République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

UNIVERSITE YAHIA FARES DE MEDEA

Faculté de Technologie

Département de Génie des Procédés et Environnement

Mémoire de fin d'études de Master

Filière : Génie des Procédés

Spécialité : Génie des Procédés Pharmaceutiques

THEME :

Utilisation de la méthode matricielle pour la prédiction de l'activité biologique des médicaments. Application au cancer de la peau

Proposé et dirigé par :

- ✓ Pr. BENKORTBI Othmane
- ✓ Dr. MADANI Achouak

Présenté par :

- ✓ KIRAT Asmaa
- ✓ FEKHAR Zineb

Promotion 2023_2024

Dédicaces

Avec l'expression de ma reconnaissance, je dédie ce modeste travail :

A l'homme, mon précieux offre du Allah, qui doit ma vie, ma réussite et tout mon respect : mon cher père Mohamed.

A la femme qui a souffert sans me laisser souffrir, qui n'a jamais dit non âmes exigences et qui n'a épargné aucun effort pour me rendre heureuse adorable mère Houria.

A mes chère sœurs Sara et Anfel et mes frères Abdellatif et Houcine qui n'ont pas cessée de me conseiller, encourager et soutenir tout au long de mes études. Que Dieu les protège et leurs offre la chance et le bonheur.

A tout ma famille. Qu'Allah leur donne une longue et joyeuse vie.

A tous les cousins, spécialement Hadil et Houda.

A ma meilleure amie jusqu'à maintenant Meriem. Merci pour ton amour et encouragement.

A ma binôme et copine Zineb pour son soutien moral, sa patience et sa compréhension tout au long de ce projet.

A la fin et sans oublier les étudiants de mon année qui ont été les meilleurs et qui ont fait de notre étude des souvenirs inoubliable.

Asmaa

Je dédie ce travail

À ceux qui ont façonné ce que je suis aujourd'hui... à ceux qui ont veillé tard et travaillé dur pour m'élever jusqu'à ce point, mes parents honorables, ma mère attentionnée, Karima, et mon père bien-aimé, Abd Al Karim. Merci infiniment pour vos soutiens, vos confiances et vos encouragements constants, qui m'ont conduit au succès dans tous mes établissements.

À ma chère sœur, Kaouther, et son mari, Djamel Al Dine Mouhoubi, ainsi qu'à mes chers frères Mohammed, Ahmed et Abbas Mustapha, pour leur soutien indéfectible. Je leur souhaite une réussite éclatante dans toutes leurs vies. Tous mes sentiments de respect et de reconnaissance pour le soutien qu'ils n'ont cessé de me porter tout au long de cette période.

À ma chère copine et binôme, Asmaa, un grand merci pour ta présence à mes côtés. Je te suis particulièrement reconnaissante pour tous les moments agréables que nous avons partagés et pour ta précieuse contribution à la réussite de ce travail. Merci pour tout ce que tu as fait.

Sans oublier ma chère copine, Meriem, qui m'a constamment soutenu et encouragé moralement tout au long de ma démarche.

À tous les étudiants de ma promotion, ainsi qu'à tous ceux que j'aime et qui m'aiment.

Zineb

Remerciement

Louange à Allah Tout-Puissant qui nous a aidés à accomplir cette recherche scientifique et qui nous a inspiré santé, bien-être et volonté d'achever ce travail malgré tous les obstacles et les difficultés.

Nous tenons à exprimer nos profonds remerciements et notre gratitude à notre encadrante, le Dr. Madani, pour son soutien continu et sa disponibilité exceptionnelle, tant sur le plan professionnel qu'humain. Ses conseils éclairés ont grandement contribué à notre compréhension tout au long de cette recherche, Sa confiance en nos capacités et sa volonté de nous pousser à donner le meilleur de nous-mêmes ont été des moteurs essentiels pour atteindre nos objectifs et dans le développement de nos compétence notre savoir-faire.

Nous souhaitons également remercier chaleureusement notre directeur de mémoire, le Pr. Benkortbi, Maître de conférences à l'université de Médéa, qui nous a proposé le sujet de ce mémoire et qui a accepté de nous encadrer. Son dévouement, sa patience et son expertise ont été des piliers essentiels dans la réalisation de ce travail. Ses orientations précieuses, ses encouragements constants et son soutien inconditionnel ont été une source d'inspiration tout au long de notre parcours académique.

Nos sincères remerciements vont également aux membres du jury, le Pr LAIDI et le Dr BELAISSA, pour avoir accepté de juger et d'examiner ce travail

Nous tenons à exprimer notre gratitude à tous les professeurs de notre département pour leur aide précieuse qui nous a accompagnés tout au long de ces cinq années et qui a grandement contribué à notre cheminement académique.

ملخص

يعد سرطان الجلد الميلانوما و هو نوع عدواني من سرطان الجلد تحديا كبيرا في مجال الاورام. على الرغم من التقدم الكبير في العلاج فإن الحاجة إلى استراتجيات علاجية جديدة تظل ضرورية. تستخدم هذه الدراسة خوارزمية تصفية تعاونية تعتمد على عامل المصفوفة للتنبؤ بتفعلات الدواء مع الهدف. يظهر النموذج دقة عالية حيث بلغ MSE=0. 31و MAE=0.34, RMSE=0.57 متفوقا على نماذج QSAR من حيث الموثوقية. تقدم هذه النتائج افاقا واعدة لإعادة توجيه الأدوية و تحديد العلاجات الجديدة.

كما ساعدتنا النتائج أيضا في تحديد الوظائف المساعدة للتفاعل مع الهدف مثل الإيثر و الميثوكسيل .

الكلمات المفتاحية : *تفاعلات , تفكك المصفوفات , ميلانوما , QSAR , توجيه الادوية*

Résumé :

Le mélanome, un cancer de la peau agressif, demeure un défi en oncologie. Bien que des avancées aient été réalisées, de nouvelles stratégies thérapeutiques sont essentielles. Cette étude utilise un algorithme de filtrage collaboratif basé sur la factorisation matricielle pour prédire les interactions médicament cible. Le modèle démontre une précision élevée avec des scores MSE = 0.31, MAE = 0.34, et RMSE = 0.57, surpassant les modèles QSAR en fiabilité. Ces résultats offrent des perspectives prometteuses pour le repositionnement de médicaments et l'identification de nouveaux traitements.

Les résultats nous ont également aidées à identifier les fonctions auxiliaires pour l'interaction avec la cible comme l'éther et le méthoxyle.

Mots clés : *interactions, prédiction, factorisation matricielle, mélanome, QSAR , repositionnement de médicaments*

Summary

Melanoma, an aggressive form of skin cancer, remains a major challenge in oncology. Despite significant advancements in treatment, new therapeutic strategies are essential. This study utilizes a collaborative filtering algorithm based on matrix factorization to predict drug target interactions. The model demonstrates high accuracy with MSE = 0.31, MAE = 0.34, and RMSE = 0.57, surpassing QSAR models in reliability. These results offer promising prospects for drug repositioning and the identification of new treatments.
The results also helped us identify auxiliary functions for interaction with the target like ether and methoxyl.

Key words: *interactions, prediction, matrix factorization, melanoma, QSAR,* Drug repositioning

Contents

Liste des figures

Liste des tableaux

Liste des abréviations

Liste des symboles

INTRODUCTION GENERALE : 1

I. Cancer 3

I.1. Définition de cancer 3

I.2. Les causes 4

I.2.1. Les facteurs internes 4

I.2.2. Les facteurs externes 4

I.3. Les types de cancer 5

I.3.1. L'organe 5

I.3.2. La nature 5

I.4. Les symptômes 5

I.4.1. Les douleurs 6

I.4.2. Saignement 6

I.4.3. Perte de poids et fatigue 6

I.4.4. Gonflement ganglions lymphatique 7

I.4.5. Caillots du sang 7

I.4.6. Symptômes respiratoires 7

I.4.7. Symptômes neurologiques et musculaires 7

I.5. Le traitement 7

I.5.1. Le but 7

I.5.2. Le choix 8

I.5.3. Les types du traitement 8

II. Cancer de la peau 14

II.1. Définition 14

II.2. Les causes 14

II.3. Les symptômes 15

II.4. Les types 15

II.4.1. Carcinome basocellulaire 15

II.4.2. Carcinome épidermoïde 15

II.4.3. Mélanome malin 16

II.5. Les stades 16

II.6. Dépistage du mélanome 17
II.7. Diagnostique 17
III. Interaction médicament-cible 19
III.1. Le fonctionnement d'un médicament dans le corps 19
III.2. Les types de traitement 19
III.2.1. Traitement Curatif 19
III.2.2. Traitement symptomatique 20
III.2.3. Traitement préventif 20
III.3. Mécanisme d'action du Principe actif P.A 20
.III.4 Les cibles réceptrices 21
III.4.1. Transporteur pompe et canaux ioniques 21
III.4.2. Récepteur enzymatique 21
III.4.3. Cibles de P.A sur neurotransmission 22
III.4.4. Récepteur nucléaire 22
III.4.5. Récepteur membranaire 22
III.5. La cible V600E-BRAF : 23
III.5.1. Gène BRAF : 23
III.5.2. Mutation V600E : 23
III.5.3. Effets biologiques : 23
III.5.4. Implication clinique : 23
IV. Factorisation matricielle et systèmes de recommandation 24
IV.1. Système de recommandation 24
IV.1.1. Définition 24
IV.1.2. Les types de données de SR 25
IV.1.3. Les différentes approches de SR 26
IV.1.4. Les problèmes de SR 29
IV.2. Factorisation matricielle 30
IV.2.1. Définition 30
IV.2.2. Principe de fonctionnement du FM 30
IV.2.3. Les modèles du FM 31
IV.2.4. Le point d'intérêt 33
IV.2.5. Le rôle de la FM dans les SR 34
V. Matériel et méthodes 36
V.1. Méthodologie de calcul 36
V.2. Matériel utilisé 37

V.2.1. Micro-ordinateur 37
V.2.2. Les outils et logiciels 37
V.3. Préparation de la base de données 39
V.3.1. BindingDB 39
V.3.2. KIBA 40
V.4. Prétraitement de la base 40
V.5. Factorisation matricielle 41
V.6. Les optimiseurs (SGD, Adam, RMSprop) 42
V.6.1. SGD 42
V.6.2. RMSprop 43
V.6.3. ADAM 43
V.7. Critère de validation 44
V.8. Test externe 45
V.8.1. Collecte de la base de données 45
V.8.2. Pré procédure de la base 61
V.9. Conclusion 62
VI. Résultats et discussion 63
VI.1. Optimisation du modèle 63
VI.1.1. Les optimiseurs 63
VI.1.2. Les hyperparamètres 67
VI.2. Performance de modèle : 68
VI.2.1. BindingDB 69
VI.2.2. KIBA 71
VI.3. Test externe 74
VI.3.1. BindingDB 74
VI.3.2. KIBA 82
VI.4. Comparaison 90
VI.4.1. Entre le modèle initial et amélioré 90
VI.4.2. Entre KIBA et BindingDB 94
VI.4.3. Entre la factorisation matricielle et le QSAR 101
CONCLUSION GENERALE : 105
Bibliographie 106

Liste des figures

Figure I-1. Formulation du cancer 3
Figure I-2. Développement du cancer 4
Figure I-3 Chirurgie retirement de la masse tumorale 9
Figure I-4 la chimiothérapie 10
Figure I-5la radiothérapie externe 11
Figure I-6la curiethérapie 11
Figure I-7la radiothérapie métabolique 12
Figure I-8 : la thérapie ciblée 12
Figure I-9 : l'immunothérapie 13
Figure II-1 : différence entre les trois types de cancer de la peau 16
Figure II-2: méthode ABCDE pour mélanome 18
OrganigrammeIV-1 : les étapes d'un système de recommandation 25
Organigramme V-1:Protocole général d'établissement du modèle 36
Figure VI-1: courbe de perte BindingDB (entrainement/test) de l'optimiseur SGD 63
Figure VI-2: courbe de perte KIBA (entrainement/test) de l'optimiseur SGD 63
Figure VI-3:courbe de perte BindingDB (entrainement/test) de l'optimiseur RMSprop 64
Figure VI-4:courbe de perte KIBA (entrainement/test) de l'optimiseur RMSprop 64
Figure VI-5:courbe de perte BindingDB (entrainement/test) de l'optimiseur Adam 64
Figure VI-6:courbe de perte KIBA (entrainement/test) de l'optimiseur Adam 64
Figure VI-7: la courbe de perte quadratique moyenne de BindingDB (loss) 70
Figure VI-8: la courbe de dispersion de BindingDB 71
Figure VI-9: la courbe de perte quadratique moyenne de KIBA 72
Figure VI-10: la courbe de dispersion de KIBA 73
Figure VI-11:la courbe de perte quadratique moyenne (MSE) de modèle initial pour BindingDB 90
Figure VI-12:la courbe de dispersion de modèle initial pour BindingDB 90
Figure VI-13:la courbe de perte quadratique moyenne (MSE) de modèle initial pour KIBA 91
Figure VI-14:la courbe de dispersion de modèle initial pour KIBA 91

Liste des tableaux

Tableau IV-1:exemple de points d'intérêts et ses informations associées 33
Tableau V-1: valeurs correspondantes aux différents éléments utilisés dans la base de données. 40
Tableau V-2: critères de validation 45
Tableau V-3:les données de la base de test 46
Tableau VI-1: La moyenne des pertes (d'entraînement/de test) et le temps pris pour chaque combinaison de taille de lot et de facteur latente base BindingDB. 67
Tableau VI-2: La moyenne des pertes (d'entraînement/de test) et le temps pris pour chaque combinaison de taille de lot et de facteur latente base KIBA. 68
Tableau VI-3: les valeurs des critères de validation 68
Tableau VI-4: les prédictions de BindingDB 74
Tableau VI-5: Propriétés physiques et chimiques des meilleures 3 molécules de BindingDB 77
Tableau VI-6: Propriétés physiques et chimiques des mauvaises 3 molécules de BindingDB 78
Tableau VI-7: Propriétés physiques et chimiques des meilleures 3 molécules de KIBA 85
Tableau VI-8: Propriétés physiques et chimiques des molécules de comparaison de KIBA. 86
Tableau VI-9: Propriétés physiques et chimiques des derniers 3 molécules de KIBA 87
Tableau VI-10: Les ajustements des paramètres du modèle 90
Tableau VI-11: L'effet des ajustements des paramètres du modèle sur les critères de validation 90
Tableau VI-12: l'ordre des substances pour chaque une des bases 94

Liste des abréviations

ACD: Advanced Chemistry Development

ADMET: Absorption, Distribution, Métabolisme, Elimination, Toxicité

ATP : Adénosine Triphosphate

ADN: Acide Désoxyribonucléique

ARN : Acide Ribonucléique

BS: Batch Size

CTLA4: Cytotoxic T-lymphocyte-associated protein 4

DMSO: Diméthylsulfoxyde

DTI: Drug Target Interaction

FC : filtrage collaborative

FM: Factorisation Matricielle

GABA : Acide Gamma-Aminobutyrique

HRE : Hormone Responsive Élément

IMDb : Internet MovieDatabase

KIBA: Kinase Inhibitor Bio Activity

MAE: Mean Absolute Error

MAPE: Mean Absolute Percentage Error

MEK: Mitogen-Activated Protein Kinase

MSE: Mean Squared Error

OMS: Organisation Mondiale de la Santé

PA : Principe Actif

PBS: Phosphate Buffered Saline

PD1: Programmed cell death protein 1

QSAR: Quantitative Structure-Activity Relationship

RMSE: Root Mean Squared Error

RMSprop: Root Mean Square Propagation

RS : Recommandation System

SAR: Structure-Activity Relationship

SGD: Stochastic gradient descent

SVM: Support Vector Machine

TDC: Thérapeutique Data Commons

TDM : Tomodensitométrie

Liste des symboles

$\boldsymbol{m_w^{t+1}}$: La moyenne mobile des gradients (Adam).

n :Le nombre de données.

Qi : La fonction de perte pour l'élément (i, j) de la matrice de notation.

t: L'itération actuelle de l'algorithme.

$\boldsymbol{t}$: L'itération d'entraînement en cours

$\upsilon(\mathbf{w}, \mathbf{t})$: La moyenne mobile en termes de carré moyen (RMSprop).

W : Les matrices A et B mises à jour après une itération

$\boldsymbol{w^t}$: Les matrices A et B avant la mise à jour (Adam).

w : Les matrices A et B avant la mise à jour.

$\mathbf{Y}, \mathbf{Y_{cal}}$:La sortie réelle et la sortie attendue.

$\boldsymbol{\beta_1, \beta_2}$: Les facteurs d'oubli pour le premier moment et le second moment

η : Le taux d'apprentissage

$\boldsymbol{\epsilon}$: scalaire utilisé pour empêcher la division par 0

$\boldsymbol{\nabla_w L^t}$: Gradient de la fonction de perte par rapport aux paramètres à l'itération t .

$\boldsymbol{\gamma}$: Le facteur d'oubli dans RMSprop

υ_w^{t+1} : La moyenne mobile des carrées gradients (Adam)

INTRODUCTION GENERALE :

Le mélanome malin cutané, un cancer de la peau, est le plus mortel parmi tous les cancers cutanés, représentant 1,6 % de tous les cancers détectés. Il a connu une augmentation significative de son incidence chez les populations blanches au cours des dernières décennies (OMS)[1]. En Algérie, ce fléau a provoqué 306 nouveaux cas et 113 décès en 2020, et ces statistiques continuent d'augmenter chaque année, cela met en évidence l'importance d'approfondir les études sur ce genre de cancer.[2]

La détection précoce du mélanome malin offre la possibilité de guérir par chirurgie, mais au moment où il se propage, le pronostic du mélanome métastatique est défavorable. Il y a dix ans, de nouvelles immunothérapies et thérapies ciblées (inhibiteurs de BRAF, MEK et immunothérapies anti-CTLA4, anti-PD1) ont été introduites, ce qui a permis d'augmenter pour la première fois la survie globale des patients. Pourtant, le mélanome cutané demeure une affection mortelle qui nécessite une meilleure compréhension afin d'améliorer l'efficacité des traitements et d'obtenir des réponses cliniques durables. [1]

De nombreux travaux ont été réalisés pour améliorer la qualité du traitement du cancer de la peau. Parmi ceux-ci, une étude basée sur des molécules appelées « Terpénoïdes », d'origines végétales a révélé qu'elles possèdent des propriétés anticancéreuses. Cette étude a présenté les résultats de la recherche sur les Terpénoïdes végétaux dans les modèles de mélanome in vitro et in vivo[3]. Une autre étude a présenté les derniers développements dans le domaine des nouveaux agents anti-mélanome. Cette revue a discuté l'activité biologique de ces agents ainsi que de la relation structure-activité (SAR)[4]. D'autres recherches ont été menée pour examiner les dérivés de 2-(1H-imidazol-2-yl) pyridines en utilisant la technique du Docking pour prédire les propriétés pharmacocinétiques et toxicologiques (ADMET) en qu'inhibiteurs du V600E-BRAF[5]. En dépit de ces avancées, le mélanome demeure une maladie coûteuse qui requiert des traitements plus performants.

Dans ce cadre, l'objectif de ce travail de mémoire est focalisé sur la prédiction des inhibiteurs plus efficaces pour l'une des mutations responsables du mélanome. Pour ce faire, un processus fondamental de découverte de médicaments appelé DTI (Drug-Target Interaction) a été utilisé, impliquant la prédiction des interactions entre les médicaments et les cibles en utilisant des méthodes basées sur des algorithmes d'apprentissage automatique. Cette technique joue un rôle essentiel dans la réutilisation des médicaments, car elle permet de comprendre comment un médicament potentiel interagit avec sa cible, influençant ainsi

l'efficacité et la spécificité du traitement. De plus, la DTI offre l'avantage de nécessiter moins de coûts et de temps par rapport aux méthodes expérimentales biochimiques. Son principe repose sur le fait que des médicaments similaires peuvent partager des cibles similaires, ce qui permet d'explorer davantage des cibles potentielles pour les médicaments existants.[6]
L'utilisation de l'algorithme de factorisation matricielle a été proposée pour une forme de filtrage collaboratif comme méthode de prédiction des interactions médicaments-cible. L'avantage de cet algorithme, ne nécessite pas la connaissance des propriétés chimiques des composés, les propriétés génomiques des protéines cibles ni les réseaux d'interactions connus tels que le Docking ou le QSAR, il repose uniquement sur la similarité chimique et génomique pour son fonctionnement.

L'objectif principal de cette étude est de fournir une perspective globale et approfondie sur l'importance d'utilisation de la factorisation matricielle dans la découverte des médicaments et de développer un outil capable de prédire l'activité inhibitrice de certains composés contre une cible spécifique du mélanome, notamment BRAF-V600E. Cette prédiction permet de faciliter la recherche de nouvelles applications pour les médicaments déjà existants de manière plus précise, rapide et économique. Le mémoire est présenté de la manière suivante :

- Chapitre 1 et 2 : Étude bibliographique sur le cancer en général et sur le cancer de la peau en particulier
- Chapitre 3 et 4 : Présentation des généralités sur les interactions médicament-cible, le fonctionnement des médicaments dans l'organisme, et le système de recommandation, en mettant l'accent sur les différentes approches telles que la factorisation matricielle pour une forme de filtrage collaboratif.
- Chapitre 5 : consacré à la conception du modèle de factorisation matricielle afin que l'explication détaillée de chaque phase de réalisation du travail.
- Chapitre 6 : Présentation et discussion des résultats obtenus par le modèle établi.
- L'étude est terminée par une conclusion générale qui regroupe tous les résultats trouvés et permet de proposer des perspectives qui pourront être envisagées dans le futur.

Chapitre I : LE CANCER

I. Cancer

Le cancer est une maladie omniprésente dans la société contemporaine, exerçant chaque année son impact sur d'innombrables personnes dans le monde. Malgré des progrès notables dans la compréhension et la lutte contre ce fléau, le cancer reste un immense obstacle tant pour la santé publique que pour la recherche scientifique.

I.1. Définition de cancer

Le cancer est une maladie dans laquelle certaines cellules du corps se développent de façon incontrôlable et se propagent à d'autres parties du corps. Normalement, les cellules humaines se développent et se multiplient (par un processus appelé division cellulaire) pour former de nouvelles cellules que le corps a besoin d'elles. Lorsque les cellules vieillissent ou sont endommagées, elles meurent et de nouvelles cellules prennent leur place. Parfois, ce processus ordonné se décompose, et les cellules anormales ou endommagées se développent et se multiplient quand elles ne devraient pas.

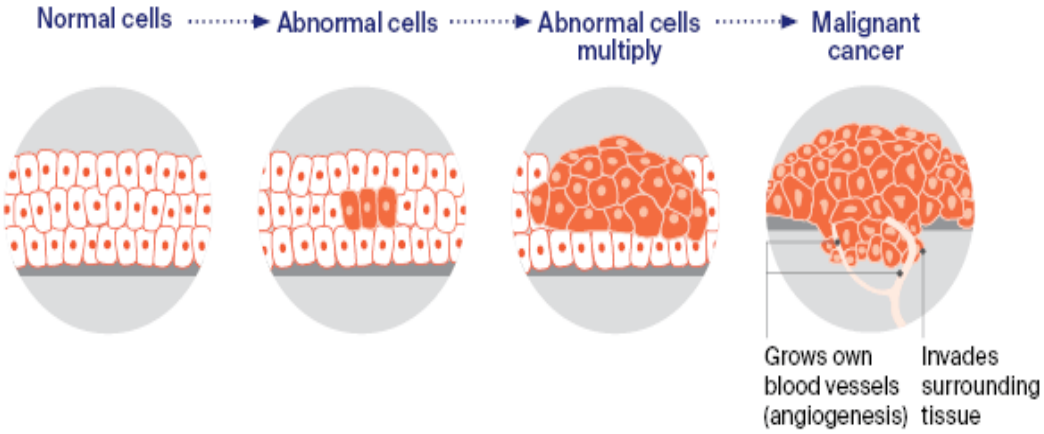


Figure I-1. *Formulation du cancer*

Les tumeurs cancéreuses se propagent ou envahissent les tissus avoisinants et peuvent se déplacer vers des endroits éloignés du corps pour former de nouvelles tumeurs (un processus appelé métastase) [7]

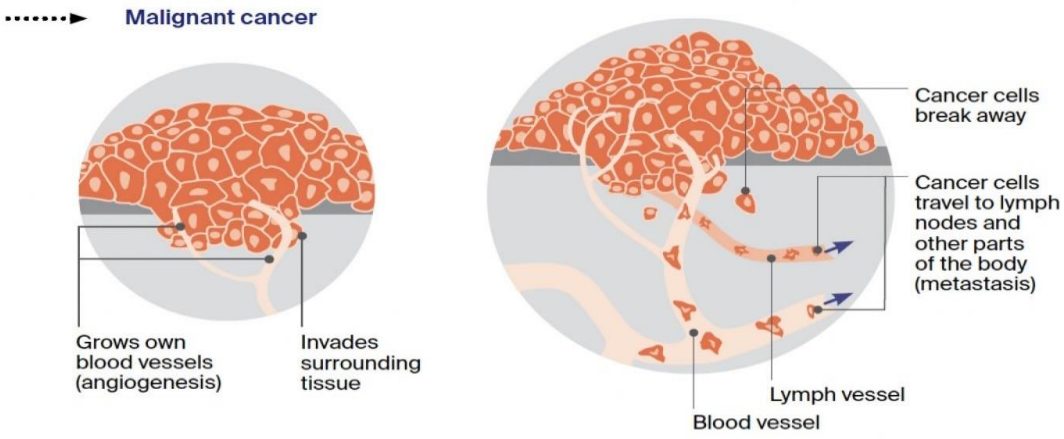


***Figure I-2**. Développement du cancer*

Selon l'organisation mondiale de la santé (OMS) : c'est un terme générique désignant un large groupe de maladies pouvant toucher n'importe quelle partie de l'organisme. On parle aussi de tumeurs malignes et de néoplasmes. L'un des traits caractéristiques du cancer est la multiplication rapide de cellules anormales à la croissance inhabituelle, qui peuvent ensuite envahir des parties voisines de l'organisme, puis migrer vers d'autres organes. On parle alors de métastases. La présence de métastases étendues est la principale cause de décès par cancer[2].

I.2. Les causes

Il existe deux types de facteurs de risque entrent en jeu, les facteurs **"internes"** et **"externes"**. En raison de leur possible combinaison, le cancer est une maladie multifactorielle[8].

I.2.1. Les facteurs internes

Sont principalement liés à l'âge et l'hérédité. Face à l'augmentation du nombre de cas chaque année, le cancer du sein est le plus évoqué quand on parle de facteur héréditaire (63% des personnes interrogées). Émerge un important besoin d'être informé sur le suivi de personnes dites "à risque" (71%), les risques à proprement dits et les tests à réaliser (46%).

I.2.2. Les facteurs externes

Sont liés à l'hygiène de vie et l'environnement : ils causent 4 cancers sur 10.

- Le tabagisme ;
- La consommation d'alcool ;
- Une alimentation déséquilibrée (consommation excessive de produits transformés, de viande, peu de légumes et fruits frais) ;

- Le manque d'activité physique ;
- Le surpoids et l'obésité.

Autres facteurs aujourd'hui avérés : l'exposition aux rayons UV sans protection, aux rayons ionisants comme dans le cas du cancer de la peau. Ou encore à certains produits cancérigènes (amiante, nickel, chrome, causant des cancers "professionnels")[9].

I.3. Les types de cancer

Chaque organe et chaque type de cellule peut être à l'origine d'une tumeur cancéreuse donc les types de cancer sont classés selon :

I.3.1. L'organe

Cancer de peau, cancer du sein, cancer de la prostate, cancer de la colorectale (gros intestin) …etc.

I.3.2. La nature

I.3.2.1. Les carcinomes

Les cellules cancéreuses apparaissent dans les cellules épithéliales, le tissu qui recouvre la surface interne (le tissu qui recouvre un organe) ou la surface externe (comme l'épiderme). Dans cette famille, on distingue les adénocarcinomes qui se développent à partir des cellules épithéliales de glandes comme le sein et la prostate.

I.3.2.2. Les sarcomes

Les cellules cancéreuses apparaissent dans un tissu « de support » comme les os, la graisse ou les muscles. On parle d'ostéosarcomes (sarcomes des os), de liposarcomes (sarcomes des tissus graisseux) et de rhabdomyome sarcomes (sarcomes des muscles striés).

I.3.2.3. Les cancers hématopoïétiques ou hématologiques

Les cellules cancéreuses apparaissent dans la moelle osseuse qui fabrique les cellules du sang (globules rouges et blancs et plaquettes) et leurs précurseurs. Elles peuvent également apparaître dans les autres organes lymphoïdes (thymus, ganglions lymphatiques, rate, amygdales…). Il existe trois familles de cancers hématologiques : les leucémies, les myélomes et les lymphomes[10].

I.4. Les symptômes

Au début, le cancer est constitué d'une petite masse de cellules et n'entraîne aucun symptôme. Au fur et à mesure que le cancer se développe, sa présence physique peut affecter les tissus adjacents. De la même manière, certains cancers sécrètent des substances ou

déclenchent des réactions immunitaires qui entraînent des symptômes dans d'autres régions de l'organisme, éloignées du cancer (syndromes paranéoplasiques).

I.4.1. Les douleurs

Le plus souvent, le cancer se développe en silence, sans faire mal. C'est ce qui le rend si difficile à diagnostiquer. La tumeur n'a donc aucune sensibilité et n'est pas douloureuse. Bien que la douleur puisse être un symptôme précoce de certains cancers, comme les tumeurs cérébrales qui causent des maux de tête et les cancers qui affectent la tête et le cou, et les cancers de l'estomac et de l'œsophage qui causent des problèmes de déglutition. Au fur et à mesure qu'ils grandissent, les premiers symptômes sont généralement un léger inconfort, qui s'aggrave ensuite jusqu'à devenir une douleur intense, transformant l'invasion en une invasion tumorale. Les douleurs causées par la tumeur sont mécaniques, inflammatoires ou neuropathiques[11].

I.4.2. Saignement

Initialement, un cancer peut saigner légèrement car ses vaisseaux sont fragiles. Ensuite, quand le cancer augmente de volume et envahit les tissus environnants, il peut proliférer dans un vaisseau adjacent et entraîner une hémorragie. Celle-ci peut être discrète et indétectable ou détectable seulement par des examens, comme c'est souvent le cas dans les stades précoces de cancer du côlon. Mais cette hémorragie peut être, notamment en cas de cancers avancés, plus importante, massive, voire mortelle.

Le siège du cancer détermine le site de l'hémorragie. Un cancer qui se développe sur un segment du tube digestif, quel qu'il soit, peut être à l'origine d'une présence de sang dans les selles (méléna [sang noir] ou rectorragies [sang rouge]). Un cancer de l'appareil urinaire peut induire la présence de sang dans les urines (hématurie). D'autres types de cancers peuvent saigner à l'intérieur de l'organisme. Le saignement dans les poumons peut provoquer l'émission de sang lors de la toux (hémoptysie)[12].

I.4.3. Perte de poids et fatigue

Souvent, les personnes atteintes de cancer subissent une perte de poids et une faiblesse (fatigue) qui s'aggrave à mesure que la maladie progresse. Certains patients perdent du poids malgré un bon appétit, tandis que d'autres souffrent de perte d'appétit voire d'anorexie ou de difficulté à avaler. Ils peuvent devenir très fins. Les personnes atteintes d'un cancer avancé sont souvent très fatiguées, La fatigue liée au cancer est beaucoup plus importante. Cette sensation d'épuisement, non liée à un effort particulier, voire d'affaiblissement général de

l'organisme (asthénie) n'est pas comparable aux ressentis d'une fatigue "normale". Elle perdure même après une nuit de sommeil.

I.4.4. Gonflement ganglions lymphatique

Lorsque le cancer commence à se propager dans tout le corps, il peut initialement affecter les ganglions lymphatiques voisins, les faisant grossir. Les ganglions lymphatiques gonflés sont souvent indolores et peuvent prendre une consistance ferme ou caoutchouteuse. Ils peuvent être mobiles ou, à un stade plus avancé, ils peuvent se coller aux tissus adjacents ou entre eux.

I.4.5. Caillots du sang

Certains cancers produisent des substances qui entraînent une formation excessive de caillots du sang, principalement dans les veines des jambes (thrombose veineuse profonde). Les caillots du sang dans les veines des jambes peuvent parfois se désolidariser et migrer dans un poumon (embolie pulmonaire), une situation qui peut être fatale. Une coagulation excessive est fréquente chez les personnes atteintes de tumeurs pancréatiques, pulmonaires et autres tumeurs solides, ainsi que chez les personnes atteintes de tumeurs cérébrales.

I.4.6. Symptômes respiratoires

Le cancer peut comprimer ou obstruer les voies respiratoires au niveau du poumon, entraînant de ce fait une gêne respiratoire, une toux ou une pneumonie. La gêne respiratoire peut également se produire lorsque le cancer provoque un important épanchement pleural, un saignement dans les poumons ou une anémie.

I.4.7. Symptômes neurologiques et musculaires

Le cancer peut infiltrer ou comprimer les fibres nerveuses ou la moelle épinière et produire des symptômes neurologiques et musculaires, comme une douleur, une faiblesse ou une altération de la sensibilité (paresthésies). Quand un cancer se développe dans le cerveau, les symptômes sont difficiles à situer avec précision, mais ils comprennent des confusions, des vertiges, des céphalées, des nausées, des troubles de la vision et des convulsions. Les symptômes neurologiques peuvent aussi faire partie d'un syndrome paranéoplasique.

I.5. Le traitement

I.5.1. Le but

Selon le cas, les traitements ont pour objectif de :

- Supprimer la tumeur ou les métastases.
- Réduire le risque de récidive.

- Ralentir le développement de la tumeur ou des métastases.
- Prévenir et traiter les symptômes et les complications engendrés par la maladie et les traitements pour assurer la meilleure qualité de vie possible.

