Journal of Pharmaceutical Analysis 14 (2024) 101022

Contents lists available at ScienceDirect

Journal of Pharmaceutical Analysis

journal homepage: www.elsevier.com/locate/jpa

Original article

Deep learning-based drug screening for the discovery of potential therapeutic agents for Alzheimer's disease

Tong Wu [a], Ruimei Lin [a], Pengdi Cui [a], Jie Yong [a], Heshui Yu [a, b, c, **], Zheng Li [a, b, c, d, *]

[a] College of Pharmaceutical Engineering of Traditional Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, 301617, China
[b] State Key Laboratory of Component-Based Chinese Medicine, Tianjin, 301617, China
[c] Haihe Laboratory of Modern Chinese Medicine, Tianjin, 301617, China
[d] National Key Laboratory of Chinese Medicine Modernization, Tianjin, 301617, China

ARTICLE INFO

Article history:
Received 14 December 2023
Received in revised form
29 May 2024
Accepted 11 June 2024
Available online 13 June 2024

Keywords:
Alzheimer's disease
Deep learning models
Drug screening
Kaixinsan formula
Acetylcholinesterase
Monoamine oxidase-A

ABSTRACT

Alzheimer's disease (AD) is gradually increasing in prevalence and the complexity of its pathogenesis has led to a lengthy process of developing therapeutic drugs with limited success. Faced with this challenge, we proposed using a state-of-the-art drug screening algorithm to identify potential therapeutic compounds for AD from traditional Chinese medicine formulas with strong empirical support. We developed four deep neural network (DNN) models for AD drugs screening at the disease and target levels. The AD model was trained with compounds labeled for AD activity to predict active compounds at the disease level, while the acetylcholinesterase (AChE), monoamine oxidase-A (MAO-A), and 5-hydroxytryptamine 6 (5-HT_6) models were trained for specific AD targets. All four models performed excellently and were used to identify potential AD agents in the Kaixinsan (KXS) formula. High-scoring compounds underwent experimental validation at the enzyme, cellular, and animal levels. Compounds like 2,4-di-*tert*-butylphenol and elemicin showed significant binding and inhibitory effects on AChE and MAO-A. Additionally, 13 compounds, including α-asarone, penetrated the blood-brain barrier (BBB), indicating potential brain target binding, and eight compounds enhanced microglial β-amyloid phagocytosis, aiding in clearing AD pathological substances. Our results demonstrate the effectiveness of deep learning models in developing AD therapies and provide a strong platform for AD drug discovery.

1. Introduction

Approximately 50 million people worldwide experience dementia, and this number is expected to exceed 150 million by 2,050 [1,2]. Of these cases, approximately 70% are caused by Alzheimer's disease (AD), which is a progressive neurodegenerative disorder that clinically manifests as memory loss, cognitive decline, behavioral abnormalities, and language impairments and severely impacts patients' lives [3,4]. Pathologically, AD leads to the loss of synapses, neuroinflammation, tau and beta-amyloid protein (Aβ) metabolic abnormalities, and brain atrophy, ultimately resulting in death [5]. It has now become the seventh leading cause of death globally [2]. Despite years of extensive research, the options for drugs that effectively manage the progression of AD remain limited. To date, only seven drugs have been approved by the United States Food and Drug Administration for treatment. Four of these drugs, i.e., donepezil, rivastigmine, galantamine, and memantine, alleviate AD symptoms by inhibiting acetylcholinesterase (AChE) or *N*-methyl-D-aspartate (NMDA) receptors [6–9]. However, the therapeutic effect of these drugs is limited; moreover, they can cause side effects, such as headaches and nausea [10]. Therefore, there is an urgent need to develop new therapeutic agents to address this global health challenge.

Previous trials for the development of AD drugs have reported largely unfavorable outcomes, with a failure rate reaching 99.6% [11]. Every failed trial of a therapeutic agent for AD based on new molecular entities consumes a significant amount of time and resources. In contrast, screening potential AD therapeutic agents from existing compound libraries is more cost-effective. Machine

Peer review under responsibility of Xi'an Jiaotong University.
* Corresponding author. College of Pharmaceutical Engineering of Traditional Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, 301617, China.
** Corresponding author. College of Pharmaceutical Engineering of Traditional Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, 301617, China.
E-mail addresses: lizheng@tjutcm.edu.cn (Z. Li), hs_yu08@163.com (H. Yu).

https://doi.org/10.1016/j.jpha.2024.101022

learning (ML), which can be used to explore extensive chemical spaces, facilitate the acquisition of molecular characterizations, and extract shared features among pharmacological molecules, has been applied to prediction of diverse molecular properties [12]. In recent years, various pharmacological compounds for various disease domains have been screened using ML methods. Stokes et al. [13] used deep neural networks (DNNs) to predict halicin, a new antibiotic, in the Drug Repurposing Hub. Furthermore, Wong et al. [14] discovered small-molecule senolytic drugs using DNNs. Wang et al. [15] applied transfer learning to identify antiviral drugs that combat the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus from the ZINC15 database. Given these successful cases, we hypothesize that ML will be able to identify potential therapeutic agents for AD within currently available compound libraries.

The successful application of ML to the screening pharmacological molecules has typically relied on extensive compound libraries comprising tens of millions of compounds [16]. However, this approach demands substantial computational resources alongside low hit rates. To address this challenge, we employed DNN models to investigate the compounds of the classic traditional Chinese medicine formula (CTCMF), Kaixinsan (KXS) formula, which has a historical track record in AD treatment. In contrast to other extensive databases, KXS comprises only hundreds of compounds, which substantially reduces screening time while utilizing equivalent computational power. Moreover, all components in KXS consist of natural compounds, which simplifies acquisition and facilitates subsequent verification and application. Lastly, KXS boasts extensive human applications and has demonstrated therapeutic efficacy against AD [17]. Therefore, screening for AD therapeutic agents in KXS will likely significantly enhance hit rates while reducing experimental costs.

In our investigation, we initially trained four DNNs. One model was designed to predict therapeutic agents for AD, whereas the other three models were trained to predict inhibitors of specific AD targets. The selected targets were AChE, 5-hydroxytryptamine-6 (5-HT_6), and monoamine oxidase-A (MAO-A). AChE inhibitors have been shown to improve cognitive function in patients with AD, 5-HT_6 receptor inhibitors may enhance neuroprotection and reduce Aβ production, and MAO-A inhibitors have shown promise in mitigating AD progression [18–20]. Following model training, we carried out hyperparameter optimization, and the optimized models were used to screen for both therapeutic agents and target inhibitors for AD within the KXS formula. After screening, we conducted experiments to assess the specificity and efficacy of the model's predictions for compounds exhibiting high scores. Experiments included *in vitro* assays of enzyme activity, cell-based experiments, and *in vivo* assessments of brain absorption. Collectively, we sought to investigate the effectiveness and feasibility of employing DNNs for the development of drugs targeting AD.

2. Materials and methods

2.1. Datasets preparation

For model training, we used four proprietary compound libraries that were made in-house: the AD set, AChE set, MAO-A set, and 5-HT_6 set. Each dataset comprised compounds represented by their simplified molecular input line entry system (SMILES), which is a standardized notation for expressing the molecular structure of chemical compounds using American Standard Code for Information Interchange (ASCII) characters [21]. The AD set served as the training set for the AD model and consisted of compounds obtained from the Anti-Alzheimer's Disease Compound Libraries downloaded from MedChemExpress and Selleck [22,23]. Additionally, the AD set encompassed natural products derived from plants, animals, and microbes and was compiled from relevant literature sources. Subsequently, data deduplication was performed, which resulted in a dataset comprising 2,465 unique compounds. The AD set was then transformed into a binary classification dataset by categorizing activity values related to the compounds and AD. Compounds were assigned the label '1' to indicate their efficacy against AD when their median effective concentration (EC_{50}) and half-maximal drug inhibitory concentration (IC_{50}) values were below 50 μM. The final AD set containing 1,313 positives and 1,152 negatives was used to train the deep learning classification model for screening potential anti-AD compounds.

The AChE set, MAO-A set, and 5-HT_6 set serve as the training set for the AChE model, MAO-A model, 5-HT_6 model. The active compound libraries for AChE, MAO-A, and 5-HT_6 receptors were sourced from ChEMBL [24]. The 'Smile' column in the downloaded tables underwent deduplication to generate the initial AChE, MAO-A, and 5-HT_6 sets. Subsequent binary classification was performed on the downloaded datasets using 'assay description' and their corresponding activity values. Compounds showing inhibition of AChE, MAO-A, and 5-HT_6 targets were assigned a positive label '1', and the remaining compounds were designated as negative data with a label '0'. The final AChE, MAO-A, and 5-HT_6 sets comprised 1609, 1349, and 454 compounds, respectively. During the model training process, the above four datasets were randomly divided into 80% for training, 10% for testing, and 10% for validation.

The compounds in the KXS formula were used as a test set in the four models to predict potentially effective anti-AD compounds. These compounds were previously identified in our previous experiments using headspace (HS) gas chromatography (GC)-ion mobility spectrometry, HS solid-phase microextraction (SPME) GC-mass spectrometry (MS), and ultra-performance liquid chromatography-quadrupole-Orbitrap MS (UPLC-Q-Orbitrap-MS) [25,26]. A total of 272 compounds were identified for the test KXS set.

Table 1 [22–26] summarizes all datasets used in the study, and detailed information for these tables is available in Table S1.

2.2. Models training, optimization, and prediction

We developed four distinct models and used those to screen potential anti-AD compounds from the KXS set. This process followed three principal steps (Fig. 1A). 1) A comprehensive anti-AD model was trained using the AD set. Specific models targeting key AD-related receptors, namely, AChE, MAO-A, and 5-HT_6, were further developed using the AChE, MAO-A, and 5-HT_6 sets, respectively. 2) RDKit, Bayesian hyperparameter optimization, and ensemble methods were used to enhance and optimize the models' performance. 3) The four optimized models were used to screen potential anti-AD compounds from the KXS set. Compounds with the highest scores were subsequently selected for laboratory experimental evaluation.

