DEMOCRATIC AND POPULAR REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
UNIVERSITY YAHIA FARES OF MEDEA

Faculty of Technology

End of study Master's Dissertation

Field: Process Engineering
Option: Pharmaceutical Process Engineering
THEME

A Comparative Analysis of Artificial Intelligence Techniques for Modelling Drug Pharmacokinetics

Submitted by:

Ms. BOUKERMA Safaa

Ms. HADDOUDA Zineb

Supervised by:

Pr: KHAOUANE Latifa

Dr: IBRIR Abdellah

The defense will take place on **22/06/2024**

Academic year 2023/2024

الحمد لله الذي بنعمته
تتم الصالحات،
وبفضله تتنزل
الخيرات والبركات
وبتوفيقه تتحقق
المقاصد والغايات

Acknowledgments

First and foremost, we extend our deepest gratitude to Allah, the Almighty, whose guidance, courage, and strength have enabled us to successfully complete this work.

This endeavor owes much to the insightful guidance and meticulous supervision of Professor KHOUANE Latifa. We are profoundly grateful for her exceptional oversight, patience, rigor, and unwavering availability throughout the preparation of this dissertation.

We also wish to express our sincere appreciation to our co-supervisor, Dr. IBRIR Abdellah, for his invaluable assistance, enlightened advice, and constant encouragement.

Our heartfelt thanks go to Dr. HANTABLI Mohamed for his support and patience in addressing our numerous queries during this undertaking.

We are deeply indebted to the jury members for their gracious willingness to evaluate this work and for providing their invaluable feedback.

We extend our sincere gratitude to all our professors over the past five years in the Department of Process and Environmental Engineering at the Faculty of Technology, University of MEDEA.

Finally, we offer our profound appreciation to all the staff and administrators of our university, as well as to all those who have encouraged and supported us, both near and far, throughout the completion of this work.

Dedications

From the bottom of my heart, I dedicate this modest work:

To my lovely mother,

Who believed in me, who sacrificed for me. She has been my sister, my friend, my companion, my strength, and everything beautiful in my life. I thank her for all that she has done for me, for her early morning prayers for me. Thank you for all the love you give me, may God preserve and keep her for me.

To my dear father,

Who has always worked for my success and happiness. He has always been proud of me in my success or defeat and especially his unwavering trust in me throughout my life. May God preserve and keep him for me.

To my dear sister Marwa,

Who encouraged me on my days of despair and who has been the best sister, the best friend, and my lucky charm, thank you for being my sister.

To my dear brothers Walid and Taki,

Whom I deeply love and may ALLAH protect them.

To my little heart Dani whom I love dearly and my second sister Souad.

To my fiance Omer,

Who has encouraged me throughout my studies. I thank him for everything he has done for me. Thank you for all the love you give me, may God preserve and keep him for me.

To my Grand paternal family BOUKERMA and maternal family BOURAHLA without exception.

To everyone who has contributed to my education since my childhood.

or from afar Everyone who knows SAFAA, whether closely.

Not forgetting my best partner ZINEB for her moral support, patience, and understanding throughout this project.

SAFAA

Dedications

From the bottom of my heart, I dedicate this modest work :

To my lovely mother,

who sacrificed to shape me into the person I am today. Being a shoulder to lean on. Her early morning prayers for my well-being. I am endlessly grateful for her presence in my life, for the countless times she believed in me when I doubted myself. Her love is a treasure beyond measure. May God bless and protect her.

To my dear father,

for his endless sacrifices, from driving me to college to bringing joy into my life. His support have been a constant source of inspiration, tirelessly working for my success and happiness. May God protect and bless him always.

To my dear grandmother Zineb Elhachemi,

who always prayed for my success, and shared her invaluable life advice. May God protect her and keep her safe for us, granting her healing.

To my dear younger sister Yasmin,

wishing for her all successes in her studies

My confidant friend Aya,

my beloved second sister who has been by my side through every triumph and tribulation, offering unwavering support and companionship.

To the university dormitory employees of Ouzra,

for their commitment to creating a supportive and safe learning environment that greatly contributed to my academic success.

To my Grand paternal family HADDOUDA and maternal family TIFOUR.

To everyone who has contributed to my education since my childhood.

Not forgetting my best partner SAFAA and her family for their moral support, patience, and understanding throughout this project.

ZINEB

الملخص:

يهدف هذا العمل إلى التنبؤ بمعامل الحركية الدوائية الذي هو الرابطة البروتينية (FB)من خلال دراسة تحليلية مقارنة تركز على العلاقة بين الدواء والبروتينات، باستخدام أفضل 208مواصفات الجزيئية لـ 263 مادة نشطة مشتقة من RDKit كعينة دراسة.حيث تستخدم الدراسة إحدى تقنيات الذكاء الاصطناعي المتمثلة تحديداً في تقنيات التعلم الآلي، بما في ذلك الشبكات العصبية الاصطناعية (ANN) وآلة الدعم المتجهين المقترنة بخوارزمية اليعسوب (DA-SVM) ، المدربة على قاعدة بيانات مخفضة إلى25مواصفات جزيئية تم الوصول إليها من خلال خوارزمية الجينات (GA). كما تستخدم قاعدة البيانات هذه، البيانات المخفضة حصراً لنماذج ANNو DA-SVM.

حيث تم التوصل لوجود علاقة تفاعلية بين الدواء والبروتينات،خاصة البروتينات البلازمية مثل الألبومين والجليكو بروتين ألفا-١ الحمضي، مما يؤثر على توزيع الدواء وأيضه وإزالته. علاوة على ذلك، تم اكتشاف أساليب التعلم العميقة الجديدة وهي الشبكات العصبية التكرارية ذات الطبقات المتعددة مع الشبكات العصبية المتعرجة (CNN-MLP) والشبكات العصبية التكرارية مع DA-SVM (CNN-DA-SVM).

يظهر النموذج CNN-MLP بشكل ملحوظ قدرة تنبؤ ممتازة ومتينة، مع قيم المعايير لمعاملات تحديد R^2= 0.99و RMSE=0.0314. تسهم هذه الدراسة في تقدم النمذجة الحركية الدوائية وتسلط الضوء على إمكانيات تقنيات التعلم العميق في البحث الصيدلاني.

الكلمات المفتاحية: *الحركية الدوائية؛ الأدوية؛ الرابطة البروتينية (FB)؛ النمذجة؛ المواصفات جزيئية ؛ الذكاء الاصطناعي(RNA؛ DA-SVM؛ CNN-MLP؛ CNN-DA-SVM).*

Abstract:

This work aims to predict the pharmacokinetic parameter which is Protein Binding (FB) through a comparative analytical study focusing on the relationship between drug and proteins, using the best 208 molecular descriptors of 263 active substances derived from RDKit as a sample. The study uses an artificial intelligence technique, namely machine learning techniques, including artificial neural networks (ANN) and a support vector machine coupled with the dragonfly algorithm (DA-SVM), trained on a reduced database of 25 molecular descriptors accessed through the genetic algorithm (GA). This database also uses reduced data exclusively for ANN and DA-SVM models.

It has been found that there is an interactive relationship between the drug and proteins, especially plasma proteins such as albumin and alpha-1 acid glycoprotein, which affects the distribution, metabolism and elimination of the drug. Moreover, new deep learning methods have been discovered namely convolutional neural networks with multi-layer perceptron (CNN-MLP) and convolutional neural networks with DA-SVM (CNN-DASVM).

The CNN-MLP model remarkably shows excellent and robust prediction ability, with criteria values of determination coefficient R^2= 0.99 and RMSE= 0.0314. This study contributes to the advancement of pharmacokinetic modelling and highlights the potential of deep learning techniques in pharmaceutical research.

Key words: *Pharmacokinetics; Drugs; Protein binding (FB); Modeling; molecular descriptors; Artificial Intelligence (RNA; DA-SVM; CNN-MLP; CNN-DASVM).*

Résumé :

Ce travail vise à prédire les paramètres pharmacocinétiques lequel est la liaison protéique (FB) par une étude analytique comparative axée sur la relation entre le médicament et les protéines, en utilisant comme échantillon les meilleures 208 descripteurs moléculaires de 263 substances actives dérivées de RDKit. L'étude utilise une technique d'intelligence artificielle, à savoir des techniques d'apprentissage automatique, notamment des réseaux neuronaux artificiels (ANN) et une machine à vecteur de support couplée à l'algorithme de la libellule (DASVM), entraînée sur une base de données de 25 descripteurs moléculaires réduite accessible par l'algorithme génétique (GA). Cette base de données utilise également des données réduites exclusivement pour les modèles ANN et DA-SVM.

Il a été observé qu'il existe une interaction entre les médicaments et les protéines, notamment les protéines plasmatiques comme l'albumine et l'alpha-1 glycoprotéine acide, qui affecte la distribution, le métabolisme et l'élimination du médicament. En outre, de nouvelles méthodes d'apprentissage en profondeur ont été découvertes, à savoir les réseaux neuronaux récursifs multicouches avec réseaux neuronaux convolutifs (CNN-MLP) et les réseaux neuronaux récursifs avec DA-SVM (CNN-DASVM).

Le modèle CNN-MLP montre remarquablement une capacité de prédiction excellente et robuste, avec des valeurs des critères de coefficients de détermination R^2= 0,99et RMSE= 0.0314. Cette étude contribue à l'avancement de la modélisation pharmacocinétique et met en évidence le potentiel des techniques d'apprentissage automatique et d'apprentissage profond dans la recherche pharmaceutique.

Mots Clés : *Pharmacocinétique ; Médicaments ; Liaison protéine (FB) ; Modélisation ; Descripteurs moléculaires ; Intelligence artificielle (RNA ; DA-SVM ; CNN-MLP ; CNN-DA-SVM).*

Table of Contents

Table of Contents …..... I
Abbreviations …..... V
Nomenclature and symbol…..... IX
List of tables..... X
List of figures…..... XI
General introduction …..... 1
Bibliographic References..... 3

Part I: Conceptual Framework

I.1. Introduction 04
I.2. Pharmacokinetics…..... 04
I.3. Pharmacokinetics processes..... 05
I.3.1. Absorption..... 06
I.3.1.1. Routes of administration..... 06
I.3.1.1.1. Oral administration (PO)..... 06
I.3.1.1.2. Intravenous administration (IV)..... 06
I.3.1.1.3. Transdermal administration (TD)..... 07
I.3.1.1.4. Intramuscular administration (IM)..... 08
I.3.1.1.5. Rectal administration (PR)..... 09
I.3.1.1.6. Inhalation administration (IA)..... 09
I.3.1.1.7. Sublingual administration (SA)..... 10
I.3.1.2. Absorption Mechanism..... 10
I.3.1.2.1. Passive diffusion..... 10
I.3.1.2.2. Facilitated diffusion..... 11
I.3.1.2.3. Active diffusion..... 11
I.3.2. Distribution..... 12
I.3.2.1. Plasma protein binding (PPB)..... 12
I.3.2.2. Tissue binding (TB)..... 14
I.3.2.3. Fat as a Reservoir:..... 14
I.3.3. Metabolism..... 15
I.3.3.1. First-passe effect..... 16
I.3.4. Excretion (Elimination)..... 17
I.3.4.1. Renal excretion..... 17
I.3.4.2. Biliary excretion..... 18
I.3.4.3. Pulmonary pathway..... 19
I.4. Pharmacokinetics parameters 19
I.4.1. Bioavailability..... 19
I.4.1.1. Types of Bioavailability..... 20
I.4.1.1.1. Absolute Bioavailability..... 20
I.4.1.1.2. Relative Bioavailability..... 20
I.4.1.1.3. Bioequivalence..... 20
I.4.2. Volume of Distribution..... 21
I.4.3. Clearance..... 21
I.4.4. Half-life time..... 21

I.4.5. Protein Binding Fraction (FB) 21
I.4.5.1. Parameters of Drug-Protein Binding 21
I.4.5.1.1. Affinity 22
I.4.5.1.2. Binding Sites 22
I.4.5.1.3. Fraction of Binding (FB 23
I.4.5.2. Measuring the drug bound fraction by equilibrium dialysis Method 23
I.4.5.3. Factors Influencing Protein Binding 24
I.4.5.3.1. Physiologic Factors 24
I.4.5.3.2. Pathologic Factors 25
I.4.5.3.3. Drug-Induced Changes 25
I.5. Factors influencing Pharmacokinetics 26
I.5.1. Age 26
I.5.2. Gender 26
I.5.3. Chronical diseases 26
I.5.3.1. Diabetes 26
I.5.3.2. Heart failure 27
I.6. Conclusion 28
Bibliographic references 29

Part II: Methodologies, Results and Discussions.

II.1. Introduction 37
II.2. Methodology (Contribution I) 37
II.2.1 Numerical modelling using AI methods for the prediction of protein binding fraction based on descriptors variation 37
II.2. Database collection 37
II.2.3. Modelling of protein binding fraction using descriptors 39
II.2.3.1. Modelling by Artificial Neural Networks (ANN) 39
II.2.3.1.1. Calculation Methodology 40
II.2.3.1.2. Normalization of data 41
II.2.3.1.3. De-normalization of data 41
II.2.3.1.4. Design of Artificial Neural Network 41
II.2.3.2. Modelling by Support Vector Machine coupled with the Dragonfly Algorithm (DA-SVM) 41
II.2.4. Modelling of protein binding fraction using descriptors features 48
II.2.4.1. Modelling by Convolutional Neural Networks and Multilayer Perceptron(CNN-MLP) 48
II.2.4.1.1 The design of CNN-MLP 49
II.2.4.2. Modelling by Convolutional Neural Networks coupled with Dragonfly Algorithm-Support Vector Machine (CNN-DASVM) 51
II.2.5. Statistical Performance Criteria (Analysis and Evaluation of a Model) 51
II.3. Results and discussions (Contribution II) 53

II.3.1. Modelling of protein binding fraction using descriptors..... 53
II.3.1.1. Statistical study of protein binding fractions and descriptors..... 53
II.3.1.2. Prediction of protein binding fractions using artificial neural networks (ANN)..... 54
II.3.1.2.1. Choice of ANN parameters 54
II.3.1.2.2. Concept of feedforward neural network (MLP)..... 55
II.3.1.2.3. Results interpretation..... 60
II.3.1.3. Prediction of protein binding fractions using Support the hybrid Dragonfly algorithm coupled with Support Vector Machine (DA-SVM) 64
II.3.2. Modelling of protein binding fraction using descriptors features..... 69
II.3.2.1. Prediction of protein binding fractions using Convolutional neural network-Multiple layer perceptron (CNN-MLP)..... 69
II.3.2.1.1 Choice of CNN parameters..... 69
II.3.2.1.2. Results interpretation..... 72
II.3.2.2. Prediction of protein binding fractions using Convolutional Neural Networks coupled with Dragonfly Algorithm-Support Vector Machine (CNN-DASVM)..... 75
II.3.3 Comparative Study..... 79
II.3.3.1. A comparative study between models used descriptors (ANN and DA-SVM)..... 79
II.3.3.2. A comparative study between models used descriptors features (CNN-MLP and CNN-DASVM)..... 80
II.3.3.3. An overall comparative study (ANN, DA-SVM, CNN-MLP and CNN-DASVM)..... 81
II.3.4.User Interface of the Best Model (CNN-MLP)..... 84
II.4. Conclusion..... 85
Bibliographic references..... 86
General Conclusion and Perspectives..... 88
Annexe..... 90

Abbreviations

Abbreviation	Signification
ML	**M**achine **L**earning
DL	**D**eep **L**earning
AI	**A**rtificial **I**ntelligence
ADME	**A**bsorption, **D**istribution, **M**etabolism, **E**limination
QSPR	**Q**uantitative **S**tructure-**P**roperty **R**elationship
ANN	**A**rtificial **N**eural **N**etwork
AARD	**A**verage **A**bsolute **R**elative **D**eviation
RMSE	**R**oot **M**ean **S**quare **E**rror
SVMr	**S**upport **V**ector **M**achine **R**egression
DA	**D**ragonfly **A**lgorithm
DA-SVM	Support Vector Machine coupled with the Dragonfly Algorithm
CNNMLP	**C**onvolutional **N**eural **N**etworks with **M**ulti-**L**ayer
PK	**P**erceptron **P**harmaco**k**inetic
DDI	**D**rug-**D**rug **I**nteractions
PO	**P**er **O**s
pH	**H**ydrogen **P**otential
GIT	**G**astro **I**ntestinal **T**ract
IV	**I**ntra**v**enous
TD	**T**rans**d**ermal
TDD	**T**ransdermal **D**rug **D**elivery
TDDS	**T**ransdermal **D**rug **D**elivery **S**ystem
IM	**I**ntramuscular **A**dministration
PR	**P**er **R**ectum
IA	**I**nhalation **A**dministration
SA	**S**ublingual **A**dministration
MDIs	**M**etered-**D**ose **I**nhalers
DPIs	**D**ry **P**owder **I**nhalers
COPD	**C**hronic **O**bstructive **P**ulmonary **D**isease
ATP	**A**denosine **T**ri**p**hosphate
PPB	**P**lasma **P**rotein **B**inding
HSA	**H**uman **S**erum **A**lbumin
AAG	**A**lpha-1-**A**cid **G**lycoprotein

TB	**T**issue **B**inding
HSA	**H**uman **S**erum **A**lbumin
AAG	**A**lpha-1-**A**cid **G**lycoprotein
TB	**T**issue **B**inding
NSAIDs	**N**on-**S**teroidal **A**nti-**I**nflammatory **D**rugs
Vd	**V**olume of **d**istribution
C	**C**oncentration
$\mathbf{T}_{1/2}$	Half-life
FB	Protein **B**inding **F**raction
D	**D**rug **c**oncentration
DP	**D**rug-**P**rotein
NMR	**N**uclear **M**agnetic **R**esonance
FU	**U**nbound **F**raction
HPLC	**H**igh **P**erformance **L**iquid **C**hromatography
ED	**E**quilibrium **D**ialysis
GA	**G**enetic **A**lgorithm
Des	**Des**criptor
MLP	**M**ulti-**L**ayer **P**erceptron
GRBF	**G**aussian **R**adial **B**asis **F**unction
FC	**F**ully **C**onnected
Conv	**Conv**olutional
ReLU	**R**ectified **L**inear **U**nit
MSE	**M**ean **S**quared **E**rror
NRMSE	**N**ormalized **R**oot **M**ean **S**quared **E**rror
MAE	**M**ean **A**bsolute **E**rror
ARD	**A**bsolute **R**elative **D**eviation
Max	**Max**imum
Min	**Min**imum
Var	**Var**iance
STD	**St**andard **D**eviation
FFNN	**F**eed **f**orward **N**eural **N**etwork
OANN	**O**ptimized **ANN**
Exp	**Exp**erimental
Cal	**Cal**culated

Nomenclatures and Symboles

Lecture	Designation	Unit
R^2	Coefficient of determination	(-)
R	Coefficient of correlation	(-)
f_{up}	**U**nbound**f**raction in **p**lasma	(-)
FB	Protein**B**inding **F**raction	%
T	**T**emperature	K or °C
C	Regularizationparameter	(-)
Iter	Currentiterationnumber	(-)
r	Neighbourhood**r**adius	Mm
s	Number of individuals (dragonflies)	(-)
a	**A**ttractivenessparameter	(-)
c	Regularizationparameters	(-)
f	**F**itnessfunction	(-)
e	**E**xploration rate	(-)
1D	**1 D**imensions	(-)
2D	**2 D**imensions	(-)
3D	**3 D**imensions	(-)
K	**K**ernel	(-)
w	**W**eight	(-)
b_i	**Bi**ais	(-)
bp	Biasvector	(-)
Bz	Biasvector	(-)
V_d	**V**olume of **d**istribution	L
Cl	**Cl**earance	mL/min
$T_{1/2}$	**H**alf-life **T**ime	min or h
FU	**U**nbound**F**raction	%
Ka	**A**ffinityconstant	L/mol
DP	Concentration of drug-protein complex	mol/L
D	Concentration of drug	mol/L
BD	Concentration of bound drug	mol/L
FD	Concentration of free drug	mol/L
n	**N**umber of binding site	(-)

Greek Letters

Lecture	Designation	Unit
ε	Insensitivity coefficient	(-)
σ	Bandwidth of the Gaussian RBF Kernel	(-)
α	Slope	(-)
β	Y-intercept	(-)

Exponents/Indices

Lecture	Designation
$\mathbf{Y_{i\ cal}}$	Calculated Output
$\mathbf{Y_i}$	Experimental Input
ΔX	Step size
0	Initial

List ofTables

Number of table	Table title	Page
	Part I	
Table I.1	Routes of administration, bioavailability, and general characteristics	20
	Part II	
Table II.1	Ranges of initial parameters for DA-SVM	44
Table II.2	Statistical description of the entire dataset	53
Table II.3	Optimized ANN structure	54
Table II.4	Established neural network model {(Input layer – 1st hidden layer)}	57
Table II.5	Established neural network model {(1st hidden layer – 2nd hidden layer)}	58
Table II.6	Established neural network model {(2nd hidden layer - output layer)}	59
Table II.7	The coefficient and statistical errors of the predicted results (ANN) compared to the experimental results	64
Table II.8	Structure of the DA-SVM model optimized by the Dragonfly algorithm	65
Table II.9	The coefficient and statistical errors of the predicted results (DA-SVM) compared to the experimental results	69
Table II.10	The coefficient and statistical errors of the predicted results compared to the experimental results	75
TableII.11	The coefficient and statistical errors of the predicted results (DA-SVM) compared to the experimental results	79
Table II.12	Comparison between the CNN-MLP model presented and existing models.	83

List of Figures

Figure	Figure title	Page
	Part I	
Figure I.1	The four pharmacokineticsprocesses	05
Figure I.2	Simplified pharmacokinetic model of oral, transdermal, intravenous route of administration.	08
Figure I.3	Types of drugs diffusion	12
Figure I.4	Compositions of blood in an average person	13
Figure I.5	Metabolic alteration of drug	15
Figure I.6	Metabolism of active drugs and pro-drugs	16
Figure I.7	First-pass effect on IV, oral, sublingual	17
Figure I.8	Elimination of hydrophobic and hydrophilic drug molecule	19
	Part II	
Figure II.1	Representation of database division	38
Figure II.2	Feature Construction Flowchart	39
Figure II.3	Flowchart of the methodology for calculating Artificial Neural Networks (ANN)	40
Figure II.4	Flowchart of SVM	44
Figure II.5	Organigramme de algorithm Dragonfly	45
Figure II.6	Schematic representation of the DA-SVMr model applied to FB prediction	46
Figure II.7	Flowchart of the DA-SVM model	47
Figure II.8	Basic CNN architecture	48
Figure II.9	CNN-MLP architecture for the prediction of protein binding fractions	50
Figure II.10	Feedforward artificial neural network	55
Figure II.11	The difference in error for the numbers of neurons in each hidden layer	56
Figure II.12	Linear regression plot of ANN predicted data versus experimental data for the global phase database	60
Figure II.13	Linear regression plot of ANN predicted data versus experimental data for the learning phase database	61
Figure II.14	Linear regression plot of ANN predicted data versus experimental data for the Test phase database	62
Figure II.15	Linear regression plot of ANN predicted data versus experimental data for the validation phase database	63
Figure II.16	Linear regression plot of DA-SVM predicted data versus experimental data for the global phase database	66
Figure II.17	Linear regression plot of DA-SVM predicted data versus experimental data for the learning phase database	67
Figure II.18	Linear regression plot of DA-SVM predicted data versus experimental	68

	data for the Test phase database	
Figure II.19	Results of the CNN-MLP technique	71
Figure II.20	Linear regression plot of CNN-MLP predicted data versus experimental data for the global phase database	72
Figure II.21	Linear regression plot of CNN-MLP predicted data versus experimental data for the learning phase database	73
Figure II.22	Linear regression plot of CNN-MLP predicted data versus experimental data for the Test phase database	74
Figure II.23	Linear regression plot of CNN-DASVM predicted data versus experimental data for the global phase database	76
Figure II.24	Linear regression plot of CNN-DASVM predicted data versus experimental data for the learning phase	77
Figure II.25	Linear regression plot of CNN-DASVM predicted data versus experimental data for the Test phase	78
Figure II.26	The difference in terms of error and root mean square between the ANN and DA-SVM models	80
Figure II.27	The difference in terms of error and root mean square between the CNN-MLP and CNN-DASVM models	81
Figure II.28	The difference in terms of error and root mean square between the CNN-MLP, ANN, CNN-DASVM and DA-SVM models	82
Figure II.29	The Graphical User Interface of the CNN-MLP Model	84

GENERAL INTRODUCTION

General Introduction

In recent years, there has been a growing interest in understanding the molecular processes governing drug behaviour in vivo. The binding of drugs to plasma proteins, such as serum albumin and alpha-1 acid glycoprotein, is a well-established pharmacokinetic parameter, with plasma proteins controlling the free drug concentration in plasma (Trainor, 2007a). It is widely accepted that a drug's therapeutic effect correlates more closely with a patient's exposure to the unbound fraction of the drug in plasma than with the total concentration, potentially reducing the availability of free concentration for therapeutic action, especially for drugs with high binding to plasma proteins(Kratochwil and al., 2002). The importance of plasma protein binding in modulating the effective drug concentration at pharmacological target sites has been a topic of significant discussion and debate among drug development groups over the past few decades (Trainor, 2007b).

In parallel, the advent of Machine Learning (ML) and Artificial Intelligence (AI) emerged as a major tool for modelling of *in vivo* animal and human pharmacokinetics (Olga Obrezanova, 2023), to enhance drug development and personalized medicine. Understanding the intricate dynamics of drug Absorption, Distribution, Metabolism, And Excretion (ADME) is critical for optimizing drug efficacy and minimizing adverse effects.

This master's thesis undertakes a comprehensive comparative analysis of various modelling approaches. Specifically, the study focuses on employing various Artificial Intelligence (AI) techniques to predict protein Binding Fractions FB using descriptors of active substances, including Artificial Neural Networks (ANN), and Support Vector Machine (SVM) with Dragonfly Algorithm (DA-SVM), alongside innovative deep learning techniques such as Convolutional Neural Networks with Multi-Layer Perceptron (CNN-MLP) and Convolutional Neural Networks with DA-SVM (CNN-DA-SVM), that uses descriptors features for the prediction of protein Binding Fractions (FB).

An extensive literature study was conducted on the application of artificial intelligence systems based on the use of descriptors as inputs. In 2023, Khaouane and co-authers developed QSAR models using a feed-forward neural network (FFNN) to predict the fraction bound to plasma proteins (fb) for 277 drugs, starting with 1666 descriptors from drug structures. Three descriptor selection methods were used: the Filter method selected 55 descriptors, the Wrapper method selected 15, and a

convolutional neural network (CNN) treated descriptors as image-like matrices to automatically extract essential features. The CNN method's robustness and reliability, with the CN-FC15-FFNN model (15 descriptors) significantly outperforming the CN-FC55-FFNN model (55 descriptors). The CNN method proved efficient, retaining all critical information and automating feature selection with R^2= 0.903, MSE= 0.011 and MAE= 0.066.

In 2013, Their paper titled “QSAR Models for the Prediction of Plasma Protein Binding”, Taravat and Zeshan. Predictedplasma Protein Binding (FB) usingQuantitative Structure–Activity Relationships (QSAR). Protein binding values for 794 compounds were collected. A total of 450 descriptors were calculated for each compound using ACD labs/logD, MOE, and Symyx QSAR software. Several data mining techniques, including stepwise regression analysis, Classification and Regression Trees (CART), Boosted Trees, and Random Forest, were employed to construct predictive models. The best-performing model was a boosted regression tree, with a mean absolute error (MAE= 13.25) and R^2= 0.71. The study concluded that while simple regression trees and multiple linear regressions provided reasonable accuracy, boosted regression trees offered the most precise FB predictions.

In 2005, Joseph R and al, presented a QSAR Modeling of Human Serum Protein Binding with Several Modeling Techniques Utilizing Structure-Information Representation. Their study aims to predict human serum protein binding (% bound) using 20 descriptors on a dataset of 1008 compounds. Three modeling techniques were employed, which are: Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), and Support Vector Machines (SVM), with ANN showing the best performance with R^2= 0.90 and MAE = 7.6.

This work consists of two parts: the first one is theoretical and the other is practical. Theoretical part is dedicated to a conceptual framework on pharmacokinetics, while the second part describes the methodology of each method of Ai used in this work, as well as the interpretation of the results obtained, and the discussion in detail, also a comparative study between the four techniques used for the prediction. The work concludes with a general conclusion that summarizes all the findings to outline potential future perspectives.

Bibliographic References

Kratochwil, N. A., Huber, W., Èller, F. M., Kansy, M., & Gerber, P. R. (2002). Predicting plasma protein binding of drugs: a new approach. Biochemical Pharmacology (pp.1355–1356).

Olga Obrezanova. (2023). Artificial intelligence for compound pharmacokinetics prediction. Current Opinion in Structural Biology (Vol. 79, pp.12–15).

Trainor, G. L. (2007a). Chapter 31 Plasma Protein Binding and the Free Drug Principle: Recent Developments and Applications. Annual Reports in Medicinal Chemistry (Vol. 42, pp. 489–502).

Trainor, G. L. (2007b). The importance of plasma protein binding in drug discovery. Expert Opinion on Drug Discovery (Vol. 2, Issue 1, pp. 51–64).

PART I

CONCEPTUAL FRAMEWORK

I.1. Introduction:

The term pharmacokinetics was first introduced by Franz Heinrich Dost* in 1953 in his text, Der BliitspiegeI-Kinetic der Konzentrationsablaiife in der Frieslauffliissigkeit (F. H. Dost, 1953; Wagner, 1981.). The purpose of pharmacokinetics is to study the time course of drug and metabolite concentrations in the biological fluids, tissues and excreta, then construct suitable models to interpret such data. In pharmacokinetics, the data are analysed using a mathematical representation of a part or the whole of an organism. Broadly then, the purposes of pharmacokinetics are to reduce data to a number of meaningful parameter values (the drug's bioavailability, half-life, clearance rate, and volume of distribution), and to use the reduced data to predict either the results of future experiments or the results of a host of studies which would be too costly and time-consuming to complete (Wagner, 1968, 1975).

I.2. Pharmacokinetics:

Pharmacokinetics (from Ancient Greek *pharmakon* "drug" and kinetikos "moving, putting in motion"), sometimes abbreviated as **PK**. Pharmacokinetics is the field of science that is aimed at studying the fate of any substance after administration to living organisms. In a sense, pharmacokinetics has always been a part of pharmacology; but it began to be considered a new discipline when more sophisticated methods were introduced to study the kinetic properties of drugs absorption, distribution, metabolism and elimination, including the rate and extent of each of these processes (ADME). These processes are the main determinants of the time course of the drug effect after drug administration. The pharmacokinetic principles are usually applied to drugs. However, the same principles can be applied to any compound such as nutrient, hormones, toxins, pollutants, pesticides and others (De Aldo Rescigno, 2003; Hedaya, 2012).

The science of pharmacokinetics basically seeks to answer the following questions:

- ✓ How much of the drug that is given to the patient actually reaches the target organ?
- ✓ Where in the body does the drug go?
- ✓ How long does the drug stay in the body? (Kenakin, 2019).

---

* F.H. Dost (Franz Heinrich Dost) was a prominent German pharmacologist and researcher known for his contributions to pharmacokinetics and clinical pharmacology. He was born on December 6, 1907, and passed away on February 5, 1999. Dost's career focused extensively on the quantitative study of how drugs and other substances interact with the human body. His research primarily centered on pharmacokinetics, which explores the absorption, distribution, metabolism, and excretion of drugs (often abbreviated as ADME).

The scope of PK not only covers studies on healthy subjects but also includes broad research on variations under a variety of physiologic or pathological conditions and the underlying mechanisms, potential Drug-Drug Interactions (DDI), and possible strategies such as dose adjustment to achieve precision medication. Collectively, these aspects of PK allow customization of drug dosage regimens to enhance therapeutic outcomes (Currie, 2018; Li and al., 2019).

The intensity and the duration of the response usually depend on parameters such as:

- The rate and extent of uptake of the drug from its site of administration;
- The rate and extent of distribution of the drug to different tissues, including the site of action;
- The rate of elimination of the drug from the body (Waller and Sampson, 2018).

I.3. Pharmacokinetics processes:

The four stages of Pharmacokinetics are represented by the acronym ADME: Absorption, distribution, metabolite and elimination (excretion), as shown in Figure I.1.

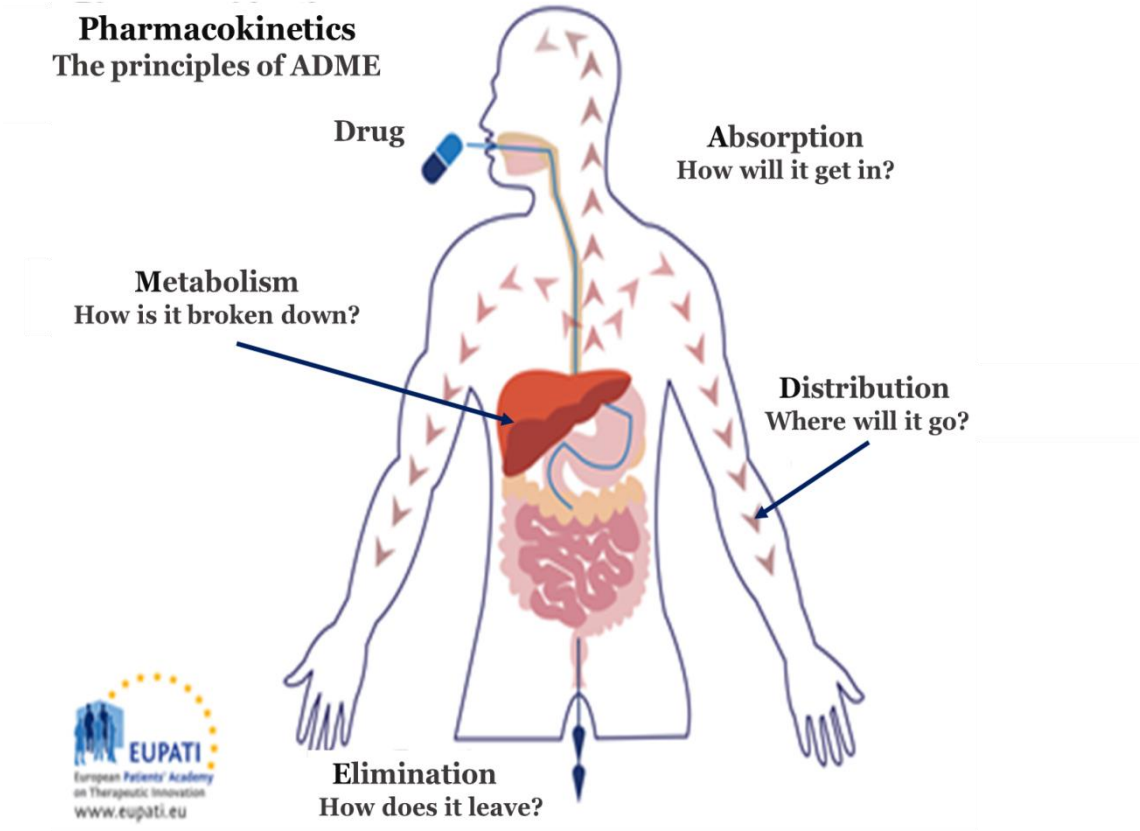


Figure I.1: The four pharmacokinetics processes (EUPATI, 2017).