I.5.2. Le choix

Le choix de traitement est adapté en fonction de chaque situation, Les médecins décident du meilleur traitement pour le cancer d'une personne en se basant sur :

- La chance de guérison (le cancer est éliminé et ne revient pas)
- L'espérance de vie la plus longue, la possibilité de mener une vie confortable
- L'efficacité du traitement sur les symptômes
- Les effets secondaires du traitement
- Les souhaits de la personne concernant le traitement

Le médecin doit discuter avec la personne des avantages et des inconvénients des différents traitements. Les protocoles de traitement sont des méthodes standardisées développées par les médecins pour traiter certains types de cancer. Ces programmes ont été testés avec rigueur au cours d'études de recherche appelées essais cliniques. Dans les essais cliniques, les médecins comparent les nouveaux médicaments ou traitements avec des traitements habituels pour savoir si les nouveaux traitements sont plus efficaces. De nombreuses personnes atteintes de cancer ont la possibilité de participer à un essai clinique. Le médecin peut expliquer les avantages et les inconvénients d'une participation à un essai clinique.

I.5.3. Les types du traitement

I.5.3.1. La chirurgie

Le principal traitement du cancer est la chirurgie. Elle est utilisée d'environ 80 % des cas uniquement quand le cancer ne s'est pas propagé (métastasé) ailleurs. L'organisme de la personne est suffisamment fort pour subir la chirurgie[13]. Ce traitement consiste à opérer le malade à fin d'éliminer les tumeurs cancéreuses solides Cette stratégie a pour objectif :

- Curative : lorsqu'elle permet de retirer 100 % des cellules tumorales
- Diagnostique : le tissu retiré est analysé pour préciser la nature de la lésion et faciliter le choix de traitements curatifs complémentaires
- Palliative : pour soulager la douleur liée à la taille de la tumeur, pour faciliter le fonctionnement de l'organe atteint[14].

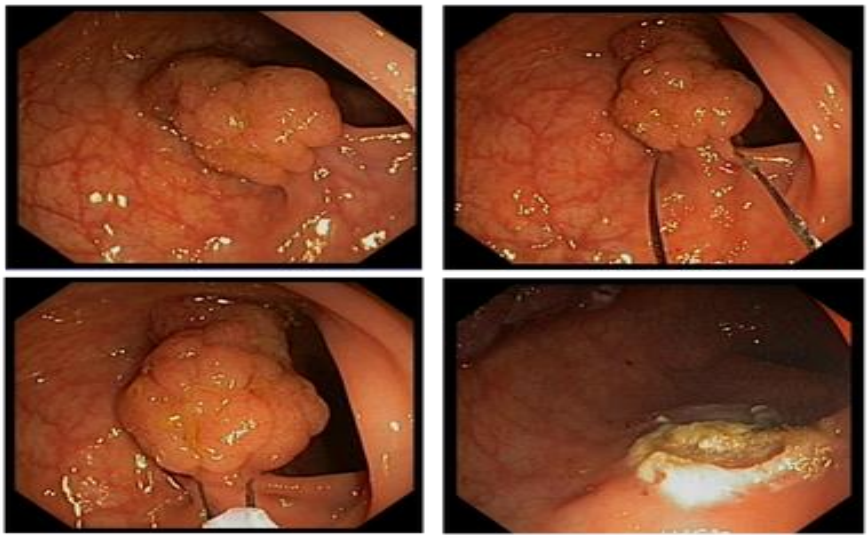


Figure I-3*Chirurgie retirement de la masse tumorale*

I.5.3.2. La chimiothérapie

La chimiothérapie est un traitement du cancer qui repose sur l'administration d'un ou plusieurs médicaments, dit « cytotoxique » Le choix des médicaments est adapté en fonction de chaque situation, chaque cancer est particulier et nécessite un traitement approprié. Elle vise à combattre et éliminer les cellules tumorales présentées dans l'organisme quel que soit ses localisations, Y compris celles qui n'ont pas été repérées par les examens d'imagerie. La chimiothérapie agit soit en les détruisant directement, soit en les empêchant de se multiplier[15].

Cette stratégie peut être proposée comme seul traitement ou être associé aux autres traitements, elle est proposée généralement dans les cas suivants :

- **Avant une chirurgie :** On parle alors de chimiothérapie néo-adjuvant. Cette chimiothérapie a pour but de diminuer la taille de la tumeur.
- **Après une chirurgie complète de la tumeur :** c'est-à-dire lorsque le chirurgien a enlevé toutes les cellules cancéreuses visibles. C'est alors une chimiothérapie adjuvante. La chimiothérapie a pour but de diminuer les risques de récidive locale ou à distance. La chimiothérapie complète alors la chirurgie.
- **Pour traiter des métastases :** c'est-à-dire des cellules cancéreuses qui se sont propagées dans d'autres parties du corps. On dit que c'est une chimiothérapie métastatique[16]



***Figure I-4** la chimiothérapie*

I.5.3.3. La radiothérapie

La radiothérapie consiste à délivrer localement des rayons, ou des radiations ionisantes, qui provoquent des dégâts majeurs au niveau de l'ADN. Comme les cellules cancéreuses ne parviennent pas à réparer ces lésions aussi bien que les cellules saines, elles ne peuvent plus se multiplier et/ou meurent. La radiothérapie peut être le seul traitement mis en place pour traiter un cancer, mais elle est généralement couplée à une ou plusieurs autres thérapies[17].

- Radiothérapie néo-adjuvant : radiothérapie administrée avant l'opération pour réduire la taille d'une tumeur afin qu'elle soit plus facile à retirer par intervention chirurgicale.
- Radiothérapie adjuvante : après l'ablation chirurgicale d'une tumeur, la radiothérapie peut être utilisée pour éliminer les éventuelles cellules cancéreuses subsistant localement.
- Chimio radiothérapie : la radiothérapie peut être pratiquée en même temps que la chimiothérapie pour en renforcer l'effet. qui rend les cellules cancéreuses plus sensibles aux rayons[18]. Cette stratégie est administrée sous trois défirent formes :

La radiothérapie externe : Les rayons sont émis par une machine appelée accélérateur linéaire de &particules, située à proximité du patient et dirigée vers la région du corps à traiter. Ces rayons traversent la peau pour atteindre la zone à traiter.

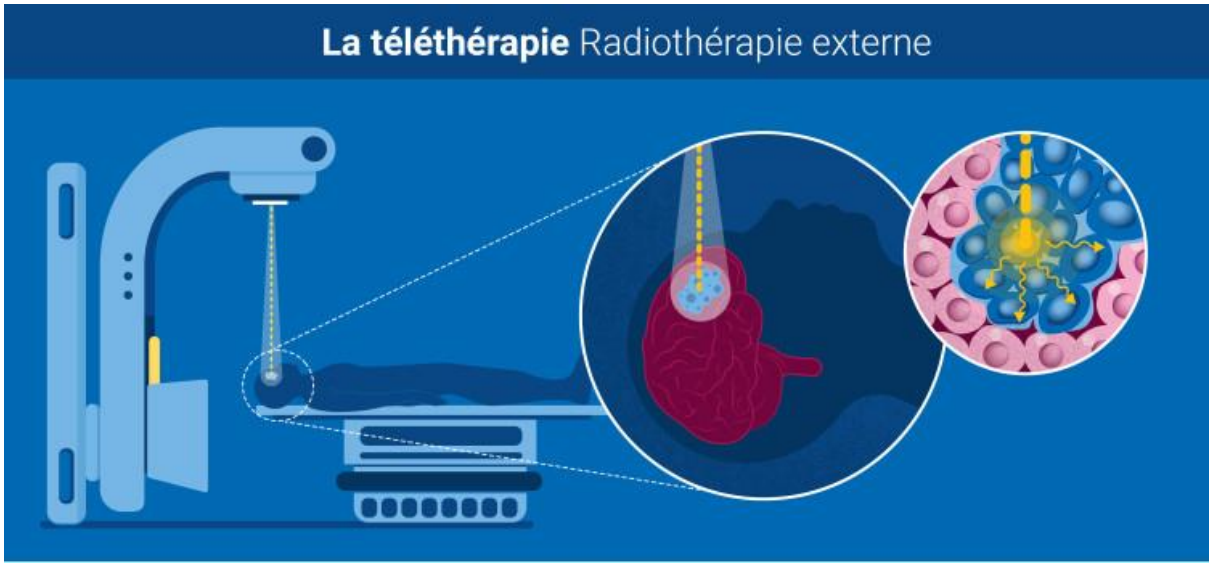


Figure I-5*la radiothérapie externe*

La curiethérapie : des sources radioactives sont mises en contact direct avec la zone à traiter à l'intérieur du corps afin d'épargner les tissus sains environnants. C'est un traitement le plus souvent dédié à des cancers localisés avec des indications spécifiques.

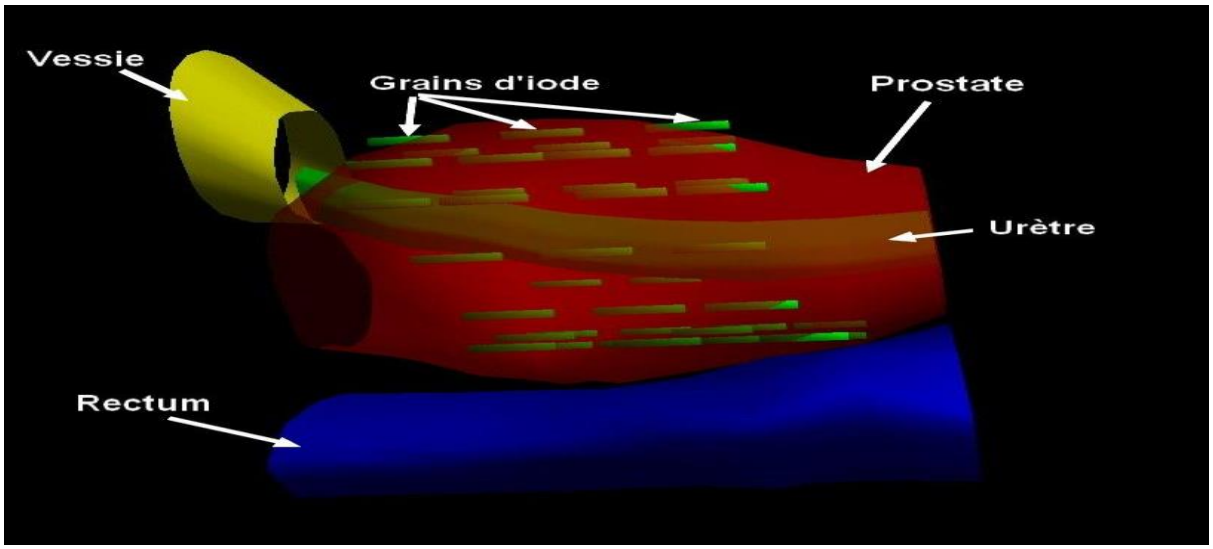


Figure I-6*la curiethérapie*

La radiothérapie métabolique : les sources radioactives sont administrées par voie orale (boisson ou capsule) ou par injection intraveineuse. Ces sources radioactives se fixent ensuite sur les cellules cancéreuses pour les détruire.



Figure I-7*la radiothérapie métabolique*

I.5.3.4. Thérapie ciblée

Les thérapies ciblées constituent une autre famille de traitements du cancer le plus souvent disponibles par voie orale. En ciblant spécifiquement certaines molécules de l'organisme, elles bloquent des mécanismes qui sont indispensables à la prolifération des cellules cancéreuses ou, plus globalement, au développement de la tumeur.

Les médicaments ciblés visent des molécules à l'intérieur ou à la surface des cellules cancéreuses pour ralentir la croissance du cancer, Selon les cancers, les thérapies ciblées peuvent être prescrites seules, en association entre elles ou avec d'autres traitements[19].

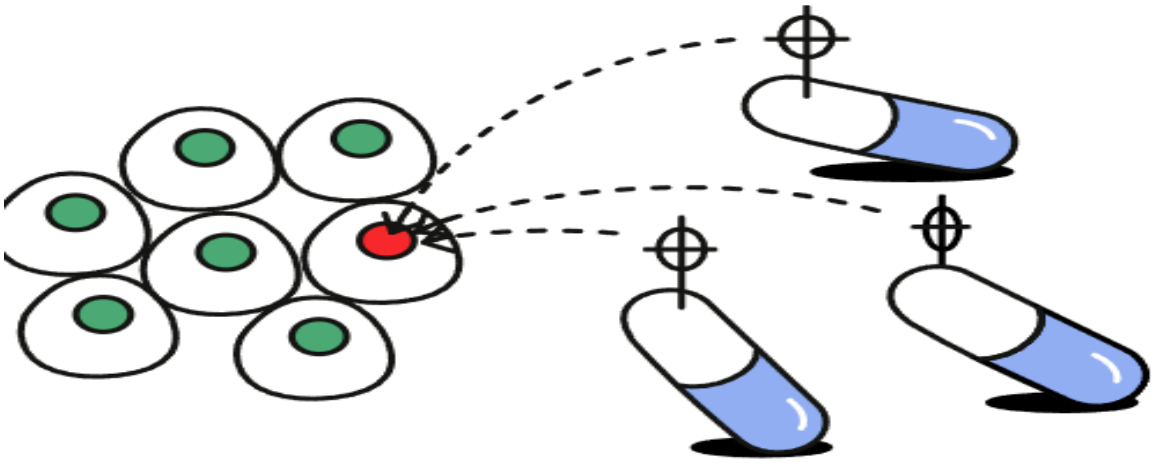


Figure I-8 : *la thérapie ciblée*

I.5.3.5. L'hormonothérapie

La croissance de certains cancers est favorisée par les hormones sexuelles produites par l'organisme : ainsi, certaines tumeurs du sein ou de l'utérus croissent sous l'action des

œstrogènes ou de la progestérone, et certains cancers de la prostate progressent sous l'action de la testostérone. Les hormones sexuelles agissent sur les cellules tumorales en se fixant à leur surface au niveau de récepteurs spécifiques. Les médicaments d'hormonothérapie bloquent la synthèse de ces hormones ou empêchent leur fixation aux récepteurs. Ils sont administrés essentiellement par voie orale.

Un traitement par hormonothérapie peut être prescrit lorsque le cancer est dit « hormonodépendant », c'est-à-dire que sa croissance dépend de l'action d'hormones sexuelles. Avant le traitement, les médecins recherchent donc la présence de récepteurs aux hormones sur les cellules tumorales. Cette vérification passe par une analyse moléculaire, conduite à partir d'échantillons de la tumeur[20].

I.5.3.6. L'immunothérapie

L'immunothérapie regroupe un ensemble de stratégies visant à mobiliser ou à renforcer les défenses immunitaires des patients de manière à ce qu'elles s'attaquent aux cellules tumorales.

a. **L'immunothérapie « spécifique » :** Consiste à bloquer spécifiquement des protéines à la surface des cellules cancéreuses ou dans leur microenvironnement pour freiner la croissance tumorale.
b. **L'immunothérapie active :** vise à stimuler le système immunitaire du patient pour qu'il réagisse contre les cellules tumorales[21].

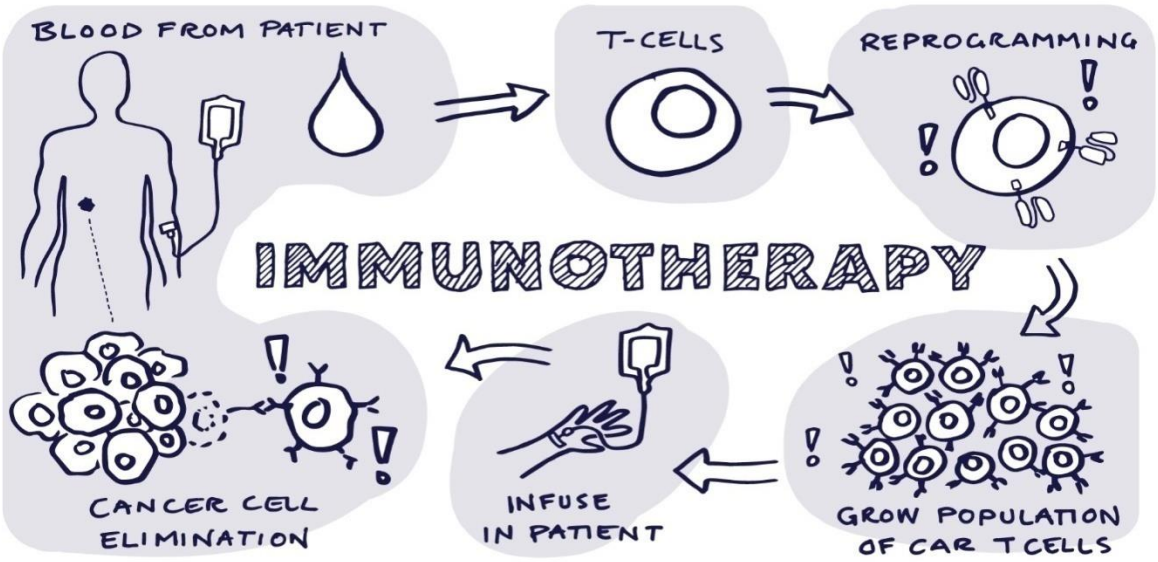


Figure I-9 : *l'immunothérapie*

Chapitre II :
LE CANCER DE LA PEAU

II. Cancer de la peau

II.1. Définition

Le cancer de la peau se manifeste par la prolifération incontrôlée et anormale de cellules de la peau, et plus précisément de l'épiderme, la couche externe de la peau. Cette multiplication pathologique est causée par des lésions non réparées de l'ADN, qui entraînent des mutations de ces cellules, venant à former une tumeur maligne.8 mélanomes sur 10 se développent sur une peau saine, qui ne présente pas de lésion ni de tache. Ce qui rend leur diagnostic délicat[22].

II.2. Les causes

Les causes exactes du cancer de la peau ne sont pas entièrement comprises, mais plusieurs facteurs de risque ont été identifiés. Les principales causes sont l4 exposition aux rayons ultraviolets (UV) du soleil et des lits de bronzage, ainsi que des facteurs génétiques Et environnementaux. Les facteurs de risque pour le cancer de la peau comprennent :

(a) ***L'exposition aux rayons UV :*** Une exposition excessive aux rayons UV du soleil ou des lits de bronzage est le principal facteur de risque de tous les types de cancer de la peau. Les rayons UV endommagent l'ADN des cellules cutanées, ce qui peut entraîner une croissance anormale et un cancer de la peau.

(b) ***Antécédents familiaux de cancer de la peau :*** Les personnes ayant des antécédents familiaux de cancer de la peau ont un risque plus élevé de développer la maladie. Certains types de cancer de la peau, tels que le mélanome, peuvent être hérités.

(c) ***Teint de peau clair :*** Les personnes ayant une peau claire sont plus susceptibles de Développer un cancer de la peau, car leur peau contient moins de mélanine, qui protège contre Les rayons UV.

(d) ***Exposition professionnelle :*** Certaines professions, telles que les agriculteurs et les Travailleurs de la construction, peuvent être exposées à des niveaux élevés de rayons UV ou à Des produits chimiques carcinogènes, augmentant ainsi leur risque de cancer de la peau.

(e) ***Maladies cutanées préexistantes :*** Les personnes atteintes de certaines maladies Cutanées, telles que la xérodermie pigmentosum, ont un risque accru de développer un cancer de la peau

(f) ***Médicaments immunosuppresseurs*** : Les personnes qui prennent des médicaments immunosuppresseurs, tels que les transplantés d'organes, sont plus susceptibles de Développer un cancer de la peau en raison de leur système immunitaire affaibli

II.3. Les symptômes

Les symptômes de mélanome peuvent varier d'une personne a l'autre. Voici quelques signes et symptômes courants :

- Changement dans la taille, la forme ou la couleur d'une tache de naissance existante.
- Apparition d'une nouvelle tâche de couleur brun foncé, noire ou multicolore sur la peau.
- Asymétrie : une moitié de la tache ou de la lésion est différente de l'autre.
- Bords irréguliers ou flous autour de la tache ou de la lésion.
- Évolution rapide de la taille ou de l'apparence de la tache ou de la lésion.
- Présence de plusieurs couleurs dans la tache ou la lésion, telles que du noir, du brun, du rouge, du bleu ou du blanc.
- Sensation de démangeaisons, de picotements ou de douleur au niveau de la tache ou de la lésion.
- Saignement ou formation de croûtes sur la tache ou la lésion.
- Ulcération ou formation d'une bosse au niveau de la tache ou de la lésion[23].

II.4. Les types

Il existe plusieurs types de cancer de la peau, chacun ayant des caractéristiques spécifiques. Les principaux types de cancer de la peau comprennent :

II.4.1. Carcinome basocellulaire

C'est le type le plus commun de cancer de la peau. Il se développe dans les cellules basales de l'épiderme, la couche supérieure de la peau. Le carcinome basocellulaire est généralement causé par une exposition au soleil excessif et est plus fréquent chez les personnes de plus de 50 ans. Il a un taux de guérison élevé et est rarement fatal.

II.4.2. Carcinome épidermoïde

Également appelé cancer de la peau à cellules squameuses, se développe dans les cellules squameuses de l'épiderme. Il est également souvent causé par une exposition excessive au soleil et peut se produire sur n'importe quelle partie de la peau, mais il est plus fréquent sur

les zones exposées au soleil. Comme le carcinome basocellulaire, le carcinome épidermoïde a un taux de guérison élevé s'il est diagnostiqué et traité tôt.

II.4.3. Mélanome malin

C'est un type plus rare de cancer de la peau, mais il est également le plus dangereux. Le mélanome se développe dans les cellules qui produisent le pigment de la peau (mélanocytes) et peut se propager rapidement à d'autres parties du corps. Le mélanome est souvent causé par une exposition au soleil, mais il peut également être causé par d'autres facteurs tels que des antécédents familiaux de mélanome.



Figure II-1 : *différence entre les trois types de cancer de la peau*

II.5. Les stades

Le stade du cancer de la peau fait référence à la gravité et à la progression de la maladie. Il est Déterminé par la taille de la lésion, la profondeur de l'invasion tumorale et la présence de métastases (Propagation de la tumeur à d'autres parties du corps). Les stades d'évolution du cancer de la peau Les plus couramment utilisés sont :

- ✓ **Stade 0 :** Le cancer de la peau est limité à la couche la plus superficielle de la peau (épiderme) et ne S'est pas propagé à d'autres parties du corps. Ce stade est souvent appelé carcinome in situ.
- ✓ **Stade I :** Le cancer de la peau mesure moins de 2 centimètres de diamètre et n'a pas encore atteint les ganglions lymphatiques ou d'autres parties du corps.
- ✓ **Stade II :** Le cancer de la peau mesure plus de 2 centimètres de diamètre et/ou s'est propagé aux ganglions lymphatiques adjacents, mais n'a pas encore atteint d'autres parties du corps.

- ✓ **Stade III :** Le cancer de la peau s'est propagé aux ganglions lymphatiques régionaux ou à des tissus voisins, mais n'a pas encore atteint d'autres parties du corps.
- ✓ **Stade IV :** Le cancer de la peau s'est propagé à des organes distants du corps, tels que le cerveau, les poumons ou le foie. Ce stade est souvent appelé cancer de la peau métastatique ou avancé Il est important de noter que les stades d'évolution peuvent varier en fonction du type de cancer de la peau[24].

II.6. Dépistage du mélanome

Le diagnostic du mélanome est souvent effectué par un dermatologue ou un médecin de soins primaires. Les étapes courantes du diagnostic sont :

(a) Examen visuel : le médecin examine la peau à la recherche de lésions cutanées qui pourraient être un signe de mélanome

(b) Biopsie : si le médecin suspecte un mélanome, il peut prélever un échantillon de tissu de la lésion cutanée pour une biopsie. Le tissu prélevé est examiné au microscope pour déterminer s'il s'agit d'un mélanome et, si oui, quel type.

(c) Imagerie : dans certains cas, une imagerie comme une radiographie, une échographie ou une tomodensitométrie (TDM) peut être utilisée pour aider à déterminer si le mélanome s'est propagé à d'autres parties du corps.

Il est important de noter que tous les cancers de la peau ne nécessitent pas une biopsie, et que les examens d'imagerie ne sont généralement pas nécessaires pour les stades précoces de la maladie. Le diagnostic précoce est crucial pour un traitement efficace du mélanome[25].

II.7. Diagnostique

Pour diagnostiquer un mélanome, le médecin traitant ou le dermatologue examinera la peau, et en particulier les grains de beauté, en appliquant la **méthode ABCDE** :

- **A comme asymétrie** : le grain de beauté n'a pas une forme symétrique ;
- **B comme bords irréguliers ;**
- **C comme couleur non homogène ;**
- **D comme diamètre supérieur à 6 mm** ;

- **E comme évolution**, c'est-à-dire que le grain de beauté change d'aspect en quelques semaines ou quelques mois.

Le médecin procède alors à une exérèse diagnostique ou une biopsie**,** c'est-à-dire qu'il retire la lésion suspecte et envoie le prélèvement en analyse pour savoir s'il est composé de cellules cancéreuses. Si le cancer de la peau est confirmé, d'autres examens sont envisagés pour constater l'étendue de la maladie et notamment sa propagation à d'autres organes ou aux ganglions lymphatiques[26].

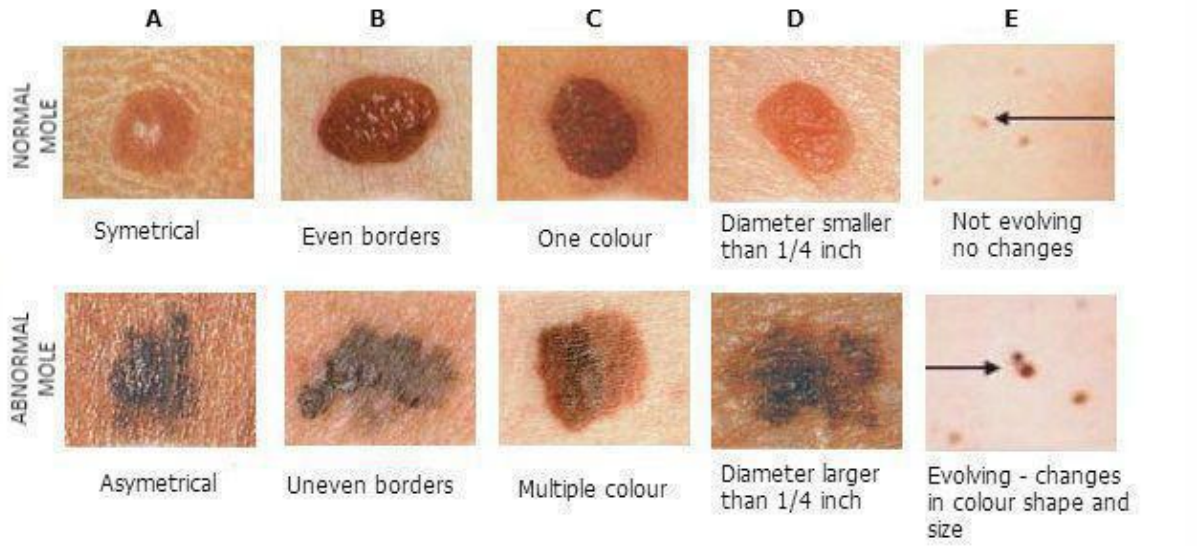


Figure II-2: *méthode ABCDE pour mélanome*

Chapitre III :
INTERACTION MEDICAMENT-CIBLE

III. Interaction médicament-cible

Le corps humain et modulent ainsi les processus physiologiques et pathologiques. Dans les domaines de la pharmacologie et de la médecine la compréhension des interactions médicament-cible est essentielle au développement de médicaments efficaces et sûrs.

Cesinteractions représentent les mécanismes par lesquels les médicaments agissent sur des ci bles biologiques dans le corps humain et modulent ainsi les processus physiologiques et path ologiques.

III.1. Le fonctionnement d'un médicament dans le corps

- Remplacer une molécule endogène déficitaire
- Diminuer le taux trop élevé d'une molécule endogène
- Agir directement sur une cible moléculaire
- Empêcher la prolifération dans l'organisme hôte
- Moduler l'expression du génome

III.2. Les types de traitement

III.2.1. Traitement Curatif

Dans le meilleur des cas, l'objectif est curatif, et a pour le but d'éliminer l'état pathologique, en influençant les réponses cellulaires et de normaliser le fonctionnement cellulaire des organes.

III.2.1.1. Un traitement étiologique : élimine la cause de la pathologie. Exemple : antibiotiques.

III.2.1.2. Un traitement symptomatique : souvent, la situation est complexe et dans ce cas-là, le traitement symptomatique élimine ou diminue les manifestations de la pathologie. Exemple : traitement antalgique contre la douleur.

III.2.2. Traitement symptomatique

Un traitement symptomatique peut avoir différents objectifs selon les circonstances :

- Soulager des symptômes transitoires. **Exemple :** mal de tête —> traitement antalgique
- En attendant l'effet du traitement étiologique. **Exemple :** abcès dentaire —> antalgique + antibiotique
- En absence de traitement étiologique : permet d'améliorer la qualité de vie du patient. **Exemple :** maladies neuro dégénératives

Un traitement symptomatique diminue les symptômes, ce qui peut rendre difficile le suivi de l'évolution de la pathologie.

III.2.3. Traitement préventif

Le but est de prévenir le développement d'une pathologie, les principaux moyens sont :

- La vaccination préventive
- La vaccination thérapeutique : favorise la production d'anticorps, afin d'aider l'organisme à mieux lutter contre la maladie

Exemples :

- Antibiothérapie préopératoire
- Traitement antipaludéen si voyage dans les pays infestés
- Traitement contraceptif par oestro-progestatifs
- Traitement anticoagulant si alitement prolongé

III.3. Mécanisme d'action du Principe actif P.A

Le médicament peut agir de plusieurs façons [27]:

- Action par interaction avec un récepteur : principal mécanisme d'action du médicament
- Action de substitution d'une molécule nécessaire à l'organisme
- Action physico-chimique : changement physique ou chimique du milieu
- Action par inhibition du métabolisme
- Action par interaction avec des transporteurs et des canaux ioniques
- Action par interaction avec des exo-biotiques
- Action au niveau des enzymes
- Action au niveau des bactéries, virus, parasites ou levures

III.4. Les cibles réceptrices

Principalement une macromolécule qui permet au principe actif d'interagir avec la cible, provoquant une réponse biologique qui produit l'effet thérapeutique souhaité, mais parfois aussi des effets indésirables. Il peut s'agir principalement de protéines, mais aussi d'enzymes ou d'acides nucléiques. Ces cibles se retrouvent partout dans l'organisme : dans le sang, dans la plupart des cellules, au niveau de la membrane cellulaire, ou encore dans le cytoplasme ou le noyau. Ces cibles biologiques jouent souvent un rôle déterminant dans l'initiation et le processus. Développement de la maladie[27]. Il s'agit de :

III.4.1. Transporteur pompe et canaux ioniques

- Récepteurs membranaires associés à un canal ionique (qui fait passer des ions)
- Présence d'un site de fixation pour un (ou plusieurs) ligand(s) sur la partie extracellulaire
- Régule l'entrée dans la cellule soit des cations sodium ou calcium, soit des chlorures pour induire une réponse cellulaire.
- Possibilité de moduler l'effet du canal activé par des modulateurs agissant sur des sites allostériques et qui ne modifient pas eux-mêmes l'ouverture du canal[27].

Exemples :

- ❖ Récepteurs nicotiniques de l'Ach (Na^+)
- ❖ Récepteurs du GABA (Cl^-),
- ❖ Effets inhibiteurs Récepteurs du glutamate (Na^+, Ca^{2+}),
- ❖ Excitateurs récepteurs de la sérotonine 5-HT3 (Na^+) Récepteurs P2X1 de l'ATP (Ca^{2+})

III.4.2. Récepteur enzymatique

Les récepteurs liés aux enzymes sont des récepteurs qui basent leur modèle fonctionnel sur l'augmentation de l'activité enzymatique. Une protéine se lie au récepteur et provoque une activité enzymatique spécifique au niveau intracellulaire. Une hormone importante qui exerce ses effets par l'intermédiaire de la tyrosine kinase est l'insuline. Les récepteurs enzymatiques sont des protéines transmembranaires avec un côté récepteur faisant face au liquide interstitiel et le côté enzymatique faisant face au cytosol. Les exemples incluent les tyrosine kinases et les guanylate cyclases.

III.4.3. Cibles de P.A sur neurotransmission

Les neurotransmetteurs sont principalement des catécholamines, ils incluent des molécules comme la noradrénaline, la dopamine, etc.

Ces neurotransmetteurs sont synthétisés à partir d'un précurseur, et sont ensuite stockées dans des vésicules, qui libèreront leur contenu lors d'une stimulation électrique, où les granules s'intègrent à la membrane, et libèrent les neurotransmetteurs. Une partie des catécholamines libérées sont récupérées par des transporteurs dans la fente synaptique.

Une fois les neurotransmetteurs capturés, ils peuvent subir une dégradation par les catéchol-O-méthyl-transférases, etc dans le cytosol. L'effet principalement recherché est l'interaction avec les récepteurs post synaptiques, qui peuvent être localisés sur des neurones, ou des cellules musculaires. Ils peuvent aussi agir sur des récepteurs pré-synaptiques, et l'interaction a pour finalité de moduler la libération des neurotransmetteurs[28].

III.4.4. Récepteur nucléaire

- Récepteurs facteurs de transcription
- Récepteurs intracellulaires de localisation nucléaire ou parfois cytoplasmique
- Contrôle de l'expression de gènes cibles en reconnaissant des séquences nucléotidiques spécifiques dans le promoteur des gènes (HRE, Hormone Responsive Élément)
- Contrôle de la transcription des gènes en ARN messages
- Contrôle de la synthèse de protéines
- Cible des hormones stéroïde sur rénaliennes et sexuelles, des hormones thyroïdiennes, vitamine A, vitamine D, les rétinoïdes[29].

