T.C.
YEDITEPE UNIVERSITY
INSTITUTE OF HEALTH SCIENCES
DEPARTMENT OF PHARMACEUTICAL CHEMISTRY

DEVELOPMENT OF A QSAR MODEL TO PREDICT iNOS INHIBITION OF SELECTED COMPOUNDS USED IN ALZHEIMER'S DISEASE TREATMENT

MASTER OF SCIENCE THESIS

SHKAR MARIWAN AHMED RASHEED

İSTANBUL-2024

T.C.
YEDITEPE UNIVERSITY
INSTITUTE OF HEALTH SCIENCES
DEPARTMENT OF PHARMACEUTICAL CHEMISTRY

DEVELOPMENT OF A QSAR MODEL TO PREDICT iNOS INHIBITION OF SELECTED COMPOUNDS USED IN ALZHEIMER'S DISEASE TREATMENT

MASTER OF SCIENCE THESIS

SHKAR MARIWAN AHMED RASHEED

SUPERVISOR

Prof. Dr. Meriç KÖKSAL AKKOÇ
Asst. Prof. Dr. Gülçin TUĞCU

İSTANBUL-2024

THESIS APPROVAL FORM

Institute : Yeditepe University Institute of Health Sciences

Program : Pharmaceutical Chemistry

Title of the Thesis : Development of a QSAR Model to Predict iNOS Inhibition of Selected Compounds Used in Alzheimer's Disease Treatment

Owner of the Thesis : Shkar MARIWAN AHMED

Examination Date : 08.01.2024

This study have approved as a Master Thesis in regard to content and quality by the Jury.

Chair of the Jury:	Prof. Dr. Hülya AKGÜN (Yeditepe University Faculty of Pharmacy)	(Signature)
Supervisor:	Prof. Dr. Meriç KÖKSAL AKKOÇ (Yeditepe University Faculty of Pharmacy)	(Signature)
Member/Examiner:	Prof. Dr. İlkay KÜÇÜKGÜZEL (Fenerbahçe University Faculty of Pharmacy)	(Signature)
Member/Examiner:	Asst. Prof. Dr. Tuğçe ÖZYAZICI (Health Sciences University Faculty of Pharmacy)	(Signature)
Cosupervisor:	Asst. Prof. Dr. Gülçin TUĞCU (Yeditepe University Faculty of Pharmacy)	(Signature)

APPROVAL

This thesis has been deemed by the jury in accordance with the relevant articles of Yeditepe University Graduate Education and Examinations Regulation and has been approved by Administrative Board of Institute with decision dated ..... and numbered .....

Prof. Dr. Bayram Yılmaz

Director of Institute of Health Science

DECLARATION

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Shkar Mariwan Ahmed

Signature …..……..……..

DEDICATION

This thesis is lovingly dedicated to my dear family - my father, mother, and my beloved wife, Zonna. Their unwavering support and endless encouragement have been the guiding lights that illuminated my path throughout this endeavor.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my esteemed thesis advisors, Prof. Dr. Meriç KÖKSAL and Asst. Prof. Dr. Gülçin TUĞCU. Their unwavering support and guidance have been instrumental throughout my research journey. I am truly grateful for the invaluable knowledge and expertise they have shared with me.

Prof. Dr. Meriç KÖKSAL and Asst. Prof. Dr. Gülçin TUĞCU have been exceptional mentors, always available and approachable whenever I encountered challenges or had inquiries regarding my research or writing. Their doors were open, and they generously offered their time and insights, steering me in the right direction whenever needed.
Working with them has been a privilege, and I consider myself fortunate to have had the opportunity to learn from their vast experience and expertise. Their dedication to my academic and personal growth has been truly inspiring, and I am deeply grateful for their support throughout the thesis process.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AD	Alzheimer's disease
Aβ	Amyloid-beta
ACC	Accuracy
AUC	Area under the ROC curve
BBB	Blood-brain barrier
CT	Computed tomography
CV	Cross-validation
DT	Decision tree
eNOS	Endothelial nitric oxide synthase
E-HOMO	Highest occupied molecular orbital energy
E-LUMO	Lowest unoccupied molecular orbital energy
FDA	Food and drug administration
iNOS	Inducible nitric oxide synthase
IQR	Interquartile range
IFN-c	Interferon-c
LPS	Lipopolysaccharide
L-NMMA	NG-monomethyl-L-arginine
L-NAME	L-NG-nitro arginine methyl ester
L-NIL	L-N6-(1-iminoethyl)-lysine
LR	Logistic regression
MRI	Magnetic resonance imaging
MI	Myocardial infarction
ML	Machine learning
MCC	Matthews Correlation Coefficient
NO	Nitric oxide
NMDA	N-methyl-D-aspartate
nNOS	Neuronal nitric oxide synthase
NB	Naïve bayes
ONOO	Peroxynitrite
PET	Positron emission tomography
QSAR	Quantitative structure-activity relationship
RF	Random forest
RES	Resveratrol
RNSs	Reactive nitrogen species
SMT	2-methyl-2-thio pseudourea sulfate
SOM	Self-organizing map

ABSTRACT

Mariwan Ahmed, S. (2023). Development of a QSAR Model to Predict iNOS Inhibition of Selected Ccompounds Used in Alzhimers Disease Treatment. Yeditepe University, Institute of health scince, Department of Pharmaceutical Chemistry, MSc thesis, Istanbul.

This thesis focuses on the development of a Quantitative Structure-Activity Relationship (QSAR) model to predict the inhibition of inducible Nitric Oxide Synthase (iNOS) by selected compounds used in Alzheimer's disease (AD) treatment. AD is a neurodegenerative disorder with a complex etiology, requiring the development of effective treatment strategies.

In this study, we have gathered 90 synthetic derivatives of iNOS inhibitors from various publications in Elsevier and Web of Science. These derivatives include [(4-(1,2,4-oxadiazol-5-yl)phenyl]-2-aminoacetamide derivatives, 7-O-1,2,3-triazole hesperetin derivatives, and 3-(4-pyridyl)-5-(4-sulfamido-phenyl)-1,2,4-oxadiazole derivatives. We have specifically chosen these compounds to develop a QSAR model for predicting the biological activity of the selected iNOS inhibitors. Additionally, we aim to assess their potential effectiveness in the treatment of AD.

The results demonstrate the successful development of a reliable QSAR model capable of predicting iNOS inhibition of the selected compounds. Key molecular descriptors contributing to the inhibitory activity are identified, providing insights into the structure-activity relationships. The importance of this research lies in its potential to aid in the discovery and development of novel compounds with enhanced iNOS inhibitory activity for the treatment of AD. Overall, this study contributes to the field of drug discovery by providing a valuable tool for predicting the inhibitory potential of compounds targeting iNOS in AD treatment. The developed QSAR model offers a cost-effective and time-efficient approach to prioritize compounds for further experimental evaluation, ultimately accelerating the drug discovery process and potentially leading to improved therapies for AD.

Keywords: Alzheimer's disease, iNOS inhibitors, neuroinflammation, QSAR, metod validation.

ÖZET

Mariwan Ahmed, S. (2023). Alzheimer Hastalığı Tedavisinde Kullanılan Seçilmiş Bileşiklerin iNOS İnhibisyonunu Tahmin Etmek İçin Bir QSAR Modeli Geliştirilmesi. Yeditepe Üniversitesi, Sağlık Bilimleri Enstitüsü, Eczacılık Kimyası Anabilim Dalı, Yüksek Lisans Tezi, İstanbul.

Bu tez, Alzheimer hastalığı (AH) tedavisinde kullanılabilecek, indüklenebilir nitrik oksit sentaz (iNOS) üzerine etkili literatürden seçilmiş bileşiklerle enzim inhibisyonunu tahmin etmek amacıyla bir kantitatif yapı-etki ilişkisi (QSAR) modelinin geliştirilmesini hedeflemektedir. AH, kompleks bir etiyolojiye sahip olan nörodejeneratif bir bozukluktur ve etkili tedavi stratejilerinin geliştirilmesini önemlidir. Bu çalışmada, Elsevier ve Web of Science veri tabanlarından taranmış farklı literatürlerinden toplam 90 adet [(4-(1,2,4-okzadiazol-5-il)fenil)-2-aminoasetamid türevi, 7-O-1,2,3-triazol hesperetin türevi ve 3-(4-piridil)-5-(4-sülfamido-fenil)-1,2,4-okzadiazol türevi sentetik iNOS inhibitörü kullanarak biyolojik aktiviteyi tahmin etmek için bir QSAR modeli oluşturulmuştur. Elde edilen sonuçlar, iNOS inhibisyonunu tahmin edebilen güvenilir bir QSAR modelinin başarılı bir şekilde geliştirildiğini göstermektedir. Modelde, inhibitör aktivitesine katkıda bulunan ana moleküler tanımlayıcılar, yapı-aktivite ilişkilerini ortaya koyacak şekilde belirlenmiştir. Sonuç olarak, bu çalışma, AH tedavisinde iNOS'u hedef alan bileşiklerin inhibisyon potansiyelini tahmin etmek için değerli bir araç sağlayarak ilaç keşfi alanına katkıda bulunmaktadır. Geliştirilen QSAR modeli, bileşikleri daha fazla deneysel değerlendirmeye önceliklendirmek için maliyet-etkin ve zaman-etkin bir yaklaşım sunmakta, ayrıca, ilaç keşif sürecini hızlandırmakta ve AH için iyileştirilmiş terapileri açığa çıkarma potansiyeli sunmaktadır.

Anahtar Kelimeler: Alzheimer hastalığı, iNOS inhibitörleri, nöroinflamasyon, QSAR, metot validasyonu

1. INTRODUCTION AND PURPOSE

Alzheimer's disease (AD) is an age-related neurodegenerative disease with a complex etiology. It is also the leading cause of dementia in the elderly, which accounts for 60-80% of all cases. According to the World Health Organization (WHO), approximately 50 million people worldwide are now living with dementia today, and this number is expected to triple by 2050 [1]. Symptoms of AD include language impairment, progressive memory loss, and visual-spatial impairment, all of which require caregiving. Early signs of AD may vary from patient to patient, but typically memory problems are one of the early signs of AD. However, the very early stages of AD may also be indicated by a deterioration in other cognitive abilities, including the ability to express oneself clearly, problems with vision or spatial awareness, and impaired reasoning or judgment [2, 3].

Although the pathology of AD has been studied for decades, the exact cause remains unknown. Recent developments in the diagnostic strategy of AD have aided to diagnose pathological abnormalities in the early stages of the disease. The current understanding of the pathological features of AD, which are the formation of neurofibrillary tangles (NFTs) inside the neurons, senile plaques outside neurons, and neuronal loss, is shown in Figure 1 [4]. However, there are other critical factors that promote the onset and development of AD, such as oxidative stress, neuroinflammation, and impaired glucose metabolism in the brain. In addition, amyloid plaques and tau protein-based neurofibrillary tangles may form years before clinical dementia manifests [5].

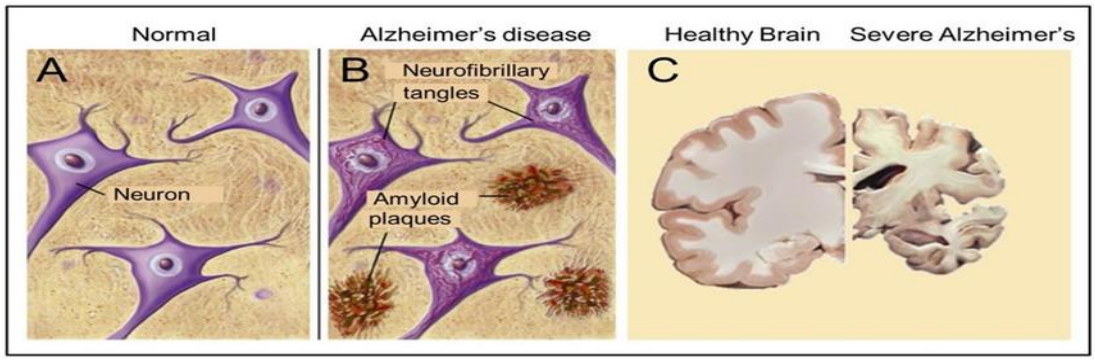


Figure 1. Difference between a healthy brain and the brain of a patient with severe AD.

In the ongoing pursuit of an effective treatment for Alzheimer's disease (AD), scientists have made significant strides in the current decade, but we still find ourselves quite distant from achieving a comprehensive strategy to halt the disease's relentless progression and development [6]. The existing treatment options primarily target the alleviation of AD symptoms, offering limited respite. Recent research endeavors have increasingly honed in on the pivotal role of neuroinflammation as a potential avenue for a breakthrough in AD treatment. Neuroinflammation stands as one of the principal catalysts in the development and exacerbation of AD, making it a prime focus for investigation. During neuroinflammatory processes, inducible nitric oxide synthase (iNOS) triggers the excessive production of nitric oxide (NO) in response to inflammatory stimuli, ultimately contributing to the progression and worsening of AD [7]. As such, the quest for effective iNOS inhibitors becomes paramount in our mission to curb and decelerate the relentless march of AD.

In the realm of our research, we embarked on a journey to identify and examine 90 synthetic chemical compounds, each exhibiting potential as iNOS enzyme inhibitors in the context of AD treatment. Curiously, despite the availability of literature detailing these inhibitors, no quantitative structure-activity relationship (QSAR) models had previously been developed to shed light on their mechanisms. Consequently, we undertook the task of selecting these compounds as the foundation for creating a classification model aimed at predicting their inhibitory efficacy within the framework of AD treatment. Our research endeavors hold the promise of contributing essential insights and guidance to advance our understanding of potential therapeutic strategies for Alzheimer's disease.

2. LITERATURE REVIEW

2.1 Alzheimer’s Disease

Alzheimer's disease is a neurodegenerative disorder that leads to the deterioration of brain cells, and it’s a leading cause of dementia [8], characterized by cognitive decline, impaired thinking, and independence in an individual's daily activities. Alzheimer’s disease is known to be the most common cause of dementia in people 65 years and older. Moreover, AD affects 50% of people over the age of 80 years, and it affects more than 10% of people over the age of 65 years [9]. The current knowledge about the pathological features of Alzheimer’s disease is the formation of neurofibrillary tangles (NFTs) inside the neurons, senile plaque outside neurons, and neuronal loss. However, there are other significant factors that contribute to the onset and progression of AD, such as oxidative stress, disruption of glucose metabolism in the brain, and neuroinflammation [10].

2.1.1 Stages of Alzheimer’s Disease

Alzheimer's disease can be divided into three stages: first, the pre-dementia stage: this early stage of AD is difficult to distinguish from issues related to stress or normal aging [11, 12]. One of the first and most common signs of the pre-dementia stage is a decline in episodic memory. At this stage, there is mild impairment of verbal and visuospatial functions, but, no failure of sensory and motor function. During this early stage of AD, the person retains their independence and does not require care [12]. Furthermore, memory loss in the early stages of AD affects recent declarative memories more than short-term and implicit memories [13].

Second, in the moderate stage, recent memory continues to decline. During this stage, the patient appears to be living in the past due to their inability to create new memories [14]. The patient may still be able to perform basic daily activities but may require assistance in areas such as dressing and grooming. At this point, patients often lose sight of their disease and become delusional [13, 14].

The third stage of AD is characterized by symptoms such as agitation, cognitive impairment, and depression. According to the findings of a longitudinal research carried out in 1993, patients in the third stage of AD have the greatest need for caregiving [15]. However, early memories may be lost in this severe stage of the disease. Basic activities of daily living are now being affected and gradually decline. Communication deteriorates and language becomes significantly impaired [16, 17].

2.1.2 Diagnosis of Alzheimer's Disease

The diagnostic strategy for AD is carried out in several ways, first, the patient's history should include information obtained from family members and relatives, then a mental status test should be done. This should involve a validated cognitive function test, while the physical examination should primarily concentrate on identifying neurological and vascular indicators, supported by investigations and the patient's medical history [2]. Brain scans are an effective way to diagnose AD by using neuroimaging to detect structural abnormalities of the brain. Scans include PET, MRI, and CT (Fig. 2), which are effective techniques in disease detection. Each type of scanning system has its own strategy for detecting abnormalities in specific areas of the brain. Although brain scanning is not the only way to diagnose AD, recent research has yielded promising results that could change the way physicians diagnose the disease. Based on a study conducted in 2002, which involved gathering MRI scans from 50 participants experiencing cognitive impairment, the MRI was utilized to detect hippocampal volume and determine its significance as an indicator of AD neuropathology. The findings suggest that brain imaging techniques may have the potential to identify elderly individuals with Alzheimer's disease, even in the absence of memory impairment symptoms [18].

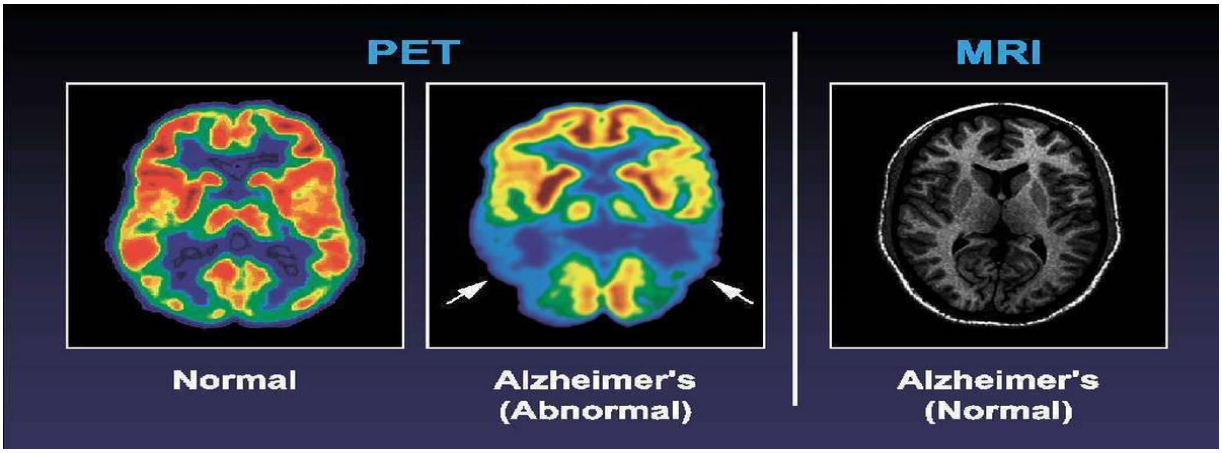


Figure 2. Pet and MRI scan of AD patient**.**

2.1.2.1 Positron Emission Tomography (PET) Scan

A PET scan is a medical imaging technique that produces a three-dimensional color image of the human body's tissues [19]. To perform a PET scan, a radiotracer, which is a radioactive medicine bound to a naturally occurring chemical radioactive medicine, is injected into the bloodstream of the patient. Glucose is a chemical that is often used to learn more about Alzheimer's disease. The radiotracer is taken to the organs that get their energy from that particular chemical. Positrons are given off when the molecule is broken down. Then the PET scan picks up the energy from these positrons and uses it to make a picture on the output screen. This picture shows how well the radiotracer is broken down in the body tissue of the patient, the amount of energy given off by positrons makes a range of colors and intensities that show the activity of the brain. A PET scan can find differences in the brain's blood flow, metabolism, and cellular communications to each other, as well as other things going on in the brain [19].

Small *et al.* conducted a study in 1996 described how to use a PET scan to detect changes in glucose metabolism in an AD patient's brain. Irregularly low metabolic rate of glucose is observed in the parietal, temporal, and posterior cortices. The rate was reduced even further in patients with advanced diseases that affected more areas of the brain (Fig. 3) [20]. Furthermore, another study have been done on the effectiveness of PET scans in diagnosing AD, which reveal that a PET scan has the capability to identify alterations in glucose

metabolism long before the manifestation of clinical symptoms. Additionally, PET images could also be used to determine the effectiveness of AD treatments [21].

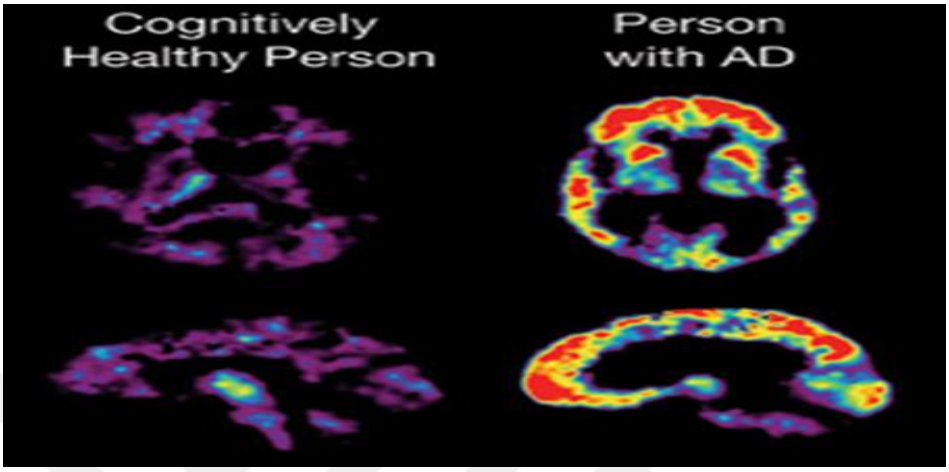


Figure 3. PET scan of AD patient.