I.3.1. Absorption:

Absorption is the transfer of a drug from its site of administration to the bloodstream. The rate and extent of absorption depends on the route of administration, the formulation and chemical properties of the drug, and physiologic factors that can impact the site of absorption (Joy Bellis Sakai, 2008).

I.3.1.1. Routes of administration:

Drug's absorption depends on the route of administration, it can be administrated though various routes (Figure I.2). The choice of route of drug delivery is vastly dependent on drug properties, disease states, site of action and patient compliance (Gupta and al., 2019).

I.3.1.1.1. Oral administration (PO):

On account of its easy application, being painless, low expense, wide drug assimilation/distribution and high patient compliance, the oral route is the most preferred route for systemic and local drug delivery. However, the efficiency of many oral medicines is still limited due to various physiological barriers, resulting in low permeability, and drug degradation (Fox and al., 2015).

Different pH environments and digestive enzymes were regarded as the main biochemical barriers for oral drug delivery systems. The pH varies distinctly in different parts of the GIT (Gastro Intestinal Tract), it rises gradually from the stomach to the colon in the range from 1 to 8 (Barbari and al., 2017; Koziolek and al., 2015). The variation from acidic to alkaline environment affects the drug's activities and bioavailability. pH variation not only affects drug delivery, but also a route for targeting the design of oral drugs. The existence of various enzymes will critically influence the bioavailability of drugs in the GIT, especially for protein drugs (Lou and al., 2023).

It can be mentioned that The GIT exerts a low permeability to the bloodstream and extraneous substances, which restricts the bioavailability and absorption of drugs. Since drugs are also susceptible to colonic enzymes and generate biotransformation, the "pro-drug" approach is often used for the colon-specific drug delivery (Amidon and al., 2015).

I.3.1.1.2. Intravenous administration (IV):

Intravenous (IV) is a method of administering concentrated medications (diluted or undiluted) directly into the vein using a syringe through a needless port on an existing IV line or a saline lock. The direct IV route usually administers a small volume of fluid/medicine (max 20 ml) that is pushed manually into the patient. Medications given by IV are usually administered to treat emergent concerns. Medications administered by

direct IV route are given very slowly over AT LEAST 1 minute. This route of administration also eliminates the process of drug absorption and breakdown by directly depositing it into the blood. This results in the immediate elevation of serum levels and high concentration in vital organs, such as the heart, brain, and kidneys. Both therapeutic and adverse effects can occur quickly with direct intravenous administration (Amidon and al., 2015).

I.3.1.1.3. Transdermal administration (TD):

The most common routes of drug delivery are the oral and parenteral routes with the majority of small molecule drugs conventionally delivered orally (Anselmo and Mitragotri, 2014; Han and Das, 2015). However, most therapeutic peptides or proteins are not delivered by the oral route, due to rapid degradation in the stomach and size-limited transport across the epithelium (Schoellhammer and al., 2014). And because of the invasive nature of injections eliciting pain and lower acceptance/compliance by patients, in addition to the requirement for administration by a trained administrator (Kermode, 2004; Schoellhammer and al., 2014). Those routes can be replaced by advanced drug delivery methodologies such as Transdermal Drug Delivery (TDD) (McCrudden and al., 2013).

TDDS has significantly influenced the delivery of various therapeutic agents, especially in pain management, hormonal therapy, and treatment of diseases of the cardiovascular and central nervous systems (Leppert and al., 2018; Roohnikan and al., 2019). TDDS does not involve passage through the gastrointestinal tract; therefore, there is no loss due to first-pass metabolism, and drugs can be delivered without interference from pH, enzymes, and intestinal bacteria. In addition, TDDS can be used to control drug release according to usage restrictions, thereby contributing to the high persistence of this method (Jeong and al., 2021). Most importantly, because TDDS is a non-invasive administration method and involves minimal pain on the patient, drugs can be safely and conveniently administered to children or the elderly (Akhtar and al., 2020; Ruby and al., 2014).

The administration of the drug by TDD starts by applying a drug formulation onto intact and healthy skin (Han and Das, 2015; Schoellhammer and al., 2014). The drug initially penetrates through the stratum corneum and then passes through the deeper epidermis and dermis without drug accumulation in the dermal layer. When drug reaches the dermal layer, it becomes available for systemic absorption via the dermal microcirculation (Kretsos and Kasting, 2007).

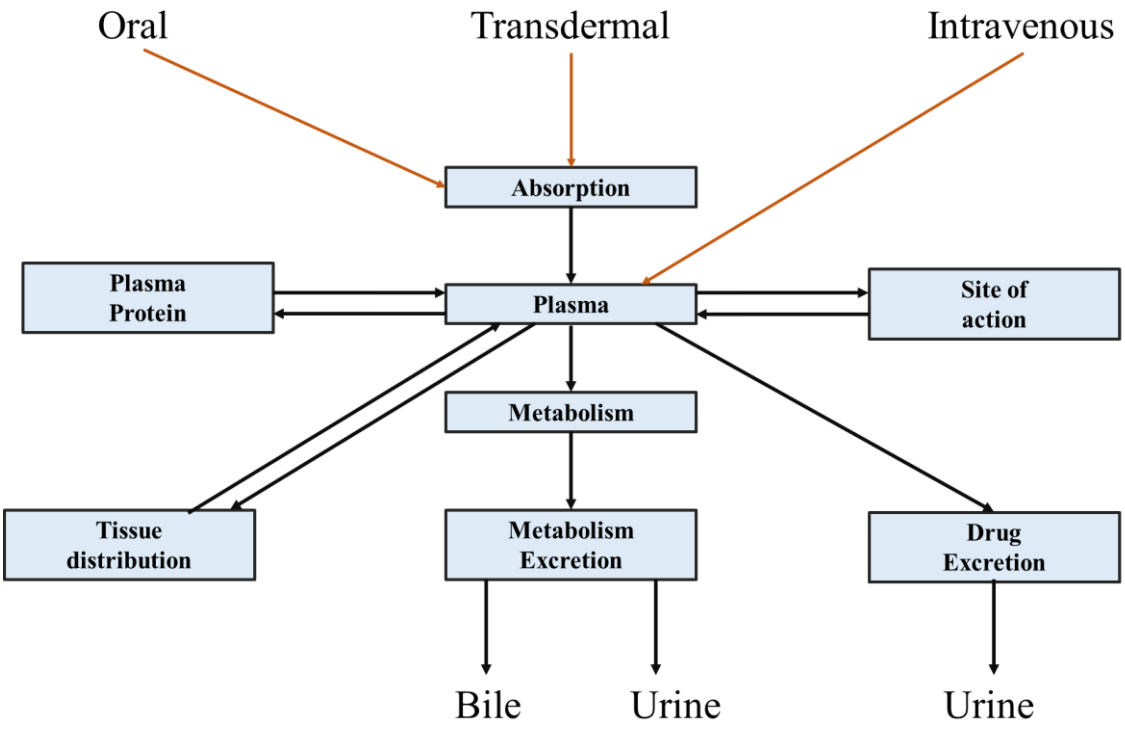


Figure I.2: Simplified pharmacokinetic model of oral, transdermal, intravenous route of administration.

I.3.1.1.4. Intramuscular administration (IM):

Intramuscular administration involves injecting medication directly into a muscle. This method is typically performed by a healthcare professional, although some individuals may self-administer IM injections. The injection site is commonly located in the deltoid muscle of the upper arm. Prior to injection, the skin at the site is cleaned with an alcohol swab. Following injection, the medication is absorbed into the bloodstream from the muscle tissue. The rate of absorption can vary depending on factors such as the formulation of the medication, the size and condition of the muscle, and the injection technique employed. Generally, IM injections result in a more rapid onset of action compared to oral medications but may have a slower onset compared to intravenous (IV) injections (Kretsos and Kasting, 2007).

IM injections are commonly used for a variety of medications. Vaccines, including those for influenza and hepatitis, are frequently administered via IM injection. Antibiotics, such as penicillin and cephalosporins, may be given IM for the treatment of bacterial infections. Additionally, analgesics, such as opioids and nonsteroidal anti-inflammatory drugs (NSAIDs), may be administered IM for pain relief. Hormones, such

as testosterone is also commonly administered IM for hormonal therapy (Amidon and al., 2015).

I.3.1.1.5. Rectal administration (PR):

Involve the insertion of medication into the rectum via the anus. This method can be performed by a healthcare professional or self-administered by the patient. Typically, the medication is in the form of suppositories or rectal solutions. Drugs administered rectally are absorbed through the rectal mucosa into the bloodstream, by passing the liver's first-pass metabolism. The absorption rates may vary depending on factors such as the formulation of the medication, the condition of the rectal mucosa, and the presence of stool in the rectum. Rectal administration provides relatively rapid absorption and onset of action compared to oral medications but may have slower absorption compared to intravenous (IV) administration. This route is commonly used for various medications, including those for gastrointestinal conditions, fever, pain, and constipation. While rectal administration offers many benefits, there are several considerations to keep in mind. Proper technique and hygiene are essential to minimize the risk of contamination and infection during PR administration (Schoellhammer and al., 2014).

I.3.1.1.6. Inhalation Administration (IA):

Inhalation Administration (IA) involves the delivery of medication (drug) directly into the respiratory tract via inhalation. This method allows for the rapid absorption of medication into the bloodstream through the lungs, providing systemic or localized effects. Inhalation Administration (IA) can be achieved through various devices, including Metered-Dose Inhalers (MDIs), Dry Powder Inhalers (DPIs) and nebulizers. The choice of device depends on factors such as the type of medication, the patient's ability to coordinate inhalation, and personal preference.

Medications administered via inhalation are absorbed directly into the bloodstream through the alveoli in the lungs. The large surface area and rich blood supply of the lungs allow for rapid absorption and onset of action. Inhalation administration bypasses the digestive system and liver metabolism, resulting in higher bioavailability and quicker therapeutic effects compared to oral medications. The distribution of inhaled medications throughout the body depends on factors such as the drug's properties, particle size, and pulmonary blood flow. This route is commonly used for the treatment of respiratory conditions such as asthma, Chronic Obstructive Pulmonary Disease (COPD) and respiratory infections. Antibiotics are among the medications commonly delivered via inhalation. They treat respiratory infections.

Inhalation administration offers several advantages over other routes of drug administration. It allows for the targeted delivery of medication to the lungs, resulting in high local concentrations and minimal systemic side effects. Inhalation therapy provides rapid relief of respiratory symptoms, making it particularly effective for acute exacerbations of asthma. It also allows for lower doses of drug to be used compared to oral or parenteral routes, reducing the risk of systemic side effects. While inhalation therapy offers many benefits, there are several considerations to keep in mind. Proper technique and device use are essential to ensure optimal drug delivery and therapeutic effects. Patients should receive thorough education and training on how to use their inhalation devices correctly. Healthcare providers should also assess patient's inhalation technique regularly and provide feedback and support as needed. Patients should be monitored for adverse effects such as bronchospasm, cough, throat irritation, and systemic effects of medication. Device maintenance and cleaning are important to prevent contamination and ensure proper functioning of inhalation devices (Schoellhammer and al., 2014).

I.3.1.1.7. Sublingual administration (SA):

The sublingual route of administration involves placing drug under the tongue, where it is absorbed into the bloodstream through the mucous membranes. This method allows for rapid absorption and bypasses the digestive system, leading to quicker onset of action compared to oral ingestion. Sublingual Administration (SA) is commonly used for medications such as certain types of nitro-glycerine for angina, as well as some drugs for conditions like allergies, pain, and hormone replacement therapy (Gupta and al., 2019).

I.3.1.2. Absorption Mechanism:

I.3.1.2.1. Passive diffusion:

Passive diffusion is the process by which drug molecules diffuse by a concentration gradient from a region of higher concentration to a region of lower concentration. It is the most important mechanism for passage of drugs through membranes. Lipid soluble drugs penetrate lipid membranes. Polar molecules and all ionized compounds partition poorly into lipids, and are not able to pass through membranes, or do so at a much lower rate (Wermuth, 2003). This process doesn't require energy and not specific.

I.3.1.2.2. Facilitated diffusion:

Facilitated diffusion is a form of carrier transport that does not require the expenditure of cellular energy. In these conditions, physiological substrates enter cells down a concentration gradient, but at a faster rate than anticipated from their lipid solubility or molecular size. Facilitated diffusion mediates the absorption of some simple sugars, steroids, amino acids and pyrimidines from the small intestine and their subsequent transfer across cell membranes (Leslie Z. Benet, 2007). If a drug is too **hydrophilic** (easily absorbed by or dissolved in water) it will have trouble crossing the cell's lipid membrane. If a drug is too **lipophilic** (fat-soluble) it will not be soluble enough to cross the water layer surrounding the cell (Elaine N. Marieb and Katja Hoehn, 2016). Since most of the drugs are either weak acids or bases, unionized species represents more lipid-soluble and ready to be absorbed form. Therefore, rate of passive absorption is proportional to fraction of the total drug that exists in unionized solution form (Zhu and al., 2017).

I.3.1.2.3. Active diffusion:

Active transporters in active diffusion typically move molecules against their electrochemical gradients; such process is entropically unfavourable and therefore needs ATP as an energy source. Due to energy consumption, drug molecules can be moved against concentration gradient. The transporters are located in different parts of GIT and help in absorption of both ionized and unionized drug molecules based on their affinity for the substrate (Gupta and al., 2019). Figure I.3 presents the types of diffusion mentioned previously.

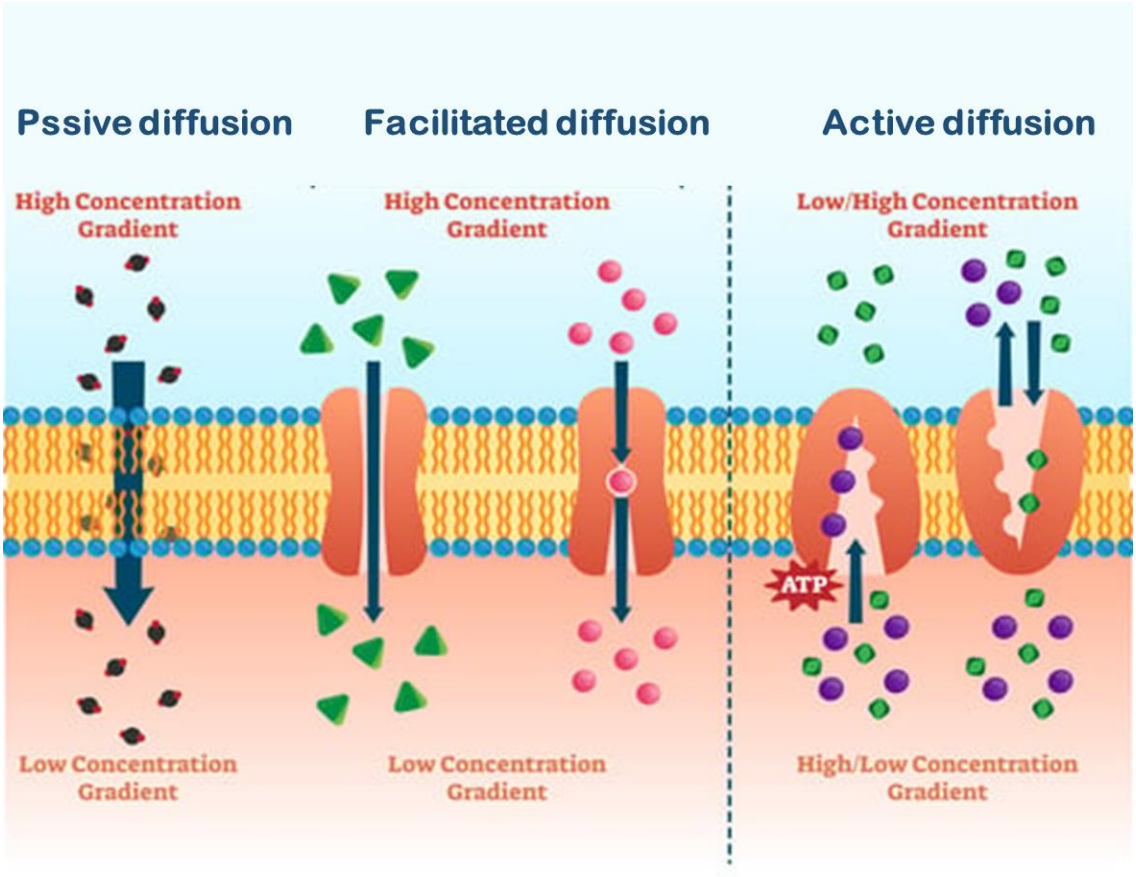


Figure I.3: Types of drugs diffusion (Future learn, 2015).

I.3.2. Distribution:

After the absorption of the drug, this latest will be dissolved in the plasma after it enters the bloodstream. In the plasma the drug will be in two forms: bound and unbound. The pharmacokinetics and pharmacodynamic of the drug will be affected because only the unbound fraction of the drug can distribute to its target site and interact with its target to give a pharmacological effect.

I.3.2.1. Plasma protein binding (PPB):

The human plasma contains many different proteins that are responsible for various functions such as immunity and blood coagulation, etc. However, some of these plasma proteins are capable of binding to drug molecules, and the binding is usually reversible. Only free fraction of the drug which is pharmacologically active and exert the therapeutic effect, whether the bound fraction is pharmacologically inactive, this inactivity is only temporary because the bound and free forms are in reversible equilibrium. As the free form disappears (through elimination), there is a transition from the bound form to the free form (Loichot C, 2005).

Plasma is composed of approximately 90 % water, with the remaining 10 % consisting of ions, proteins, nutrients, waste, and dissolved gases. The three significant fractions of plasma proteins are albumin, globulin (Figure I.4). In the blood, the drug exists in two forms: a free form and a bound form, with the latter primarily binding to these plasma proteins (Mohammad S., 2021).

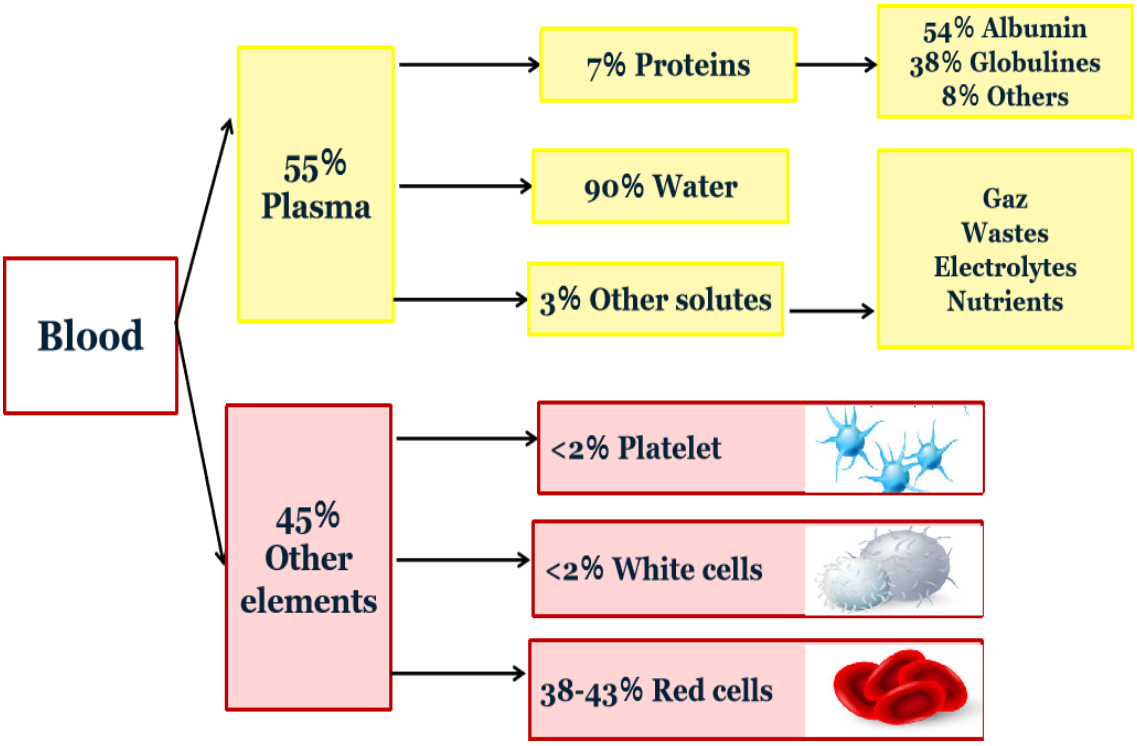


Figure I.4: Compositions of blood in an average person.

Although the plasma is known to have more than 60 different proteins, only a few of them are considered important in the drug binding phenomena (Frank W Putnam, 1975). The main proteins known to bind a high number of drugs in plasma are:

- Human Serum Albumin (HSA).
- Alpha-1-Acid Glycoprotein (AAG) (Schmidt and al., 2010).

Hydrophobic and acidic drugs (e.g., Warfarin, Ketoprofen, Ibuprofen, Diazepam) bind preferably to Human Serum Albumin (HSA), while Alpha-1-Acid Glycoprotein (AAG) connects with the basic ones (e.g., Bupivacaine, Clindamycin) (Ghuman and al., 2005; Otagiri, 2005; Yamasaki and al., 2013). It should be noted that other plasma proteins such as transthyretin and antibodies can also participate in drug binding, but usually, they have affinities for specific drugs only (Chao Han, 2010; Otagiri, 2005).

Notably, it is the unbound fraction that be metabolized and/or excreted. For example, the "fraction bound" of the anticoagulant warfarin is 97 %. This means that of the amount of warfarin in the blood, 97 % is bound to plasma proteins. The remaining

3% (the fraction unbound) is the fraction that is actually active and may be excreted (Leon Shargel, 2005).

In practice, protein binding should only be considered if it is high (> 90 %) and if the drug has a narrow therapeutic margin (toxic concentration close to effective concentration) (Pr. Philippe Lechat, 2006).

I.3.2.2. Tissue binding (TB):

Similar to how drugs bind to plasma proteins, drugs also reversibly bind to proteins and phospholipids in tissues such as liver, kidney.

The tissues to which drugs bind act as a reservoir of these drugs. They keep releasing the drug into the central compartment slowly and increase the elimination half-life of the drug. Some drugs can get accumulated in tissues and remain for a very long time. For example, Tetracyclines accumulate in the bones and thyroids (Marshall and Raveendran, 2019).

The tissue binding of drugs is as important as the plasma proteins binding and has a significant impact on the pharmacokinetics and pharmacodynamics of the drug as it affects both the distribution and elimination processes of the drug in addition to affecting the pharmacological effect of the drug. As it was mentioned before, the greater of the free fraction of the drug, the higher the rate of elimination, while extensive tissue binding will result in a lower elimination rate and a high volume of distribution. Following drug distribution phase, equilibrium will be established between the drug concentration in the plasma and the concentration in tissues. Tissue binding of drugs is different according to the type of tissue because drugs usually have different affinities for different kinds of tissues. And sometimes, binding of a drug in a particular organ is not homogeneous, i.e., the drug concentration in a specific site in an organ is greater than the concentration in another site in the same organ. The liver is the organ with which most of the drugs show highest binding affinities regarding tissues binding, other organs with which drugs are found to bind to a lesser extent are the kidney and the lungs (Deb and al., 2018).

Measurement of tissue binding of drugs is more difficult than the measurement of plasma protein binding, because of the difficulty of measuring drug unbound and bound concentrations inside the tissues (Kotsiou and Tesseromatis, 2011).

I.3.2.3. Fat as a Reservoir:

Many lipid-soluble drugs are stored by physical solution in the body fat. Fat may serve as a reservoir for lipid-soluble drugs. For example, as much as 70 % of the highly lipid-soluble barbiturate thiopental may be present in body fat 3 hours after

administration. Fat is a rather stable reservoir because it has a relatively low blood flow (Laurence L.Brunton, 2011).

I.3.3. Metabolism:

Once a drug has been absorbed and distributed in the body, it will then be broken down by a process known as metabolism (Michael Patrick Adams and Norman Holland, 2020). It involves enzymes that modify the drug in various cells (e.g., for the liver, in hepatocytes) (David E. Golan, 2012). Most drugs are formulated to be lipid-soluble so that they can cross the phospholipid bilayer membranes and be suitable for oral absorption. Metabolism of lipid-soluble drugs to water-soluble structures is needed for effective renal elimination (Humphrey Rang, 2008).

Modifications of drugs and pro-drugs are referred to as metabolism and can be categorized as phase I or phase II (figure I.5) (David E. Golan, 2012; Humphrey Rang, 2008):

- Phase I (reactions of functionalization) involves the creation of a functional group or modification of an existing one by oxidation, reduction or hydrolysis reactions may be referred to as precomputation. Oxidation generally adds a polar group to the chemical structure of a drug by adding an oxygen molecule, reduction tends to add a hydrogen molecule, and hydrolysis adds water (David E. Golan, 2012; Humphrey Rang, 2008). It is a biotransformation (metabolism) of the drugs resulting metabolites that are polar and chemically more reactive than the parent compound (the drug administrated), which can be eliminated directly or continue the metabolization processes of phase II (Laurence L.Brunton, 2011).
- Phase II (reactions of conjugation) the conjugation and hydrolysis reactions generally facilitate attachment of the drug to a polar molecule. These highly polar conjugates generally are inactive and are excreted rapidly in the urine and bile (Laurence L.Brunton, 2011). Most drugs undergo both phase I and phase II metabolism; however, some drugs will undergo only phase I or phase II (Humphrey Rang, 2008).



Figure I.5: Metabolic alteration of drug.

Couple a drug or a metabolite to an endogenous conjugating molecule such as glucuronic acid, sulfuric acid, acetic acid, glutathione, etc. From a pharmacological point of view, it is essential to check the pharmacodynamic consequences of these metabolic reactions. Often but far from always, biotransformation involves inactivation or detoxification. Activation concerns pro-drugs, but also active compounds (drugs) giving rise to active metabolites. The latter may exhibit a PK profile different from that of the parent drug, and/or a qualitatively different activity (D Tremblay, 2007), can be a toxic metabolite as it is shown in Figure I.6.

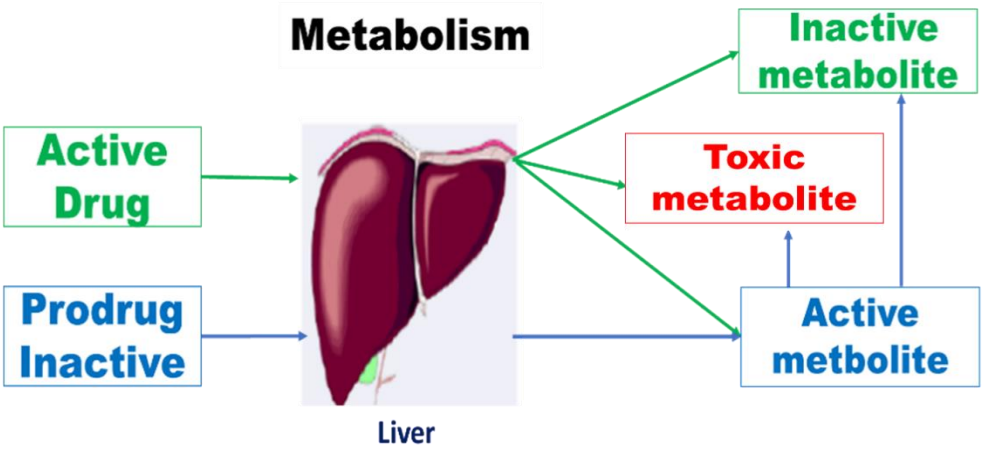


Figure I.6: Metabolism of active drugs and pro-drugs.

I.3.3.1. First-pass effect:

Hepatic first pass effect is a phenomenon concerns drugs given by the oral route especially, which are absorbed from the stomach and the small intestine into the hepatic portal vein. This blood vessel goes directly to the liver. The process of biotransformation begins, and the drug will start to be metabolized in preparation for excretion (elimination) from the body. The drug molecules are now metabolized by the liver enzymes in the normal fashion. This 'first pass effect' reduces the fraction of the dose administered, which then goes on to reach the systemic circulation and become available for therapeutic effect as it's shown in Figure I.7 (Robertson, 2017).

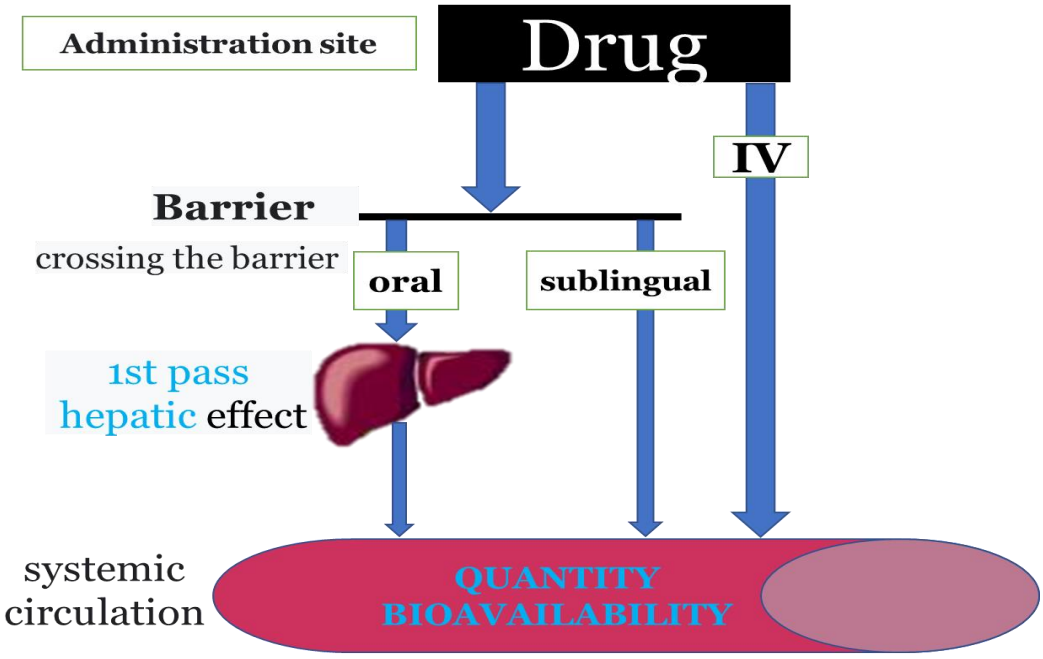


Figure I.7: First-pass effect on IV, oral, sublingual.

I.3.4. Elimination:

Renal elimination is the most common route of drug elimination. However, many drugs are excreted into bile via the liver and some volatile substances (primarily gaseous anesthetics) can be excreted via the lungs. Saliva often contains very small quantities of drug and this may be regarded as a mechanism of excretion although the amount is inconsequential in companion animal species. The gastrointestinal tract can also be involved in elimination of some drugs (Maddison and al., 2008).

Hydrophobic drugs, to be excreted, must undergo metabolic modification making them more polar. Hydrophilic drugs, on the other hand, can undergo excretion directly, without the need for metabolic changes to their molecular structures. Although many sites of metabolism and excretion exist, the principal organ of metabolism is the liver, while the organ primarily tasked with excretion is the kidney shown in Figure I.8. Any significant dysfunction in either organ can result in the accumulation of the drug or its metabolites in toxic concentrations (Aaron Z. Garza; Sharon B. Park; Remek Kocz., 2023). The primary routes of drug excretion are:

I.3.4.1. Renal excretion:

Elimination occurs through the kidneys (urine), it involves three mechanisms: glomerular filtration, tubular secretion, and passive reabsorption.

- Glomerular filtration: This is the process by which blood is filtered in the glomerulus, a network of capillaries in the kidney. Small molecules, including water, electrolytes, and waste products like urea and creatinine, are filtered from

the blood into the renal tubules to form urine. However, large molecules such as proteins and highly protein-bound drugs are not filtered effectively.

- Tubular secretion: In this process, certain substances are actively transported from the blood into the renal tubules, further adding them to the urine. This mechanism allows the elimination of substances that were not effectively filtered in the glomerulus, including drugs and waste products. Both organic anions and cations can be secreted, allowing for the elimination of a wide range of substances.
- Passive reabsorption: While most of the substances filtered by the glomerulus are intended for elimination in the urine, some important substances, such as water and electrolytes, need to be reabsorbed back into the bloodstream to maintain proper balance. Passive reabsorption refers to the movement of substances from the renal tubules back into the blood through diffusion, driven by concentration gradients (Muir, 2015).

I.3.4.2. Biliary excretion:

Bile acid secretion is required for efficient lipid absorption, and bile acids also possess powerful direct and indirect antimicrobial functions in the small intestine (Hofmann, 2007).

A study of nine compounds with differing physicochemical/pharmacokinetic properties suggested that compounds with low molecular weight and ionic character are eliminated renally and not through the bile, whereas drugs with larger molecular weight, more lipophilic, unionized compounds are not eliminated renally, but through the bile. Compounds that are highly protein bound cannot be filtered by the glomerulus in the kidney and are thus excreted in the bile (Leon Shargel Andrew B. C. Yu, 1999).