III.4.5. Récepteur membranaire

Les récepteurs transmembranaires sont des protéines qui couvrent toute l'épaisseur de la membrane plasmique de la cellule, les protéines transmembranaires, avec une extrémité du récepteur à l'extérieur de la cellule (domaine extracellulaire) et une autre extrémité du récepteur à l'intérieur (domaine intracellulaire). Lorsque le domaine extracellulaire reconnaît une hormone, le récepteur entier subit un changement dans sa composition structurelle qui affecte le domaine intracellulaire, lui donnant une nouvelle action. Dans ce cas, l'hormone (ou un autre ligand) ne traverse pas la membrane plasmique pour pénétrer dans la cellule. Bien qu'un seul récepteur puisse transduire un certain signal après la liaison au ligand, le plus souvent, la liaison au ligand provoque l'association de plusieurs molécules de récepteurs[30].

III.5. La cible V600E-BRAF :

La cible V600E BRAF fait référence à une mutation spécifique du gène BRAF, qui joue un rôle crucial dans la régulation de la croissance et de la division cellulaire via la voie de signalisation MAPK/ERK. Voici une définition détaillée :

III.5.1. Gène BRAF :

Le gène BRAF code pour une protéine kinase qui fait partie de la voie de signalisation intracellulaire MAPK/ERK. Cette voie est essentielle pour la régulation de la croissance, la division et la survie des cellules.

III.5.2. Mutation V600E :

La mutation V600E implique un changement d'acide aminé dans la protéine BRAF, où la valine (V) à la position 600 est remplacée par un acide glutamique (E). Cette substitution entraîne une activation constitutive (continue) de la protéine kinase BRAF, indépendamment des signaux régulateurs externes.

III.5.3. Effets biologiques :

- Activation constante de la voie MAPK/ERK : La protéine BRAF mutée V600E active en permanence la voie de signalisation MAPK/ERK, ce qui conduit à une prolifération cellulaire incontrôlée.
- Prolifération et survie cellulaire : Cette activation aberrante favorise la croissance et la survie des cellules tumorales, contribuant ainsi au développement et à la progression du cancer[31].

III.5.4. Implication clinique :

- Prévalence dans les cancers : La mutation V600E BRAF est fréquemment observée dans divers types de cancers, notamment le mélanome, le cancer colorectal, et le cancer de la thyroïde.
- Cible thérapeutique : En raison de son rôle clé dans la prolifération des cellules cancéreuses, la mutation V600E BRAF est une cible thérapeutique importante. Les inhibiteurs de BRAF, tels que le vemurafenib et le dabrafenib, sont utilisés pour traiter les cancers porteurs de cette mutation, en bloquant l'activité de la protéine BRAF mutée et en freinant la croissance tumorale.

En résumé, la cible V600E BRAF désigne une mutation spécifique du gène BRAF qui entraîne une activation continue de la voie MAPK/ERK, conduisant à une prolifération cellulaire incontrôlée et à la progression du cancer. Cette mutation est une cible importante pour les thérapies ciblées dans divers types de cancers[32].

Chapitre IV : FACTORISATION MATRICIELLE ET SYSTEMES DE RECOMMANDATION

IV. Factorisation matricielle et systèmes de recommandation

Les vastes quantités de données échangées sur Internet constituent aujourd'hui un atout d'acquisition d'informations sans précédent. En même temps, cela pose également une difficulté, car cela nécessite de nombreux traitements de filtrage pour que les données renvoyées soient pertinentes pour l'utilisateur le plus rapidement possible ; il semble que les capacités humaines ne permettent pas une analyse exhaustive de toutes ces informations ; Pour aider les humains dans le processus de sélection, des systèmes de recommandation des consommateurs sont apparus au cours de la dernière décennie du XXe siècle.

Dans ce chapitre, nous définissons d'abord ce qu'est un système de recommandation, puis présentant les trois principales méthodes de filtrage permettant de faire des recommandations, et introduisant la factorisation matricielle, ses modèles et son rôle. Enfin nous évoquons les inconvénients de ces systèmes et leur devenir.

IV.1. Système de recommandation

IV.1.1. Définition

Les systèmes de recommandation (RS) sont des outils logiciels et des techniques fournissant des suggestions pour les articles qui peuvent être utiles à un utilisateur. Les suggestions concernent divers processus de décision, tels que les articles à acheter, la musique à écouter ou les nouvelles en ligne à lire[33].

D'après Gediminas et Alexander, « le système de recommandation aide à faire face à la surcharge d'informations et à fournir des recommandations, du contenu et des services personnalisés ». Un système de recommandation peut etre défini aussi comme une forme spécifique de filtrage de l'information visant à présenter les éléments d'information susceptibles d'intéresser l'utilisateur.

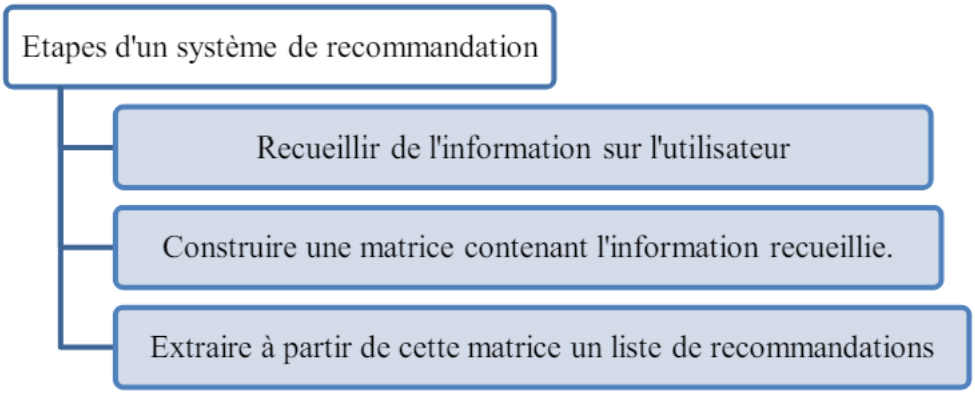
Les systèmes de recommandation peuvent être aussi vus comme une réponse donnée aux utilisateurs ayant des difficultés à prendre une décision dans le cadre d'utilisation d'un système de recherche d'information. Ils ont pour objectif de fournir à un utilisateur des ressources pertinentes en fonction de ses préférences.

Les systèmes de recommandation jouent un rôle important dans des sites Internet aussi bien classés que Amazon.com, YouTube, Netflix, Yahoo, TripAdvisor, Last.fm et IMDb.

En outre, de nombreuses entreprises de médias développent et déploient actuellement des systèmes de recommandation dans le cadre des services qu'ils fournissent à leurs abonnés. Par exemple Netflix, le service de location de films en ligne, a décerné un prix de 1 million de dollars à l'équipe qui a réussi à améliorer sensiblement les performances de son système de recommandation. L'entreprise a intégré les propositions les plus pertinentes dans sa version du système de recommandation mis en production[34]. Il existe deux entités de base dans les systèmes de recommandation sont :

- ✓ « Item » est le terme général utilisé pour dénoter ce que le système recommande aux utilisateurs.
- ✓ « Utilisateur » est la personne qui utilise un système de recommandation, donne son avis sur les différents items et reçoit les nouvelles recommandations du système.

Un système de recommandation requiert généralement 3 étapes :



OrganigrammeIV-1 : *les étapes d'un système de recommandation*

IV.1.2. Les types de données de SR

Nous distinguons deux types de collecte des données :

IV.1.2.1. Collecte explicite

Dans ce cas, les utilisateurs sont sollicités pour émettre leurs avis sur des produits/objets. Ils peuvent le faire via un système de notation (ex. une grille de 5 étoiles, un questionnaire de satisfaction), ou bien en publiant leurs avis sur un élément donné (ex. La fonction "J'aime" sur le réseau social Facebook).

IV.1.2.2. Collecte implicite

La collecte implicite s'intéresse aux interactions des utilisateurs sur le système. Les exemples de cette collecte incluent la surveillance du nombre de visites sur une page, le nombre de vues sur une vidéo, le temps passé sur une section donnée ou de l'historique des achats sur une plateforme de e-commerce[35].

IV.1.3. Les différentes approches de SR

Les systèmes de recommandation doivent être capables de prédire les intérêts des gens utilisateurs. Il est donc nécessaire de pouvoir collecter un certain nombre de données afin de créer des profils personnels pour chaque utilisateur[36]:

IV.1.3.1. Basé sur le contenu

Les systèmes de recommandation basés sur le contenu (content based) fonctionnent en analysant les caractéristiques des items à recommander (lieux, produits, etc.) puis en les regroupant. Dans ce genre de SR, des mots-clés sont utilisés afin de décrire les ressources ; un profil est ensuite créé afin d'indiquer pour chaque utilisateur le type de ressources qu'il aime. Par exemple, nous construisons un profil d'un utilisateur A qui préfère les séries dont les genres préférés sont actions et romances (cas de Netflix). Et, nous essayons de recommander des produits qui de la même section que A préfère.

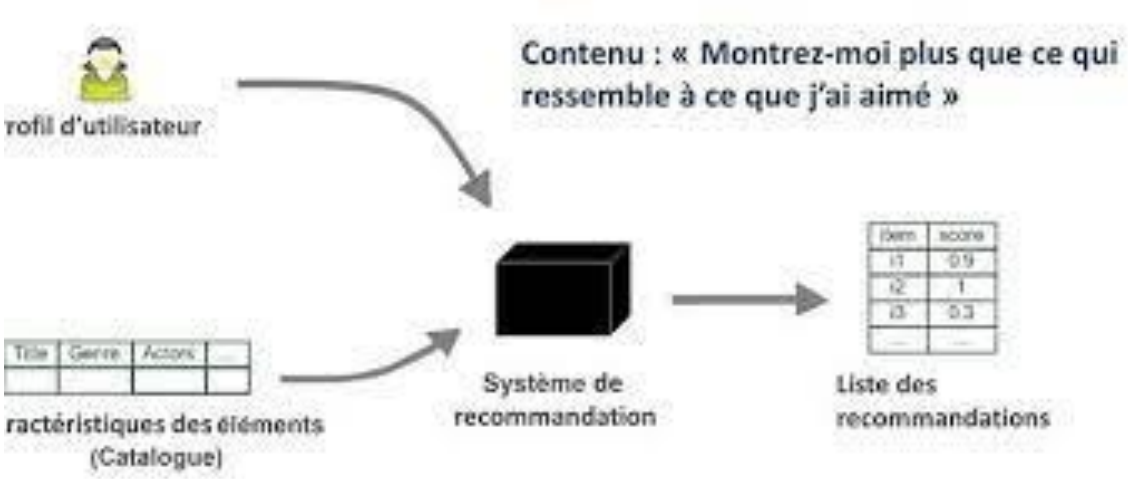


Figure IV-1*système de recommandation basée sur le contenu*

IV.1.3.2. Filtrage hybride

Un système de recommandation hybride utilise des composants de différents types d'approches de recommandation ou s'appuie sur leur logique. En général, l'hybridation s'effectue en deux phases :

- Appliquer séparément le filtrage collaboratif et autres techniques de filtrage pour générer des recommandations candidates.
- Combiner ces ensembles de recommandations préliminaires selon certaines méthodes telles que la pondération, la mixtion, la cascade, la commutation, etc. afin de produire les recommandations finales pour les utilisateurs[33].

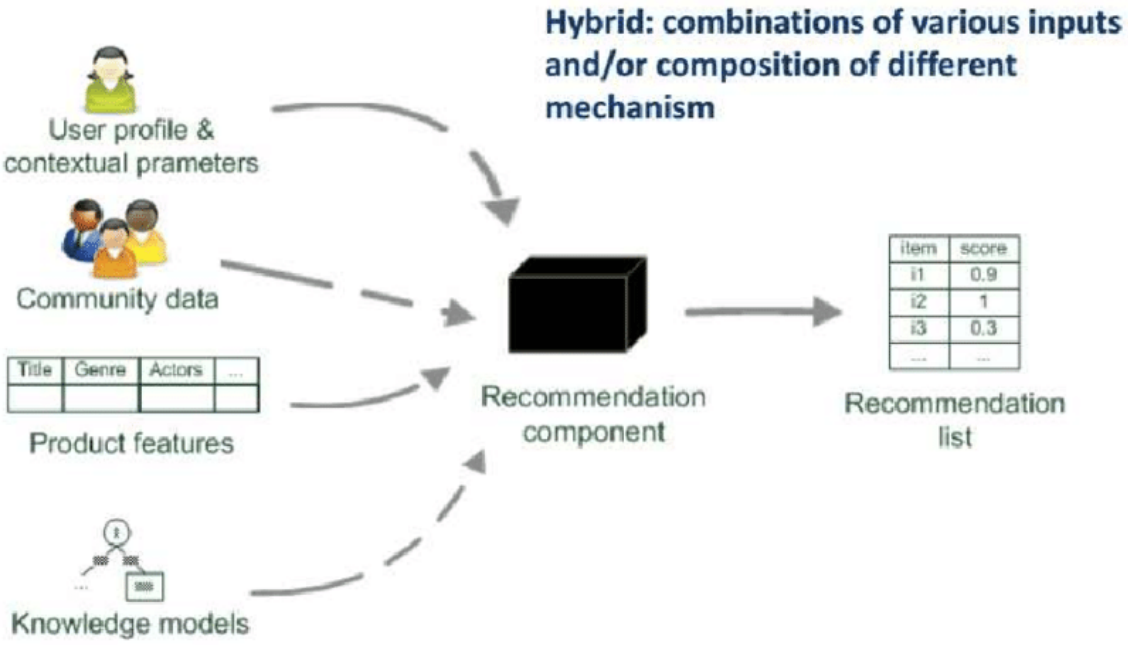


***Figure IV-2** filtrage hybride*

IV.1.3.3. Filtrage collaboratif

Les méthodes basées sur le filtrage collaboratif sont basées sur la collection et l'analyse d'informations sur le comportement des utilisateurs, leurs activités et leurs préférences pour ensuite prédire ce que les utilisateurs sont susceptibles d'aimer d'après leurs similarités avec d'autres utilisateurs. L'idée clef est que la note d'un utilisateur U pour un nouvel item i est susceptible d'être similaire à celle donnée par un autre utilisateur V, si U et V ont noté d'autres items d'une manière similaire. De même, U est susceptible de noter deux items i et j de la même façon, si d'autres utilisateurs ont donné des notes similaires à ces deux items. Les approches collaboratives dépassent certaines limitations des approches basées sur le contenu. Par exemple, des items dont le contenu n'est pas défini, peuvent quand même être recommandés aux utilisateurs grâce aux feedbacks des autres utilisateurs. De plus, les recommandations collaboratives sont basées sur la qualité des items évaluée par les utilisateurs, au lieu de s'appuyer sur le contenu qui peut être un mauvais indicateur de qualité[37].

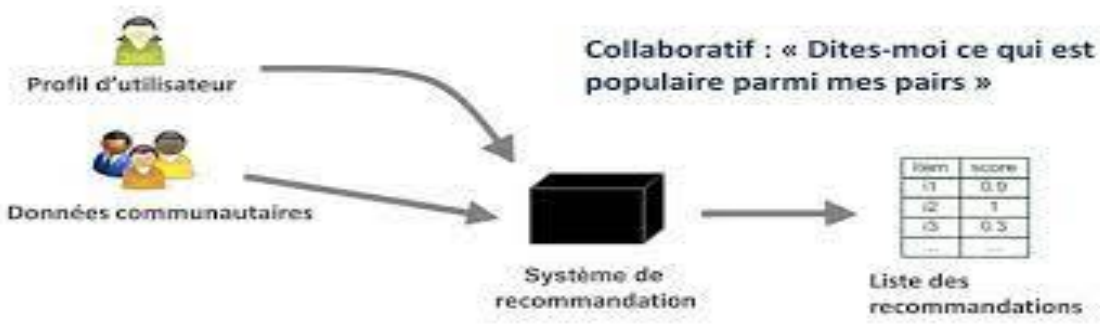
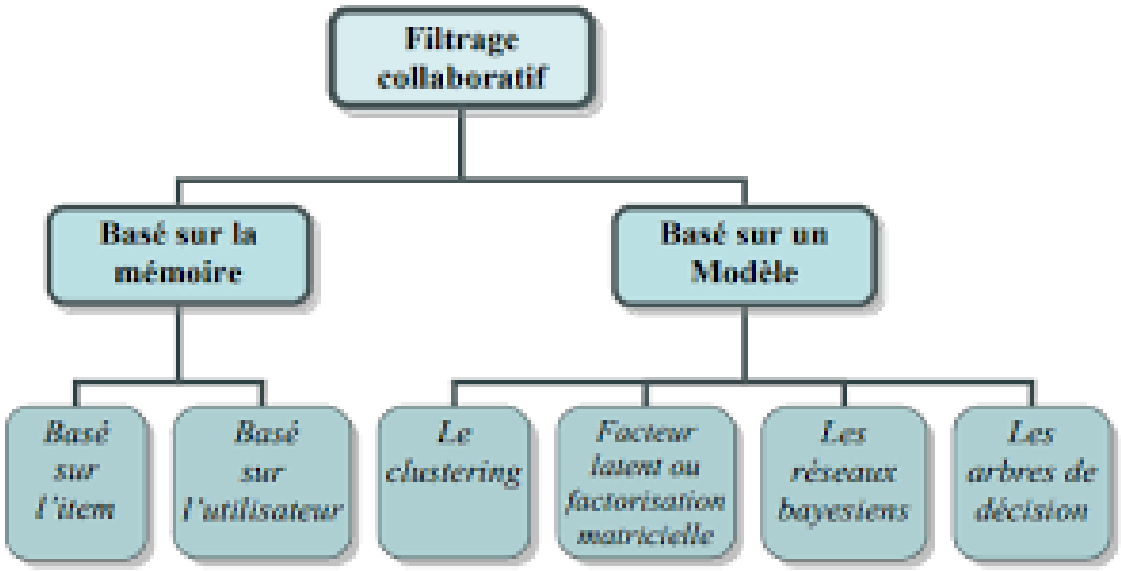


***Figure IV-3** filtrage collaboratif*

Le filtrage collaboratif est composé de deux méthodes :



***OrganigrammeIV-2 :** Les méthodes de filtrage collaboratif*

(a) Basé sur le mémoire

Cette méthode utilise un système de notation pour prédire les préférences d'un utilisateur en tenant compte des préférences d'un utilisateur similaire, ou du « voisin ». Il existe deux façons de calculer les préférences ici, le FC basé sur l'utilisateur et le FC basé sur l'item. Le FC à base de mémoire souffre essentiellement de deux problèmes : le passage à l'échelle et le manque de données. Le passage à l'échelle fait que le temps de calcul est énormément long lors de son application sur des données nombreuses. Le manque de données vient du fait qu'il n'est pas toujours facile de trouver assez de notes communes entre les utilisateurs, particulièrement sur des grandes listes d'items[38].

(b) Basé sur un modèle

Les méthodes basées sur un modèle ont été intégrées aux systèmes de recommandation pour améliorer et remédier aux problèmes des méthodes basées sur la mémoire. Les algorithmes basés sur le modèle se basent aussi sur les évaluations précédentes (les profils) des usagers, mais cette méthode ne calcule pas directement les prédictions, elle classifie les usagers suivant des groupes ou d'apprendre les modèles à partir de leurs données Pour la construction du modèle plusieurs méthodes sont utilisées. En général, les méthodes basées sur le modèle utilisent les techniques d'apprentissage automatiques, telles que le clustering, la factorisation matricielle, les réseaux bayésiens, les arbres de décision, etc.

Dans le présent travail, une attention particulière est donnée à la factorisation matricielle aussi dite décomposition matricielle. Elle consiste en la décomposition d'une matrice en plusieurs autres matrices. Pour retrouver la matrice originale, il suffira de faire le produit de ces matrices entre elles. La factorisation matricielle a donné de bons résultats dans les systèmes de recommandation [37].

IV.1.4. Les problèmes de SR

Malgré le succès des systèmes de recommandation, ils présentent certains problèmes

- ✓ ***Manque de données :*** l'identification des appréciations des utilisateurs est l'un des piliers de base du processus de recommandation. Ces appréciations sont soit renseignées explicitement par les utilisateurs eux-mêmes ou bien induites par le système sur la base de l'analyse des interactions de ces utilisateurs avec le système. Dans les deux cas, souvent les données relatives aux appréciations des utilisateurs manquent et s'avèrent insuffisantes pour le bon fonctionnement du système de recommandation. En effet, la quantité de données ou de notes disponible demeure toujours insuffisante pour pouvoir prédire correctement les notes manquantes. Donc, en raison de ce manque de données, la modélisation des utilisateurs devient complexe. Les modèles utilisateurs deviennent ainsi peu fiables.
- ✓ ***Rareté (Sparsity) Le gray Sheep :*** Quand des utilisateurs ont des gouts atypiques (qui varient de la norme), ils n'auront pas beaucoup d'utilisateurs en tant que voisins. Cela mènera à des recommandations pauvres. Ce problème est également connu comme gray sheep. La popularité est un problème qui se produit fréquemment dans le filtrage collaboratif.
- ✓ ***Le shilling :*** C'est l'action malveillante d'influencer la recommandation, en créant de faux profils pour voter et favoriser/défavoriser certains items. Par exemple, un

vendeur qui, à travers des faux profils, valorise ses produits pour que le système les recommande pour les clients en ligne.

- ✓ ***Démarrage à froid :*** Un problème des plus connus dans les systèmes de recommandation est lorsque l'on ne détient aucune ou peu d'informations sur un utilisateur, il y a deux grands types de ce problème :
  - ♣ Cas du nouvel utilisateur : est qu'un nouvel utilisateur qui n'a pas encore accumulé suffisamment d'évaluations ne peut pas avoir de recommandations pertinentes.
  - ♣ Cas du nouvel item : le système doit avoir suffisamment d'évaluations pour que l'item soit pris en considération dans le processus de recommandation[38].

IV.2. Factorisation matricielle

Factorisation matricielle est parmi les algorithmes utilisés dans le filtrage collaboratif qui appartient aux méthodes basées sur le modèle. Elle permet de donner généralement une représentation plus compacte que l'apprentissage de la matrice complète. Dans ce qui suit, le principe et les différents modèles de FM et son rôle dans les systèmes de recommandation sera envisagé.

IV.2.1. Définition

La matrice de factorisation est une matrice d'évaluation qui permet d'identifier les utilisateurs et ses interattractions sous forme des notes par rapport aux items dans une matrice de taille N*M tel que M les notes des utilisateurs et N la listes des items. Son principe est de décomposer une matrice en deux ou plusieurs autres matrices, et de calculer par la suite le produit de ces matrices pour obtenir la matrice des prédictions.

IV.2.2. Principe de fonctionnement du FM

La plupart des modèles de factorisation matricielle sont basés sur le modèle du facteur latent. C'est l'approche la plus utilisé pour réduire le problème de sparsity et le manque de données dans la matrice d'un système de recommandation[34]. Tout d'abord, nous avons un ensemble d'utilisateurs T et un ensemble des films F. Soit R la matrice de taille (m x n) qui contient toutes les notes que les utilisateurs ont attribuées aux films (rating matrix).La méthode consiste à trouver deux matrices U (m x d) et V (d x n), tel que d est le nombre facteur latent (un entier positif) et leur produit qui se rapproche étroitement de la matrice originale R est :

$R \approx V^t \times U$

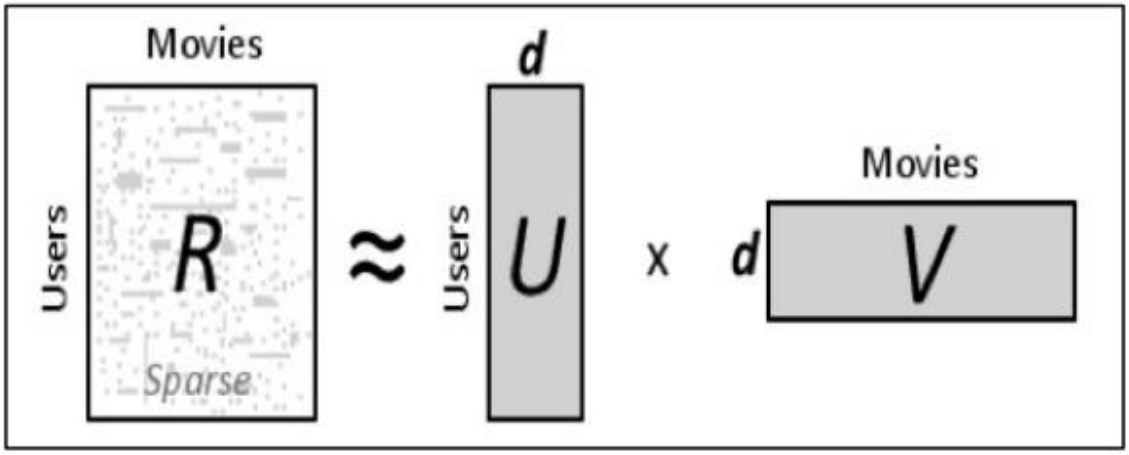


***Figure IV-4** factorisation matricielle*

IV.2.3. Les modèles du FM

L'identification de facteurs latents peut se faire à travers plusieurs modèles de factorisation matricielle, dont certains sont : la décomposition en valeur singulière (SVD), l'analyse des composantes principales (PCA), factorisation matricielle probabiliste (PMF), factorisation matricielle non négative (NMF) etc.

IV.2.3.1. La décomposition de la valeur singulière SVD

La décomposition de la valeur singulière, en Anglais (Singular Value Decomposition SVD) est la technique la plus puissante de réduction de la dimensionnalité. Il réduit d'abord la dimensionnalité d'une matrice, avant d'utiliser le résultat de ce processus pour rapprocher les scores originaux. SVD réduit une matrice (m x n) en trois matrices, U, S et V.

$$R=U\ S\ V'$$

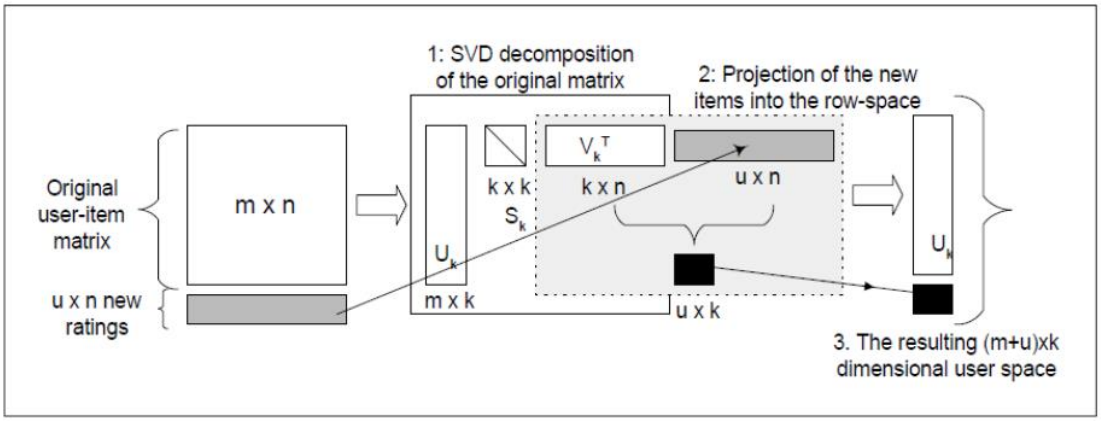


***Figure IV-5 :** modèle SVD*

U et V sont deux matrices orthogonales de taille $(m\ x\ k)$ et $(n\ x\ k)$ où k est le rang de la matrice R. S est une matrice k x k et contient toutes les valeurs singulières de R comme ses entrées diagonales. La dimensionnalité des matrices peut être réduite en choisissant S pour ne contenir que les k plus grandes valeurs singulières de R. Cela fait de S_k une matrice $k\ x\ k$, et puisque la réduction peut également être appliquée aux matrices U et V en supprimant les colonnes (r_k) des lignes U et (r_k) de V, la dimensionnalité De USV' peut être considérablement réduit en $U_k S_k V'_k$, ce qui permet d'économiser de l'espace et de la puissance de calcul en plus de supprimer les données de bruit de R.

Le résultat de l'exécution de SVD sur R est doncr_k, une version de dimension réduite de R.Ces matrices peuvent ensuite être utilisées pour prédire une note d'un utilisateur à un élément i.et c'est la technique que nous allons utiliser dans notre application[37].

IV.2.3.2. Analyse des composantes principales PCA

L'analyse des composantes principales (ACP) (en Anglais Principal Component Analysis PCA) est une réalisation particulière de l'approche de la factorisation de la matrice (FM). L'ACP est une procédure statistique qui utilise une transformation orthogonale pour convertir un ensemble d'observations de variables éventuellement corrélées en un ensemble de valeurs de variables linéairement non corrélées appelées composantes principales. Le nombre de variables d'origine est supérieur ou égal aux composants principaux.

Cette transformation est définie de telle sorte qu'une projection linéaire de données dimensionnelles élevées dans un sous-espace dimensionnel inférieur tel que la variance retenue est maximisée et l'erreur de reconstruction de moindre carré est minimisée. Les principaux composants sont orthogonaux car ils sont les vecteurs propres de la matrice de covariance. PCA est sensible à la mise à l'échelle relative des variables d'origine. PCA permet d'obtenir une liste ordonnée de composants qui représentent la plus grande quantité de variance à partir des données en termes d'erreurs de moindres carrés.

La quantité de variance saisie par le premier composant est supérieure à la quantité de variance sur le deuxième composant et ainsi de suite. Nous pouvons réduire la dimensionnalité des données en négligeant ces composants [37].

IV.2.3.3. La factorisation probabiliste matricielle PMF

La factorisation probabiliste de la matrice (en Anglais Probabilistic Matrix Factorization PMF) est un modèle linéaire probabiliste avec un bruit d'observation gaussien. La matrice de préférence de l'utilisateur est représentée comme le produit des matrices d'objets et d'éléments

de base inférieurs de filature dans la factorisation probabiliste de la matrice (PMF). Supposons que nous avons N utilisateurs et M films. R_{IJ} la valeur nominale de l'utilisateur i pour le film j, Ui et Vj représentent respectivement des vecteurs de caractéristique latente spécifiques à l'utilisateur spécifiques à la D-dimension et au film. Ensuite, la distribution conditionnelle sur les notes observées $R \in R^{n\times m}$ et les distributions antérieures sur$U \in R^{n\times d}$ et $V \in R^{d\times m}$ sont données dans l'ensemble des équations présentés ci-dessous :

$$p(R|U,V,\sigma^2) = \prod_{i=1}^{N}\prod_{j=1}^{M}\left[N\left(R_{ij}\middle|U_i^T V_j,\sigma^2\right]^{Iij} \qquad eqt\ 1$$

$$p(U|\sigma_U^2) = \prod_{i=1}^{N} N\left(U_i|0,\sigma_U^i I\right) \qquad eqt\ 2$$

$$p(V|\sigma_V^2) = \prod_{j=1}^{M} N\left(V_i|0,\sigma_V^2 I\right) \qquad eqt\ 3$$

N (x|μ, Q²) désigne la distribution gaussienne avec la moyenne μ et la variance. I ij est la variable indicatrice égale à 1 si l'utilisateur a évalué le film j et égal à 0 sinon.

IV.2.4. Le point d'intérêt

Un point d'intérêt (en Anglais Point OF Interest) est un site spécifique identifié de manière unique généralement associé à une catégorie spécifique d'activités. De même pour les enregistrements dans les réseaux sociaux basés sur la localisation (LBSN), un point d'intérêt est également associé à un contenu qui correspond à l'ensemble de tous les commentaires, images, opinions que les utilisateurs ont téléchargés lors des enregistrements qu'ils ont effectués. Par exemple, dans le tableau IV.1, le point d'intérêt de la Tour Eiffel est associé à certaines des catégories auxquelles il aspire. Cependant, dans de nombreux cas pratiques, les catégories ou autres descriptions de points d'intérêt ne sont pas divulguées pour diverses raisons (par exemple, vie privée, confidentialité, etc.). C'est pourquoi dans nos approches nous avons supposé que nous ne connaissions que les emplacements, c'est-à-dire les paires (latitude, longitude) pour tous les points d'intérêt[39].

Tableau IV-1:*exemple de points d'intérêts et ses informations associées*

Le point d'intérêt	Tour Eiffel
Checkins	7.097.302
Location	Latitude : 48.858° Longitude : N, 2.294° E
Catégories	Vues panoramiques, monument, divertissement, loisirs

IV.2.5. Le rôle de la FM dans les SR

La factorisation matricielle joue un rôle essentiel dans les systèmes de recommandation, une application clé de l'apprentissage automatique qui vise à prédire les préférences ou les intérêts d'un utilisateur pour des éléments tels que des produits, des films, des articles de presse, etc. En particulier, la factorisation matricielle est largement utilisée dans les systèmes de recommandation collaboratifs, qui exploitent les informations sur les interactions passées entre utilisateurs et éléments pour générer des recommandations personnalisées.