We describe the methods of one of the models below; however, all four deep-learning models were trained using the same methods. The model was constructed using chemprop (https://github.com/chemprop/chemprop), a directed-message passing neural network (D-MPNN) for the prediction of molecular properties from the molecular graph structure, where atoms represent nodes and bonds represent edges [27]. In a D-MPNN, the molecular graph for each compound's SMILES string in the training sets is reconstructed using RDKit to identify the set of atoms and bonds. A feature vector is then initialized for each atom and bond, as described previously [13]. Subsequently, the model underwent multiple message-passing steps to aggregate information from neighboring atoms and bonds to construct a representation of local chemistry. During each message-

Table 1
Summary of the datasets.

Dataset	Description	Refs.
AD set	Contains 1313 compounds against AD and 1152 inactive compounds	[22,23]
AChE set	Contains 1177 compounds exhibiting inhibitory activity against the AChE receptor and 432 inactive compounds	[24]
MAO-A set	Contains 171 compounds exhibiting inhibitory activity against the MAO-A receptor and 1178 inactive compounds	[24]
5-HT_6 set	Contains 248 compounds exhibiting inhibitory activity against the 5-HT_6 receptor and 206 inactive compounds	[24]
KXS set	Contains 272 compounds identified in KXS formula	[25,26]

AD: Alzheimer's disease; AChE: acetylcholinesterase; MAO-A: monoamine oxidase-A; 5-HT_6: 5-hydroxytryptamine-6; KXS: Kaixinsan.

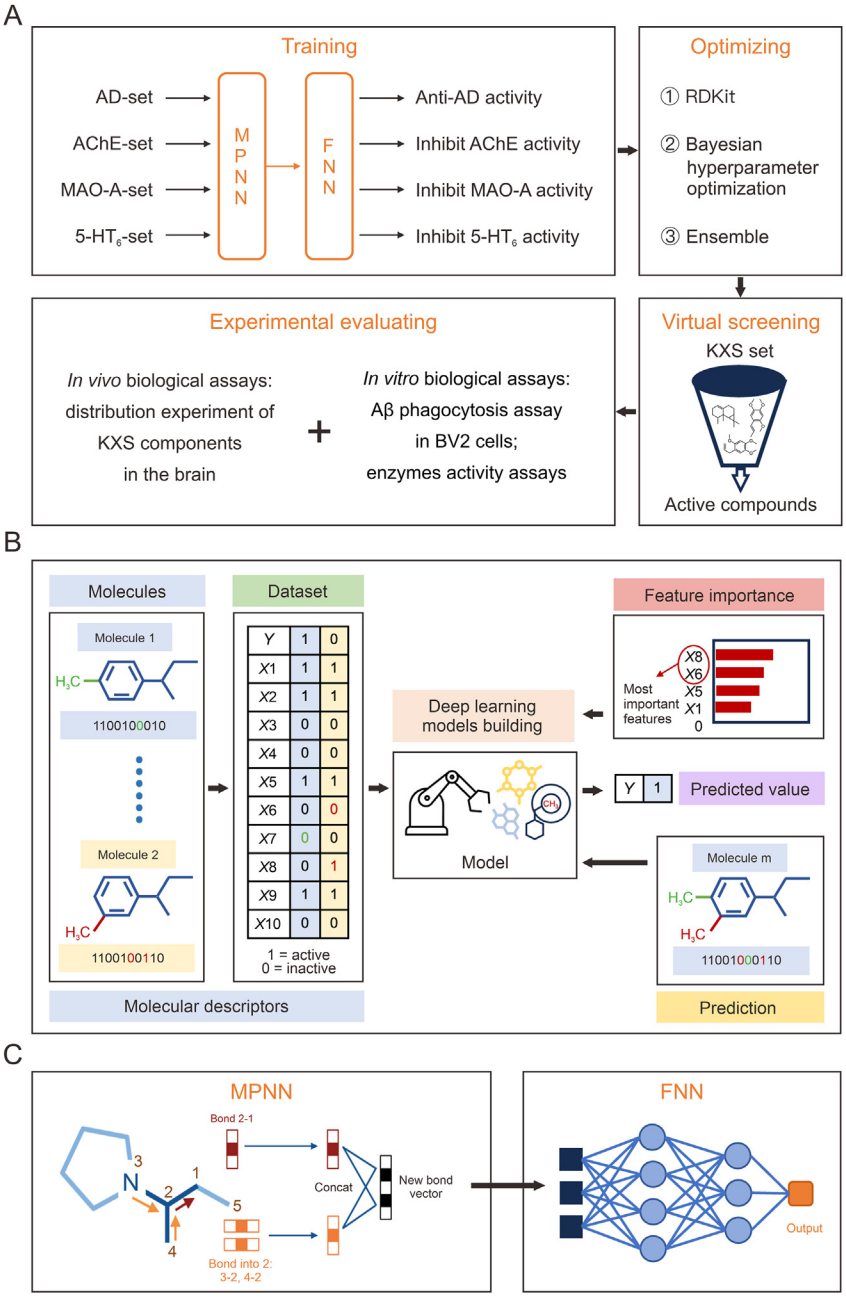

Fig. 1. The prediction models' process, details, and framework. (A) The workflow of virtual screening models based on deep learning for identifying anti-Alzheimer's disease (AD) compounds in the Kaixinsan (KXS) formula. (B) The details flow of the model trained and predicted process. In the green dataset part, blue line represents molecule 1 and yellow line represents molecule 2. *X* represents the pass bonds and edges features, *Y* is the X features corresponding labels. Using a set with an *X* tag of 1 as the molecular feature set of the interested properties, where different features have different weights in the set, a machine learning (ML) model is used to learn and distinguish the best feature weight set for the training set (red feature importance), construct a feature prediction model, and then predict the interested molecules in the model, using the *X* features of the molecules to predict the *Y* tag. (C) The model mainly consists of message passing neural network (MPNN) and fully connected neural network (FNN). MPNN is mainly used to learn the *X* features of molecules and construct feature models. FNN is mainly used for prediction after the new molecules after inputting in the model. AChE: acetylcholinesterase; MAO-A: monoamine oxidase-A; 5-HT_6: 5-hydroxytryptamine-6; Aβ: beta-amyloid protein.

passing step, the featurization of individual bonds underwent an updating process. This involved summing the features of neighboring bonds, applying a single neural network layer, concatenating the current bond's message with the sum, and applying a rectified linear unit (ReLU) activation function. Following a defined number of message-passing steps, the learned featurization across the molecule was summed to produce a single representation for the entire molecule. Finally, the resulting featurization was input into a feed-forward neural network, which generated a prediction for the property of interest. The details of the model prediction and framework are depicted in Figs. 1B and C, respectively.

Additionally, to enhance the performance of the model and address the limitation of message-passing steps in conveying information across large molecules, the learned feature from message passing was augmented with 200 additional features calculated at the molecular level using RDKit. The model was trained for 30 epochs on the training set, and the performance of the model was evaluated at each epoch using the validation set. The parameters from the epoch demonstrating the best performance on the validation set were chosen as the final model parameters. The model was then tested and evaluated on the test set. To optimize hyperparameters (i.e., the number of message-passing steps, the hidden size of the neural network, the number of feed-forward layers, and dropout probability), a Bayesian hyperparameter optimization approach was used during the initial stages of the training process. The ensemble method was applied to enhance model performance because empirical evidence has shown that it is effective in improving the performance of deep learning models [28]. The dataset was randomly split 20 times with different random seeds, and each split was used to train an individual model. The predictions of all 20 models were averaged to produce the final results.

Finally, we used the optimized models to make separate predictions from the KXS set. The resultant prediction was a real number ranging from 0 to 1, which reflected the probability of KXS set compounds being identified as potential anti-AD and inhibitors of AD targets. Further analysis will be carried out on the compounds predicted to be in the top 50 with scores exceeding 0.7.

2.3. In vitro biological assays: enzyme activity and Aβ phagocytosis assays

2.3.1. Determination of the inhibition rate of compounds on AChE

The enhanced Ellman's method [29] was used to assess the inhibitory effect of compounds with high model prediction scores on AChE. The purchasing details of the experimental samples and reagents are provided in Table S2. All samples were stored in dimethyl sulfoxide (DMSO; Beijing Solarbio Science & Technology Co., Ltd., Beijing, China), with galantamine (Shanghai Yuanye Biotechnology Co., Ltd., Shanghai, China) serving as the positive control sample. We added 145 μL of phosphate-buffered saline (PBS) (pH 8.0; Beijing Solarbio Science & Technology Co., Ltd.), 20 μL of AChE (C0663; Sigma-Aldrich, St. Louis, MO, USA) solution (0.1 U/mL, soluble in PBS), and 5 μL of different concentrations of the sample/reference substance/standard solution to each 96-well plate well. The standard solution did not contain the sample, but its DMSO concentration matched that of the sample group. We preincubated the samples at 37 °C for 20 min. Next, we added 10 μL of acetylthiocholine iodide (0.2 mmol/L; Shanghai Yuanye Biotechnology Co., Ltd.) and 10 μL of 5,5′-dithiobis-(2-nitrobenzoic acid) (0.25 mmol/L; Shanghai Yuanye Biotechnology Co., Ltd.) to each well, mixed thoroughly, and incubated at 37 °C for 30 min. We then stopped the enzymatic reaction by adding 10 μL of 0.4% sodium dodecyl sulfate (SDS) and measured the absorbance at 405 nm using a SpectraMax iD5 microplate reader (Molecular Devices, San Jose, CA, USA). The AChE inhibition rate was calculated using the following formula: $(A_{standard} - A_{sample})/(A_{standard} - A_{black}) \times 100\%$, where $A_{standard}$ is the absorbance of the standard solution, A_{sample} is the absorbance of the sample and reference, and A_{black} is the absorbance of the standard solution without AChE. Each experiment was conducted in triplicate.

