5,1,2,1) et
le nomme y
plot(x,y) % trace y en fonction de x et joint les points (x,y) par
des segments de droites.

```

Chapitre 4 :

Implémentation et Résultats

4.1 Introduction

Ce chapitre présente l'implémentation et les résultats obtenus dans le cadre du regroupement des noms de médicaments similaires issus de la liste FDA. Deux approches principales ont été utilisées : la mesure de similarité Bi-SIM et la distance de Levenshtein. L'objectif est d'identifier automatiquement des groupes de médicaments dont les noms sont proches orthographiquement, réduisant ainsi les risques d'erreurs médicamenteuses dues à des confusions entre noms ressemblants.

Les résultats sont présentés sous forme de groupes (clusters), accompagnés de visualisations graphiques permettant d'analyser les similarités détectées.

4.2 Description des données

4.2.1 Objectif du travail

L'objectif principal est de développer un système MATLAB capable de :

- Calculer la similarité entre les noms de médicaments à l'aide de deux méthodes : BiSIM et Levenshtein.
- Regrouper automatiquement les médicaments similaires en fonction d'un seuil de similarité prédéfini.
- Visualiser les résultats sous forme de graphes, matrices de similarité.

4.2.2 Description des entrées

Les données d'entrée consistent en une liste de **39 médicaments FDA**, caractérisés par :

- Des variations orthographiques complexes (majuscules/minuscules, suffixes/préfixes communs).
- Des longueurs variables (de 6 à 16 caractères).
- Des similarités partielles (ex: *HYDROmorphone* et *HYDROcodone*).

La figure représente un échantillon des médicaments extraites de la liste FDA, qu'on va traiter en utilisant les deux mesures de similarités.

1	sulfaSALazine	14	sulfISOXAZOLE	27	NovoLIN
2	DAPTOmycin	15	eriBULin	28	CeleXA
3	hydrALAZINE	16	NexAVAR	29	epiRUBicin
4	romiDEPsin	17	hydrOXYzine	30	levOCARNitine
5	HYDROmorphone	18	oxyCODONE	31	CeleBREX
6	levoFLOXacin	19	hydroCHLOROthiazide	32	romiPLOstim
7	NexIUM	20	idaruCIZumab	33	HYDROcodone
8	penicillAMINE	21	SUFentanil	34	SORafenib
9	fentaNYL	22	valACYclovir	35	DACTINomycin
10	predniSONE	23	acetoHEXAMIDE	36	methazolAMIDE
11	NIFEdipine	24	HumuLIN	37	acetaZOLAMIDE
12	oxyMORphone	25	chlordiazepOXIDE	38	methyIPREDNISolone
13	valGANciclovir	26	penicillin	39	HumaLOG

Figure.3.8 Liste des noms de médicaments

4.2.3 Description des sorties

Les sorties générées comprennent :

1. **Groupes de médicaments similaires** (clusters).
2. **Visualisations graphiques** :
 - Matrice binaire de similarité.
 - Diagramme en barres des tailles de groupes. □ Graphe des connexions entre médicaments.
3. **Export des résultats** (fichiers texte et Excel).

4.3 Modules développés (code MATLAB)

4.3.1 Module `regrouper_medicaments_bisim`

Ce module utilise la **similarité Bi-SIM** pour regrouper les médicaments.

Fonctionnement :

1. **Calcul de la similarité** entre chaque paire de noms.
2. **Seuillage** pour obtenir une matrice binaire (1 = similaire, 0 = non similaire).
3. **Regroupement** via un parcours en largeur (BFS) pour identifier les composantes connexes.

4.3.2 Module `regrouper_medicaments_levenshtein`

Ce module repose sur la distance de Levenshtein, normalisée et convertie en similarité.

Fonctionnement :

1. **Calcul de la distance** entre chaque paire de noms.
2. **Normalisation** par la longueur maximale des deux noms.
3. **Conversion en similarité** ($\text{similarity} = 1 - \text{distance_normalisée}$).
4. **Seuillage et regroupement** comme dans le module Bi-SIM.

4.4 Résultats

4.4.1 Prétraitement

Avant de calculer les matrices de similarité, vue la nature des données bruitées, il est essentiel d'appliquer un prétraitement qui consiste en une :

- Conversion des noms en minuscules pour une comparaison homogène.