La factorisation matricielle contribue aux systèmes de recommandation comme suite :

- ***Modélisation des Interactions Utilisateur-Élément :*** Dans un système de recommandation, les interactions passées entre utilisateurs et éléments sont souvent représentées sous forme de matrice, où chaque ligne représente un utilisateur, chaque colonne représente un élément, et les valeurs de la matrice indiquent le degré d'interaction (par exemple, une note attribuée par un utilisateur à un élément). La factorisation matricielle permet de décomposer cette matrice en deux matrices de facteurs, une pour les utilisateurs et une pour les éléments, où les produits scalaires entre les vecteurs de facteurs des utilisateurs et des éléments peuvent être utilisés pour estimer les interactions manquantes.
- ***Réduction de Dimensionnalité :*** La factorisation matricielle permet de réduire la dimensionnalité de la matrice d'interactions utilisateur-élément en approchant celle-ci par un produit matriciel de dimensions réduites. Cela permet de modéliser de manière plus efficace les tendances et les modèles sous-jacents dans les données, tout en réduisant la complexité du modèle et en améliorant sa capacité à généraliser à de nouveaux utilisateurs et éléments.
- ***Personnalisation des Recommandations :*** En utilisant les matrices de facteurs obtenues par la factorisation, les systèmes de recommandation peuvent générer des recommandations personnalisées pour chaque utilisateur en calculant les éléments les plus pertinents pour cet utilisateur spécifique. Cela permet de fournir des recommandations plus précises et pertinentes, en tenant compte des préférences individuelles de chaque utilisateur.
- ***Surpassement du Problème de la Froideur de l'Article et du Nouvel Utilisateur:*** La factorisation matricielle permet de surmonter le problème de la "froideur" de l'article (lorsqu'un nouvel élément n'a pas encore reçu suffisamment d'interactions pour être recommandé de manière fiable) et le problème du nouvel utilisateur (lorsqu'un nouvel

utilisateur n'a pas encore fourni suffisamment d'interactions pour que des recommandations personnalisées puissent être générées), en exploitant les informations agrégées des autres utilisateurs et éléments.

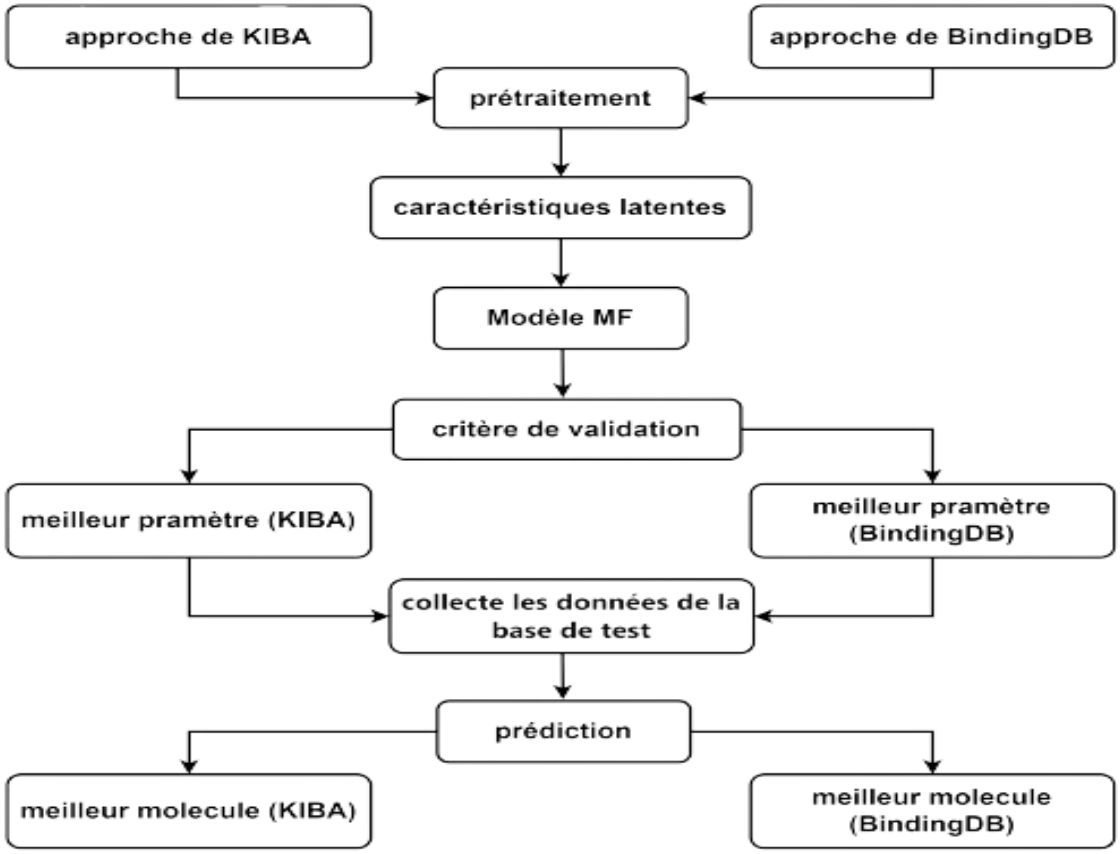
Chapitre V :
MATERIEL ET METHODES

V. Matériel et méthodes

Les modèles d'apprentissage automatique capables de prédire avec précision l'affinité jouent un rôle crucial dans la découverte des médicaments. Comme nous l'avons décrit dans le chapitre précédent, les modèles de factorisation matricielle sont particulièrement adaptés à cette étude. Dans ce chapitre, nous allons présenter l'architecture globale de notre système de factorisation matricielle, depuis la collecte des données jusqu'à la prédiction de la meilleure molécule. Ensuite, nous décrirons en détail chaque étape de notre processus.

V.1. Méthodologie de calcul

La méthodologie d'évaluation de notre modèle, adoptée pour trouver l'affinité de liaison de la meilleure molécule est schématisée dans l'organigramme ci-dessous :



***Organigramme V-1**:Protocole général d'établissement du modèle*

V.2. Matériel utilisé

V.2.1.Micro-ordinateur

Dans cette étude, nous avons utilisé un micro-ordinateur ayant un Processeur AMD Ryzen 5 4500U with Radeon Graphics et une mémoire de ram installé 8,00 Go (7,37 Go utilisable), sous le système d'exploitation Windows 11 Pro 64-bit version 23H2.

V.2.2.Les outils et logiciels

V.2.2.1. Python

Python est un langage de programmation informatique en haut niveau, développé par Guido van Rossum, Python a été introduit pour la première fois en 1991. Il est disponible en open source et gratuitement, et est utilisé pour créer divers programmes. (Lynch, 2023)

Notre travail a été développé à l'aide de Google Colab, c'est une plateforme de développement Python basée sur le cloud qui offre un accès gratuit aux ressources informatiques. Le code est implémenté en utilisant Python 3.10.12 et des bibliothèques supplémentaires chacune a son rôle :

- ✓ RDKit : Manipuler et visualiser des structures moléculaires.
- ✓ PyTDC : prédiction multi-cibles en chimie et pharmacologie.
- ✓ Numpy: effectuer des calculs numériques
- ✓ Pandas: manipuler des données.
- ✓ Tqdm : afficher des barres de progression lors de l'exécution de boucles.
- ✓ Matplotlib: la création de visualisations graphiques.
- ✓ PyTorch: construire et entraîner des modèles d'apprentissage en profondeur.

V.2.2.2. PubChem

PubChem (https://pubchem.ncbi.nlm.nih.gov/) est une base de données ouverte sur les composés chimiques et leurs propriétés biologiques. Cette base contient trois bases de données sont les suivantes : Substance, Composé et BioAssay. Les données chimiques déposées par les contributeurs de données individuels à PubChem sont stockées dans la base de données Substance, tandis que la base de données Composés stocke des structures chimiques uniques extraites de la base de données Substance. La base de données BioAssay contient les informations sur l'activité biologique des substances chimiques analysées lors d'expériences d'analyse.[40]

V.2.2.3. Chem sketch

Chem Sketch ACD Lab, est un logiciel disponible en versions gratuite et payante, particulièrement efficace pour l'apprentissage de la chimie, notamment pour les matières nécessitant des dessins ou des schémas. grâce à des outils facilement accessibles sur la barre de travail, Ce logiciel permet de dessiner des structures chimiques en 2D ou 3D, des réactions et des schémas chimiques, qui permet également de nommer des petites molécules, de calculer des propriétés moléculaires, de préparer des rapports professionnels et des présentations, et de rechercher dans des bases de données ACD Lab[41]. Voici le logo du logiciel dans la figure ci-dessous :

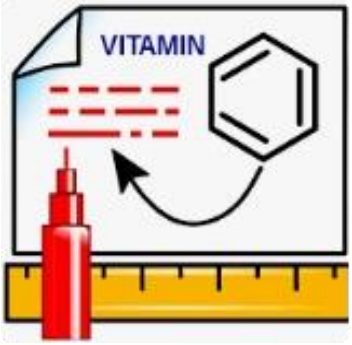


Figure V-1 : *logo de Chem sketch*

V.2.2.4. Data Warrior

La plateforme Data Warrior allie la visualisation graphique interactive et le filtrage de données à l'aide d'une intelligence chimique. Différentes caractéristiques moléculaires sont codées par les descripteurs chimiques, comme la structure et les propriétés pharmacophores. (https://openmolecules.org/datawarrior/). Voici le logo ainsi qu'une capture d'écran du logiciel dans les figures suivantes :



Figure V-2 : *logo de Data Warrior*

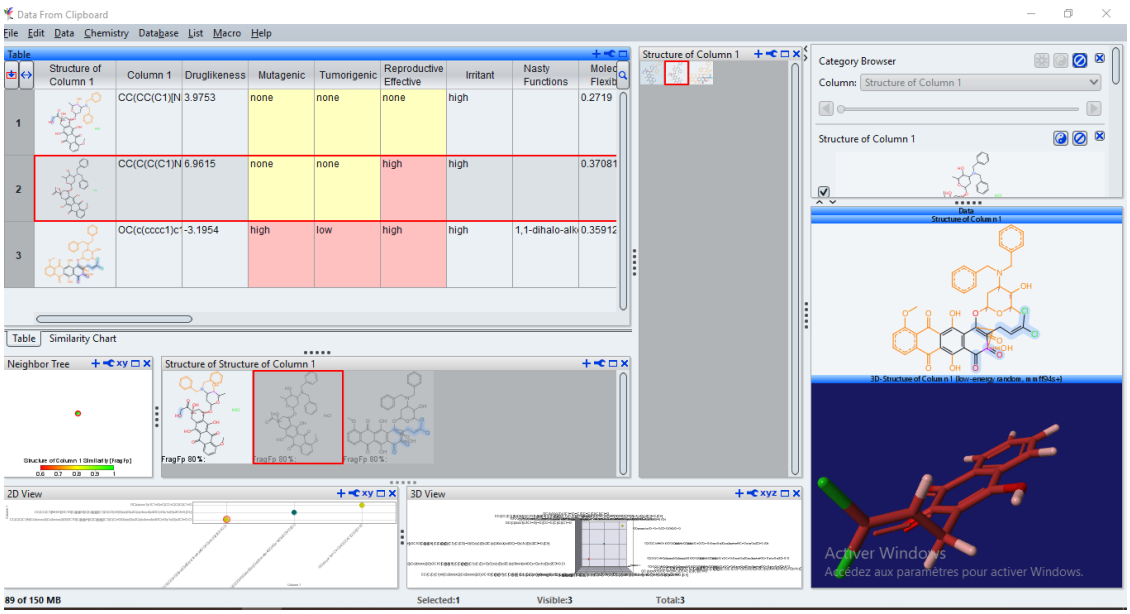


Figure V-3: *Capture d'écran de Data Warrior*

V.3. Préparation de la base de données

Deux bases de données d'entraînement ont été fournies par le Thérapeutique Data Commons (TDC). L'une est BindingDB et l'autre est KIBA. (https://tdcommons.ai/multi_pred_tasks/dti/)

V.3.1.BindingDB

Cette base regroupe les affinités de liaison mesurées, principalement axées sur les interactions entre des protéines identifiées comme cibles médicamenteuses et de petites molécules présentant des similitudes avec des médicaments. Les données de BindingDB sont extraites des articles scientifiques, de plus en plus, de brevets américains appropriés.

BindingDB propose des outils spécialisés exploitant sa base de données étendue pour générer des hypothèses sur les cibles protéiques associées à un composé bioactif, ainsi que sur les composés liés à une nouvelle protéine de séquence connue. En outre, il propose des méthodes de criblage de composés virtuels basées sur des critères tels que la similarité chimique maximale, ce qui facilite la découverte et l'analyse des interactions moléculaires, contribuant ainsi à la recherche et au développement de nouveaux médicaments.[42]

Actuellement, cette base de données compte environ 20 000 enregistrements de liaison, comprenant environ 11 000 ligands de petites molécules et 110 cibles médicamenteuses distinctes[43].

Dans BindingDB, il existe trois ensembles de données : Kd, Ic50 et Ki. Comme Kd est une mesure importante dans l'étude des interactions médicament-protéine, nous avons chargé

l'ensemble de données Kd, lequel contient 52 284 paires DTI, 10 656 médicaments et 1 413 protéines.

V.3.2.KIBA

C'est une évaluation méthodique des profils de sélectivité des cibles dans trois récents tests biochimiques à grande échelle d'inhibiteurs de kinases. Elle contient des informations supplémentaires sur différents types d'activités biologiques (IC50, Ki, Kd). Cette démarche peut classer les cibles des inhibiteurs de kinases et identifier les erreurs potentielles dans les interactions médicament-cibles[44]. Cette base de données contient 117 657 paires DTI, 2 069 médicaments et 230 protéines. Le tableau suivant résume les informations concernant la base de données choisie :

Tableau V-1: *valeurs correspondantes aux différents éléments utilisés dans la base de données.*

	Nombre de médicaments	**Nombre de cibles**	**Nombre de paires DTI**
BindingDB	10656	1413	52221
KIBA	2069	230	117538

Dans le traitement du cancer, les inhibiteurs de kinases sont fréquemment employés car ils ciblent spécifiquement les protéines kinases qui jouent un rôle dans la croissance et la propagation des tumeurs.[45]. Par conséquent, l'ensemble de données KIBA est le mieux adapté à nos objectifs, car il fournit des informations spécifiques sur les interactions médicament-cible liées au cancer. En comparaison, la base de données BindingDB est la plus générale en raison de sa grande taille et de sa richesse en informations sur divers contextes biologiques et chimiques ainsi que sur différentes classes thérapeutiques.

Chaque fois que l'un de ces deux ensembles soit utilisé, celui-ci est divisé en sous-ensembles un ensemble de données de formation et un ensemble de données de test à l'aide de la fonction de répartition aléatoire disponible dans la bibliothèque TDC. Après la division des informations et la détermination du rôle de chaque sous-ensemble, nous avons recombiné les deux ensembles en un seul ensemble de données.

V.4. Prétraitement de la base

La méthode de factorisation matricielle que nous utilisons nécessitera un index distinct pour chaque médicament et chaque cible. Ainsi, la première étape du prétraitement consiste à établir des correspondances entre les identifiants et les noms des médicaments et des cibles.

Pour ce faire, les noms des médicaments et des cibles sont remplacés par leurs identifiants respectifs, dans notre modèle, chaque médicament et chaque cible unique est associé à un numéro, commençant à zéro. Ce dictionnaire de correspondance est ensuite appliqué à l'ensemble d'apprentissage, de sorte que chaque médicament et chaque cible dans l'ensemble de données soit représenté par son numéro d'identification. Ensuite, les données sont préparées pour être utilisées dans un Data Loader PyTorch, et les données préparées ainsi que les étiquettes correspondantes sont renvoyées. Cette étape simplifiera le traitement ultérieur des données et l'indexation pendant l'entraînement du modèle.

L'étape suivante, consiste à charger les données d'entraînement et de test, à les préparer à l'aide de la fonction « data Loader », puis à stocker les caractéristiques et les étiquettes dans des variables correspondantes qui seront utilisées pour entraîner et évaluer le modèle. Les ensembles de données d'entraînement et de test sont désormais configurés sous forme de tableaux (N, 2), où N est le nombre d'entrées dans l'ensemble de données. La première colonne représente les médicaments et la deuxième colonne représente les cibles.

La dernière étape du prétraitement consiste à utiliser la fonction PyTorch Data Loader afin de préparer les données à transmettre au modèle PyTorch. Cette démarche implique la création d'objets Data Loader contenant les ensembles de données d'entraînement et de test, prêts à être utilisés pour l'entraînement et l'évaluation du modèle. Les données sont converties en tenseurs PyTorch puis encapsulées dans ces objets Data Loader, simplifiant ainsi le traitement et l'entraînement des modèles avec PyTorch

V.5. Factorisation matricielle

Le modèle de factorisation matricielle de base est largement utilisé pour prédire l'affinité de liaison entre les médicaments et leurs cibles protéiques. Il a été démontré que ces méthodes surpassent d'autres techniques d'apprentissage automatique. Ses performances supérieures en font une approche de choix dans ce domaine.[46]

Dans cette étude, nous avons utilisé un code de factorisation matricielle d'une forme de filtrage collaboratif à l'aide de la fonction « PyTorch » avec biais, qui contient une technique essentielle pour prédire les interactions médicament-cible (DTI). Cette technique permet de projeter à la fois les médicaments et les cibles dans un espace de facteurs latents. Dans cet espace, les interactions sont modélisées comme des produits scalaires entre les vecteurs de facteurs latents des médicaments et des cibles. Pour la prédiction des interactions (DTI), la factorisation matricielle décompose la matrice des interactions observées en deux matrices de

rang réduit : l'une contenant les vecteurs de facteurs latents représentant les propriétés des médicaments, et l'autre correspondant aux vecteurs de facteurs latents des cibles protéiques[47]Ceci est illustré à lafigure suivante[48] :

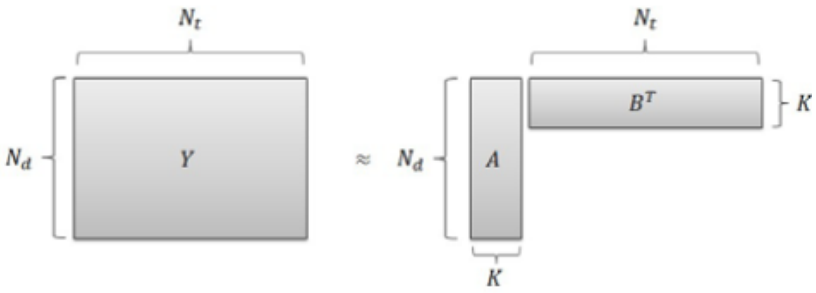


Figure V-4 : *méthode de factorisation matricielle*

Soit :

- A = {d1, d2, …, dNd} un ensemble donné de médicaments, où A est une matrice de dimensions (Nd, f), où Nd est le nombre de médicaments uniques et f est le nombre de facteurs latents
- B = {t1, t2, …, tNt} un ensemble donné de cibles, où B est une matrice de dimensions (Nt, f), où Nt est le nombre de cibles uniques.

La multiplication matricielle donne alors Y qui a des dimensions (Nd, Nt) où chaque valeur contient l'affinité de liaison prédite du médicament et de la cible respectifs.

V.6. Les optimiseurs (SGD, Adam, RMSprop)

V.6.1. SGD

SGD optimise en utilisant des échantillons aléatoires (stochastiques). Au lieu de calculer le gradient en utilisant l'ensemble de données complet, cet optimiseur sélectionne quelques échantillons au hasard dans l'ensemble de données pour chaque itération. Il calcule alors le gradient du paramètre en utilisant ces échantillons en se basant sur un seul exemple d'entraînement ou un petit sous-ensemble d'exemples d'entraînement, ce qui permet une mise à jour des paramètres plus fréquente et plus rapide. L'équation V.1 montre la mise à jour de chaque exemple d'entrainement.

$$W = w - \eta \nabla Q_i(w) \qquad (V.1)$$

Où :

- W : Les matrices A et B mises à jour après une itération
- w : Les matrices A et B avant la mise à jour.
- ŋ : Le taux d'apprentissage, qui est l'un des hyperparamètres du modèle

- Q_i : La fonction de perte pour l'élément (i, j) de la matrice de notation généralement définie comme l'erreur quadratique moyenne (MSE) entre les prédictions du modèle et les valeurs réelles.

V.6.2. RMSprop

RMSprop est un optimiseur similaire à l'algorithme de descente de gradient avec élan. Il a pour avantage de limiter les oscillations brusques des paramètres au cours de l'apprentissage. Cela permet d'augmenter le taux d'apprentissage, ce qui aide l'algorithme à faire des progrès significatifs et à converger plus rapidement vers une solution optimale. La principale différence entre RMSprop et la descente de gradient réside dans la méthode de calcul des gradients. La moyenne mobile en termes de carré moyen est calculée comme indiqué dans l'équation suivante :

$$\upsilon(w,t) = \gamma\, \upsilon\,(w,t-1) + (1-\gamma)\,(\nabla Q_i(w))^2 \qquad (V.2)$$

L'équation (V.3) montre la mise à jour des paramètres :

$$W = w - \frac{\eta}{\sqrt{\upsilon(w,t)}}\nabla Q_i(w) \qquad (V.3)$$

Où les paramètres W, w, ŋ et Qi sont les mêmes que dans l'équation 1, et les autres paramètres sont les suivants :

- $\upsilon(w,t)$: La moyenne mobile en termes de carré moyen des matrices A et B à l'itération t.
- γ : Le facteur d'oubli dans RMSprop
- t : L'itération actuelle de l'algorithme.

V.6.3. ADAM

ADAM est un algorithme d'optimisation, comme alternative à la descente de gradient stochastique classique (SGD) pour mettre à jour les poids du réseau pendant l'entraînement. ADAM (Adaptive Moment Estimation) combine les avantages de deux autres algorithmes d'optimisation : Adagrad et l'optimiseur Momentum. Il ajuste les taux d'apprentissage individuellement pour chaque paramètre et utilise des moyennes mobiles des gradients et des carrés des gradients pour stabiliser les mises à jour. En effet, pour chaque itération de l'algorithme ADAM, il y a deux estimations :

- La moyenne mobile exponentielle du premier moment (moyenne mobile des gradients) m_w^{t+1} après l'itération t+1 :

$$m_w^{t+1} = \beta_1 m_w^t + (1-\beta_1)\nabla_w L^t \qquad (V.4)$$

- La moyenne mobile exponentielle du second moment (moyenne mobile des carrés des gradients) υ_w^{t+1} après l'itération t+1 :

$$\upsilon_w^{t+1} = \beta_2 \upsilon_w^t + (1-\beta_2)(\nabla_w L^t)^2 \quad \text{(V.5)}$$

Le biais est corrigé dans le premier et deuxième estimation :

$$\hat{m}_w = \frac{m_w^{t+1}}{1-\beta_1^{t+1}} \quad \text{(V.6)}$$

$$\hat{\upsilon}_w = \frac{\upsilon_w^{t+1}}{1-\beta_2^{t+1}} \quad \text{(V.6)}$$

À l'aide des moments ajustés, les paramètres de prédiction mis à jour sont calculés et appliqués :

$$w^{t+1} = w^t - \eta \frac{\hat{m}_w}{\epsilon + \sqrt{\hat{\upsilon}_w}} \quad \text{(V.7)}$$

Où :

- β_1, β_2 : Les facteurs d'oubli pour le premier moment et le second moment, respectivement.
- t : L'itération d'entraînement en cours
- $\nabla_w L^t$: Gradient de la fonction de perte par rapport aux paramètres à l'itération t .
- w^t : Les matrices A et B avant la mise à jour.
- ϵ : Un petit scalaire utilisé pour empêcher la division par 0[49].

V.7. Critère de validation

Le processus de validation statistique permet de déterminer la fiabilité et la pertinence du modèle élaboré pour la cible précise.[50]

La capacité prédictive du modèle de factorisation matricielle a été évaluée en analysant les courbes de perte quadratique moyenne (MSE) sur les ensembles de données d'entraînement et de test à chaque époque, et en utilisant quatre indices statistiques. En général, ces mesures incluent MAE, MSE, RMSE et MAPE.

L'erreur absolue moyenne **MAE** (Mean Absolute Error) est une mesure statistique basée sur la moyenne des écarts entre chaque prédiction de note et sa valeur réelle.[51]

L'erreur quadratique moyenne **MSE** (Mean Squared Error) est une mesure statistique basée sur la moyenne des carrés des différences entre les valeurs prédites et les valeurs réelles, elle la variabilité des prévisions les erreurs. Le MAE et le MSE sont tous des mesures de prévision qui dépendent de l'échelle de précision, c'est-à-dire que leurs valeurs sont calculées en fonction de l'originale unité de mesure.

La racine carrée de l'erreur quadratique moyenne **RMSE** (Root Mean Squared Error) mesure la moyenne de l'erreur de mesure. Les différences entre les valeurs prédites et les valeurs observées correspondantes sont mises au carré individuellement puis moyen sur l'échantillon. Enfin, la racine carrée de la moyenne est prise. Étant donné que les erreurs sont mises au carré avant d'être moyennes, le RMSE accorde un poids relativement élevé aux erreurs importantes. Cela signifie que RMSE est particulièrement utile lorsque des erreurs importantes sont particulièrement indésirables.

La racine carrée de l'erreur quadratique moyenne **MAPE** (Mean Absolute Percentage Error) est l'indicateur statistique le plus objectif car il mesure des pourcentages relatifs et n'est pas affecté par les unités de la séquence de prévision. Plus le MAPE est proche de zéro, meilleure est la prédiction. [52]

Les critères de validation statistique sont détaillés dans le tableau V-3

Tableau V-2: *critères de validation*

Métrique	Nom	Formule
MAE	Erreur absolue moyenne	$\frac{1}{n}\sum_{i=1}^{n} /Y_{cal} - Y/$
MSE	Erreur quadratique moyenne	$\frac{1}{n}\sum_{i=1}^{n} (Y - Y_{cal})^2$
RMSE	Racine carrée de l'erreur quadratique moyenne	$\sqrt{\frac{1}{n}\sum_{i=1}^{n} (Y_{cal} - Y)^2}$
MAPE	Racine carrée de l'erreur quadratique moyenne	$\frac{1}{n}\sum_{i=1}^{n} / \frac{Y - Y_{cal}}{Y} /* 100\%$

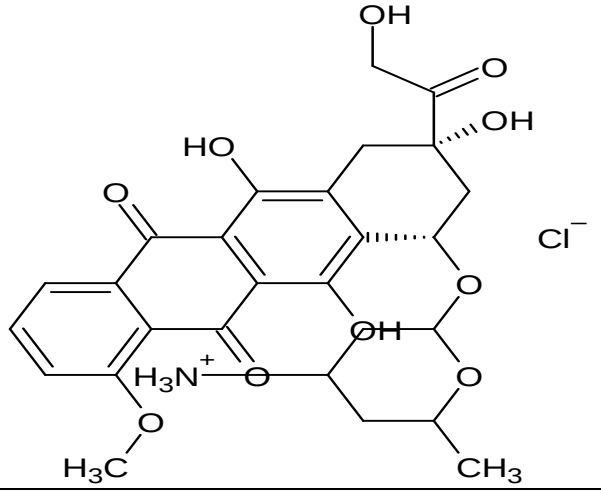
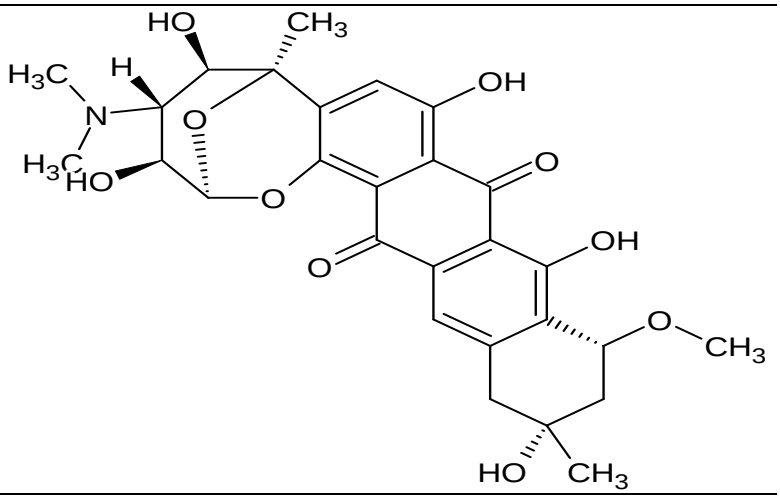
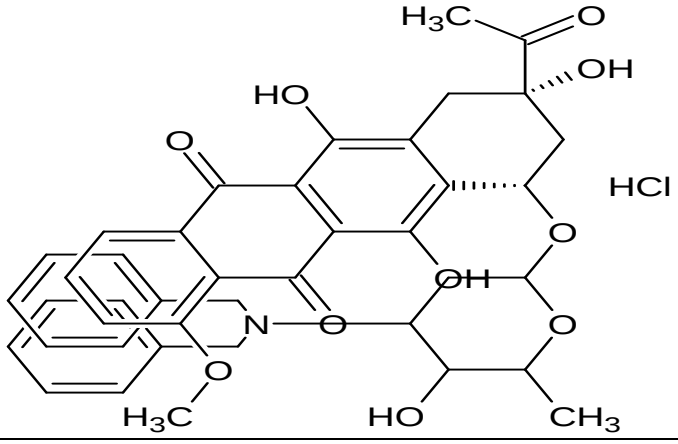
Où **Y** et $\mathbf{Y_{cal}}$ sont respectivement la sortie réelle et la sortie attendue (la valeur prédite par le modèle), et **n** est le nombre de données.

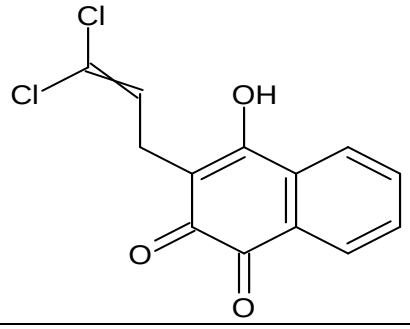
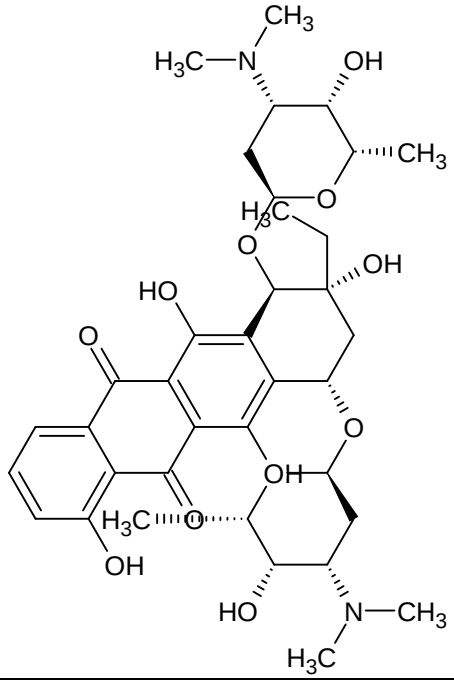
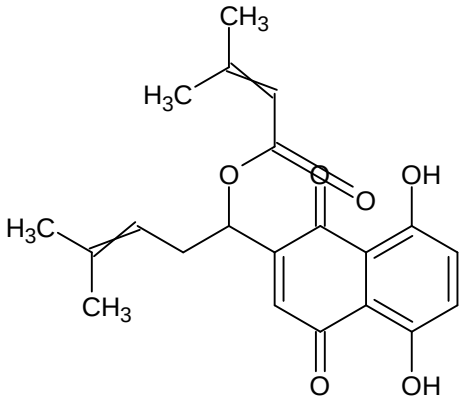
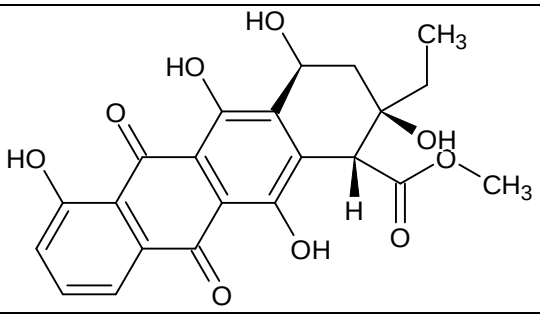
V.8. Test externe

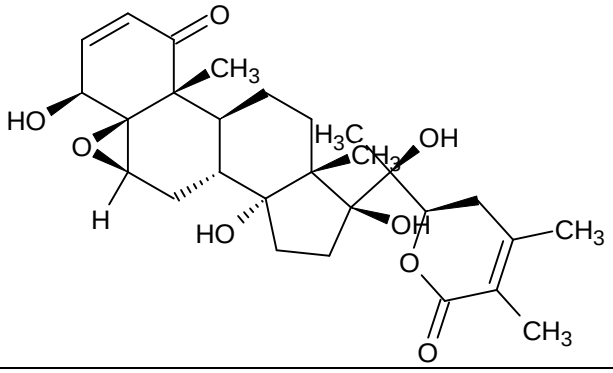
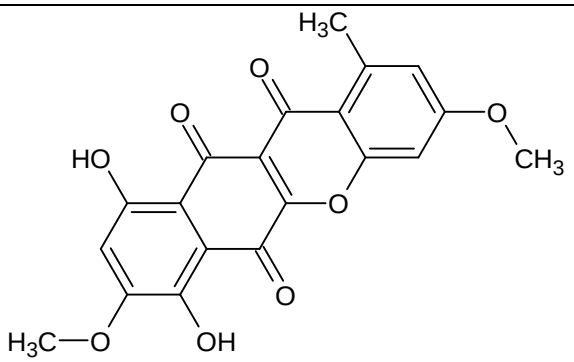
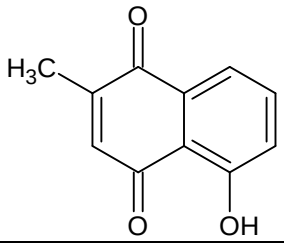
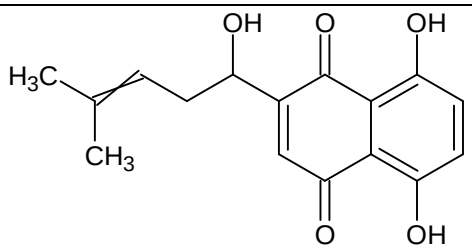
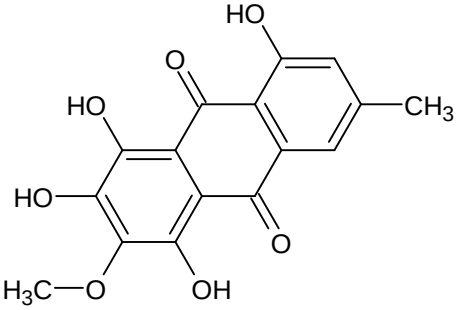
V.8.1. Collecte de la base de données

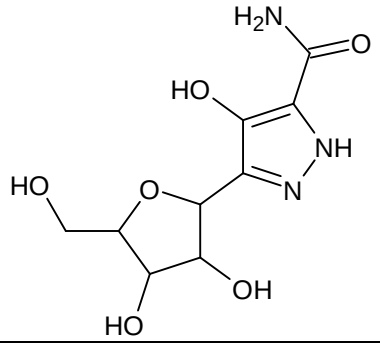
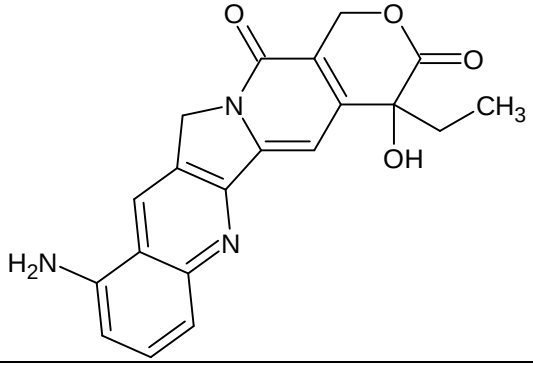
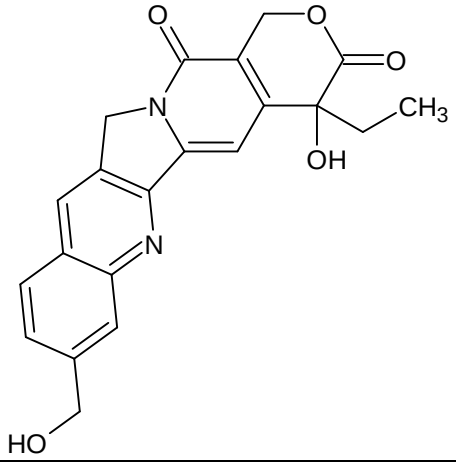
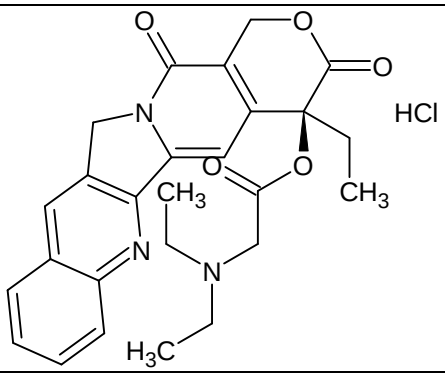
La base de données du National Cancer Institute (NCI) a été utilisée pour obtenir soixante-onze ensembles de composés et leurs activités contre les cellules de mélanome SK-MEL-2, afin d'identifier de nouveaux composés potentiellement efficaces contre le cancer, en particulier contre la protéine cible V600E-BRAF, responsable du cancer du mélanome[53]. Les données ont été recueillies en utilisant PubChem. On a effectué une analyse et un traitement de ces données en utilisant le logiciel Data Warrior. On a également supprimé les molécules doubles.