2.3.2. Determination of the inhibition rate of compounds on MAO-A

The inhibitory effect of compounds with high predicted scores on MAO-A was assessed using the enhanced Weissbach's method [30]. The purchasing information for the experimental samples and reagents is provided in Table S2. The samples were stored in DMSO, and clorgyline (Sigma-Aldrich) was used as the control. To each well of a 96-well plate, we added 158 μL of potassium phosphate buffer (PPB; 0.2 mol/L, pH 7.4; Shanghai Yuanye Biotechnology Co., Ltd.), 30 μL of MAO-A (M7316; Sigma-Aldrich) solution (0.09 mg/mL, soluble in PPB), and 2 μL of sample/reference/standard solution. The standard solution did not contain the sample, but the DMSO concentration remained consistent with the sample group. We preincubated the mixture at 37 °C for 30 min. We then added 20 μL of kynuramine (0.5 mmol/L; Sigma-Aldrich) and mixed well for 30 s. We measured the initial absorbance value (A_1) of the substrate and sample mixture at a wavelength of 360 nm, followed by incubation at 37 °C for 60 min and measuring the absorbance value again (A_2). The change in absorbance value of the standard control was denoted as ΔA_0. Finally, we calculated the monoamine oxidation inhibition rate (%) using the following formula: $1 - [(A_1 - A_2)/\Delta A_0] \times 100\%$. Each experiment was conducted in triplicate.

2.3.3. Affinity detection among compounds with AChE and MAO-A

In this study, surface plasmon resonance (SPR) interaction experiments were conducted at 25 °C using the BIAcore T200 system (GE Healthcare, Chicago, IL, USA).

For the AChE binding assay, AChE (C2888; Sigma-Aldrich) was initially dissolved in PBS to create a stock solution at a concentration of 1 mg/mL. This stock solution was then diluted with sodium acetate buffer (100 mM, pH 4.5) to achieve a working concentration of 50 μg/mL. CM7 sensor chip (Cytiva, Marlborough, MA, USA) was coupled with AChE using an amine coupling kit (GE Healthcare). Methyl isoeugenol, 2,4-di-*tert*-butylphenol, 7-*O*-methylmangiferin, sibiricaxanthone B, onjixanthone I, camphor, and galantamine were diluted to various concentrations for testing by PBS (Beijing Solarbio Science & Technology Co., Ltd.). Thereafter the diluted samples were introduced to the sensor chip with a contact time of 60 s, followed by a dissociation phase at a flow rate of 30 μL/min for 60 s.

For the MAO-A binding assay, the recombinant MAO-A from human was initially dissolved in PBS to create a stock solution at a concentration of 1 mg/mL. The stock solution was diluted with sodium acetate buffer (100 mM, pH 4.5) to achieve a working concentration of 50 μg/mL. CM7 sensor chip was coupled with MAO-A using an amine coupling kit. Methyl eugenol, estragole, elemicin, and clorgyline were diluted to various concentrations by PBS-P buffer. Then the diluted samples were introduced to the sensor chip with a contact time of 90 s, followed by a dissociation phase at a flow rate of 30 μL/min for 120 s.

The 1:1 Langmuir binding model in the Biacore evaluation software (T200 version 1.0, GE Healthcare) was then used to calculate the affinity constants.

2.3.4. Cell culture

BV2 cells were generously provided by the Professor Hong Guo's group at Tianjin University of Traditional Chinese Medicine, China. They were maintained in Dulbecco's modified Eagle's medium (DMEM; Gibco, Grand Island, NY, USA), supplemented with 10% fetal bovine serum (FBS; Gibco) and 1% penicillin/streptomycin (Gibco). The cells were cultured in a humidified incubator

(Shanghai Lishen Scientific Equipment Co., Ltd., Shanghai, China) at 37 °C with 5% CO_2.

2.3.5. Preparation of fluorescein isothiocyanate labeled-$A\beta_{1-42}$ (FITC-$A\beta_{1-42}$) peptide

Initially, 1 mg of FITC-$A\beta_{1-42}$ powder (GL Biochem (Shanghai) Ltd., Shanghai, China) was dissolved in 220 μL of hexafluoroisopropanol (HFIP; Sigma-Aldrich) to obtain a solution with a concentration of 1 mM. The solution was subjected to a 10-min ultrasound treatment in a water bath and subsequently incubated in the dark inside a fume hood for 24 h until the HFIP evaporated. We then used 44 μL of DMSO to redissolve FITC-$A\beta_{1-42}$ and prepared a 5-mM solution. The solution was subsequently stored in a freezer at −20 °C until further use.

2.3.6. Aβ phagocytosis assay in BV2 cells

Upon reaching 90% confluency, the cells were trypsinized and subsequently seeded onto a six-well plate (BIOFIL, Shanghai, China) at an appropriate density. The cells were then incubated at 37 °C for 24 h. Various compounds (Shanghai Yuanye Biotechnology Co., Ltd.) from the KXS formula were added to the designated wells according to Table S2. The control group was treated with DMEM solution only, and triplicate wells were established for each experimental group. The cells were then placed in a 5% CO_2 incubator and allowed to incubate for an additional 24 h. The culture medium was aspirated from each well, and the cells were incubated with 1 μM FITC-$A\beta_{1-42}$ solution in the dark for 2 h. After the incubation period, the FITC-$A\beta_{1-42}$ solution was removed, and the cells underwent two gentle washes with PBS to remove any residual extracellular FITC-$A\beta_{1-42}$. Finally, 1 mL of PBS was added to each well, and the cells were observed using an inverted fluorescence microscope (DS-Ri2; Nikon, Tokyo, Japan). The ImageJ software (National Institutes of Health, Bethesda, MD, USA) was used for the analysis of fluorescence phagocytic content.

2.4. In vivo biological assays: distribution experiment of KXS compounds in the brain

2.4.1. Preparation of freeze-dried KXS formula powder

The following four Chinese herbal pieces were purchased from Anhui Yuankang Chinese Herbal Medicine Co., Ltd. (Bozhou, China). We combined powdered *Ginseng Radix et Rhizoma* (180 g), *Polygala Radix* (180 g), *Poria* (360 g), and *Acorus tatarinowii Rhizoma* (180 g) after each powder was passed through a number six sieve. The mixture was soaked in 75% ethanol water at 10 times the volume for 1 h, followed by two 2-h heat-reflux extraction cycles. The mixture was filtered, and the supernatant was collected and filtered again. The filtrates were combined and evaporated at 60 °C until dry. The concentrated solution was pre-frozen at −80 °C for 24 h and transferred to a freeze-drying machine for three days to produce freeze-dried KXS powder. We prepared a 4-g/mL aqueous suspension solution as a backup.

2.4.2. Sample collection of Sprague-Dawley (SD) rat brains

We used male SD rats from Beijing Vital River Laboratory Animal Technology Co., Ltd. (Beijing, China) for the experiment. The rats were acclimated for one week at room temperature (20 ± 5 °C) and a relative humidity of 40%–60%, with a 12-h light-dark cycle and free access to food and water. The rats were fasted for 12 h before the experiment. The 27 rats were divided into nine groups, with three rats in each group. During the experiment, the control group was administered carboxymethyl cellulose sodium, and the other eight groups were administered an aqueous suspension solution of freeze-dried KXS powder at a dose of 10 mL/kg (4 g/mL), which was equivalent to a clinically effective dose [17]. The solutions were orally administered to the rats for seven consecutive days. The rats then fasted for 12 h before the gavage on the 8th day, with free access to water. On the 8th day, the 24 rats in the treatment groups were perfused with normal saline into the heart at 0.167, 0.5, 1, 1.5, 2, 3, 6, and 12 h after gavage. The control group rats were also perfused with normal saline into the heart on the 8th day immediately after gavage. After perfusion, whole brain tissue was collected, stored in a cryotube, placed in liquid nitrogen, and stored at −80 °C for later use. All animal studies were approved by the Laboratory Animal Ethics Committee of Tianjin University of Traditional Chinese Medicine, China (Permit No.: TCM-LAEC2022230).

2.4.3. Preprocessing of SD rat brain tissue samples

Before the GC-MS experiment, the frozen brain tissue samples obtained in Section 2.4.2 were thawed and homogenized by adding three times the amount of physiological saline. Subsequently, the brain tissue homogenates from the control and experimental groups were combined according to the time groupings specified in Section 2.4.2. At each time point, 100 μL of brain tissue was homogenized from both the control and experimental groups for subsequent analysis using GC-MS. The samples were injected using HS-SPME. Three sets of samples were prepared for each group to facilitate three experimental analysis repeats.

Before the UPLC-Q-Orbitrap-MS analysis, frozen brain tissue samples obtained in Section 2.4.2 were thawed and homogenized in three times the volume of physiological saline. The brain tissue homogenates from the control and experimental groups were subsequently pooled according to the abovementioned time points, with 100 μL taken from each group at every time point. Then, 600 μL of methanol solution was added, and the samples were vortexed for 5 min before centrifugation at 14,000 rpm for 10 min. The resulting supernatant was collected, subjected to nitrogen drying at 4 °C, and redissolved in 50% methanol. After vortexing for 5 min and centrifugation at 14,000 rpm for 10 min, the supernatant was extracted for further analysis. Each sample group was triplicated, and subsequent analyses were repeated three times.

2.4.4. Preparation of reference solution

Ten reference compounds were prepared into 1 mg/mL mother liquor using ethyl acetate (Fisher, Fair Lawn, NJ, USA) for the GC-MS analysis. The mixture was then diluted to 1 μg/mL in a mixed standard solution. The brain tissue homogenate was added to the solution, and the mixture was thoroughly vortexed. We then applied the HS-SPME-GC-MS sample pre-treatment method outlined in Section 2.4.3, and 1 μL was obtained for analysis.