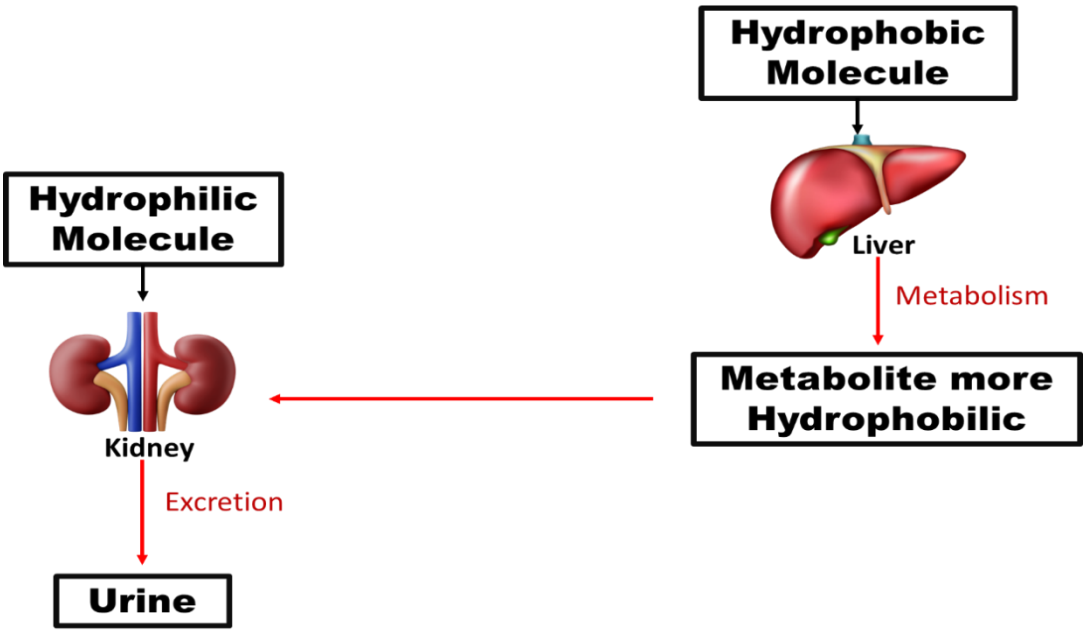


Figure I.8: Elimination of hydrophobic and hydrophilic drug molecule.

I.3.4.3. Pulmonary pathway:

The pulmonary pathway allows for the elimination of volatile compounds such as anesthetics, alcohol, and essences in exhaled air. Additionally, it's important not to overlook other elimination pathways such as saliva, tears, and the skin (Muir, 2015).

I.4. Pharmacokinetics parameters:

Several parameters are crucial in understanding the absorption, distribution, metabolism, and elimination of drugs. Some key pharmacokinetic parameters include:

I.4.1. Bioavailability (BA):

Bioavailability is defined as the fraction of unchanged drug reaching the systemic circulation following administration by any route represented in Table I.1. For an intravenous dose, bioavailability is assumed to be equal to unity. Bioavailability is expressed as the percentage of the total drug dose administered that reaches the circulation (Table I.1). For a drug taken orally, the ‘first-pass effect’ of hepatic metabolism reduces bioavailability. The bioavailability calculations include both free and bound forms of the drugs (Katzung, 2020).

Table I.1: Routes of administration, bioavailability, and general characteristics (Tusom, 2015).

Route	Biovailability (%)	Characteristics
Intravenous (IV)	100	Most rapid onset
Intramuscular (IM)	75 to ≤100	Large volume administration May be painful
Transdermal (TD)	80 to ≤100	Usually very slow absorption Used for lack of first-pass Prolonged duration of action
Rectal (PR)	30 to <100	Less first-pass effect than oral
Inhalation (IA)	5 to <100	Often very rapid onset
Oral (PO)	5 to <100	Most convenient First-pass effect may be important

I.4.1.1. Types of Bioavailability:

I.4.1.1.1. Absolute Bioavailability:

Absolute bioavailability is the dose of extravascular route compared to intravenous route. Therefore, a drug given the by the intravenous route will have an absolute bioavailability of 100 %, whereas drugs given by other routes usually have an absolute bioavailability of less than one (Abo-El-Sooud and al., 2017). When the bioavailability of extravascular route is very close to unity, this indicates that the absorption of the tested drug from this site is nearly completed. If we compare the two different dosage forms having same active ingredients and compare the two drugs bioavailability, it is called relative bioavailability (Abo-EL-Sooud, 2018).

1.4.1.1.2. Relative Bioavailability:

Relative bioavailability measures the bioavailability of a formulation (A) of a certain drug when compared with another formulation (B) of the same drug, usually an established standard or through administration via a different route. When the standard consists of an intravenously administered drug, this is known as absolute bioavailability (Abo-EL-Sooud, 2018; Wesch, 2011).

1.4.1.1.3. Bioequivalence:

Bioequivalence is the property of two dosage forms or active ingredients with similar blood concentration levels that produce the same effect at the site of physiologic activity (Osakwe, 2016). Bioequivalence tests are very important role in the development of pharmaceuticals and match the accepted standard to ensure the therapeutic performance of drugs that have undergone changes in manufacturing

processes, formulation modifications and also for the approval of generic drugs. In the last three decades, bioequivalence tests are so widely used, however, in recent years a new standard, applicable to a number of drugs, was developed based on the classification of drugs according to their biopharmaceutical properties (da Silva Júnior and al., 2017; Takagi and al., 2006).

I.4.2. Volume of Distribution (V_d):

Volume of distribution (V_d) is a theoretic concept that relates the amount of drug in the body (dose) to the concentration (C) of drug that is measured (in blood, plasma, and unbound in tissue water) (Alsanosi and al., 2014). Thus, V_d reflects the extent to which the drug is present in extravascular tissues but not in plasma. Lipid solubility can affect V_d, as highly lipid-soluble drugs have good cell penetration, resulting in high V_d. Plasma protein binding, particularly to albumin, reduces the V_d, while tissue binding increases it (Davis and al., 2011). In addition, V_d is helpful in predicting the initial maximum concentration for an IV route (Toutain and Bousquet-me Lou, 2004).

I.4.3. Clearance (Cl):

The drug present in the plasma can be removed from the body primarily through the kidneys and liver (Toutain and al., 2004). Drug clearance is defined as the volume of plasma cleared of a drug over a specified time period. Clearance by the kidneys is called renal clearance, and that by all other organs is referred to as non renal clearance. The latter most often represents clearance by the liver. Total clearance is the sum of all body clearances (Bhosle and al., 2017).

I.4.4. Half-life time ($T_{1/2}$):

Half-life ($T_{1/2}$) is defined as the amount of time required for the drug concentration measured in plasma (or other biological matrices) to be reduced to exactly half of its starting concentration or amount (50 %), measured by hours or minutes (Schrag and Regal, 2013). This parameter is crucial for determining estimating drug accumulation or clearance and predicting the duration of action or potential side effects.

I.4.5. Protein Binding fraction (FB):

I.4.5.1. Parameters of Drug-Protein Binding:

The binding of drugs to plasma proteins is a dynamic phenomenon that depends on the number of binding sites on the protein, the affinity constant, and the molar concentration of the protein. The modes of binding differ according to the chemical structure and physicochemical characteristics of the drug. There are several types of binding to plasma proteins, which explain the different affinities. It is possible to

classify drug-protein binding into two categories according to their strength, that is, their affinity:

- Weak or labile bonds, such as hydrogen or hydrophobic bonds.
- Strong bonds, such as ionic bonds.

Knowledge of these protein-binding parameters is essential for evaluating the pharmacokinetic implications of drug binding (P. E. Rolan, 1994; Wright et al., 1996).

I.4.5.1.1. Affinity:

The binding of a drug D to a protein P results from a reversible equilibrium between the free form [D] and the bound form [DP], which follows the law of mass action (Eq.I.1):

$$\mathbf{[D] + [P] \leftrightarrow [DP]} \qquad \mathbf{(I.1)}$$

Only the free form [D] is active and can diffuse across membranes. The bound form [DP] can be considered a storage form that does not cross membranes in mole per liters. Affinity defines the strength of the drug's binding to the protein. With Ka being the affinity constant of the drug for the protein expressed in liters per mole (Oravcova and Lindner, 1996), at equilibrium we have (Eq.I.2):

$$\mathbf{K_a} = \frac{[\mathbf{DP}]}{([\boldsymbol{M}] \ *[\mathbf{D}])} \qquad \mathbf{(I.2)}$$

K_a: Affinity constant in (L/mol);

DP: Concentration of drug-protein complex in (mol/L);

D: Concentration of drug in (mol/L);

P: Concentration of protein in (mol/L).

I.4.5.1.2. Binding Sites:

The binding of drugs to proteins occurs at sites located on the surface of the proteins. The number of binding sites is generally denoted as **(n)**. There may be several families of sites on a protein, and for a given drug, a protein offers one or more binding sites belonging to one or more classes, each characterized by different affinities. Competitive binding phenomena can develop between multiple drugs when a protein has a limited number of common sites for the substances present. If two drugs of the same type, binding to the same sites, are administered simultaneously, the one with the higher affinity will preferentially bind, displacing or preventing the binding of the other. This competition phenomenon can also occur with endogenous substances, such as bilirubin, which also binds to albumin and can compete with acidic drugs, leading to protein unbinding of these drugs (Oravcov and Lindner, 1996).

To identify the binding sites of drugs to plasma proteins, various experimental and computational methods can be employed such as Nuclear Magnetic Resonance (NMR) Spectroscopy.

I.4.5.1.3. Fraction of Binding (FB):

Two common ways are used to express the plasma protein binding of drugs. The first is to describe it as the Unbound Fraction in plasma **(FU)**, which is the ratio of the unbound (free) drug concentration. The second describes the Bound Fraction of drug in the plasma **(FB)** (Deb and al., 2018b):

The most commonly used binding parameter is the percentage of drug binding to plasma proteins. The bound fraction **(FB)** and the unbound or free fraction **(FU)** will have values ranging from 0 to 100 %. If the value of FB is greater than 90 %, protein binding is considered significant, whereas FB values below 20 % indicate that binding is weak or negligible. The values of FB and FU are determined by the following relationships (Eq.I.5; I.6) (Oravcov and Lindner, 1996):

$$\mathbf{FB} = \frac{[\boldsymbol{BD}]}{[\mathbf{FD}]+[\mathbf{BD}]} * \mathbf{100} \qquad \textbf{(I.3)}$$

FB: Fraction of binding in %;

BD: Concentration of bound drug in mol/L;

FD: Concentration of free drug in mol/L.

$$\mathbf{FU = 100 - FB} \qquad \textbf{(I.4)}$$

This percentage of binding is directly related to the product n*[P]*K_a. With n is the number of binding sites; K_a represents the affinity constant of the drug for the protein, and [P] is the molar concentration of the protein.

I.4.5.2. Measuring the drug bound fraction by equilibrium dialysis Method:

To measure the drug bound fractions, numerous separation and study techniques are used, that are continuously developed, and well described. Equilibrium dialysis is the most frequently used methodology for determination of plasma protein binding in basic research (Gleeson, 2007). This method employs a dialysis cell which contains two reservoirs separated by a semipermeable dialysis membrane (usually cellulose), which is available in various molecular weight cut-offs and should be chosen based on the molecular weights of the compounds and proteins of interest. The system should be at physiological temperature (37 °C) and allowed to reach equilibrium (J. E. D. Wright and al., 1996). One reservoir contains a solution of proteins found in the blood (usually

serum) and the drug. In reservoir, a buffer solution is added, which helps keep the pH stable. Over time, the drug molecules in the solution will interact with the proteins, some will bind to them, and others will remain unbound or "free". The membrane allows only the free drug molecules to move between the chambers. Once equilibrium is reached (the concentration of free drug is the same in both reservoirs), after a suitable incubation period, typically several hours or overnight. Samples from the buffer solution and free drug reservoir are collected and analysed for drug concentration by using High-Performance Liquid Chromatography HPLC (Deb and al., 2018b). Noting that preliminary experiments should be conducted to determine the time required to reach equilibrium. When no change is detected in protein and buffer concentrations between multiple time points, equilibrium is said to be reached. By comparing the total drug concentration with the free drug concentration, the fraction of bounded drug to protein can be calculated easily.

According to Fick's law, the rate of drug diffusion depends on the membrane's surface area and thickness, the concentration gradient, the compartment volume, and the diffusion coefficient. This latter quantity, which characterizes a particular drug molecule, depends on the drug's molecular weight and temperature (the dialysis rate increases with temperature) (Gleeson, 2007). The membrane must be chosen to achieve the highest ligand diffusion rate with the highest retention capacity for all proteins. Although there is no "standard method" for measuring binding, Equilibrium Dialysis (ED) is often considered the "gold standard" for determining the drug-protein binding profile. Davis and Klotz were the first to use equilibrium dialysis (Davis', 1943; Klotz, 1946).

I.4.5.3. Factors Influencing Protein Binding:

Factors that can cause alterations in protein binding of drugs can be divided into three classes: physiologic conditions; pathological conditions; and drug-drug interaction. The most common ones are changes in the binding protein concentration, or a change in the structure of the binding protein that would affect the binding affinity of the protein (John Wiley and Sons, 2011; Joseph and Hage, 2010).

I.4.5.3.1. Physiologic Factors:

Physiologic factors that can influence drug protein binding include:

- Age: Concentrations of both albumin and alpha-1-acid glycoprotein in neonates are lower than adults. Albumin concentration in the elderly is also lower compared to normal adults (Fanali and al., 2012a).

- Gender: the concentration of both albumin and alpha-1-acid glycoprotein are found to be lower in adult females (nonpregnant) as compared to adult males (Hanley and al., 2010).

1.4.5.3.2. Pathologic Factors:

Many pathological conditions can influence drug protein binding. Several of this cause a change in the concentration of a binding protein(s), e.g., conditions characterized by a lowered plasma albumin concentration, such as hypoalbuminemia or analbuminemia, cause an increase in the unbound fraction of albumin binding drugs. On the other hand, pathological conditions can also cause an increase in an endogenous substance concentration that binds the binding protein, such as the increased bilirubin concentration in hyperbilirubinemia, where bilirubin binds to albumin and causes an increase in the unbound fraction of albumin bound drugs (Gatta and al., 2012). In a similar fashion, unbound fractions of alpha-1-acid glycoprotein binding drugs will be affected by pathological conditions that alter the concentration of alpha-1-acid glycoproteins, e.g., hepatitis, hepatic cirrhosis, pancreatic cancer, and nephrotic syndrome are diseases associated with a decrease in alpha-1-acid glycoprotein concentration and thus, an increase in the unbound fraction of alpha-1-acid glycoprotein binding drugs (Stangier and al., 2010; Vivekanandan-Giri and al., 2011).

1.4.5.3.3. Drug-Drug Interaction:

The coadministration of one drug can affect the protein binding of another administrated drug and thus change its unbound fraction. The mechanism by which one drug can alter the binding of another drug to a plasma protein is through direct displacement of the drug from the binding site. If the competing drug has a higher affinity for the protein, then it will displace the other one from the binding site and thus increase its unbound fraction which may lead to toxicity. For example, in patients taking the anticoagulant drug warfarin, the coadministration of phenylbutazone that binds to the same binding site of warfarin in albumin will cause displacement of warfarin from the binding site, which in turn may result in an increased free fraction of warfarin and an increase in prothrombin time with increased risk of bleeding (Ansari, 2010; Fanali and al., 2012b; Hines and Murphy, 2011).

I.5. Factors influencing Pharmacokinetics:

Pharmacokinetics can be influenced by various factors including:

I.5.1. Age:

One of the effects of the ageing process is the physiological and biochemical changes that occur within the body with a general decline in the functional efficiency of the body organs, such alterations might be expected to affect drug responsiveness by changing the process of drug disposition. Age related changes can impede absorption due to decreased blood flow to the tissues and the GI tract and changes in gastric pH. Age related variations in plasma proteins may also cause significant problems with medications that are highly protein bound, such as Phenytoin and Levodopa/Carbidopa. Blood levels can vary, especially if food intake and dosing are not consistent (Elaine N. Marieb and Katja Hoehn, 2016). Decreased cardiac output in older adults and those with chronic conditions may reduce subcutaneous and intramuscular drug absorption, thus affecting the pharmacokinetics of injectable medications. Transdermal medications are absorbed through subcutaneous fat, which is reduced with aging (Anita Lee Wynne, 2002).

I.5.2. Gender:

Females generally have a lower body weight and thus lower organ size and blood flow, as well as a higher percentage of body fat (Fletcher et al., 1994; Miriam del Carmen Carrasco-Portugal, 2011). On the contrary, several fluoroquinolones, including Ofloxacin, Fleroxacin and Grepafloxacin, have smaller volumes of distribution and, in part, smaller oral clearances in women compared with men (Ejthymiopoulos and al., 1997; Koch-Weser, 1982). Also, gastric acid secretion has been reported to be lower in women than men, and is accompanied by an additional reduction in gastric emptying (Kanda and al., 1995; Mesnil et al., 1998). Lastly Women have an approximately 10 % lower glomerular filtration rate than men (Barry M. Brenner, 2005; C. E. Wright and al., 1997).

I.5.3. Chronical diseases:

I.5.3.1. Diabetes:

The majority of diabetic patients are between 45 and 65 years of age in developing countries, while they are 65 years of age in developed countries. Diabetes may influence the pharmacokinetics of numerous drugs by affecting their ADME system micro and macrovascular changes are probably the main long-term complication of diabetes. Gastric mucosal blood flow has been reported to be significantly reduced in diabetic patients compared with non-diabetic subjects (Dostalek and al., 2012). Changed

mucosal blood flow might explain gastric pH changes in diabetic patients (Hasler and al., 2008). Should be mentioned also that the volume of distribution of a drug correlates with the degree of obesity, and because obesity is one of the most important factors in the development of insulin resistance and diabetes, the volume of distribution of lipophilic drugs is affected by the disease (Kashyap & De Fronzo, 2007; Svacina, 2020). On the other hand, obesity does not affect serum albumin and drug binding to albumin. Diabetic patients have higher level of circulating glucose in the blood, leads to changes in the structure of albumin, which can increase the free fraction of acidic drugs in patients with diabetes (Keller et al., 1984). It also leads to reduction in protein binding capacity (Pascale and al., 1990). Lastly nephropathy develops in approximately 40% of diabetic patients (Sasso and al., 2006). Diabetes causes micro and macrovascular changes and leads to hyperfiltration and, hence, an increased glomerular filtration rate (Gill and al., 1994).

1.5.3.2. Heart failure:

The pathophysiological changes that occur in congestive heart failure may lead to alterations in drug pharmacokinetics. The volume of distribution of drugs is reduced in heart failure. Loading doses should therefore be smaller for drugs such as Lignocaine, Quinidine, Disopyramide and Tocainide. Clearance of drugs may be impaired due to hypoperfusion to the liver and kidneys, and also due to liver congestion. Administration by the intravenous route is recommended in severe congestive heart failure, both to achieve a prompt response and to titrate therapy effectively (Mangoni and Jarmuzewska, 2019).

I.6. Conclusion:

In conclusion, this part has provided a comprehensive overview of pharmacokinetics, shedding light on how drugs are absorbed, distributed, metabolized, and eliminated within the body. Possessing an understanding of these processes is crucial for optimizing drug therapy, it also allows practitioners the flexibility to prescribe and administer medications that will provide the greatest benefit at the lowest risk and allow them to make adjustments as necessary, given the varied physiology and lifestyles of patients.

Bibliographic references:

Aaron z. Garza, Sharon B. Park, Remek Kocz. (2023). Drug elimination. StatPearls (vol.13, p.57).

Abo-el-sooud, K. (2018). Absolute and relative bioavailability. In drug discovery and evaluation: methods in clinical pharmacology (2nd ed.). Springer international publishing. Berlin, Heidelberg (pp.1–7).

Abo-el-sooud, K. Samar, Mouneir, M., & Fahmy. (2017). Curcumin ameliorates the absolute and relative bioavailabilities of marbofloxacin after oral administrations in broiler chickens. Wulfenia journal (vol. 24, issue 3, p.285).

Akhtar, N. Singh, V. Yusuf, M., & Khan, A. (2020). Non-invasive drug delivery technology: development and current status of transdermal drug delivery devices, techniques and biomedical applications. Biomedizinische Technik (vol. 65, issue 3, pp.243–272).

Alsanosi, M. M., Skiffington, C., & Padmanabhan, S. (2014). Pharmacokinetic pharmacogenomics. In handbook of pharmacogenomics and stratified medicine (1st ed.). Elsevier's Academic Press (pp.341–364).

Amidon, S., Brown, J. E., & Dave, V. S. (2015). Colon-targeted oral drug delivery systems: design trends and approaches. Aaps pharmscitech (vol.16, Issue 4, pp.731–741).

Anita lee Wynne, T. M. (2002). Pharmacotherapeutics for nurse practitioner prescribers (1st ed.). F.a. davis company. United States (p.1241).

Ansari, J. (2010). Drug interaction and pharmacist. Journal of young pharmacists (vol. 2, issue 3, pp.326–331).

Anselmo, C., & Mitragotri, S. (2014). An overview of clinical and commercial impact of drug delivery systems. Journal of controlled release (vol. 190, pp.15–28).

Barbari, R., Dorkoosh, A., Amini, M., S Harifzadeh, M., Atyabi, F., Balalaie, S., tehrani, R., & Tehrani, R. (2017). A novel nanoemulsion-based method to produce ultrasmall, water-dispersible nanoparticles from chitosan. International journal of nanomedicine (vol. 12, pp.347–348).

Barry, M., Brenner, Jr., & M. Brenner. (2005). The kidney (4th ed.). Lippincott williams & wilkins. United States (pp.515–516).

Bhosle, K., Altit, G., Autmizguine, J., & Chemtob, S. (2017). Drugs that cause significant displacement of bilirubin from albumin in vitro. Fetal and neonatal physiology (5th ed.). McGraw Hill. United States (p.314).

Chao Han, W. (2010). Evaluation of drug candidates for preclinical development: pharmacokinetics, metabolism, pharmaceutics, and toxicology (1st ed.). John Wiley & Sons. Canada (p.27).

Currie, G. M. (2018). Pharmacology, part 2: introduction to pharmacokinetics. Journal of nuclear medicine technology (vol.46, p.222).

Dtremblay, J.P. T. (2007). Clinical pharmacokinetic criteria for drug research. Comprehensive Medicinal Chemistry (2nd ed). Elsevier Ltd. France (pp.16–22)

Da silva junior, w. F., de oliveira pinheiro, J. G., moreira, A., De Souza, J., & De lima, A. A. N. (2017). Alternative technologies to improve solubility and stability of poorly water-soluble drugs. Multifunctional systems for combined delivery, biosensing and diagnostics (1st ed). Elsevier. Amsterdam, Netherlands (pp.300).

David E. Golan, A. A. (2012). Principles of pharmacology: the pathophysiologic basis of drug therapy (3rd ed.). Williams & wilkins. U.S.A (p.321).

Davis, B. D. (1943). The binding of sulfonamide drugs by plasma proteins. A factor in determining the distribution of drugs in the body. The Chemical Division (vol.22, issue 5, p.754).

Davis, P. J., Bosenberg, A., Davidson, A., Jimenez, N., Kharasch, E., Lynn, M., Tofovic, P., & Woelfel. (2011). Pharmacology of pediatric anesthesia. Smith's anesthesia for infants and children: expert consult (8th ed). Mosby. Amsterdam, Netherlands (pp.1301–1307).

De Aaldo Rescigno. (2003). Foundations of pharmacokinetics (3rd ed.). Kluwer Academic/Plenum. New York. (p.150).

Deb, P. K., Al-attraqchi, O., Prasad, R., & Tekade, K. (2018). Protein and tissue binding (1st ed.). Dosage form design considerations. Academic Press. United States (p.54).

Dostalek, M., Akhlaghi, F., & Puzanovova, M. (2012). Effect of diabetes mellitus on pharmacokinetic and pharmacodynamic properties of drugs (vol 51, issue 8, pp.483–499).

Ejthymiopoulos, C., Bramer, L., & Maroli, A. (1997). Effect of age and gender on the pharmacokinetics of grepafloxacin. Clinical pharmacokinetics. (vol.33, issue 1, pp.2–9).

Elaine N. Marieb and Katja Hoehn. (2016). Human anatomy and physiology (11th ed.). Pearson education. London, England. (pp.611–623).

Eupati. (2017). Pharmacokinetics [pk]. Https://toolbox.eupati.eu/glossary/pharmacokinetics/. (Accessed 13/01/2023).

F. H. Dost. (1953). Der blutspiegel, kinetik der konzentrationsabläufe in der kreislaufflüssigkeit. Springer (vol.11, issue 3, pp.374–401).

Fanali, G., Di Masi, A., Trezza, V., Marino, M., Fasano, M., & Ascenzi, P. (2012). Human serum albumin: from bench to bedside. Molecular aspects of medicine (vol. 33, issue 3, pp.209–290).

Fletcher, V., Acosta, E. I., & Strykowski, J. M. (1994). Gender differences in human pharmacokinetics and pharmacodynamics. Clinical pharmacokinetics (vol.48, issue 3, pp.143–157).

Fox, B., Kim, J., le, V., Nemeth, L., Chirra, D., & Desai, T. A. (2015). Micro-nanofabricated platforms for oral drug delivery. Journal of controlled release (vol. 21. pp.431–444).

Frank w Putnam. (1975). The plasma proteins (2nd ed.). Academic press. Amsterdam, Netherlands. (p.312).

Future learn. (2015). Cell survival, moving materials around. Cell survival – moving materials around (futurelearn.com). (Accessed 18/04/2023).

Gatta, A., Verardo, A., & Bolognesi, M. (2012). Hypoalbuminemia. Internal and emergency medicine (vol. 7, issue 3, pp.193–199).

Ghuman, J., Zunszain, A., Petitpas, I., Bhattacharya, A., Otagiri, M., &Curry, S. (2005). Structural basis of the drug-binding specificity of human serum albumin. Journal of molecular biology (vol. 353, issue 1, pp.38–52).

Gill,V., Hardy, J., Patrick, W., & Masterson, A. (1994). Random blood glucose estimation in type 2 diabetes: does it reflect overall glycaemic control. Diabetic medicine (vol.11, issue 7, pp.705–708).

Gleeson, M. P. (2007). Plasma protein binding affinity and its relationship to molecular structure: An in-silico analysis. Journal of Medicinal Chemistry (vol.50, issue 1, pp.101–112).

Gupta, D., Gupta, V., &Yang, N. (2019). Understanding the routes of administration. In handbook of space pharmaceuticals. Springer international publishing (vol. 12, pp.1–25).

Han, T., & Das, B. (2015). Potential of combined ultrasound and microneedles for enhanced transdermal drug permeation. European journal of pharmaceutics and biopharmaceutics. Elsevier (vol. 89, pp.312–320).

Hanley, J., Abernethy, R., & Greenblatt, J. (2010). Effect of obesity on the pharmacokinetics of drugs in humans. Journal of clinical pharmacokinetics (vol. 49, issue 2, pp.71–80).

Hasler, L., Coleski, R., Chey, D., Koch, L., Mccallum, W., M., Kuo, B., Sitrin,D., Katz, A., Hwang, J., Semler, R., & Parkman, P. (2008). Differences in intragastric pH in diabetic vs idiopathic gastroparesis: relation to degree of gastric retention. American journal of gastrointestinal and liver physiology (vol.249, issue 6, p.1389).

Hedaya, M. A. (2012). Basic pharmacokinetics (2nd ed.). CRC Press. United States (p. 162).

Hines, l. E., & Murphy, J. E. (2011). Potentially harmful drug-drug interactions in the elderly. American journal geriatric pharmacotherapy (vol. 9, issue 6, pp.367–371).

Hofmann, F. (2007). Biliary secretion and excretion in health and disease: current concepts. Annals of hepatology (vol. 6, issue 1, p.20).

Humphrey Rang, R., & F. (2008). Rang and dale's pharmacology (6th ed.). Churchill Livingstone. United Kingdom (p.509).

Jeong, Y., Kwon, M., Choi, E., & Kim, S. (2021). Recent advances in transdermal drug delivery system. Biomaterials research. Biomed central (vol.25, pp.2–9).

John Wiley & Sons, J. (2011). Oral bioavailability: basic principles, advanced concepts, and applications (1st ed.). Wiley. United Kingdom (p.734).

Jhoseph, K. S., & Hage, S. (2010). The effects of glycation on the binding of human serum albumin to warfarin and l-tryptophan. Journal of pharmaceutical and biomedical analysis (vol.53, issue 3, pp.813–815).

Joy Bellis Sakai. (2008). Practical pharmacology for the pharmacy technician (1st ed.). jones & bartlett learning. United States (p.452).

Kanda, C., Yonkers, A., & Cole, O. (1995). Gender as a risk factor for adverse events to medications. Drugs (vol. 50, issue 1, pp.2–5).

Kashyap, S. R., & Defronzo, A. (2007). The insulin resistance syndrome: physiological considerations. Diabetes and vascular disease research (vol. 4, issue 1, pp.13–19).

Katzung, G. (2020). Chapter 1: introduction: the nature of drugs & drug development & regulation. Basic & clinical pharmacology. (15th ed.). McGraw-Hill. United State (pp.25–30).

Keller, F., Maiga, M., Neumayer, H., Lode, H., & Distler, A. (1984). Pharmacokinetic effects of altered plasma protein binding of drugs in renal disease. European journal of drug metabolism and pharmacokinetics (vol. 9, issue 3, p. 277).

Kenakin, P. (2019). Pharmacokinetics. A pharmacology primer (5th ed.). Elsevier. United states (pp.246–290).

Kermode, M. (2004). Unsafe injections in low-income country health settings: need for injection safety promotion to prevent the spread of blood-borne viruses. Health promotion international (vol. 19, issue 1, p.95).

Koch-weser, J., M., & I. (1982). Gender differences in human pharmacokinetics and pharmacodynamics. Clin Pharmacokinet (vol. 48, issue 3, pp.144–149).

Kotsiou, A., & Tesseromatis, C. (2011). Protein binding of drugs. Oral bioavailability (1st ed.). Wiley. United State pp.147–159).

Koziolek, M., Grimm, M., Becker, D., Iordanov, V., Zou, H., Shimizu, J., Wanke, C., Garbacz, G., & Weitschies, W. (2015). Investigation of pH and temperature profiles in the GI tract of fasted human subjects using the intellicap system. Journal of pharmaceutical sciences (vol. 104, pp.2855–2863).

Kretsos, K., & Kasting, B. (2007). A geometrical model of dermal capillary clearance. Mathematical biosciences (vol. 208, issue 2, pp.432–450).

Laurence Brunton. (2011). Goodman and gilman's the pharmacological basis of therapeutics (12th ed.). Mcgraw hill. United States (pp.319–498).

Leon Shargel. (2005). Applied biopharmaceutics & pharmacokinetics (5th ed). Mcgraw-hill. United States (pp.209–536).

Leon Shargel Sndrew C. (1999). Applied biopharmaceutics and pharmacokinetics (4th ed.). Appleton & lange. United States (pp.498–594).

Leppert, W., Malec-milewska, M., Zajaczkowska, R., & Wordliczek, J. (2018). Transdermal and topical drug administration in the treatment of pain. Molecules, MDPI (vol. 23, issue 3, p.681).

Leslie Z. Benet, H. (1995). Drug absorption, distribution, and elimination. Toxicology and pathology (vol. 23, issue 2, pp.115–123).

Li, Y., Meng, Q., Yang, M., Liu, D., Hou, X., Tang, L., Wang, X., Lyu, Y., Chen, X., Liu, K., Yu, M., Zuo, Z., & Bi, H. (2019). Current trends in drug metabolism and pharmacokinetics. Acta pharmaceutica sinica B (vol. 23, issue 6, p.1114).

Loichot C, G. M. (2005). Module of general Pharmacology. Medicine Faculty of Strasbourg (p. 25).

Lou, J., Buan, H., Qin, Q., Teng, Z., Gan, F., Zhou, X., &Zhou, X. (2023). Advances in oral drug delivery systems: challenges and opportunities. Pharmaceutics (vol. 15, issue 2, pp.4–20).

Maddison, E., Page, W., & Dyke, M. (2008). Clinical pharmacokinetics. Small animal clinical pharmacology (2th ed.). Saunders Ltd. United Kingdom (pp.345–478).

Mangoni, A., & Jarmuzewska, A. (2019). The influence of heart failure on the pharmacokinetics of cardiovascular and non-cardiovascular drugs: a critical appraisal of the evidence. British journal of clinical pharmacology (vol. 85, pp.20–32).

Marshall, G., & Raveendran, R. (2019). Introduction to basics of pharmacology and toxicology (1st ed.). Springer. Amsterdam, Netherlands (pp.356–393).

Mccrudden, C., Singh, R., Migalska, K., & Donnelly, F. (2013). Strategies for enhanced peptide and protein delivery. Therapeutic delivery. Newlands press (vol. 4, issue 5, pp.593–610).

Mesnil, F., Mentre, F., Dubruc, C., Thenot, P., & Mallet, A. (1998). Population pharmacokinetic analysis of mizolastine and validation from sparse data on patients using the nonparametric maximum likelihood method. Journal of pharmacokinetics and biopharmaceutics (vol. 26, issue 2, pp.135–150).

Michael Patrick Adams & Norman Holland. (2020). Nursing pharmacology (1st ed.). Pearson. United Kingdom (pp.203–498).

Miriam Del Carmen Carrasco-portugal, M. (2011). Gender differences in the pharmacokinetics of oral drugs. Pharmacology & pharmacy (vol. 2, pp.32–40).

Mohammad S. (2021). Physiology Of blood. Phyiology2021 (vol. 10, p.64).

Muir, W. (2015). Factors influencing analgesic drug selection, dose, and routes of drug administration. Handbook of veterinary pain management (3rd ed.). Mosby. United States (p.650).

Osakwe, O. (2016). Clinical development: ethics and realities. Social aspects of drug discovery, development and commercialization (1st ed.). Elsevier. United States (p.467).

Otagiri, M. (2005). A molecular functional study on the interactions of drugs with plasma proteins. In drug metabolism and pharmacokinetics (vol. 20, issue 5, pp.310–312).

OravcovS, J., & Lindner, W. (1996). Drug-protein binding studies trends in analytical and experimental methodology. Journal of Chromatography (Vol. 677, issue 1, p.4).

Pascale, Z., Barre, J., & Tillemellf, P. (1990). Disease-induced variations in plasma protein levels implications for drug dosage regimens (vol. 19, issue 2, pp.152–157).

Pr. Philippe Lechat. (2006). Module of Pharmacology. University Pierre and Marie Curie. Paris, France (pp.63–75).

P. E. Rolan. (1994). Plasma protein binding displacement interactions, why are they still regarded as clinically important. British Journal of Clinical Pharmacology (vol. 37, pp.125–128).

Robertson, D. (2017). Insulin's Discovery: New Insights on Its Hundredth Birthday: From Insulin Action and Clearance to Sweet Networks. Nurse prescribing (vol. 22, issue 3, pp.2–10).

Roohnikan, M., Laszlo, E., Babity, S., & Brambilla, D. (2019). A snapshot of transdermal and topical drug delivery research in Canada. Pharmaceutics (vol. 11, issue 256, pp.4–10).

Ruby, K., Pathak, M., & Aggarwal, D. (2014). Critical attributes of transdermal drug delivery system (TDDS) a generic product development review. Drug development and industrial pharmacy (vol. 40, issue 11, pp.1421–1428).