2.1.2.2 Magnetic Resonance Imaging (MRI)

MRI is a medical imaging technique that creates 2D or 3D images of the body to aid in diagnosing diseases and injuries. The MRI system comprises a superconducting magnet that produces a stable and large magnetic field, as well as gradient magnets that produce weaker magnetic fields to scan various body tissues. The hydrogen atoms in the billions of atoms present in each cell of the human body are affected by the magnetic field. While hydrogen molecules align themselves with the MRI's magnetic field direction, hydrogen atoms spin randomly around an axis outside it, with half pointing at the patient's head and half at their feet. A few of these atoms are not cancelled out from every million atoms. The system then releases a hydrogen-specific radio frequency pulse, which alters the direction of these protons' spin. When the protons stop spinning, they emit energy that the system interprets. An MRI image shows each tissue type's distinctive grayscale shade when using a contrast dye [22].

Researchers can determine whether an MRI can detect cellular death and structural changes in AD patients' brains by understanding how the system works. In 2002, a research study was conducted involving 50 participants with varying degrees of cognitive impairments.

Postmortem MRI scans were collected from these individuals, specifically focusing on the hippocampal volume. The aim of the study was to investigate the significance of hippocampal volume as an indicator of neuropathological changes associated with AD. The findings suggested that non-demented elderly with AD neuropathology could be identified using the scans, even before memory impairment presents itself, potentially allowing physicians to slow Alzheimer's disease progression by identifying at-risk patients [23].

2.1.2.3 Computed Tomography (CT) Scan

A CT scan is a medical imaging procedure that generates a sequence of cross-sectional images of the body. These images are combined and processed by a computer to create a comprehensive and detailed image. CT scans provide information on tissue density across the body and various areas of the brain. Contrast dye can be used to differentiate between similar tissues and improve image clarity [24].

2.1.3 Current Treatment Options for Alzheimer's Disease

To date, there are six medication therapies approved by the US Food and Drug Administration (FDA) for AD. The drugs are divided into two groups, first group is cholinesterase inhibitors such as (tacrine, donepezil, galantamine, and rivastigmine) (Fig. 4) which may delay clinical decline with cognitive benefits [4]. On the other hand, the second group of drugs such as memantine (Fig. 4) which is an N-methyl-D-aspartate (NMDA) receptor antagonist, which only treats dementia symptoms without stopping and delaying the disease progression, while aducanumab ($C_{6472}H_{10028}N_{1740}O_{2014}S_{46}$) is an anti-amyloid antibody that could delay the clinical decline. Donepezil, galantamine, and rivastigmine are the most commonly used drugs for mild to moderate dementia symptoms, but, memantine is preferred for more severe cases of dementia because it regenerates the function of the damaged neurons. Since 2003, only aducanumab has been approved as a new therapy for AD. It targets amyloid β, a protein fragment that accumulates in the brain and forms plaques, thereby interfering with neuronal communication and possibly inducing inflammation. In addition, in recent years, the FDA has approved a combination therapy consisting of donepezil plus memantine at a fixed dose for moderate to severe cases of AD [25].

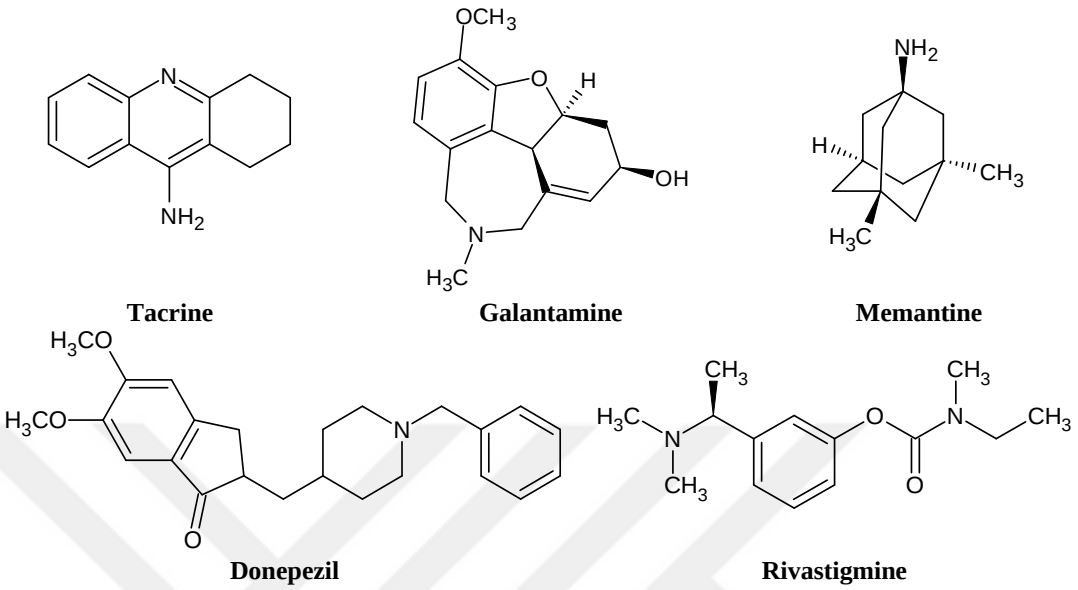


Figure 4. The structures of the approved drugs for AD.

2.2 Neuroinflammation and Alzheimer's Disease

Recent investigations have demonstrated that neuroinflammation is one of the major causes of the development and exacerbation of AD [26, 27]. Interestingly, neuroinflammation could be a beneficial immune response that limits and slows disease progression [28]. For example, in its normal range, neuroinflammation modulates microglial activation and involves in the clearance of Aβ and cellular debris via microglial phagocytosis [29, 30]. On the other hand, some researchers believe that neuroinflammation is merely a byproduct of the disease process and may not significantly alter its course. However, evidence from clinical studies and animal experiments has demonstrated that enhanced neuroinflammatory cascades mediated by primed microglial cells also contribute to AD pathogenesis [31, 32].

Neuroinflammation during AD is characterized by microglial activation and astrocyte release of various chemokines and cytokines, which impair blood-brain barrier (BBB) function and stimulate cognitive impairment [33, 34]. Furthermore, during neuroinflammation, NO is

overproduced by the enzyme iNOS in response to inflammatory stimuli, leading to an exacerbation of AD and dementia [35].

2.2.1 Nitric Oxide

Nitric oxide (NO) serves as a mediator in numerous physiological and pathophysiological processes in the human body, enzymes known as nitric oxide synthase (NOS) are responsible for converting the amino acid L-arginine into L-citrulline in order to create it. NOS enzymes can be found in three different isoforms, each of which corresponds to a different part of the body where it is found: type I is known as neuronal NOS (nNOS), type II is inducible NOS (iNOS), and lastly, type III is endothelial NOS (eNOS) (Fig. 5). While constitutive NOS enzymes are regulated by calcium-calmodulin protein binding, iNOS enzymes are regulated by transcription and do not depend on calcium-calmodulin [36].

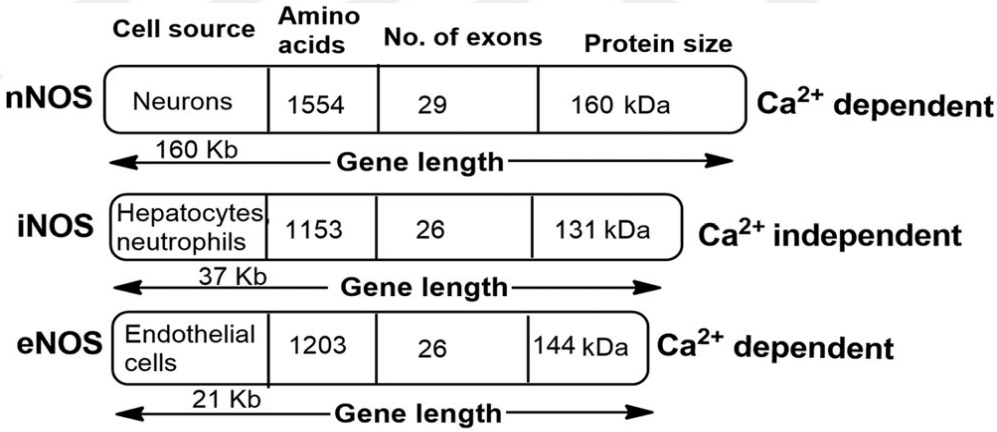


Figure 5. Structural difference of nitric oxide synthase isoforms.

2.2.2 Inducible Nitric Oxide Synthase

iNOS is predominantly expressed in immune cells like neutrophils and macrophages, a variety of stimuli such as cytokines, bacterial endotoxins, or oxidative stress can trigger the production of iNOS [37]. Unlike nNOS and eNOS, is capable of producing high quantities of NO for a prolonged period. iNOS plays a role in the immune response by aiding macrophage defense mechanisms. However, when produced in non-immune cells, it can

cause harm to tissues. Excessive iNOS production leads to an increased generation of nitric oxide (NO), which can react with the superoxide radical to form reactive nitrogen species (RNSs) such as peroxynitrite (ONOO) and dinitrogen trioxide (N_2O_3). These RNSs have the potential to induce inflammation and oxidative stress, contributing to the development of numerous pathological conditions such as multiple sclerosis, stroke, Alzheimer's disease, rheumatoid arthritis, Parkinson's disease, diabetes, and celiac disease [38].

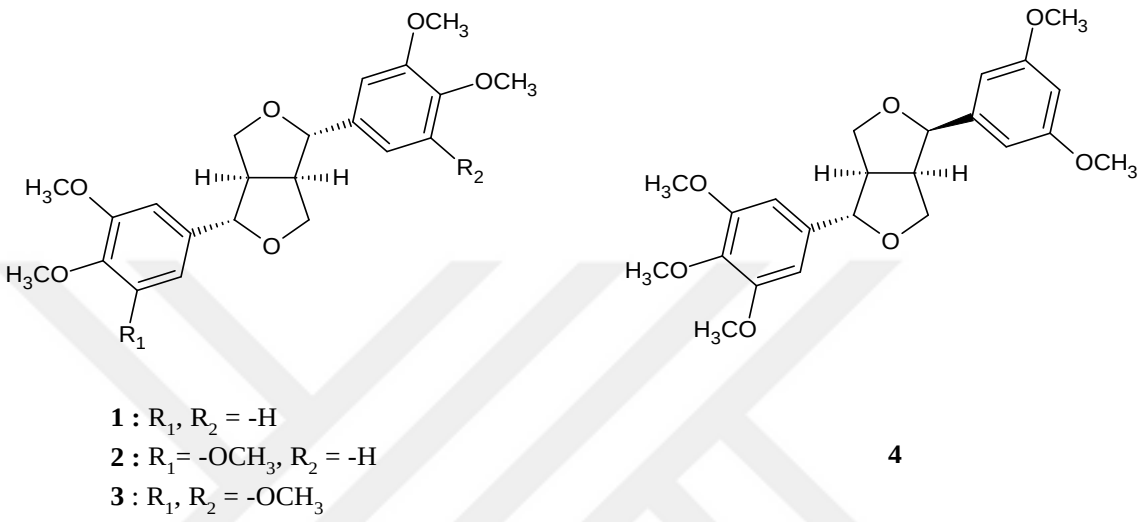
2.3 iNOS Inhibitors Treatment for Alzheimer's Disease

Numerous studies have focused on investigating the role of neuroinflammation in the pathogenesis of AD. It has been identified that chronic and prolonged activation of microglia, particularly interleukin 1β (IL-1β), is the main cause of the increase in inflammatory factors such as NO, TNF-α, and IL-1β. These inflammatory substances can lead to local inflammatory response as well as direct neuronal death. Additionally, the inflammatory environment may promote the development of NFTs and senile plaques, contributing to the progression of AD. The second isoform of nitric oxide synthases (iNOS) plays a significant role in producing NO in response to neuronal inflammation, leading to exacerbation of AD and dementia. Therefore, the development of iNOS inhibitors is crucial for the treatment of AD [7, 39].

2.3.1 iNOS Inhibitors from Natural Sources

In recent decades, plants have been used as medicines to treat various diseases, and many plants have been used as anti-inflammatory agents that act by inhibiting the production and expression of NO. Furthermore, compounds isolated from these plants provided a medical chemist with a crucial lead to find effective and safe iNOS inhibitors for the treatment of AD [29]. The anti-neuroinflammatory effect of four ligands, (+)-eudesmin (**1**), (+)-magnolin (**2**), (+)-yangambin (**3**), and epimagnolin B (**4**) isolated from *Magnolia fargesii* flower buds were tested for their iNOS inhibitory activity against neurodegenerative disorders such as AD (Fig. 6). These four compounds showed IC_{50} values ranging from 10.9-30.0 μM. Among the group, compound **4** showed the most potent inhibition of LPS-induced NO production with an IC_{50}

value of 10.9±1.6 μM compared to the positive control drug NG-monomethyl-L-arginine (L-NMMA) IC_{50}=19.2 ± 1.8 μM [40].

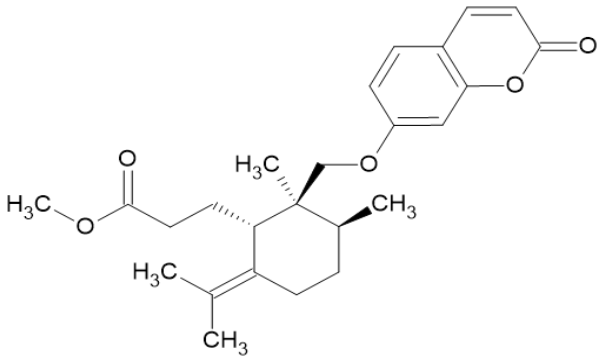


1 : R_1, R_2 = -H
2 : R_1= -OCH_3, R_2 = -H
3 : R_1, R_2 = -OCH_3

4

Figure 6. Absolute configuration of natural iNOS inhibitors isolated from *Magnolia fargesii.*

Kohno *et al.* investigated the potential inhibitory effects of various terpenoid coumarins such as (methyl galbanate, galbanic acid, and farnesiferol A) that have been extracted from *Ferula szowitsiana DC.* Specifically, the study aims to investigate whether these compounds can inhibit the production of NO in RAW264.7 mouse macrophage cells that have been stimulated with lipopolysaccharide (LPS) and interferon-c (IFN-c). Among the terpenoid coumarins that studied, it was observed that in the presence of methyl galbanate **5** (Fig. 7), LPS/IFN-c-induced iNOS mRNA expression was significantly reduced to 52% of the level found with LPS/IFN-c stimulation alone [41].



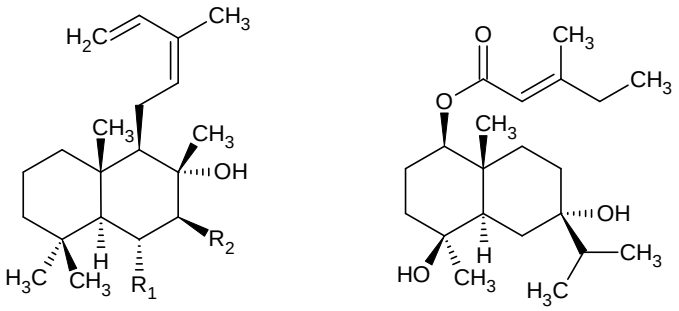
5

Figure 7. Chemical structure of methyl galbanate.

In recent years, two other studies for the development of iNOS inhibitors from natural sources were performed by Liu *et al*. In both studies, the anti-neuroinflammatory effect of compounds isolated from plant sources was evaluated by inhibiting LPS-induced NO release in murine microglial BV-2 cells. In addition, 2-methyl-2- thiopseudourea, sulfate (SMT) was used as a positive control in the studies.

In the first study, the anti-neuroinflammatory effects of twelve daphnane diterpenoids isolated from the twigs of *Trigonostemon thyrsoideus* were investigated. The result of this study demonstrated that all the compounds had an inhibitory activity on LPS-induced NO production with an IC_{50} value range of 3.19-24.9 μM [42]. In the second study of the group, isolated sesquiterpenes and terpenes from the flowers of *I. japonica* were evaluated as anti-neuroinflammatory agents for AD through NO inhibition. All of the compounds showed inhibitory activity against NO production, however, five of them possessed a more inhibitory effect on LPS-induced NO release with lower IC_{50} values than 10 μM [43].

Ma *et al.* investigated the potential inhibitory effects of one new labdane diterpenoid and three new guaiane sesquiterpenoids, alongside ten known compounds from *Blumea balsamifera*. The study focused on assessing the anti-neuroinflammatory effects by impeding the release of NO in murine microglial BV-2 cells induced by LPS. Among these derivatives, three compounds **6-8** showed IC_{50} values ranging from 15.4-22.7 μM, respectively, indicating notable inhibition against LPS-induced NO production in BV-2 cells, with their IC_{50} values falling below 30 μM (Fig. 8) [44].

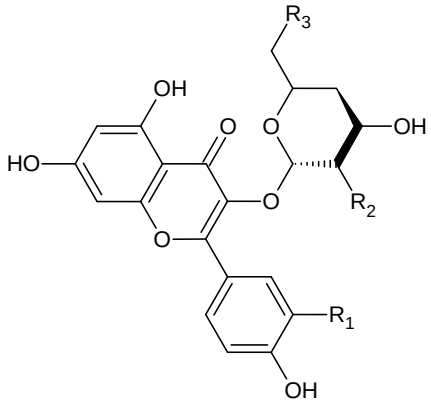


6 : R_1 = -OH, R_2 = -H
7 : R_1 = -H, R_2 = -OH

8

Figure 8. iNOS inhibitors as anti-neuroinflammatory agents for AD isolated from *Blumea balsamifera.*

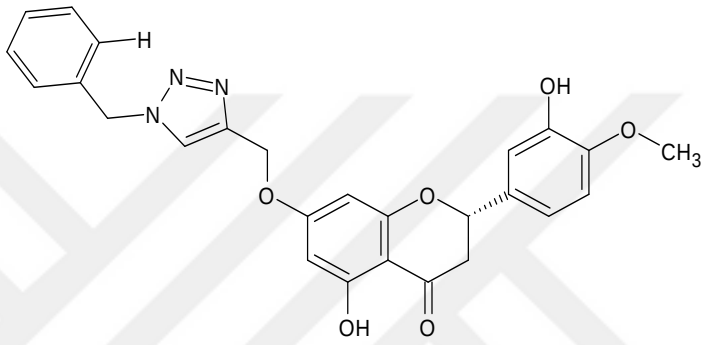
In 2021, two crucial studies have been done to test the anti-neuroinflammatory activity of numerous glycosides isolated from the leaves and stems of *Neoshirakia japonica*, and 7-O-1,2,3-triazole hesperetin derivatives. In the first study, the anti-neuroinflammatory effects of all the isolates were evaluated by inhibiting NO production against LPS-induced BV-2 microglial cells. Notably, three compounds **9-11** exhibited IC_{50} values ranging from 2.7 to 5.5 μM, respectively. These values indicated a higher degree of inhibitory activity compared to the positive control, minocycline, which had an IC_{50} value of 15.6 μM. (Fig. 9) [45].



9 : R_1 = -H, R_2 = -OH, R_3 = -S_4
10 : R_1 = -H, R_2 = -OH, R_3 = -S_3
11 : R_1, R_2 = -OH, R_3 = -S_2

Figure 9. iNOS inhibitors as anti-neuroinflammatory agents for AD isolated from *Neoshirakia japonica.*

In the second study, the anti-neuroinflammatory of 7-O-1,2,3-triazole hesperetin derivatives was evaluated by Wang *et al*. Most of the hesperetin derivatives showed better NO inhibitory activity compared to hesperetin (IC_{50}= 49.56 ± 2.39 μM). Furthermore, resveratrol (Res) was used as a positive control, and compound **12** (IC_{50}= 1.04 ± 0.31 μM) had an eight fold higher NO inhibitory capacity compared to the positive control Res (IC_{50}= 7.86 ± 1.49 μM) (Fig. 10) [7].



12

Figure 10. Hesperetin derivative potent iNOS inhibitor.

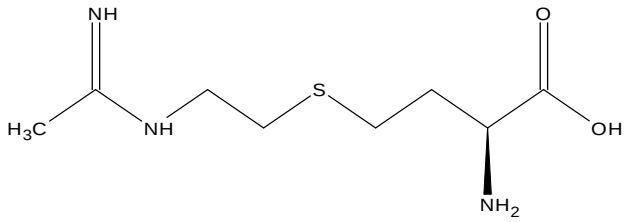
2.3.2 iNOS inhibitors from Synthetic Sources

Synthetic iNOS inhibitors have been extensively studied for their potential therapeutic applications in a variety of disease, including neurodegenerative disorders. These inhibitors are designed to target the iNOS enzyme, which is responsible for the overproduction of NO in response to neuroinflammation. Overexpression of NO can lead to tissue damage and may play a role in the etiology of a number of diseases such as AD. A number of approaches, including structure-based drug design and high-throughput screening, have been used to develop synthetic iNOS inhibitors. These inhibitors have shown promising results in preclinical research, currently being evaluated in clinical trials to determine the efficacy and safety of iNOS inhibitors [38]. Synthetic iNOS inhibitors are classified based on their chemical structures into amidinic and guanidine-based compounds, quinone-based compound, oxadiazole-based iNOS inhibitors, steroidal compounds, and others.