For the UPLC-Q-Orbitrap-MS analysis, 35 reference compounds were prepared into 1 mg/mL mother liquor using pure methanol (Fisher). We added the blank brain tissue homogenate to the mixed standard solution at a concentration of 1 μg/mL and vortexed thoroughly to ensure proper mixing. The sample was then processed using the UPLC-Q-Orbitrap-MS method as described in Section 2.4.3, and 1 μL was obtained for injection analysis. Details of the reference substance are available in Table S2.

2.4.5. Profiling KXS compounds in SD rat brain tissue using HS-SPME-GC-MS

The sample solution presented in Section 2.4.3 and the mixed standard solution described in Section 2.4.4 were utilized for HS-SPME-GC-MS analysis, with the mixed standard solution serving as the control group. Each group of sample solutions were prepared in triplicate for analysis. The analytical procedure is outlined below.

Solid-phase extraction conditions were established as follows: an extraction column with 65 μM polydimethylsiloxane/divinylbenzene (PDMS-DVB) was selected, and extraction was conducted at 60 °C for 30 min. The GC system (7980D; Agilent Technologies, Santa Clara,

CA, USA) used an HP-5MS capillary column (30 m × 0.25 mm, 0.25 μm; Agilent Technologies). Helium served as the carrier gas with a flow rate of 1 mL/min. The injection port temperature was set to 250 °C, and we used a solvent delay of 10 min. The injection volume was 1 μL. The programmed temperature gradient is detailed in Table S3. The spectrometer (7000D; Agilent Technologies) for MS was operated in the positive ion mode of the electron ionization source, with the ion source temperature set to 230 °C and ionization energy set to 70 eV. The scanning range was adjusted to *m/z* 50–600, and data processing was performed using the Masshunter software (Agilent Technologies).

2.4.6. Profiling KXS compounds in SD rat brain tissue using UPLC-Q-Orbitrap-MS

The sample solution presented in Section 2.4.3 and the mixed standard solution described in Section 2.4.4 were used for the UPLC-Q-Orbitrap-MS analysis, with the mixed standard solution serving as the control group. Each group of sample solutions was prepared in triplicate for analysis.

The liquid phase system used in this experiment was a Waters ACQUITY™ UPLC I Class system (Waters Corporation, Milford, MA, USA). The pre-column used was ACQUITY UPLC® BEH C_{18} VanGuard™ column (2.1 mm × 5 mm, 1.7 μm). We used the ACQUITY UPLC® BEH C_{18} (2.1 mm × 100 mm, 1.7 μm) chromatographic column. The column temperature was set to 40 °C, and the mobile phase consisted of 0.1% formic acid water (A) and 0.1% formic acid methanol (B). The elution gradient was as follows: 0–10.0 min, 98%–80% A; 10–18 min, 80%–64% A; 18–20 min, 64%–58% A; 20–30 min, 58%–48% A; 30–37 min, 48%–39% A; 37–48 min, 39%–31% A; and 48–53 min, 31%–2% A. The flow rate was set to 0.3 mL/min, and the injection volume was 1 μL. For MS, the Q Exactive™ hybrid quadrupole Orbitrap mass spectrometer (Thermo Fisher Scientific Inc., San Jose, CA, USA) was used with a heated electrospray ionization (HESI) ion source in negative ion mode. The sheath gas flow rate was set to 35 arb, and the spray voltage was set to −3.0 kV. The capillary and auxiliary gas temperatures were set to 350 °C, and the normalized energy was set to 20, 30, and 50 eV. Data were acquired using the Full MS/data-dependent (dd)-MS^2 data collection mode for non-targeted analysis, and subsequent processing of the raw data was performed using the Compound Discover 3.1 software (Thermo Fisher Scientific Inc., Waltham, MA, USA).

3. Results

3.1. The training and optimization results of the four AD-related models

Our model training was conducted according to an established algorithm framework [27], and our objective was to develop deep-learning models that could predict the anti-AD activity of uncharacterized compounds. Chemprop is a DNN framework designed to predict molecule properties based on chemical structures. It has been shown to accurately predict a range of properties, encompassing antibiotic activity, drug toxicity, senolytic activity, and drug side effects [13,14]. Four deep-learning models (i.e., AD, AChE, MAO-A, and 5-HT_6 models) related to AD were trained using the chemprop framework with personalized datasets (i.e., AD, AChE, MAO-A, and 5-HT_6 sets).

Initially, we applied Bayesian hyperparameter optimization to determine the optimal hyperparameters. The resulting hyperparameters of each model are presented in Table 2. Subsequently, the predictive ability of each model for anti-AD activity or inhibiton of AD-related targets was evaluated by training, testing, and validating on randomly segmented subsets of the self-built database, with a distribution of 80% for training, 10% for testing, and 10% for validation. However, the AD model exhibited a relatively low receiver operating characteristic curve-area under the curve (ROC-AUC) of 0.771 on the test set. To improve the performance of the models, we integrated additional vectors containing molecular features calculated using RDKit. These vectors were then concatenated with the molecular representations generated by the chemprop module. The new models all showed a significant improvement in ROC-AUCs on the test sets, which indicated that the enhanced data representation greatly enhanced model performance. Because of the limited size of the training data, test performance varied substantially when trained on different data splits. To bolster the robustness of the models, we constructed an ensemble of 20 models and evaluated their performance. The findings revealed a notable enhancement in the model's performance via the application of ensemble methods. The four final models demonstrated optimal performance on the testing sets, as indicated in Table 2. Moreover, the high performance of the models indicated that these models could effectively distinguish between compounds that are active and inactive against AD. The models were all subjected to 10-fold cross-validation, and the final optimized ROC-AUC plots of the four models shown in Fig. S1.

3.2. Prediction results of KXS formula screened by four models

The KXS formula is a CTCMF with a long historical background that offers a robust foundation for treating AD. We used four models to predict 272 identified compounds within the KXS formula. This enabled the discovery of compounds that are potentially active against AD, the exploration of specific KXS formula compounds that show high efficacy, and the identification of the primary targets of these active compounds. The predicted results are available in Table S4. We then established criteria for screening compounds: the prediction of a score exceeding 0.70 and a top 50 ranking by a minimum of two models. We aimed to narrow the screening range for AD-active compounds and identify compounds with relatively high predictive activity. Ultimately, a total of 12

Table 2
Hyperparameter value and area under the receiver operating characteristic curve (ROC-AUC) scores of the four models.

Hyperparameter	AD model	AChE model	MAO-A model	5-HT_6 model
Number of message-passing steps	5	2	6	6
Neural network hidden size	1400	1400	2000	2300
Number of feed-forward layers	3	3	1	3
Dropout probability	0.3	0.35	0.15	0.05
ROC-AUC scores of models under different optimization methods	–	–	–	–
Bayesian optimization	0.771	0.960	0.988	0.94
Bayesian optimization + RDKit	0.961	0.988	0.99	0.966
Bayesian optimization + RDKit + ensemble	0.966	0.996	1.0	0.970

–: no data. AD: Alzheimer's disease; AChE: acetylcholinesterase; MAO-A: monoamine oxidase-A; 5-HT_6: 5-hydroxytryptamine-6.

compounds were identified that satisfied both criteria. The ranking outcomes for the compounds are presented in Table 3.

To comprehensively assess the screening outcomes, we performed a Venn diagram analysis on the 12 compounds. The corresponding graphical representation is shown in Fig. 2. Notably, three models simultaneously predicted four compounds: estragole, methyl isoeugenol, methyl eugenol, and elemicin. In previous work, all four compounds have demonstrated good activity against neurological diseases. Of these, estragole, a naturally occurring compound, serves as the primary constituent in certain aromatic plants and has been studied for its potential anti-inflammatory and antioxidant properties. Research has suggested that estragole can reduce neuroinflammation and oxidative stress reactions, which may have a protective effect on the nervous system [31]. Methyl isoeugenol and methyl eugenol are structurally similar compounds derived from eugenol, a natural compound found in certain plants and has been investigated for its anti-AD effects [32]. Methyl eugenol has been found to exhibit antioxidant properties by eliminating free radicals, increasing antioxidant levels, and inhibiting the release of inflammatory factors [33]. These properties likely contribute to its potential to treat inflammation and delay aging. Elemicin is a natural compound belonging to the phenylpropanoid class and has been studied for its antioxidant and anti-inflammatory effects. Research has shown that elemicin can pass through the blood-brain barrier (BBB), which indicates its ability to affect the brain [34].

Additionally, as shown in Fig. 2, a total of eight compounds satisfied established criteria and were predicted by two models. These compounds include 2,4-di-*tert*-butylphenol, α-asarone, β-asarone, γ-asarone, calarene, acoramone, onjixanthone I, and β-caryophyllene. Previous research has shown that 2,4-di-*tert*-butylphenol exhibits neuroprotective effects, which are attributed to the compound's antioxidant capacity. Specifically, this compound effectively protects PC12 cells from neurotoxicity induced by H_2O_2 and Aβ, which contributes to the deceleration of cognitive decline and the progression of AD [35]. The antioxidant properties of 2,4-di-*tert*-butylphenol play a crucial role in mitigating the detrimental effects of oxidative stress on neuronal cells, underscoring its potential as a therapeutic agent for AD. α-asarone, β-asarone, and γ-asarone are natural compounds characterized by their three highly similar chemical structures. α-asarone and β-asarone are isomers of each other, and both are composed of a benzene ring and an oxygen heterocycle. In contrast, γ-asarone is an isomer of these three compounds and featues two benzene rings and a more intricate structure. In animal models of AD, α-asarone has been demonstrated to enhance cognitive function and memory [36]. Similarly, β-asarone can inhibit Aβ by stimulating autophagy in AD cell models and promote synaptic plasticity by upregulating the expression of synaptophysin (*SYP*) and glutamate receptor 1 (*GluR1*), thereby mitigating neural damage in AD [37]. β-caryophyllene is a naturally occurring bicyclic sesquiterpene compound that has shown promise in reducing neuroinflammatory reactions associated with AD by activating cannabinoid receptor 2. This unique property of β-caryophyllene makes it an attractive molecule for the development of new drugs aimed at treating AD [38].