Sasso, C., De Nicola, L., Carbonara, O., Nasti, R., Minutolo, R., Salvatore, T., Conte, G., & Torella, R. (2006). Cardiovascular risk factors and disease management in type 2 diabetic patients with diabetic nephropathy for the nid-2 (nephropathy in diabetes-type 2) study group. Clinical care (vol.29, issue 3, p.498).

Schmidt, S., Gonzalez, D., & Derendorf, H. (2010). Significance of protein binding in pharmacokinetics and pharmacodynamics. Journal of pharmaceutical sciences (vol. 99, issue 3, p.1108).

Schoellhammer, M., Blankschtein, D., & langer, R. (2014). Skin permeabilization for transdermal drug delivery: recent advances and future prospects. Expert opinion on drug delivery (vol. 11, issue 3, pp. 395–403).

Schrag, M., & Regal, K. (2013). Pharmacokinetics and toxicokinetics. A comprehensive guide to toxicology in preclinical drug development (1st ed.). Academic press. United states (pp. 71–92).

Stangier, J., Rathgen, K., Stahle, H., & Mazur, D. (2010). Influence of renal impairment on the pharmacokinetics and pharmacodynamics of oral dabigatran etexilate an open-label, parallel-group. Clinical pharmacokinetics (vol.49, issue 4, p.661).

Svacina, S. (2020). Obesity and cardiovascular disease. Vnitrni lekarstvi (vol. 66, issue 2, p.89).

Takagi, T., Ramachandran, C., Bermejo, M., Yamashita, S., Yu, X., & Amidon, L. (2006). A provisional biopharmaceutical classification of the top 200 oral drug products in the United States, great britain, spain, and japan. Molecular pharmaceutics (vol. 3, issue 6, pp.634–639).

Toutain, L., & Bousquet-me Lou, A. (2004). Volumes of distribution. Journal of veterinary pharmacology (vol. 27, issue 6, pp.443–450).

Toutain, L., Bousquet-me Lou, A., & Toutain, P. (2004). Plasma clearance. Journal of veterinary pharmacology (vol. 27, issue 6, p.417).

Tusom. (2015). Introduction to pharmacokinetics. Pharmwikis (vol.1, pp.5–10).

Vivekanandan-giri, A., Slocum, L., Buller, L., Basrur, V., W., Pop-busui, R., Lubman, M., Kretzler, M., & Pennathur, S. (2011). Urine glycoprotein profile reveals novel markers for chronic kidney disease. International journal of proteomics (vol.201, pp. 11–12).

Wagner, G. (1981). History of pharmacokinetics. Pergamon press Ltd (vol. 12, pp.539–550).

Wagner, G. (1968). Kinetics of pharmacologic response: Proposed relationships between response and drug concentration in the intact animal and man. Journal of Theoretical biology (vol. 20, issue 2, pp.174–178).

Wagner, G. (1975). Fundamental of clinical pharmacokinitics (1st ed.). Drug intelligence publications. United States (p. 311)

Waller, G., & Sampson, P. (2018). Pharmacokinetics. In medical pharmacology and therapeutics (5th ed.). Elsevier. United States (p.743).

Wermuth, G. (2003). The practice of medicinal chemistry (2nd ed.). Academic press. United States. (pp. 209–597).

Wesch, R. (2011). B.15 absolute and relative bioavailability. Methods in clinical pharmacology (1st ed.). Elsevier. Netherlands. (pp. 173–189).

Wright, E., Sisson, L., Ichhpurani, K., & Peters, R. (1997). Steady-state pharmacokinetic properties of pramipexole in healthy volunteers. Journal of clinical pharmacology (vol.37, issue 6, pp.520–524).

Wright, J. E. D., Douglas Boudinotl, F., & Ujhelyi, M. R. (1996). Measurement and Analysis of Unbound Drug Concentrations. Clin.Pharmacokinet. (vol. 30, issue 6, pp. 446–452).

Yamasaki, K., chuang, G., Maruyama, T., & Otagiri, M. (2013). Albumin-drug interaction and its clinical implication. In biochimica et biophysica acta - general subjects (vol. 1830, issue 11, p. 5436).

Zhu, L., Lu, L., Wang, S., Wu, J., Shi, J., Yan, T., Xie, M., &Liu, Z. (2017). Oral absorption basics: pathways and physicochemical and biological factors affecting absorption. Developing solid oral dosage forms: pharmaceutical theory and practice (2nd ed.). Academic Press. United Kingdom (p.206).

PART II

METHODOLOGIES, RESULTS AND DISCUSSIONS

II.1. Introduction:

With the aim of improving modelling for prediction of protein Binding Fraction (FB), this part is divided into two contributions:

- 1st contribution: Presenting the methodologies of this work.
- 2nd contribution: Dedicated for results and discussions obtained from using Machine Learning (ML) techniques based on Artificial Intelligence (AI) such as: ANN and DA-SVM, and a Deep Learning (DL) technique which is CNN-MLP develop correlative and/or predictive models to calculate protein Binding Fraction (FB).

II.2. Methodology (Contribution I):

II.2.1 Numerical modelling using artificial intelligence methods for the prediction of protein binding fraction based on descriptors variation:

Artificial Intelligence (AI) methods have become a powerful tool for scientific computing in several science and engineering disciplines, daily life, and various scientific disciplines in recent years (Shen 2018). In this contribution, we aim through a comparative study between the proposed prediction methods: (ANN, CNN-MLP, and DA-SVM), to find the best model that can predict protein Binding Fraction (FB) based on variations in descriptors, with an error approaching 0 and a coefficient of determination R^2 approaching unity.

II.2.2. Database collection:

To improve the modelling of (FB), we used artificial intelligence methods to develop correlational/predictive models related to nonlinear problems based on variations in descriptors. We constructed the database architecture, which is the backbone of this modelling. The approach used in this work is based on a new technique that transforms descriptor values into descriptors features. In this work, we used two types of databases; the first database is formed of normal descriptors while the second database is represented by descriptors features.

The origin of the database used for the two prementioned techniques comes from the chemoinformatic tool-book (RDKit) imported by PYTHON programming language within the SPIDER® when we introduced the SMILES notations of the molecules of active substance (active pharmaceutical ingredient) that we obtained from the PubChem website. The RDKit tool-book provides the top 208 descriptors among the available descriptors. To achieve this, we used the Genetic Algorithm (GA) to reduce the number of descriptors from 208 to 25 descriptors, and this is for the first database. This latter is

used in modelling by the 2 artificial intelligence methods (ANN and DA-SVM) and consists of the values of 25 descriptors for 263 active substances from different classes (view annex) (Brunton 2011; Laurence 2005) with a rectangular matrix of dimension (263; 25) for the input range and (263; 1) for the output range, distributed as follows: 75% for the learning phase and 25% for the test and validation phase (Figure II.1).

The second database consists of descriptors features obtained according to Figure II.2. It is realized by a rectangular matrix of dimension (263; 25) for the input range and (263; 1) for the output range, distributed as follows: 75% for the learning phase and 25% for the test and validation phase. The numerical experiment was conducted (L.-L. Li et al., 2020) on the Microsoft Windows 10 operating system and the software (PYTHON and MATLAB® R2018a). The computer configuration is specifically described as follows: (Intel® Core™ i5-4210M processor, 2.6 GHz, 4 GB RAM, 64-bit).

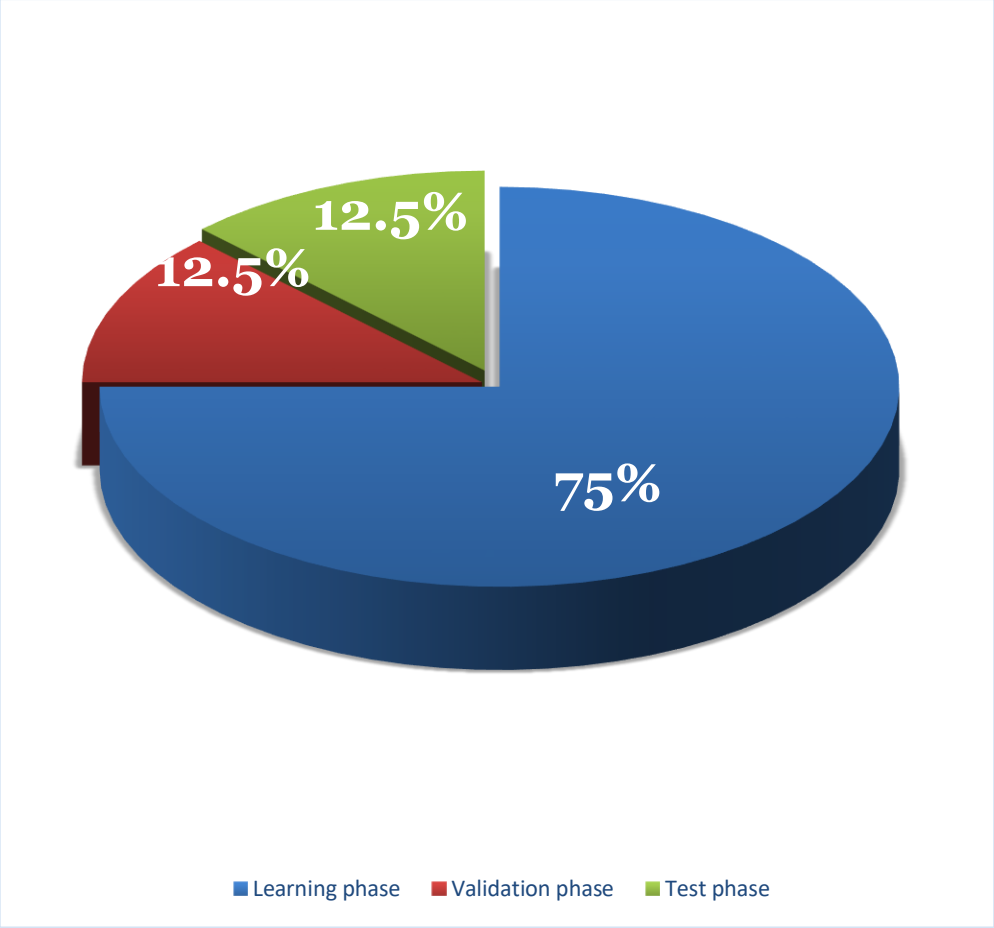


Figure II.1: Representation of database division.

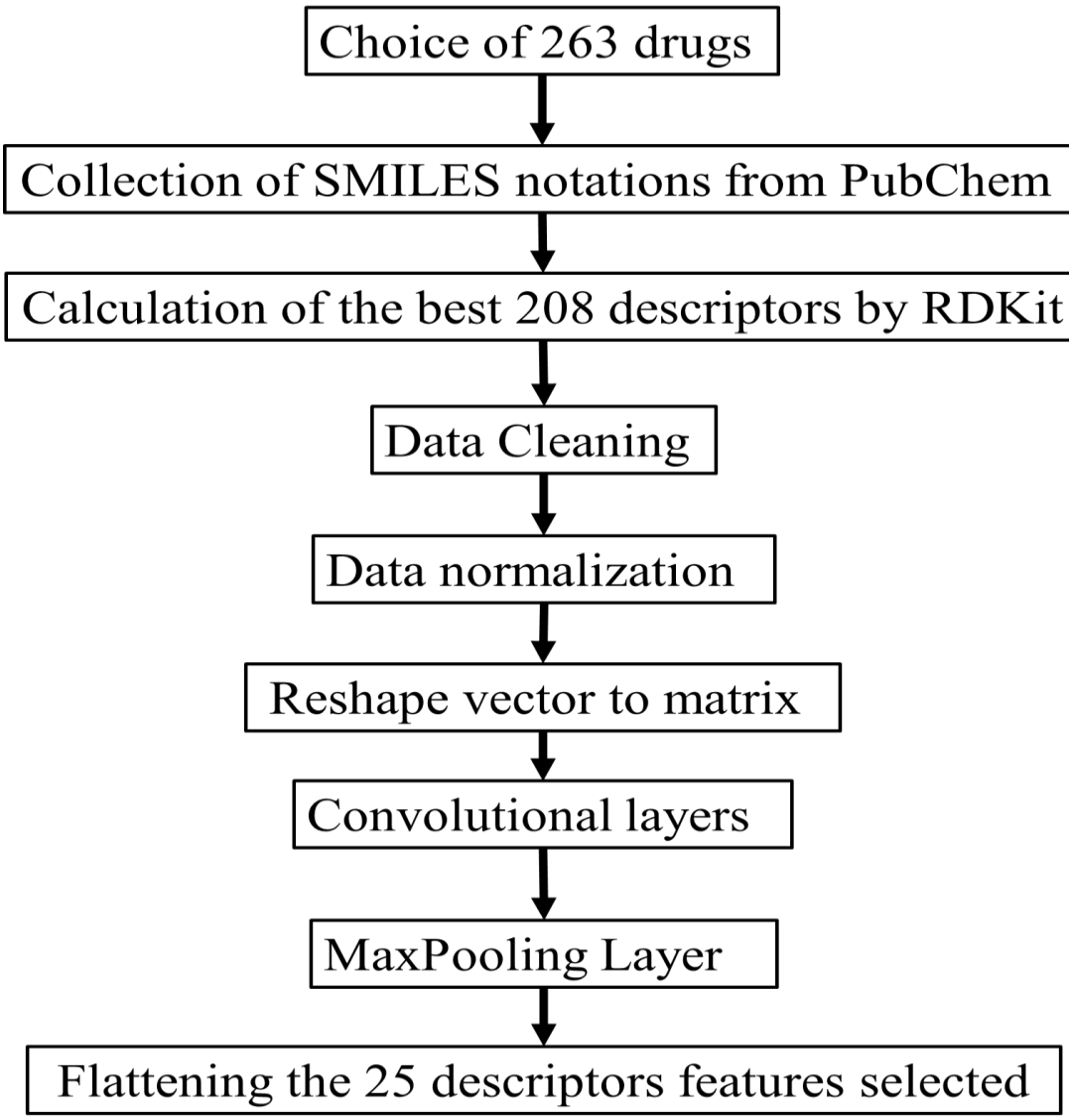


Figure II.2: Feature Construction Flowchart.

II.2.3. Modelling of protein binding fraction using descriptors:

II.2.3.1. Modelling by Artificial Neural Networks (ANN):

Experimental knowledge of FB is often influenced by descriptors such as molecular weight and Lipophilicity. Although ANNs are part of black-box models that represent an empirical relationship between input and output sets (Kanungo et al. 2006), the choice of input and output variables is decisive for the neural model to approximate deterministic models. The selection and collection of data on which artificial intelligence methods are trained constitute the first step in neural modelling. The choice of input and output variables and data pre-processing constitute the second step. Then normalization and division of the database into three phases (learning, testing, and validation). The next step concerns the selection of parameters of the neural network (type of neural network, number of hidden layers and number of neurons in each hidden

layer, transfer function, learning algorithm, and number of iterations) that will allow the network to learn the relationship between the inputs and the chosen output from the selected data. The third step is devoted to the learning, testing, and validation of the chosen network, and then the RMSE is considered as stopping criterion for optimal neural network. Finally, the performance of the network can be, if necessary, compared with those of other types of modelling.

II.2.3.1.1. Calculation Methodology:

The diagram in Figure II.3 illustrates the calculation principle of protein Binding Fractions (FB) using neural networks.

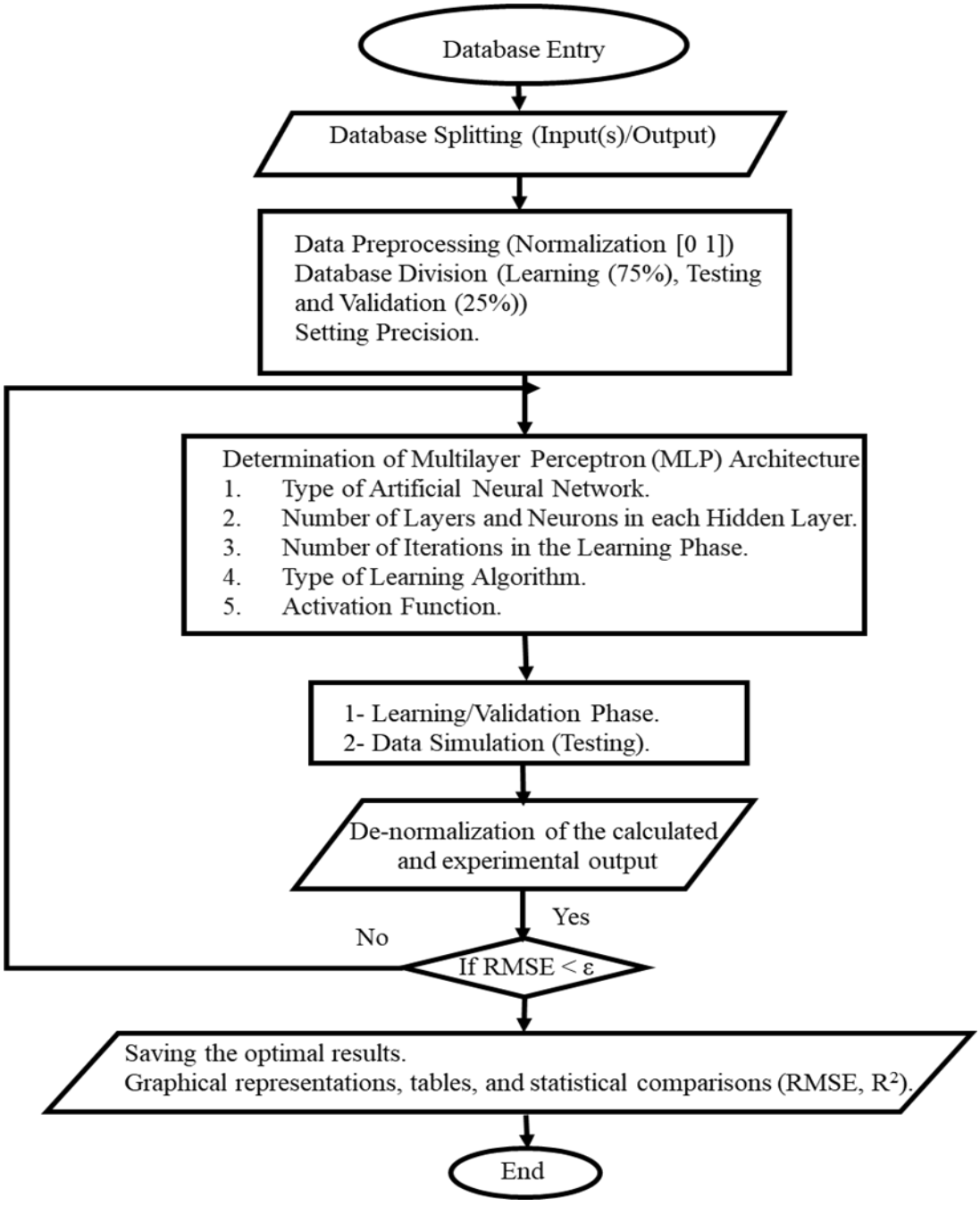


Figure II.3: Flowchart of the methodology for calculating Artificial Neural Networks (ANN).

II.2.3.1.2. Normalization of data:

The data is divided into 26 columns, corresponding respectively to: {Des1, Des2, …, Des25, FB}. The numerical values have been normalized to improve optimization. Normalization is done per row; each input and output parameters (xi) are normalized ($x_{i\ normalized}$) by the following equation (II.1) (MATLAB's premnmx) formula:

$$x_{i\,normalized} = 2\,\frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} - 1 \qquad \text{(II.1)}$$

II.2.3.1.3. De-normalization of data:

All numerical values will be de-normalized to represent these values in their real state. Each normalized input and output parameter ($x_{i\ normalized}$) is transformed into (x_i) by equation (II.2) (Postmnmx in MATLAB):

$$x_i = \frac{1}{2} * (x_{i\,normalized} + 1) * (\max(x_i) - \min(x_i)) + \min(x_i) \qquad \text{(II.2)}$$

II.2.3.1.4. Design of Artificial Neural Network:

Supervised learning is employed in this study, which involves providing the network with sets of input data (Des1, Des2, …, Des25) for which the desired responses (FB) are known. An error criterion is then constructed by comparing the predicted response of (FB) by the ANN model to the experimental (FB) expected response. A synaptic weight correction algorithm will then reduce this error and thus improve the ANN response. The implementation of this technique requires the construction of a leaning (training) database consisting of the entire dataset and associated outputs.

There are no fixed rules for determining the number of hidden layers or the number of neurons in each hidden layer in a MLP neural network for a given problem, except for limitations imposed by some statistical laws based on the heuristic rule that dictates that the number of connections between neurons, which is directly proportional to the number of weights and biases, should never exceed the size of the data matrix. Therefore, a trial-and-error strategy is the most reasonable approach. To obtain the best neural structure, the number of neurons in each hidden layer is varied by the modeller using a MATLAB program to optimize the structure through tests based on statistical error formulas. In this study, the number of neurons in each hidden layer was tested for one hidden layer then two hidden layers with different transfer functions. While the number of neurons was systematically optimized in the ANN algorithm. The number of neurons was varied between 1 and 20 neurons for the two hidden layers.

However, each architecture was repeated 5000 times to avoid convergence to a local minimum. The optimal number of hidden neurons is defined as the number for which the Root Mean Square Error (RMSE) and the generalization cost are minimal. Artificial neurons are arranged in layers, where the input layer receives input signals (u_i) from the experimental world and each successive layer receives the weighted outputs ($w_{j,i}$) from the previous layer as its resulting input, thus forming a feedforward-type ANN (case of Multi-Layer Perceptron MLP), in which each output of a layer is fed forward to the succeeding layer where it is processed. The output of the last layer constitutes the predictions by the ANN model.

In such a feedforward ANN, a neuron in a hidden layer or an output layer has two tasks, it adds the bias to the weighted inputs of several connections and then applies an activation function (also called transfer function or threshold function) to this sum as given by equation (II.3) (for neuron "j" in the first hidden layer).

$$z_j = f\left[\sum_{i=1}^{n} w_{j,i}.u_i + b_j\right], j = 1,2,\dots,m \qquad \text{(II.3)}$$

Then, it propagates the resulting value through outgoing connections to the neurons of the next layer where it undergoes the same calculation process given by equation (II.4):

$$v_k = f\left[\sum_{j=1}^{m} w_{j,k}.z_j + b_k\right], k = 1,2,\dots\dots,l \qquad \text{(II.4)}$$

By combining equations II.3 and II.4, we obtain the relationship (equation II.5) between the output v_k and the inputs of the ANN:

$$v_k = f\left[\sum_{j=1}^{m} w_{j,k}.f\left[\sum_{i=1}^{n} w_{i,j}.u_i + b_j\right] + b_k\right] \qquad \text{(II.5)}$$

Finally, it propagates the resulting value through outgoing connections to the neurons of the next layer where it undergoes the same calculation process given by equation II.6:

$$u_p = f\left[\sum_{k=1}^{l} w_{k,p}.v_k + b_p\right], p = 1,2,\dots\dots,h \qquad \text{(II.6)}$$

By combining equations II.5 and II.6, we obtain the relationship between the output up and the inputs of the ANN (Eq. II.7):

$$u_p = f\left[\sum_{k=1}^{l} w_{k,p}.f\left[\sum_{j=1}^{m} w_{j,k}.f\left[\sum_{i=1}^{n} w_{i,j}.u_i + b_j\right] + b_k\right] + b_p\right] \qquad \text{(II.7)}$$

At the end, the best results in terms of error are saved.

II.2.3.2. Modelling by Support Vector Machine coupled with the Dragonfly Algorithm (DA-SVM):

In this section, the hybrid model of DA-SVM is utilized, where the proposed model is based on the Support Vector Machine (SVM) learning algorithm coupled with the Dragonfly Algorithm (DA) for the optimization of its hyperparameters. This model is employed for the nonlinear modelling of Protein Binding Fractions based on descriptors using the same experimental data used to develop the ANN model. The data is randomly partitioned into two phases: 75% for the learning phase and 25% for the testing phase.

The DA-SVM model generates a nonlinear relationship between the 25 inputs, which are (fr_nitro_arom, MolLogP, fr_Ndealkylation2, MaxAbsPartialCharge, fr_alkyl_halide, fr_quatN, TPSA, PEOE_VSA13, PEOE_VSA6, fr_epoxide, Chi2n, VSA_EState8, qed, fr_halogen, Kappa3, SMR_VSA4, fr_amide, fr_hdrzone, fr_aryl_methyl, fr_lactone, PEOE_VSA8, MolWt, Chi3v, EState_VSA10, FpDensityMorgan1), and the output, which is protein Binding Fraction (FB). Therefore, a kernel function is required for the implicit mapping of the learning patterns into the feature space. A kernel function can be any mathematical function that satisfies the Mercer's theorem. Obtaining a better structure of the predictive model is based on systematically testing its regularization parameters during the optimization process by the modeller using a MATLAB program, through statistical error formulas. However, this can be a challenging task and may lead to overfitting of the model selection criterion (usually based on cross-validation) when more than a few hyper parameters need to be adjusted. In the present SVM modelling, the kernel function used is the Gaussian Radial Basis Function (GRBF). The mathematical expression of the tested kernel function is as follows. (Eq II.8) (Min and Lee 2005; Scholkopf and Smola 2001; Keskes et al. 2020):

Gaussian Radial Basic Function (GRBF):

$$k(x_i . x_j) = exp\left(-1/2\,\sigma^2 * \left|x_i - x_j\right|^2\right) \qquad \text{(II.8)}$$

Where σ is the width of the RBF function defined by the user. The flowchart of the SVM method is presented in Figure II.4.

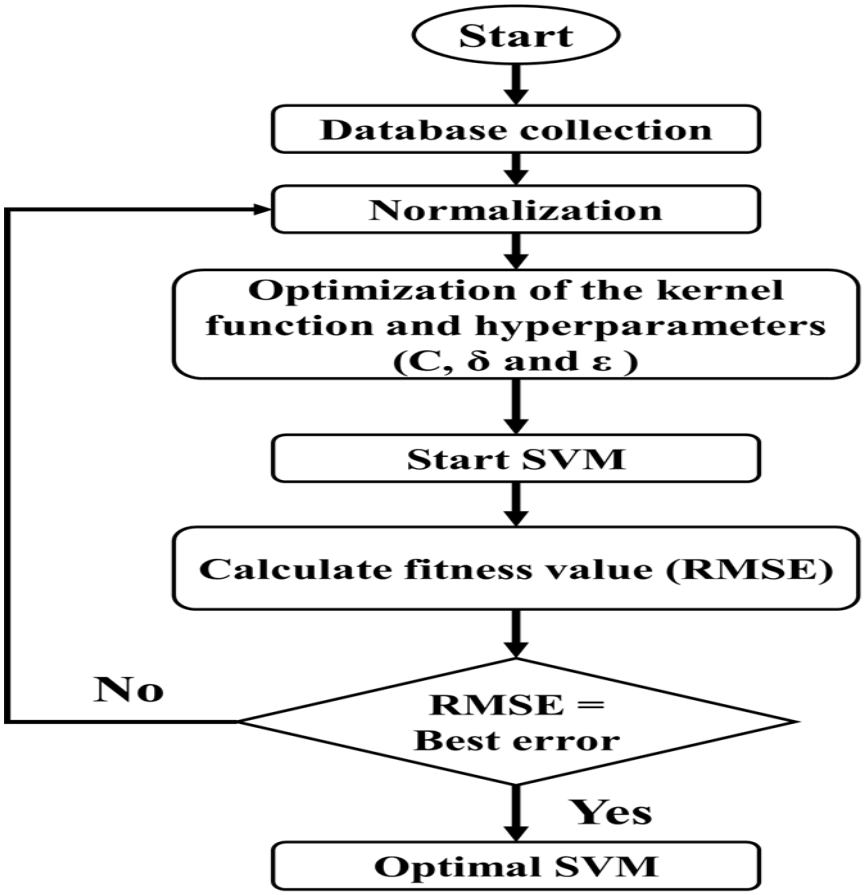


Figure II.4: Flowchart of SVM.

To obtain a high-performing SVM model with good predictive capacity, there are three parameters that need to be carefully chosen (X. li Li et al. 2013; Kanungo et al. 2006). These parameters, known as the SVM model's hyperparameters, include the regularization parameter C, the epsilon-insensitivity coefficient ε, and the kernel parameters (bandwidth of the Gaussian RBF kernel σ in this study). These hyperparameters play a significant role in the regression accuracy and generalization performance of the SVM model (X. li Li et al. 2013; Kanungo et al. 2006). Therefore, optimal values of σ, C, and ε are obtained through trial-and-error testing. Table II.1 represents the ranges of variation of the hyperparameters used in this modelling.

Table II.1: Ranges of initial parameters for DA-SVM.

hyperparameters of DA-SVM	C	σ	ε
Inferior limit	20	0.3	0
Superior limit	100	0.6	10^{-2}
Number of agents		40	
Maximum number of iterations		120	
Learning phase (%)		75	
Testing phase (%)		25	

The coefficient of determination R^2 and the Root Mean Square Error RMSE are the statistical parameters used for evaluating the performance of these models.

The flowchart of the Dragonfly Algorithm (DA) metaheuristic algorithm is presented in Figure II.5.

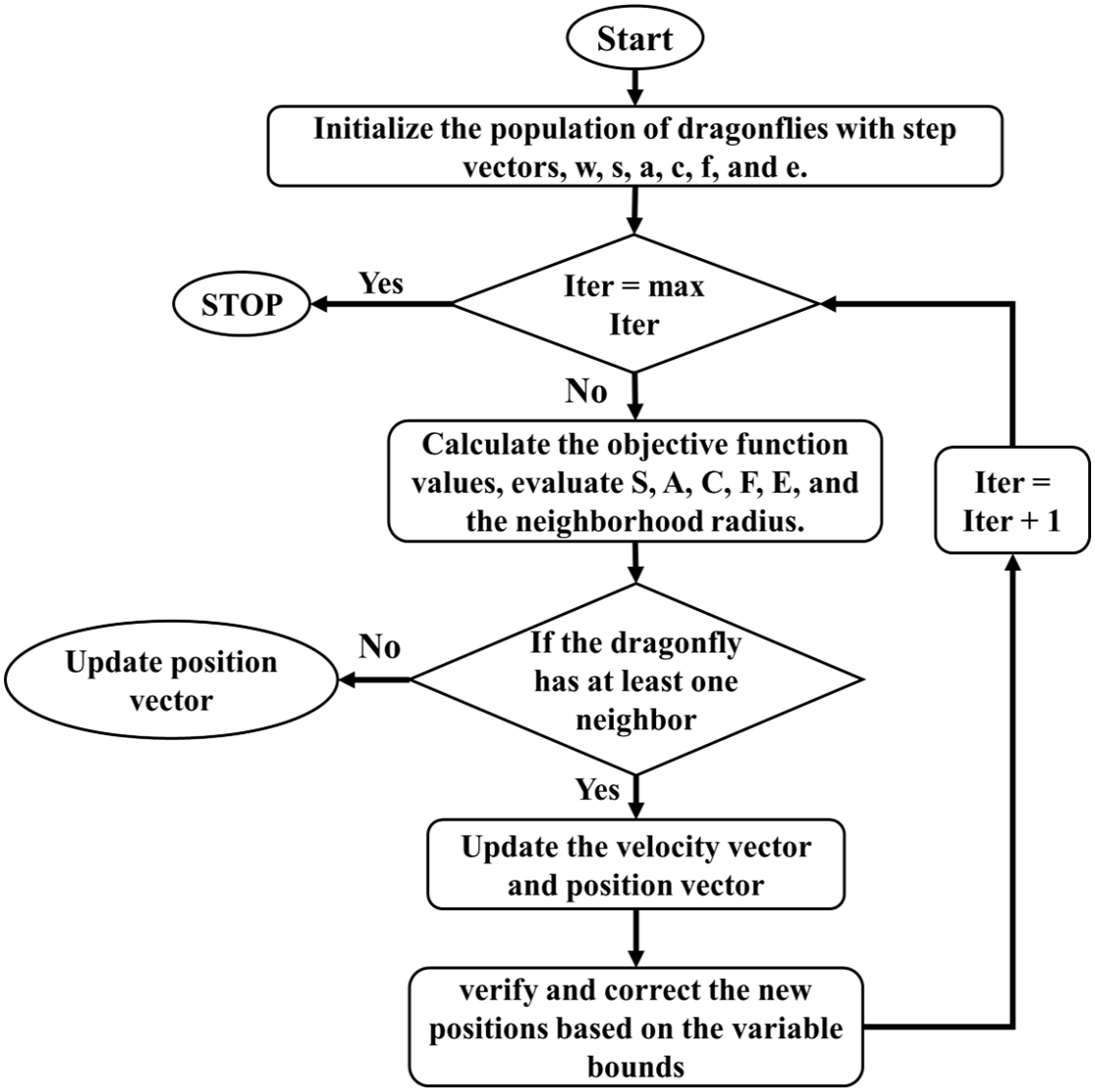


Figure II.5: The flowchart of the Dragonfly Algorithm (DA).

The DA feeds the SVM with random values of hyperparameters within their predefined range.

Flow of the Dragonfly Algorithm coupled with SVM represented in 8 steps:

- Step 01: Define the values of the DA parameters such as:
  - Dimension of parameters to optimize;
  - Number of dragonflies;
  - Maximum number of iterations;
  - Upper and lower limits of C, σ, and ε.

- Step 02: Initialize the step vectors ΔXi (i = 1, 2, ..., n), the values of s, a, c, f, e, then initialize the inertia weight (w) and the SVM hyperparameters values (C, σ, and ε).
- Step 03: Train the SVM model using the training set and calculate the physical value of each dragonfly. The Root Mean Square Error (RMSE) is used as the fitness function.
- Step 04: Update the values of s, a, c, f, e, and w.
- Step 05: For all individuals, calculate the values of S, A, C, F, and E.
- Step 06: Update the neighbourhood radius (r).
- Step 07: Correct the new positions taking into account the upper and lower bounds of the variables C, σ, and ε.
- Step 08: Check the stopping criterion: if its criterion is met, the best position for all individuals including the optimized SVM hyperparameters is selected, then the SVM model is tested and evaluated. Otherwise, return to step 3.

The diagram of the DA-SVMr model applied for predicting the protein binding fractions presented in Figure II.6.