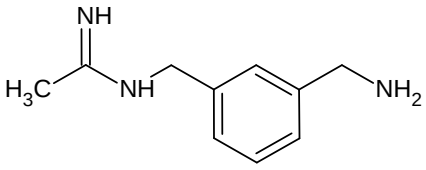
2.3.2.1 Amidine and Guanidine-based Compounds

Amidine-based iNOS inhibitors, as the name suggests, have an amidine group (-C(=NH)-NH_2) in their structure, which can participate in hydrogen bonding and electrostatic interactions with the active site of iNOS. It has been discovered that these compounds are being investigated as potential treatments for Alzheimer's disease [38].

An example of an amidine-based iNOS inhibitor is GW274150 **13** and 1400W **14**, which are selective inhibitors of iNOS that have shown promising results in preclinical studies related to AD (Fig.11). GW274150 **13** specifically inhibit the activity of iNOS by interfering with its enzymatic function, reducing the production of NO. Also, it has demonstrated anti-inflammatory effects and has been found to attenuate neuroinflammation and neurodegeneration in animal models of AD. Additionally, GW274150 **13** has shown neuroprotective properties by reducing oxidative stress and preserving neuronal function. Similarly, 1400W **14** which is another amidine-based iNOS inhibitor is a potent and selective inhibitor of iNOS. It effectively suppresses iNOS activity and the subsequent production of NO. Preclinical studies utilizing 1400W have shown promising results in attenuating neuroinflammation, reducing amyloid-beta (Aβ) deposition, and improving cognitive function in animal models of AD. The inhibition of iNOS by 1400W has been associated with decreased oxidative stress and reduced neurodegeneration [7, 38].



13



14

Figure 11. Amidine-based iNOS inhibitors.

Guanidine-based iNOS inhibitors have a guanidine group (-C(=NH)-NH-C(=NH)-NH_2) in their structure, which can also interact with the active site of iNOS. Examples of guanidine-based iNOS inhibitors include aminoguanidine **15**, and L-NG -nitro arginine methyl ester (L-NAME) **16** (Fig. 12).

Two pivotal studies were conducted on guanidine-based compounds in mouse microglial for investigating their iNOS inhibitory activity in AD. The first study was conducted in 2003, during which scientists tested the inhibitory activity of two different guanidine-based iNOS inhibitors for inhibition of AGE-induced NO production in mouse microglial. Compound **15** exhibited the highest rate of iNOS inhibition, reaching 90%. On the other hand, the other guanidine-based iNOS inhibitors, compounds **16** demonstrated inhibition values of 80% for NO production in AD [46].

In addition, Esposito *et al.* investigated the stimulation of PC12 cells with Aβ (1–42) (1 g/mL), which resulted in a significant increase in NO formation by iNOS enzymes compared to unstimulated cells. Both selective and nonselective iNOS inhibitors significantly reduced the effect of Aβ (1–42). For example, S-methyl-isothiourea (SMT) **17** (Fig. 12), a selective iNOS inhibitor, demonstrated the highest level of iNOS inhibition activity in AD, achieving an impressive 80% inhibition rate. Conversely, L-NAME **16**, a non-selective iNOS inhibitor, exhibited a lower inhibition rate of 45.6% for iNOS [47].

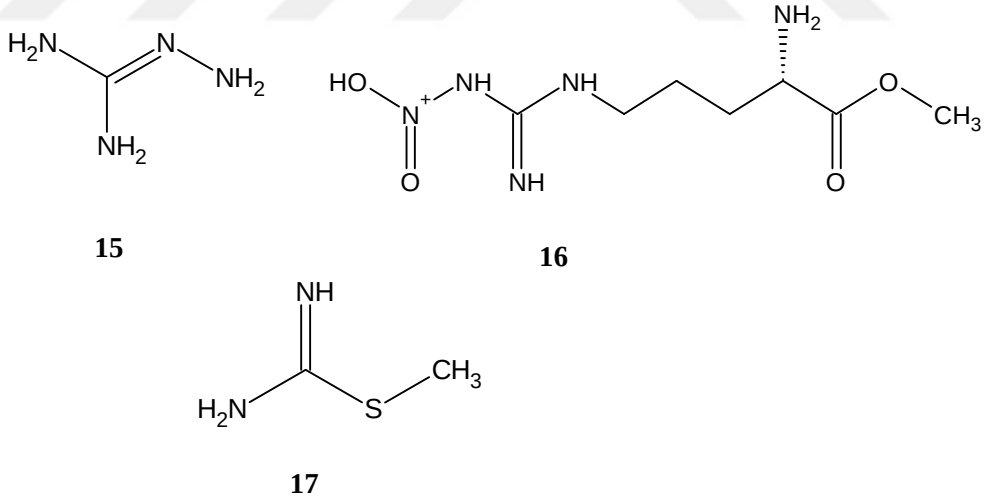
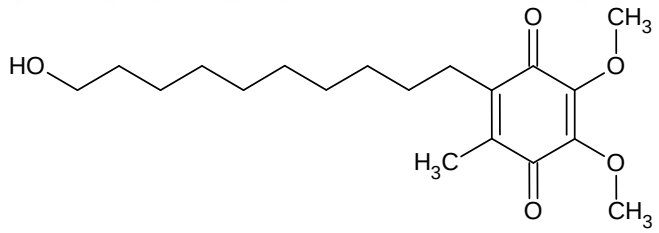


Figure 12. S-methyl-isothiourea and guanidine-based iNOS inhibitors.

2.3.2.2 Quinone-based iNOS Inhibitors

Quinone-based compounds are a class of organic compounds characterized by the presence of a quinone structure in their chemical composition. Quinones are aromatic compounds that

consist of a benzene ring conjugated with two carbonyl groups (C=O) in different positions. Quinone-based derivatives have been investigated as iNOS inhibitors for AD due to their additional antioxidant properties. These compounds are designed to scavenge reactive oxygen species (ROS) and prevent oxidative damage, which is thought to contribute to the pathogenesis of AD. An example of a quinone-based iNOS inhibitor is idebenone **18** shown in (Fig. 13), which has been evaluated in preclinical and clinical studies for its potential as a therapeutic agent for AD. Yan *et al.* conducted a study to investigate the iNOS inhibitory effects of idebenone **18** on LPS-activated BV2 cells for the treatment of neurodegenerative diseases such as AD. The researchers examined the cytotoxicity of idebenone by analyzing its effects at different concentrations (1, 2.5, and 5 μM) on LPS-activated BV2 cells. The findings revealed that idebenone exhibited a dose-dependent reduction in the LPS-stimulated NO production, and decreased mRNA expression of TNF-α, IL-1β, iNOS, and IL-6 in LPS-stimulated BV2 cells. In addition, the highest inhibition of iNOS activity was observed at the concentration of 5 μM [48].



18

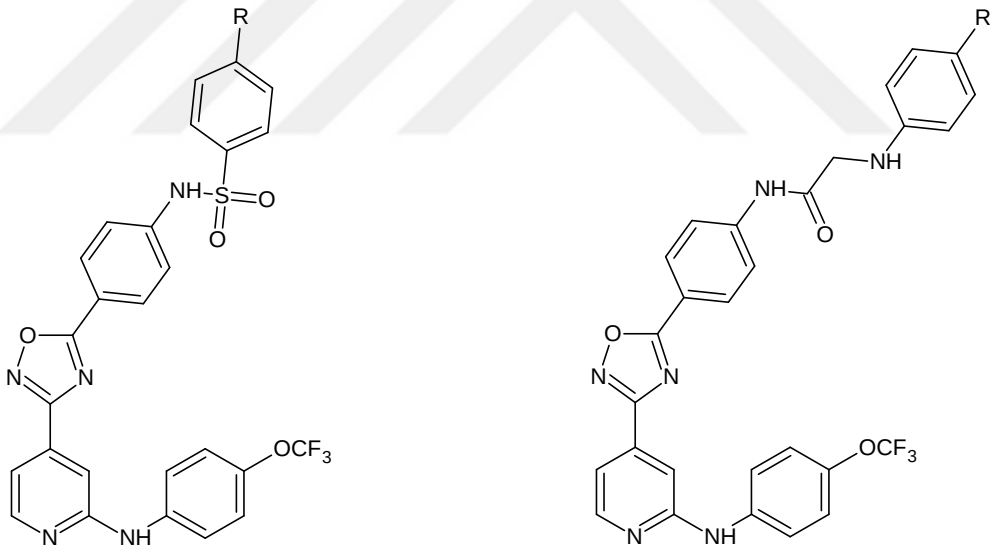
Figure 13. Quinone-based iNOS inhibitors.

2.3.2.3 Oxadiazole-based iNOS Inhibitors

Oxadiazole-based iNOS inhibitors are a class of compounds being developed as potential therapeutics for various inflammatory diseases such as neuroinflammation. These compounds have the potential to inhibit iNOS activity and reduce NO production, which may have therapeutic implications for Alzheimer's disease and other inflammatory conditions [49, 50].

Two important studies were conducted on 1,2,4-oxadiazole-containing compounds to investigate their NO inhibitory activity in AD. In the first study, a series of novel 3-(4-

pyridyl)-5-(4- sulfamido-phenyl)-1,2,4-oxadiazole derivatives were synthesized, and the anti-neuroinflammatory activity of the compounds was assessed in LPS-induced BV2 microglial cells. All compounds showed better activity against NO compared to the positive drug, Resveratrol (Res) (IC_{50}= 10.16 ± 0.12 μM). Compounds **19** and **20** showed the best inhibition rate against NO production, with an IC_{50} value range of 0.47 to 0.72 μM, which was about ten times higher than that of the positive drug Res [49]. In the second study, researchers investigated the potential anti-neuroinflammatory effect of the novel (4-(1,2,4-oxadiazol-5-yl)phenyl)-2-aminoacetamide derivatives in LPS-induced BV2 microglial cells. Among the derivatives tested, compounds **21** and **22** exhibited particularly noteworthy results, displaying more than 20-fold greater iNOS inhibitory activity compared to Res. Notably, compounds **21** and **22** demonstrated IC_{50} values ranging from 0.42 to 0.67 μM, highlighting their potent inhibitory effects on iNOS expression and NO production in AD (Fig.14) [50].



19 : R = -$COCH_3$
20 : R = -$C(CH_3)_2$

21 : R = -H
22 : R = -F

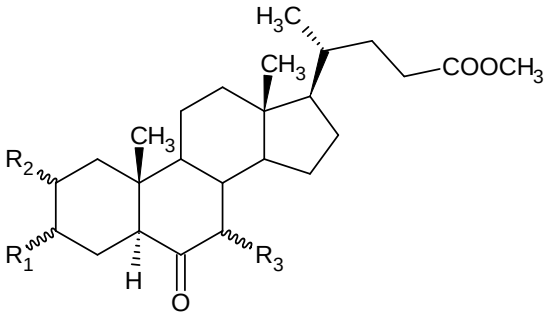
Figure 14. Oxadiazole-based iNOS inhibitors.

2.3.2.4 Steroidal Compound

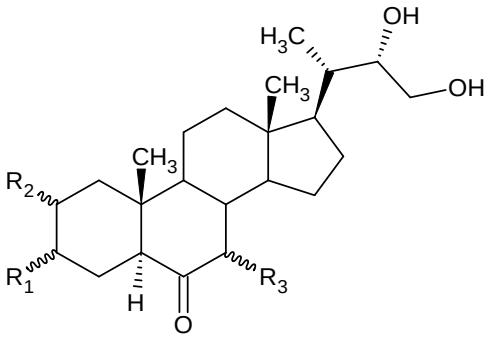
Steroid compounds are a class of organic compounds that have a characteristic structure consisting of a sterane ring. These compounds have numerous biological activities, including anti-neuroinflammatory activity in AD [38].

Yang *et al.* synthesized novel steroidal derivatives, from readily available hyodeoxycholic acid ($C_{24}H_{40}O_4$). In the study, the anti-neuroinflammatory activity of 5α-cholestan-6-one derivatives was evaluated in LPS-stimulated BV-2 microglial cells. As a result of the study, five compounds (**23-27**) strongly inhibited LPS-induced NO production with a percentage inhibition range of 73.6 – 60%, without causing cell toxicity while dehydroepiandrosterone, which was used as a reference drug in the study, has 55.3% percentage inhibition (Fig. 15).

Microglial activation is known to increase the expression of iNOS and COX-2, which are responsible for the production of PGE-2, and NO, and activated microglial cells also increase the production of proinflammatory cytokines such as IL-1β and TNF-α. As a further study, the effects of compound **23** on the mRNA expression of these cytokines were investigated. The result of the investigation clearly indicated that compound **23** which strongly inhibited LPS-induced expression of iNOS, IL-1β, TNF-α, and COX-2 in a dose-dependent manner [51].



23 : R_1 = OH , R_2 = -H , R_3 = OH
24 : R_1 = F , R_2 = -H , R_3 = OH
25 : R_1 = OH , R_2 = F , R_3 = F



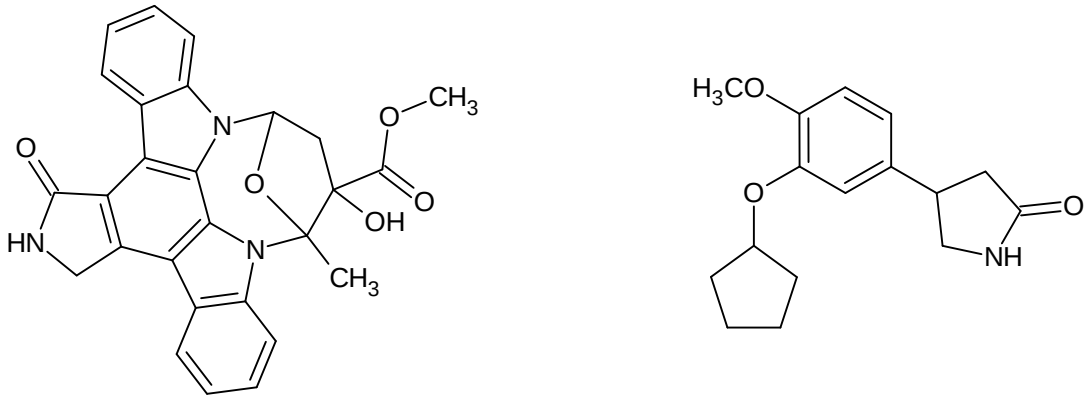
26 : R_1 = OH , R_2 = -H , R_3 = -H
27 : R_1 = OH , R_2 = -H , R_3 = F

Figure 15. Steroidal compounds as inhibitors of iNOS.

2.3.2.5 Other Compounds

Other studies have been conducted to test the iNOS inhibitory ability of various synthetic compounds, such as the study by Watterson *et al*. that investigated the role of the anti-inflammatory drug K252a **28** (Fig.16), which is a CaMKII inhibitor, that exhibits dose-dependent inhibition of lipopolysaccharide-induced increases in iNOS production and NO accumulation by the BV-2 microglial cell line. Furthermore, the study showed that the anti-inflammatory K252a inhibits the accumulation of NO and IL-1, and other cytokines during neuroinflammation in AD [52].

Zhou *et al.* explored the effects of rolipram **29** which is a Phosphodiesterase-4 (PDE4) inhibitor, in AD by using bilateral Aβ25-35 injections into the hippocampus of rats (Fig. 16). Different doses of rolipram were administered daily for 25 days after Aβ 25-35 injections. The results of the study demonstrated that rolipram significantly inhibited NO and iNOS pathways in the hippocampus. Furthermore, rolipram improved memory and learning abilities in the Aβ25-35-induced AD rat model [53].



28 **29**

Figure 16. Other synthetic compounds as inhibitors of iNOS.

2.4 QSAR Study

Quantitative structure-activity relationship (QSAR) is commonly used in silico-modeling methods, that link the physicochemical properties of chemical compounds to their biological

activity. QSAR models are developed using statistical and mathematical techniques, which is considered a crucial strategy in the drug discovery process, as it enables the prediction of the activity of new compounds based on their structural and physicochemical properties. In addition, the development of QSAR models has significantly improved the efficiency of the drug discovery process by reducing the time and cost involved in the identification of lead compounds [54, 55].

The development of a QSAR model offers several significant advantages in the field of pharmaceutical research. Firstly, it aids in the identification and optimization of lead compounds, streamlining the drug discovery process. Additionally, it assists in prioritizing chemicals, allowing researchers to focus on the most promising candidates. This approach is both economically and time-saving, as it minimizes the need for extensive and costly laboratory experiments. Moreover, QSAR models contribute to the reduction of animal experimentation by providing valuable insights into chemical behavior. Furthermore, they facilitate risk assessment of chemicals, enhancing the safety profile of potential drug candidates. Finally, QSAR models play a crucial role in exploring the mechanism of action, providing a deeper understanding of how compounds interact with their biological targets.

2.4.1 QSAR Modelling Approaches

There are two QSAR modeling approaches; classification model and regression model: classification model is used when a chemical compound has categorical biological activity against a certain target, which means that it's either active or inactive against a certain biological target. The aim of the classification model is to categorize the selected chemical compound into two or more categories based on their calculated molecular descriptors [56].

On the other hand, regression model is used when the biological activity of a chemical compound is continuous and when it can take on a range of values. The aim of a regression model is to predict the numerical value of the biological activity of a chemical compound based on its molecular descriptors [56].

2.5 Machine Learning

Machine learning (ML), a field within artificial intelligence (AI), involves the development of algorithms that enable computers to learn from data without the need for explicit programming. The purpose of machine learning is to enable machines to make decisions, and predictions, or take actions based on patterns or relationships that they learn from data [57]. In machine learning, a model is developed using a set of training data that contains input features and output values. The model learns to recognize patterns in the input features and predicts the output value for new, unseen data. The performance of the model is evaluated using a set of test data that contains input features and actual output values. The model is trained to minimize the difference between its predicted output values and the actual output values. Additionally, ML is subdivided into supervised machine learning and unsupervised machine learning [58].

2.5.1 Supervised Machine Learning

Supervised learning is a learning method based on the supervision of both input and output which maps input data into output data to develop predictive models using the algorithm Y = f(x), in other notation, output = f (input). Computer learns from labelled data by supervised learning. It uses classification and regression techniques to develop ML models [57]. The key points of the ML model developed under supervision are a selection of features, selection of algorithm and the quality of training set [59]. Supervised learning is a type of task-driven approach to predict a category or continuous numeric features. With the help of supervised learning, two types of ML models can be built. They are the classification model and regression model

2.5.1.1 Classification Model

The classification model predicts the different class labels in the data. If output variable or experimental data is categorical such as, active-inactive, or true-false, classification algorithms are used to build an ML classification model. In classification, the supervised task is to separate data into categories. In ML, the most common and popular classification

algorithms are Random Forest (RF), Naive Bayes (NB), Logistic Regression (LR), and Decision Tree (DT) [60].

2.5.1.2 Regression Model

Regression models predict continuous quantities [60]. When output variable or experimental data is continuous numerical, regression algorithms are used to build an ML regression model. The regression model is used in numerous fields such as prediction of drug response modelling, trend analysis, cost estimation, financial forecasting, and marketing. Some of the common regression algorithms are polynomial regression, simple or multiple linear regression, ridge regression, LASSO regression [55, 58].

2.5.2 Unsupervised Machine Learning

Unsupervised learning is a learning method based on only input data in order to group and interpret the data. Computer learns from unlabeled data without the need for human intervention in unsupervised learning. Unsupervised learning is a type of data-driven approach for discovering the data structure. It can be divided into groups such as clustering, pattern/structure recognition, dimension reduction, and feature learning [60].

2.5.2.1 Clustering

Clustering, an unsupervised learning method, is a kind of data analysis technique for identifying and grouping data according to their similarities, such in some sense more similar to each other than objects in other groups. In ML and data science, the most popular clustering algorithms are mean-shift clustering, agglomerative clustering, and k-means clustering [60].

2.5.2.2 Kohenen network

Kohonen network clustering, also known as self-organizing maps (SOM), is an unsupervised machine learning technique used for clustering and visualization of high-dimensional data. The algorithm involves mapping high-dimensional data onto a low-dimensional grid of

nodes, where each node represents a cluster. The nodes are organized in such a way that nearby nodes in the grid represent similar patterns in the input data. The Kohonen network clustering algorithm allows for exploratory data analysis, dimensionality reduction, and visualization of complex data patterns, making it useful for a wide range of applications [61, 62]

3. MATERIALS AND METHOD

The process of QSAR modeling consists of several steps, including data collection, calculation of appropriate molecular descriptors based on chemical structures, model development, and validation (Fig. 17). Multiple software's were used to support the QSAR modeling study [56].

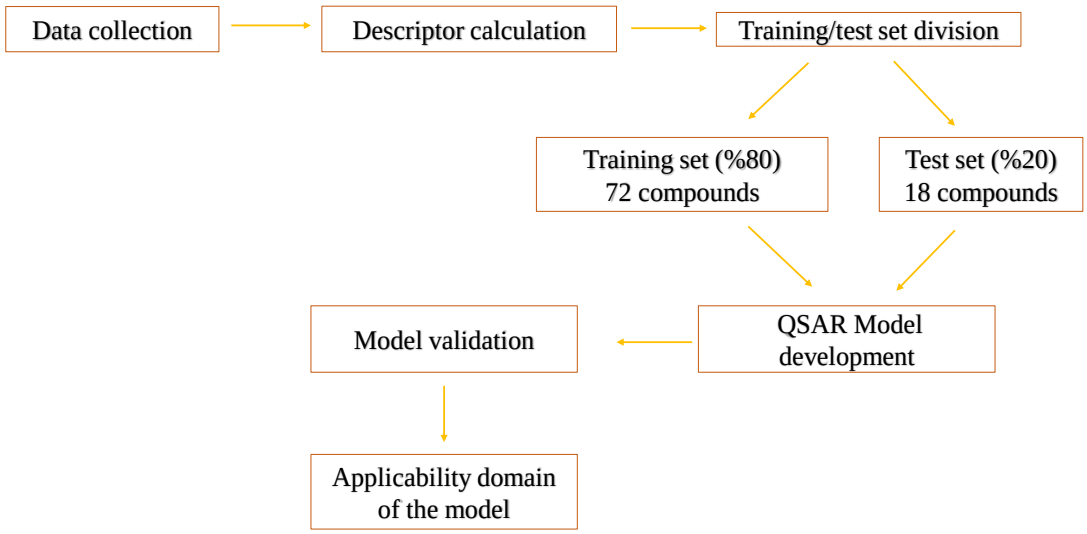


Figure 17. The flowchart of QSAR modeling.