We quantified the chemical relationships among 12 molecules and the other molecules in the KXS formula using the Tanimoto similarity, which measures the proportion of shared chemical substructures. The Morgan fingerprints for each compound were computed using RDKit to obtain the Tanimoto similarity. Subsequently, *t*-distributed stochastic neighbor embedding (*t*-SNE) analysis was performed to generate a *t*-SNE map via scikit-learn's implementation of *t*-SNE (Fig. 3). Each point in the map corresponds to a compound, with the inter-point distances indicating the Tanimoto similarity among compounds. A shorter distance signifies a higher proportion of shared substructures between

Table 3
The ranking of compounds predicted by the various models.

Compounds	AD model	AChE model	MAO-A model	5-HT_6 model
Estragole	8	–	1	9
Methyl isoeugenol	10	36	–	3
Methyl eugenol	7	–	2	2
2,4-Di-*tert*-butylphenol	47	17	–	–
α-asarone	19	–	–	8
β-asarone	33	–	–	7
γ-asarone	13	–	–	6
Calarene	30	32	–	–
Elemicin	18	–	9	1
Acoramone	–	23	4	–
Onjixanthone I	–	7	–	11
β-Caryophyllene	24	39	–	–

AD: Alzheimer's disease; AChE: acetylcholinesterase; MAO-A: monoamine oxidase-A; 5-HT_6: 5-hydroxytryptamine-6.

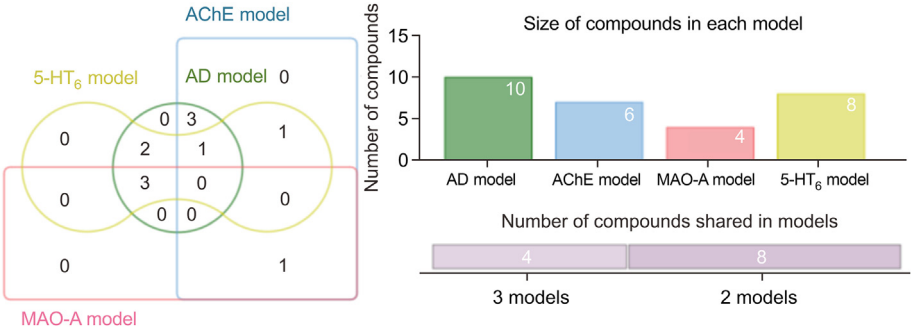

Fig. 2. Venn diagram of the prediction results of four models. AChE: acetylcholinesterase; 5-HT_6: 5-hydroxytryptamine-6; AD: Alzheimer's disease; MAO-A: monoamine oxidase-A.

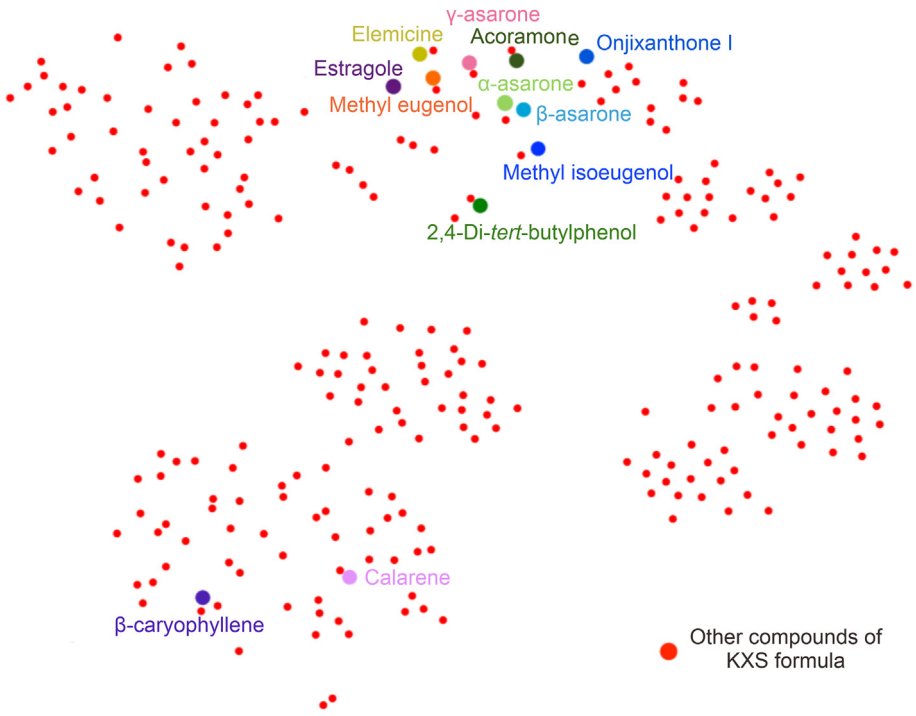

Fig. 3. The *t*-distributed stochastic neighbor embedding (*t*-SNE) map of Kaixinsan (KXS) formula. The graph illustrates the chemical spatial relationships among compounds in KXS formula. Compounds with high predicted scores are accentuated through larger dots, wherein the proximity of the dots signifies heightened chemical similarity.

compounds. Fig. 3 illustrates the chemical spatial distribution of KXS formula. All predicted active compounds, apart from calarene and β-caryophyllene, exhibited comparable and smaller distances, which suggested a higher proportion of shared chemical substructures. Thus, these common substructures may possess anti-AD activity, which warrants further in-depth investigation.

In conclusion, numerous studies have demonstrated that the compounds screened by the models have anti-AD effects or potentially beneficial effects on AD, which substantiates the efficacy of the screening models. Thus, our findings offer a valuable reference and support for developing therapeutic drugs for AD and various other diseases.

3.3. Validation of the in vitro assays for the predicted results

To validate the effectiveness and specificity of the model's predictions, enzyme- and cell-level validations were performed for the predictions of the AChE, MAO-A, and AD models.

3.3.1. AChE inhibition assay

The excessive activation of AChE in the brains of AD patients leads to a reduction in acetylcholine levels, which, in turn, impairs synaptic transmission function. Therefore, inhibiting AChE activity is considered as a viable strategy for slowing AD progression [18]. Our experiment validated the efficacy of the AChE model in predicting potential compounds using *in vitro* experiments aimed at inhibiting AChE.

A total of seven commercially available compounds with prediction scores greater than 0.9 were used for validation in the AChE model inhibition test. Galantamine served as the positive control group. The experimental findings are depicted in Fig. 4A. Notably, 2,4-di-*tert*-butylphenol showed a markedly high inhibitory rate and potent inhibitory effect on AChE. At a concentration of 25 μM, its inhibitory rate surpassed that of galantamine. Furthermore, in line with the model's predictions, both camphor and methyl palmitate demonstrated AChE inhibition rates exceeding 80%, surpassing that of galantamine. Although the inhibitory effect of 7-*O*-methylmangiferin and sibiricoxanthone B on AChE was relatively low, they still exhibited some degree of inhibition. With the exception of onjixanthone I, all compounds exhibited higher inhibition rates than the negative control estragole, which further affirmed the relative effectiveness and accuracy of the AChE model.

The binding affinity between the compounds and AChE was measured using a SPR assay. The dissociation constant (K_D) values of the compounds, determined based on the affinity fitting curves, ranged from 10^{-5} to 10^{-8} M, indicating strong binding ability to AChE. The IC_{50} and K_D values of the compounds are shown in Table 4.

3.3.2. MAO-A inhibition assay

MAO-A plays a key role in neurotransmitter metabolism. Excessive activation of MAO-A can result in abnormal reductions in neurotransmitter levels, thereby contributing to the onset and progression of AD. Inhibiting MAO-A activity can attenuate neurotransmitter degradation, elevate neurotransmitter levels, and ameliorate neurological function and symptoms in AD patients [19]. The anti-MAO-A effects of the compounds predicted by the MAO-A model were validated using *in vitro* inhibition experiments to evaluate the model's accuracy.

In this study, three commercially available compounds were selected for experimental validation from 10 compounds with prediction scores exceeding 0.7. Clorgyline served as the positive control. The experimental results are shown in Fig. 4B. With increasing sample concentration, the inhibition rate of MAO-A showed a pronounced dose-dependent increase. The inhibition rate of MAO-A within the body should not be excessively high because MAO-A is responsible for the regulation of the human nervous system. Excessive inhibition can lead to an increase in dopamine, which can induce various symptoms, such as accelerated respiration and cardiac arrhythmia. Therefore, an inhibition rate greater than 15% for MAO-A indicates biologically significant inhibition [39]. Among the

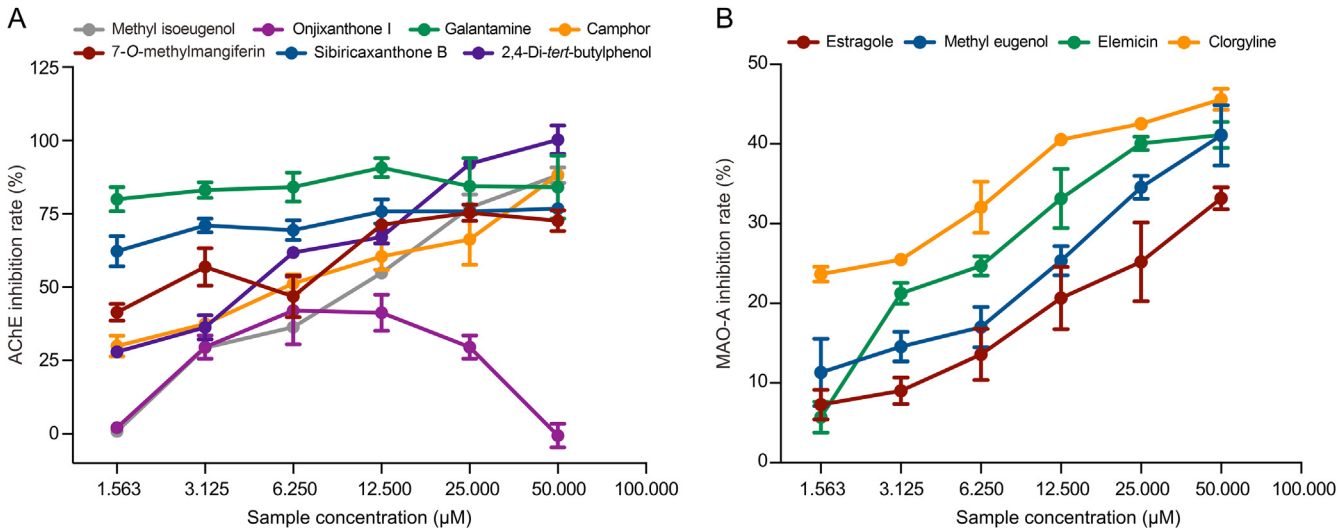

Fig. 4. Inhibition rates of different compounds on acetylcholinesterase (AChE) and monoamine oxidase A (MAO-A). (A) Inhibition rate of AChE by compounds at different concentrations. (B) Inhibition rate of MAO-A by compounds at different concentrations.