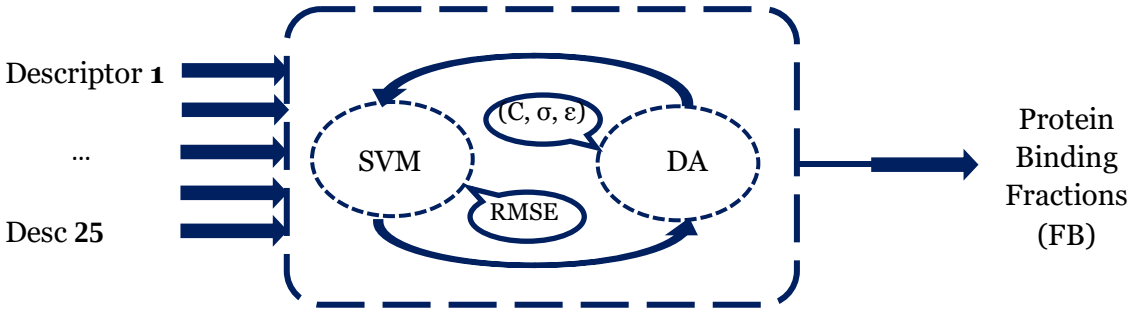


Figure II.6: Schematic representation of the DA-SVMr model applied to FB prediction.

The flowchart of the steps leading to the development of the hybrid DA-SVM model is illustrated in Figure II.7.

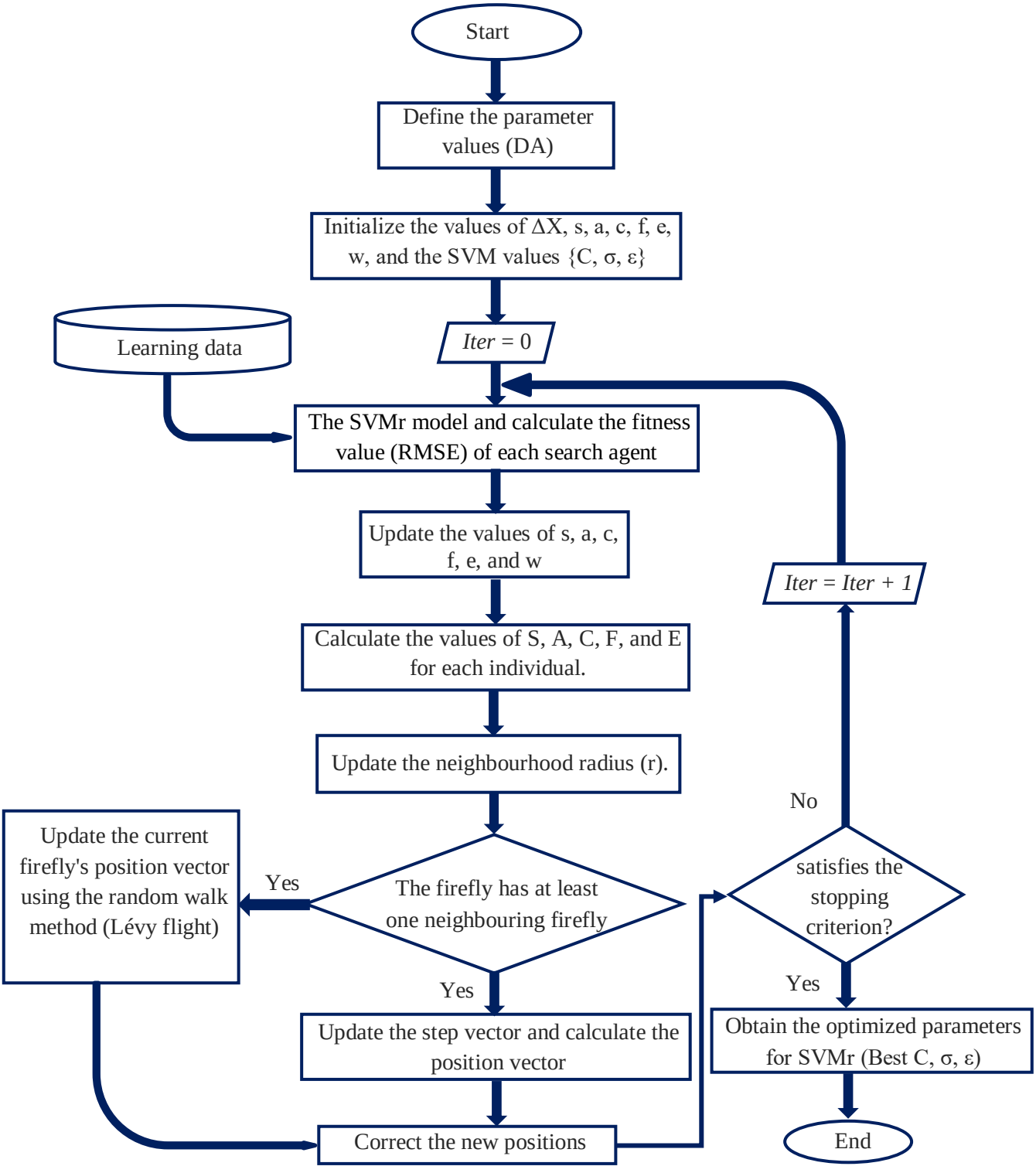


Figure II.7: Flowchart of the DA-SVM model.

II.2.4. Modelling of protein binding fraction using descriptors features:

II.2.4.1. Modelling by Convolutional Neural Networks and Multilayer Perceptron (CNN-MLP):

The CNN is a type of deep neural network. The main benefit of CNN compared to its pre-existing methods is that it automatically identifies the features without any human supervision (Gu and al. 2018). CNN have been extensively applied in a range of different fields, such as Face Recognition (Hubel and Wiesel 1962). The structure of CNN was inspired by neurons in human and animal brains, more specifically, in a cat's brain (H. C. Li, Deng, and Chiang 2020). The CNN uses 3D input data structures like image signals. A commonly used type of CNN, which is similar to the Multi-Layer Perceptron (MLP), consists of numerous convolution layers preceding sub-sampling (pooling) layers, while the ending layers are Fully Connected (FC) layers. An example of a basic CNN architecture is illustrated in Figure II.8. The input of each layer in a CNN model is organized in three dimensions: height, width, and depth, or m × m × r, where the height (m) is equal to the width. Several kernels (filters) available in each convolutional layer are denoted by K and also have three dimensions (n × n × q), similar to the input. Furthermore, Multiple layers perceptron MLP is a type of artificial neural network that includes multiple hidden layers, which each layer is fully connected to the next (FC) (Alzubaidi and al. 2021).

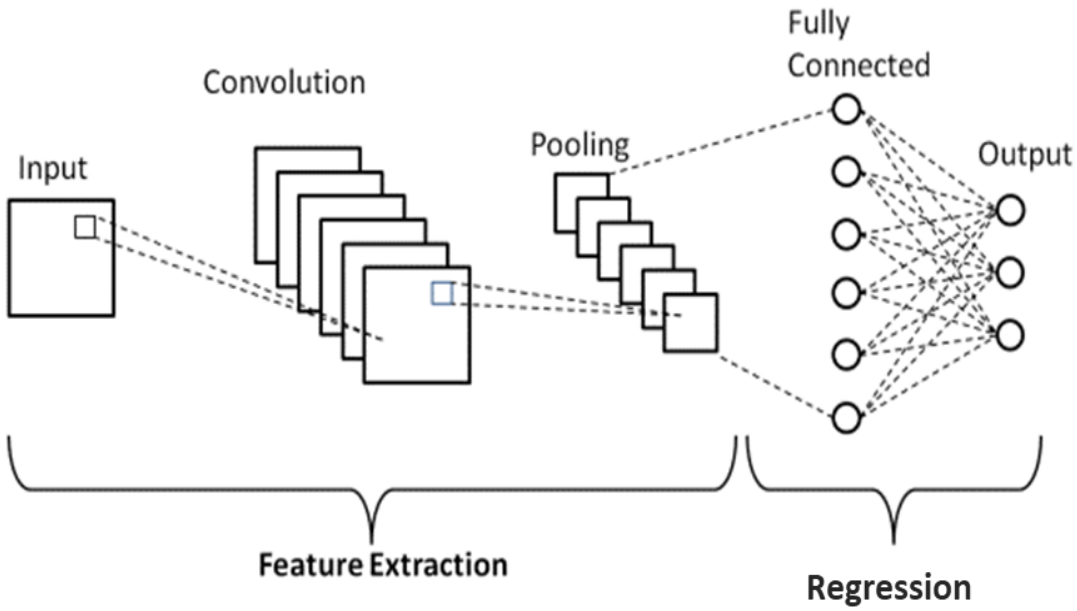


Figure II.8: Basic CNN architecture.

II.2.4.1.1. The design of CNN-MLP:

The 208 descriptors calculated by the RDKit tool-book are used as inputs to predict the output which is the protein Binding Fractions (FB) by combining CNN and MLP. After the data cleaning, where missing values in the databases are replaced with the value 0, which is a neutral value that does not affect the calculations. The next step is the normalization and division of the database into three phases (learning, test, validation) same as the previous method ANN. The inputs variables in CNN should be used in a 2D matrix. The data base includes 263 active substances, each substance was described by 1D vector which contains 208 descriptors, this later is reshaped into 2D matrix with the size of (15,15) input for active substance.

Unlike the previous Machine Learning (ML) methods used (ANN and DA-SVM), the CNN in this case learns automatically the meaningful features from the inputs, the suitable convolutional kernels (filters) with a size of (3,3) designed to process each matrix individually for the features extraction in two convolutional layers (Conv2D), with the use of ReLU (Rectified Linear Unit) as an activation function with the learning algorithm gradient descent. The spatial dimensions (width and height) of the resulting feature maps are reduced by a pooling window with the size of (2,2) used to compute the maximum local value in a one Maxpooling layer. After that the outputs of the Maxpooling layer are converted into a 1D vector, this vector is fed into five dense layers. The purpose of these layers is to further process the features extracted by the convolutional and pooling layers. These dense layers contain 1024, 512, 256, 128, 64 neurons in each layer respectively for the prediction of a single output that is the protein Binding Fraction (FB) as it's shown in figure II.9, using ReLU with the same activation function used in the previous ANN, noting that the weights and biases change. By the following equation (II.9):

$$z_j = \mathrm{ReLU}\left[\sum_{i=1}^{n} w_{j,i}.\, u_i + b_j\right], j = 1,2,\dots,m \qquad \text{(II.9)}$$

Then, it propagates the resulting value through outgoing connections to the neurons of the next layer where it undergoes the same calculation process given by equation (II.10):

$$v_k = \mathrm{ReLU}\left[\sum_{j=1}^{m} w_{j,k}\,.\, z_j + b_k\right], k = 1,2,\dots\dots,l \qquad \text{(II.10)}$$

By combining equations II.9 and II.10, we obtain the relationship equation (II.11) between the output v_k and the inputs of the NN:

$$v_k = \mathrm{ReLU}\left[\sum_{j=1}^{m} w_{j,k}.\, f\left[\sum_{i=1}^{n} w_{i,j}.\, u_i + b_j\right] + b_k\right] \qquad \text{(II.11)}$$

Finally, it propagates the resulting value through outgoing connections to the neurons of the next layer where it undergoes the same calculation process given by equation (II.12):

$$u_p = \mathrm{ReLU}\left[\sum_{k=1}^{l} w_{k,p}.\, v_k + b_p\right], p = 1,2, \ldots \ldots, h \qquad \text{(II.12)}$$

By combining equations II.11 and II.12, we obtain the relationship between the output u_p and the inputs of NN (II.13):

$$u_p = \mathrm{ReLU}\left[\sum_{k=1}^{l} w_{k,p}.\, f\left[\sum_{j=1}^{m} w_{j,k}.\, f\left[\sum_{i=1}^{n} w_{i,j}.\, u_i + b_j\right] + b_k\right] + b_p\right] \qquad \text{(II.13)}$$

By the same method, the five hidden layers are connected until with output (FB) presented in Figure II.9.

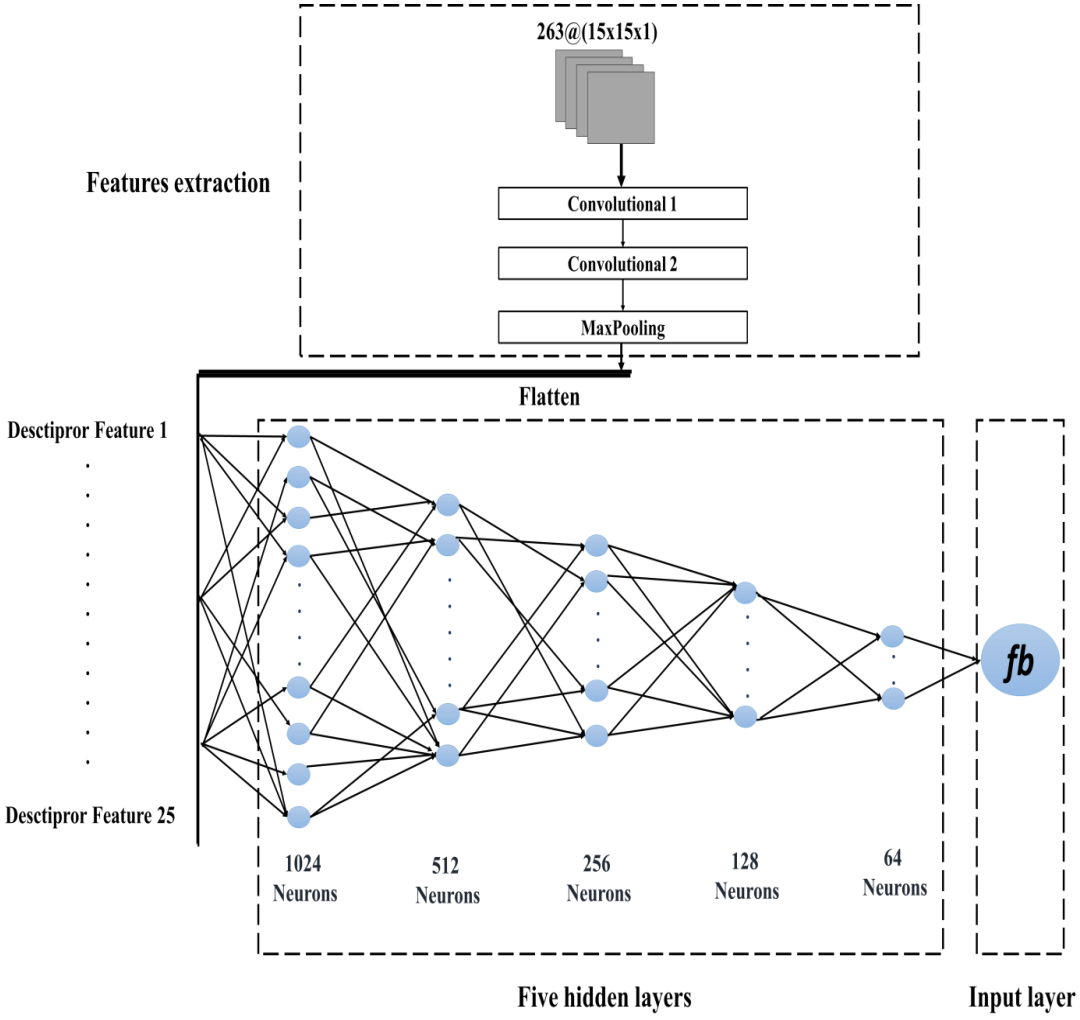


Figure II.9: CNN-MLP architecture for the prediction of protein binding fractions.

II.2.4.2. Modelling by Convolutional Neural Networks coupled with Dragonfly Algorithm-Support Vector Machine (CNN-DA-SVM):

Combining Convolutional Neural Networks (CNN) with the Dragonfly Algorithm (DA) and Support Vector Machine (SVM), known as DA-SVM, involves utilizing the CNN for descriptors features extraction and then feeding the extracted features into the SVM, which is optimized using the Dragonfly Algorithm. This coupling allows for the effective utilization of CNN capabilities in extracting complex features from the database. The CNN first processes the input data, extracting relevant features through convolutional and pooling layers. These features are then flattened. The flattened features serve as input to the DA-SVM regression model, which learns to predict continuous output variables. The Dragonfly Algorithm optimizes the hyperparameters of the SVM, such as regularization parameters and kernel parameters, enhancing the model's performance. This coupled approach integrates the feature learning capabilities of CNNs with the optimization power of DA-SVM.

II.2.5. Statistical Performance Criteria (Analysis and Evaluation of a Model):

The reliability and accuracy of models developed from artificial intelligence methods rely on various statistical indices. The evaluation of the models proposed in this study was conducted using seven statistical performance criteria to test predictive power. Commonly used evaluation indices include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Normalized Root Mean Squared Error (NRMSE), Mean Absolute Error (MAE), the Absolute Relative Deviation (ARD), correlation coefficient (R), and determination coefficient (R^2), given by equations (II.14) to (II.20).

$$MSE = \frac{1}{n} * \sum_{i=1}^{n} (Y - Y_{cal})^2 \quad \text{(II.14)}$$

$$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^{n} (Y_{cal} - Y)^2} \quad \text{(II.15)}$$

$$NRMSE = \sqrt{\frac{n \sum_{i=1}^{n} (Y - Y_{cal})^2}{\left(\sum_{i=1}^{n} Y\right)^2}} \quad \text{(II.16)}$$

$$MAE = \frac{1}{n} * \sum_{i=1}^{n} |Y_{cal} - Y| \quad \text{(II.17)}$$

$$ARD = \left| \frac{(Y_{Cal} - Y)}{Y_{Cal}} \right| * 100 \quad \text{(II.18)}$$

$$R = \frac{\sum_{i=1}^{n} ((Y - \bar{Y}) * (Y_{cal} - \bar{Y}_{cal}))}{\sqrt{\sum_{i=1}^{n} (Y - \bar{Y})^2 * \sum_{i=1}^{n} (Y_{cal} - \bar{Y}_{cal})^2}} \quad \text{(II.19)}$$

$$R^2 = 1 - \left(\frac{\sum_{i=1}^{n} ((Y - Y_{cal})^2)}{\sum_{i=1}^{n} ((Y - \bar{Y})^2)} \right) \quad \text{(II.20)}$$

II.3. Results and discussions (*Contribution II*):

II.3.1. Modelling of protein binding fraction using descriptors:

II.3.1.1. Statistical study of protein binding fractions and descriptors:

The variable selection procedure which is Genetic Algorithm (GA), generated 25 descriptors. Descriptive statistics are used to quantitatively describe or summarize the characteristics of the data in order to explain further their implication. This typically involves calculating various statistical measures such as: the maximum and minimum, mean, variance, STD. Mean is a statistical term introduced to explain the average trend of data. The variance is the average of the squared differences from the mean (Eq. II.21). To determine the variance, we first calculate the difference between each point; then, we square and divide the sum by N-1. Standard Deviation (STD) is a measure to quantify the amount of variation or dispersion of data values (Eq. II.22), it is also a statistic that examines how much distance from the mean in a group of numbers, using the square root of the variance.

A higher STD shows that the data points are spread over a wider range of values, a low STD indicates that the data points tend to be close to the mean of the dataset (Wang et al. 2017).

$$Variance = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N-1} \qquad (II.21)$$

$$STD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N-1}} \qquad (II.22)$$

Where x_i is the experimental values, $\bar{x}$ is the mean of the experimental data and N represents the size of the experimental data.

The results are represented in Table II.2

Table II.2: Statistical description of the entire dataset.

Category of variables	Variables	Variable ranges (Min - Max)	MEAN	STD	Variance
Inputs	fr_nitro_arom	0 -1	0.019	0.136	0.018
	MolLogP	-8.4242 - 8.948	2.303	1.955	3.824
	fr_Ndealkylation2	0 – 2	0.106	0.332	0.110
	MaxAbsPartialCharge	0.240997 - 0.572597	0.432	0.071	0.005
	fr_alkyl_halide	0 – 6	0.156	0.766	0.587
	fr_quatN	0 – 2	0.0228	0.173	0.030
	TPSA	3.24 - 331.94	79.198	45.94	2110.5
	PEOE_VSA13	0 - 16.02530659	2.5124	4.136	17.107
	PEOE_VSA6	0 - 78.34063068	22.948	18.080	326.860
	fr_epoxide	0 – 1	0.004	0.061	0.004
	Chi2n	0.826603 - 18.73759	6.215	2.812	7.912
	VSA_EState8	-2.21285 - 17.72703	3.452	3.332	11.101
	Qed	0.095298 - 0.940384	0.639	0.207	0.043
	fr_halogen	0 – 7	0.410	0.913	0.835
	Kappa3	0.539771 - 10.28456	3.555	1.662	2.763
	SMR_VSA4	0 - 50.24366	5.040	8.826	77.906
	fr_amide	0 – 6	0.593	0.977	0.956
	fr_hdrzone	0 – 1	0.007	0.086	0.007
	fr_aryl_methyl	0 – 2	0.281	0.562	0.316
	fr_lactone	0 – 1	0.023	0.150	0.022
	PEOE_VSA8	0 - 90.93631	18.676	14.290	204.17
	MolWt	126.004 - 807.89	341.28	113.65	12917
	Chi3v	0.658608 - 15.85828	5.017	2.530	6.401
	EState_VSA10	0 - 45.64676	11.623	8.716	75.976
	FpDensityMorgan1	0.647058824 - 1.8333	1.180	0.206	0.042
Output	FB	0 - 0.999	0.685	0.334	0.111

II.3.1.2. Prediction of protein binding fractions using Artificial Neural Networks (ANN):

MATLAB® software is used in this work to predict the protein binding fractions in function of descriptors. ANN architecture optimization consist the determination of:

- The number of layers and the neurons in each hidden layer.
- The learning algorithm, by comparing the performance of the NNs based on learning, testing and validation in terms of the Root Mean Square Error (RMSE).
- The activation functions of the hidden layer and the output layer (Reynel-Avila and al. 2015).

II.3.1.2.1. Choice of ANN parameters:

The best architecture of the neuronal model obtained is composed of 25 neurons in the input layer and two hidden layers, using the **newff** function, which corresponds to the NN type feedforward neural network (FFNN), with the Levenberg-Marquardt backpropagation learning algorithm (MATLAB® **trainlm** function), the sigmoid tangent activation function (MATLAB® **tansig** function) for neurons of the hidden layers and the purelin function (MATLAB® **purelin** function) for the output layer.

The structure parameters of the Optimized ANN (OANN) for the prediction of protein binding fractions are summarized in table II.3.

Table II.3: Optimized ANN structure.

Network type	**Learning Algorithm**	**Input Layer**	**Hidden Layer**			**Output layer**
		Nb. Of neurons	**Structure**	**Hidden Layer 1**	**Hidden Layer 2**	
FFNN-BP (MATLAB **newff** function).	Levenberg-Marquardt. (MATLAB **trainlm** function).		**Nb. Of neurones**	8	18	1
		25	**Activation Function**	(MATLAB **tansig** Function).	(MATLAB **tansig** Function).	(MATLAB **purelin** Function).

II.3.1.2.2. Concept of feed forward neural network type MLP:

Artificial neurons are organized in layers, where the input layer receives input signals (u_i) from the real world and each successive layer receives the weighted outputs ($w_{i,j}$) of the layer former. The outputs of the last layer constitute the outputs to the real world. The diagram of ANN type MLP is used to model the protein binding fraction and is shown in Figure II.10.

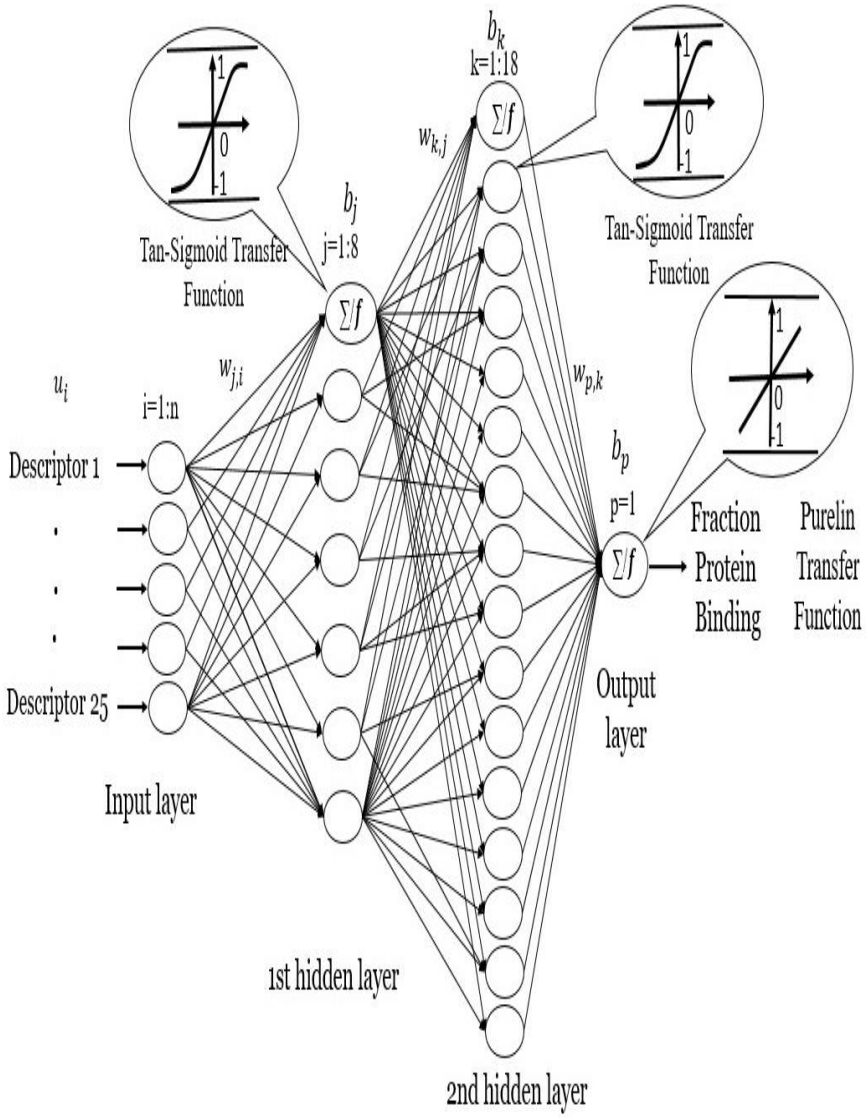


Figure II.10: Feedforward artificial neural network

The difference in error for the numbers of neurons in each hidden layer shows in Figure II.11.

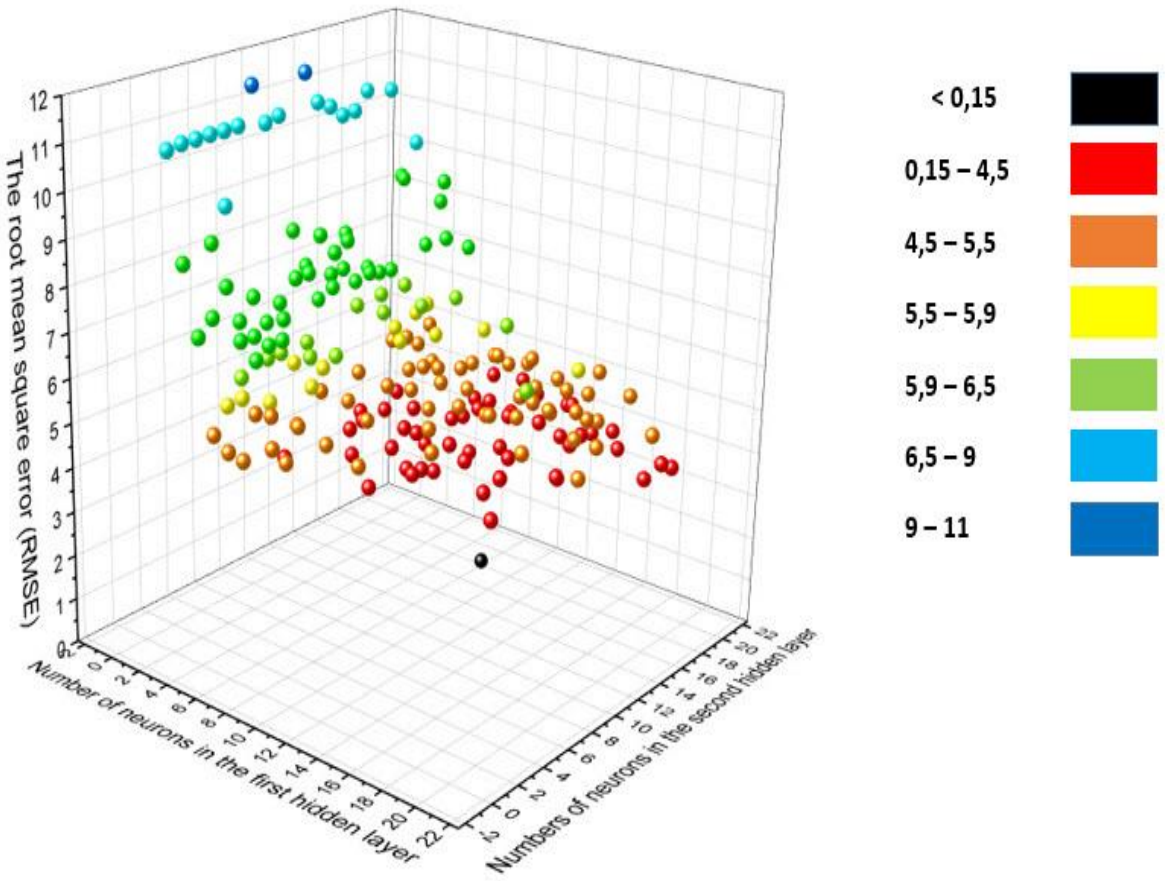


Figure II.11: The difference in error for the numbers of neurons in each hidden layer.

Tables II.4 to II.6 summarize the weights and biases of the two hidden layers and the optimized neural network output layer.

Table II.4: Established neural network model {(Input layer – 1st hidden layer)}.

Connexion [input layer – 1st hidden layer]																							
Weight matrix																							**Bias vector**
$W_{i,1}$	$W_{i,2}$	$W_{i,3}$	$W_{i,4}$	$W_{i,5}$	$W_{i,6}$	$W_{i,7}$	$W_{i,8}$	$W_{i,9}$	$W_{i,10}$	$W_{i,11}$	$W_{i,12}$	$W_{i,13}$	$W_{i,14}$	$W_{i,15}$	$W_{i,16}$	$W_{i,17}$	$W_{i,18}$	$W_{i,19}$	$W_{i,20}$	$W_{i,21}$	$W_{i,22}$	$W_{i,23}$	$W_{i,24}$	$W_{i,25}$	b_j
-0.0557	-1.22660	1.3020	-0.22242	0.0601	0.0243	0.6665	0.0668	-0.03381	-0.01114	0.1230	-1.06685	0.1811	-0.05551	0.0010	0.3748	-0.7876	-0.0636	-0.1433	-0.0269	-0.7299	-0.0900	-0.4686	0.0006	-0.1637	-0.0268
-0.3533	0.0802	-0.1049	-0.1127	0.2231	-0.0844	-0.1049	0.0852	0.2059	-0.38636	0.2803	0.4294	-0.9892	0.1275	-0.0048	-0.0968	0.2674	-0.3872	0.0817	-0.2165	0.4924	0.5243	0.2157	-1.0423	-0.3038	0.3503
-0.2108	0.7948	1.0541	-0.3720	-0.1081	0.1188	0.6111	0.6393	0.4945	-0.08335	-0.00045	0.1747	-0.2930	0.5471	-0.6956	-0.6809	-0.6196	0.0237	-0.0379	-0.1905	-0.2456	-0.5430	-0.8042	0.3984	-0.0581	0.1200
-0.0737	0.0873	0.3653	-0.0166	-0.2747	0.1890	-0.5384	0.4461	-0.6418	-0.0737	-0.3044	0.0103	-0.0404	0.5310	0.5462	-0.2202	0.7570	-0.3255	-0.6187	-0.1486	0.7536	-0.4941	-0.1672	0.8147	0.1655	0.0578
0.0250	0.4091	-0.0810	-0.1382	-0.0725	-0.3428	-0.4308	0.6903	1.2753	-0.01118	-0.4475	-0.1895	0.2528	0.1891	-0.3337	-0.3091	-0.0688	0.0277	-0.0775	0.1634	0.2860	-0.2844	-0.0939	-0.3776	0.5897	-0.0029
-0.2430	0.2780	0.0188	0.4064	-0.1570	0.0753	0.2381	-0.2452	0.4696	-0.34666	0.4321	-0.9207	0.7957	-0.5806	-0.1146	1.3558	0.9306	-0.2915	-0.1058	-0.2416	-0.0648	0.3693	-0.1783	-0.2177	-0.4912	0.3789
0.1482	0.2619	-0.2021	-0.0743	-0.6747	0.2989	-0.1011	-0.7379	0.2205	0.1081	0.6529	0.1285	0.8026	0.0478	-0.6752	0.7597	0.2953	0.2959	0.1797	-0.2059	-1.0711	-0.3824	-0.2577	0.4189	0.1445	-0.1303
0.1388	-0.2689	-0.3725	-0.6270	0.0352	0.3350	-0.6701	-0.3177	-0.5980	0.0543	-0.2622	-0.3093	-0.4770	0.0291	0.3421	0.8485	0.4920	-0.0020	-0.4487	-0.1903	0.0991	0.0350	-0.3233	-0.3404	0.7630	-0.0316

Table II.5: Established neural network model {(1st hidden layer – 2nd hidden layer)}.

Connection [1st hidden layer – 2nd hidden layer]									
Weight matrix									Bias vector (bp)
weight	1	2	3	4	5	6	7	8	
$W_{m,1}$	-0.1094	-0.0185	-0.0716	-0.1084	-0.0324	0.0738	-0.0556	0.0968	-0.0066
$W_{m,2}$	0.1099	0.0185	0.0719	0.1089	0.0325	-0.0740	0.0558	-0.0971	0.0065
$W_{m,3}$	0.0947	0.0169	0.0613	0.0928	0.0273	-0.0666	0.0490	-0.0866	0.0078
$W_{m,4}$	0.0748	0.0139	0.0479	0.0727	0.0208	-0.0550	0.0391	-0.0704	0.0077
$W_{m,5}$	1.6898	1.1716	-1.9551	1.5626	0.9234	-1.6478	1.3304	0.2093	0.0660
$W_{m,6}$	-0.4556	0.4431	-0.1331	0.0636	0.1258	-0.6689	0.2378	-0.3376	-0.4162
$W_{m,7}$	-0.1933	-0.7518	-0.1623	0.0662	-0.0516	-0.0087	-0.4039	1.8368	-0.3851
$W_{m,8}$	-0.0979	-0.0174	-0.0636	-0.0962	-0.0284	0.0683	-0.0505	0.0890	-0.0076
$W_{m,9}$	-0.1249	-0.0191	-0.0826	-0.1256	-0.0381	0.0799	-0.0614	0.1052	-0.0036
$W_{m,10}$	-0.2137	-0.1409	0.2384	-0.3484	0.8523	0.0117	-0.7588	-0.6631	-0.4396
$W_{m,11}$	-0.2825	-0.0897	0.8984	0.2275	-0.9388	-0.2207	-0.1011	0.1652	0.2784
$W_{m,12}$	0.1030	0.0179	0.0671	0.1015	0.0301	-0.0709	0.0528	-0.0926	0.0073
$W_{m,13}$	-0.1032	-0.0179	-0.0672	-0.1017	-0.0302	0.0710	-0.0529	0.0927	-0.0072
$W_{m,14}$	0.0865	0.0158	0.0558	0.0845	0.0246	-0.0621	0.0450	-0.0802	0.0080
$W_{m,15}$	0.0825	0.0152	0.0531	0.0804	0.0233	-0.0598	0.0430	-0.0769	0.0080
$W_{m,16}$	-0.0998	-0.0176	-0.0649	-0.0982	-0.0290	0.0693	-0.0514	0.0903	-0.0075
$W_{m,17}$	-0.5868	0.0453	-0.0180	-0.2316	-0.4074	0.0595	0.6843	0.2181	-0.4182
$W_{m,18}$	0.1015	0.0178	0.0660	0.0999	0.0296	-0.0701	0.0521	-0.0915	0.0074

Table II.6: Established neural network model {(2nd hidden layer - output layer)}.