- Suppression des caractères spéciaux si nécessaire.

4.4.2 Codage des données

4.4.2.1 Similarité Bi-SIM

Cette similarité consiste à :

- Appliquer une mesure basée sur la comparaison de caractères adjacents.
- Appliquer un Score normalisé entre 0 et 1.

4.4.2.2 Distance Levenshtein

Cette similarité consiste à :

- Mesurer le nombre minimum d'opérations (insertion, suppression, substitution) pour passer d'un nom à l'autre.
- Appliquer une normalisation pour obtenir un score de similarité.

4.4.3 Regroupement

4.4.3.1. Mesure de similarité Bi-SIM

Pour l'évaluation de la mesure de similarité Bi-Sim, et le regroupement des noms de médicaments similaires, on a appliqué des seuils varies entre 0.3 et 0.8. Pour chaque seuil on a trouvé un groupe de médicaments jugés similaires. Chaque groupe contient une liste des noms de médicaments similaires.

Le tableau suivant représente l'ensemble des seuils avec le nombre de groupe de médicaments.

Tableau 4.1 Seuils de similarité en fonction du nombre de groupe de médicaments.

Seuil	0,3	0,4	0,5	0,6	0,7	0,8
Groupe	7	16	22	29	33	39

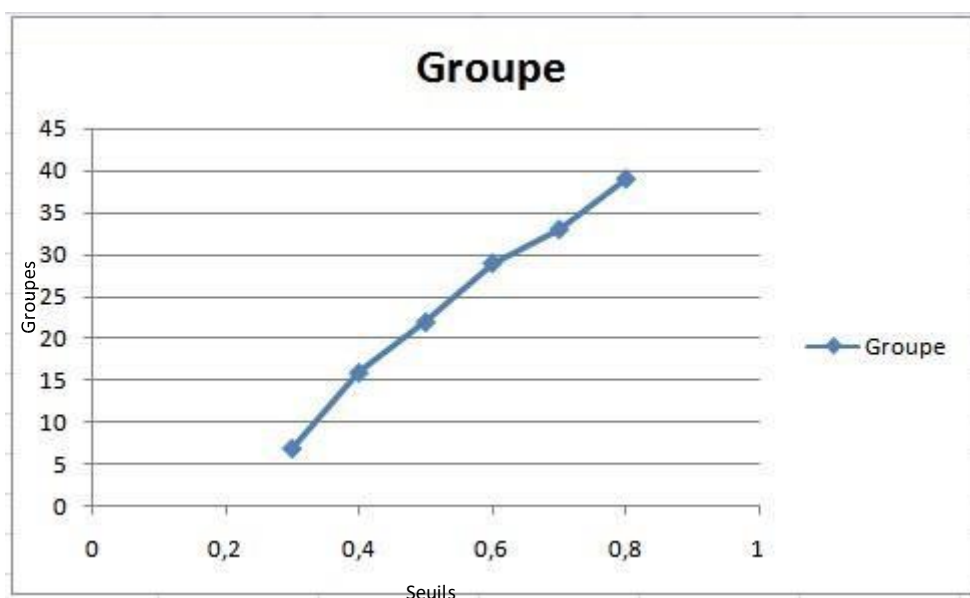


Figure3.9 Nombre de groupe en fonction des seuils

Tableau 4.2 Liste des noms de médicaments Seuil 0.2

Groupe 1 (36 médicaments)	sulfaSALAZINE, hydrALAZINE, levoFLOXacin, penicillAMINE, NIFFedipine, valGANciclovir, sulfiSOXAZOLE, eriBULin, hydrOXYzine, hydroCHLOROthiazide, SUFentanil, valACYclovir, acetoHEXAMIDE, HumuLIN, chlordiazePOXIDE, penicillin, NovoLIN, levOCARNitine, SORafenib, methazolAMIDE, acetaZOLAMIDE, HumaLOG, HYDROmorphone, epiRUBicin, HYDROcodone, methylPREDNISolone, NexAVAR, CeleXA, romiPLOstim, DACTINomycin, fentaNYL, predniSONE, oxyCODONE, idaruCIZUmab, CeleBREX, oxyMORphone
Groupe 2 (1 médicament)	DAPTOmycin
Groupe 3 (1 médicament)	romiDEPsin
Groupe 4 (1 médicament)	NexIUM

Tableau 4.3 Liste des noms de médicaments Seuil0.4

Groupe 1 (27 médicaments)	sulfaSALAzine, hydrALAZINE, penicillAMINE, sulfiSOXAZOLE, HYDROmorphone, hydrOXYzine, hydroCHLOROthiazide, HYDROcodone, NIFFedipine, penicillin, levOCARNitine, methazolAMIDE, acetaZOLAMIDE, oxyMORphone, DAPTOmycin, oxyCODONE, chlordiazePOXIDE, predniSONE, eriBULin, epiRUBicin, levoFLOXacin, acetoHEXAMIDE, DACTINomycin, methylPREDNISolone, HumuLIN, NovoLIN, HumaLOG
Groupe 2 (2 médicaments)	romiDEPsin, romiPLOStim
Groupe 3 (4 médicaments)	NexIUM, NexAVAR, CeleXA, CeleBREX
Groupe 4 (3 médicaments)	fentaNYL, SUFentanil, SORAfenib
Groupe 5 (2 médicaments)	valGANciclovir, valACYclovir
Groupe 6 (1 médicament)	idaruCIZUmab