four samples, the highest inhibition rates (all exceeding 15%) were observed at the final concentration of 50 μM, which indicated a significant inhibitory effect. The group treated with clorgyline consistently demonstrated the highest inhibition rate, and the elemicin group exhibited a comparable inhibition rate. Conversely, the negative control group exhibited consistently lower inhibition rates than other groups, which validated the accuracy of the MAO-A model predictions. Furthermore, given the inhibition and concentration increase up to 50 μM, the decision to not further increase the concentration was based on the established inhibitory effect of MAO-A. Although elevating the concentration could potentially enhance the *in vitro* inhibition rate of MAO-A, such effect may not translate directly to *in vivo* animal experiments.

The K_D between the compounds and MAO-A were measured using the SPR method. The K_D values ranged from 10^{-6} M, indicating a strong binding ability to MAO-A. The IC_{50} and K_D values of the compounds are presented in Table 4. Further investigation is needed to extend these findings.

3.3.3. *Validation of compounds in the KXS formula in* in vitro *microglial cell phagocytosis*

The Aβ protein forms aggregates and plaques in the brains of individuals with AD and is one of the pathological hallmarks of AD. These aggregates can be highly toxic to nerve cells, causing oxidative stress, disruption of cell signaling pathways, and the accumulation of hyperphosphorylated tau protein in the brain. Consequently, this damage to nerve cells contributes to neuroinflammation and the progression of AD [40]. BV2 cells are a glial cell line derived from the cerebellum of mice. It is commonly used in the study of neuroinflammation and neurodegenerative diseases. BV2 cells are macrophage-like cells with phagocytic capabilities, which allow them to engulf and eliminate any surrounding abnormal proteins and cellular debris. When BV2 cells phagocytose Aβ, they aid in the clearance of excessive Aβ deposits in the brain. This clearance mechanism helps to mitigate the toxic effects caused by Aβ and prevent the formation of plaques, ultimately slowing down the progression of AD [41]. We used an AD model of BV2 cells induced by oligomeric FITC-Aβ_{1-42} to investigate the effects of the compounds predicted by the model on cells' ability to phagocytose Aβ.

We evaluated the effects of the 11 compounds of the KXS formula on the phagocytic function of BV2 cells. There were three replicates for each group, including a blank control group. The phagocytic ability of BV2 cells was assessed by measuring fluorescence intensity after the phagocytosis assay. Higher fluorescence intensity indicated stronger phagocytic ability of the BV2 cells. Fig. 5 shows the experiment results. The BV2 cells exhibited significantly enhanced phagocytic ability than the control group when exposed to elemicin disturbance. Additionally, a-asarone, 2,4-di-*tert*-butylphenol, β-asarone, and estragole demonstrated notable improvements in phagocytic activity and promising

Table 4
Inhibitory effects of compounds toward acetylcholinesterase (AChE) and monoamine oxidase-A (MAO-A).

Compounds	AChE IC_{50} (μM)	Affinity K_D (M)	MAO-A IC_{50} (μM)	Affinity K_D (M)
Estragole	–	–	144.43 ± 7.43	2.349×10^{-6}
Methyl isoeugenol	9.682 ± 0.353	1.847×10^{-8}	–	–
Methyl eugenol	–	–	96.06 ± 9.87	5.828×10^{-6}
2,4-Di-*tert*-butylphenol	4.512 ± 0.191	1.334×10^{-7}	–	–
Camphor	6.230 ± 0.567	2.515×10^{-8}	–	–
7-*O*-methylmangiferin	2.985 ± 0.245	1.691×10^{-5}	–	–
Sibiricaxanthone B	0.119 ± 0.102	1.199×10^{-6}	–	–
Elemicin	–	–	62.24 ± 1.08	4.686×10^{-6}
Onjixanthone I	13.853 ± 1.717	8.770×10^{-5}	–	–
Galantamine	0.139 ± 0.012	1.211×10^{-5}	–	–
Clorgyline	–	–	59.38 ± 4.51	4.344×10^{-6}

IC_{50}: half maximal inhibitory; K_D: dissociation constants.

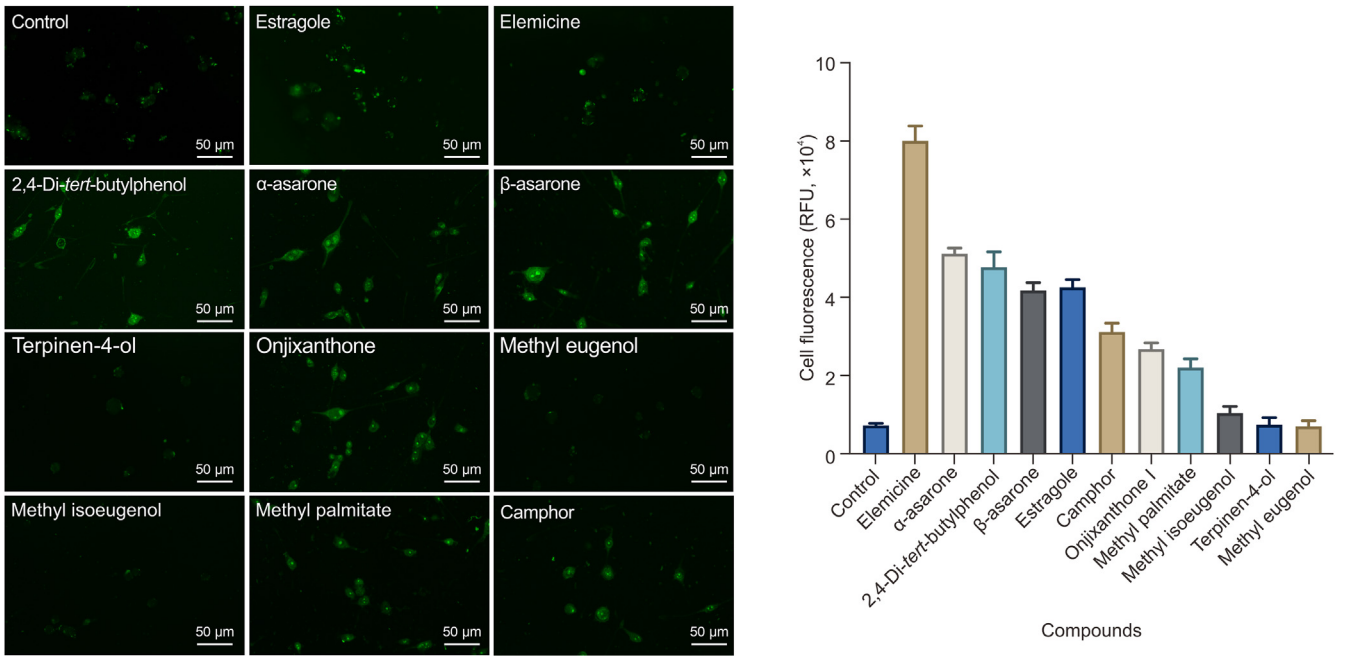

Fig. 5. Phagocytic activity of BV2 cells towards fluorescein isothiocyanate labeled beta-amyloid protein $(A\beta)_{1-42}$ (FITC-$A\beta_{1-42}$) under the influence of various compounds. The left image displays fluorescence microscopy observations of BV2 cells phagocytizing FITC-$A\beta_{1-42}$ in the control group and those treated with various compounds. The right image illustrates the fluorescence intensity values of BV2 cells phagocytizing FITC-$A\beta_{1-42}$ following treatment with diverse compounds. Higher values indicate increased phagocytosis of FITC-$A\beta_{1-42}$. RFU: relative fluorescence unit.

outcomes in our model predictions. However, methyl isoeugenol, terpiene-4-ol, and methyl eugenol exhibited weaker phagocytic ability towards BV2 cells than that of the aforementioned compounds, albeit superior to that of the control group.

It is worth noting that the compounds showing stronger BV2 cell phagocytic ability, particularly those ranked high, were mostly predicted by the 5-HT_6 model. The 5-HT_6 model was designed to predict the activity of 5-HT_6 receptor inhibition. Studies have indicated that the activation of the 5-HT_6 receptor promotes the production and release of Aβ, which leads to an increased accumulation of Aβ in the brain. In contrast, administering antagonists that target the 5-HT_6 receptor has demonstrated potential therapeutic effects in AD. These antagonists mitigate the production and release of Aβ, which, in turn, improves cognitive and memory dysfunctions linked to AD. Taken together, our findings demonstrate the accuracy of the 5-HT_6 model [42].

3.4. Validation of compounds in the KXS formula in in vivo rat brain absorption

The absorption of compounds by the brain is crucial for their application as a therapeutic agent for AD. We conducted a comprehensive analysis of brain-penetrating compounds in the KXS formula using a combination of UPLC-Q-Orbitrap-MS and HS-SPME-GC-MS techniques. A total of 23 brain-absorbing compounds were successfully identified, with six non-volatile compounds detected via UPLC-Q-Orbitrap-MS and 17 volatile compounds identified via HS-SPME-GC-MS. Detailed findings, encompassing a thorough overview of the identified compounds, are presented in Tables 5 and 6.