Connection [2nd hidden layer – output layer]																		
Weight matrix																		**Bias vector (bz)**
weight	**1**	**2**	**3**	**4**	**5**	**6**	7	**8**	**9**	**10**	11	12	13	14	15	16	17	18	
$W_{m,1}$	-0.2196	0.2204	0.1920	0.1531	-3.0182	0.8433	1.7323	-0.1982	-0.2468	-1.1541	-0.9321	0.2077	-0.2081	0.1762	0.1684	-0.2018	-0.8835	0.2049	0.0686

II.3.1.2.3. Results interpretation:

The performance results of the neural network model established for the learning, testing, validation and global phases for the prediction of protein Binding Fractions (FB) rates are presented in Figures II.12 to II.15 represent the regression curves of the data calculated by the ANN model in relation to the experimental data, with a coefficient of determination (R^2) value indicating low correlation [α (slope), β (y-intercept), R^2(coefficient of determination)] = [0.62, 0.35, 0.25] $_{Test.}$

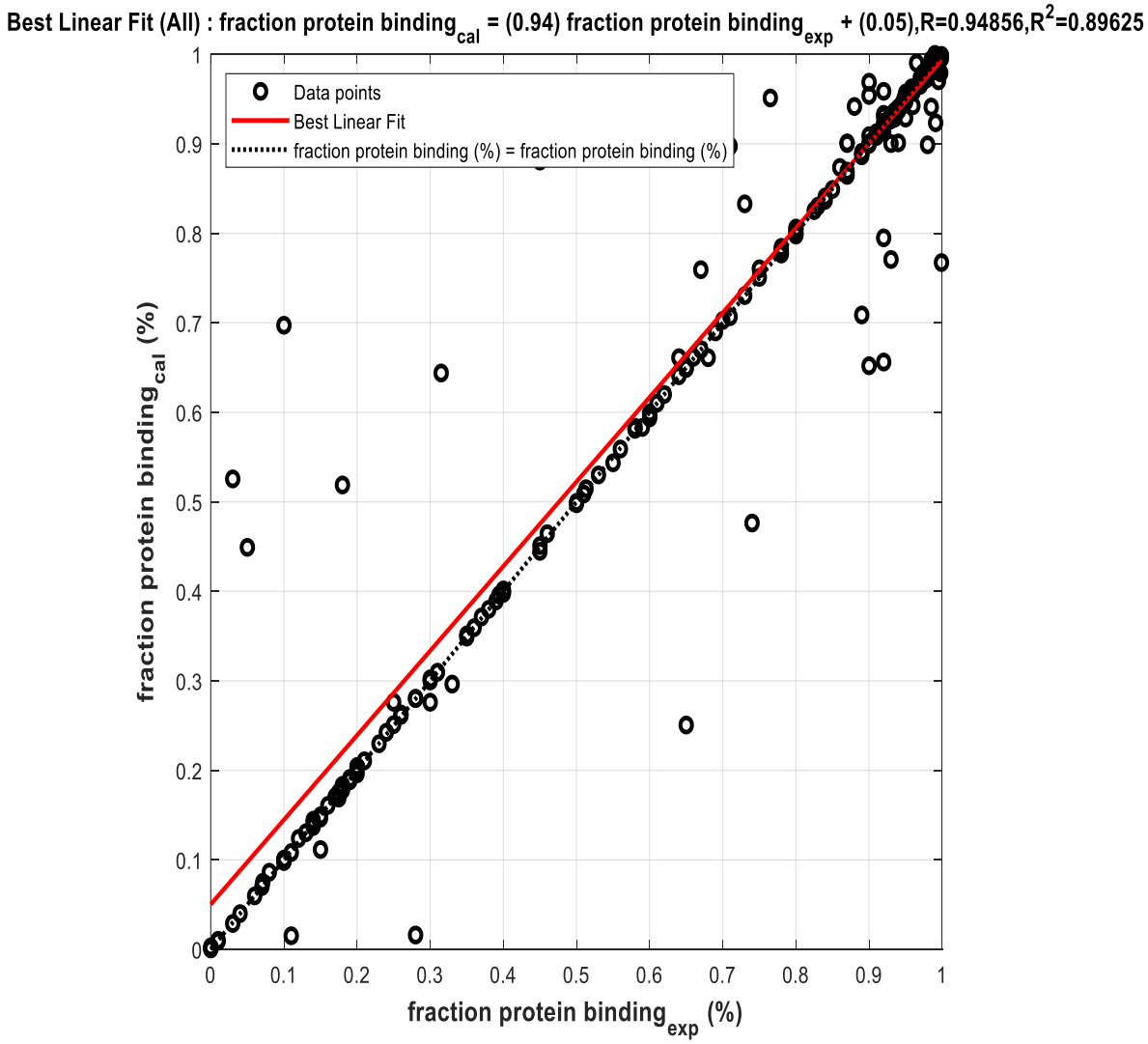


Figure II.12: Linear regression plot of ANN predicted data versus experimental data for the global phase database.

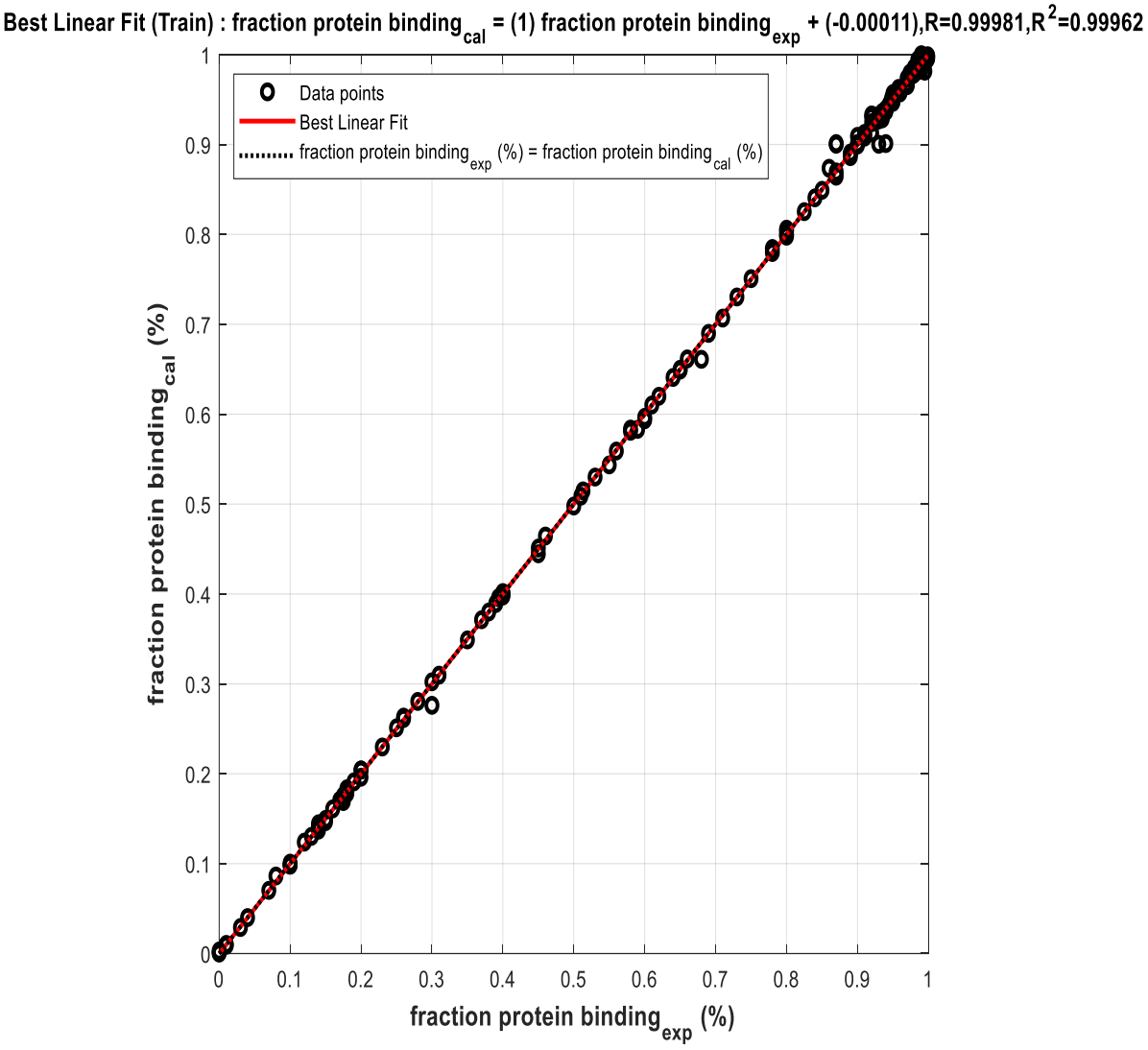


Figure II.13: Linear regression plot of ANN predicted data versus experimental data for the learning phase database.

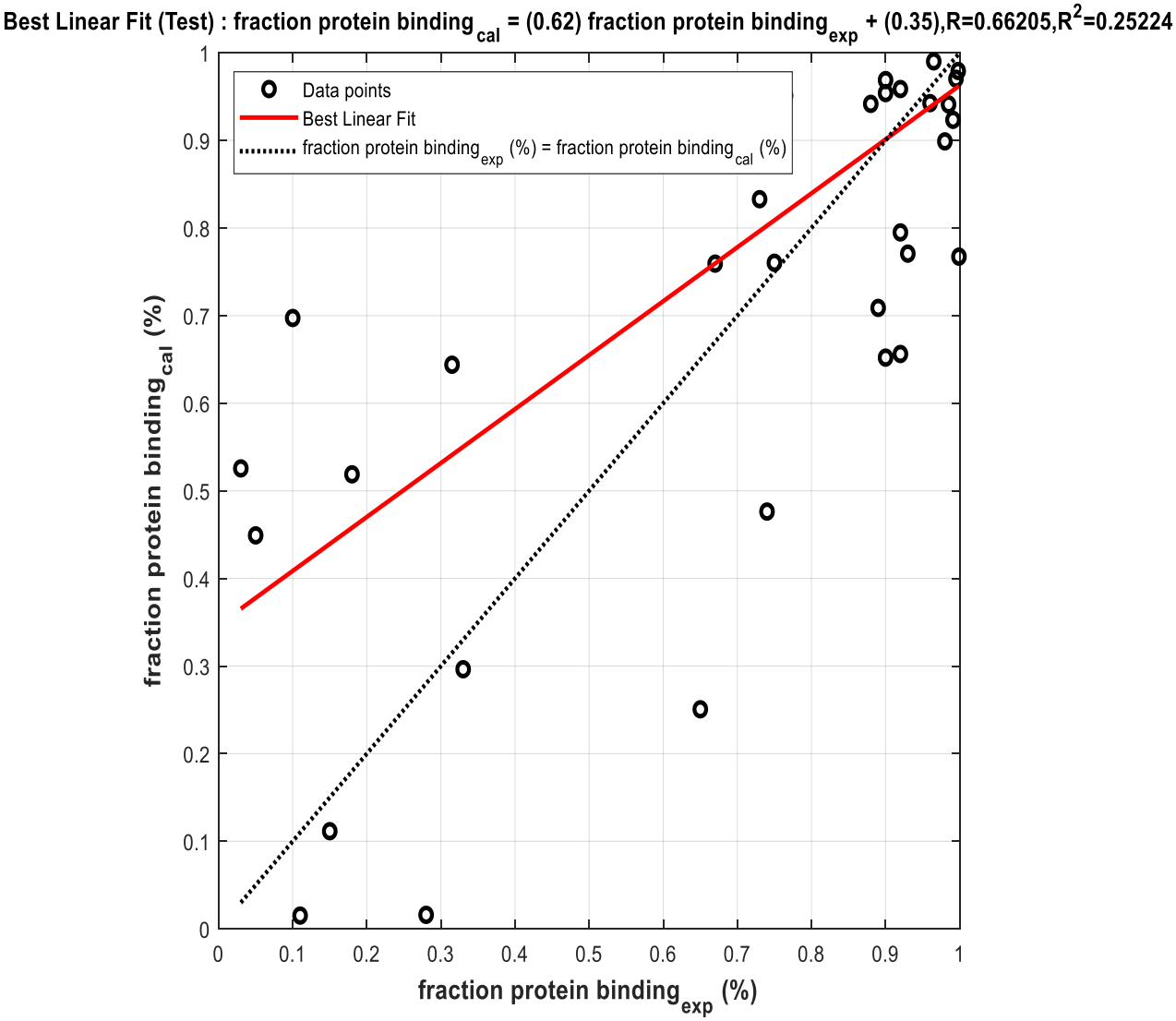


Figure II.14: Linear regression plot of ANN predicted data versus experimental data for the Test phase database.

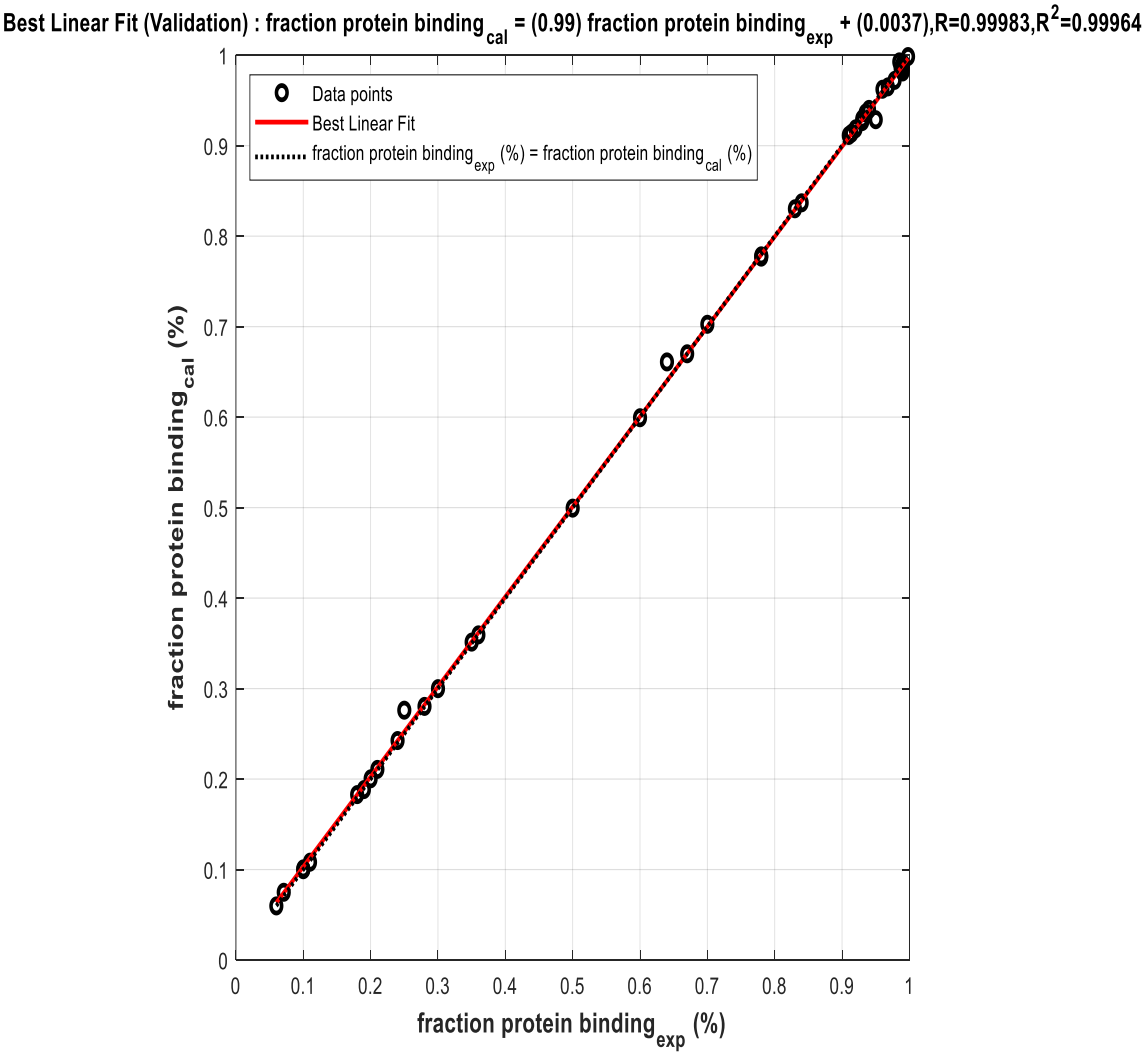


Figure II.15: Linear regression plot of ANN predicted data versus experimental data for the validation phase database.

To analyse the performance of the ANN more precisely, all errors: MSE; RMSE; NRMSE; MAE and ARD and R^2 are calculated for the predicted values of protein Binding Fractions (FB). The calculation results are shown in Table II.7. The ANN shows the calculation predicted values of protein Binding Fractions (FB).

Table II.7: The coefficient and statistical errors of the predicted results (ANN) compared to the experimental results.

Error type	Phase (Learn)	Phase (Test)	Phase (Validation)	Phase (Global)
MSE	$9.5489.10^{-7}$	0.2315	$2.8717.10^{-6}$	0.0387
RMSE	0.0065	0.2627	0.0066	0.1076
NRMSE	0.0014	0.6738	0.0026	0.2871
MAE	0.0033	0.1887	0.0033	0.0344
ARD	$4.2228.10^{-5}$	0.0690	$4.3565.10^{-5}$	0.0116
α	1	0.62	0.99	0.94
β	-0.00011	0.35	0.0037	0.05
R	0.9998	0.6620	0.9998	0.9485
R^2	0.9996	0.2522	0.9996	0.8962

II.3.1.3. Prediction of protein binding fractions using Support the hybrid Dragonfly algorithm coupled with Support Vector Machine (DA-SVM):

In this part, the DA-SVM is used for nonlinear modelling for the prediction of protein Binding Fractions (FB) based on the same 25 descriptors used to develop the ANN model. Different permutations of meta-parameters, namely the regularization parameter C and the epsilon insensitivity coefficient ε, as well as the kernel parameter σ, are adopted on a trial-and-error basis. Combinations of parameters with the radial basis function (RBF Gaussian) kernel offer better performance; hence they have been recorded and defined as optimal kernel parameters.

The optimal results of the method used in this study are summarized in Table II.8.

Table II.8: Structure of the DA-SVM model optimized by the Dragonfly algorithm.

Model	C	ε	Σ	kernel function	Quantity of support vectors	Time (Hours)	RMSE
DA-SVM	64.8	0.458	0.087	RBF Gaussian	267	0.3	0.7401

To assess the predictive ability of the optimized DA-SVM model, the MATLAB function "postreg" is used to plot the entire learning, testing, and overall predicted data against the experimental data. The figures from II.16 to II.18 clearly show that the majority of points are located on the first bisector, with an s coefficient of determination (R^2) value: [α (slope), β (intercept), R^2 (coefficient of determination)] = [0.81, 0.16, and 0.71] $_{Test}$.

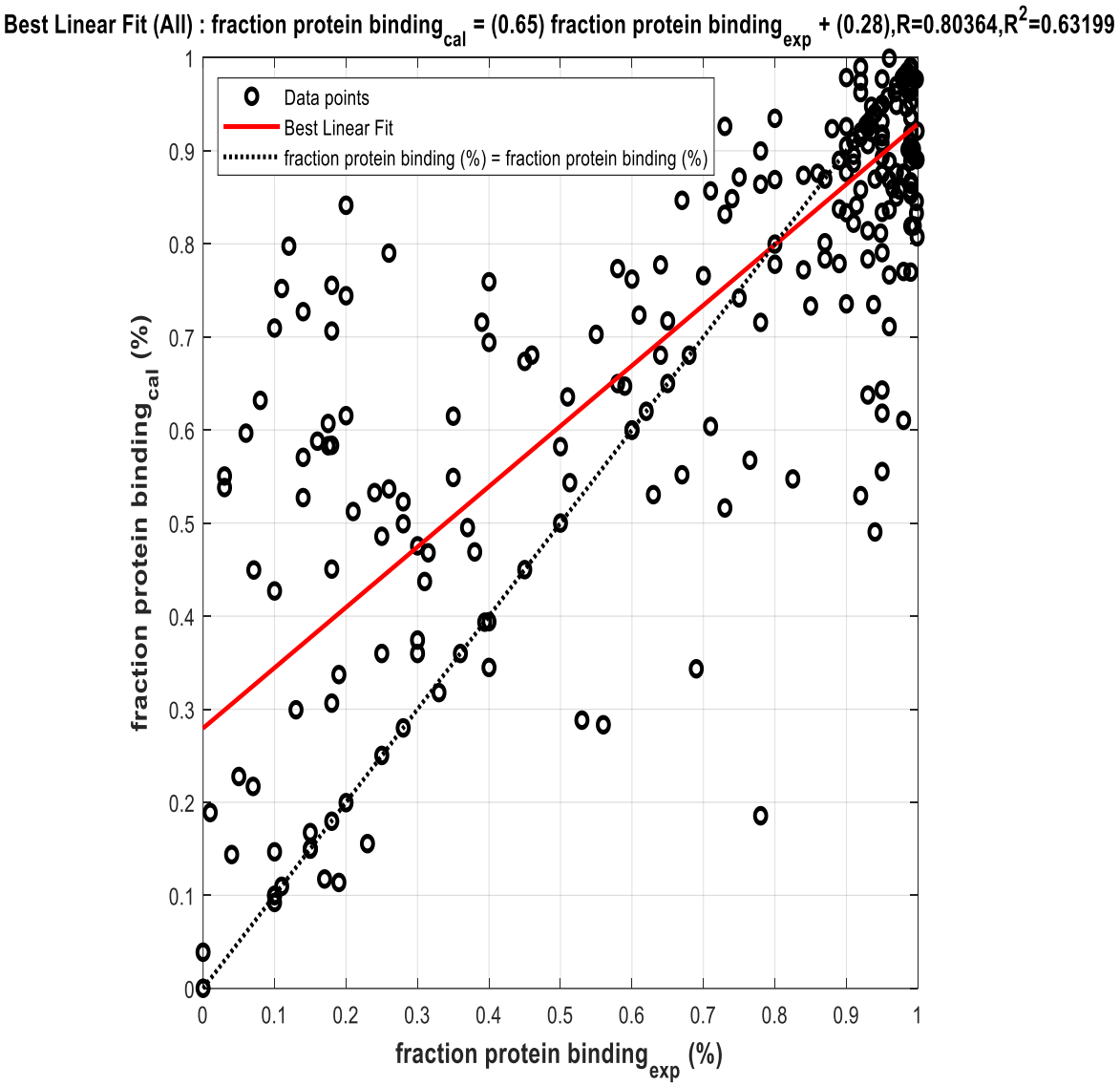


Figure II.16: Linear regression plot of DA-SVM predicted data versus experimental data for the global phase database.

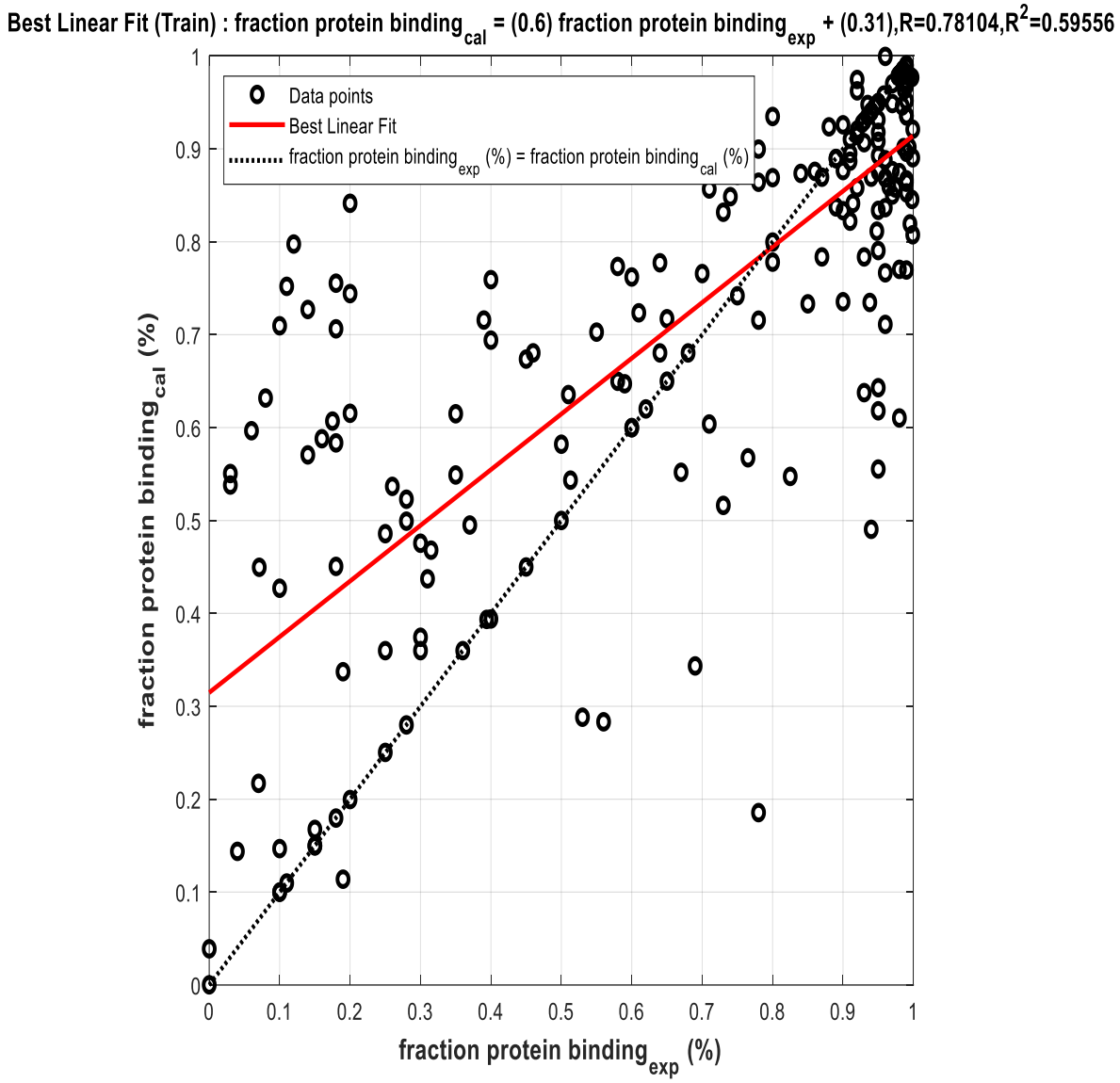


Figure II.17: Linear regression plot of DA-SVM predicted data versus experimental data for the learning phase database.

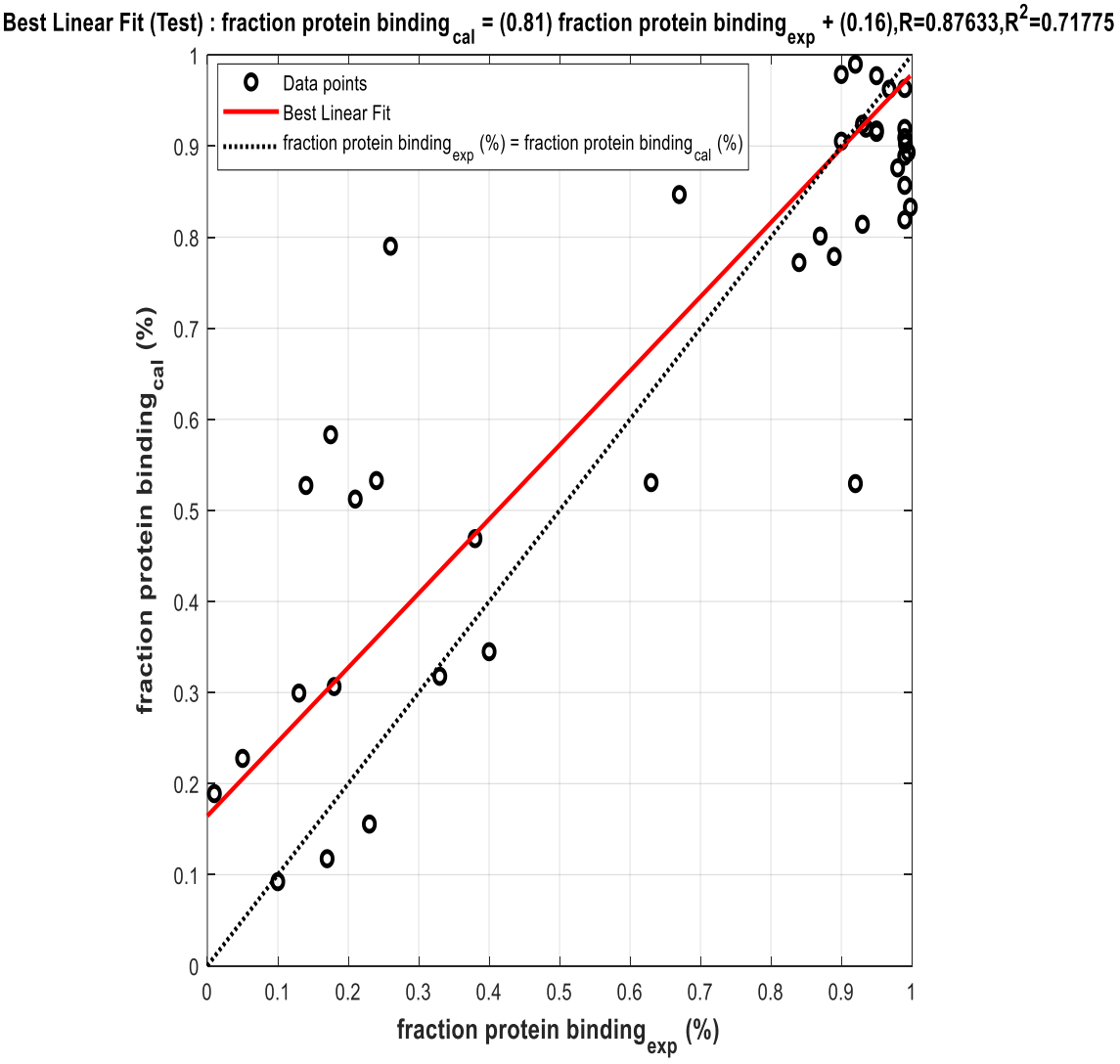


Figure II.18: Linear regression plot of DA-SVM predicted data versus experimental data for the Test phase database.

To analyse more precisely the performance of the DA-SVM model, a set of error formulas such as: MSE; RMSE; NRMSE; MAE and ARD are calculated for the predicted values of protein Binding Fractions (FB). The calculation results are shown in Table II.9.

Table II.9: The coefficient and statistical errors of the predicted results (DA-SVM) compared to the experimental results.

Error type	Phase (Learn)	Phase (Test)	Phase (All)
MSE	0.3229	1.4721	0.4061
RMSE	0.2081	0.7401	0.2027
NRMSE	0.8251	0.0614	0.9300
MAE	0.1335	0.2522	0.1342
ARD	0.0433	0.5477	0.0411
α	0.6	0.81	0.65
β	0.31	0.16	0.28
R	0.7810	0.8763	0.8036
$\mathbf{R^2}$	0.5955	0.7177	0.6319

II.3.2. Modelling of protein binding fraction using descriptors features:

II.3.2.1. Prediction of protein binding fractions using Convolutional neural network-Multiple layer perceptron (CNN-MLP):

Python programming language is used within SPIDER® software is used in this work to predict the protein binding fractions in function of descriptors features.

II.3.2.1.1 Choice of CNN parameters:

The feature extraction process begins with a 2D input of size (15, 15,1), which is reduced to (13,13,8) in the first convolutional layer. Subsequently, in the second convolutional layer, the size is further reduced to (11, 11,1). Finally, after passing through the Maxpooling layer, the matrix size becomes (5, 5,1), which is then flattened into a 1D vector containing 25 features for each active substance.

After flattening the output from the convolutional and pooling layers into a 1D vector, the subsequent step in the Convolutional Neural Network (CNN) architecture involves passing this vector through fully connected layers, also known as dense layers. The following

results in figure II.19 shows that the best architecture of the dense layers obtained is composed of 25 extracted features of a one active substance from our database used, are initially connected to a dense layer comprising 1024 neurons. Subsequently, the output of this layer is further processed through subsequent dense layers, progressively reducing the number of neurons to 512, then 256, and finally 128 in the subsequent layers. Following this sequence, the output of the fourth dense layer with 128 neurons is then linked to a last hidden layer containing 64 neurons that finally leads to the output dataset and desired model performance.