3.1 Data Collection

In the initial phase, we collected 90 synthetic iNOS inhibitors from various literature sources. These inhibitors belonged to specifically [(4-(1,2,4-oxadiazol-5-yl)phenyl]-2-amino-acetamide derivatives, 7-O-1, 2, 3-triazole hesperetin derivatives, and 3-(4-pyridyl)-5-(4-sulfamidophenyl)-1,2,4-oxadiazole derivatives given in Appendix Table_S1 [7, 49, 50]. To ensure data consistency, we prioritized studies with consistent test procedures and the use of a common reference compound resveratrol. We then compiled a table containing these selected inhibitors along with their respective IC_{50} values for iNOS inhibition. The data set was further classified into active and inactive compounds based on the median IC_{50} value of the data set.

3.2 Drawing Compound

The iNOS inhibitors were drawn by using ACD/Chem sketch (v.14.54.1.64577) [63]. The drawn chemicals were saved as MDL sdf files and SMILES files given in Appendix Table_S2 for descriptor calculation.

3.3 Descriptor Calculation

Molecular descriptors were calculated using two software programs: Spartan'20 [64], and alvaDesc (v.4.2.1) [65]. Spartan was initially employed to minimize each structure with a two-step calculation: conformer distribution with molecular mechanics and then equilibrium geometry at ground state in gas with semi-empirical PM3. Gemometry-optimized structures were used to obtain quantum-chemical descriptors: the highest occupied molecular orbital energy ((E HOMO (eV)) and the lowest unoccupied molecular orbital energy ((E LUMO (eV)), dipole moment (debye), volume, and area. The other descriptors were also added to the descriptor pool using the calculated Spartan descriptors: GAP (ELUMO-EHOMO), Hardness (GAP/2), Softness (1/Hardness), Electronegativity (-(EHOMO+ELUMO)/2), and Electrophilicity (Electronegativity2/2*Hardness). Finally, alvaDesc was utilized to calculate Log P and other molecular descriptors.

3.4 Training and Test Set Division

We partitioned the dataset into a training set and a test set. To create a representative test set, firstly, we made systematic sampling to select the test set based on the IC_{50} activity values of the compounds. Furthermore, we utilized Kohonen networks in MATLAB R2022b (The MathWorks, Inc.) to cluster compounds to identify groupings in the data set considering their structural similarities [66]. During training, each data point in the data set was presented to the network, and the network calculated its distance from each neuron in the topological map. The neuron with the closest distance to each data point was designated as the "winner," and its weights were adjusted to better represent that specific data point. This process was repeated for all data points in the data set [67].

This clustering facilitated a deeper understanding of the compound landscape. To enhance the representativeness of the test set, we strategically adjusted its composition based on clustering results, ensuring balanced representation of clusters with varying compound activities.

3.5 The Selection of Descriptors for QSAR Models

The selection of descriptors for the QSAR study was carried out using WEKA v.3.8.6 software [68]. The descriptor selection process consisted of two main parts: an attribute evaluator and a search method. In this particular study, WrapperSubsetEval with J48 classifier was employed as the attribute evaluator, and BestFirst was used as the search method to identify the most important descriptors from a pool of 2940 descriptors based on learning-based feature selection (Table 1). WrapperSubsetEval assesses attribute sets using a learning scheme, analyzes attribute sets using a classifier, and employs cross-validation to calculate the accuracy of the learning scheme for each set [69]. The J48 classifier, which applies the C4.5 algorithm, was used in the attribute evaluation to build the decision tree in WEKA, selecting the attribute with the highest normalized information gain. Additionally, five significant descriptors given in Appendix Table_S3 were selected for developing the QSAR model.

Table 1. Setting parameters of learning-based feature selection method.

Attribute Evaluator	WrapperSubSetEval
Classification Algorithm	J48 classifier
Search Strategy	BestFirst (Direction:Forward)
SubSet Evaluation	Classification Acuracy
Accuracy Estimation	5-fold cross validation
Learning scheme	Weka. classifiers. trees. J48
Scheme options	C 0.25 -M 2

3.6 Model Development

Various machine-learning classifiers were tested using WEKA software to classify compounds as active or inactive. The performance of these classifiers was evaluated using metrics such as classification accuracy (ACC), Matthews' correlation coefficient (MCC), and Receiver Operating Characteristic (ROC). Random forest (RF), Kstar, and IBK emerged as the top-performing classifiers. The training set chemicals exclusively were used to train the models. Subsequently, internal and external validation procedures were carried out on the developed models as described in the following subsection.

3.7 Model Validation

Model validation was performed according to OCED principles through internal and external validation. Internal validation was conducted through 10-fold cross-validation. In this process, the training set data was divided into ten equal parts. One part (10%) was set aside as the test set, while the remaining served as the training set. This process was repeated ten times, creating ten different models that were evaluated and trained. The final accuracy of the validated model was determined by averaging the accuracy results from the ten iterations. The prediction ability of the classification model was assessed via external validation [70, 71]. External validation involved the test set, with an equal representation of active and inactive compounds. The model was utilized to predict the classes of each compound in the test set correctly.

In addition, both specificity and sensitivity for both the cross-validated model and the test datasets have been calculated. This process of calculating sensitivity and specificity is vital as it provides valuable insights into the model's performance and reliability. Sensitivity, or the true positive rate, evaluates the model's competence in correctly recognizing positive instances, ensuring the effective detection of relevant factors or compounds. Specificity, representing the true negative rate, gauges the model's proficiency in accurately classifying negative instances, thus minimizing false alarms. A balanced model with high sensitivity and specificity ensures both the identification of potential candidates and the avoidance of false conclusions, enhancing the model's utility and trustworthiness in critical applications [72].

3.8 Applicability Domain

The determination of the applicability domain is essential for ensuring reliable and accurate predictions, as well as identifying the specific type of chemical compounds that can be accurately predicted by the model [54]. The applicability domain was established based on molecular descriptors, allowing for valid and reliable predictions within the molecular descriptor space covered by the compounds in the training set. As a result, by identifying outliers, it becomes possible to define the boundaries of the applicability domain [73].
The boxplot method was employed to identify the outliers of the selected descriptors by using MS Excel software [74]. For each descriptor in the training and test sets, the minimum and maximum values of the descriptor were determined. Then, the interquartile range (IQR) of each descriptor was calculated. Furthermore, the outliers of each descriptor in the training and test sets were determined by calculating the lower and upper outlier limits.

3.9 Drug-likeness properties

Drug-likeness refers to the set of physicochemical and structural characteristics that make a compound more likely to be developed into a successful drug candidate. It involves assessing the molecule's properties, and adherence to certain rules like Lipinski's Rule of Five. In this study, drug-likeness properties were performed for the data-set compounds using established guidelines, including Lipinski's Rule of Five. The analysis focused on molecular weight, LogP, hydrogen bond donors and acceptors, all pivotal factors in predicting oral bioavailability. Furthermore, the ADMET properties related to absorption, distribution, metabolism, and excretion of the selected compounds were identified utilizing pkCSM software. This computational tool enabled the prediction and evaluation of these crucial pharmacokinetic properties, offering insights into the compounds' behavior within biological systems and their potential suitability for drug development.

4. RESULTS

4.1 The Training and Test Sets

Dividing the dataset into training and test sets is a pivotal step in developing a robust QSAR model. To achieve this, we followed a systematic approach. Initially, we first sorted the IC50 activity values of the compounds, and allocated every fifth compound, starting from the second compound in the sorted list. This process ensured that approximately 20% of the compounds were reserved for testing.

Subsequently, we employed a Kohonen network to analyze and cluster the data set based on the molecular descriptors. The network was configured with a 4x4 square map, normal bound, no scaling, and underwent 100 epochs of training. The primary objective here was to identify patterns and groupings within the dataset, particularly focusing on similarities in the structural profiles of compounds, ensuring a balanced representation of both active and inactive compounds (Fig. 18).

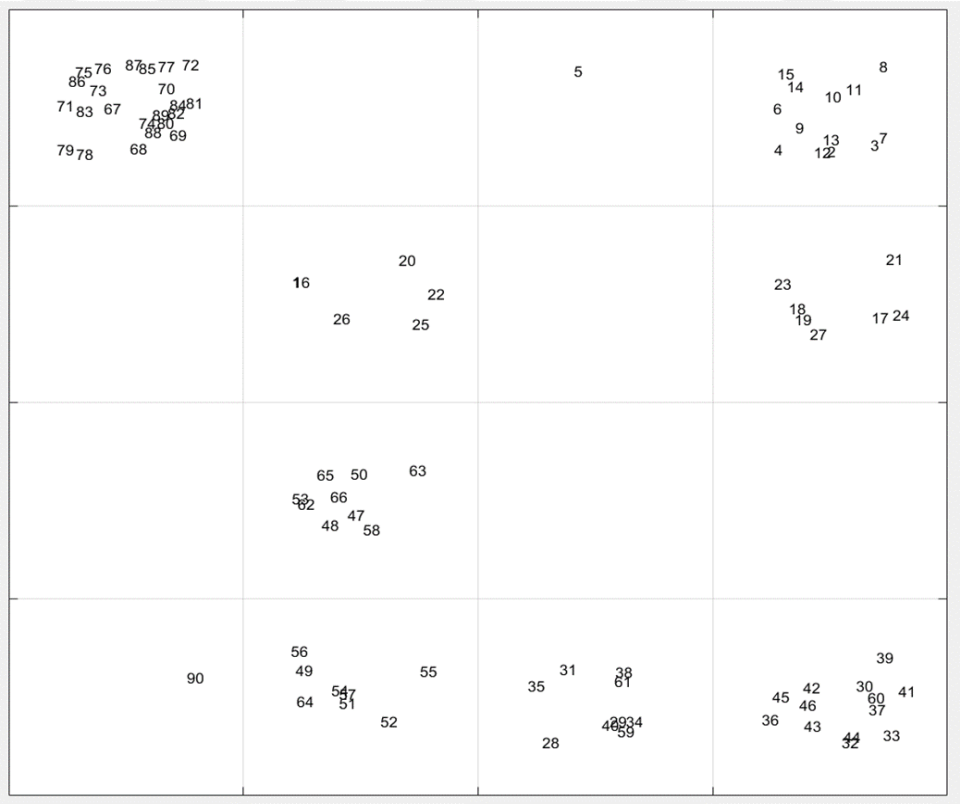


Figure 18. Kohonen network map.

After dividing the dataset, the Kohonen network is trained using the training set. During training, each data point in the training set is presented to the network, and its distance from each neuron in the top-map is calculated. The neuron with the closest distance to each data point is selected as its "winner", and its weights are adjusted to better represent that data point. Once training is complete, the performance of the Kohonen network on both sets is evaluated using numerous evaluation metrics such as precision, accuracy, F1 score, recall. These metrics measure how well the algorithm has clustered similar data points together and separated dissimilar ones. In order to better represent clusters with a significant number of compounds, we decided to change some of the test set compounds according to the distribution on the map. Finally, 20% of the compounds were allocated to the test set, and the remaining compounds were used in the training set.

4.2 Division of the Compounds as Active and Inactive for QSAR Model Development

Following the clustering process using the Kohnen network, the molecules within each cluster were assigned to the training and test sets, approximately in an 80%-20% ratio. Then in terms of iNOS inhibitory activity for AD, the compounds were distributed equally to active and inactive groups according to their IC_{50} values. The allocation of the 90 chemicals into the training and test sets, as well as the distribution of their classes as active and inactive, are given in Table 2.

Table 2. Distribution of the dataset to Active and Inactive in both training set and test set.

	Active	**Inactive**	**Total**
Training set	36	36	72
Test set	9	9	18
Total	45	45	90

4.3 Descriptor selection

In the realm of QSAR studies, feature selection is a pivotal step for selecting the best descriptors in developing effective classification models. In this study, we employed WEKA software for learning-based feature selection [75].

4.3.1 Descriptor Selection

Selecting the best attributes (descriptors) by WEKA software has been done based on Learning-based feature selection via WrapperSubsetEval with J48 classifier as an attribute evaluator and BestFirst as the search method. As a result, five descriptors which are (MATS6p, Chi1_EA(dm), Mor17s, NsssCH, and SHED_AL) given in Appendix TableS.1 were selected as the best descriptor among 2940 descriptors by WEKA for the development of QSAR models.

4.3.2 Comparison of Classifiers

The performance of different classifiers was assessed using our QSAR-ready dataset in WEKA. During preliminary tests, RF, IBK, and Kstar emerged as the top three classifiers. Consequently, using these classifiers, a total of three models were developed. Table 3 displays the evaluation results and performances of the top three classifiers for each model.

Table 3. The performance of the top three classifiers for each model.

Classifier	Set	ACC (%)	AUC	MCC	KAPPA	ROC	TP rate
KStar	Train	100	1.00	1.00	1.00	1.00	1.00
	CV (10-fold)	68.8	0.80	0.37	0.37	0.80	0.64
	Test	66.6	0.87	0.31	0.30	0.87	0.66
RF	Train	100	1.00	1.00	1.00	1.00	1.00
	CV (10-fold)	76.6	0.80	0.53	0.53	0.80	0.76
	Test	88.8	0.88	0.80	0.78	0.88	0.88
IBK	Train	100	1.00	1.00	1.00	1.00	1.00
	CV (10-fold)	77.7	0.77	0.55	0.55	0.77	0.77
	Test	83.3	0.81	0.69	0.64	0.81	0.83

All results are given in " weighted average".

Abbreviations: CV: cross-validation, ACC: accuracy, AUC: area under the ROC curve, MCC: Matthews' correlation coefficient.

4.4 Applicability Domain for Classification Model

The applicability domain serves as a critical boundary that defines the range of chemical compounds for which the model's predictions remain reliable. The border of AD was determined using the alvaDesc model with five descriptors, which was found to be the best model in the study.

To determine the applicability domain, we utilized boxplots, a standardized method for visualizing data distribution by quartiles. This graphical representation helps us gain insights into how values are distributed within our datasets. Specifically, we employed the 1.5 * IQR (Interquartile Range) rule, where any data point falling below the first quartile minus 1.5 times the IQR or above the third quartile plus 1.5 times the IQR is considered a mild outlier or a potential outlier.

In our analysis, six compounds in the training set were identified as outliers across five descriptors, while one compound in the test set fell outside the applicability domain boundary. To provide a visual representation of these findings, we included boxplots for each descriptor (MATS6p, Chi1_EA(dm), Mor17s, NsssCH, and SHED_AL) for both the training and test sets (Fig. 19-23).

Descriptor-specific outliers can be listed as follows:

- MATS6p Descriptor (Fig. 19): The boxplot analysis for MATS6p revealed four outliers in the training data (compounds 7, 12, 14, and 19), whereas in the test data, only one compound (compound 14) was found outside the applicability domain.
- Chi1_EA(dm) Descriptor (Fig. 20): The boxplot for Chi1_EA(dm) identified one outlier (compound 90) in the training data, with no outliers detected in the test data.
- SHED_AL Descriptor (Fig. 23): Analysis of the SHED_AL descriptor indicated two outliers (compounds 47 and 90) in the training set.
- Mor17s and NsssCH Descriptors: These descriptors did not reveal any outliers in either the training or test set data.

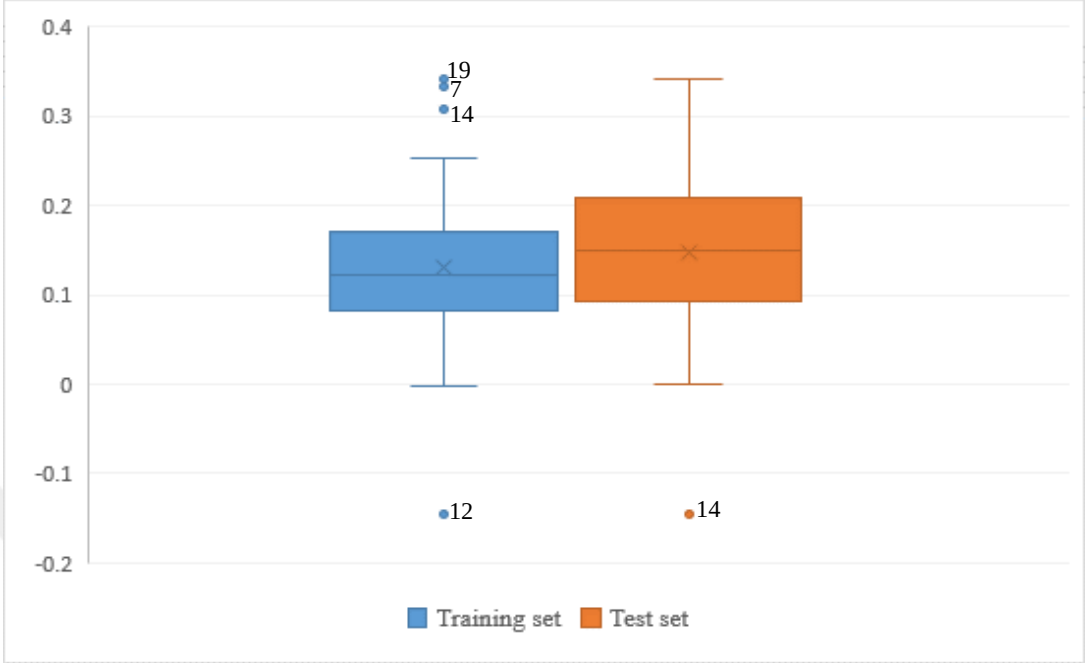


Figure 19. MATS6p descriptor boxplot identified outliers in the training data (7, 12, 14, and 19), and in the test set data, only one compound (14) fell outside the applicability domain.

Figure 20. Boxplot graph of Chi1_EA(dm) descriptor showed one outlier which is compound 90 in the training data.

Figure 21. Boxplot graph of Mor17s descriptor did not show outliers in both the training and test set data.

Figure 22. Boxplot graph of NsssCH descriptor did not show outliers in both the training and test set data.

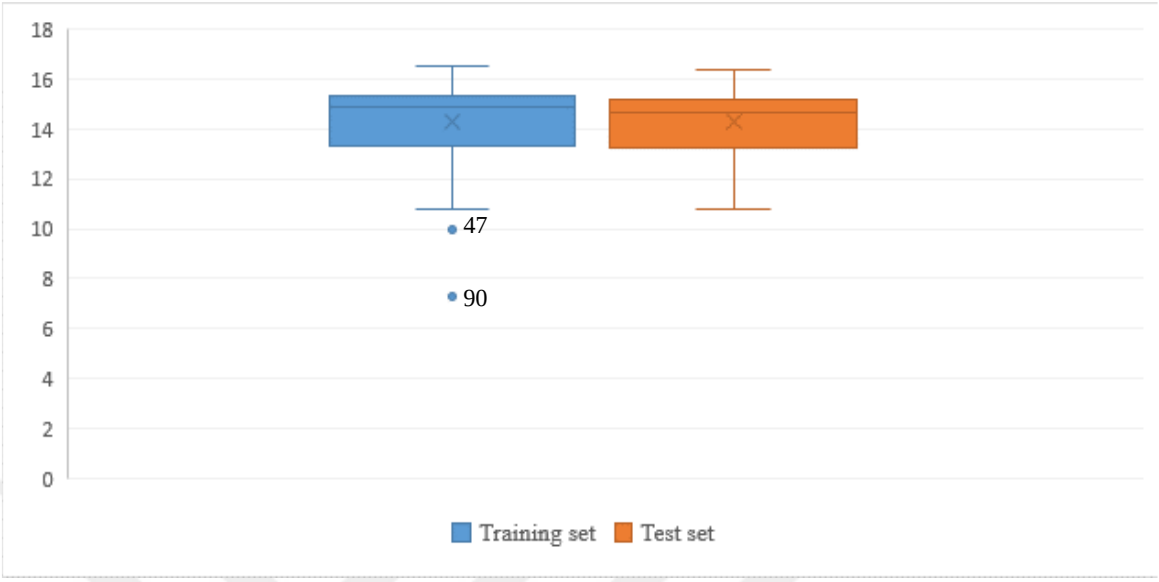


Figure 23. Boxplot graph of SHED_AL descriptor detected two outliers (47, 90) in the training set data.