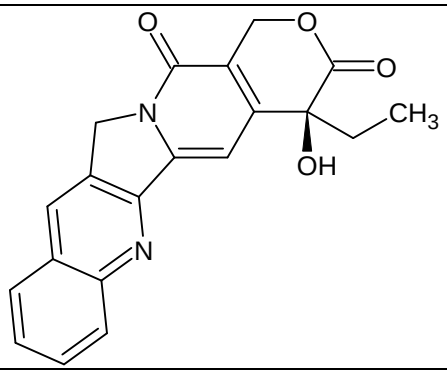
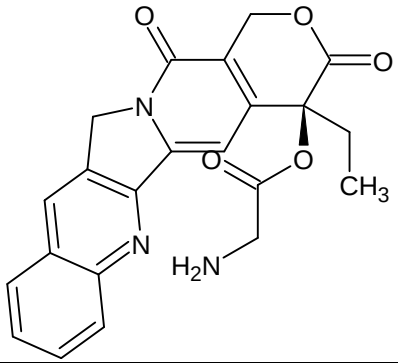
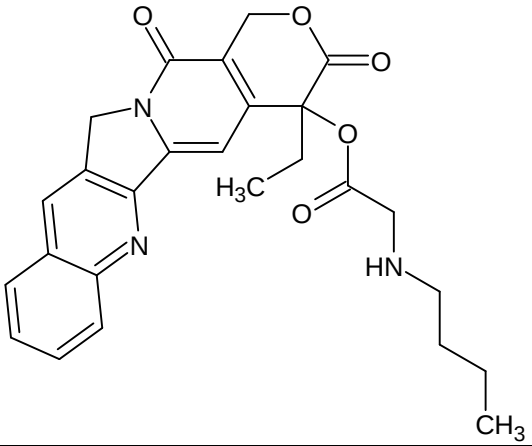
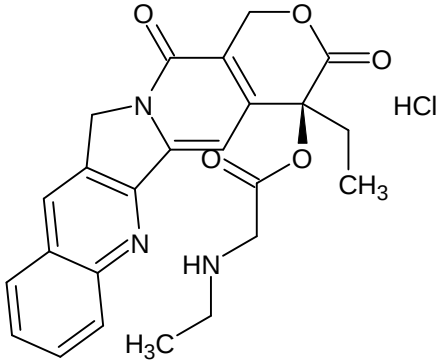
Tableau V-3:*les données de la base de test*

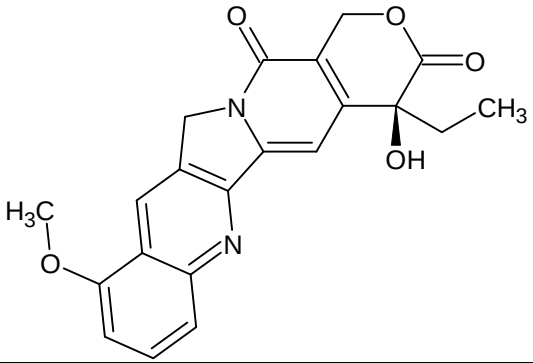
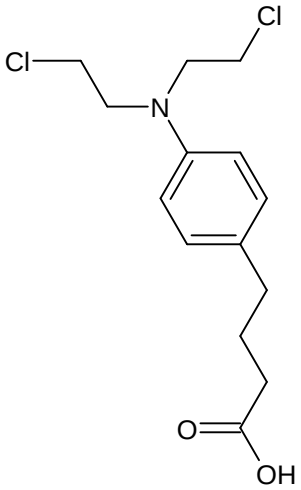
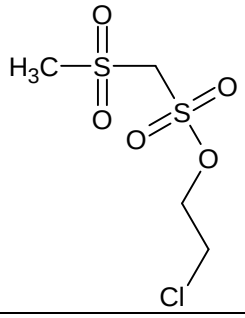
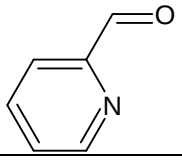
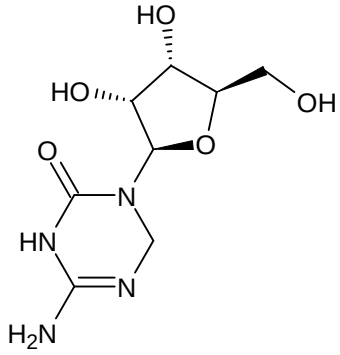
N°	*Nom Chimique*	*Structure Semi-Développée*
1	Deoxydoxorubicin	
2	Menogaril	
3	N,N-Dibenzyldaunorubicin Hydrochloride	

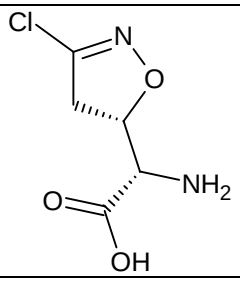
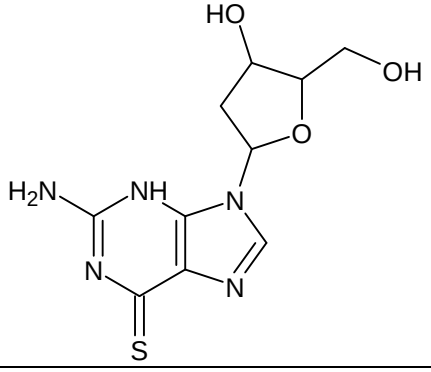
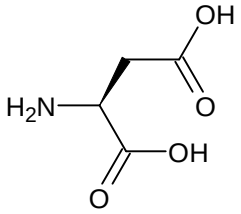
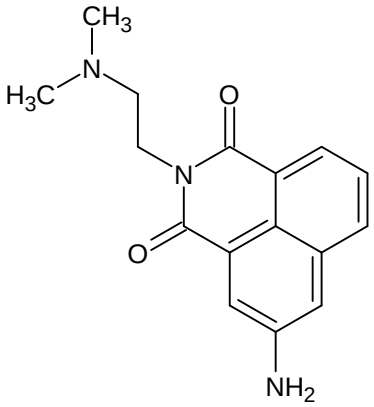
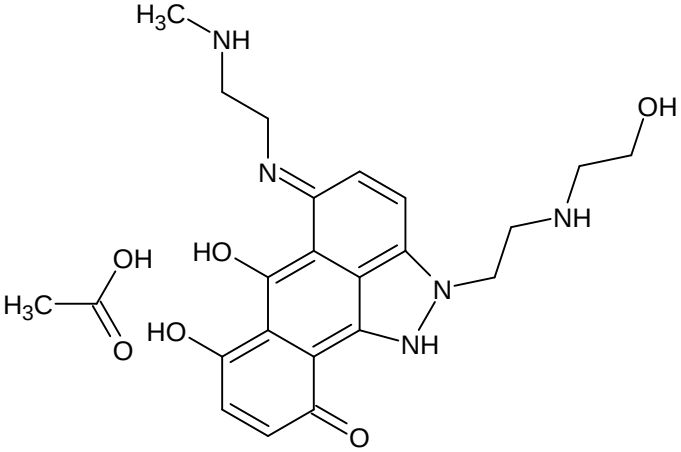
4	Dichloroallyl Lawson	
5	Rhodomycin A	
6	Arnebin 1	
7	Epsilon.-Rhodomycinone	

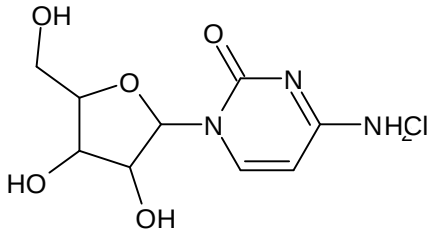
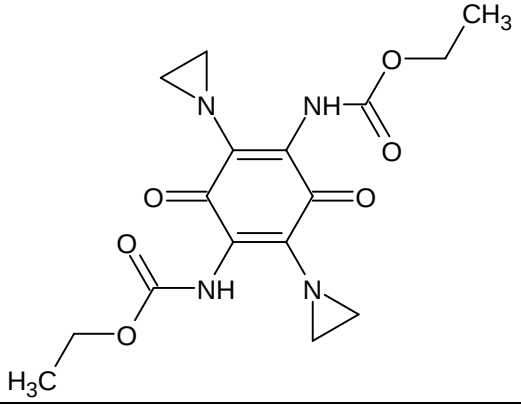
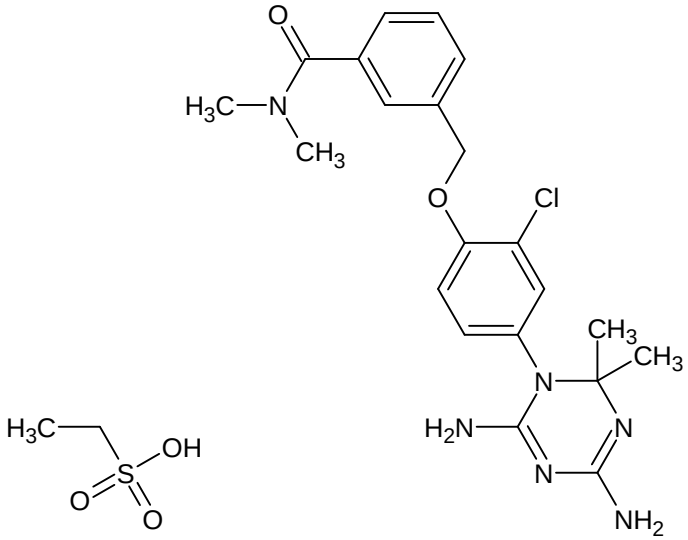
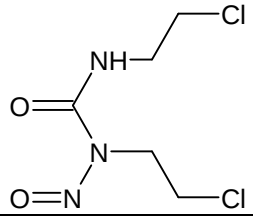
8	4BETA-Hydroxywithanolide	
9	Bikaverin	
10	Plumbagin	
11	Shikalkin	
12	Dermocybin	

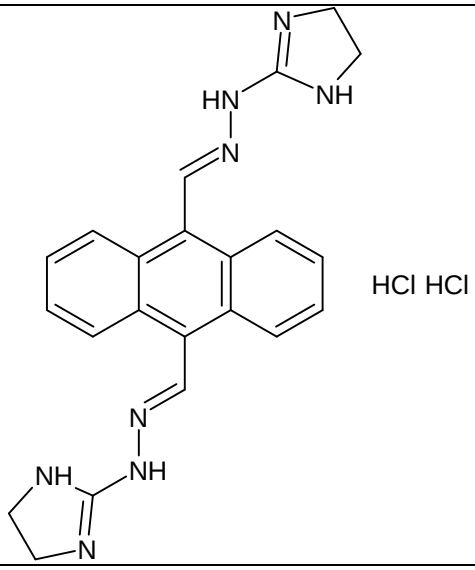
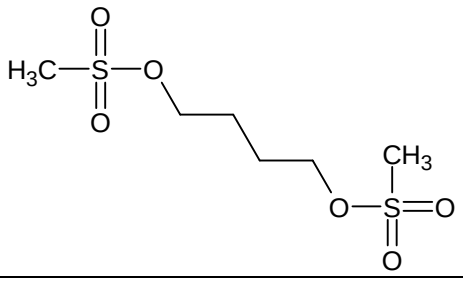
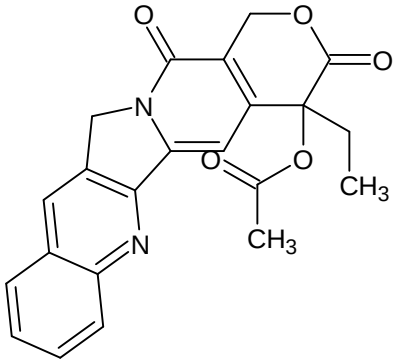
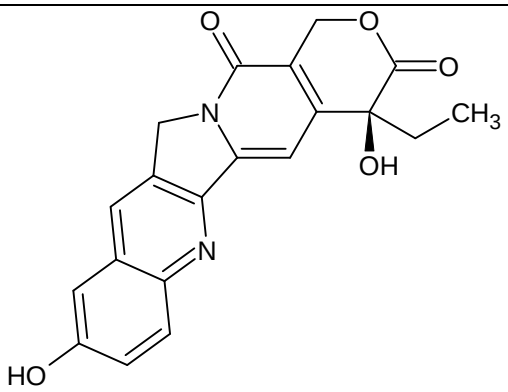
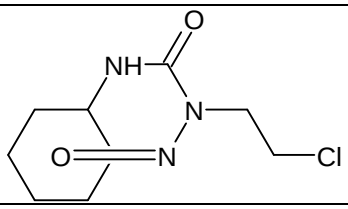
13	Pyrozofurin	
14	9-Aminocamptothecin (R,S)	
15	11-Hydroxymethyl-20(Rs)-Camptothecin	
16	Camptothecin,(N-Diethyl) Glycinate	

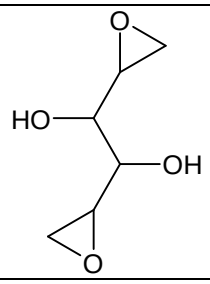
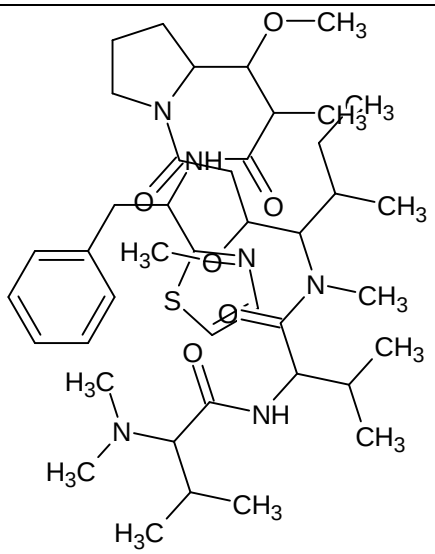
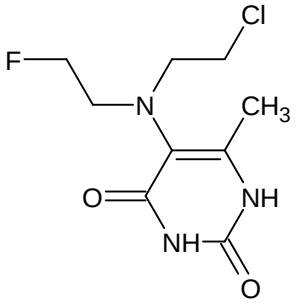
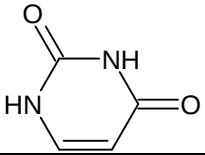
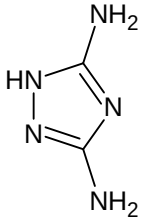
17	Camptothecin	
18	Campothecin Analog	
19	Camptothecin Butylglycinate Ester Hydrochloride	
20	Camptothecinethylglycinate Esterhydrochloride	

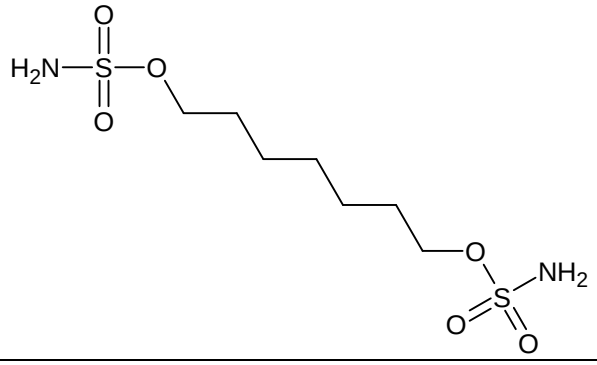
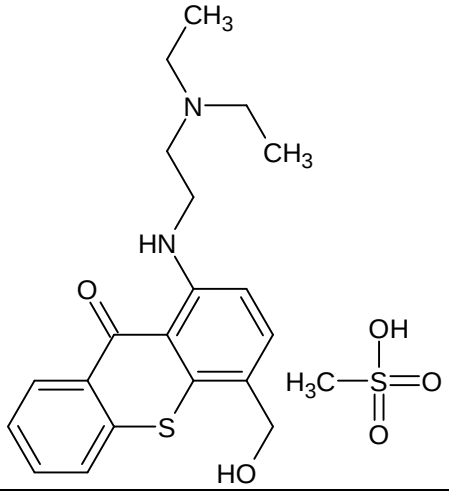
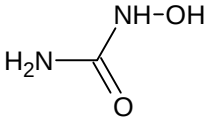
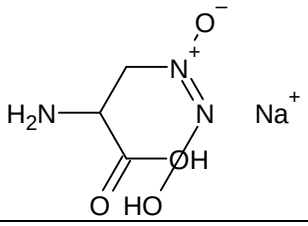
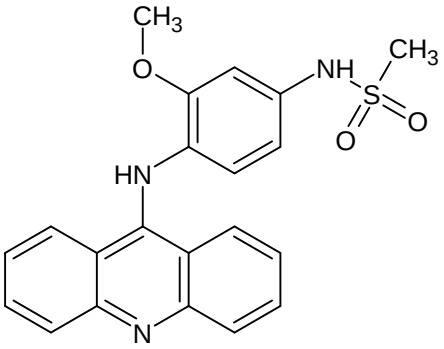
21	9-Methoxycamptothecin	
22	Chlorambucil	
23	Clomesone	
24	Picolinaldehyde	
25	Dihydro-5-Azacytidine	

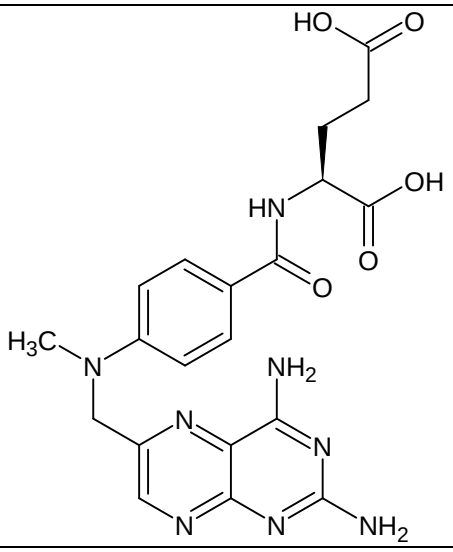
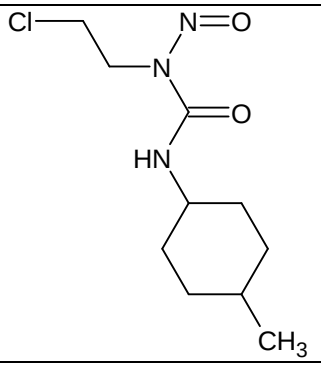
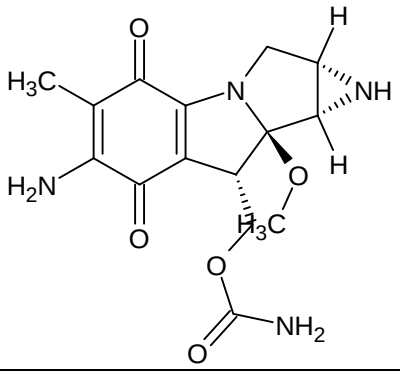
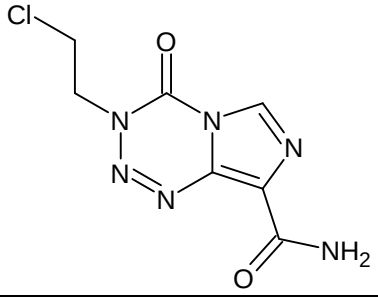
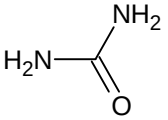
26	Acivicin	
27	Alpha.-Thiodeoxyguanosine	
28	L-Aspartic Acid	
29	Amonafide	
30	Anthra[1,9-Cd]Pyrazol-6(2H)-One Der	

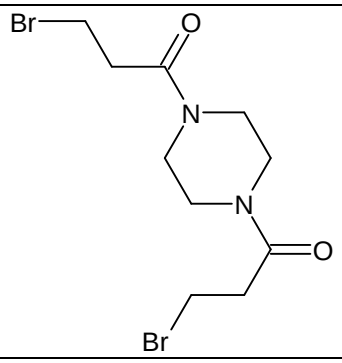
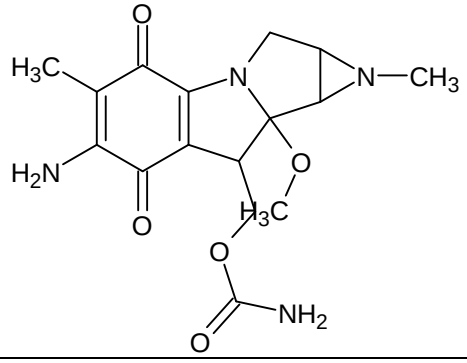
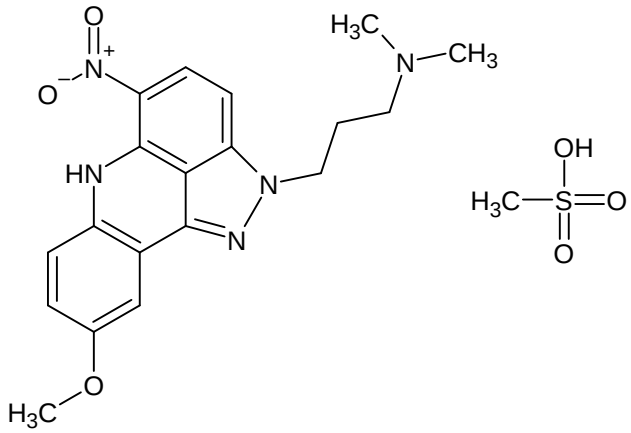
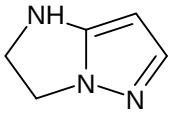
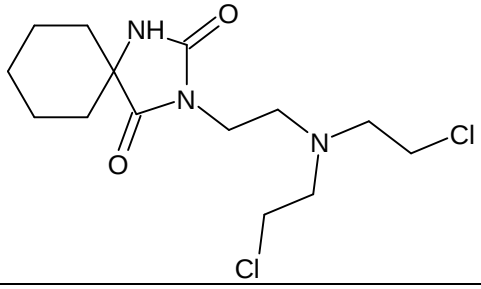
31	Cytosine, Monohydrochloride	
32	Diaziquone	
33	Triazinate	
34	Carmustine	

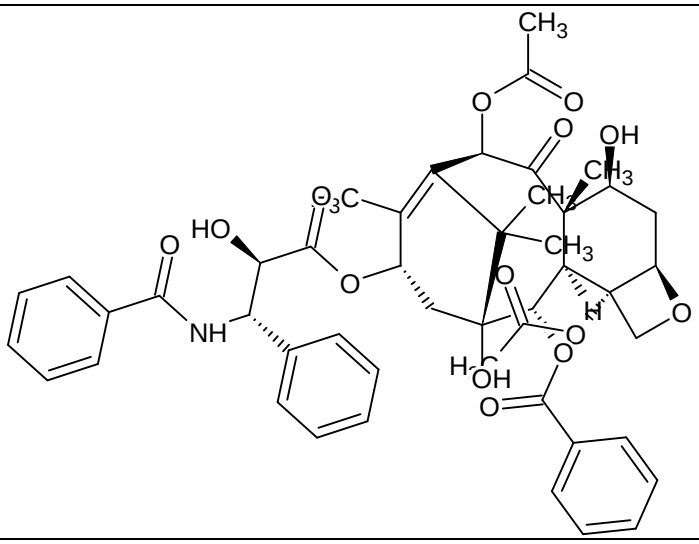
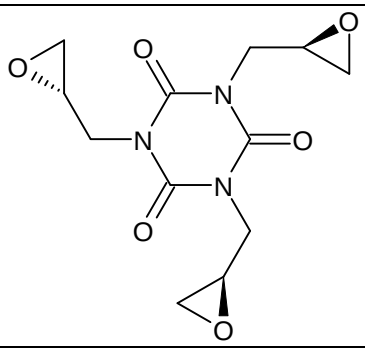
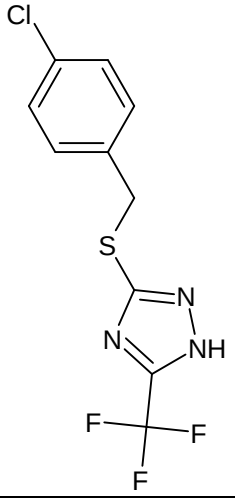
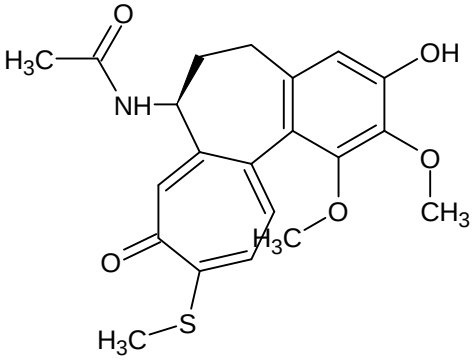
35	Bisantrene Hydrochloride	 HCl HCl
36	Busulfan	
37	Camptothecin, Acetate	
38	10-Hydroxycamptothecin	
39	Lomustine	

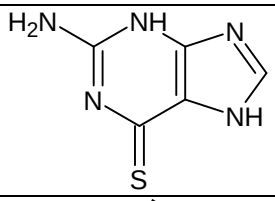
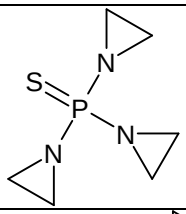
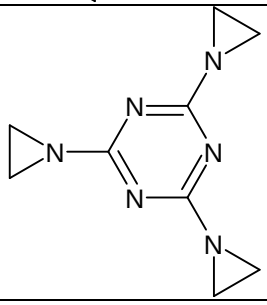
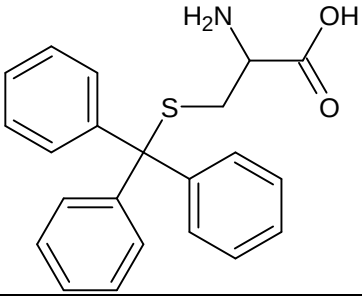
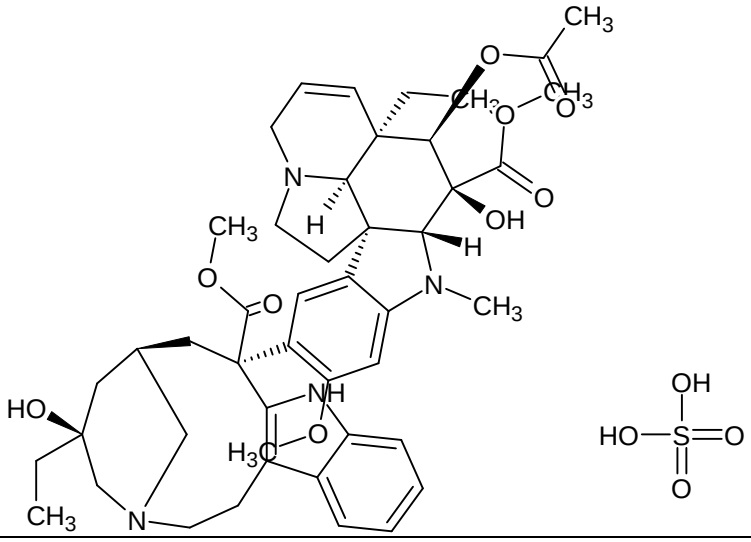
40	Dianhydrodulcitol	
41	Ac1l2oas	
42	Fluorodopan	
43	Uracil	
44	Guanazole	

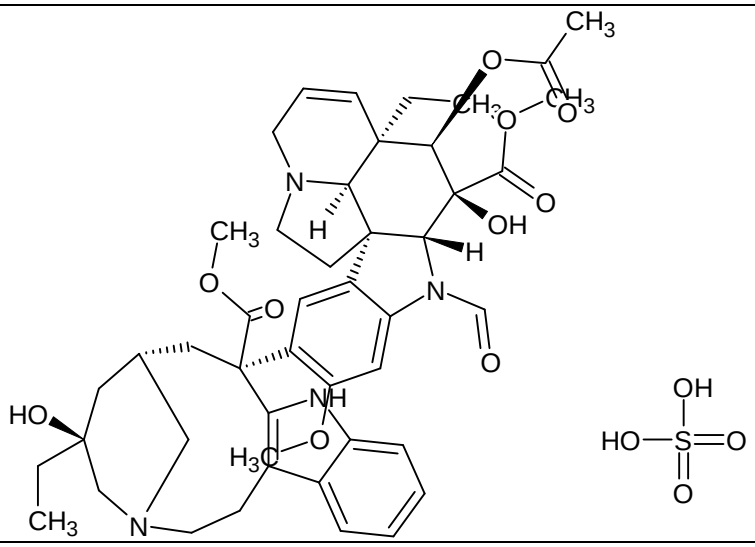
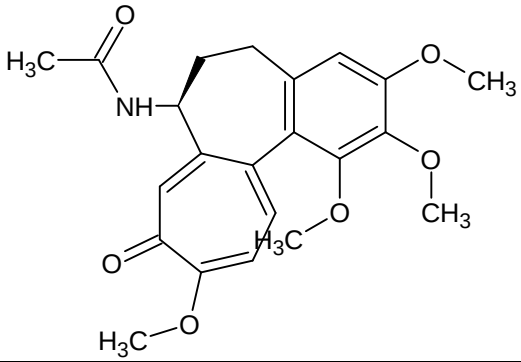
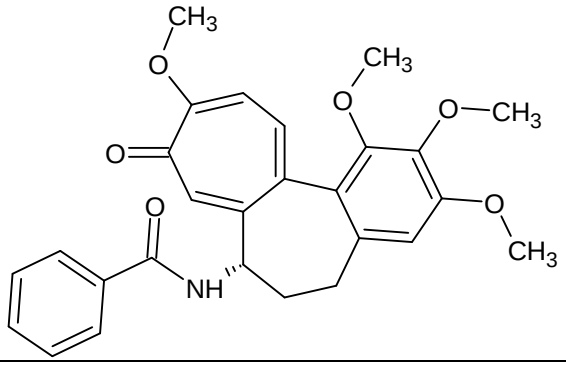
45	Hepsulfam	
46	Hycanthone Mesylate	
47	Hydroxyurea	
48	Alanosine Monosodium Salt	
49	Amsacrine	

50	Methotrexate	
51	Semustine	
52	Mitomycin C	
53	Mitozolomide	
54	Urea	

55	Pipobroman	
56	Profiromycin	
57	Pyrazoloacridine Mesylate	
58	Pyrazoloimidazole	
59	Spiromustine	

60	Paclitaxel	
61	Teroxirone	
62	5-((4-Chlorobenzyl)Thio)-3-(Trifluoromethyl)-1H-1,2,4-Triazole	
63	3-Demethylthiocolchicine	

64	6-Thioguanine	
65	Thiotepa	
66	Triethylenemelamine	
67	Tritylcysteine	
68	Vinblastine Sulfate	

69	Vincristine Sulfate	
70	Colchicine	
71	N-Benzoyl-Deacetylcolchicine	

V.8.2. Pré procédure de la base

Les données de test sont préparées de la même manière que Les données d'entrainement, mais elles varient en termes de détails en fonction des caractéristiques particulières des données et des exigences du modèle, permet les principales différences entre les deux bases on trouve :

- ✓ La première base est composée de médicaments et de cibles, tandis que la seconde est uniquement composée de médicaments. Ainsi, l'indexation des médicaments diffère et est réalisée à l'aide d'un nouveau dictionnaire.
- ✓ Les différentes fonctions employées pour générer un objet DataLoader différent car il n'y a pas d'étiquettes dans cette base.