Tableau 4.4 Liste des noms de médicaments Seuil 0.5

Groupe 1 (8 médicaments)	sulfaSALAzine, hydrALAZINE, sulfiSOXAZOLE, hydrOXYzine, HYDROcodone, HYDROmorphone, oxyCODONE, oxyMORphone
Groupe 2 (2 médicaments)	DAPTOmycin, DACTINomycin
Groupe 3 (2 médicaments)	romiDEPsin, romiPLOstim

Groupe 4 (2 médicaments)	levoFLOXacin, levOCARNitine
Groupe 5 (1 médicaments)	NexIUM
Groupe 6 (2 médicaments)	penicillAMINE, penicillin
Groupe 7 (3 médicaments)	fentaNYL, SUFentanil, SORAfenib
Groupe 8 (2 médicaments)	predniSONE, methylPREDNISolone
Groupe 9 (1 médicaments)	NIFFedipine
Groupe 10 (2 médicaments)	valGANciclovir, valACYclovir

Groupe 11 (3 médicaments)	eriBULin, HumuLIN, HumaLOG
Groupe 12 (1 médicaments)	NexAVAR
Groupe 13 (2 médicaments)	hydroCHLOROthiazide, chlordiazePOXIDE
Groupe 14 (1 médicaments)	idaruCIZUmab
Groupe 15 (3 médicaments)	acetoHEXAMIDE, methazolAMIDE, acetaZOLAMIDE
Groupe 16 (1 médicaments)	NovoLIN
Groupe 17 (2 médicaments)	CeleXA, CeleBREX
Groupe 18 (1 médicaments)	epiRUBicin

Tableau 4.5 Liste des noms de médicaments Seuil 0.6

<u>GROUPE 1:</u> sulfaSALAzine	<u>Groupe 2:</u> DAPTOmycin, DACTINomycin	<u>Groupe 3:</u> hydrALAZINE, hydrOXYzine, HYDROcodone, HYDROmorphone, oxyCODONE, oxyMORphone
<u>Groupe 4:</u> romiDEPsin	<u>Groupe 5:</u> levoFLOXacin	<u>Groupe 6:</u> NexIUM
<u>Groupe 7:</u> penicillAMINE, penicillin	<u>Groupe 8:</u> fentaNYL, SUFentanil	<u>Groupe 9:</u> predniSONE
<u>Groupe 10:</u> NIFFedipine	<u>Groupe 11:</u> valGANciclovir, valACYclovir	<u>Groupe 12:</u> sulfiSOXAZOLE
<u>Groupe 13:</u> eriBULin	<u>Groupe 14:</u> NexAVAR	<u>Groupe 15:</u> hydroCHLOROthiaz ide
<u>Groupe 16:</u> idaruCIZUmab	<u>Groupe 17:</u> acetoHEXAMID E, acetaZOLAMID E, methazolAMID E	<u>Groupe 18:</u> HumuLIN
<u>Groupe 19:</u> chlordiazepOXI DE	<u>Groupe 20:</u> NovoLIN	<u>Groupe 21:</u> CeleXA
<u>Groupe 22:</u> epiRUBicin	<u>Groupe 23:</u> levOCARNitine	<u>Groupe 24:</u> CeleBREX

<u>Groupe 25:</u> romiPLOstim	<u>Groupe 26:</u> SORAfenib	<u>Groupe 27:</u> methylPREDNISolone
<u>Groupe 28:</u> HumaLOG		

4.4.4 Discussion

En considérant la liste de 39 médicaments de la FDA comme référence pour le regroupement des médicaments, on observe que, les paires de la liste FDA sont bien regroupées avec un seuil égal à 0.4 et un nombre de groupe égale à 6. Ce résultat indique que le choix de ce seuil permet de capturer efficacement les similarités orthographiques entre les noms de médicaments potentiellement confondus. La majorité des paires listées par la FDA apparaissent dans les mêmes clusters, ce qui valide la pertinence de l'approche employée.