4.5 Comperative analysis with existing QSAR models

In this study, we conducted a comprehensive evaluation of our classification model, designed to predict the inhibitory activity of selected compounds in AD, specifically targeting the iNOS enzyme. To assess the performance and uniqueness of our model, we compared it with existing QSAR models, recognizing that there is a scarcity of classification models explicitly tailored for iNOS inhibition. Instead, we turn our attention to regression QSAR models, a widely established approach in cheminformatics and drug discovery. These regression models establish quantitative relationships between chemical compound structures and their associated biological or chemical activity levels, offering continuous numerical predictions for factors such as biological activity or property measurements, as opposed to classification models.

To further assess our classification model's performance relative to other QSAR models, we evaluated its prediction accuracy. Our results revealed an impressive accuracy rate of 88.8% in predicting the inhibitory activity of the selected compounds.

4.6 Drug-likeness and ADMET profile

In the study, it was found that thirty-nine compounds out of the ninety compounds violated more than one of Lipinski's rules, suggesting potential challenges in terms of oral bioavailability. However, fifty-one compounds adhered to Lipinski's rules, demonstrating suitability for solubility in aqueous medium shown in table 4.

Assessing the ADMET properties using the pkCSM website, is pivotal in understanding the pharmacokinetic behavior of compounds. Within this study, the absorption, distribution, metabolism, excretion, and toxicity profiles of the chosen compound were precisely delineated and provided in Supplementary File Table S1. This thorough analysis offers valuable insights into how the compound behaves in terms of absorption, distribution, metabolism, excretion, and potential toxicity.

Table 4. Drug-likeness properties of the compounds that did not violated more than one lipinski rule.

Compounds Number	MW	log P	RB	HBA	HBD
28	473.4	4.031	7	9	2
29	487.5	4.340	7	9	2
30	491.4	4.170	7	9	2
31	507.9	4.685	7	9	2
32	757.4	3.206	7	9	2
34	487.5	4.340	7	9	2
35	507.9	4.685	7	9	2
36	757.4	3.206	7	9	2
38	487.5	4.340	7	9	2
39	491.4	4.170	7	9	2
40	507.9	4.685	7	9	2
42	501.4	3.866	8	10	2
43	519.4	4.005	8	10	2
44	535.9	4.519	8	10	2
46	519.4	4.005	8	10	2
47	397.3	2.520	5	9	2
48	411.4	3.003	6	9	2
49	425.4	3.393	7	9	2
50	409.3	1.915	5	9	2
51	439.4	3.783	8	9	2

52	437.4	3.492	6	9	2
53	437.4	3.708	6	9	2
54	451.4	4.098	6	9	2
55	465.5	4.488	6	9	2
56	427.4	1.975	7	10	3
57	441.4	2.365	8	10	3
58	474.4	3.968	6	10	3
59	488.5	4.276	6	10	3
61	508.9	4.621	6	10	3
62	440.4	3.389	7	10	3
63	426.4	2.998	6	10	3
64	454.4	3.779	8	10	3
65	452.4	3.488	6	10	3
66	466.4	4.094	6	10	3
67	436.3	3.711	9	8	3
68	450.4	4.102	10	8	3
69	450.4	3.957	9	8	3
70	434.3	3.464	8	8	3
71	462.4	4.244	8	8	3
72	470.4	4.824	8	8	3
73	500.4	4.833	9	9	3
74	500.4	4.833	9	9	3
75	488.4	4.963	8	8	3
76	488.4	4.963	8	8	3
82	484.4	5.132	8	8	3
83	484.4	5.132	8	8	3
84	484.4	5.132	8	8	3
85	484.4	4.502	9	8	3
86	502.4	4.641	9	8	3
88	502.4	4.641	9	8	3
90	228.2	2.973	2	3	3

5. DISSCUSION AND CONCLUSION

The assumption in all QSAR studies is that similar compounds have comparable biological activities. To create an effective predictive QSAR model, the test set points must resemble those in the training set in terms of molecular representation in the descriptor space [76, 77]. In this study, we clustered the data sets using the Kohonen network. After we had done clustering by the Kohnen network, the molecules in each cluster were distributed to the training and test sets by keeping the ratio approximately 80%-20%. Specifically, we ensured that both active and inactive compounds were evenly distributed within each set. As a result, the final composition of the training and test sets consisted of 72 compounds in the training set, evenly divided between 36 active and 36 inactive compounds. The test set comprised 18 compounds, with an equal representation of 9 active and 9 inactive compounds, maintaining a balanced 1:1 ratio for both active and inactive compounds.

Following the division of training and test sets, we performed feature selection, a crucial step in identifying the most pertinent descriptors for building a robust QSAR model using WEKA [78]. We evaluated two approaches: the correlation-based method (CfsSubSet) and the learning-based method (Wrapper with the J48 classifier) with the J48 decision tree classifier [79].

Correlation-based Method (CfsSubSet): This approach assesses attribute subsets that exhibit high correlation with the class while minimizing intercorrelations among themselves. While it offers a valuable perspective, the learning-based method proved to be more interpretable and efficient for our dataset.

Learning-based Method (Wrapper with J48 Classifier): The learning-based feature selection, employing the J48 decision tree classifier, evaluated attribute subsets and utilized cross-validation to estimate the classification accuracy. This approach resulted in a simpler and more interpretable model compared to the correlation-based method.

The learning-based feature selection method within WEKA identified the top descriptors from a pool of 2940 descriptors given in Appendix Table_S2. The selected descriptors for building our QSAR model, are as follows:

MATS6p (Mean Atomic Weights for Specific Property): MATS6p provides insights into the weighted average atomic characteristics of a molecule, aiding in predicting biological or chemical activities based on atomic properties.

Chi1_EA(dm) (Chi1 Clustera and EA (Element and Atom Count)): This topological descriptor captures the structural attributes of a molecule by considering specific element and atom counts, which are vital for understanding its chemical and biological behavior.

Mor17s (Moran Autocorrelation): Mor17s quantifies the spatial arrangement and distribution of atoms within a molecule, offering insights into spatial relationships among atoms.

NsssCH (Number of Non-Hydrogen Single Bonds): NsssCH represents the count of non-hydrogen single bonds in a molecule, providing information about the molecule's connectivity and stability.

SHED_AL (Sum of Heaviside Function Values for Atoms): SHED_AL computes the summation of Heaviside function values for atoms within a molecule, offering insights into the molecule's shape and size.

Another crucial step in the QSAR model development by WEKA is the selection of the best classifier for the model development. To evaluate the performance of different classifiers, we utilized our QSAR-ready dataset in WEKA, leading to the identification of three top-performing classifiers: Kstar, RF, and IBK. Subsequently, three models were developed using these classifiers.

Table 3 provides a comprehensive overview of the evaluation results and performance metrics for these top three classifiers. Random Forest (RF) demonstrated the highest predictive ability, achieving an accuracy of 88.88%. This result aligns with the common preference for RF in QSAR studies due to its strong discriminatory performance and high accuracy in classification models [80-82]. Moreover, RF exhibited superior performance in various evaluation metrics, including AUC, MCC, KAPPA, and Area under ROC, both in cross-validated models and test sets, outperforming Kstar and IBK. This highlights RF's efficacy in predicting the target variable accurately. Consistent with our findings, RF exhibited the highest predictive ability in our study, achieving an impressive accuracy of 88.88%. Conversely, IBK demonstrated an accuracy of 83.33%, while Kstar showed an accuracy of 66.66%. Furthermore, in terms of cross-validated results, RF displayed better

performance in terms of AUC and ACC compared to the other classifiers, indicating its superior discriminatory power and accuracy in predicting the target variable, as shown in Table 3, moreover, RF showed the best performance compared to Kstar and IBK when considering the other evaluation metrics such as (ACC, MCC, KAPPA, and Area under ROC) for both cross-validated models and test sets, developed by the top three classifiers [83]. We, therefore used RF to develop the QSAR model to predict iNOS inhibition of selected compounds used in the treatment of AD.

Our QSAR models underwent both internal and external validation in accordance with OCED principles [70, 71]. For internal validation, we performed 10-fold cross-validation, as detailed in Table 3. External validation involved testing the model's predictive ability on an independent test set composed of an equal distribution of active and inactive compounds. The model correctly predicted the classes of the test set compounds with an accuracy of 88.88%. Additionally, the sensitivity and specificity results of our QSAR classification model are noteworthy. The sensitivity, which measures the ability of the model to correctly identify positive cases, is robust, with a value of 0.733 for the cross-validated model and a value of 1 for the test dataset, indicating excellent performance in detecting true positives. Additionally, the model exhibited commendable specificity scores of 0.8 for both cross-validation and the test set, underscoring its ability to correctly identify true negatives. These results collectively highlight the model's proficiency in distinguishing between positive and negative cases in both internal and external validation.

Furthermore, the evaluation of the applicability domain is another crucial step of a QSAR study, and it is a important way to ensure that the model predictions are reliable and valid [84]. In this step, we discuss the establishment of the applicability domain for our QSAR model, which utilized five descriptors, identified as the best model in our study. The applicability domain serves as a critical boundary that defines the range of chemical compounds for which the model's predictions remain reliable.

To determine the applicability domain, we utilized boxplots, a standardized method for visualizing data distribution by quartiles. This graphical representation helps us gain insights into how values are distributed within our datasets. Specifically, we employed the 1.5 * IQR

(Interquartile Range) rule, where any data point falling below the first quartile minus 1.5 times the IQR or above the third quartile plus 1.5 times the IQR is considered a mild outlier or a potential outlier.

It's important to note that our model was constructed exclusively for oxadiazole derivatives and triazole derivatives to predict their inhibitory activities. Consequently, any compounds falling outside this defined scope may be considered outliers and require special attention. The outliers identified in our model primarily belong to the category of structural outliers. This classification arises from the fact that these outliers exhibit distinct and often unique structural characteristics that distinguish them from the rest of the data points. It is worth mentioning that these outliers, while deviating from the typical data pattern, are not extreme outliers; they fall within the category of potential outliers, as they lie beyond the boundaries defined by the 1.5 * IQR rule. In our analysis, six compounds in the training set were identified as outliers across five descriptors, while one compound in the test set fell outside the applicability domain boundary.

In addition, to assess the performance and uniqueness of our model, we compared it with existing QSAR models, recognizing that there is a scarcity of classification models explicitly tailored for iNOS inhibition. Instead, we turn our attention to regression QSAR models, a widely established approach in cheminformatics and drug discovery. These regression models establish quantitative relationships between chemical compound structures and their associated biological or chemical activity levels, offering continuous numerical predictions for factors such as biological activity or property measurements, as opposed to classification models.

An initial point of comparison pertains to the number of descriptors employed in model development. Our classification model employs a streamlined selection of five descriptors, prioritizing those demanding minimal computational resources. Contrarily, a previous study conducted by Wei et al. employed a larger set of descriptors for constructing their regression QSAR model [85]. This contrast underscores the efficiency and conciseness of our classification model in terms of descriptor selection. To further assess our classification model's performance relative to other QSAR models, we evaluated its prediction accuracy. Our results revealed an impressive accuracy rate of 88.8% in predicting the inhibitory activity

of the selected compounds. This outcome establishes our classification model as superior in terms of predictive accuracy when compared to other QSAR models.

Finally, in this study the drug-likeness and ADMET properties of the compounds were defined using pkCSM software, which was found that thirty-nine compounds out of the ninety compounds violated more than one of Lipinski's rules, suggesting potential challenges in terms of oral bioavailability. However, fifty-one compounds adhered to Lipinski's rules, demonstrating suitability for solubility in aqueous medium shown in table 4. Consequently, these compounds exhibit favorable drug-likeness properties, positioning them as promising candidates for further consideration as oral drug candidates. Assessing the ADMET properties using the pkCSM website, is pivotal in understanding the pharmacokinetic behavior of compounds. In this study, the fifty-one compounds displayed impressive oral absorption capabilities, efficiently entering the systemic circulation after intestinal absorption. Notably, compounds **17-19, 21-27** demonstrated the ability to reach the central nervous system, potentially affecting therapeutic outcomes. On the other hand, understanding metabolic pathways is crucial; notably, compounds **42**, **47-52**, **56**, **57**, **59**, **61-65**, **67-71**, **90** were identified as non-CYP3A4 substrates, implying reduced susceptibility to metabolism via the liver's CYP3A4 enzyme. Nonetheless, a majority of the compounds underwent metabolism via this enzyme, necessitating caution due to potential drug interactions with other CYP3A4-metabolized drugs. Interestingly, none of the compounds were renal OCT2 substrates, suggesting that renal clearance may not play a significant role in their elimination from the body, Supplementary File Table S1.

Additionally, evaluating toxicity profiles is critical for safety assessment. Analysis using the pkCSM website [49] revealed that compounds **32**, **36**, **42-47**, **50**, **51**, **60**, **62**, **90** exhibited no toxic effects, whereas other compounds demonstrated hepatotoxic effects. This comprehensive understanding of ADMET properties and toxicity evaluations is fundamental for determining the safety and efficacy of the identified compounds in potential therapeutic applications.

In summary, our study introduces an innovative and highly effective classification model tailored for predicting iNOS inhibitory activity in AD. Notably, this model utilizes a minimal number of descriptors, optimizing computational efficiency while achieving exceptional prediction accuracy. These findings underscore the potential of our classification model as a valuable asset in drug discovery and development efforts aimed at targeting iNOS inhibition in AD treatment.

In conclusion, we have successfully developed a QSAR model utilizing alvaDesc descriptors to predict the iNOS inhibition activity of selected compounds crucial in Alzheimer's disease treatment. Our model exhibits commendable predictive accuracy, achieving an accuracy rate of 76.67% for the cross-validated model and an even more impressive 88.88% accuracy for the test set. These results highlight the model's robustness and reliability in forecasting the inhibitory activity of these compounds, underlining its potential as a valuable tool in Alzheimer's disease research and drug development.

Moreover, this study represents a pivotal initial step in advancing our ability to predict iNOS inhibitory activity for compounds relevant to AD. The future promises even greater strides in this direction as we augment our model with additional data and harness the evolving capabilities of in silico modeling methods. By doing so, we can pave the way for the development of non-animal-based alternative models, addressing concerns related to animal rights and ethical considerations. These emerging models will provide a more compassionate and humane approach to safety evaluation for iNOS inhibitors in the context of Alzheimer's disease, aligning with contemporary values and ethical standards in biomedical research.

6. REFERENCES

[1] Lane CA, Hardy J, Schott JM. Alzheimer's disease. *European Journal of Neurology*. 2017;25(1):59-70. doi:10.1111/ene.13439

[2] Li L, He R, Yan H, Leng Z, Zhu S, Gu Z. Nanotechnology for the diagnosis and treatment of alzheimer's disease: A Bibliometric analysis. *Nano Today*. 2022;47:101654. doi:10.1016/j.nantod.2022.101654

[3] De-Paula VJ, Radanovic M, Diniz BS, Forlenza OV. Alzheimer's disease. *Protein Aggregation and Fibrillogenesis in Cerebral and Systemic Amyloid Disease*. Published online 2012:329-352. doi:10.1007/978-94-007-5416-4_14

[4] De Loof A, Schoofs L. Alzheimer's disease: Is a dysfunctional mevalonate biosynthetic pathway the master-inducer of deleterious changes in cell physiology? *OBM Neurobiology*. 2019;3(4):1-1. doi:10.21926/obm.neurobiol.1904046

[5] Kumar V, Saha A, Roy K. In silico modeling for dual inhibition of acetylcholinesterase (ache) and butyrylcholinesterase (Buche) enzymes in alzheimer's disease. *Computational Biology and Chemistry*. 2020;88:107355. doi:10.1016/j.compbiolchem.2020.107355

[6] Yiannopoulou KG, Anastasiou AI, Zachariou V, Pelidou S-H. Reasons for failed trials of disease-modifying treatments for alzheimer disease and their contribution in recent research. *Biomedicines*. 2019;7(4):97. doi:10.3390/biomedicines7040097

[7] Wang M, Fang L, Liu T, et al. Discovery of 7-O-1, 2, 3-triazole hesperetin derivatives as multi-target-directed ligands against alzheimer's disease. *Chemico-Biological Interactions*. 2021;342:109489. doi:10.1016/j.cbi.2021.109489

[8] de Mendonça LJ, Ferrari RJ. Alzheimer's disease classification based on graph kernel svms constructed with 3D texture features extracted from Mr Images. *Expert Systems with Applications*. 2023;211:118633. doi:10.1016/j.eswa.2022.118633

[9] Liu Y, Mazumdar S, Bath PA. An unsupervised learning approach to diagnosing alzheimer's disease using brain magnetic resonance imaging scans. *International Journal of Medical Informatics*. 2023;173:105027. doi:10.1016/j.ijmedinf.2023.105027

[10] Zhang L, Hou S, Movahedi F, et al. Amyloid-β/tau burden and neuroinflammation dual-targeted nanomedicines synergistically restore memory and recognition of alzheimer's disease mice. *Nano Today*. 2023;49:101788. doi:10.1016/j.nantod.2023.101788

[11] Förstl H, Kurz A. Clinical features of alzheimer's disease. *European Archives of Psychiatry and Clinical Neuroscience*. 1999;249(6):288-290. doi:10.1007/s004060050101

[12] Almkvist O. Neuropsychological features of early alzheimer's disease: Preclinical and clinical stages. *Acta Neurologica Scandinavica*. 1996;94(S165):63-71. doi:10.1111/j.1600-0404.1996.tb05874.x

[13] Chou E. Alzheimer's disease: Current and future treatments. A Review. *International Journal of Medical Students*. 2014;2(2):56-63. doi:10.5195/ijms.2014.85

[14] Galasko D. An integrated approach to the management of alzheimer's disease: Assessing cognition, function and behaviour. *European Journal of Neurology*. 1998;5(S4). doi:10.1111/j.1468-1331.1998.tb00444.x

[15] Haupt M, Kurz A. Predictors of nursing home placement in patients with alzheimer's disease. *International Journal of Geriatric Psychiatry*. 1993;8(9):741-746. doi:10.1002/gps.930080906

[16] Burns A. Psychiatric phenomena in dementia of the alzheimer type. *International Psychogeriatrics*. 1992;4(3):43-54. doi:10.1017/s1041610292001145

[17] Beard CM, Kokmen E, Sigler C, Smith GE, Petterson T, O'Brien PC. Cause of death in alzheimer's disease. *Annals of Epidemiology*. 1996;6(3):195-200. doi:10.1016/1047-2797(95)00068-2

[18] Panza F, Lozupone M, Logroscino G, Imbimbo BP. A critical appraisal of amyloid-β-targeting therapies for alzheimer disease. *Nature Reviews Neurology*. 2019;15(2):73-88. doi:10.1038/s41582-018-0116-6

[19] DALES MJ. Medicare may cover PET scan for cervical cancer patients. *Family Practice News*. 2009;39(15):58. doi:10.1016/s0300-7073(09)70684-x

[20] Small GW. Neuroimaging and genetic assessment for early diagnosis of Alzheimer's disease. *J Clin Psychiatry*. 1996;57 Suppl 14:9-13.

[21] Nordberg A. PET studies and cholinergic therapy in Alzheimer's disease. *Rev Neurol (Paris)*. 1999;155 Suppl 4:S53-S63.