Seventeen brain compounds were detected using GC-MS. Among these, β-asarone and α-asarone were found in brain samples, which indicated their ability to penetrate the BBB owing to their small molecule size and fat solubility. α-asarone has been shown to improve cognitive function in APPswe/PSEN1dE9 (*APP*/*PS1*) transgenic mice by reducing hippocampal $A\beta_{42}$ and p-tau protein levels, suppressing glial fibrillary acidic protein (*GFAP*) expression, and alleviating neuroinflammation. Furthermore, α-asarone promotes the survival of hippocampal neurons and has shown promise as a candidate drug for the treatment of AD. In contrast, β-asarone exerts its effects by inhibiting Beclin-1-dependent autophagy via the phosphatidylinositol 3-kinase (PI3K)/protein kinase B (Akt)/mechanistic target of rapamycin (mTOR) pathway, leading to enhanced learning and memory abilities in *APP*/*PS1* transgenic mice [43]. These findings highlight the therapeutic potential of these compounds in the treatment of neurodegenerative diseases. Moreover, we identified six compounds in brain tissue using UPLC-Q-Orbitrap-MS, which confirmed their ability to enter the brain. Ginsenoside was found to reduce the expression of the amyloid precursor protein and Aβ

Table 5
The compounds absorbed by brain detected by ultra-performance liquid chromatography-quadrupole-Orbitrap mass spectrometry (UPLC-Q-Orbitrap-MS).

Compounds	Formula	Relative retention time (min)	Measured value	$[M-H]^-$	Secondary fragments
Polygalaxanthone III	$C_{25}H_{28}O_{15}$	15.8	567.1372	567.1355	461.8638, 345.0627, and 197.0390
Tenuifoliside A	$C_{31}H_{38}O_{17}$	19.3	681.2048	681.2036	652.1328, 443.1193, 281.0662, and 239.0553
3,6′-Disinapoylsucrose	$C_{34}H_{42}O_{19}$	17.8	753.2287	753.2248	547.1667, 529.1586, 367.1026, and 205.0491
Ginsenoside	$C_{42}H_{72}O_{14}$	25.7	799.4901	799.4849	799.4816, 637.4307, 475.3794, and 534.7444
Tumulosic acid	$C_{31}H_{50}O_4$	50.4	485.3644	485.3636	409.2725, 340.9807, 322.9902, and 254.9902
Poricoic acid A	$C_{31}H_{46}O_5$	51.2	497.3279	497.3272	479.3198, 462.0006, 419.2976, and 351.2325

Table 6
The compounds absorbed by brain detected by headspace solid-phase microextraction gas chromatography-mass spectrometry (HS-SPME-GC-MS).

Compounds	Formula	Relatvie retention time (min)	Measured mass (*m/z*)
(+)-Camphor	$C_{10}H_{16}O$	13.5	152
Terpinen-4-ol	$C_{10}H_{18}O$	14.0	154
Estragole	$C_{10}H_{12}O$	14.4	148
Methyleugenol	$C_{11}H_{14}O_2$	19.5	178
α-muurolene	$C_{15}H_{24}$	20.5	204
Calarene	$C_{15}H_{24}$	20.9	204
α-acorenol	$C_{11}H_{14}O_2$	21.4	222
1,2-Dimethoxy-4-(1-propenyl)-benzene	$C_{11}H_{14}O_2$	22.4	178
2,4-Di-*tert*-butylphenol	$C_{14}H_{22}O$	22.8	206
Elemicine	$C_{12}H_{16}O_3$	25.0	208
β-asarone	$C_{12}H_{16}O_3$	26.6	208
α-asarone	$C_{12}H_{16}O_3$	28.4	208
Asarylaldehyde	$C_{10}H_{12}O_4$	29.6	196
Isocalamenediol	$C_{15}H_{26}O_2$	30.4	238
Methyl palmitate	$C_{17}H_{34}O_2$	35.9	270
Palmitic acid	$C_{16}H_{32}O_2$	36.7	256
Palmitic acid ethyl ester	$C_{18}H_{36}O_2$	37.9	284

deposition, thereby restoring neuronal damage and exhibiting anti-AD properties. Polygalaxanthone III and tenuifoliside A have been shown to exert an antidepressant effect by increasing the release of brain-derived neurotrophic factor and promoting neuronal survival, differentiation, and regeneration. Although the effects of poricoic acid A on the brain are currently poorly understood, pachymic triterpene acid has been shown to regulate 5-hydroxytryptamine, which suggests therapeutic potential for depression as well as AD.

We conducted a Venn diagram analysis to identify the intersection between compounds penetrating the brain and those ranked in the top 50 and possessing a prediction score exceeding 0.7 in at least one model. The results of this analysis are depicted in Fig. 6. Specifically, 13 compounds entering the brain showed favorable prediction results in at least one classifier, and eight compounds met the criteria outlined in Section 3.2. Although only one classifier yielded positive sorting outcomes for terpinen-4-ol, α-acorenol, camphor, and methyl palmitate, these compounds still possess potent anti-AD properties. As stated in Section 3.3, terpinen-4-ol, camphor, and methyl palmitate effectively enhanced the phagocytic capability of BV2 cells against Aβ, which further emphasizes their therapeutic potential in combating AD.

Collectively, the comparison of the *in vivo* and *in vitro* experimental verification results with the model predictions indicates that the effective compounds identified by the model are capable of brain absorption and enhancement of microglia phagocytic ability. This confirms the potential of these compounds in treating AD. Moreover, our findings indirectly demonstrate the utility of our models and form a preliminary platform for AD drug development.

4. Discussion

In this investigation, four DNN models were trained using distinct datasets associated with AD activity. These models were subsequently used to identify structurally diverse compounds in the KXS formula exhibiting anti-AD activity and serving as inhibitors for AD targets. The four datasets were meticulously curated from reputable public databases, including ChEMBL. Traditionally, data on compound activity are acquired via labor-intensive laboratory experiments, which demand considerable time investment (often more than 1 month) to compile comprehensive training datasets. The use of readily available annotated active compounds from public databases as training sets expedites the initial phase of research. Furthermore, these chemical entities have been implemented in previous experiments, which ensures that the active ingredients constituting our training dataset have been rigorously experimentally validated.

During the model training process, we strategically incorporated Bayesian hyperparameter optimization and ensemble techniques to bolster the efficacy and resilience of our models. It is notable that an ensemble consisting of 20 models consistently demonstrated superior performance. Nevertheless, in the context of scrutinizing extensive compound repositories, a judicious reduction in the number of models within the ensemble may be

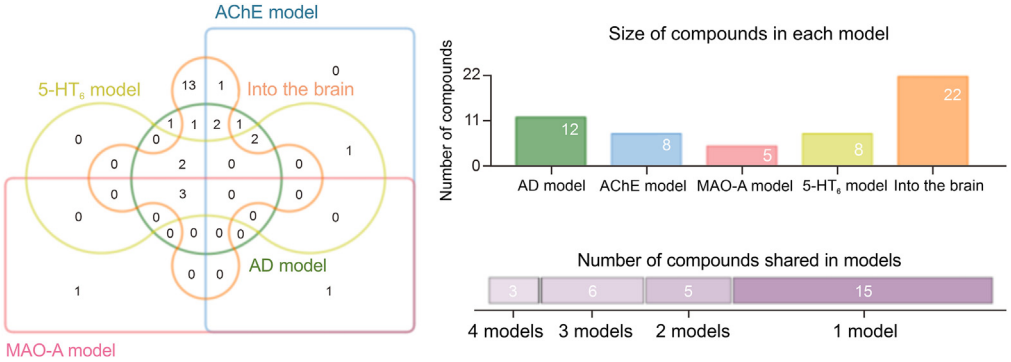

Fig. 6. Venn diagrams of brain absorption compounds in different models. AChE: acetylcholinesterase; 5-HT_6: 5-hydroxytryptamine-6; MAO-A: monoamine oxidase-A; AD: Alzheimer's disease.

considered, primarily as a measure to mitigate the computational overheads associated with large-scale screening endeavors.

The decision to predict effective compounds in the KXS formula stems from its status as a CTCMF with a long history of medicinal use, extensive human application, ample experimental research data, and a solid clinical foundation in AD treatment. Thus, there is a degree of overlap between the chemical spaces of the KXS formula and anti-AD compounds, resulting in a higher success rate for predicting active compounds for AD within the KXS formula. Furthermore, it indirectly elucidates the material foundation underlying the anti-AD properties of the KXS formula. To broaden the screening scope, future studies could extend predictions to other formulations akin to the KXS formula because they would exhibit similar effects and have the potential to augment the chemical diversity within the test set.

We conducted *in vitro* experiments to assess the inhibitory effects of the compounds on AChE and MAO-A activity and evaluate the specificity and accuracy of the model predictions. Additionally, at the disease level, we measured the phagocytic activity of microglia targeting Aβ. Aβ is a challenging pathological entity in AD. However, the identified compounds exhibited the capacity to enhance microglial phagocytosis of Aβ, which indicates their potential for slowing down the pathological progression of AD. Moreover, drug penetration into the brain is crucial for treating brain-related diseases. Thus, we conducted *in vivo* experiments to assess brain absorption and verify the brain permeability of compounds in the KXS formula. The integration of *in vivo* and *in vitro* experiments facilitated the discovery of compounds with potential pharmacological effects in animal brains, and the experimental outcomes were largely consistent with previous research findings. However, our study has inherent limitations. At the target level, we did not obtain a direct measurement of the compound's antagonistic effect on the 5-HT_6. At the animal level, we did not perform comprehensive evaluations of the compound's effects on animal physiology. Therefore, future research endeavors should encompass more extensive investigations of the predicted compounds at the molecular, cellular, organoid, and animal model levels.