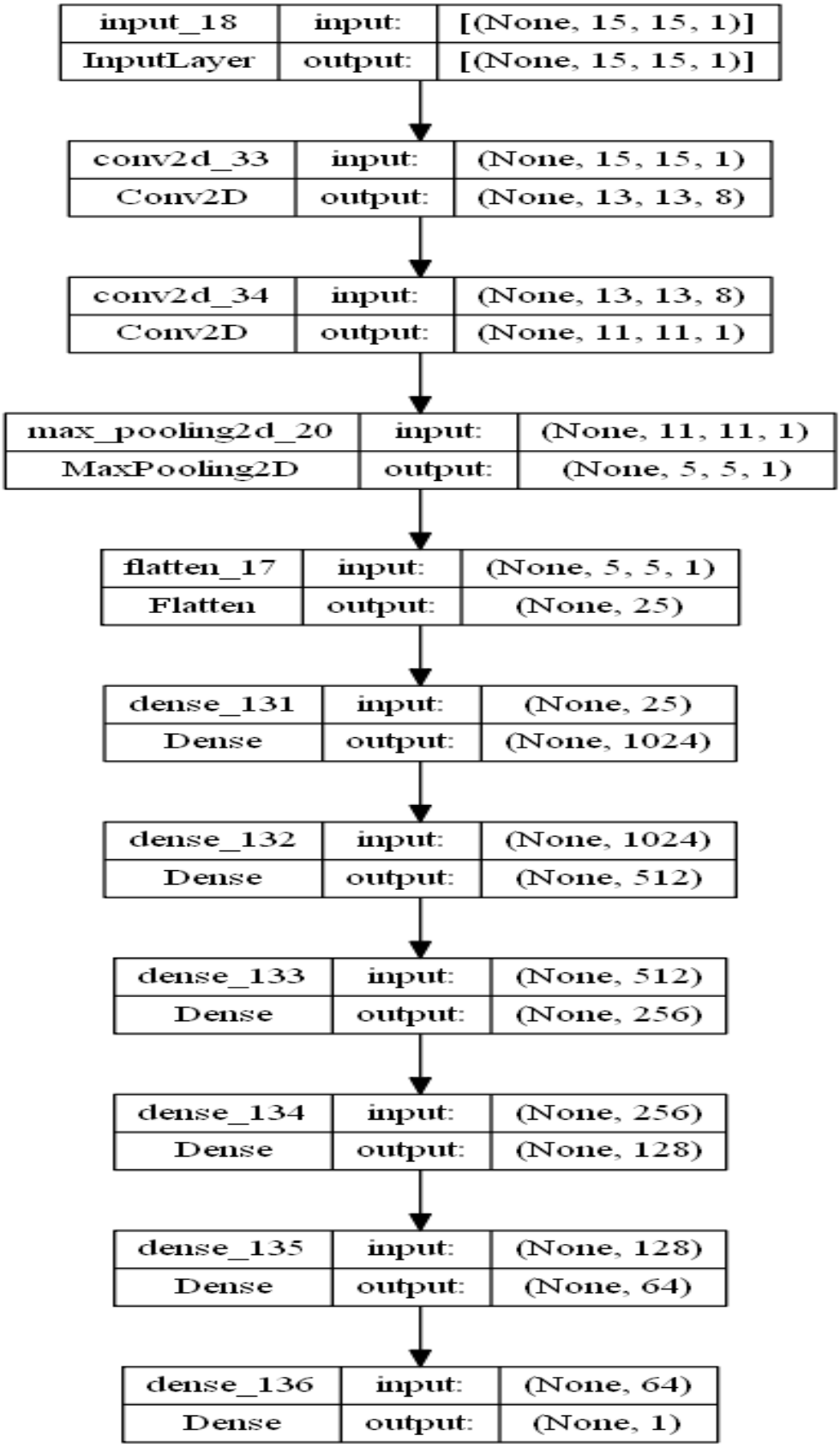


Figure II.19: Results of the CNN-MLP technique.

II.3.2.1.2. Results interpretation:

The performance results of the Convolutional Neural Network Multi-Layer Perceptron model (CNN-MLP) established for the learning, testing and Global phases (Learning + Testing) for the prediction of protein binding fractions rates are presented in Figures II.20 to Figure II.22 Presents the comparison between the experimental values and the values calculated by the CNN-MLP. Clearly shows that the majority of points are located on the first bisector, with a coefficient of determination (R^2) value approaching the ideal: [α (slope), β (y-intercept), R^2 (coefficient of determination)] = [1, -0.005, 0.99] $_{Test.}$ This confirms the robustness of the CNN-MLP model.

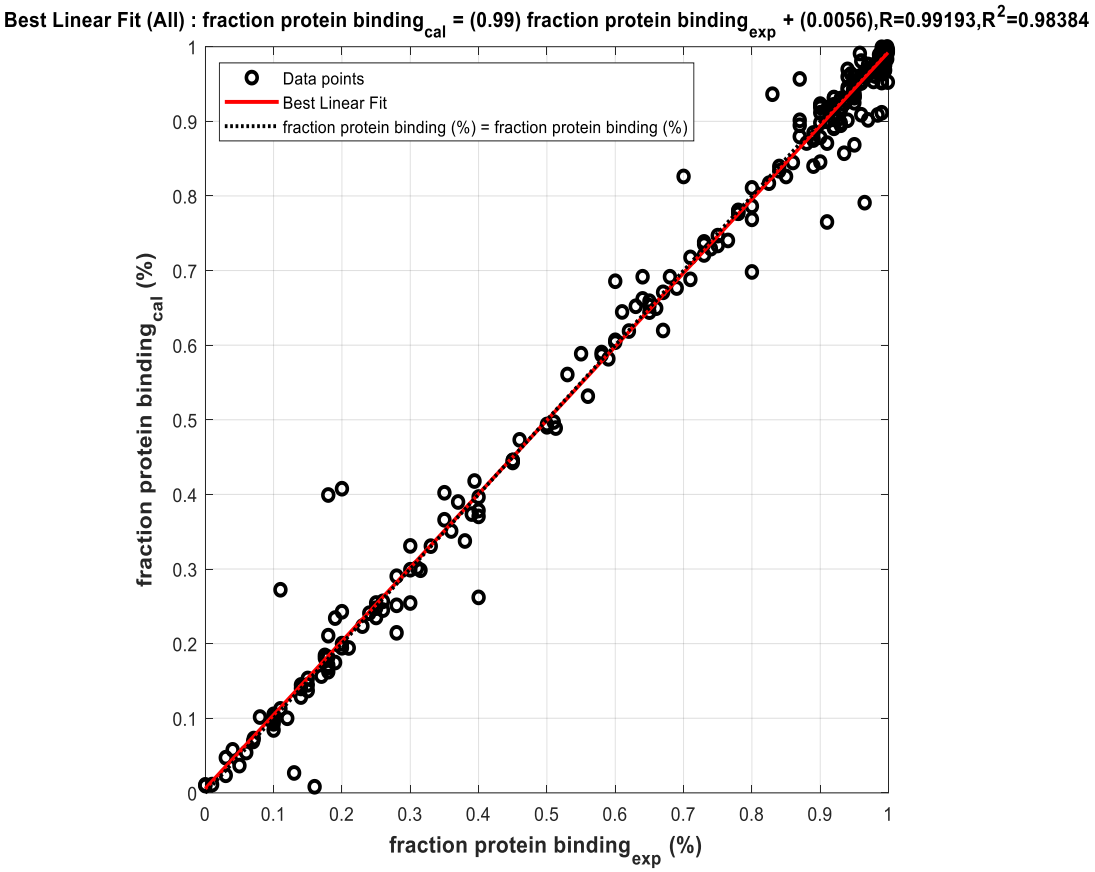


Figure II.20: Linear regression plot of CNN-MLP predicted data versus experimental data for the global phase database.

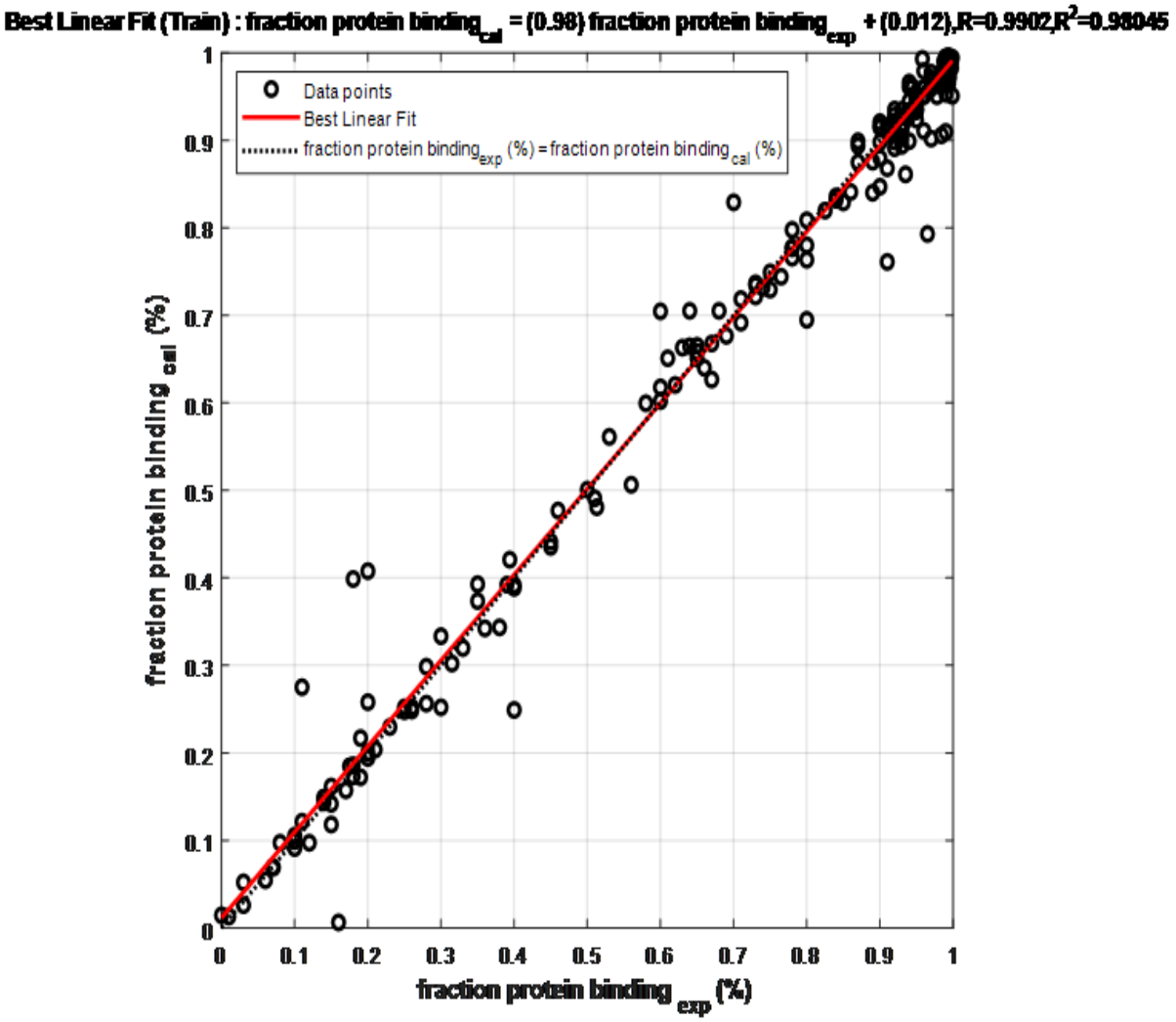


Figure II.21: Linear regression plot of CNN-MLP predicted data versus experimental data for the learning phase database.

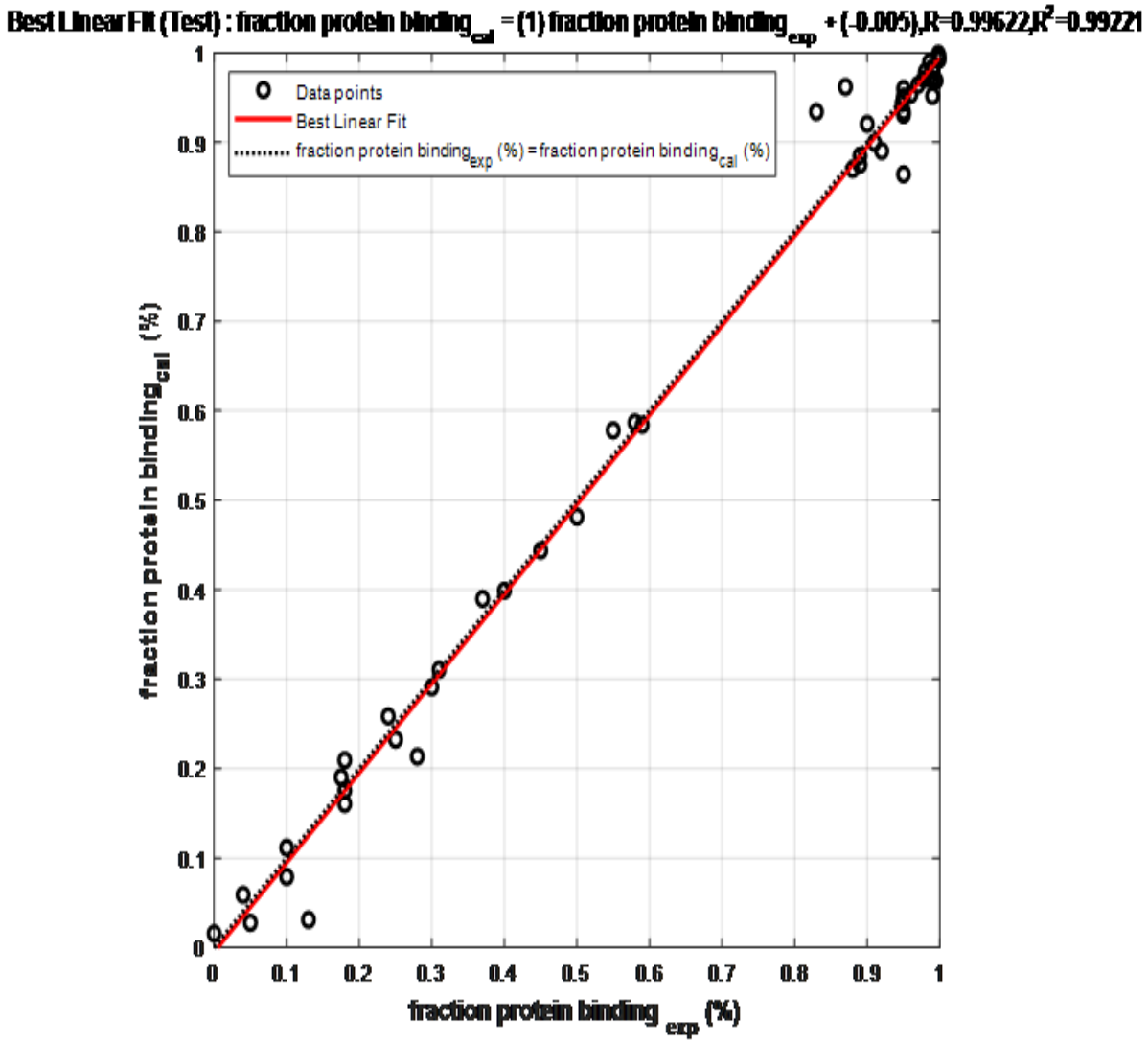


Figure II.22: Linear regression plot of CNN-MLP predicted data versus experimental data for the Test phase database.

To analyse more precisely the performance of the CNN-MLP model, a set of error formulas such as: MSE; RMSE; NRMSE; MAE and ARD are calculated for the predicted values of protein binding fractions (FB). The calculation results are shown in Table II.10.

Table II.10: The coefficient and statistical errors of the predicted results compared to the experimental results by CNN-MLP.

Type Of error	Phase (Learn)	Phase (Test)	Phase (All)
MSE	$8.6782.10^{-04}$	0.0013	0.0022
RMSE	0.0459	0.0314	0.0425
NRMSE	0.0427	0.0536	0.0691
MAE	0.0252	0.0199	0.0230
ARD	0.0021	$9.8759.10^{-04}$	0.0018
α	0.98	1	0.99
β	0.012	- 0.005	0.0056
R	0.9902	0.99622	0.99193
R^2	0.98045	0.99221	0.98384

II.3.2.2. Prediction of protein binding fractions using Convolutional Neural Networks coupled with Dragonfly Algorithm-Support Vector Machine (CNN-DA-SVM):

In this part, the DA-SVM is used for nonlinear modelling for the prediction of protein Binding Fractions (FB) based on the same 25 descriptors features generated by the CNN technique. The figures from II.23 to II.25 represent the regression curves of the data calculated by the DA-SVM model in relation to the experimental data, with a coefficient of determination (R^2) value indicating a moderate correlation: [α (slope), β (intercept), R^2 (coefficient of determination)] = $[1.3, -0.26, 0.63]_{Test}$.

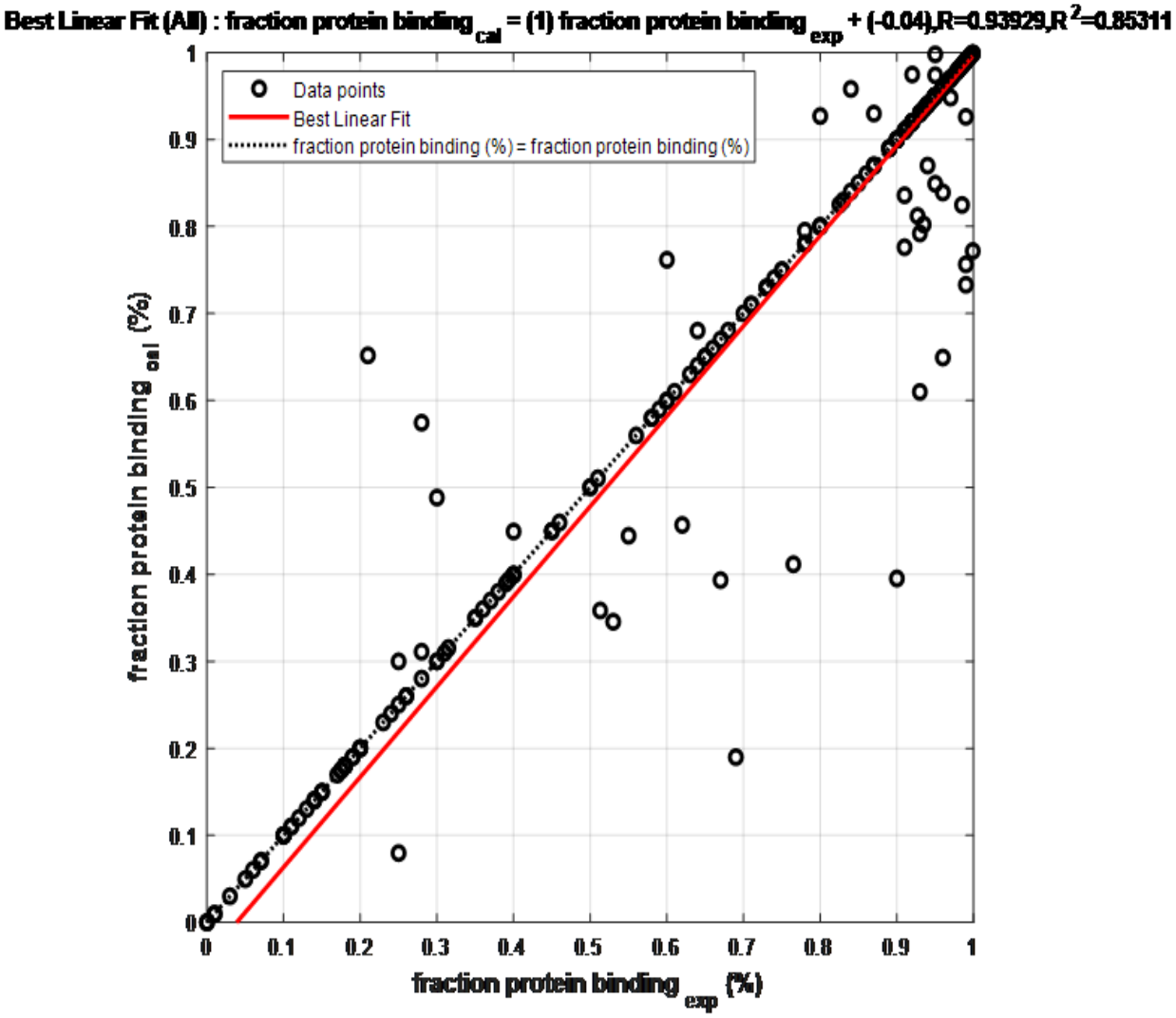


Figure II.23: Linear regression plot of CNN-DA-SVM predicted data versus experimental data for the global phase database.

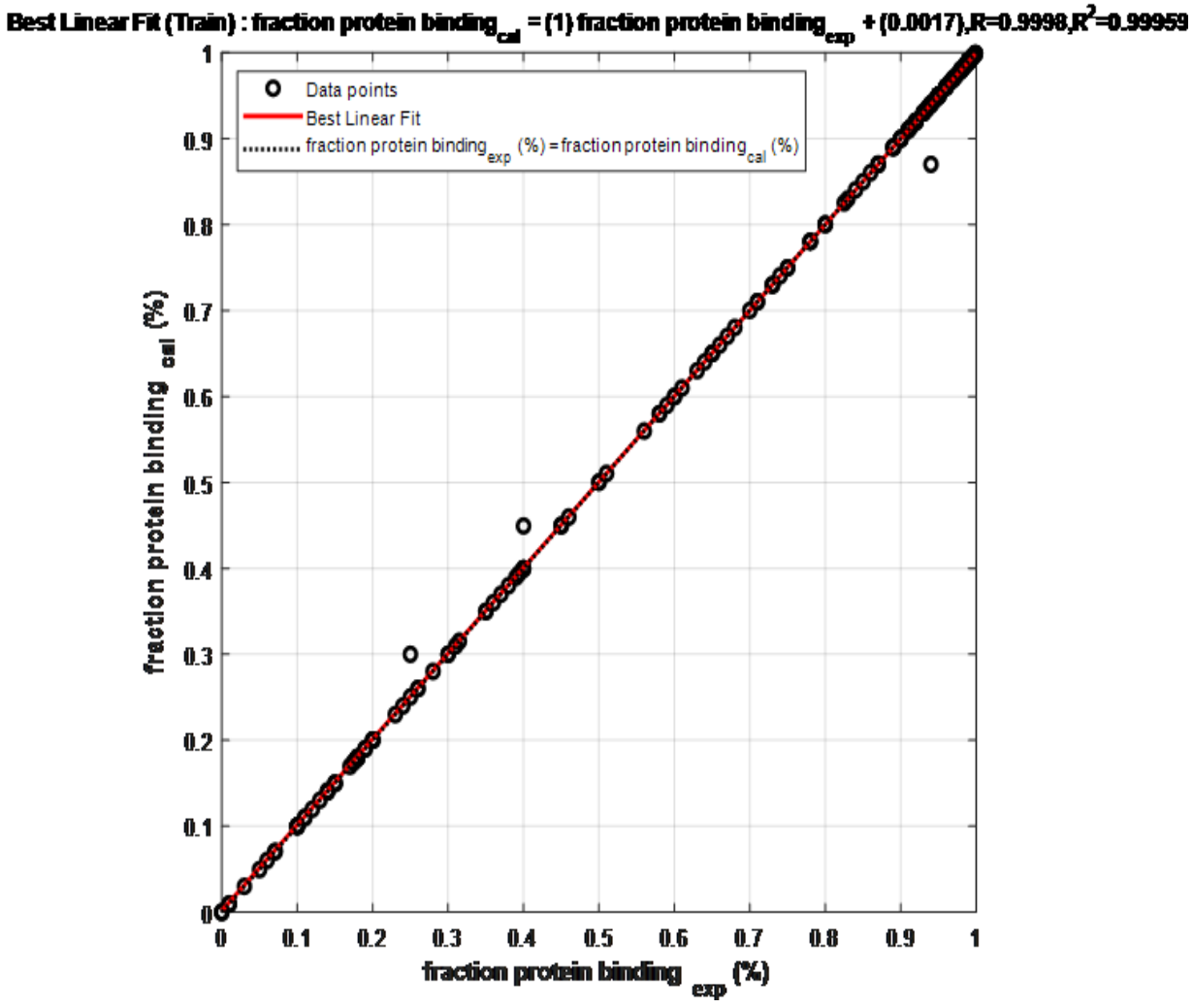


Figure II.24: Linear regression plot of CNN-DA-SVM predicted data versus experimental data for the learning phase.

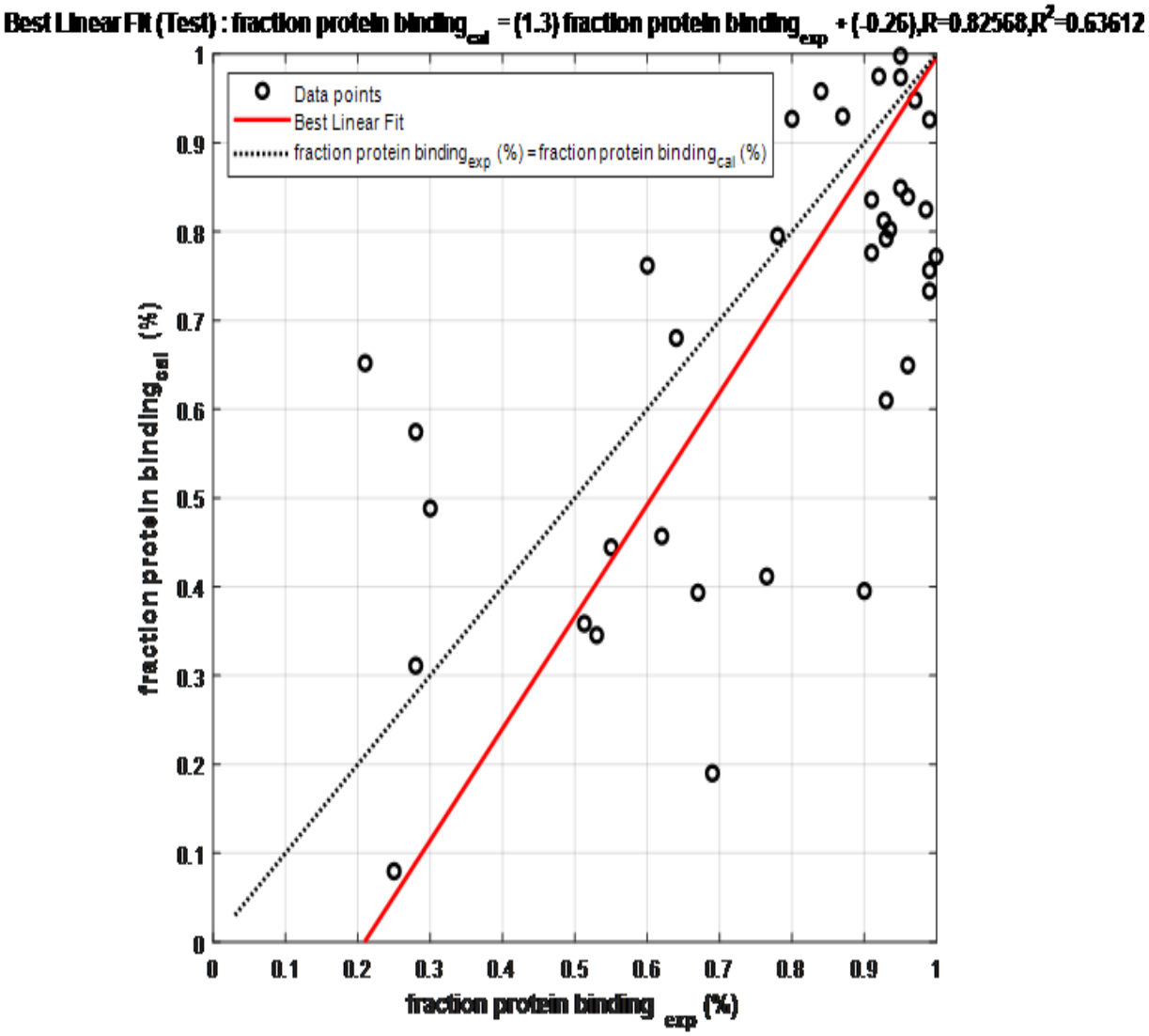


Figure II.25: Linear regression plot of CNN-DA-SVM predicted data versus experimental data for the Test phase.

Table II.11: The coefficient and statistical errors of the predicted results (CNN-DA-SVM) compared to the experimental results.

Error Type	Phase (Learn)	Phase (Test)	Phase (All)
MSE	$5.7468.10^{-06}$	0.2989	0.0581
RMSE	0.0068	0.2877	0.1281
NRMSE	0.0035	0.7575	0.3516
MAE	0.0010	0.2189	0.0441
ARD	$4.6716.10^{-05}$	0.0828	0.0164
α	1	1.3	1
β	0.0017	-0.26	-0.04
R	0.9998	0.8256	0.9392
R^2	0.9995	0.6361	0.8531

II.3.3. Comparative Study:

In this section, the two categories of models were compared: those utilizing descriptors generated by genetic algorithm (ANN and DA-SVM) and those incorporating descriptors features (CNN-MLP and CNN-DA-SVM) for predicting protein Binding Fractions (FB). The evaluation is based on error and statistical coefficients obtained previously in the testing phase.

II.3.3.1. A comparative study between models used descriptors (ANN and DA-SVM):

Based on the results in terms of error and statistical coefficient presented in Tables II.7 and II.9. The Dragonfly Algorithm-Support Vector Machine (DA-SVM) exhibits a robust performance in predicting protein binding fractions (FB). Its ability to capture complex relationships within the data leads to accurate the prediction of FB, as evidenced by low error and high statistical coefficients. In contrast, the Artificial Neuron Network (ANN) does not perform as effectively when compared to DA-SVM. While ANN performed poorly, delivering inaccurate results in predicting protein Binding Fractions (FB), it falls short in achieving the same level of accuracy and predictive power as DA-SVM, as it is shown in figure II.26.

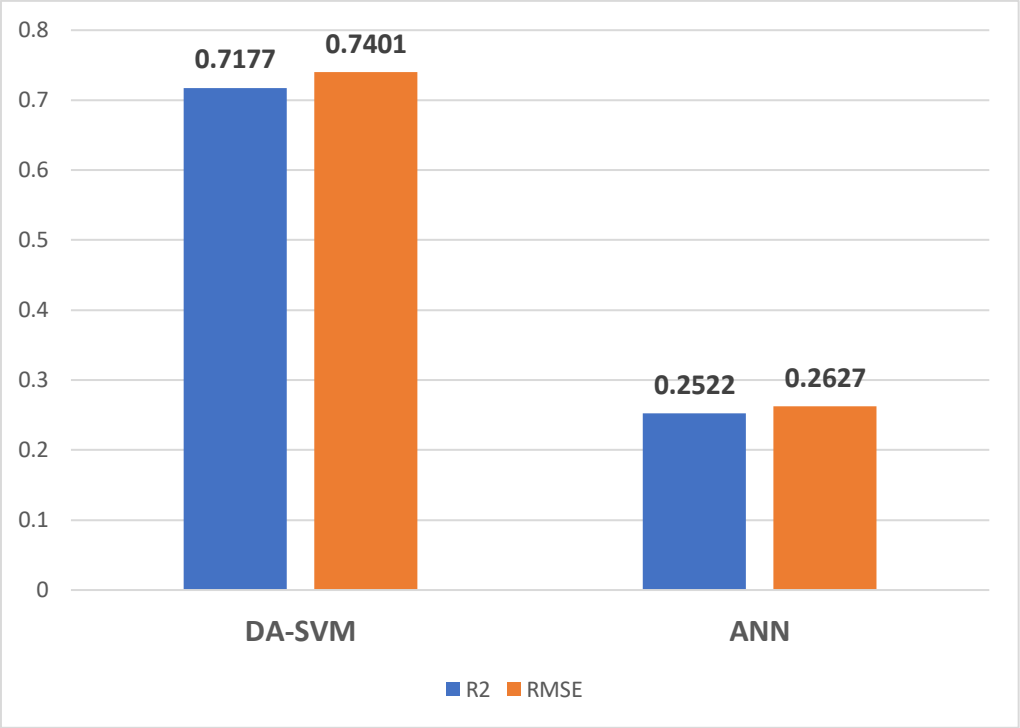


Figure II.26: The difference in terms of root mean square and coefficient of determination between the ANN and DA-SVM models.

II.3.3.2. A comparative study between models used descriptors features (CNN-MLP and CNN-DA-SVM):

Based on the results presented in Tables II.10 and II.11, the Convolutional Neural Network - Multi-Layer Perceptron (CNN-MLP) demonstrates robust performance in predicting protein Binding Fractions (FB). Its ability to extract descriptors features leads to an accurate prediction of FB, with a low error rates and high statistical coefficients. In contrast, the Convolutional Neural Network - Dragonfly Algorithm-Support Vector Machine (CNN-DA-SVM) does not perform as effectively when compared to CNN-MLP. While CNN-DA-SVM demonstrates competitive performance in predicting protein Binding Fractions (FB), it falls short in achieving the same level of accuracy and predictive power as CNN-MLP, as shown in Figure II.27.

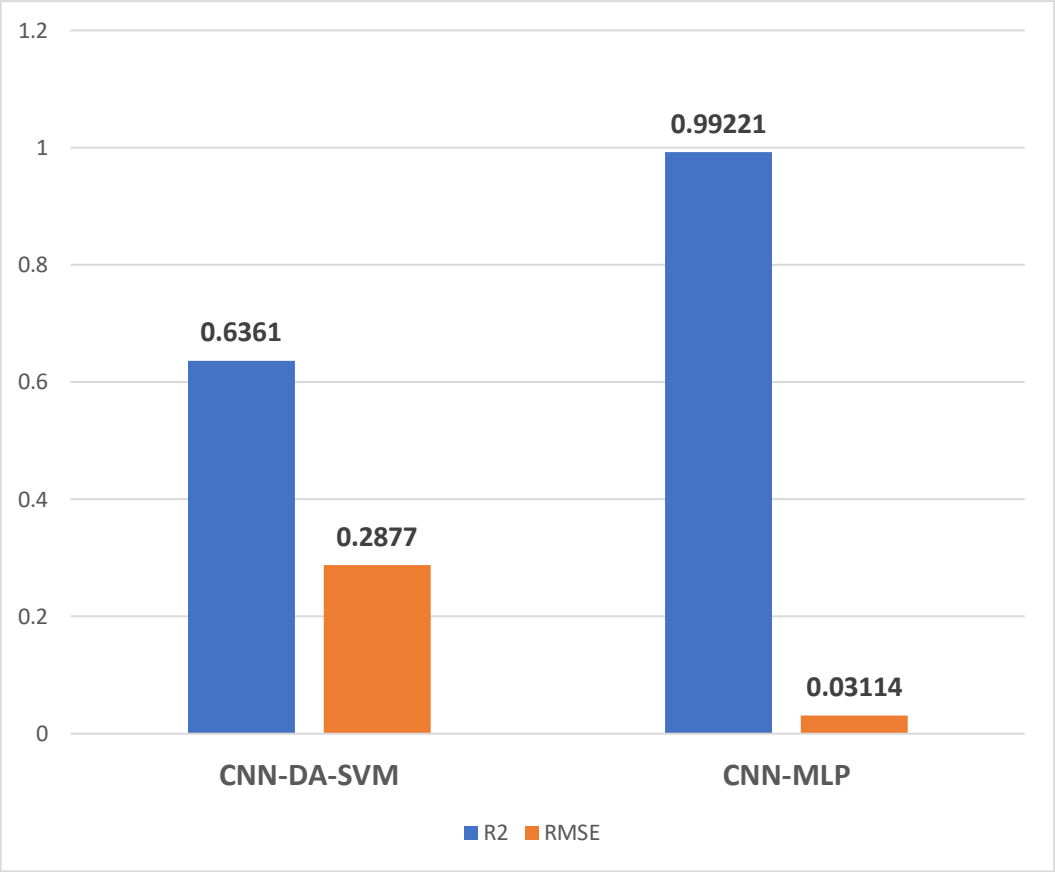


Figure II.27: The difference in terms of root mean square and coefficient of determination between the CNN-MLP and CNN-DA-SVM models.