[22] Fortea A, Pinzón-Espinosa J, Ilzarbe D, et al. Radiological findings in brain MRI scans in youth with early-onset psychosis: A controlled study. *Journal of Psychiatric Research*. 2022;156:151-158. doi:10.1016/j.jpsychires.2022.10.016

[23] Gosche KM, Mortimer JA, Smith CD, Markesbery WR, Snowdon DA. Hippocampal volume as an index of alzheimer neuropathology: Findings from the nun study. *Neurology*. 2002;58(10):1476-1482. doi:10.1212/wnl.58.10.1476

[24] Morris J. Diagnosis of alzheimer's disease. *Alzheimer's Disease*. Published online 1996. doi:10.1201/9781420048407.ch7

[25] Umar T, Meena R, Mustehasan, Kumar P, Khan AA. Recent updates in the development of small molecules as potential clinical candidates for alzheimer's disease: A review. *Chemical Biology & Drug Design*. 2022;100(5):674-681. doi:10.1111/cbdd.14133

[26] Heneka MT, Kummer MP, Stutz A, et al. NLRP3 is activated in alzheimer's disease and contributes to pathology in APP/PS1 mice. *Nature*. 2012;493(7434):674-678. doi:10.1038/nature11729

[27] Welikovitch LA, Do Carmo S, Maglóczky Z, et al. Early intraneuronal amyloid triggers neuron-derived inflammatory signaling in app transgenic rats and human brain. *Proceedings of the National Academy of Sciences*. 2020;117(12):6844-6854. doi:10.1073/pnas.1914593117

[28] Newcombe EA, Camats-Perna J, Silva ML, Valmas N, Huat TJ, Medeiros R. Inflammation: The link between comorbidities, genetics, and alzheimer's disease. *Journal of Neuroinflammation*. 2018;15(1). doi:10.1186/s12974-018-1313-3

[29] Xue F, Du H. Trem2 mediates microglial anti-inflammatory activations in alzheimer's disease: Lessons learned from transcriptomics. *Cells*. 2021;10(2):321. doi:10.3390/cells10020321

[30] Yang Y, Zhang Z. Microglia and Wnt pathways: Prospects for inflammation in alzheimer's disease. *Frontiers in Aging Neuroscience*. 2020;12. doi:10.3389/fnagi.2020.00110

[31] Hampel H, Caraci F, Cuello AC, et al. A path toward precision medicine for neuroinflammatory mechanisms in alzheimer's disease. *Frontiers in Immunology*. 2020;11. doi:10.3389/fimmu.2020.00456

[32] Ndoja A, Reja R, Lee S-H, et al. Ubiquitin ligase COP1 suppresses neuroinflammation by degrading C/EBPB in Microglia. *Cell*. 2020;182(5). doi:10.1016/j.cell.2020.07.011

[33] Blanchard JW, Bula M, Davila-Velderrain J, et al. Reconstruction of the human blood–brain barrier in vitro reveals a pathogenic mechanism of APOE4 in pericytes. *Nature Medicine*. 2020;26(6):952-963. doi:10.1038/s41591-020-0886-4

[34] Sweeney MD, Sagare AP, Zlokovic BV. Blood–brain barrier breakdown in alzheimer disease and other neurodegenerative disorders. *Nature Reviews Neurology*. 2018;14(3):133-150. doi:10.1038/nrneurol.2017.188

[35] Whitson HE, Colton C, El Khoury J, et al. Infection and inflammation: New perspectives on alzheimer's disease. *Brain, Behavior, & Immunity - Health*. 2022;22:100462. doi:10.1016/j.bbih.2022.100462

[36] Umar S, van der Laarse A. Nitric oxide and nitric oxide synthase isoforms in the normal, hypertrophic, and failing heart. *Molecular and Cellular Biochemistry*. 2009;333(1-2):191-201. doi:10.1007/s11010-009-0219-x

[37] Lowenstein CJ, Padalko E. Inos (NOS2) at a glance. *Journal of Cell Science*. 2004;117(14):2865-2867. doi:10.1242/jcs.01166

[38] Minhas R, Bansal Y, Bansal G. Inducible nitric oxide synthase inhibitors: A comprehensive update. *Medicinal Research Reviews*. 2019;40(3):823-855. doi:10.1002/med.21636

[39] Cinelli MA, Do HT, Miley GP, Silverman RB. Inducible nitric oxide synthase: Regulation, structure, and inhibition. *Medicinal Research Reviews*. 2019;40(1):158-189. doi:10.1002/med.21599

[40] Kim JY, Lim HJ, Lee DY, et al. In vitro anti-inflammatory activity of lignans isolated from Magnolia fargesii. *Bioorganic & Medicinal Chemistry Letters*. 2009;19(3):937-940. doi:10.1016/j.bmcl.2008.11.103

[41] Kohno S, Murata T, Sugiura A, et al. Methyl galbanate, a novel inhibitor of nitric oxide production in mouse macrophage raw264.7 cells. *Journal of Natural Medicines*. 2011;65(2):353-359. doi:10.1007/s11418-010-0505-7

[42] Liu F, Yang X, Ma J, et al. Nitric oxide inhibitory daphnane diterpenoids as potential anti-neuroinflammatory agents for AD from the twigs of Trigonostemon Thyrsoideus. *Bioorganic Chemistry*. 2017;75:149-156. doi:10.1016/j.bioorg.2017.09.007

[43] Liu F, Dong B, Yang X, et al. No inhibitors function as potential anti-neuroinflammatory agents for ad from the flowers of inula japonica. *Bioorganic Chemistry*. 2018;77:168-175. doi:10.1016/j.bioorg.2018.01.009

[44] Ma J, Ren Q, Dong B, et al. No inhibitory constituents as potential anti-neuroinflammatory agents for AD from Blumea balsamifera. *Bioorganic Chemistry*. 2018;76:449-457. doi:10.1016/j.bioorg.2017.12.008

[45] Zhao H-Y, Wang Y-Q, Li Y-C, et al. Flavonol glycosides and phenylpropanoid glycosides with inhibitory effects on microglial nitric oxide production from Neoshirakia japonica. *Fitoterapia*. 2021;151:104877. doi:10.1016/j.fitote.2021.104877

[46] Dukic-Stefanovic S, Gasic-Milenkovic J, Deuther-Conrad W, Münch G. Signal transduction pathways in mouse microglia n-11 cells activated by advanced glycation endproducts (ages). *Journal of Neurochemistry*. 2003;87(1):44-55. doi:10.1046/j.1471-4159.2003.01988.x

[47] Esposito G, De Filippis D, Maiuri MC, De Stefano D, Carnuccio R, Iuvone T. Cannabidiol inhibits inducible nitric oxide synthase protein expression and nitric oxide production in β-amyloid stimulated PC12 neurons through P38 MAP kinase and NF-KB involvement. *Neuroscience Letters*. 2006;399(1-2):91-95. doi:10.1016/j.neulet.2006.01.047

[48] Yan A, Liu Z, Song L, et al. Idebenone alleviates neuroinflammation and modulates microglial polarization in LPS-stimulated BV2 cells and MPTP-induced parkinson's disease mice. *Frontiers in Cellular Neuroscience*. 2019;12. doi:10.3389/fncel.2018.00529

[49] Wang M, Liu T, Chen S, Wu M, Han J, Li Z. Design and synthesis of 3-(4-pyridyl)-5-(4-sulfamido-phenyl)-1,2,4-oxadiazole derivatives as novel GSK-3B inhibitors and evaluation of their potential as multifunctional anti-alzheimer agents. *European Journal of Medicinal Chemistry*. 2021;209:112874. doi:10.1016/j.ejmech.2020.112874

[50] Liu T, Chen S, Du J, Xing S, Li R, Li Z. Design, synthesis, and biological evaluation of novel (4-(1,2,4-oxadiazol-5-yl)phenyl)-2-aminoacetamide derivatives as multifunctional agents for the treatment of alzheimer's disease. *European Journal of Medicinal Chemistry*. 2022;227:113973. doi:10.1016/j.ejmech.2021.113973

[51] Yang Y-X, Zheng L-T, Shi J-J, et al. Synthesis of 5α-cholestan-6-one derivatives and their inhibitory activities of no production in activated microglia: Discovery of a novel neuroinflammation inhibitor. *Bioorganic & Medicinal Chemistry Letters*. 2014;24(4):1222-1227. doi:10.1016/j.bmcl.2013.12.055

[52] Watterson DM, Mirzoeva S, Guo L, et al. Ligand modulation of glial activation: Cell permeable, small molecule inhibitors of serine-threonine protein kinases can block induction of interleukin 1β and nitric oxide synthase II. *Neurochemistry International*. 2001;39(5-6):459-468. doi:10.1016/s0197-0186(01)00053-5

[53] Zhuo Y, Guo H, Cheng Y, et al. Inhibition of phosphodiesterase-4 reverses the cognitive dysfunction and oxidative stress induced by AB25–35 in rats. *Metabolic Brain Disease*. 2016;31(4):779-791. doi:10.1007/s11011-016-9814-1

[54] Raies AB, Bajic VB. *in silico* toxicology: Computational methods for the prediction of chemical toxicity. *WIREs Computational Molecular Science*. 2016;6(2):147-172. doi:10.1002/wcms.1240

[55] Myatt GJ, Ahlberg E, Akahori Y, et al. In silico toxicology protocols. *Regulatory Toxicology and Pharmacology*. 2018;96:1-17. doi:10.1016/j.yrtph.2018.04.014

[56] Michielan L, Moro S. Pharmaceutical perspectives of nonlinear qsar strategies. *Journal of Chemical Information and Modeling*. 2010;50(6):961-978. doi:10.1021/ci100072z

[57] Priya V, Prabu S. Application of AI in Healthcare. *Smart Systems for Industrial Applications*. Published online 2022:225-248. doi:10.1002/9781119762010.ch9

[58] Tiwari AK. Introduction to machine learning. *Deep Learning and Neural Networks*. Published online 2020:41-51. doi:10.4018/978-1-7998-0414-7.ch003

[59] Mao J, Akhtar J, Zhang X, et al. Comprehensive strategies of machine-learning-based quantitative structure-activity relationship models. *iScience*. 2021;24(9):103052. doi:10.1016/j.isci.2021.103052

[60] Sarker IH. Machine learning: Algorithms, real-world applications and Research Directions. *SN Computer Science*. 2021;2(3). doi:10.1007/s42979-021-00592-x

[61] Parunakian D, Efitorov A, Shirokii V. Analysis of Mercury's magnetosphere states based on Messenger Data by Kohonen neural network and other clustering ALGORITHMS1. *Procedia Computer Science*. 2016;88:499-504. doi:10.1016/j.procs.2016.07.471

[62] Kiang MY. Extending the kohonen self-organizing map networks for clustering analysis. *Computational Statistics & Data Analysis*. 2001;38(2):161-180. doi:10.1016/s0167-9473(01)00040-8

[63] ACD/chem sketch. Available at: Draw Chemical Structures, Molecules, and Reactions | ChemSketch | ACD/Labs (acdlabs.com) (Accessed: November 9, 2022)

[64] SPARTAN 20 descriptor calculation Available at: Spartan'20 | wavefunction (Accessed: December 1, 2022)

[65] ALVADESC descriptor calculation Available at: alvaDesc - Alvascience (Accessed: January 1 , 2023)

[66] Ballabio D, Consonni V, Todeschini R. The kohonen and CP-ann toolbox: A collection of MATLAB modules for self organizing maps and counterpropagation artificial neural networks. *Chemometrics and Intelligent Laboratory Systems*. 2009;98(2):115-122. doi:10.1016/j.chemolab.2009.05.007

[67] Ran J, Xu Y, Tong Y. The eis-based kohonen neural network for high strength steel coating degradation assessment. *Journal of Chemometrics*. 2017;31(11). doi:10.1002/cem.2919

[68] WEKA software available at: Weka - Browse /weka-3-8/3.8.6 at SourceForge.net (Accesssed: February 1 , 2023)

[69] Witten IH, Frank E. Data Mining. *ACM SIGMOD Record*. 2002;31(1):76-77. doi:10.1145/507338.507355

[70] Gramatica P. On the development and validation of QSAR models. *Methods in Molecular Biology*. Published online 2012:499-526. doi:10.1007/978-1-62703-059-5_21

[71] Gramatica P. Principles of QSAR models validation: Internal and External. *QSAR & Combinatorial Science*. 2007;26(5):694-701. doi:10.1002/qsar.200610151

[72] Baidya SK, Amin SkA, Banerjee S, Adhikari N, Jha T. Structural exploration of arylsulfonamide-based ADAM17 inhibitors through validated comparative multi-qsar modelling studies. *Journal of Molecular Structure*. 2019;1185:128-142. doi:10.1016/j.molstruc.2019.02.081

[73] Mathea M, Klingspohn W, Baumann K. Chemoinformatic classification methods and their applicability domain. *Molecular Informatics*. 2016;35(5):160-180. doi:10.1002/minf.201501019

[74] Microsoft Excel availabile at: Microsoft Excel Spreadsheet Software | Microsoft 365 (Accessed: March 1 , 2023)

[75] Ishii K, Muelhauser F, Liebl U, et al. Subacute no generation induced by alzheimer's β-amyloid in the living brain: Reversal by inhibition of the inducible no synthase. *The FASEB Journal*. 2000;14(11):1485-1489. doi:10.1096/fj.99-0786com

[76] Leonard JT, Roy K. On selection of training and test sets for the development of predictive QSAR models. *QSAR & Combinatorial Science*. 2006;25(3):235-251. doi:10.1002/qsar.200510161

[77] Omran MGH, Engelbrecht AP, Salman A. An overview of clustering methods. *Intelligent Data Analysis*. 2007;11(6):583-605. doi:10.3233/ida-2007-11602

[78] B. Demisse G, Tadesse T, Bayissa Y. Data mining attribute selection approach for drought modelling : A case study for greater horn of africa. *International Journal of Data Mining & Knowledge Management Process*. 2017;7(4):01-16. doi:10.5121/ijdkp.2017.7401

[79] Witten IH, Frank E. Data Mining. *ACM SIGMOD Record*. 2002;31(1):76-77. doi:10.1145/507338.507355

[80] Varsou D-D, Melagraki G, Sarimveis H, Afantitis A. MouseTox: An online toxicity assessment tool for small molecules through Enalos Cloud Platform. *Food and Chemical Toxicology*. 2017;110:83-93. doi:10.1016/j.fct.2017.09.058

[81] Gregorutti B, Michel B, Saint-Pierre P. Correlation and variable importance in random forests. *Statistics and Computing*. 2016;27(3):659-678. doi:10.1007/s11222-016-9646-1

[82] Oztan Akturk S, Tugcu G, Sipahi H. Development of a QSAR model to predict comedogenic potential of some cosmetic ingredients. *Computational Toxicology*. 2022;21:100207. doi:10.1016/j.comtox.2021.100207

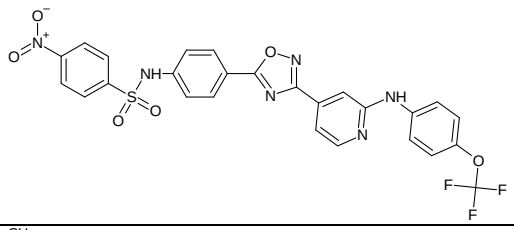
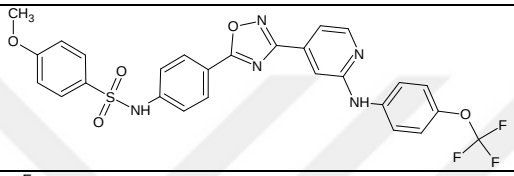
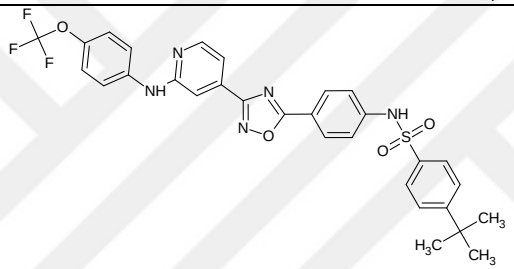
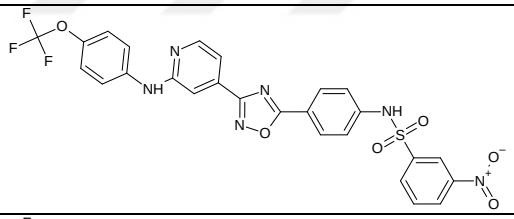
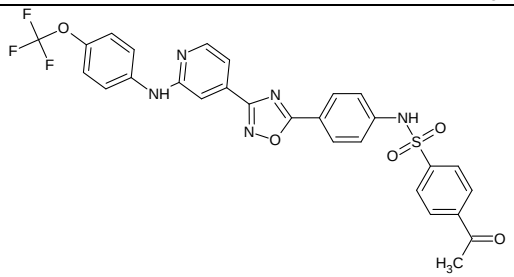
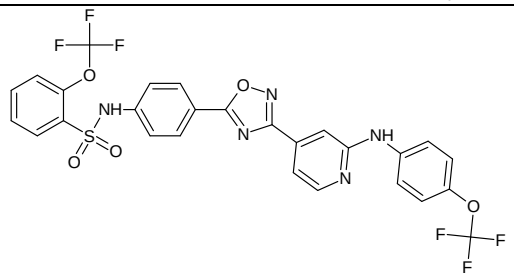
[83] Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*. 2015;10(3). doi:10.1371/journal.pone.0118432

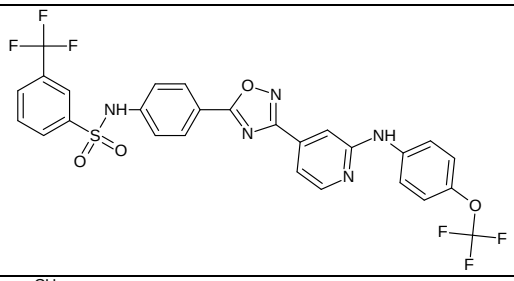
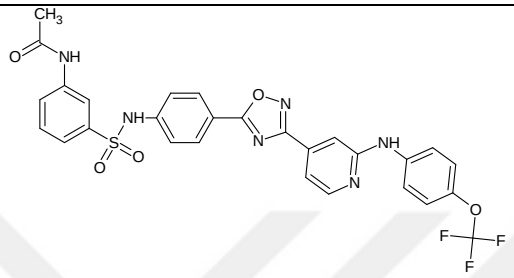
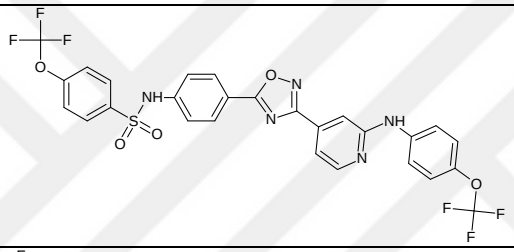
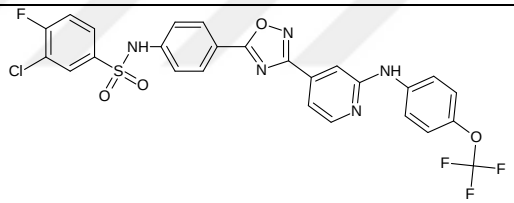
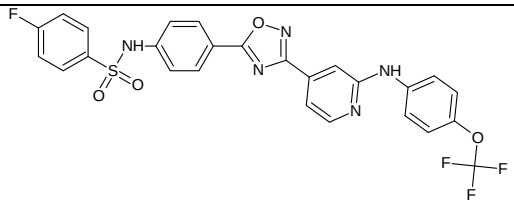
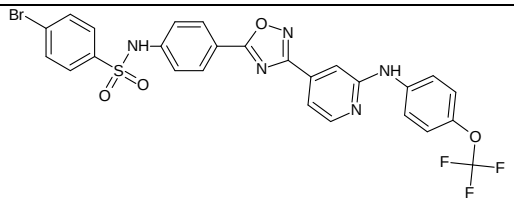
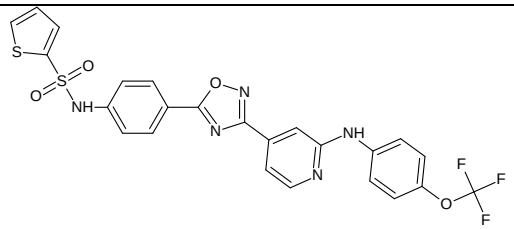
[84] Kühne R, Ebert R-U, Schüürmann G. Chemical Domain of QSAR models from Atom-centered fragments. *Journal of Chemical Information and Modeling*. 2009;49(12):2660-2669. doi:10.1021/ci900313u

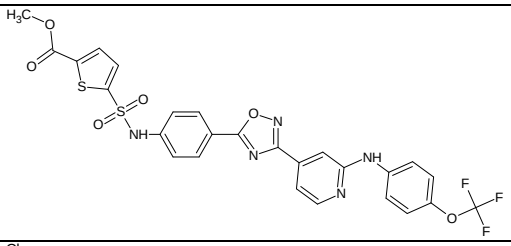
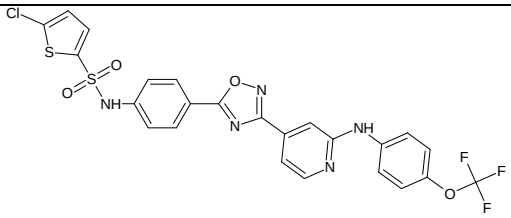
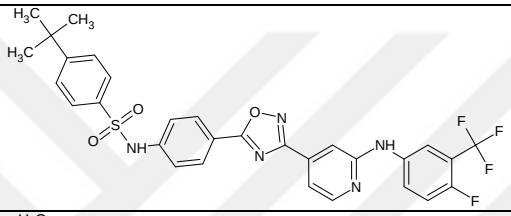
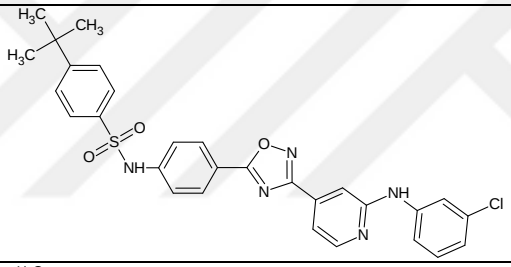
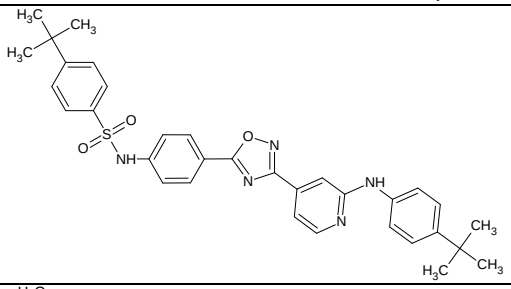
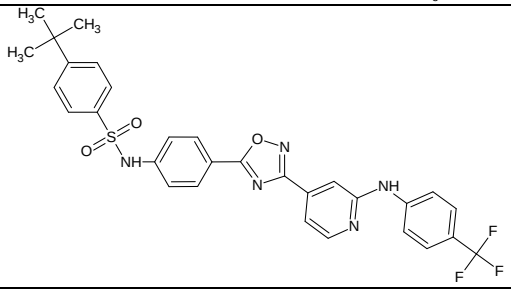
[85] Long W, Xiang J, Wu H, et al. QSAR modeling of inos inhibitors based on a novel regression method: Multi-stage adaptive regression. *Chemometrics and Intelligent Laboratory Systems*. 2013;128:83-88. doi:10.1016/j.chemolab.2013.07.011