V.9. Conclusion

Dans ce chapitre, nous avons détaillé toutes les phases par lesquelles passe notre système. Nous avons présenté l'ordinateur et les logiciels utilisés, ainsi que les différentes étapes de préparation des bases de données que nous avons choisies. Nous avons également exposé les outils de développement employés, les étapes de mise en œuvre, ainsi que les interfaces de notre application.

Chapitre VI :
RESULTATS ET DISCUSSION

VI. Résultats et discussion

Sur la base d'une description de tous les éléments constitutifs et les processus impliqués dans le développement et la réalisation du présent système, les meilleures molécules dans la base de test à recommander sont basées sur celles qui ont obtenu les meilleurs scores. Pour ce faire, les critères de validation utilisés pour évaluer les performances du modèle seront présentés. Les résultats obtenus trouvés à partir des critères de validation seront discutés.

Les meilleures molécules ont été identifiées en mettant en évidence leurs caractéristiques et leurs performances présentant des affinités plus élevées.

L'étude sera terminée par une comparaison des performances obtenues entre chaque base de données d'entraînement (BindingDB/KIBA) ainsi que les résultats du modèle établi par rapport à ceux du modèle QRAR

VI.1. Optimisation du modèle

VI.1.1. Les optimiseurs

Dans ce contexte, divers optimiseurs : SGD (l'optimisateur d'origine du modèle), Adam et RMSprop ont été testés, afin de déterminer lequel était le plus performant. Les pertes (entraînement/test) ont été enregistrées en fonction des époques pour chacun. Les résultats de ce teste sont illustrés dans les figures suivantes :



***Figure VI-1**: courbe de perte BindingDB (entrainement/test) de l'optimiseur SGD*

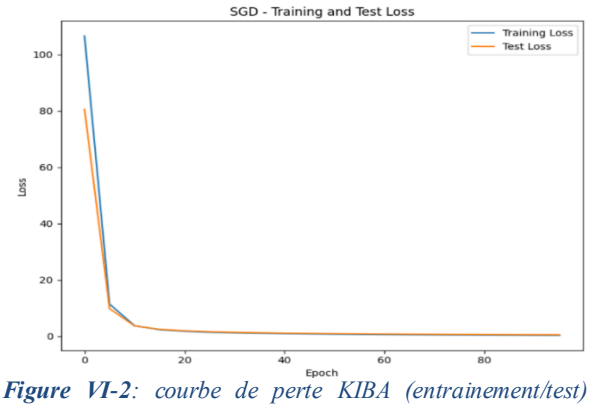


***Figure VI-2**: courbe de perte KIBA (entrainement/test) de l'optimiseur SGD*

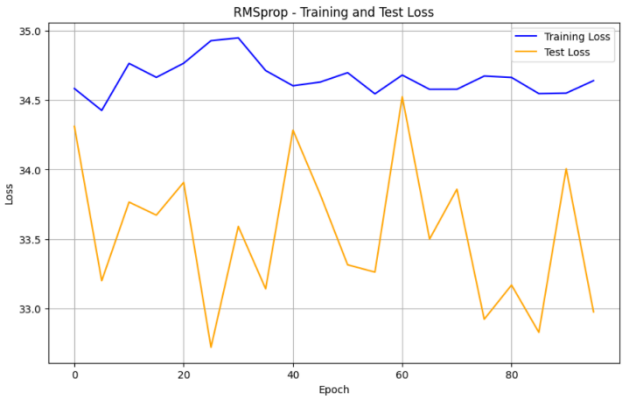


***Figure VI-3**:courbe de perte BindingDB (entrainement/test) de l'optimiseur RMSprop*

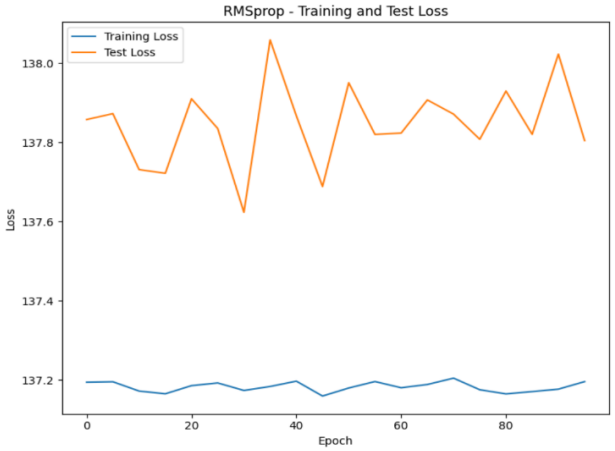


***Figure VI-4**:courbe de perte KIBA (entrainement/test) de l'optimiseur RMSprop*



***Figure VI-5**:courbe de perte BindingDB (entrainement/test) de l'optimiseur Adam*



***Figure VI-6**:courbe de perte KIBA (entrainement/test) de l'optimiseur Adam*

VI.1.1.1. SGD (Stochastic Gradient Descent) :

Les conclusions tirées à partir des deux bases :

- Le graphe montre qu'au début, il y a une diminution rapide de la perte d'entraînement et de test. Cette diminution rapide indique que le modèle a rapidement appris à partir des données d'entraînement et à généraliser sur les données de test.
- Une stabilisation à une valeur très basse a été observée. Cela suggère que le modèle a atteint une performance élevée et qu'il n'y a pas de surapprentissage significatif.
- Des trajectoires similaires La perte d'entraînement et de test, ce qui indique que le modèle se comporte de manière cohérente sur les deux ensembles
- Convergence rapide à une faible perte est visible, suggérant que SGD a bien fonctionné pour optimiser les paramètres du modèle.

VI.1.1.2. RMSprop :

Les résultats trouvés mettent en lumière des différences significatives dans les performances du modèle entre les bases de données KIBA et BindingDB.

- Les différences dans la performance entre les bases de données : La stabilité et la valeur faible de la perte d'entraînement avec la base KIBA, par rapport à la stabilité élevée de la perte d'entraînement avec la base BindingDB, indiquent des variations dans la difficulté d'apprentissage entre les deux ensembles de données. Cette disparité peut être due à des différences dans la distribution des données, la qualité des données ou d'autres facteurs spécifiques à chaque base de données.
- Surapprentissage avec la base KIBA **:** La forte variabilité et l'élévation persistante de la perte tout au long de l'entraînement et par rapport la base BindingDB on remarque des fluctuations comme celles de KIBA. Cela indique que bien que le modèle s'entraîne correctement (basse perte d'entraînement), il ne généralise pas bien sur les données de test, ce qui suggère un problème de surapprentissage (overfitting) au niveau de la base de KIBA.
- Problèmes avec la base BindingDB **:** La faible perte de test par rapport à la perte d'entraînement dans la base BindingDB, ainsi que la stabilité de la perte d'entraînement à une valeur élevée, suggèrent des problèmes d'optimisation du modèle. Cela pourrait être dû à des hyperparamètres mal ajustés, comme le taux d'apprentissage de l'algorithme Adam soulignant l'importance d'explorer différentes méthodes d'optimisation et de faire des ajustements appropriés pour améliorer les performances du modèle.

VI.1.1.3. Adam :

L'utilisation de l'algorithme RMSprop conduit également à des observations similaires en ce qui concerne les performances du modèle sur les bases de données KIBA et BindingDB :

- Stabilité de la perte d'entraînement **:** La stabilité et la faible valeur de la perte d'entraînement dans la base KIBA avec RMSprop, par rapport à la stabilité élevée de la perte d'entraînement dans la base BindingDB, suggèrent une tendance similaire à ce qui a été observé avec Adam. Cela indique que l'algorithme RMSprop peut également rencontrer des difficultés à optimiser efficacement le modèle dans certaines situations.
- Surapprentissage dans KIBA **:** Bien que moins prononcé qu'avec Adam, la différence significative entre la perte de test et la perte d'entraînement dans la base KIBA avec

RMSprop suggère toujours la présence d'un certain niveau de surapprentissage. Il pourrait être nécessaire d'explorer des techniques de régularisation ou d'autres approches pour améliorer la capacité de généralisation du modèle.

- Optimisation inefficace dans BindingDB : Les observations de la stabilité de la perte d'entraînement à une valeur relativement élevée dans la base BindingDB avec RMSprop, ainsi que les fluctuations persistantes dans la perte de test, suggèrent des problèmes similaires à ceux observés avec Adam. Il est possible que le choix d'un taux d'apprentissage inapproprié ou d'autres hyperparamètres nécessite des ajustements pour améliorer les performances du modèle.

Sur la base des preuves présentées, il semble que la descente de gradient stochastique (SGD) surpasse les deux autres optimiseurs en termes d'efficacité et d'efficience soit au niveau de la base KIBA ou de la base BindingDB.

L'optimiseur SGD a montré une diminution rapide des pertes et a atteint un état de perte minimale pour les phases de formation et de test à laquelle Leeds a une convergence rapide et constante aussi Généralisation compétente qu'elle était illustré par les trajectoires de perte des données d'entrainement et de test présentent une proximité remarquable, ce qui signifie la capacité du modèle à généraliser sur l'ensemble de données de test.

Lorsqu'ils sont appliqués à cet ensemble de données spécifique, RMSprop et Adam présentent des indications de surajustement, comme en témoignent des pertes de test élevées et fluctuantes, diminuant ainsi leur fiabilité dans la réduction des pertes d'entraînement.

La forte variabilité observée dans les pertes de test pour RMSprop et Adam peut être attribuée à plusieurs facteurs liés à la nature de ces optimisateurs et à la manière dont ils mettent à jour les poids du modèle peuvent être :

- ✓ La complexité de modèle
- ✓ Sensibilité aux hyperparamètres comme le taux d'apprentissage
- ✓ La nature des données (contient beaucoup de bruit ou pas bien équilibrées)

Enfin on dit que RMSprop et Adam sont des optimiseurs puissants, mais leur sensibilité aux hyperparamètres et leur adaptation aux gradients peuvent parfois conduire à une forte variabilité des pertes de tests, surtout si le modèle ou les données ne sont pas bien adaptés à ces méthodes, on utilise donc toujours le SGD optimiseur comme optimiseur parfait pour nos modèles

VI.1.2. Les hyperparamètres

Pour améliorer la performance du modèle établi, quelques paramètres ont été modifiés, parmi lesquels on peut citer :

- Les fractions des données d'entrainement : Ce sont des fractions utilisées pour déterminer les sous-ensembles spécifiques (entraînement/test).
- Facteurs latents : Les facteurs latents sont des caractéristiques cachées qui capturent les propriétés des médicaments et des cibles protéiques. La matrice d'origine est décomposée en ces facteurs latents, ce qui permet de prédire les interactions médicament-cible manquantes de manière plus précise.
- Taille du lot BS : C'est le nombre d'échantillons d'entraînement utilisés lors de chaque itération (époque) pour mettre à jour des paramètres du modèle.

En effet, les fractions des données ne sont pas spécifiquement des hyperparamètres, mais après le changement de ces fractions, la valeur d'erreur a clairement diminué. L'algorithme qui réalise cette division crée trois parties : l'entraînement, la validation et le test. Cependant, le code élaboré ne continue qu'avec l'entraînement et le test ce qui explique la réduction de la fraction de validation afin de permettre à évaluer la capacité prédictive du modèle en optimisant les hyperparamètres. Pour ce faire, différents facteurs latents (50, 100 et 200) avec différentes tailles de lot (32, 64 et 128) ont été testés. Les pertes (entraînement/test) en fonction des époques pour chacun, ainsi que le temps pris ont été enregistrées. Les résultats obtenus sont présentés dans les tableaux suivants :

Tableau VI-1: *La moyenne des pertes (d'entraînement/de test) et le temps pris pour chaque combinaison de taille de lot et de facteur latente base BindingDB.*

Taille de lot	***Facteur latente***	***Perte d'entraînement moyenne***	***Perte de test moyenne***	***Temps (Seconds)***
32	*50*	0.3191296	0.8318129	314.29914
	100	0.2671291	0.862937	411.90377
	200	0.2262124	0.8864707	1532.7082
64	*50*	0.7861922	0.9344834	198.55281
	100	0.7043977	0.9322398	245.01747
	200	0.6006704	0.9383409	400.92282
128	*50*	1.6173057	1.3349756	143.03265
	100	1.5416074	1.3085126	174.81006
	200	1.4389121	1.2807228	276.72504

Tableau VI-2: *La moyenne des pertes (d'entraînement/de test) et le temps pris pour chaque combinaison de taille de lot et de facteur latente base KIBA.*

Taille de lot	***Facteur latente***	***Perte d'entrainement moyenne***	***Perte de test moyenne***	***Temps (Seconds)***
32	*50*	0.0395	0.3105	493.89
32	*100*	0.0190	0.3158	516.84
32	*200*	0.0094	0.3401	590.25
64	*50*	0.1164	0.3235	318.52
64	*100*	0.0827	0.3229	342.70
64	*200*	0.0567	0.3490	384.30
128	*50*	0.2839	0.4575	242.86
128	*100*	0.2351	0.4473	256.08
128	*200*	0.1925	0.4442	280.76

Il a été observé que la plus petite taille de lot (32) a présenté les moyennes des pertes (entraînement/test) les plus faibles (BindingDB : 0.319 ; 0.267 ; 0.226 / 0.832 ; 0.863 ; 0.886, KIBA : 0.0365 ; 0.019 ; 0.0094 / 0.3105 ; 0.3158 ; 0.34). Pour cette taille de lot, un nombre plus faible de facteurs latents (50) a donné une perte (entraînement/test) plus élevée, tandis que les valeurs 100 et 200ont permis d'améliorer les pertes (entraînement/test). Cependant, avec 200 facteurs latents, la perte de test moyenne est légèrement plus élevée (BindingDB : 0.886, KIBA :0.340), indiquant un problème de surapprentissage.

En conséquence, la configuration optimale des paramètres pour une meilleure performance du modèle est une taille de lot de 32 avec 100 facteurs latents, cette petite taille de lot étant suffisante pour mener à une meilleure généralisation ainsi que 100 facteurs latents suffisants pour capturer des relations plus complexes pour les deux bases d'entrainement.

VI.2. Performance de modèle :

Les résultats des mesures d'évaluation présentées dans le chapitre précédent ont été extraits pour valider ce travail. Une analyse de la courbe de perte quadratique moyenne (MSE) a été effectuée pour chaque règle d'entraînement. Les métriques MSE, MAE, RMSE et MAPE ont été calculées. Enfin, une étude de la courbe de régression. Les résultats du MAE, du MSE, du RMSE et du MAPE sont résumés dans le tableau suivant :

Tableau VI-3: *les valeurs des critères de validation*

Les critères	**MSE**	**MAE**	**RMSE**	**MAPE(%)**
BindingDB	0.858	0.55	0.926	9.1276
KIBA	0.3184	0.3411	0.5695	2.8295

VI.2.1. BindingDB

La figure VI.9 présente l'évolution de la perte quadratique moyenne (MSE) sur les ensembles d'entraînement et de test en fonction du nombre d'époques (itérations). Cette visualisation permet d'analyser la capacité du modèle à apprendre à partir des données d'entraînement et à généraliser sur de nouvelles données, ainsi que de détecter tout signe de surapprentissage ou de sous-apprentissage.

L'évolution de ce graphe se devise en deux phases :

1. **Première phase (0-60 époques) :**
    - Pendant cette phase initiale, une diminution rapide des pertes à la fois sur l'ensemble d'entraînement et sur l'ensemble de test a été observée. Elle indique que le modèle apprend efficacement à partir des données d'entraînement et est capable de réduire rapidement l'erreur de prédiction. La convergence rapide des pertes sur les ensembles d'entraînement et de test suggère que le modèle généralise bien sur de nouvelles données dès le début de l'entraînement.
2. **Deuxième phase (60-100 époques) :**
    - L'entraînement progresse, la diminution de la perte ralentit, et les courbes de perte sur les ensembles d'entraînement et de test commencent à se stabiliser, ce qui indique que les erreurs dans cette phase diminuent plus lentement par rapport à la phase initiale. Par conséquent, il n'est pas nécessaire de continuer l'entraînement ou d'augmenter le nombre d'itérations, car cela ne réduira pas significativement l'erreur.

D'autre part, on observe que les deux courbes (train/test) sont proches l'une de l'autre, ce qui signifie qu'à mesure que le nombre d'époques augmente, les valeurs d'erreur des deux groupes sont similaires. Cela suggère que le modèle est bien équilibré et se généralise bien, ce qui lui permet d'apprendre des relations importantes sans sur apprendre les détails spécifiques aux données d'entraînement. Si la courbe de test est nettement supérieure à la courbe d'entraînement, cela indique un sur apprentissage.

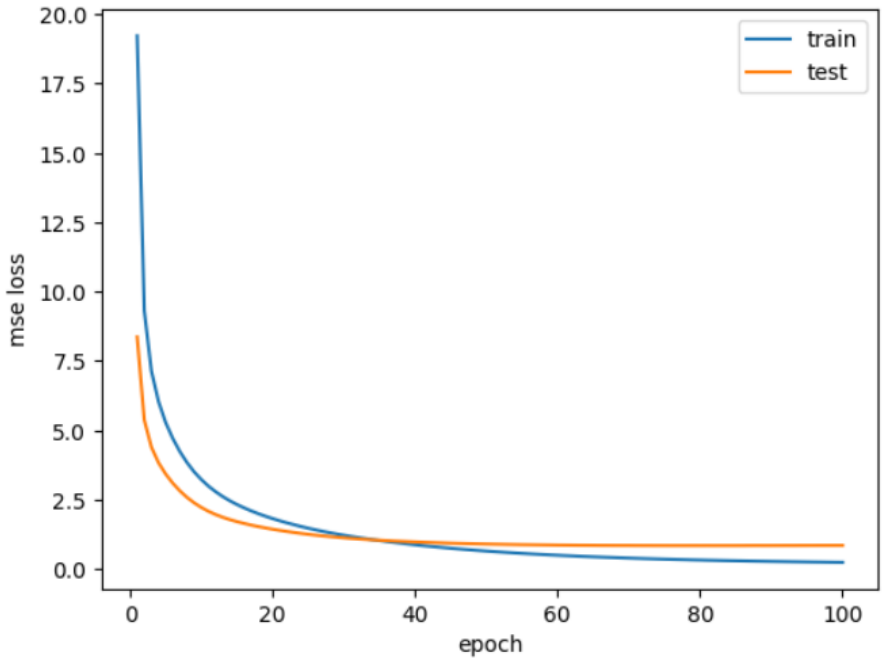


Figure VI-7: *la courbe de perte quadratique moyenne de BindingDB (loss)*

Concernant les résultats des mesures statistiques (tableau VI-1), les valeurs de MAE, MSE et RMSE sont relativement faibles, ce qui indique que la plupart des erreurs de prédiction sont raisonnables et que les prédictions du modèle sont exactes. De plus, le pourcentage de MAPE se rapproche de 0, ce qui suggère que les résultats des prédictions sont précis. Par conséquent, les valeurs des différentes métriques ont montré que le modèle établi de FM fonctionne bien et que l'erreur de prédiction n'affecte pas de manière disproportionnée les résultats.

Pour évaluer plus justement l'exactitude du modèle et sa capacité à prédire, la courbe de régression a été utilisée pour examiner la relation entre les valeurs réelles et les valeurs prédites. Les résultats du modèle sont représentés sur la figure (VI-10).

L'évolution de la courbe est comme suit :

- La plupart des points qui se superposent à la droite bissectrice indiquent une corrélation positive, ce qui signifie que les prédictions sont assez similaires aux valeurs réelles.
- Certains points qui se dispersent en dehors de la droite bissectrice indiquent une corrélation négative, ce qui signifie que les prédictions sont assez éloignées des valeurs réelles.

D'après le graphe de régression, on remarque qu'il y a des valeurs réelles identiques pour des valeurs prédites différentes (2.3/5/9), Ce problème semble provenir de la composition de notre base d'entraînement. En effet, un médicament peut interagir avec différentes

cibles, ce qui complexité notre base d'entraînement où certains médicaments ont plusieurs cibles. Cette observation suggère que le modèle n'a pas bien appris ces cas et prédit des valeurs incorrectes. Cela ne pose pas de problème majeur, il suffit donc d'enrichir notre jeu de données avec davantage de données d'entraînement.

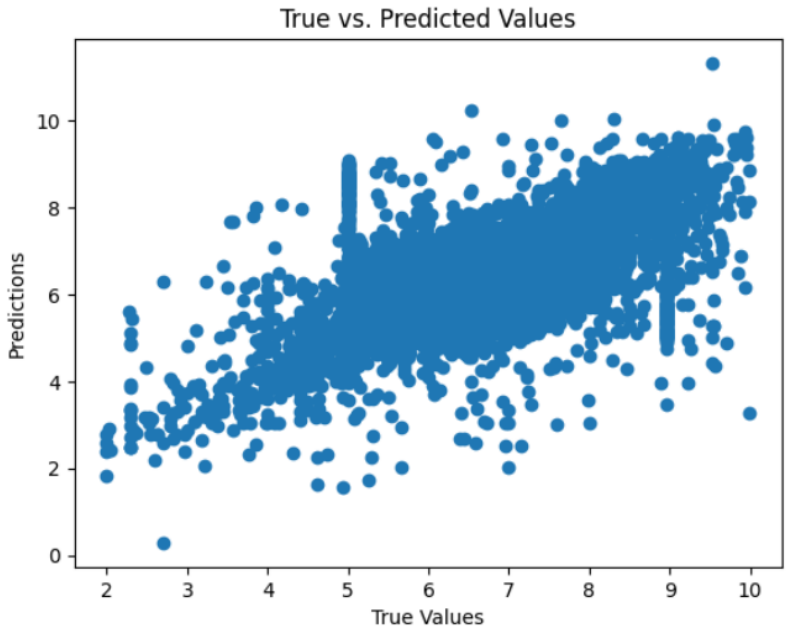


Figure VI-8: *la courbe de dispersion de BindingDB*

Après avoir validé le modèle de FM, il a été constaté que les courbes et les valeurs des statistiques d'évaluation sont satisfaisantes, avec des erreurs raisonnables. En conclusion, on peut affirmer que la performance de notre modèle est excellente, et que le modèle a réussi à recommander des cibles médicamenteuses correspondant à leurs propriétés.

VI.2.2. KIBA

Cette courbe représente l'évolution de la perte quadratique moyenne (MSE, Mean Squared Error) pour les ensembles d'entraînement (en bleu) et de test (en orange) au fil des époques d'entraînement de notre modèle.

- **Diminution rapide de la perte initiale** : au début de l'apprentissage, la perte dans les deux ensembles (entraînement et test) diminue rapidement, ce qui est typique car le modèle apprend rapidement les principales caractéristiques des données.
- **Stabilisation de la perte** : après presque 20 époques, les pertes pour les deux groupes se sont stabilisées et ont continué à diminuer très lentement, indiquant que le modèle avait atteint un stade de convergence et que les améliorations supplémentaires étaient limitées.

- **Pas de surapprentissage évident** : Les courbes de perte d'entraînement et de test présentent une proximité remarquable et affichent des modèles parallèles, indiquant l'absence de surapprentissage notable. En fait, si le modèle était enclin à un apprentissage excessif, on s'attendrait à une augmentation des pertes aux tests tandis que les pertes à la formation poursuivraient leur baisse.

- **Bonne généralisation :** La proximité des courbes de perte pour les données de formation et de test suggère une forte capacité du modèle à généraliser à des données inconnues. Le modèle fonctionne presque aussi bien sur les ensembles de données de formation que de test, ce qui constitue une indication favorable.

En conclusion, on peut en déduire que cette courbe démontre l'exécution réussie du processus de formation de notre modèle, démontrant une convergence louable et une généralisation efficace. Le modèle ne présente aucun signe de surajustement et les pertes observées dans l'ensemble de test s'alignent harmonieusement avec celles de l'ensemble d'entraînement.

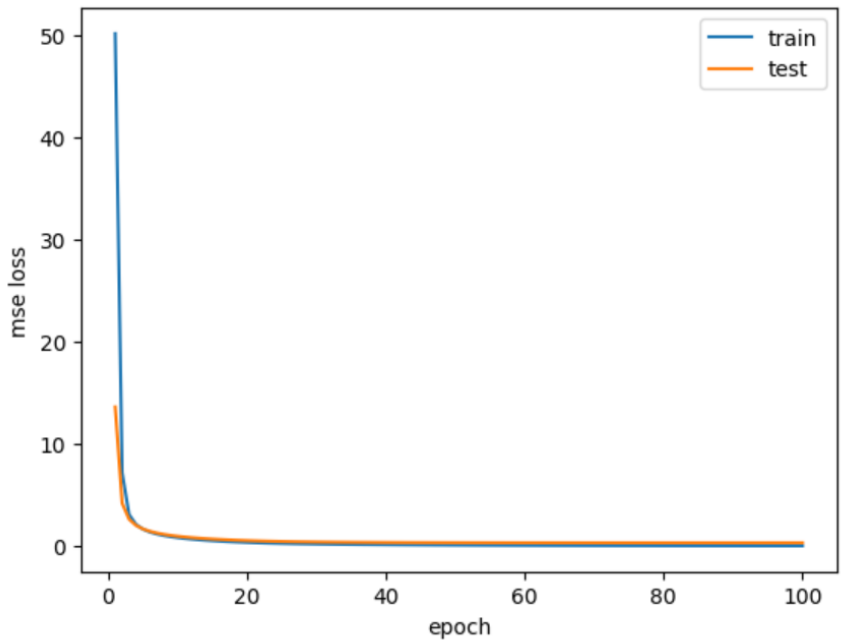


***Figure VI-9**: la courbe de perte quadratique moyenne de KIBA*

Concernant les résultats des mesures statistiques (tableau VI-9), on observe que :

- Faible erreur relative : Un MAPE de 2.8% est super excellent, signifiant que les erreurs de prédiction sont très petites par rapport aux valeurs réelles.

- Faible erreur moyenne absolue et quadratique : Les valeurs de MAE et MSE montrent que les erreurs de prédiction sont faibles en moyenne.
- Bon équilibre entre différentes métriques : Les valeurs de RMSE, MAE et MSE sont tous faibles, indiquant une cohérence dans la performance du modèle et une gestion efficace des erreurs, y compris des erreurs plus importantes.

En conclusion, les métriques indiquent que notre modèle est bien calibré et offre des prédictions précises avec une erreur relativement faible.

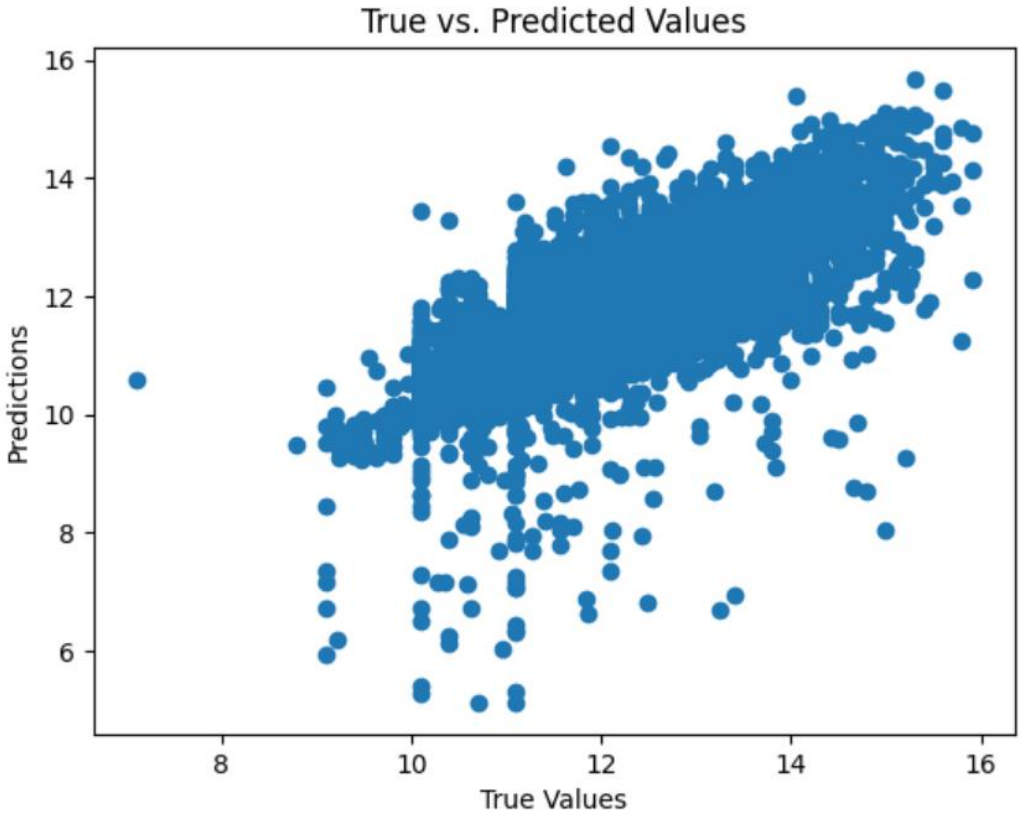


***Figure VI-10**: la courbe de dispersion de KIBA*

Cette courbe de dispersion (scatter plot) présente les valeurs réelles (True Values) par rapport aux valeurs prédites (Prédictions) de notre modèle, et on remarque que :

- **Alignement générale :** la majeure partie des points de données se regroupe près de la diagonale, ce qui signifie une forte corrélation entre les valeurs prédites et réelles. Cela implique que le modèle présente des performances satisfaisantes dans ses capacités prédictives.
- **Dispersion des points :** une dispersion perceptible des points de données peut être observée autour de la diagonale, en particulier pour les valeurs réelles supérieures à

10. Cela indique la présence d'une variabilité dans les prédictions, avec des écarts plus importants associés à des valeurs plus élevées.

- **Biais du modèle :** Si les points se regroupaient principalement au-dessus ou en dessous de la ligne y = x, cela indiquerait un biais systématique dans les prédictions (sous-estimation ou surestimation). Ici, la dispersion ne semble pas systématiquement orientée dans une direction particulière, ce qui est un bon signe.
- **Les points aberrants :** une poignée de points de données s'écartent considérablement de la tendance attendue (les valeurs 9 /10.1/11), ce qui suggère des inexactitudes substantielles dans les prévisions. Ces valeurs aberrantes peuvent être attribuées à des cas exceptionnels ou à des défauts du modèle ou dans la base de données

À la fin, le modèle présente une bonne capacité de prédiction avec des performances globales solides, malgré quelques écarts pour des valeurs extrêmes.

VI.3. Test externe

VI.3.1. BindingDB

Les scores de BindingDB sont exprimés en valeurs de Kd, une mesure de l'affinité de liaison. La constante de dissociation Kd correspond à la concentration de ligand nécessaire pour occuper 50 % des récepteurs. Plus le Kd est bas, plus l'affinité du ligand pour le récepteur est élevée. À l'inverse, un Kd élevé indique une faible affinité du ligand pour le récepteur. Ainsi, ce principe est utilisé pour identifier les molécules présentant la plus forte affinité de liaison, caractérisées par les valeurs de Kd les plus faibles.