II.3.3.3. Comparative study of the top models for predicting FB: First vs. Second approach (DA-SVM and CNN-MLP):

In the overall comparison, based on the previous comparisons, CNN-MLP emerges as the most effective model for predicting protein binding fractions, with the lowest RMSE value comparing to the rest of the models and the highest R^2 value. Its superior performance over all other models evaluated is attributed to its ability to leverage both descriptor features and complex neural network architectures, resulting in highly accurate predictions. Following CNN-MLP, DA-SVM stands out as the second most effective model, demonstrating strong predictive capabilities. (Figure II.28).

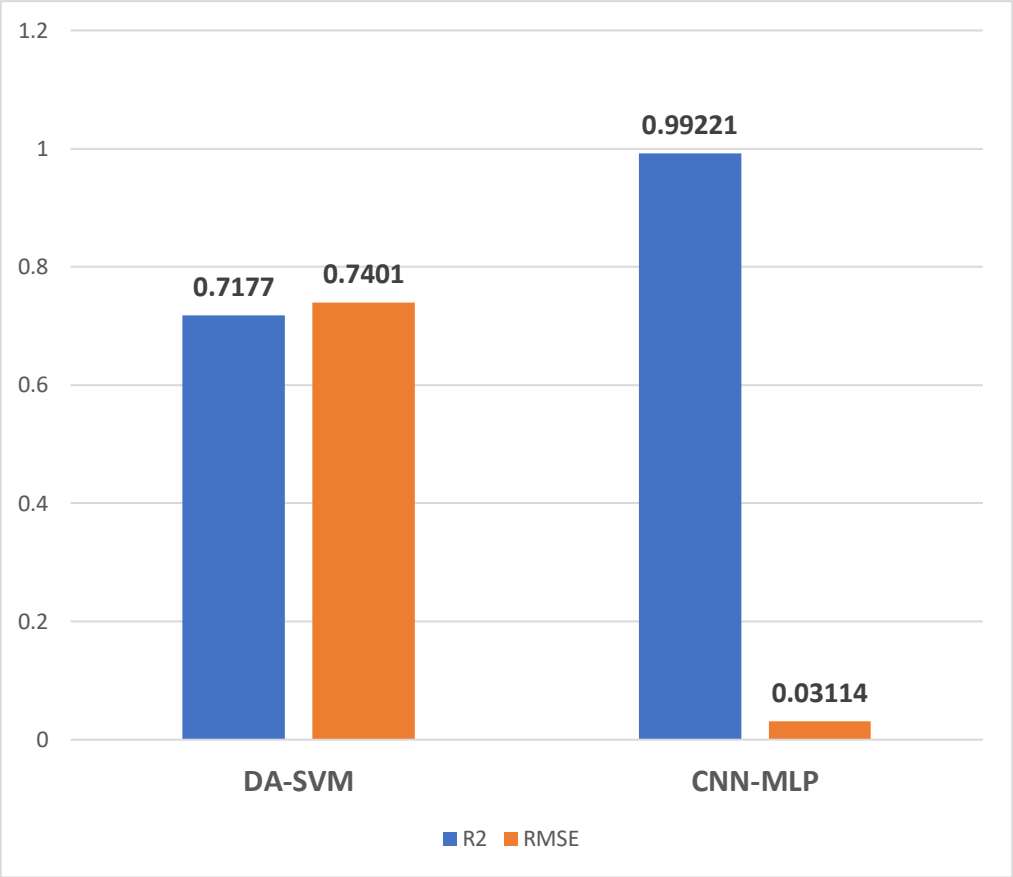


Figure II.28: The difference in terms of root mean square and coefficient of determination between the CNN-MLP and DA-SVM models.

The performance of the best model developed in this work (CNN-MLP) is compared to the performances of other models in the literature based on various types of learning algorithms, namely Artificial Neural Network, Boosted Regression Tree, and Feed-Forward Neural Network. A comparison was made between the statistical results of the CNN-MLP model and other predictive models for protein binding fractions. According to Table II.12, the value of the coefficient of determination R^2 is close to 1, and the MAE, RMSE, ARD, NRMSE, and MSE values are close to 0. This indicates that the proposed model (CNN-MLP) significantly outperforms the five reference models.

Table II.12: Comparison between the CNN-MLP model presented and existing models.

Models	**Method**	**Number of Inputs**	**Number of Outputs**	**Statistical parameters**						
				R	***R^2***	***MAE***	***RMSE***	**ARD**	***NRMSE***	***MSE***
Actual Work	**CNN-MLP**	**25**	**1**	**0.996**	**0.992**	**0.019**	**0.031**	**$9.8.10^{-4}$**	**0.053**	**0.001**
(Joseph R and al. 2005)	**ANN**	**20**	**1**	(-)	0.90	(-)	(-)	(-)	(-)	7.600
(Zvetanka and Irini. 2012)	**QSPPBR**	**178**	**1**	(-)	0.77	(-)	(-)	(-)	(-)	(-)
(Taravat and Zeshan. 2013)	**Boosted Regression Tree**	**450**	**1**	(-)	0.71	13.25	(-)	(-)	(-)	(-)
(Lixia Sun and al. 2018)	**MLR**	**26**	**1**	(-)	0.43	0.17	0.27	(-)	(-)	(-)
(Khaouane. 2023)	**feed-forward neural network**	**1666**	**1**	0.95	0.90	0.06	(-)	(-)	(-)	0.01

II.3.4. User Interface of the Best Model (CNN-MLP):

To simplify the complexity of the mathematical model developed by the best obtained model in the present work, a graphical user interface (also known as a user interface) is developed to enhance the usability of the developed CNN-MLP model. This user interface is a tool that allows for the direct and effortless use of the CNN-MLP model to calculate the protein Binding Fractions FB, this interface was created to calculate the protein Binding Fractions FB within the range of input variables. For this purpose, a MATLAB program was designed with a screenshot of the resulting interface (Figure II.29).

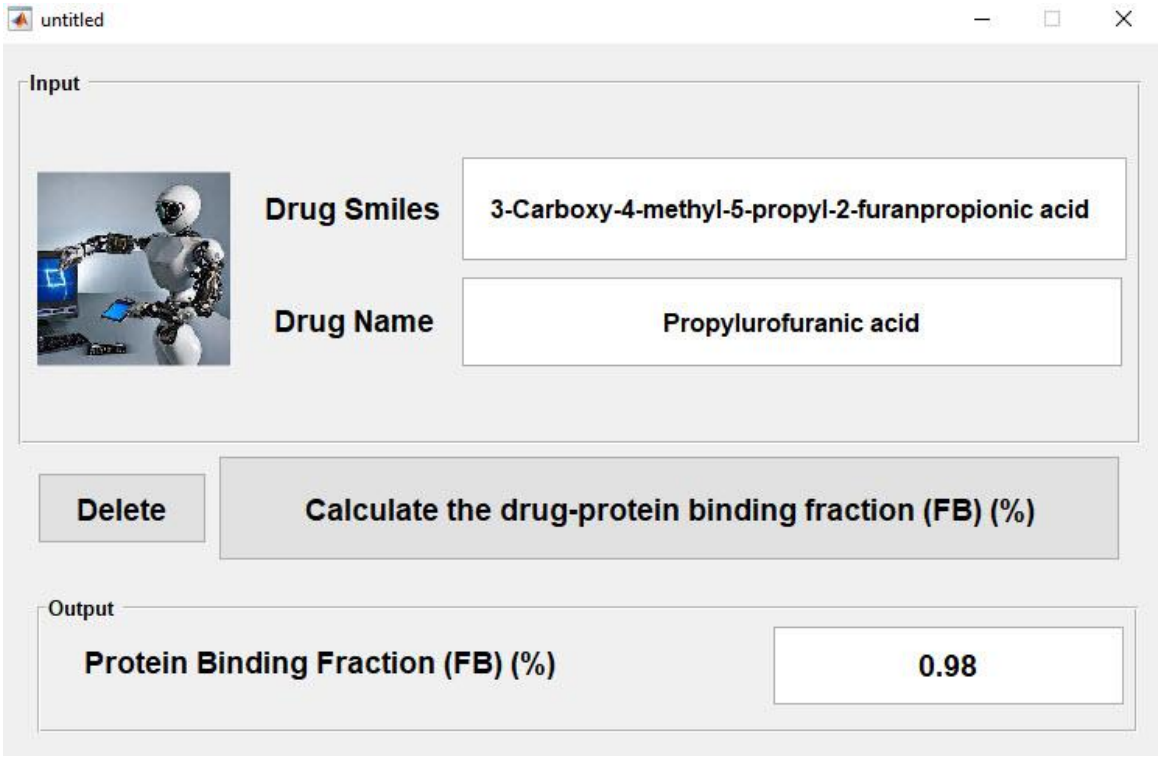


Figure II.29: The Graphical User Interface of the CNN-MLP Model

II.4. Conclusion:

In this part, extensive study into various models for predicting protein Binding Fractions (FB) revealed several key findings. Among the models evaluated, the Convolutional Neural Network-Multi Layer Perceptron (CNN-MLP) emerged as the top performer, showcasing superior predictive accuracy. Leveraging both descriptor features and intricate neural network architectures, CNN-MLP demonstrated highly accurate predictions. Following closely, the Dragonfly Support Vector Machine (DA-SVM) stood out as the second most effective model, despite relying on descriptors generated by the Genetic Algorithm (GA). The Convolutional Neural Network coupled with Dragonfly Support Vector Machine (CNN-DA-SVM) exhibited competitive performance, yet fell short of the accuracy achieved by CNN-MLP and ANN. Conversely, the Artificial Neural Network (ANN) performed poorly, delivering inaccurate results and ranked as the least effective model among those examined. Its limitations in capturing the complexities of protein binding fraction prediction became apparent when compared to the more sophisticated CNN-MLP and DA-SVM models. Overall, these findings underscore the importance of employing deep learning techniques, particularly convolutional neural network-based approach CNN-MLP, for accurate prediction of protein Binding Fractions (FB), paving the way for enhanced understanding and analysis in this field. Additionally, it significantly surpasses the five reference models in terms of error and the predictive capacity for plasmatic protein Binding Fractions (FB). The results obtained confirm the robustness of the established CNN-MLP model.

Bibliographical References

Alzubaidi, Laith, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, and Laith Farhan. (2021). Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. Journal of Big Data (vol. 8, issue 53, pp.13–44).

Brunton, Laurence, L. Joh S. (2011). The Pharmacological Basis of Therapeutics. (12th ed.). McGraw-Hill Professional. United Kingdom (pp.1290–1321).

Gu, Jiuxiang, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, and al. (2018). Recent Advances in Convolutional Neural Networks. Pattern Recognition (vol. 77, pp. 354–370).

Hubel, D H, and Ad T N Wiesel. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Great Britain (vol. 160, p.107).

Kanungo, D. P., M. K. Arora, S. Sarkar, and R. P. Gupta. (2006). A Comparative Study of Conventional, ANN Black Box, Fuzzy and Combined Neural and Fuzzy Weighting Procedures for Landslide Susceptibility Zonation in Darjeeling Himalayas. Engineering Geology (vol. 85, pp. 347–359).

Keskes, Sonia, Salah Hanini, Mohamed Hentabli, and Maamar Laidi. (2020). Artificial Intelligence and Mathematical Modelling of the Drying Kinetics of Pharmaceutical Powders. Kemija Ind (vol. 69, issue 3, pp. 347–359).

Laurence, Brunton, Bruce Chabner, Bjorn Knollman. (2005). Goodman and Gilman's The Pharmacological Basis of Therapeutics (11th ed.). McGraw-Hill Professional. United States (p.1827).

Li, Hsiao Chi, Zong Yue Deng, and Hsin Han Chiang. (2020). Lightweight and Resource-Constrained Learning Network for Face Recognition with Performance Optimization. Sensors (vol. 20, issue 21, pp. 4–10).

Li, Xiao Li, Li Hong Li, Bao Lin Zhang, and Qian Jin Guo. (2013). Hybrid Self-Adaptive Learning Based Particle Swarm Optimization and Support Vector Regression Model for Grade Estimation. Neurocomputing (vol. 118, pp.179–90).

Min, J, and Y Lee. (2005). Bankruptcy Prediction Using Support Vector Machine with Optimal Choice of Kernel Function Parameters. Expert Systems with Applications (vol. 28, issue 4, pp.603–14).

Reynel-Avila, Hilda E., Didilia I. Mendoza-Castillo, Adrian Bonilla-Petriciolet, and Joaquin Silvestre-Albero. (2015). Assessment of Naproxen Adsorption on Bone Char in Aqueous Solutions Using Batch and Fixed-Bed Processes. Journal of Molecular Liquids (vol. 209, pp.187–190).

Scholkopf, Bernhard, and Alexander J. Smola. (2001). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (2nd ed.). The MIT Press, Cambridge. London, England (p.24).

Shen, Chaopeng. (2018). A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. Water Resources Research. Blackwell Publishing (vol. 10, pp.8559–8587).

GENERAL CONCLUSION

General Conclusion and Prescriptives

This study aims to establish a suitable model using artificial intelligence techniques to predict protein Binding Fractions (FB). The primary objective of this thesis is to conduct a comparative analysis of these AI techniques. Specifically, the study aims to evaluate the performance of various AI models, including Artificial Neural Network (ANN), Dragonfly-Support Vector Machine (DA-SVM), Convolutional Neural Network-Multi Layer Perceptron (CNN-MLP), and Convolutional Neural Network combined with Dragonfly-Support Vector Machine (CNN-DA-SVM). By comparing these models using different methods, such as descriptors treated by Genetic Algorithm (GA) and descriptors features, the goal is to identify the most effective AI technique for accurately predicting protein Binding Fractions (FB). The results of our analysis revealed that CNN-MLP yielded the best performance among all models, achieving an impressive value of coefficient of determination R^2= 0.99221 and a Root Mean Square Error RMSE= 0.0314. Following CNN-MLP, DA-SVM and CNN-DA-SVM also demonstrated competitive performance, while ANN delivered inaccurate results, respectively with values of R^2= [0.7177, 0.6361, 0.2522] and RMSE= [0.7401, 0.2877, 0.2627]. Additionally, from the comparison between the CNN-MLP model presented and existing predictive models for protein binding fractions in the literature based on various types of learning algorithms, CNN-MLP significantly surpasses the five reference models in terms of error and the predictive capacity for plasmatic protein Binding Fractions (FB). These results highlight the effectiveness of utilizing novel methodologies that integrate descriptor features, such as CNN-MLP, in accurately predicting (FB). This approach enables a more comprehensive analysis of complex relationships between molecular descriptors and the pharmacokinetic parameter, facilitating enhanced predictive modeling capabilities.

The impact of these findings on the pharmaceutical industry is significant. Accurate prediction of protein binding of active substances is crucial for drug development, as it influences factors such as bioavailability, efficacy and safety of pharmaceutical products. By employing AI techniques like CNN-MLP, pharmaceutical companies can streamline drug discovery processes, reduce both costs and time and accelerate the development of new therapeutics. Additionally, accurate predictions enhance decision-making in drug design, optimizing drug candidate selection and dosage regimen planning. This accelerated process enables pharmaceutical companies to bring new therapeutics to market more rapidly, meeting the urgent needs of patients and healthcare providers.

The results of this study pave the way for future advancements in predictive modeling for pharmaceutical applications such as:

- Broaden the base;
- Using this method for the prediction of other pharmacokinetic parameters;
- Exploring the use of other Artificial Intelligence (AI) techniques.

Table II.1: Therapeutic class of drugs

Drug	Therapeutic Class
3-Carboxy-4-methyl-5-propyl-2-furanpropionic acid	Anti-inflammatory
Acebutolol	Antihypertensive
Acenocoumarin	Anticoagulant
Acetaminophen	Analgesics and antipyretics
Acyclovir	Antiviral
Alendronate	Bisphosphonate
Alfentanil	Opioid
Allopurinol	Xanthine oxidase inhibitor
Alprazolam	Antianxiety
Amikacin	Antibiotic
Amiodarone	Antiarrhythmic
Amitriptyline	Antidepressant
Amlodipine	Antihypertensive
Amoxicillin	Antibiotic
Antipyrine	Analgesic / Antipyretic
Aprepitant	Antineoplastic
Atenolol	Antihypertensive
Atropine	Bronchodilator
Azapropazone	Anti-inflammatory
Azathioprine	Immunomodulator
Benoxaprofen	Anti-inflammatory
Benzylpenicillin / Penicillin G	Antibiotic
Bepridil	Antiarrhythmic

Betamethasone	Corticosteroid
Betaxolol	Antihypertensive
Bilirubin	Antihypertensive
Binedaline	Bronchodilator
Bisoprolol	Antihypertensive
Bosentan	Antihypertensive
Bromocriptine	Stimulant
Bufuralol	Antihypertensive
Bupivacaine	Anesthetic
Buprenorphine	Opioid
Buspirone	Antianxiety
Butorphanol	Opioid
Caffeine	Stimulant
Calcitriol	Vitamin D analog
Camptothecin	Anticancer
Candesartan	Antihypertensive
Carbamazepine	Anticonvulsant
Carbenoxolone	Anti-inflammatory
Carprofen	Anti-inflammatory
Carvedilol	Antihypertensive
Cefadroxil	Antibiotic
Cefamandole	Antibiotic
Cefazolin	Antibiotic
Cefepime	Antibiotic
Cefixime	Antibiotic
Cefoperazone	Antibiotic

Cefotaxime	Antibiotic
Cefoxitin	Antibiotic
Cefprozil	Antibiotic
Cefsulodin	Antibiotic
Ceftazidime	Antibiotic
Ceftizoxime	Antibiotic
Ceftriaxone	Antibiotic
Cefuroxime	Antibiotic
Celecoxib	Anti-inflammatory
Cephalexin	Antibiotic
Cephaloridine	Antibiotic
Cephalothin	Antibiotic
Cephradine	Antibiotic
Cetiedil	Antihypertensive
Cetirizine	Antihistamine
Chlorothiazide	Antihypertensive
Chlorpheniramine	Antihistamine
Chlorpromazine	Antipsychotic
Chlorpropamide	Antidiabetic
Chlorprothixene	Antipsychotic
Chlorthalidone	Antihypertensive
Cicletanine	Antihypertensive
Cimetidine	Antihistamine
Cimoxatone	Antidepressant
Cinacalcet	Calcimimetic

Cinoxacin	Antibiotic
Ciprofloxacin	Antibiotic
Clofibric acid	Lipid-lowering
Clometacin	Anti-inflammatory
Clozapine	Antipsychotic
Cyclophosphamide	Antineoplastic
Dapsone	Antibiotic
Desipramine	Antidepressant
Dexamethasone	Anti-inflammatory
Diazepam	Antianxiety
Diclofenac	Anti-inflammatory
Dicloxacillin	Antibiotic
Didanosine	Antiviral
Diflunisal	Anti-inflammatory
Digitoxigenin	Cardiac glycoside
Digitoxin	Cardiac glycoside
Diltiazem	Antihypertensive
Diphenhydramine	Antihistamine
Disopyramide	Antiarrhythmic
Docetaxel	Antineoplastic/Anticancer
Dofetilide	Antiarrhythmic
Domperidone	Gastrointestinal
Entacapone	Antipsychotic
Ethambutol	Antibiotic
Etodolac	Anti-inflammatory
Etoposide	Antineoplastic

Famotidine	Antihistamine
Fenbufen	Anti-inflammatory
Fenoprofen	Anti-inflammatory
Fentanyl	Opioid
Fentiazac	Anti-inflammatory
Finasteride	Steroid hormone
Flecainide	Antiarrhythmic
Fluconazole	Antifungal
Flumazenil	Antianxiety
Fluorouracil	Antineoplastic
Fluphenazine	Antipsychotic
Flurbiprofen	Anti-inflammatory
Foscarnet	Antiviral
Fulvestrant	Antineoplastic
Furosemide	Antihypertensive
Fusidic acid	Antibiotic
Gabapentin	Anticonvulsant/antiepileptic
Galantamine	cholinesterase inhibitor
Gemfibrozil	Lipid-lowering
Gentamicin	Antibiotic
Glimepiride	Antidiabetic
Glyburide	Antidiabetic
Granisetron	Antiemetic
Halofenate	Lipid-lowering
Hydrochlorothiazide	Antihypertensive

Hydromorphone	Opioid
Ibuprofen	Anti-inflammatory
Imatinib	Anticancer
Indomethacin	Anti-inflammatory
Indoprofen	Anti-inflammatory
Irbesartan	Antihypertensive
Isosorbide dinitrate	Vasodilator
Isosorbide-5-mononitrate	Vasodilator
Isradipine	Antihypertensive
Itanoxone	Anti-inflammatory
Ketoprofen	Anti-inflammatory
Lansoprazole	Gastrointestinal
Letrozole	Antineoplastic
Levetiracetam	Anticonvulsants/ antiepileptic
Levofloxacin	Antibiotic
Lidocaine	Anesthetic
Linezolid	Antibiotic
Lomefloxacin	Antibiotic
Lorazepam	Antianxiety
Losartan	Antihypertensive
Meloxicam	Anti-inflammatory
Melphalan	Antineoplastic
Meperidine	Opioid
Mepirzepine / mirtazapine	Antidepressant
Mepivacaine	Anesthetic
Mercaptopurine	Antineoplastic

Metformin	Antidiabetic
Methadone	Opioid
Methicillin	Antibiotic
Methotrexate	Antineoplastic
Methylprednisolone	Anti-inflammatory
Metoclopramide	Gastrointestinal
Metoprolol	Antihypertensive
Metronidazole	Antibiotic
Midazolam	Antianxiety/Antineoplastic
Montelukast	Anti-asthmatic
Morphine	Opioid
Moxifloxacin	Antibiotic
Moxisylyte	Antihypertensive
Nadolol	Antihypertensive
Nafcillin	Antibiotic
Nalidixic acid	Antibiotic
Naloxone	Opioid
Naproxen	Anti-inflammatory
Neostigmine	Cholinesterase inhibitor
Nicergoline	Cholinesterase inhibitor
Nifedipine	Antihypertensive
Nimesulide	Anti-inflammatory
Nitrendipine	Antihypertensive
Nitrofurantoin	Antibiotic

Norepinephrine	Vasopressor
Norfloxacin	Antibiotic
Nortriptyline	Antidepressant
Novobiocin	Antibiotic
Ofloxacin	Antibiotic
Omeprazole	Gastrointestinal
Ondansetron	Antiemetic
Oxacillin	Antibiotic
Oxazepam	Antianxiety
Oxycodone	Opioid
Oxyphenbutazone	Anti-inflammatory
Pancuronium	Neuromuscular blocking
pentobarbital	Anticonvulsant / Sedative
Phenacetin	Analgesic
Phencyclidine	Anesthetic
Phenobarbital	Anticonvulsant / Sedative
Phenylbutazone	Anti-inflammatory
Phenytoin	Anticonvulsants/ antiepileptic
Pindolol	Antihypertensive
Pipotiazine	Antipsychotic
Pirenzepine	Gastrointestinal
Piretanide	Antihypertensive
Pirprofen	Anti-inflammatory
Pravastatin	Lipid-lowering
Prazosin	Antihypertensive

Prednisone	Immunomodulator
Pregnenolone	Steroid hormone
Procainamide	Antiarrhythmic
Procaine	Anesthetic
Propafenone	Antiarrhythmic
Propanolol	Antihypertensive
Propofol	Sedative/Anesthetic
Propranolol	Antihypertensive
Pyrimethamine	Antiparasitic
Quetiapine	Antipsychotic
Quinidine	Antiarrhythmic
Quinine	Antiparasitic
Ranitidine	Gastrointestinal
Repaglinide	Antidiabetic
Riluzole	Anticonvulsant
Risedronate	Bisphosphonate
Risperidone	Antipsychotic
Rizatriptan	Triptan
Rocuronium	Anesthetic
Ropivacaine	Anesthetic
Salicylic acid	Dermatological
Scopolamine	Antiemetic
Sildenafil	Phosphodiesterase
Sotalol	Antiarrhythmic
Sufentanil	Opioid

Sulfaphenazole	Antibiotic
Sulfinpyrazone	Uricosuric
Sulfisoxazole	Antibiotic
Sulindac	Anti-inflammatory
Sumatriptan	Triptan
Suprofen	Anti-inflammatory
Tamsulosin	Antineoplastic
Tegaserod	Gastrointestinal
Temazepam	Antianxiety
Tenofovir	Antiviral
Tenoxicam	Anti-inflammatory
Terazosin	Antihypertensive
Terbutaline	Bronchodilator
Testosterone	Hormone
Tetracycline	Antibiotic
Theophylline	Bronchodilator
Thiopental	Anesthetic
Ticlopidine	Anticoagulant /antiplatelet
Tolamolol	Antihypertensive/antiarrhythmic
Tolazamide	Antidiabetic
Tolbutamide	Antidiabetic
Tramadol	Opioid
Trazodone	Antidepressant
Triazolam	Antianxiety
Trimethoprim	Antibiotic

Urapidil	Antihypertensive
Valproic acid	Anticonvulsant / antiepileptic
Valsartan	Antihypertensive
Vecuronium	Neuromuscular blocking
Verapamil	Antihypertensive/antiarrhythmic
Vinorelbine	Antineoplastic
Vinpocetine	Dietary supplement
Warfarin	Anticoagulant
Zaleplon	Sedative/hypnotic
Zidovudine	Antiviral
Ziprasidone	Antipsychotic
Zolpidem	Sedative/hypnotic
Zomepirac	Anti-inflammatory

Table II.2: The set of errors for each number of neurons in the case of two hidden layers.

I	**j**	**Erreur**	2	9	8.4376	3	18	5.6246	5	7	6.1514
1	1	10.8550	2	10	7.5394	3	19	5.7188	5	8	5.7819
1	2	10.9066	2	11	6.6547	3	20	8.0811	5	9	5.9295
1	3	10.90820856	2	12	7.1173	4	1	5.0750	5	10	4.7595
1	4	10.9195	2	13	7.1387	4	2	5.5920	5	11	4.3897
1	5	10.9102	2	14	6.7291	4	3	6.0785	5	12	29.8019
1	6	10.9235	2	15	6.9517	4	4	5.1404	5	13	4.1600
1	7	11.71067	2	16	6.7196	4	5	6.5377	5	14	4.4612
1	8	10.814214	2	17	6.6712	4	6	6.5342	5	15	4.8563
1	9	10.887993	2	18	8.7303	4	7	7.7963	5	16	6.2392
1	10	21.35196	2	19	9.4814	4	8	7.7952	5	17	4.7996
1	11	11.6649	2	20	6.9520	4	9	4.9995	5	18	7.6075
1	12	10.9249	3	1	7.0922	4	10	8.0350	5	19	6.0851
1	13	10.7428	3	2	7.3919	4	11	8.1971	5	20	7.1843
1	14	10.4591	3	3	7.9535	4	12	5.0569	6	1	4.7617
1	15	10.4765	3	4	6.6442	4	13	29.8019	6	2	29.7959
1	16	10.8514	3	5	6.6276	4	14	6.2097	6	3	5.4709
1	17	24.6713	3	6	6.0158	4	15	5.7409	6	4	4.4181
1	18	10.7091	3	7	6.0169	4	16	5.3858	6	5	4.9678
1	19	8.6117	3	8	5.6643	4	17	5.1034	6	6	5.7287
1	20	19.4730	3	9	6.0140	4	18	5.4506	6	7	29.8019
2	1	8.5598	3	10	8.3243	4	19	8.7409	6	8	29.8019
2	2	29.8019	3	11	7.0336	4	20	29.8019	6	9	4.3668
2	3	8.7879	3	12	8.1603	5	1	4.8230	6	10	4.4365
2	4	9.4823	3	13	6.3853	5	2	5.8764	6	11	29.8019
2	5	6.8502	3	14	7.0325	5	3	6.5732	6	12	4.9706
2	6	7.2812	3	15	6.4129	5	4	5.5224	6	13	5.8841
2	7	6.5902	3	16	5.2036	5	5	7.2342	6	14	4.6177
2	8	6.9174	3	17	6.4128	5	6	4.7496	6	15	5.0347

6	16	5.6674	8	6	29.8017	9	16	4.2690	11	6	29.8019
6	17	29.8019	8	7	29.8019	9	17	4.3124	11	7	29.8018
6	18	29.80199	8	8	4.9462	9	18	29.8019	11	8	4.1135
6	19	29.7430	8	9	29.8019	9	19	3.7848	11	9	29.8019
6	20	4.7979	8	10	4.0483	9	20	4.4355	11	10	29.8019
7	1	29.8171	8	11	4.3657	10	1	29.8019	11	11	4.3623
7	2	5.0083	8	12	4.1154	10	2	29.8019	11	12	29.8041
7	3	4.5622	8	13	4.5411	10	3	29.8019	11	13	29.8019
7	4	29.8019	8	14	5.0216	10	4	29.8019	11	14	4.6364
7	5	29.8019	8	15	4.0458	10	5	4.5895	11	15	29.8019
7	6	4.5454	8	16	3.9660	10	6	29.8019	11	16	4.3366
7	7	29.8019	8	17	4.2597	10	7	29.8018	11	17	7,8786
7	8	4.0442	**8**	**18**	**0.1076**	10	8	29.8019	11	18	5.4230
7	9	29.8019	8	19	5.0279	10	9	29.8019	11	19	29.8019
7	10	29.8019	8	20	4.6963	10	10	3.8063	11	20	29.8007
7	11	29.8019	9	1	29.8004	10	11	3.6227	12	1	29.8019
7	12	29.8019	9	2	29.8019	10	12	29.8019	12	2	29.8019
7	13	4.4382	9	3	29.8019	10	13	4.8345	12	3	29.7987
7	14	3.4564	9	4	29.8019	10	14	3.6387	12	4	29.8019
7	15	5.1249	9	5	29.8019	10	15	29.8038	12	5	29.8019
7	16	29.8085	9	6	29.8014	10	16	29.7373	12	6	29.8019
7	17	4.8633	9	7	3.6988	10	17	6.2042	12	7	29.801
7	18	4.8592	9	8	29.8003	10	18	29.8019	12	8	5.2496
7	19	5.5371	9	9	29.8019	10	19	5.0749	12	9	29.8019
7	20	4.3218	9	10	3.6949	10	20	4.1984	12	10	29.8019
8	1	29.7919	9	11	29.8019	11	1	29.8019	12	11	4.1134
8	2	29.8019	9	12	29.7840	11	2	29.8019	12	12	29.8019
8	3	29.8019	9	13	29.8038	11	3	29.8019	12	13	4.8828
8	4	29.8019	**9**	14	29.8019	11	4	29.8019	12	14	4.0093
8	5	29.8019	9	15	4.8684	11	5	29.8019	12	15	29.8019

12	16	29.8005	14	6	29.8019	15	16	4.1580	17	6	29.8019
12	17	4.9951	14	7	29.8019	15	17	4.2472	17	7	29.8019
12	18	29.8019	14	8	29.8018	15	18	4.1587	17	8	29.8019
12	19	4.9444	14	9	29.8019	15	19	29.80194	17	9	5.2039
12	20	4.1927	14	10	3.7996	15	20	3.9268	17	10	29.8019
13	1	29.8019	14	11	29.8019	16	1	29.8019	17	11	29.8019
13	2	29.8019	14	12	4.3013	16	2	29.7990	17	12	4.2570
13	3	29.8019	14	13	29.8027	16	3	29.8019	17	13	29.8019
13	4	29.80199	14	14	5.2822	16	4	29.8019	17	14	29.8019
13	5	29.8019	14	15	29.8019	16	5	29.8019	17	15	29.8019
13	6	29.8019	14	16	29.8019	16	6	29.8019	17	16	29.8019
13	7	4.9974	14	17	5.1689	16	7	29.8019	17	17	29.8230
13	8	29.8019	14	18	5.5033	16	8	3.7564	17	18	29.8083
13	9	29.8019	14	19	4.8190	16	9	29.8019	17	19	29.8019
13	10	29.8021	14	20	5.2048	16	10	29.8019	17	20	29.8074
13	11	29.8019	15	1	29.7875	16	11	29.8151	18	1	29.8001
13	12	29.8007	15	2	29.8019	16	12	29.8019	18	2	29.8019
13	13	29.7974	15	3	29.8019	16	13	29.8001	18	3	29.8019
13	14	4.8208	15	4	29.8019	16	14	29.8019	18	4	29.8019
13	15	29.8016	15	5	29.8019	16	15	5.1779	18	5	29.8019
13	16	4.4321	15	6	29.7884	16	16	4.8682	18	6	29.8019
13	17	4.5209	15	7	29.8019	16	17	4.6935	18	7	29.8019
13	18	3.8307	15	8	29.8019	16	18	29.8019	18	8	29.8019
13	19	4.4477	15	9	29.8019	16	19	3.7767	18	9	29.8129
13	20	29.8019	15	10	4.2517	16	20	4.8762	18	10	29.7767
14	1	29.8019	15	11	29.8019	17	1	29.8019	18	11	29.8019
14	2	29.8019	15	12	5.9261	17	2	29.8019	18	12	5.2228
14	3	29.8019	15	13	29.8098	17	3	29.8019	18	13	29.8019
14	4	29.8019	15	14	5.2195	17	4	29.7768	18	14	4.7674
14	5	29.8019	15	15	3.5261	17	5	29.8019	18	15	29.8019

18	16	29.8019
18	17	29.8065
18	18	29.8019
18	19	29.8019
18	20	29.8088
19	1	29.8019
19	2	29.8019
19	3	29.8019
19	4	29.8019
19	5	29.8019
19	6	29.7999
19	7	29.8019
19	8	29.8019
19	9	29.8099
19	10	29.80199
19	11	4.6259
19	12	29.8019
19	13	29.8019
19	14	29.80199
19	15	29.8019
19	16	29.8019
19	17	3.7466
19	18	29.8019
19	19	29.8019
19	20	29.8019
20	1	29.8019
20	2	29.8019
20	3	29.8019
20	4	29.8019
20	5	29.8019
20	6	29.8019
20	7	29.8019
20	8	29.8019
20	9	29.8019
20	10	29.8019
20	11	29.8019
20	12	29.8019
20	13	29.8019
20	14	29.8019
20	15	29.8019
20	16	5.0090
20	17	4.2241
20	18	3.9966
20	19	29.8019
20	20	29.8019