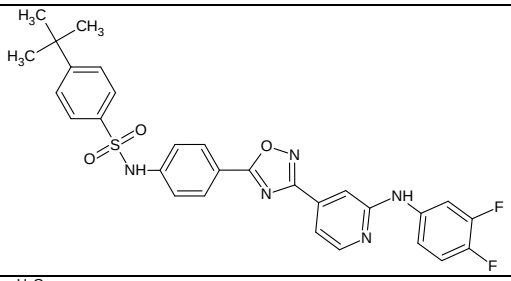
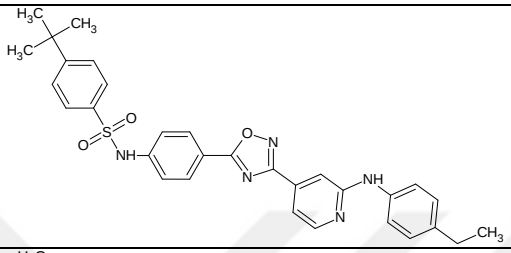
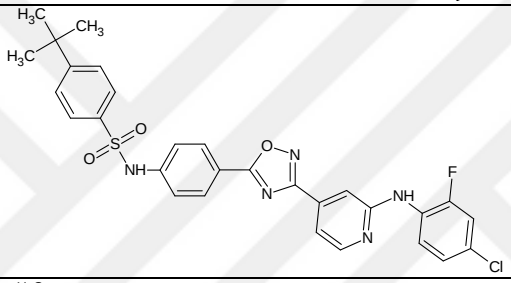
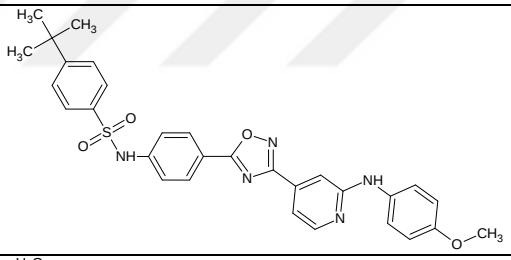
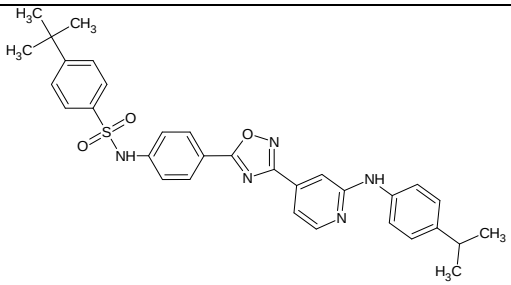
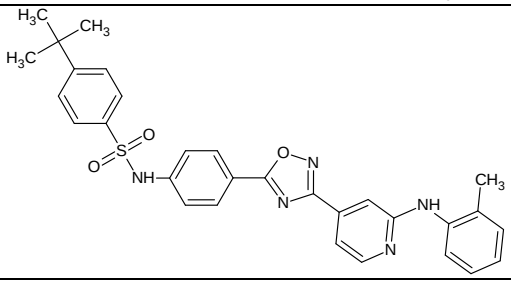
7. APPENDICES

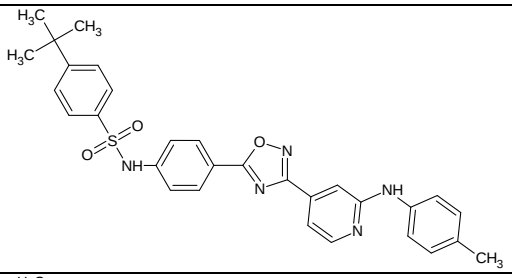
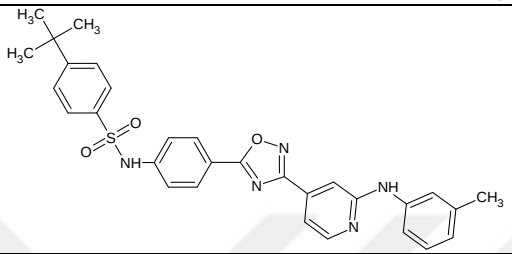
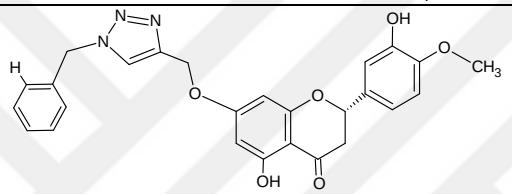
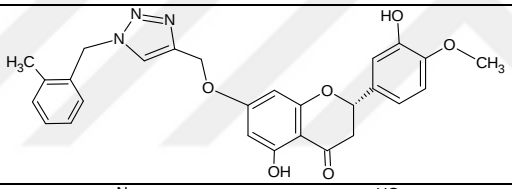
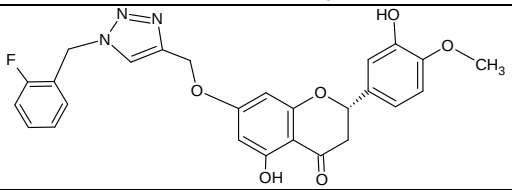
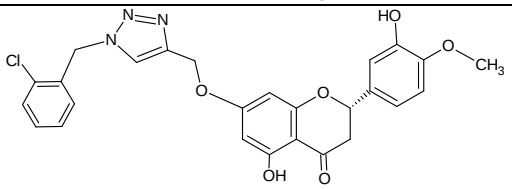
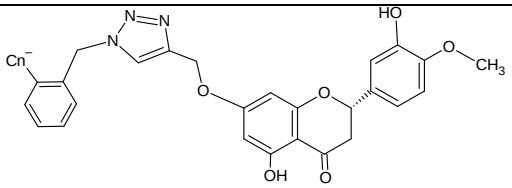
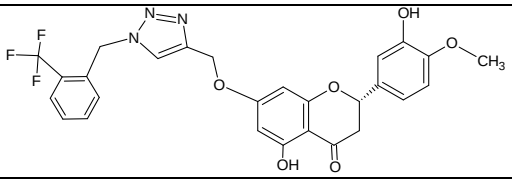
Appendix 7.1_TableS1.Dataset

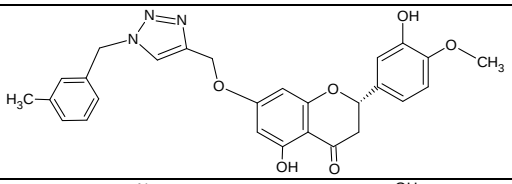
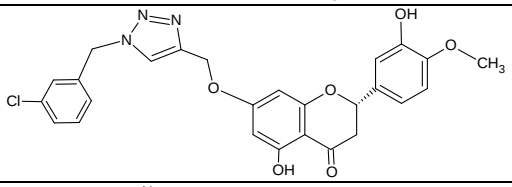
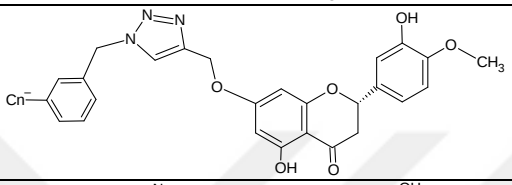
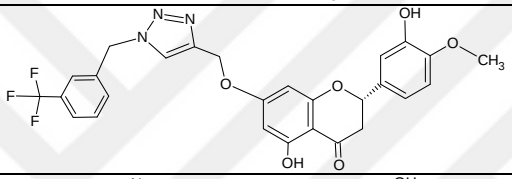
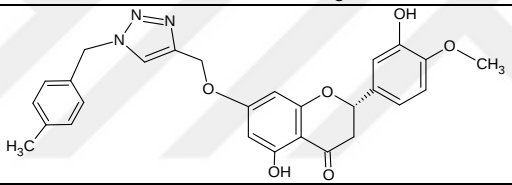
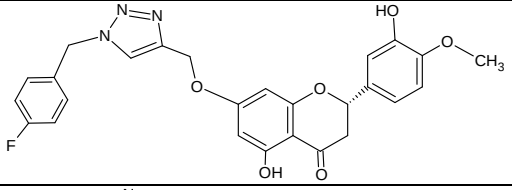
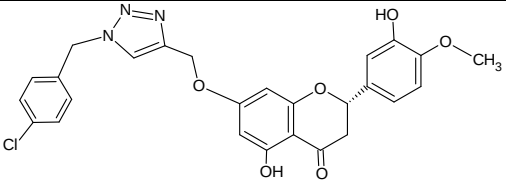
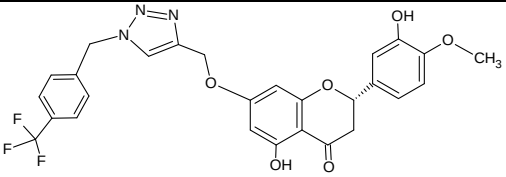
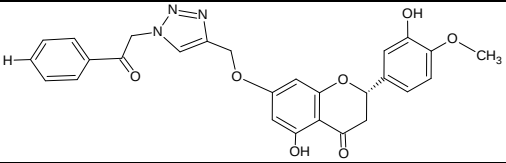
Com. no	Chemical structure	IC_{50} µM	Class
1		7.59	inactive
2		3.23	active
3		0.72	active
4		2.15	active
5		0.47	active
6		2.53	active

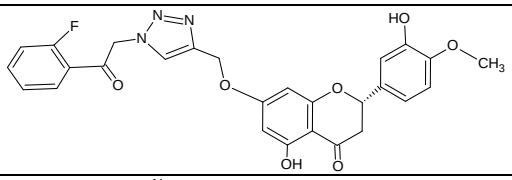
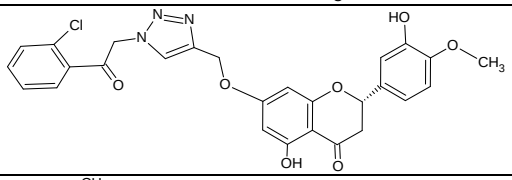
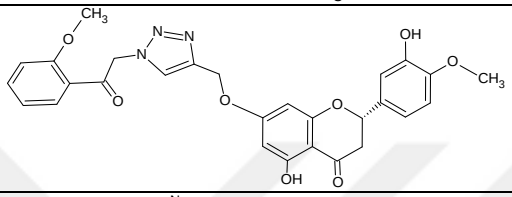
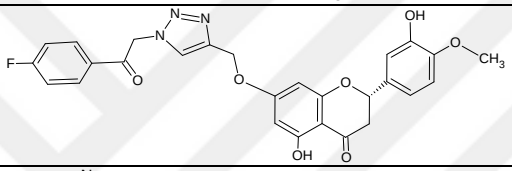
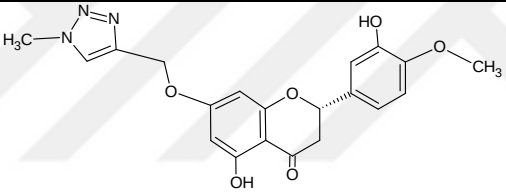
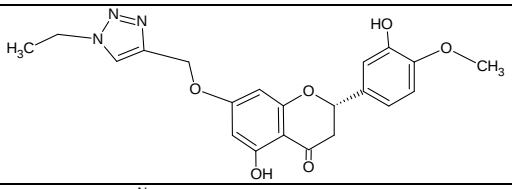
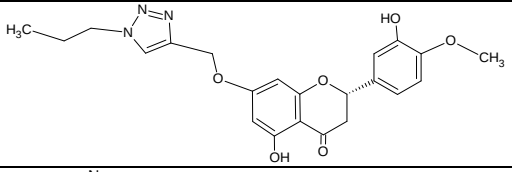
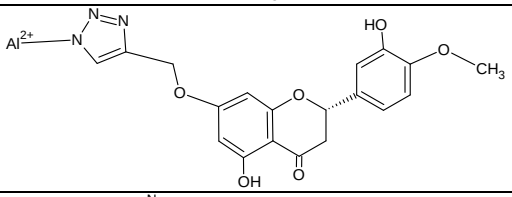
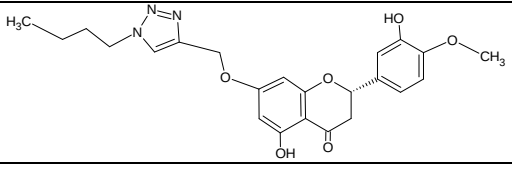
7		1.49	active
8		7.55	inactive
9		4.22	active
10		4.96	active
11		4.41	active
12		5.16	active
13		1.30	active

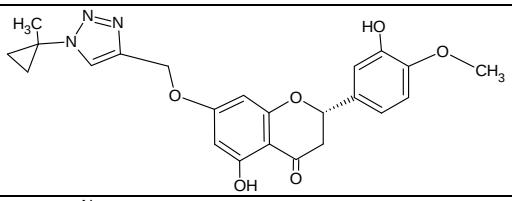
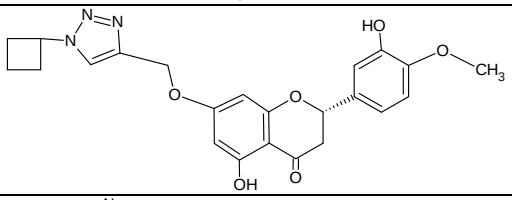
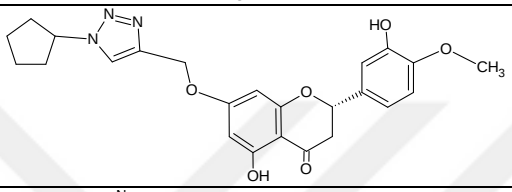
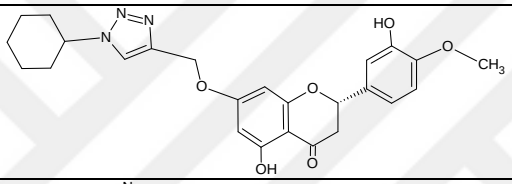
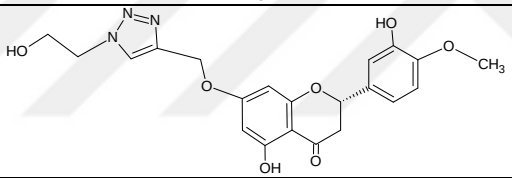
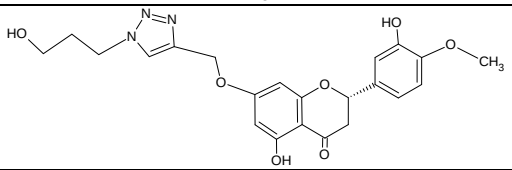
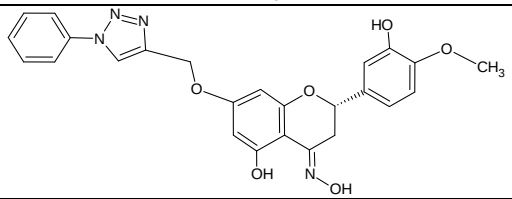
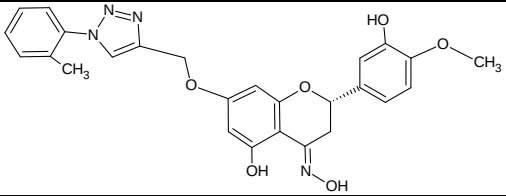
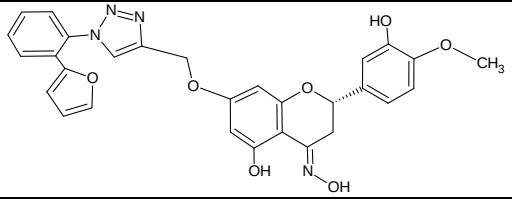
14		1.31	active
15		6.23	inactive
16		8.94	inactive
17		6.94	inactive
18		10.89	inactive
19		8.69	inactive

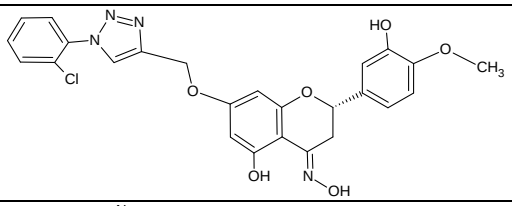
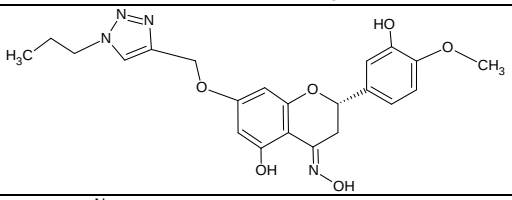
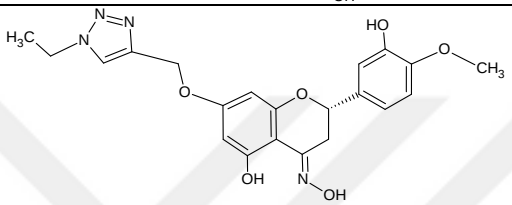
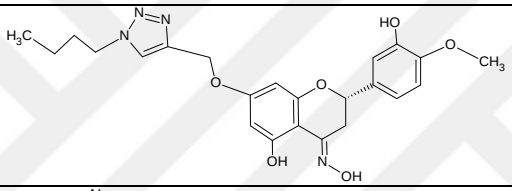
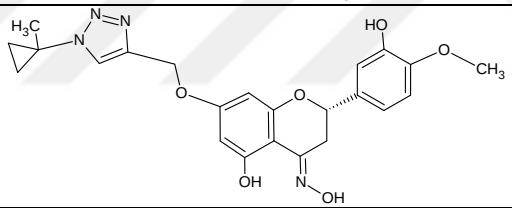
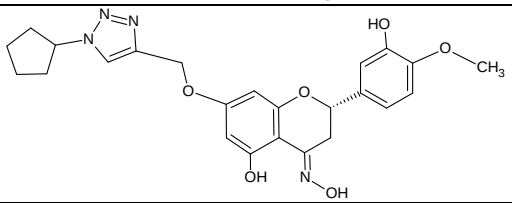
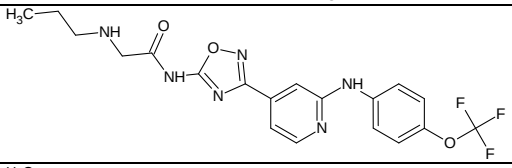
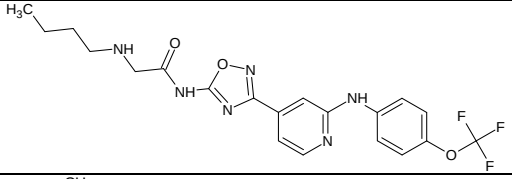
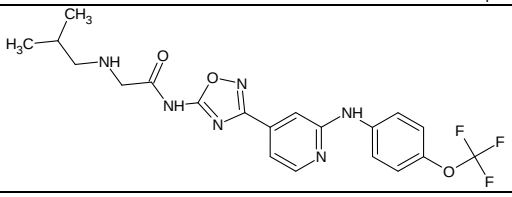
20		2.44	active
21		5.40	active
22		4.40	active
23		6.18	inactive
24		7.82	inactive
25		6.07	active

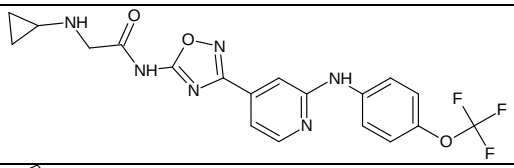
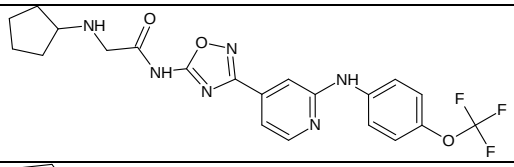
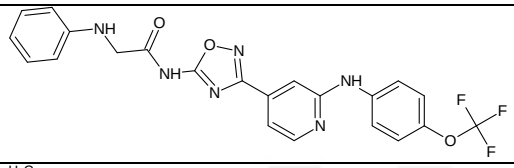
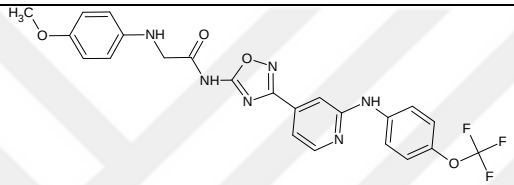
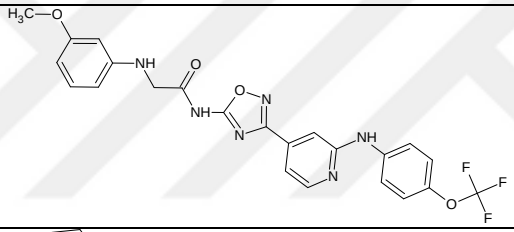
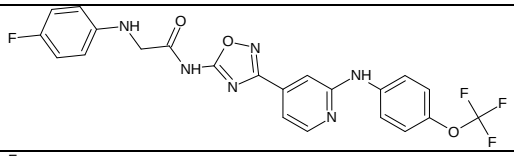
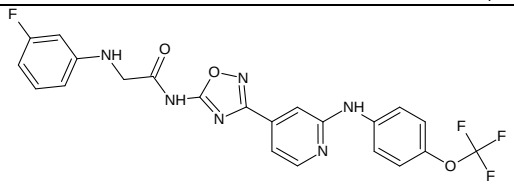
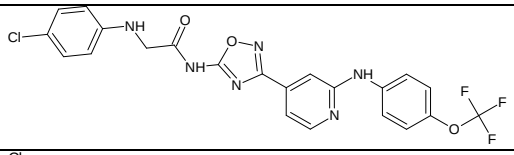
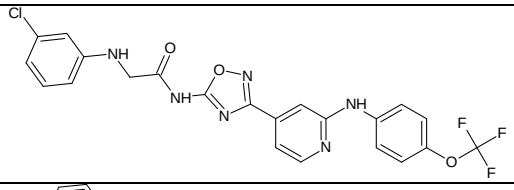
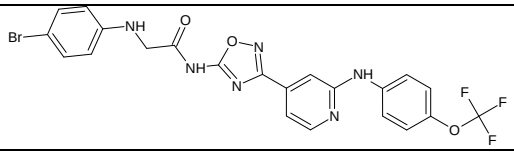
26		1.50	active
27		10.59	inactive
28		1.04	active
29		13.69	inactive
30		4.79	active
31		6.63	inactive
32		7.74	inactive
33		8.95	inactive