Ces résultats confirment l'intérêt de combiner des mesures comme **BiSim** avec des techniques de clustering pour détecter automatiquement les noms de médicaments à risque de confusion. Une telle approche peut constituer un outil d'aide pour les autorités de santé et les pharmaciens afin de prévenir les erreurs de dispensation liées aux noms de médicaments similaires.

4.5. Conclusion

Ce chapitre présente les résultats d'application de notre approche qui consiste à regrouper les noms de médicaments jugés similaires, en utilisant une mesure de similarité Bi-Sim et une méthode de regroupement binaire.

Ces résultats obtenus avec différents paramètres (seuils de similarité et nombre de groupes) au cours de ce chapitre ont permis de valider l'efficacité de l'approche proposée pour le regroupement des noms de médicaments à risque de confusion

CONCLUSION ET PERSPECTIVES

Conclusion

Ce travail a permis d'explorer et de mettre en œuvre des solutions innovantes pour identifier et prévenir les risques de confusion entre les médicaments look-alike. À travers une analyse approfondie de l'état de l'art, nous avons mis en évidence les causes et les conséquences de ces erreurs, tout en recensant les bonnes pratiques et les réglementations en vigueur. Les méthodes de similarité textuelle, notamment la similarité Bi-Sim, elle s'est révélée efficace pour calculer les similitudes entre les noms de médicaments et l'application du regroupement binaire pour regrouper les noms similaires présentant des risques de confusion.

L'implémentation de ces méthodes dans l'environnement MATLAB a démontré leur applicabilité pratique, avec des résultats visuels clairs et exploitables sous forme de matrices de similarité, de graphes et de clusters, où chaque cluster représente un nombre de médicaments similaires.

Perspectives :

Cependant, des pistes d'amélioration restent ouvertes, comme :

- L'intégration de la similarité phonétique Sound-Alike,
- Appliquer la distance Levenshtein qui a été définie dans le chapitre 3.
- Appliquer des mesures de cohérences pour comparer les résultats de regroupement, comme celle de la mesure de Silhouette.

Références

- [1] Bernardi, L. (2008). *Identification des médicaments injectables et risque d'erreurs : état des lieux*. Université de Genève.
- [2] Agence Nationale de Sécurité du Médicament (ANSM). (2016). *Attention aux confusions de noms de médicaments*.
- [3] Promis, A. S. (2014). *Les médicaments à risque dans la sécurisation de la prise en charge médicamenteuse*. Université de Bordeaux.
- [4] Levenshtein, V.I., « Binary codes capable of correcting deletions, insertions, and reversals », *Sov. Phys. Dokl.*, vol. 10, n° 8, p. 707–710, 1966.
- [5] Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8), 707–710.
- [6] Duroch S. (2011). *Similitudes visuelles et sonores des médicaments: confusions et risques dans le cadre de l'exercice officinal*, Thèse, Nancy 1.
- [7] KONDRAK, Grzegorz et DORR, Bonnie J. *A similarity-based approach and evaluation methodology for reduction of drug name confusion*. 2003.
- [8] SADIAH, Halimah Tus, ISHLAH, Muhamad Saad Nurul, ROKHMAH, Nisa Najwa, et al. Autocorrect on drugs e-dictionary search module using levenshtein distance algorithm. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 2020, vol. 4, no 1, p. 64-69.