Le tableau suivant montre la disposition des molécules existantes dans la base de données de prédiction, des molécules à faible score aux molécules à score élevé

Tableau VI-4: *les prédictions de BindingDB*

N°	*Nom de la substance*	*Prédiction*
39	Lomustine	0.28912782669067383
37	Camptothecin, Acetate	0.3225897252559662
65	Thiotepa	0.32798704504966736
53	Mitozolomide	0.32980015873908997
71	Colchicine	0.3301047086715698
67	Tritylcysteine	0.3316592276096344
34	Carmustine	0.33180078864097595
56	Profiromycin	0.3362348675727844
38	10-Hydroxycamptothecin	0.3382590115070343

62	5-((4-Chlorobenzyl)Thio)-3-(Trifluoromethyl)-1h-1,2,4-Triazole	0.34272733330726624
14	9-Aminocamptothecin (R,S)	0.3443114459514618
15	11-Hydroxymethyl-20(Rs)-Camptothecin	0.3443341851234436
10	Plumbagin	0.34603333473205566
58	Pyrazoloimidazole	0.3549492359161377
33	Triazinate	0.3579704761505127
60	Paclitaxel	0.3587341010570526
2	Menogaril	0.3587959408760071
7	Epsilon.-Rhodomycinone	0.3598419725894928
12	Dermocybin	0.3599000871181488
16	Camptothecin,(N-Diethyl) Glycinate	0.36056530475616455
11	Shikalkin	0.36206623911857605
6	Arnebin 1	0.36306291818618774
5	Rhodomycin A	0.3631356656551361
13	Pyrozofurin	0.3635620176792145
4	Dichloroallyl Lawson	0.36513566970825195
47	Hydroxyurea	0.36714333295822144
40	Dianhydrodulcitol	0.36911579966545105
8	4beta-Hydroxywithanolide	0.37531453371047974
9	Bikaverin	0.3838804364204407
55	Pipobroman	1.6009349822998047
57	Pyrazoloacridine Mesylate	1.7831355333328247
19	Camptothecin Butylglycinate Ester Hydrochloride	1.8755582571029663
69	Vincristine Sulfate	1.9491897821426392
44	Guanazole	2.4737346172332764
59	Spiromustine	2.8266453742980957
61	Teroxirone	3.308807849884033
63	3-Demethylthiocolchicine	3.3698179721832275
26	Acivicin	3.3700389862060547
29	Amonafide	3.7583601474761963
25	Dihydro-5-Azacytidine	3.8763298988342285
17	Camptothecin	3.9419257640838623
54	Urea	3.975724220275879
70	N-Benzoyl-Deacetylcolchicine	4.228569984436035
51	Semustine	4.328340530395508
27	Alpha.-Thiodeoxyguanosine	4.336771011352539
66	Triethylenemelamine	4.473781108856201
23	Clomesone	4.474382400512695
46	Hycanthone Mesylate	4.529837131500244
20	CamptothecinethylglycinateEsterhydrochloride	4.538515090942383
24	Picolinaldehyde	4.874126434326172

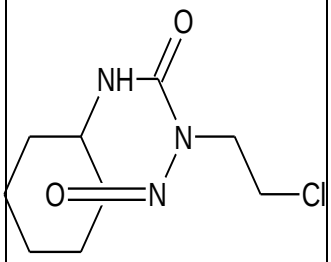
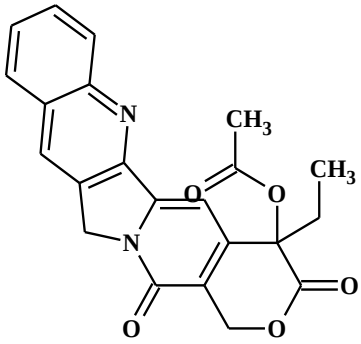
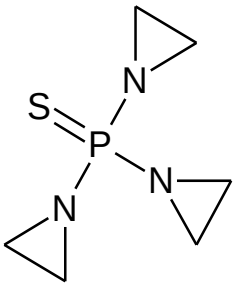
48	AlanosineMonosodium Salt	4.882874011993408
68	Vinblastine Sulfate	5.100278854370117
21	9-Methoxycamptothecin	5.134801387786865
64	6-Thioguanine	5.326044082641602
28	L-AsparticAcid	5.366934299468994
22	Chlorambucil	5.390965938568115
32	Diaziquone	5.592546463012695
36	Busulfan	5.624603748321533
42	Fluorodopan	5.658953666687012
49	Amsacrine	5.870268821716309
41	AC1L2OAS	6.019615173339844
18	Campothecin Analog	6.092503547668457
52	Mitomycin C	6.429431915283203
45	Hepsulfam	6.503969192504883
43	Uracil	6.571624755859375
31	Cytosine, Monohydrochloride	6.6257405281066895
35	Bisantrene Hydrochloride	6.7565412521362305
30	Anthra[1,9-Cd]Pyrazol-6(2h)-One Der	6.875985622406006
50	Methotrexate	7.1340460777282715
3	N,N-Dibenzyldaunorubicin Hydrochloride	7.729925632476807
1	Deoxydoxorubicin	9.198776245117188

Les résultats de nos prédictions indiquent que les molécules les plus efficaces pour le traitement du cancer de la peau, en particulier contre la protéine cible BRAF-V600E, sont la Lomustine, le Camptothécine de l'Acétate et la Thiotepa. Ces molécules ont montré une haute affinité pour cette protéine cible. En revanche, des molécules telles que le méthotrexate, le chlorhydrate de N,N-Dibenzyldaunorubicine et la Déoxydoxorubicine, qui présentent des valeurs de Kd plus élevées, ne peuvent pas se lier efficacement à cette cible spécifique et ne sont donc pas adaptées pour le traitement du cancer de la peau avec cette mutation.

VI.3.1.1. Propriétés des molécules choisies

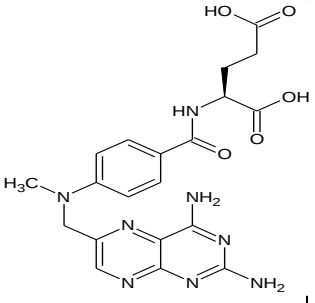
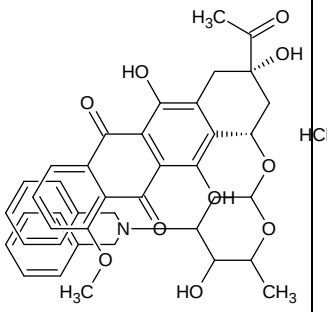
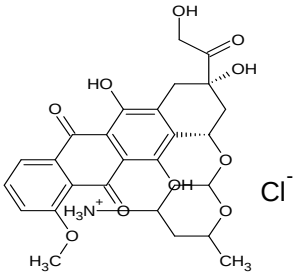
Les propriétés physiques et chimiques des meilleures molécules sont résumées dans le tableau VI.5 afin de mieux comprendre la capacité de ces molécules à traiter le cancer de la peau.

Tableau VI-5: *Propriétés physiques et chimiques des meilleures 3 molécules de BindingDB*

Caractéristique	***Lomustine***	***Camptothécine Acetate***	***Thiotepa***
Structure			
Formule chimique	$C_9H_{16}ClN_3O_2$	$C_{22}H_{18}N_2O_5$	$C_6H_{12}N_3PS$
Complexité	219	846	194
Types de Cycles	Cyclohexane	Pentacyclique, Quinoléine, Indole et Lactone	Aziridinyl
Groupes fonctionnels	Nitrosourée, Chloroéthyle et Cyclohexyle	Lactone (ester cyclique), Amide, Cétones, Éther et Acétate (ester)	Aziridine, Thiophosphoryle
Nombre de donneurs de liaisons hydrogène	1	0	0
Nombre d'accepteurs de liaison hydrogène	3	6	4
Nombre de liaisons rotatives	3	3	3
Solubilité	Soluble dans l'éthanol, l'acétone, Librement soluble dans le chloroforme	Diméthylsulfoxyde (DMOS), Méthanol, éthanol et acétone	Soluble dans l'eau à 25 °C Librement soluble dans l'alcool, l'éther, le chloroforme ; soluble dans le benzène
Mécanisme d'action	Agent alkylant	Inhibition de la topoisomérase i	Agent alkylant
Applications thérapeutiques	Des tumeurs cérébrales, du cancer du poumon, du mélanome malin et d'autres tumeurs solides	Les cancers du sein et des poumons et le cancer des ovaires	Le cancer du sein, le cancer des ovaires et le cancer de la vessie

Pour approfondir notre étude, on a également résumé certaines des propriétés physiques et chimiques des dernières molécules dans le tableau suivant :

Tableau VI-6: *Propriétés physiques et chimiques des mauvaises 3 molécules de BindingDB*

Caractéristique	**Méthotrexate**	**Chlorhydrate de N,N-Dibenzyldaunorubicin**	**Déoxydoxorubicine**
Structure			
Formule Chimique	$C_{20}H_{22}N_8O5$	$C_{41}H_{42}ClNO_{10}$	$C_{27}H_{30}ClNO_{10}$
Complexité	704	1260	945
Types de Cycles	Aromatiques benzéniques	Aromatiques benzéniques	Aromatiques benzéniques
Groupes Fonctionnels	Amine et carboxylate.	Hydroxyles, méthyles, benzyle	Hydroxyles (des structures aromatiques condensées)
Nombre de donneurs de liaisons hydrogène	5	5	5
Nombre d'accepteurs de liaison hydrogène	12	11	11
Nombre de liaisons rotatives	9	9	5
Solubilité	Insoluble dans l'eau, l'alcool, le chloroforme, l'éther ; légèrement soluble dans l'acide chlorhydrique dilué ; soluble dans les solvants dilués d'hydroxydes et de carbonates alcalins. (Hydrophile avec de faibles propriétés lipophiles)	Présence de groupes hydrophiles et hydrophobes	Contient des motifs hydrophiles et des charges

Mécanisme d'Action	Inhibiteur de la dihydrofolate réductase (DHFR), empêchant la synthèse de l'ADN et de l'ARN	Inhibiteur de la topoisomérase II, intercalant dans l'ADN	Inhibiteur de la topoisomérase II, intercalant dans l'ADN
Applications Thérapeutiques	Utilisé pour traiter certains types de cancer (comme la leucémie et le lymphome), ainsi que des maladies auto-immunes (comme la polyarthrite rhumatoïde et le psoriasis)	Utilisé pour traiter certains types de cancer, en particulier les leucémies et les lymphomes	Utilisé pour traiter divers types de cancer, en particulier les cancers du sein et les cancers hématologiques

VI.3.1.2. La comparaison entre les molécules BindingDB

✓ **Les meilleures molécules :**

- *Types de groupes fonctionnels* : chaque molécule possède des groupes fonctionnels spécifiques : la Lomustine contient un groupe Cyclohexane, la Camptothécine Acetate comprend des groupes lactone, amide, cétones, éther et acétate, et la Thiotepa contient des groupes aziridine, thiophosphoryle.
- *Complexité Structurale* : Le Camptothécine Acetate, Lomustine et Thiotepa présentent des structures différentes qui influencent leurs interactions avec les cibles thérapeutiques. Lomustine et Thiotepa ont des structures simples avec des cycles tels que le Cyclohexane et l'Aziridinyl. Par contre, le Camptothécine Acetate possède une structure complexe contenant plusieurs cycles, notamment Pentacyclique, Quinoléine, Indole et Lactone.

La Camptothécine acétate comporte deux structures importantes pour le développement de nouvelles thérapies ciblées contre la progression du mélanome : la quinoléine et l'indole. Les dérivés de la quinoléine exercent une activité antitumorale en inhibant la polymérisation de la tubuline, la réparation de l'ADN, et l'angiogenèse, en bloquant des enzymes clés pour la croissance tumorale, et en s'intercalant dans l'ADN. Les dérivés de l'indole possèdent également des propriétés antitumorales en inhibant la polymérisation de la tubuline et la topoisomérase, en induisant l'apoptose, et en inhibant l'aromatase et la kinase[4]. Cela renforce le potentiel de l'activité inhibitrice du

Camptothécine Acetate contre la cible BRAF-V600E. De plus, le fait que les molécules Lomustine et Thiotépa montrent une forte affinité pour BRAF-V600E suggère que les groupes fonctionnels de ces molécules pourraient également contribuer au développement de nouvelles thérapies ciblées contre la mutation BRAF-V600E.

- *Propriétés anticancéreuses* : Les trois molécules (Lomustine, Camptothécine Acétate, et Thiotepa) sont des agents chimio thérapeutiques couramment utilisés pour traiter différents types de cancers.
- *Les accepteurs de liaisons hydrogène* : chaque molécule possède un nombre spécifique d'accepteurs de liaisons hydrogène. La Camptothécine Acétate en a le plus grand nombre avec six accepteurs, suivie de la Thiotepa avec quatre, puis de la Lomustine avec trois. Ces accepteurs de liaisons hydrogène favorisent les interactions spécifiques avec la cible BRAF-V600E, peut contribuer à la stabilité et à l'affinité de liaison aux sites actifs des protéines.
- *Indicateur de la flexibilité de la molécule* : Les trois molécules ont les mêmes nombres de liaisons rotatives « 3 », ceci suggère qu'elles ont une certaine flexibilité moléculaire pourrait affecter la capacité de ces molécules à adopter une conformation favorable pour se lier à la cible BRAF-V600E.
- *Solubilité* : Ces molécules (Lomustine, Camptothécine Acétate et Thiotepa) sont toutes solubles dans les solvants organiques, ce qui peut favoriser une meilleure solubilité dans les solvants biologiques. Cette propriété permet une absorption plus efficace et une meilleure distribution dans les tissus cibles, potentiellement améliorant la distribution vers la cible BRAF-V600E.
- *Mécanisme d'action :* Les deux molécules, Lomustine et Thiotepa, sont des agents alkylants dont les métabolites interfèrent avec la fonction de l'ADN. Elles arrêtent la croissance tumorale en endommageant l'ADN et en empêchant les cellules de se diviser. D'autre part, le Camptothecin Acétate inhibe sélectivement l'enzyme nucléaire ADN topoisomérase de type I, empêchant ainsi la réplication de l'ADN, ce qui entraîne l'apoptose.

D'autre part, selon les résultats, une faible affinité a été observée de la Camptothécine pure avec la cible BRAF-V600E par rapport au Camptothécine Acetate, principalement en raison des limites pharmacologiques de la Camptothécine pure (voir Tableau VI.8). En effet, celle-ci présente une faible solubilité dans l'eau et des problèmes de stabilité à pH physiologique, ce qui limite son utilisation clinique directe. En revanche, la modification par acétylation offre

une meilleure stabilité chimique sous certaines conditions et réduit potentiellement la toxicité par rapport à la Camptothécine pure. Le groupe acétate peut protéger la molécule de la dégradation et contribuer à maintenir voire à améliorer son activité anticancéreuse contre les cellules tumorales ciblées.

- ✓ **Les trois dernières molécules :**

- En ce qui concerne la structure chimique et le mécanisme d'action, le chlorhydrate de N,N-Dibenzyldaunorubicine et la Déoxydoxorubicine sont étroitement liés. Cependant, ils se distinguent par de légères modifications de leur composition chimique, ce qui impacte leur spectre d'activité et leur profil de toxicité. Ces différences expliquent leur capacité à être utilisés dans le traitement de diverses formes de cancer. En revanche, le Méthotrexate est structurellement différent par rapport aux deux composés, le chlorhydrate de N,N-Dibenzyldaunorubicine et la Déoxydoxorubicine, D'autre part, ces types de structures volumineuses et rigides des anthracyclines peuvent limiter leur capacité à interagir de manière spécifique et efficace avec la protéine BRAF, ce qui influence leur affinité pour BRAF-V600E.
- Le chlorhydrate de N,N-Dibenzyldaunorubicin, le méthotrexate et la Déoxydoxorubicine, ayant le même nombre de donneurs de liaisons hydrogène (5 chacun), peuvent suggérer des capacités similaires à former des liaisons hydrogène en tant que donneurs. Cependant, le méthotrexate a une plus grande capacité d'accepteurs de liaisons hydrogène (12) par rapport aux deux autres composés (11). Il est possible que l'augmentation de la capacité de liaison hydrogène améliore les interactions avec certaines cibles biologiques, mais elle peut également augmenter les interactions non spécifiques, ce qui entraîne une diminution de l'affinité pour BRAF-V600E. De même, l'amélioration de la capacité des accepteurs de liaisons hydrogène peut favoriser la formation d'interactions spécifiques avec les donneurs de liaisons hydrogène présents dans la cible biologique, augmentant ainsi l'affinité pour celle-ci. Toutefois, une capacité excessive d'accepteurs peut également augmenter les interactions non spécifiques, ce qui diminue la spécificité et l'efficacité inhibitrice.
- En termes de la flexibilité moléculaire, le méthotrexate et le chlorhydrate de N,N-Dibenzyldaunorubicin présentent un nombre plus élevé de liaisons rotatives (9 chacune) comparé à la Désoxydoxorubicine (5). La flexibilité accrue peut favoriser l'adaptation de la molécule à la cible, mais elle pourrait également entraîner une diminution de la spécificité.

- La solubilité du méthotrexate dans les solvants organiques courants est limitée, mais il est soluble dans des conditions basiques. Le chlorhydrate de N,N-Dibenzyldaunorubicine présente une solubilité modérée dans les milieux aqueux et organiques. En revanche, la Déoxydoxorubicine est probablement plus soluble dans l'eau, ce qui pourrait entraîner des interactions différentes dans les milieux biologiques.

D'après cette comparaison, les trois molécules Lomustine, Acétate de Camptothécine et Thiotepa présentent plusieurs similitudes et différences dans leur structure chimique ainsi que dans les groupes fonctionnels spécifiques à leur mécanisme d'action. Bien que ces molécules n'aient pas été spécifiquement développées pour traiter le mélanome porteur de la mutation BRAF-V600E, elles possèdent des propriétés anticancéreuses qui pourraient être utilisées pour traiter cette forme de cancer. La Lomustine a déjà été utilisée avec succès pour le traitement du mélanome malin, en particulier en raison de son mécanisme d'alkylation de l'ADN, qui pourrait avoir des effets bénéfiques sur les mutations BRAF-V600E. De même, l'inhibiteur de la topoisomérase I (Acétate de Camptothécine) pourrait théoriquement affecter les cellules porteuses de la mutation BRAF-V600E en inhibant la réplication des fragments d'ADN, conduisant ainsi à la rupture de l'ADN. Les métabolites du Thiotepa agissent comme des agents alkylants et peuvent également interagir avec l'ADN de ces cellules.

Les propriétés physicochimiques (la solubilité, la capacité de liaison hydrogène et la flexibilité moléculaire) des trois composés (méthotrexate, chlorhydrate de N,N-Dibenzyldaunorubicine et Déoxydoxorubicine) indiquent des différences significatives dans leur potentiel d'interaction avec BRAF-V600E. De plus, un équilibre optimal entre les donneurs et les accepteurs de liaison hydrogène est crucial pour maximiser l'affinité et la spécificité de l'inhibiteur pour cette cible spécifique.

En conclusion, des molécules plus petites, flexibles et structurées peuvent suffire pour cibler spécifiquement les kinases. Cependant, des recherches supplémentaires sont nécessaires pour vérifier l'efficacité et la sécurité spécifiques de ces molécules (Lomustine, Acétate de Camptothécine et Thiotepa) dans le traitement du mélanome porteur de cette mutation spécifique.

VI.3.2. KIBA

Dans les scores d'affinité des inhibiteurs de kinase dans la base de données KIBA, une affinité plus élevée entre l'inhibiteur et la kinase cible est généralement indiquée par un score

plus faible. Cela signifie que plus le score est bas, meilleure est l'affinité, ce qui indique une interaction plus forte et une plus grande puissance de l'inhibiteur.

La signification est :

- ✓ ***Score faible*** : indique une meilleure affinité entre l'inhibiteur et la kinase cible. Cela signifie que l'inhibiteur est plus efficace pour bloquer l'activité de la kinase.
- ✓ ***Score élevé*** : indique une faible affinité entre l'inhibiteur et la kinase cible. Cela signifie que l'inhibiteur est moins efficace pour bloquer l'activité de la kinase.

Ce principe est couramment utilisé dans les bases de données d'interactions biologiques, où de faibles valeurs pour des métriques telles que Ki, Kd ou IC50 correspondent à une inhibition plus efficace.

Le tableau VI.6 représente la disposition des molécules présentes dans la base de données de prédiction de la molécule de faible score a l'une de score élevé :

Tableau VI-6 : *les prédictions de KIBA*

N°	***Nom de la Substance***	***Prédiction***
16	Camptothecin,(N-Diethyl) Glycinate	8,04933
44	Guanazole	9,99481
68	Vinblastine Sulfate	10,5375
43	Uracil	10,5524
7	Epsilon,-Rhodomycinone	10,6023
42	Fluorodopan	10,6383
21	9-Methoxycamptothecin	10,6542
9	Bikaverin	10,6679
30	Anthra[1,9-Cd]Pyrazol-6(2H)-One Der	10,8525
32	Diaziquone	10,8543
14	9-Aminocamptothecin (R,S)	10,8944
34	Carmustine	10,9043
63	3-Demethylthiocolchicine	10,9185
53	Mitozolomide	10,9424
61	Teroxirone	10,9481
66	Triethylenemelamine	10,9609
1	Deoxydoxorubicin	10,9611
64	6-Thioguanine	10,9929
58	Pyrazoloimidazole	11,0179
28	L-AsparticAcid	11,045
67	Tritylcysteine	11,0454
22	Chlorambucil	11,065
45	Hepsulfam	11,0656
12	Dermocybin	11,0665
17	Camptothecin	11,08
8	4BETA-Hydroxywithanolide	11,0879
2	Menogaril	11,0925
13	Pyrozofurin	11,0964

20	CamptothecinethylglycinateEsterhydrochloride	11,0974
15	11-Hydroxymethyl-20(Rs)-Camptothecin	11,1064
47	Hydroxyurea	11,1272
23	Clomesone	11,1413
71	N-Benzoyl-Deacetylcolchicine	11,1458
40	Dianhydrodulcitol	11,1517
25	Dihydro-5-Azacytidine	11,173
62	5-((4-Chlorobenzyl)Thio)-3-(Trifluoromethyl)-1H-1,2,4-Triazole	11,1907
4	Dichloroallyl Lawson	11,1952
48	AlanosineMonosodium Salt	11,2033
57	Pyrazoloacridine Mesylate	11,2254
35	Bisantrene Hydrochloride	11,2363
10	Plumbagin	11,2908
70	Colchicine	11,3017
54	Urea	11,3175
6	Arnebin 1	11,3221
59	Spiromustine	11,3721
69	Vincristine Sulfate	11,3957
31	Cytosine, Monohydrochloride	11,4256
65	Thiotepa	11,4333
51	Semustine	11,4475
33	Triazinate	11,452
26	Acivicin	11,4686
60	Paclitaxel	11,4993
56	Profiromycin	11,5037
29	Amonafide	11,5182
11	Shikalkin	11,5226
39	Lomustine	11,5353
27	Alpha,-Thiodeoxyguanosine	11,5472
36	Busulfan	11,5508
24	Picolinaldehyde	11,5737
52	Mitomycin C	11,7088
19	Camptothecin Butylglycinate Ester Hydrochloride	11,7874
55	Pipobroman	11,8378
41	Ac1l2oas	11,8468
50	Methotrexate	11,8636
49	Amsacrine	12,1402
37	Camptothecin, Acetate	12,299
5	Rhodomycin A	12,5039
46	Hycanthone Mesylate	12,6835
38	10-Hydroxycamptothecin	12,8566
18	Campothecin Analog	13,4713
3	N,N-Dibenzyldaunorubicin Hydrochloride	13,6612

D'après ce tableau on trouve que les molécules Camptothecin, (N-Diethyl) Glycinate, Guanazole et Vinblastine Sulfate eu une grande affinité avec notre cible étudiée V600E-

BRAF Donc sont les plus adaptées pour le traitement de mélanome. On note aussi que les molécules10-Hydroxycamptothecin, Campothecin Analog et N,N-Dibenzyldaunorubicin Hydrochloride ayant les scores plus élevées dans notre liste donc on peut dire que leur affinité avec la cible V600E-BRAF est très faible.

VI.3.2.1. Propriétés des molécules choisies

On commence par les propriétés des trois meilleures molécules :

Tableau VI-7: *Propriétés physiques et chimiques des meilleures 3 molécules de KIBA*

Caractéristique	***Camptothécine, (N-Diethyl) Glycinate***	***Guanazole***	***Vinblastine Sulfate***
Formule Chimique	$C_{26}H_{28}ClN_3O_5$	$C_2H_5N_5$	$C_{46}H_{56}N_4O_{10}, H_2O_4S$
Complexité de la Structure	939	63,1	1830
Types de Cycles	Système Pentacyclique	Un seul cycle triazole	Polycyclique avec plusieurs types de cycles
Groupes Fonctionnels	Lactone, méthyl, cycles aromatiques	Amino, triazole	Méthoxyle, hydroxyle, éther, hétérocycles
Nombre de donneurs de liaisons hydrogène	1	3	5
Nombre d'accepteurs de liaison hydrogène	7	4	16
Nombre de liaisons rotatives	7	0	10
Stabilité	Stable, améliorée par modification chimique	Stable à l'air, réagit avec les acides	Stable en solution aqueuse à température ambiante ou inférieure
Solubilité	0,25 mg/ml dans DMSO (pH 7.2)	Soluble dans l'eau	0.5 mg/ml dans PBS (pH 7.2), 50 mg/ml dans l'eau

Toxicité	Réduite par rapport à la Camptothécine pure, effets GI possibles	Toxicité modérée, fièvre, érythème noueux avec arthralgies	Suppression de la moelle osseuse, toxicité gastro-intestinale, activité vésicante, blessures d'extravasation
Mécanisme d'Action	Inhibition de la topoisomérase I	Inhibition de la synthèse de l'ADN et division cellulaire	Liaison à la tubuline et inhibition de la mitose
Applications Thérapeutiques	Traitement des cancers du côlon et de l'ovaire	Traitements anticancéreux spécifiques mais moins courants	Traitement des lymphomes de Hodgkin, cancers du sein et du testicule

Tableau VI-8: *Propriétés physiques et chimiques des molécules de comparaison de KIBA*

Caractéristique	***Vincristine Sulfate***	***Camptothecin***
Formule Chimique	$C_{46}H_{58}N_4O_{14}S$	$C_{20}H_{16}N_2O_4$
Complexité de la Structure	1830	742
Types de Cycles	Indole, pipéridine, tétrahydroindole	Pyrrolo[3,4-b]quinoline, lactone
Groupes Fonctionnels	Ester, amine, alcool, sulfonamide	Lactone, amide, alcool
Nombre de donneurs de liaisons hydrogène	5	1
Nombre d'accepteurs de liaison hydrogène	16	5
Nombre de liaisons rotatives	10	1
Stabilité	Relativement stable	Sensible à l'hydrolyse en milieu basique

Solubilité	Soluble dans l'eau	Faiblement soluble dans l'eau
Toxicité	Neurotoxique, myélosuppression	Myélosuppression, gastro-intestinale
Mécanisme d'Action	Inhibe la polymérisation des microtubules	Inhibition de la topoisomérase I
Applications Thérapeutiques	Utilisé dans le traitement des leucémies, lymphomes, et autres cancers	Utilisé dans le traitement des cancers colorectal, de l'ovaire, et du poumon

On passe maintenant aux propriétés des trois dernières molécules :

Tableau VI-9: *Propriétés physiques et chimiques des derniers 3 molécules de KIBA*

Caractéristique	***10-Hydroxycamptothecin***	***Campothecin Analog***	***N,N-Dibenzyldaunorubicin Hydrochloride***
Formule Chimique	$C_{20}H_{16}N_3O_5$	$C_{22}H_{19}N_3O_5$	$C_{41}H_{42}ClNO_{10}$
Complexité de la Structure	774	865	1260
Types de Cycles	Système pentacyclique	Système pentacyclique	Polycyclique avec plusieurs types de cycles
Groupes Fonctionnels	Lactone, hydroxy, cycles aromatiques	Lactone, amide, cycles aromatiques	Amino, méthoxyle, éther, hétérocycles, chlorure
Nombre de donneurs de liaisons hydrogène	2	1	5
Nombre d'accepteurs de liaison hydrogène	6	7	11

Nombre de liaisons rotatives	1	4	9
Stabilité	Relativement stable, sensible à la lumière	Variable selon le substituant	Relativement stable, sensible à l'hydrolyse
Solubilité	Faiblement soluble dans l'eau, soluble dans les solvants organiques	Variable, souvent faiblement soluble dans l'eau	Faiblement soluble dans l'eau, soluble dans les solvants organiques
Toxicité	Toxicité dose-dépendante, effets secondaires GI et hématologiques	Toxicité dose-dépendante, effets secondaires variables	Toxicité élevée, suppression de la moelle osseuse, cardiotoxicité
Mécanisme d'Action	Inhibition de la topoisomérase I	Inhibition de la topoisomérase I	Intercalation de l'ADN, inhibition de la topoisomérase II
Applications Thérapeutiques	Traitement du cancer du côlon et de l'ovaire	Traitement du cancer, en développement	Traitement des leucémies et des lymphomes

VI.3.2.2. La comparaison entre les molécules KIBA

Le sulfate de vincristine ($C_{46}H_{58}N_4O_{14}S$) et le sulfate de vinblastine ($C_{46}H_{56}N_4O_{10}$, H_2O_4S) sont deux alcaloïdes obtenus à partir de Catharanthus roseus, une plante réputée pour sa composition chimique complexe (1830). Leurs structures polycycliques complexes et divers groupes fonctionnels jouent un rôle crucial dans leur interaction avec la cible V600E-BRAF, un oncogène fréquemment impliqué dans certains types de cancer. La vincristine possède un répertoire de groupes fonctionnels comprenant des esters, des amines, des alcools et des sulfamides, tandis que la vinblastine présente des méthoxyles, des hydroxyles, des éthers et des hétérocycles. Ces groupes fonctionnels contrastés peuvent potentiellement influencer la manière dont chaque molécule s'engage avec la cible V600E-BRAF.

La présence de groupes méthoxyle et hydroxyle dans la vinblastine lui permet d'établir des interactions plus robustes et précises avec la poche de liaison du V600E-BRAF, contrairement à la vincristine. Ces disparités structurelles subtiles mais significatives jouent un rôle central en déterminant l'efficacité et l'affinité des molécules envers leurs cibles. D'un point de vue clinique, cela se traduit par une application plus ciblée de la vinblastine pour le

traitement des cancers associés au V600E-BRAF, tandis que la vincristine est utilisée pour un plus large éventail de cancers sans affinité spécifique pour cette cible particulière.

Il est clair que chacune de 10-Hydroxycamptothecin, Campothecin Analog, Camptothecin et Camptothecin (N-Diethyl) Glycinate présente des caractéristiques structurales distinctes qui peuvent influencer leur affinité avec la cible V600E-BRAF. La première molécule, le 10-Hydroxycamptothecin, se distingue par sa complexité de structure relativement élevée, une stabilité variable et une toxicité dose-dépendante, notamment des effets secondaires gastro-intestinaux et hématologiques. La deuxième molécule, une analogie du Camptothecin, affiche une variabilité en termes de stabilité, de solubilité et de toxicité, ce qui peut résulter de substituants spécifiques. La troisième molécule, le Camptothecin pur, présente une stabilité moindre et une solubilité faible dans l'eau, avec une toxicité associée principalement à des effets myélosuppressifs et gastro-intestinaux. Enfin, le Camptothecin (N-Diethyl) Glycinate se distingue par sa stabilité améliorée, une toxicité réduite par rapport au Camptothecin pur, et une structure complexe qui peut favoriser une affinité plus forte avec la cible. Ces différences structurelles et de propriétés chimiques peuvent donc influencer l'affinité de ces molécules avec la cible V600E-BRAF, la quatrième molécule montrant potentiellement une affinité plus grande en raison de sa stabilité accrue et de sa structure complexe.

La différence entre la structure des molécules 1, 2 et 4 par rapport à la molécule 3 réside principalement dans leur composition chimique et leur complexité structurale. Les molécules 1, 2 et 4 appartiennent toutes à la classe des Camptothécines, caractérisées par un noyau Pentacyclique et une lactone. Cependant, elles diffèrent dans leurs substituants et leurs groupes fonctionnels. Par exemple, la molécule 1, le 10-Hydroxycamptothecin, comporte des groupes hydroxyle supplémentaires par rapport à la molécule 3, le Camptothecin pur. La molécule 2, une analogie du Camptothecin, peut présenter des substituants différents, tels que des amides, qui peuvent altérer ses propriétés chimiques et biologiques. La molécule 4, le Camptothecin (N-Diethyl) Glycinate, possède des substituants diéthyl et glycinate, ce qui peut modifier sa solubilité et sa réactivité par rapport à la molécule 3. De plus, la molécule 4 peut présenter une complexité structurale supérieure avec des ramifications supplémentaires, ce qui peut influencer son interaction avec la cible V600E-BRAF

VI.4. Comparaison

Afin d'approfondir notre compréhension du modèle de factorisation matricielle, nous avons réalisé quelques comparaisons concernant les performances et les modèles à l'aide des résultats obtenus,

VI.4.1. Entre le modèle initial et amélioré

À partir de cette comparaison, il est possible d'évaluer l'effet des changements effectués aux paramètres de notre modèle de factorisation matricielle

Tableau VI-10: *Les ajustements des paramètres du modèle*

Modèle	***Fraction***	***Facteur latent***	***Taille de lot***
Initial	0,6/0,05/0, 35	20	100
Amélioré	0,749/0,001/0,25	100	32

Tableau VI-11: *L'effet des ajustements des paramètres du modèle sur les critères de validation*

La Base	***Le modèle***	***MSE***	***MAE***	***RMSE***	***MAPE %***
BindingDB	***Initial***	1,268	0,698	1,129	11,373
	Amélioré	0,8558	0,459	0,925	9,119
KIBA	***Initial***	0,5287	0,7271	0,7271	3,5038
	Amélioré	0,3184	0,3411	0,5643	2,8295

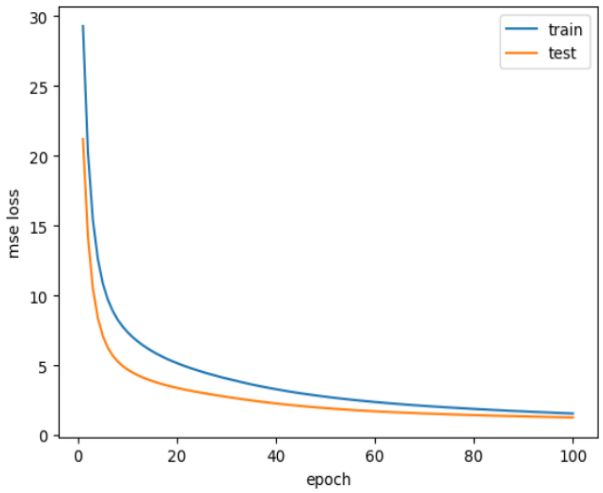


Figure VI-11*:la courbe de perte quadratique moyenne (MSE) de modèle initial pour BindingDB*

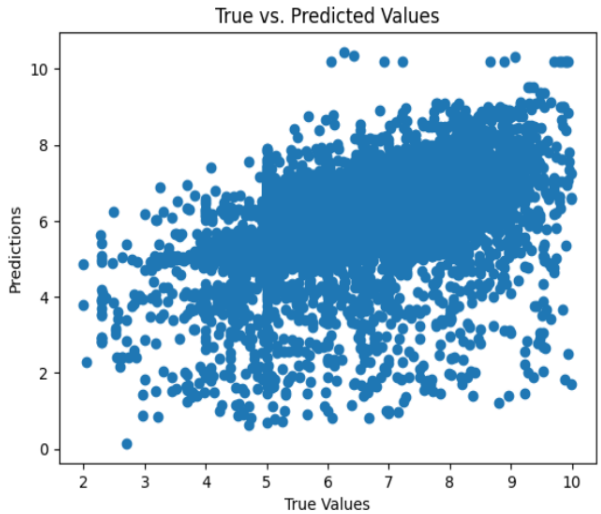


Figure VI-12*:la courbe de dispersion de modèle initial pour BindingDB*

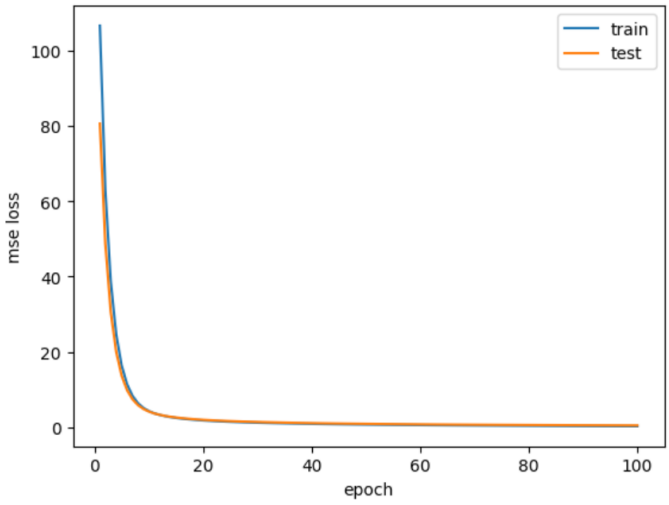


***Figure VI-13**:la courbe de perte quadratique moyenne (MSE) de modèle initial pour KIBA*

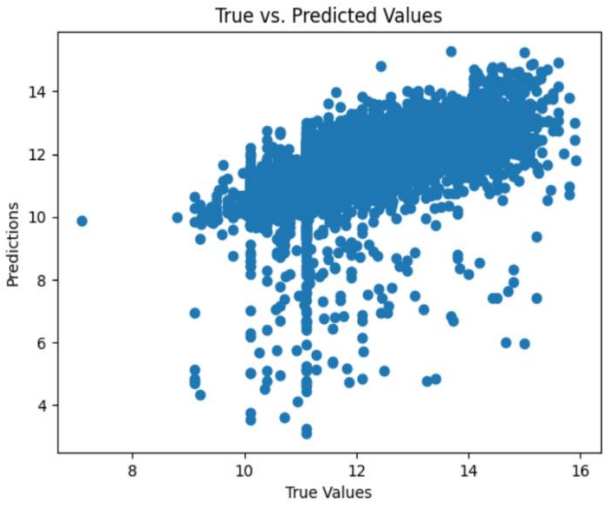


***Figure VI-14**:la courbe de dispersion de modèle initial pour KIBA*

Ces résultats nous ont permis de comparer entre le modèle amélioré et le modèle initial (tableau VI-6):On remarque une réduction notoire dans le MSE pour les deux bases aussi en remarque une amélioration significative de l'erreur moyenne de prédiction MAE d'autre part on marque un abaissement remarquable dans le RMSE accompagné avec une réduction substantielle au niveau de MAPE et tout ça pour les deux base donc en peut dire que on a une excellente amélioration sur toutes les métriques de performance comparé au modèle initial ce qui nous renforce l'idée que le modèle amélioré a une meilleur précision prédictive et des résultats plus fiables que le modèle initial.