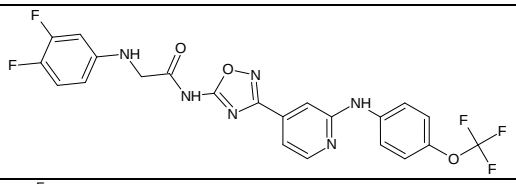
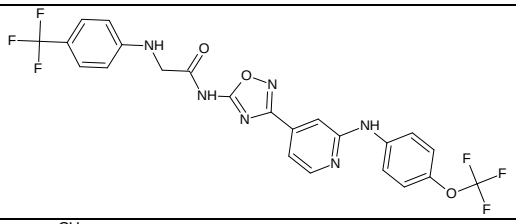
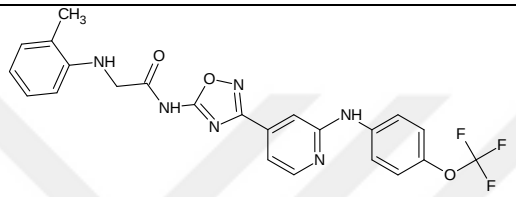
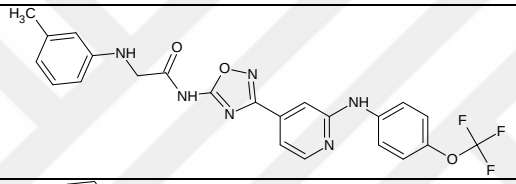
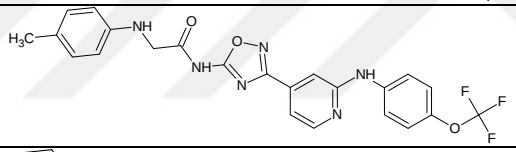
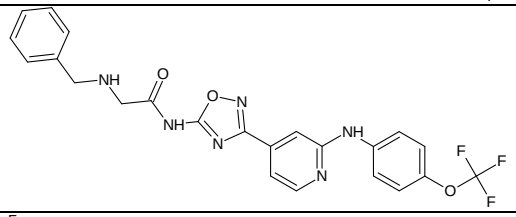
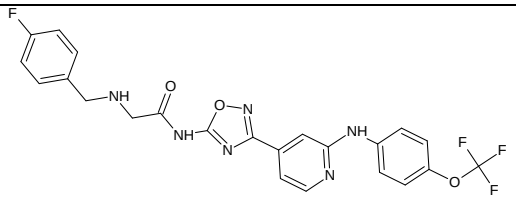
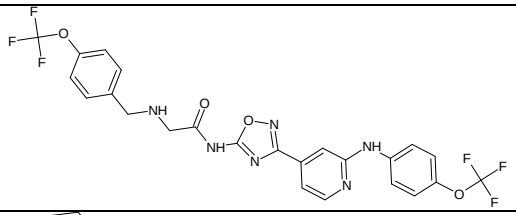
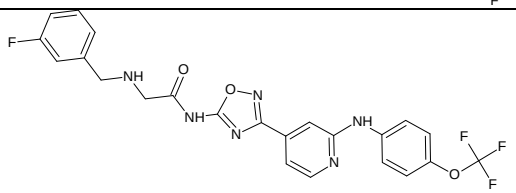
34		2.21	active
35		2.91	active
36		20.32	inactive
37		12.87	inactive
38		35.63	inactive
39		13.41	inactive
40		16.03	inactive
41		3.38	active
42		26.3	inactive

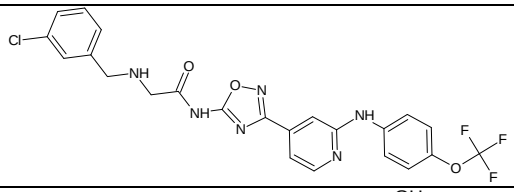
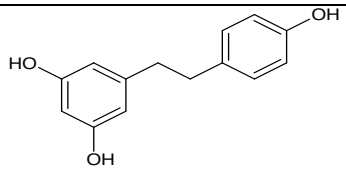
43		8.93	inactive
44		17.25	inactive
45		199.38	inactive
46		27.02	inactive
47		48.22	inactive
48		35.49	inactive
49		12.38	inactive
50		14.88	inactive
51		3.23	active

52		2.55	active
53		2.04	active
54		2.26	active
55		4.19	active
56		12.98	inactive
57		6.60	inactive
58		6.22	inactive
59		10.69	inactive
60		4.86	active

61		3.65	active
62		3.95	active
63		6.60	inactive
64		5.92	active
65		5.88	active
66		4.95	active
67		7.82	inactive
68		8.24	inactive
69		3.88	active

70		5.67	active
71		2.66	active
72		0.42	active
73		6.95	inactive
74		13.15	inactive
75		0.67	active
76		3.16	active
77		2.58	active
78		3.66	active
79		4.22	active

80		4.80	active
81		1.61	active
82		8.74	inactive
83		7.29	inactive
84		5.97	active
85		7.48	inactive
86		9.23	inactive
87		8.48	inactive
88		12.95	inactive

89		14.57	inactive
90		9.7	Inactive

Appendix 7.2_TableS2. SMILES and the IC_{50} Value of the Selected Compounds with their Classes.

NO	SMILES	IC_{50} (µM)	CLASS
1	[O-][N+](=O)c1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(OC(F)(F)F)cc1	7.59	inactive
2	COc1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(OC(F)(F)F)cc1	3.23	active
3	CC(C)(C)c1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(OC(F)(F)F)cc1	0.72	active
4	[O-][N+](=O)c1cc(ccc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(OC(F)(F)F)cc1	2.15	active
5	CC(=O)c1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(OC(F)(F)F)cc1	0.47	active
6	FC(F)(F)Oc1ccccc1S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(OC(F)(F)F)cc1	2.53	active
7	FC(F)(F)c1cc(ccc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(OC(F)(F)F)cc1	1.49	active
8	CC(=O)Nc1cc(ccc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(OC(F)(F)F)cc1	7.55	inactive
9	FC(F)(F)Oc1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(OC(F)(F)F)cc1	4.22	active
10	Fc1ccc(cc1Cl)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(OC(F)(F)F)cc1	4.96	active
11	Fc1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(OC(F)(F)F)cc1	4.41	active
12	Brc1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(OC(F)(F)F)cc1	5.16	active
13	O=S(=O)(Nc1ccc(cc1)c1nc(no1)c1ccnc(Nc2ccc(OC(F)(F)F)cc2)c1)c1cccs1	1.30	active
14	O=S(=O)(Nc1ccc(cc1)c1nc(no1)c1ccnc(Nc2ccc(OC(F)(F)F)cc2)c1)c1ccc(s1)C(=O)OC	1.31	active
15	O=S(=O)(Nc1ccc(cc1)c1nc(no1)c1ccnc(Nc2ccc(OC(F)(F)F)cc2)c1)c1ccc(Cl)s1	6.23	inactive
16	CC(C)(C)c1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1cc(c(F)cc1)C(F)(F)F	8.94	inactive
17	CC(C)(C)c1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1cccc(Cl)c1	6.94	inactive
18	CC(C)(C)c1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(cc1)C(C)(C)C	10.89	inactive
19	CC(C)(C)c1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(cc1)C(F)(F)F	8.69	inactive
20	CC(C)(C)c1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(F)c(F)c1	2.44	active
21	CC(C)(C)c1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(CC)cc1	5.40	active
22	CC(C)(C)c1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(Cl)cc1F	4.40	active
23	CC(C)(C)c1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(OC)cc1	6.18	inactive
24	CC(C)(C)c1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(cc1)C(C)C	7.82	inactive
25	CC(C)(C)c1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccccc1C	6.07	active
26	CC(C)(C)c1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1ccc(C)cc1	1.50	active
27	CC(C)(C)c1ccc(cc1)S(=O)(=O)Nc1ccc(cc1)c1nc(no1)c1cc(ncc1)Nc1cc(C)ccc1	10.59	inactive
28	COc1ccc(cc1O)[C@@H]1CC(=O)c2c(cc(cc2O)OCc2cn(Cc3ccccc3)nn2)O1	1.04	active
29	Cc1ccccc1Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	13.69	inactive
30	Fc1ccccc1Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	4.79	active
31	Clc1ccccc1Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	6.63	inactive
32	[Cn-]c1ccccc1Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	7.74	inactive

NO	SMILES	IC_{50} (μM)	CLASS
33	FC(F)(F)c1ccccc1Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	8.95	inactive
34	Cc1cccc(c1)Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	2.21	active
35	Clc1cccc(c1)Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	2.91	active
36	[Cn-]c1cc(ccc1)Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	20.32	inactive
37	FC(F)(F)c1cccc(c1)Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	12.87	inactive
38	Cc1ccc(cc1)Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	35.63	inactive
39	Fc1ccc(cc1)Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	13.41	inactive
40	Clc1ccc(cc1)Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	16.03	inactive
41	FC(F)(F)c1ccc(cc1)Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	3.38	active
42	O=C(Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1)c1ccccc1	26.3	inactive
43	Fc1ccccc1C(=O)Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	8.93	inactive
44	Clc1ccccc1C(=O)Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	17.25	inactive
45	COc1ccccc1C(=O)Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	199.38	inactive
46	Fc1ccc(cc1)C(=O)Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	27.02	inactive
47	Cn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	48.22	inactive
48	CCn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	35.49	inactive
49	CCCn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	12.38	inactive
50	[Al+2]n1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	14.88	inactive
51	CCCCn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	3.23	active
52	CC1(CC1)n1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	2.55	active
53	COc1ccc(cc1O)[C@@H]1CC(=O)c2c(cc(cc2O)OCc2cn(nn2)C2CCC2)O1	2.04	active
54	COc1ccc(cc1O)[C@@H]1CC(=O)c2c(cc(cc2O)OCc2cn(nn2)C2CCCC2)O1	2.26	active
55	COc1ccc(cc1O)[C@@H]1CC(=O)c2c(cc(cc2O)OCc2cn(nn2)C2CCCCC2)O1	4.19	active
56	OCCn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	12.98	inactive
57	OCCCn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](CC2=O)c1ccc(OC)c(O)c1	6.60	inactive
58	COc1ccc(cc1O)[C@@H]1C/C(=N\O)c2c(cc(cc2O)OCc2cn(nn2)c2ccccc2)O1	6.22	inactive
59	Cc1ccccc1n1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](C/C2=N\O)c1ccc(OC)c(O)c1	10.69	inactive
60	COc1ccc(cc1O)[C@@H]1C/C(=N\O)c2c(O)cc(cc2O1)OCc1cn(nn1)c1ccccc1c1ccco1	4.86	active
61	Clc1ccccc1n1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](C/C2=N\O)c1ccc(OC)c(O)c1	3.65	active
62	CCCn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](C/C2=N\O)c1ccc(OC)c(O)c1	3.95	active
63	CCn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](C/C2=N\O)c1ccc(OC)c(O)c1	6.60	inactive
64	CCCCn1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](C/C2=N\O)c1ccc(OC)c(O)c1	5.92	active
65	CC1(CC1)n1cc(nn1)COc1cc(O)c2c(c1)O[C@@H](C/C2=N\O)c1ccc(OC)c(O)c1	5.88	active
66	COc1ccc(cc1O)[C@@H]1C/C(=N\O)c2c(cc(cc2O)OCc2cn(nn2)C2CCCC2)O1	4.95	active

NO	SMILES	IC_{50} (μM)	CLASS
67	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNCCC)on1	7.82	inactive
68	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNCCCC)on1	8.24	inactive
69	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNCC(C)C)on1	3.88	active
70	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNC2CC2)on1	5.67	active
71	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNC2CCCC2)on1	2.66	active
72	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNc2ccccc2)on1	0.42	active
73	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNc2ccc(OC)cc2)on1	6.95	inactive
74	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNc2cccc(OC)c2)on1	13.15	inactive
75	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNc2ccc(F)cc2)on1	0.67	active
76	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNc2cccc(F)c2)on1	3.16	active
77	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNc2ccc(Cl)cc2)on1	2.58	active
78	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNc2cccc(Cl)c2)on1	3.66	active
79	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNc2ccc(Br)cc2)on1	4.22	active
80	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNc2ccc(F)c(F)c2)on1	4.80	active
81	FC(F)(F)c1ccc(cc1)NCC(=O)Nc1nc(no1)c1cc(ncc1)Nc1ccc(OC(F)(F)F)cc1	1.61	active
82	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNc2ccccc2C)on1	8.74	inactive
83	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNc2cc(C)ccc2)on1	7.29	inactive
84	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNc2ccc(C)cc2)on1	5.97	active
85	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNCc2ccccc2)on1	7.48	inactive
86	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNCc2ccc(F)cc2)on1	9.23	inactive
87	FC(F)(F)Oc1ccc(cc1)CNCC(=O)Nc1nc(no1)c1cc(ncc1)Nc1ccc(OC(F)(F)F)cc1	8.48	inactive
88	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNCc2cccc(F)c2)on1	12.95	inactive
89	FC(F)(F)Oc1ccc(cc1)Nc1nccc(c1)c1nc(NC(=O)CNCc2cccc(Cl)c2)on1	14.57	inactive
90	Oc1cc(/C=C/c2ccc(O)cc2)cc(O)c1	9.70	inactive

Appendix 7.3_TableS3. Selected Descriptors

Com. no	Training/Test	MATS6p	Chi1_EA(dm)	Mor17s	Nsss CH	SHED_AL
1	Training	0.046967014	31.83952053	-2.862677052	0	13.6906309
2	Test	0.24218374	25.6537668	-1.211453548	0	13.29446956
3	Test	0.133895063	32.13266345	-3.68227014	0	16.08497638
4	Test	0.252924989	26.4353884	-0.522395203	0	14.63276999
5	Training	0.152810067	22.33189463	-1.083805547	0	14.73600638
6	Training	0.234086419	27.66607069	-2.730219087	0	15.0008495
7	Test	0.333468611	26.72093115	-1.817677531	0	13.72720074
8	Test	0.000153951	28.41175846	-1.984815997	0	15.39855172
9	Training	0.143638465	26.71867159	-2.056862271	0	13.50447041
10	Training	0.242174236	25.97916868	-5.426072162	0	13.36413737
11	Test	0.197261313	24.81899783	-0.272875101	0	13.16736556
12	Test	-0.146231595	24.83248335	-2.643047338	0	14.57137245
13	Training	0.114752594	23.02570919	-1.386875509	0	12.70629962
14	Training	0.307825966	24.86226553	-6.539171443	0	13.31025132
15	Training	0.062863901	22.91894075	0.738848819	0	13.55320658
16	Training	0.202453399	32.91413691	-0.063661045	0	14.55566719
17	Training	0.213379968	30.44597129	0.791302554	0	14.97441716
18	Training	0.171195692	37.54665531	1.158173513	0	16.533857
19	Test	0.342076975	31.73448097	2.018237157	0	14.97808239
20	Training	0.177213859	29.41980112	-2.703033418	0	14.32662227
21	Training	0.19064751	31.02866991	3.460131794	0	16.01453058
22	Training	0.170461017	26.36608165	-3.752648833	0	14.6892928
23	Training	0.164850704	31.06775866	6.142437084	0	16.50585601
24	Test	0.180685225	31.56894847	1.224991336	0	16.35260566
25	Test	0.188574729	26.40632649	-5.399214969	0	15.03372723
26	Training	0.181955859	28.66711155	-6.335158101	0	15.11366067
27	Training	0.187792667	31.45979223	1.27262608	0	15.35045015
28	Training	0.162363094	49.61869778	-2.74701725	1	14.90675472
29	Training	0.148613892	49.66623827	-0.862309578	1	15.11535902
30	Training	0.160416427	48.6684909	-0.290901367	1	14.65361959
31	Training	0.163277366	48.63234517	-1.387605067	1	14.86521208
32	Training	0.166209738	50.35782358	-3.378604618	1	15.23156555
33	Training	0.143517353	52.18247638	1.542313276	1	14.90675472
34	Training	0.22615883	52.55987426	-1.846420421	1	15.46545265
35	Training	0.186434102	51.23064297	-0.083895516	1	15.0905098
36	Training	0.186991766	53.09351075	-1.483045442	1	15.25844265
37	Training	0.231997756	55.24299293	0.880086621	1	14.90675472
38	Training	0.095425213	51.08928602	-1.785263093	1	15.74503676
39	Test	0.163870377	50.28094944	0.10773806	1	13.90375921
40	Training	0.117412588	50.25972343	1.275588426	1	14.92526336
41	Training	0.086760653	53.77240469	2.463122137	1	14.90675472
42	Training	0.145935372	44.87680957	0.033273192	1	16.1859846
43	Training	0.142603985	46.00276023	1.185573395	1	15.91009013
44	Test	0.169473969	45.97404263	1.734596572	1	16.06949013

NO	Training/Test	MATS6p	Chi1_EA(dm)	Mor17s	NsssCH	SHED_AL
45	Training	0.111794333	50.14240661	1.56368027	1	16.28971152
46	Training	0.143529874	45.53906122	0.484216615	1	15.42625986
47	Training	0.209480339	38.56104719	1.12396016	1	9.999127244
48	Test	0.11707431	38.56104719	1.10394127	1	10.80076418
49	Test	0.124354197	38.56104719	0.262075905	1	11.86815113
50	Training	0.145119936	46.23069708	1.654036265	1	11.86815113
51	Training	0.094587941	38.56104719	-0.590607914	1	12.98628939
52	Training	0.143620083	38.56104719	-0.750328973	2	12.37725851
53	Training	0.061748287	41.06104719	-0.251499355	2	12.09859469
54	Training	0.066308083	41.06104719	0.338620041	2	12.58725704
55	Training	0.047524508	41.06104719	-0.224648838	2	13.54664737
56	Training	0.115091899	38.56104719	0.736085047	1	11.91066005
57	Training	0.113554481	40.26603304	0.85417468	1	11.94914863
58	Training	0.136975198	50.8509723	-1.112289956	1	15.34057668
59	Training	0.125651928	50.89851278	-2.451341041	1	15.5805228
60	Training	0.135221027	49.90076542	-2.678020837	1	15.05528989
61	Training	0.139647041	49.86461969	-0.348791429	1	15.32106698
62	Test	0.098211795	39.79332171	-1.138918901	1	12.36015664
63	Training	0.090152407	39.79332171	-1.334974289	1	11.17514293
64	Training	0.072507842	39.79332171	-2.31524944	1	13.45160542
65	Training	0.116617279	39.79332171	-2.38520533	2	12.96645167
66	Training	0.046611367	42.29332171	-1.03680996	2	13.28516969
67	Training	0.107803153	37.14239202	-5.009483423	0	13.23623949
68	Training	0.107803153	37.14239202	-5.090627317	0	13.23623949
69	Training	0.103449241	37.14239202	-5.559419928	1	13.93598215
70	Training	0.118641488	42.14239202	-5.225142349	1	12.48269097
71	Training	0.113732911	42.14239202	-5.202969796	1	14.19955093
72	Training	0.080844439	42.14239202	-4.579625652	0	14.93676959
73	Test	0.096302217	46.79355481	-4.373358974	0	14.7461103
74	Training	0.066268828	48.49854066	-3.593145206	0	14.97648933
75	Training	0.080078517	42.80464367	-2.67251791	0	14.19955093
76	Training	0.081986509	44.0913562	-2.065812043	0	14.4363359
77	Test	0.079509096	42.78341766	-3.074439203	0	15.03468787
78	Training	0.052792493	44.04934187	-4.101373959	0	14.84180078
79	Training	0.070411742	42.81806769	-4.65165815	0	15.03468787
80	Training	0.081211808	44.36566982	-1.046189097	0	13.59638207
81	Training	0.083436866	46.29609893	2.393817651	0	14.93676959
82	Training	0.105519122	41.35219518	-3.928669066	0	15.18966886
83	Training	0.082310164	45.53039272	-3.450500428	0	15.28116229
84	Training	0.084135604	43.61298025	-6.074372055	0	15.65684682
85	Training	0.067943447	40.87373112	-6.152481163	0	15.70779937
86	Test	0.071203316	41.53598278	-4.503811879	0	15.05510806
87	Training	-0.001605097	45.02743803	-3.296397325	0	15.70779937
88	Training	0.067289183	42.52284624	-5.345754567	0	15.44728847
89	Training	0.067474542	42.48567631	-5.478195377	0	15.77017011
90	Training	0.05223016	4.651162791	1.961432558	0	7.311634425

8. CURRICULUM VITAE

Personal Informations

Name	Shkar Mariwan	**Surname**	Ahmed

Education

Degree	**Department**	**The name of the Institution Graduated From**	**Graduation year**
University	Pharmacy	**Istanbul Altinbas University**	2021
High school	Science school	**Rzgary High School**	2016

Languages	**Grades (#))**
English	Upper Intermediate
Turkish	Intermediate

Work Experience (Sort from present to past)

Position	**A) Institute**	**Duration (Year - Year)**
		-
		-

Computer Skills

Program	**Level**
Microsoft Office	Intermediate
Microsoft Power Point	Intermediate