Après on passe aux figures (VI-11 et VI-13) et les comparé avec les figures (VI-7 et VI-9) :

- Comparaison entre la figure VI-11 et la figure VI-7 ou autrement dit entre la courbe de perte quadratique moyenne entre le modèle initial et amélioré sur la base de BindingDB quand a remarqué que l'évaluation des graphes des pertes quadratiques moyennes (d'entrainement/de test) est presque similaire, mais la différence fondamentale réside dans les valeurs de perte. Les pertes du modèle initial varient de 28.041 à 1.412 pour l'entraînement et de 19.129 à 1.206 pour le test, tandis que les pertes du modèle amélioré varient de 19.716 à 0.281 pour l'entraînement et de 8.783 à 0.858 pour le test. Cela signifie que la capacité de prédiction de notre modèle est meilleure par rapport au modèle initial. Une autre différence réside dans l'évaluation des graphiques après la deuxième phase (60-100 époques). Les pertes d'entraînement de notre modèle deviennent inférieures

aux pertes de test, contrairement au modèle initial qui ne présente pas ce comportement. Cela indique qu'il n'y a pas de surapprentissage dans notre modèle.

- Quand on a comparé entre la figure VI-9 et la figure VI-13 on observe que Dans le modèle initial, la perte commence à un niveau élevé d'environ 80, où on trouve dans le modèle amélioré, la perte initiale est d'environ 50, aussi en remarque que les deux modèles montrent une convergence rapide de la perte dans les premières étapes d'entraînement (environ les 20 premiers époques). Mais Le modèle amélioré converge plus rapidement et de manière plus stable vers une perte plus faible. Dans les deux cas, l'écart entre les pertes d'entraînement et de test est minime, indiquant un bon ajustement des modèles mais le progrès du modèle amélioré est visible à cause de la superposition de la courbe de test et la courbe d'entrainement. On ajoute à tout ça que la valeur finale dans le modèle initial est plus grande que le modèle amélioré quelle était presque nulle indiquant une performance beaucoup meilleure. Enfin, par rapport au modèle initial, le modèle amélioré a considérablement réduit la perte initiale, une vitesse de convergence plus rapide et une meilleure capacité de généralisation. Ces améliorations sont visibles à travers des valeurs MSE plus faibles et des courbes plus stables.

A la fin on compare les figures (VI-12 et VI-14) avec les figures (VI-8 et VI-10) :

- On commence par la comparaison entre la figure VI-10 et la figure VI-14 ou en observe que : Il est clair que le modèle amélioré surpasse le modèle initial. Le graphe initial montre que les points autour de la diagonale sont très dispersés, avec de nombreuses valeurs aberrantes (outliers), ce qui indique que les prédictions sont moins précises et moins fiables. Les points de ce tracé ne sont pas bien alignés sur la droite y = x, révélant des erreurs systématiques dans le modèle initial. En revanche, le deuxième graphique représentant le modèle amélioré présente une meilleure concentration de points le long de la diagonale, indiquant que les prédictions sont plus cohérentes avec les valeurs réelles. La dispersion diminue et il y a moins de valeurs aberrantes. Cela indique que le modèle amélioré fournit des prédictions plus précises et des performances globales supérieures par rapport au modèle initial

- A partir des figures VI-10 et VI-14 les points suivants ont été soulevés :

  ✓ Sur le graphique du modèle initial, on observe une plus grande dispersion des points, avec des prédictions étalées sur une plus large gamme de valeurs pour des vraies valeurs similaires. Cela indique que le modèle initial a une variabilité plus élevée et est moins précis. En revanche, sur le graphique du modèle amélioré, les points sont plus concentrés autour de la diagonale idéale (ligne y = x), ce qui montre une meilleure correspondance entre les valeurs prédites et les valeurs réelles.

  ✓ Le modèle initial montre des signes d'erreurs systématiques, visibles par la dispersion verticale plus large. Cela suggère que pour certaines valeurs réelles, les prédictions peuvent varier considérablement, indiquant un biais ou une erreur non-aléatoire dans les prédictions. Quand le modèle amélioré présente moins de dispersion verticale, suggérant une réduction des erreurs systématiques et une amélioration de la fiabilité des prédictions.

  ✓ Le modèle initial semble également prédire des valeurs qui sont parfois très en dehors de la gamme des valeurs réelles observées (par exemple, certaines prédictions inférieures à 4 ou supérieures à 14 pour des vraies valeurs allant de 8 à 16). Cependant que le modèle, a une gamme de prédictions qui correspond mieux à celle des valeurs réelles, réduisant les cas extrêmes d'erreurs.

  ✓ Dans le modèle initial, les erreurs semblent plus aléatoires et dispersées, ce qui peut être le signe d'un modèle sous-ajusté qui n'a pas capturé suffisamment bien la structure des données. Au point ou Le modèle amélioré, avec des prédictions plus concentrées, indique une meilleure capture de la structure des données et moins de variabilité aléatoire dans les erreurs.

En conclusion, le modèle amélioré montre une nette amélioration par rapport au modèle initial en termes de précision des prédictions, réduction des erreurs systématiques et meilleure capture de la structure des données. Les points sur le graphique du modèle amélioré sont plus alignés le long de la diagonale, indiquant que les valeurs prédites sont plus proches des valeurs réelles, ce qui est un signe de performance améliorée du modèle.

Le modèle amélioré démontre une précision nettement supérieure par rapport au modèle initial. Ses critères de validation sont remarquables, offrant des résultats plus précis et fiables, étroitement alignés avec les valeurs réelles. Les erreurs associées au modèle amélioré sont

extrêmement minimes, soulignant son excellence et sa grande précision. Ces améliorations reflètent une avancée significative, établissant le modèle amélioré comme une solution hautement efficace et fiable pour des prédictions exactes.

VI.4.2. Entre KIBA et BindingDB

Tableau VI-12: *l'ordre des substances pour chaque une des bases*

Ordre	***N°***	***Nom de la Substance***	***Score KIBA***	***Score BindingDB***
1	16	Camptothecin,(N-Diethyl) Glycinate	8,04933	0,3605653
2	44	Guanazole	9,99481	2,47373462
3	68	Vinblastine Sulfate	10,5375	5,10027885
4	43	Uracil	10,5524	6,57162476
5	7	Epsilon,-Rhodomycinone	10,6023	0,35984197
6	42	Fluorodopan	10,6383	5,65895367
7	21	9-Methoxycamptothecin	10,6542	5,13480139
8	9	Bikaverin	10,6679	0,38388044

9	30	Anthra[1,9-Cd]Pyrazol-6(2H)-One Der	10,8525	6,87598562
10	32	Diaziquone	10,8543	5,59254646
11	14	9-Aminocamptothecin (R,S)	10,8944	0,34431145
12	34	Carmustine	10,9043	0,33180079
13	63	3-Demethylthiocolchicine	10,9185	3,36981797
14	53	Mitozolomide	10,9424	0,32980016
15	61	Teroxirone	10,9481	3,30880785
16	66	Triethylenemelamine	10,9609	4,47378111
17	1	Deoxydoxorubicin	10,9611	9,19877625
18	64	6-Thioguanine	10,9929	5,32604408
19	58	Pyrazoloimidazole	11,0179	0,35494924
20	28	L-AsparticAcid	11,045	5,3669343

21	67	Tritylcysteine	11,0454	0,33165923
22	22	Chlorambucil	11,065	5,39096594
23	45	Hepsulfam	11,0656	6,50396919
24	12	Dermocybin	11,0665	0,35990009
25	17	Camptothecin	11,08	3,94192576
26	8	4BETA-Hydroxywithanolide	11,0879	0,37531453
27	2	Menogaril	11,0925	0,35879594
28	13	Pyrozofurin	11,0964	0,36356202
29	20	CamptothecinethylglycinateEsterhydrochloride	11,0974	4,53851509
30	15	11-Hydroxymethyl-20(Rs)-Camptothecin	11,1064	0,34433419
31	47	Hydroxyurea	11,1272	0,36714333
32	23	Clomesone	11,1413	4,4743824

33	71	N-Benzoyl-Deacetylcolchicine	11,1458	4,22856998
34	40	Dianhydrodulcitol	11,1517	0,3691158
35	25	Dihydro-5-Azacytidine	11,173	3,8763299
36	62	5-((4-Chlorobenzyl)Thio)-3-(Trifluoromethyl)-1H-1,2,4-Triazole	11,1907	0,34272733
37	4	Dichloroallyl Lawson	11,1952	0,36513567
38	48	AlanosineMonosodium Salt	11,2033	4,88287401
39	57	Pyrazoloacridine Mesylate	11,2254	1,78313553
40	35	Bisantrene Hydrochloride	11,2363	6,75654125
41	10	Plumbagin	11,2908	0,34603333
42	70	Colchicine	11,3017	0,33010471
43	54	Urea	11,3175	3,97572422
44	6	Arnebin 1	11,3221	0,36306292

45	59	Spiromustine	11,3721	2,82664537
46	69	Vincristine Sulfate	11,3957	1,94918978
47	31	Cytosine, Monohydrochloride	11,4256	6,62574053
48	65	Thiotepa	11,4333	0,32798705
49	51	Semustine	11,4475	4,32834053
50	33	Triazinate	11,452	0,35797048
51	26	Acivicin	11,4686	3,37003899
52	60	Paclitaxel	11,4993	0,3587341
53	56	Profiromycin	11,5037	0,33623487
54	29	Amonafide	11,5182	3,75836015
55	11	Shikalkin	11,5226	0,36206624
56	39	Lomustine	11,5353	0,28912783

57	27	Alpha,-Thiodeoxyguanosine	11,5472	4,33677101
58	36	Busulfan	11,5508	5,62460375
59	24	Picolinaldehyde	11,5737	4,87412643
60	52	Mitomycin C	11,7088	6,42943192
61	19	Camptothecin Butylglycinate Ester Hydrochloride	11,7874	1,87555826
62	55	Pipobroman	11,8378	1,60093498
63	41	Ac1l2oas	11,8468	6,01961517
64	50	Methotrexate	11,8636	7,13404608
65	49	Amsacrine	12,1402	5,87026882
66	37	Camptothecin, Acetate	12,299	0,32258973
67	5	Rhodomycin A	12,5039	0,36313567
68	46	Hycanthone Mesylate	12,6835	4,52983713

69	38	10-Hydroxycamptothecin	12,8566	0,33825901
70	18	Campothecin Analog	13,4713	6,09250355
71	3	N,N-Dibenzyldaunorubicin Hydrochloride	13,6612	7,72992563

D'après le tableau VI-9, on remarque qu'il y a une similarité entre les deux. Cela apparaît dans la correspondance des molécules comme la molécule 9-Aminocamptothecin (R,S) ou la molécule Busulfan qu'on trouve dans l'ordre 11 et 58 dans les prédiction et on peut trouver aussi des molécules qui sont trop proches comme les molécules Urea , colchicine et la molécule N,N-Dibenzyldaunorubicin Hydrochloride qu'on a trouvé respectivement dans l'ordre 42 / 43/ 71 dans les prédiction obtenue de la base KIBA et dans l'ordre 43/42/70 dans les prédiction obtenue de la base de BindingDB et on note que c'est un bon signe, cela nous montre que le pourcentage d'erreur dans nos prédictions est presque nul aussi pour des molécules comme Lomustine et Camptothecin, acétate qui atteignent une affinité plus importante au niveau de la base de BindingDB, mais une affinité un peu faible dans KIBA. Cela est dû aux déférences suivantes :

Caractéristiques	**BindingDB**	**KIBA**
Spécificité	Généraliste pour diverses cibles biologiques	Spécifique aux kinases
Scores-Y	Mesures individuelles Kd	Score composite Y intégrant Ki, Kd, IC50

- ✓ La base de données BindingDB et KIBA sont deux bases qui se concentrent sur les interactions entre les protéines et les ligands mais la base de BindingDB vise à être plus généraliste pour diverses cibles biologiques avec une large gamme de protéines et de ligands. Le contraire de la base KIBA qui est une base spécifique aux interactions avec les kinases le même type de notre cible V600E-BRAF c'est-à-dire la base est un peu plus petite par rapport BindingDB et plus précise

- ✓ Aussi on ajoute que le score dans BindingDB est un score individuel basé juste sur le Kd par contre dans KIBA le score intégrant Ki, Kd et IC50 ce qui signifie la fiabilité des résultats de KIBA
- ✓ Les prédictions de BindingDB basé sur l'affinité générale pour BRAF-V600E, en revanche que les résultats de KIBA testés spécifiquement l'activité inhibitrice des composés pour cette mutation donc en peut dire que la base BindingDB est efficace dans le domaine chimique et KIBA pour le domaine médicinales .

Donc on peut conclure que cette déférence entre les deux bases est normale et que les résultats sont fiables pour les deux bases mais les prédictions de KIBA sont plus fiables et précises

VI.4.3. Entre la factorisation matricielle et le QSAR

Ce travail a été réalisé en utilisant une autre méthode, celle de l'établissement d'un modèle QSAR (Quantitative Structure-Activity Relationship) pour prédire l'efficacité des nouvelles molécules. L'objectif était de concevoir un outil qui pouvait prédire l'activité inhibitrice des composés de notre base de test en utilisant une machine à vecteurs de support (SVM) basée sur des descripteurs moléculaires. Les données proviennent de PubChem et ont conduit à la génération de descripteurs moléculaires en 2D. Une sélection minutieuse a permis d'identifier 10 descripteurs pertinents pour 71 molécules. [54]

Pour bien comprendre les différences entre les deux méthodes (factorisation matricielle/ QSAR), nous avons effectué une comparaison qui mis en évidence leurs caractéristiques respectives :

- Base de données d'entraînement et taille des données :
  - ✓ Factorisation matricielle : Dans notre modèle, nous avons utilisé deux types de données comme base d'entraînement. La première, BindingDB (10656 médicaments et 1413 cibles), contient plusieurs médicaments de différentes classes thérapeutiques et évalue des milliers d'interactions complexes entre ces médicaments et leurs cibles. La seconde, KIBA (2069 médicaments et 230 cibles, se compose de médicaments spécifiques pour inhiber des cibles de type kinase, évaluant des milliers d'interactions entre différents types de kinases et leurs inhibiteurs.
  - ✓ QSAR : Dans ce modèle, pour prédire l'activité inhibitrice de composés chimiques, il est nécessaire de déterminer les descripteurs pertinents de ces

molécules. À cette fin, dix descripteurs pertinents ont été identifiés pour les soixante-onze molécules de la base d'entraînement.

- La complexité des modèles :
  - ✓ Factorisation matricielle : Pour mieux gérer notre grande matrice de médicaments-cibles avec leurs interactions, l'optimiseur le plus adapté aux paramètres du modèle est le SGD (**Stochastic Gradient Descent**). Cet optimiseur est très efficace pour capturer les relations complexes entre les médicaments et les cibles.
  - ✓ QSAR : Ce modèle a été développé en utilisant SVM (Support Vector Machines), qui est souvent utilisé pour prédire l'activité biologique dans les modèles QSAR. Les SVM sont un ensemble de techniques d'apprentissage supervisé destinées à résoudre des problèmes de discrimination et de régression.[54]
- Précision des modèles :
  - ✓ Factorisation matricielle : La précision d'un modèle de factorisation matricielle dépend de la taille et de la richesse de la base de données d'entrainement. Les deux bases de données utilisées dans notre modèle sont très volumineuses, chacune contenant différentes paires de cibles médicamenteuses en fonction de leur type (BindingDB/KIBA).
  - ✓ QSAR : La précision des modèles QSAR dépend de la sélection appropriée de descripteurs moléculaires pertinents pour l'activité biologique cible. Dans ce cas, les dix descripteurs pertinents et représentatifs ont été identifiés pour le modèle.
- Les critères de validation des modèles :
  - ✓ Factorisation matricielle : Nous avons obtenu les paramètres optimaux de notre modèle et l'avons validé par les valeurs de [MSE, MAE, RMSE, MAPE %] respectivement égales à [0,8558 ; 0,459 ; 0,925 ; 9,119] pour BindingDB et [0,3184 ; 0,3411 ; 0,5643 ; 2,8295] pour KIBA. Ces résultats démontrent une excellente performance du modèle.
  - ✓ QSAR : le modèle QSAR a été validé par les valeurs des R_2 = [0,9613, 0,8230, 0,9428] et des RMSE = [0,3275, 0,5027, 0,3769], ce qui démontre une excellente performance.

- Les molécules identifiées ayant une forte activité inhibitrice contre la mutation BRAF-V600E pour chaque méthode utilisée sont :
  - ✓ Factorisation matricielle : Lomustine, Camptothecin Acetate, Thiotepa, Camptothecin (N-Diethyl) Glycinate, Guanazole et Vinblastine Sulfate.
  - ✓ QSAR : Thiotepa, Arnebin 1 et Hycanthone Mesylate.

D'après cette description :

- Les deux modèles sont adaptés à la prédiction des interactions médicaments-cibles.
- La méthode QSAR est basée sur les propriétés chimiques, tandis que la factorisation matricielle repose sur une grande matrice comportant plusieurs interactions médicaments-cibles.
- La taille limitée de la base d'entraînement du QSAR, comparée à celle plus grande de la factorisation matricielle, favorise la capture de nouvelles interactions potentielles entre médicaments et cibles dans le modèle de factorisation matricielle.
- Il n'est pas possible de comparer directement les critères de validation des deux modèles uniquement sur la base de ces valeurs, car chaque modèle a ses propres critères spécifiques en fonction des exigences du modèle. De plus, les critères utilisés pour évaluer le modèle FM (MSE, MAE, RMSE et MAPE) sont plus nombreux que ceux utilisés pour évaluer le modèle QSAR (R2, RMSE). Cela suggère que le modèle FM est potentiellement plus performant que le modèle QSAR.
- Les méthodes QSAR utilisent des descripteurs pertinents pour modéliser l'activité biologique. La sélection et le calcul de ces descripteurs peuvent être complexes et gourmands en ressources, ce qui affecte directement la précision du modèle, tandis que, la factorisation matricielle n'a pas ce problème.
- Les deux modèles identifier la molécule 65 « Thiotepa » permis les meilleures molécules.

À partir de cette comparaison, on peut conclure que les résultats du modèle QSAR, basés sur une petite base d'entraînement de soixante-onze molécules et leurs caractéristiques chimiques, sont limités. En revanche, les résultats de la factorisation matricielle (FM), basés sur une base de données beaucoup plus grande (mille fois plus que celle de QSAR), sont meilleurs et aident à découvrir de nouvelles interactions potentielles. Ainsi, bien que les

résultats du QSAR soient raisonnables, ceux de la factorisation matricielle sont plus performants

CONCLUSION GENERALE :

L’objectif de cette étude était d’apporter une contribution à la recherche sur le mélanome et la découverte de nouveaux traitements potentiels. Les conclusions tirées sont résumées comme suit :

Le défi persistant posé par le mélanome en raison de ses changements génétiques complexes et de la diversité de ses mécanismes de défense. Cela met en évidence la nécessité continue de rechercher de nouvelles stratégies thérapeutiques pour cette maladie dont l'objectif de prédire des inhibiteurs plus efficaces en utilisant des techniques d'interaction médicament-cible (DTI) par l'approche de filtrage collaboratif méthode de factorisation matricielle.

Les résultats des critères de validation, ont permis de mettre en évidence les excellentes performances obtenues et les améliorations de précision réalisées. Les résultats de test fiables et proches des valeurs réelles renforcent la confiance dans la validité de prédictions

L’identification des molécules utilisées pour le traitement d'autres cancers qui ont également une grande affinité avec la cible V600E-BRAF, suggérant qu'elles pourraient être de nouveaux médicaments potentiels pour le traitement du mélanome.

Cette découverte ouvre de nouvelles pistes pour la recherche de médicaments contre le mélanome. Il est à souligner l'importance de la base de données soigneusement collectée,f qui contient plusieurs dérivés de la même molécule, pour mieux comprendre les fonctions qui influencent l'affinité de la molécule avec la cible. Cette compréhension améliorée des interactions médicament-cible est cruciale pour le développement de thérapies plus efficaces et ciblées.

En résumé, cette étude représente une avancée significative dans la recherche sur le mélanome et offre des perspectives prometteuses pour le développement de nouveaux traitements. Les résultats et les conclusions présentés renforcent la validité de notre approche et soulignent son potentiel pour faire progresser les soins aux patients atteints de mélanome.

Bibliographie

1. Bertolotto, C. Cutaneous and uveal melanoma: Two different cancers in therapeutic needs. *Comptes Rendus - Biol.* **344**, 219–231 (2021).
2. Ferlay, J. Colombet, M., Soerjomataram, I., Parkin, D.M., Piñeros, M., Znaor, A.Cancer statistics for the year 2020: An overview. Int. J. Cancer149, 778–789 (2021).
3. Kłos, P. Plant-Derived Terpenoids : A Promising Tool in the Fight against Melanoma. (2022).
4. Kozyra, P., Krasowska, D. Pitucha, M. New Potential Agents for Malignant Melanoma Treatment— Most Recent Studies 2020–2022. *Int. J. Mol. Sci.* **23**, (2022).
5. Umar, A. B., Uzairu, A., Shallangwa, G. A. Uba, S. Computational evaluation of potent 2-(1H-imidazol-2-yl) pyridine derivatives as potential V600E-BRAF inhibitors. *Egypt. J. Med. Hum. Genet.* **21**, (2020).
6. Yue, Y. He, S. DTI-HeNE: a novel method for drug-target interaction prediction based on heterogeneous network embedding. *BMC Bioinformatics* **22**, 1–20 (2021).
7. Boudraa, A. O., Cexus, J. C., Salzenstein, F. Guillon, L. If estimation using empirical mode decomposition and nonlinear Teager energy operator. *Int. Symp. Control. Commun. Signal Process. ISCCSP* 45–48 (2004) doi:10.1109/isccsp.2004.1296215.
8. Maxence, E. Les causes du cancer. **1900**, 21 (2012).
9. Balacheff, N. & Balacheff, N. Didactique et intelligence artificielle To cite this version : HAL Id : hal-00190648 Didactique et Intelligence artificielle. (2007).
10. Wreede, L., Schetelig, J. & Putter, H. Analysis of survival outcomes in haematopoietic cell transplant studies: Pitfalls and solutions. *Bone Marrow Transplant.* **57**, 1–7 (2022).
11. Pichard, E. La douleur en cancérologie. *Soins.* 44–47 (1989).
12. François, L. Développement d'un modèle prédictif de réponse thérapeutique et de toxicité dans les cancers du col de l'utérus: de l'imagerie à la clinique». 1–241 (2021).
13. Stoeckle, E. Surgery of limb and trunk wall soft tissue sarcoma. *Cancer/Radiotherapie* **10**, 34–40 (2006).
14. Santos J.S., Costa, M.C.O., Sobrinho, C.L.N., Silva, M.C.M., Souza, K.P., Melo, B.O. Anthropometric profile and food intake of adolescents in Teixeira de Freitas - Bahia, Brazil. Rev. Nutr.18, 623–632 (2005).
15. Ligue contre le cancer. Les traitements des cancers. (2011).
16. Pérol, M., Arpin, D. Maintenance therapy in advanced non-small cell lung cancer: A new paradigm? *Press. Medicale* **40**, 404–414 (2011).
17. Corde. S. Développement méthodologique de l'application d'agents pharmacologiques renforçateurs de l'effet photoélectrique pour l'utilisation du rayonnement synchrotron en radiothérapie anticancéreuse. Ingénierie biomédicale. Université Joseph-Fourier-

Grenoble I, (2002).

18. Pignot, G., Sargos, P. Adjuvant radiotherapy after radical cystectomy for muscle-invasive bladder cancer. *Prog. en Urol.* **31**, 158–168 (2021).

19. Mongis, A. Une nouvelle stratégie d ' immunothérapie : cibler directement des immunostimulants à la surface des cellules tumorales par ligation bio-orthogonale Aline Mongis HAL Id : tel-01713120. (2017).

20. Burstein HJ, Lacchetti C, Anderson H, Buchholz TA, Davidson NE, Gelmon KA, Giordano SH, Hudis CA, Solky AJ, Stearns V, Winer EP, Griggs JJ. Adjuvant Endocrine Therapy for Women With Hormone Receptor-Positive Breast Cancer: ASCO Clinical Practice Guideline Focused Update. J Clin Oncol.32, 2255–2269 (2014).

21. Bruland, Ø. S. Cancer therapy with radiolabeled antibodies an overview. *Acta Oncol. (Madr).* **34**, 1085–1094 (1995).

22. Wyman v. Eck Clerk's Record Dckt. 43730 Idaho Supreme Court Records & Briefs. 6147. (2016).

23. McConville BJ, Steichen-Asch P, Harris R, Neudorf S, Sambrano J, Lampkin B, Bailey D, Fredrick B, Hoffmann C, Woodman D. Pediatric bone marrow transplants: psychological aspects. Can J Psychiatry.35(9):769-75. (1990).

24. Sangaré, F. R. Etude des aspects épidémiologiques, cliniques et thérapeutiques des cancers gynécologiques et mammaires au centre de santé de référence de la Commune V du district de Bamako. 87 (2003).

25. Bleunven, A. Contribution à la conception d ' un système d ' imagerie polarimétrique en vue d ' applications pour la détection précoce du mélanome To cite this version : HAL Id : tel-01523771 Contribution à la conception d ' un système d ' imagerie polarimétrique en v. (2017).

26. Aladji, M. K. Théodore, M. H. C. Caractérisation optique du Mycobacterium Ulcerans dans le cadre du diagnostic précoce de l ' ulcère de Buruli Remerciements. (2022).

27. Carcassonne, C. H. Pharmacodynamie Cibles des médicaments. (2017).

28. MOXIBUSTION, M. F.-A. E. & 2007, undefined. L'harmonisation du yinqiao/yangqiao est equivalente au traitement par clonazepam dans les insomnies. *Meridiens.Org* **6**, (2007).

29. Sellami, A. Utilisation des méthodes de criblage virtuel dans un contexte de santé humaine et environnementale : application aux récepteurs nucléaires et aux perturbateurs endocriniens. 283 (2023).

30. Mancini, A. D. & Poitout, V. Les récepteurs membranaires des acides gras de la cellule β: De nouvelles cibles thérapeutiques pour le traitement du diabète de type 2? *Medecine/Sciences* **29**, 715–721 (2013).

31. Li, C., Lee, K. C., Schneider, E. B. & Zeiger, M. A. BRAF V600E mutation and its association with clinicopathological features of papillary thyroid cancer: A meta-analysis. *J. Clin. Endocrinol. Metab.* **97**, 4559–4570 (2012).

32. da Silva, R. C. *et al.* BRAF overexpression is associated with BRAF V600E mutation

in papillary thyroid carcinomas. *Genet. Mol. Res.* **14**, 5065–5075 (2015).

33. Ricci, F. Recommender Systems: Models and Techniques. *Encycl. Soc. Netw. Anal. Min.* 2147–2159 (2018) doi:10.1007/978-1-4939-7131-2_88.

34. Picot-clémente, R. Une architecture générique de Systèmes de recommandation de combinaison d'items. Application au domaine du tourisme. Thèse de doctorat. 'Université de Bourgogne , France (2012).

35. Dahimene, M.R. Filtrage et Recommandation sur les Réseaux Sociaux. Conservatoire national des arts et metiers- CNAM, (2014).36. Soualah-AF., CAMLearn*: Une architecture de système de recommandation sémantique sensible au contexte. Application au domaine du m-learning.VL.Thèse de doctorat. (2006).

37. Bokde, D. K., Girase, S. ,Mukhopadhay, D. Role of Matrix Factorization Model in Collaborative Filtering Algorithm : A Survey. 4th Int. Conf. Adv. Comput. Commun. Control 1–12 (2015).

38. Alchiekh Haydar, C Les systemes de recommandation ` a base de confiance. Thèse de doctorat, Université de Lorraine (2014).39. Moro-Visconti, R. Artificial Intelligence Scalability: a Multilayer Network Interpretation, (2024).

40. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J, Yu B, Zhang J, Bryant SH. PubChem Substance and Compound databases. Nucleic Acids Res. 44, D1202–D1213 (2016).

41. Doloksaribu, F., Suaka, I. Y. , Yoku, F. N. Peningkatan Pemahaman Konsep Dan Keterampilan Sketsa Kimia Melalui Pelatihan Aplikasi Chemsketch Acd / Lab Bagi Guru-Guru Kimia bervariasi pada kelompok atau kategori yang sama . Penguasaan konsep peserta didik didukung merasa perlu untuk melatihkan aplik. 5, 101–106 (2024).

42. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Res. 44, D1045–D1053 (2016).

43. Liu, T., Lin, Y., Wen, X., Jorissen, R. N., Gilson, M. K. BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* **35**, 198–201 (2007).

44. Tang, J., Szwajda, A., Shakyawar, S., Xu, T., Hintsanen, P., Wennerberg, K., Aittokallio, T.Making sense of large-scale kinase inhibitor bioactivity data sets: A comparative and integrative analysis. J. Chem. Inf. Model.54, 735–743 (2014).

45. Shuel S.L. Therapies ciblées contre le cancer: Perles cliniques à l'intention des soins primaires. Can Fam Physician. 68, 199–203 (2022).

46. Bagherian M, Sabeti E, Wang K, Sartor MA, Nikolovska-Coleska Z, Najarian K. Machine learning approaches and databases for prediction of drug-target interaction: a survey paper. Brief Bioinform. 22(1):247-269. (2021).

47. Sajadi, S. Z., Zare Chahooki, M. A., Tavakol, M. Gharaghani, S. Matrix factorization with denoising autoencoders for prediction of drug–target interactions. *Mol. Divers.* **27**, 1333–1343 (2023).

48. Zheng, X., Ding, H., Mamitsuka, H. Zhu, S. Collaborative matrix factorization with multiple similarities for predicting drug-Target interactions. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* **Part F1288**, 1025–1033 (2013).

49. Sujata, D., Biswa, R A., Mamta, M., Ajith, A., Arpad, K. *Deep Learning Techniques for Biomedical and Health Informatics*. (2020). doi:10.1016/c2018-0-04781-7.

50. Madani, A., Benkortbi, O. Laidi, M. In silico prediction of the inhibition of new molecules on SARS-CoV-2 3CL protease by using QSAR: PSOSVR approach. *Brazilian J. Chem. Eng.* **41**, 427–442 (2024).

51. Rehailia R. Une approche basée sur la prédiction de liens pour les systèmes de recommandation Mémoire de Fin d'études Master, Université de 8 Mai 1945 Guelma (2021).

52. Tanwar, H. Kakkar, M. Performance comparison and future estimation of time series data using predictive data mining techniques. *2017 Int. Conf. Data Manag. Anal. Innov. ICDMAI 2017* **4**, 9–12 (2017).

53. Umar, A. B., Uzairu, A., Shallangwa, G. A. Uba, S. QSAR modelling and molecular docking studies for anti-cancer compounds against melanoma cell line SK-MEL-2. *Heliyon* **6**, e03640 (2020).

54. Metidji, C., Djabali, F. Utilisation de la méthode QSAR –SVMDA et le Docking pour la prédiction de l'activité inhibitrice des inhibiteurs et l'évaluation des interactions entre le récepteur et le ligand du Cancer de la peau. Mémoire de fin d'études de Master, Universite Yahia Fares de Medea (2023).