Although our study demonstrated promise for accurately predicting anti-AD compounds, we must acknowledge that our models in their current form cannot clarify the mechanisms underlying the anti-AD effects of the compounds. This limitation arises from the dependency of our models on phenotype-based screening for predicting compound activity. Therefore, there is a need to develop interpretable machine-learning algorithms and conduct in-depth research on the intricate mechanisms of active compounds. In addition, it may be possible to develop more intricate neural network models by building upon our models. For instance, the maximum common structure serving as a multi-target anti-AD agent can be identified, optimized, and modified to generate new and more potent molecules. Alternatively, by leveraging the predicted results of the model, collaborative predictions between components can be systematically conducted according to dose-response relationships. This approach offers a broad trajectory for addressing related diseases and developing new drugs.

In summary, the screening strategy used in our study facilitates the rapid identification of novel anti-AD compounds and contributes to future AD inhibitor development. We have established a valuable platform and useful strategies for designing drugs for AD as well as those targeting various diseases.

5. Conclusions

We developed four robust DNN models to predict potential agents that combat AD and inhibit AD targets. These models were used to predict compounds in the KXS formula for the discovery of anti-AD compounds. *In vitro* enzyme experiments revealed that five compounds predicted by the model to have high scores demonstrated strong inhibitory effects on AChE, while three compounds exhibited potent inhibition of MAO-A. Eight compounds with high comprehensive scores predicted by the four models demonstrated the capability to permeate the BBB and augment the phagocytic activity of microglia in the *in vitro* $A\beta_{1-42}$ phagocytosis and *in vivo* brain absorption assays. Our findings demonstrate the effectiveness of identifying anti-AD agents using deep learning models and provide a valuable platform for the development of drugs for AD.

CRediT authorship contribution statement

Tong Wu: Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ruimei Lin:** Validation, Methodology. **Pengdi Cui:** Writing – review & editing. **Jie Yong:** Methodology, Validation. **Heshui Yu:** Conceptualization, Project administration, Resources. **Zheng Li:** Writing – review & editing, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the Science and Technology Project of Haihe Laboratory of Modern Chinese Medicine, China (Grant No.: 22HHZYSS00003) and the Science and Technology Program of Tianjin, China (Grant No.: 22ZYJDSS00100). The authors would like to thank the support from Innovation Team and Talents Cultivation Program of National Administration of Traditional Chinese Medicine, China (Grant No: ZYYCXTD-D-202002). The authors would like to thank to Prof. Chunhua Wang from Foshan University, Foshan, China for his guidance on this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpha.2024.101022.

References

[1] A. Nandi, N. Counts, S. Chen, et al., Global and regional projections of the economic burden of Alzheimer's disease and related dementias from 2019 to 2050: A value of statistical life approach, EClinicalMedicine 51 (2022), 101580.
[2] 2023 Alzheimer's disease facts and figures, Alzheimers Dement. 19 (2023) 1598–1695.
[3] D.S. Knopman, H. Amieva, R.C. Petersen, et al., Alzheimer disease, Nat. Rev. Dis. Primers 7 (2021), 33.
[4] F. Jessen, S. Wolfsgruber, L. Kleineindam, et al., Subjective cognitive decline and stage 2 of Alzheimer disease in patients from memory centers, Alzheimers Dement. 19 (2023) 487–497.
[5] J. Hardy, D.J. Selkoe, The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics, Science 297 (2002) 353–356.
[6] R.J. Howard, E. Juszczak, C.G. Ballard, et al., Donepezil for the treatment of agitation in Alzheimer's disease, N. Engl. J. Med. 357 (2007) 1382–1392.
[7] J.S. Birks, J. Grimley Evans, Rivastigmine for Alzheimer's disease, Cochrane Database Syst. Rev. (2015), CD001191.
[8] G.K. Wilcock, S. Lilienfeld, E. Gaens, Efficacy and safety of galantamine in patients with mild to moderate Alzheimer's disease: Multicentre randomised controlled trial, Galantamine international-1 study group, BMJ 321 (2000) 1445–1449.
[9] B. Reisberg, R. Doody, A. Stöffler, et al., Memantine in moderate-to-severe Alzheimer's disease, N. Engl. J. Med. 348 (2003) 1333–1341.
[10] T.E. Golde, S.T. DeKosky, D. Galasko, Alzheimer's disease: The right drug, the right time, Science 362 (2018) 1250–1251.
[11] E. Drummond, T. Wisniewski, Alzheimer's disease: Experimental models and reality, Acta Neuropathol. 133 (2017) 155–175.
[12] C. Réda, E. Kaufmann, A. Delahaye-Duriez, Machine learning applications in drug development, Comput. Struct. Biotechnol. J. 18 (2020) 241–252.

[13] J.M. Stokes, K. Yang, K. Swanson, et al., A deep learning approach to antibiotic discovery, Cell 180 (2020) 688–702.e13.
[14] F. Wong, S. Omori, N.M. Donghia, et al., Discovering small-molecule senolytics with deep neural networks, Nat. Aging 3 (2023) 734–750.
[15] S. Wang, Q. Sun, Y. Xu, et al., A transferable deep learning approach to fast screen potential antiviral drugs against SARS-CoV-2, Brief. Bioinform. 22 (2021), bbab211.
[16] R. Gupta, D. Srivastava, M. Sahu, et al., Artificial intelligence to deep learning: Machine intelligence approach for drug discovery, Mol. Divers. 25 (2021) 1315–1360.
[17] L. Chen, L. Jiang, X. Shi, et al., Constituents, pharmacological activities, pharmacokinetic studies, clinical applications, and safety profile on the classical prescription Kaixinsan, Front. Pharmacol. 15 (2024), 1338024.
[18] V.N. Talesa, Acetylcholinesterase in Alzheimer's disease, Mech. Ageing Dev. 122 (2001) 1961–1969.
[19] S. Manzoor, N. Hoda, A comprehensive review of monoamine oxidase inhibitors as anti-Alzheimer's disease agents: A review, Eur. J. Med. Chem. 206 (2020), 112787.
[20] N. Upton, T.T. Chuang, A.J. Hunter, et al., 5-HT_6 receptor antagonists as novel cognitive enhancing agents for Alzheimer's disease, Neurotherapeutics 5 (2008) 458–469.
[21] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comp. Sci. 28 (1988) 31–36.
[22] MedChemExpress, Anti-Alzheimer's Disease Compound Library. https://www.medchemexpress.cn/screening/anti-alzheimer-s-disease-compound-library.html. (Accessed 24 December 2022).
[23] Selleck, Anti-Alzheimer's Disease Compound Library. https://www.selleck.cn/screening/anti-alzheimer-disease-compound-library.html. (Accessed 24 December 2022).
[24] A. Gaulton, A. Hersey, M. Nowotka, et al., The ChEMBL database in 2017, Nucleic Acids Res. 45 (2017) D945–D954.
[25] J. Yin, R. Lin, M. Wu, et al., Strategy for the multi-component characterization and quality evaluation of volatile organic components in Kaixin San by correlating the analysis by headspace gas chromatography/ion mobility spectrometry and headspace gas chromatography/mass spectrometry, Rapid Commun. Mass Spectrom. 35 (2021), e9174.
[26] R. Lin, J. Yin, M. Wu, et al., Global identification and determination of the major constituents in Kai-Xin-San by ultra-performance liquid chromatography-quadrupole-Orbitrap mass spectrometry and gas chromatography-mass spectrometry, J. Pharm. Biomed. Anal. 206 (2021), 114385.
[27] E. Heid, K.P. Greenman, Y. Chung, et al., Chemprop: A machine learning package for chemical property prediction, J. Chem. Inf. Model. 64 (2024) 9–17.
[28] T.G. Dietterich, Ensemble methods in machine learning, International Workshop on Multiple Classifier Systems, April 7–9, 2000, Cairo, Egypt, 2000.
[29] G.L. Ellman, K.D. Courtney, V. Andres Jr, et al., A new and rapid colorimetric determination of acetylcholinesterase activity, Biochem. Pharmacol. 7 (1961) 88–95.
[30] H. Weissbach, T.E. Smith, J.W. Daly, et al., A rapid spectrophotometric assay of mono-amine oxidase based on the rate of disappearance of kynuramine, J. Biol. Chem. 235 (1960) 1160–1163.
[31] S. Raman, M. Asle-Rousta, M. Rahnema, Protective effect of fennel, and its major component trans-anethole against social isolation induced behavioral deficits in rats, Physiol. Int. 107 (2020) 30–39.
[32] P. Taheri, P. Yaghmaei, H.S. Tehrani, et al., Effects of eugenol on Alzheimer's disease-like manifestations in insulin-and Aβ-induced rat models, Neurophysiology 51 (2019) 114–119.
[33] M. Wang, J. Zhang, J. Zhang, et al., Methyl eugenol attenuates liver ischemia reperfusion injury via activating PI3K/Akt signaling, Int. Immunopharmacol. 99 (2021), 108023.
[34] Z. Wang, Q. Wang, B. Yang, et al., GC-MS method for determination and pharmacokinetic study of four phenylpropanoids in rat plasma after oral administration of the essential oil of *Acorus tatarinowii* Schott rhizomes, J. Ethnopharmacol. 155 (2014) 1134–1140.
[35] S.J. Choi, J.K. Kim, H.K. Kim, et al., 2,4-Di-*tert*-butylphenol from sweet potato protects against oxidative stress in PC12 cells and in mice, J. Med. Food 16 (2013) 977–983.
[36] Q. Cai, Y. Li, J. Mao, et al., Neurogenesis-promoting natural product α-asarone modulates morphological dynamics of activated microglia, Front. Cell. Neurosci. 10 (2016), 280.
[37] S.-J. Liu, C. Yang, Y. Zhang, et al., Neuroprotective effect of β-asarone against Alzheimer's disease: Regulation of synaptic plasticity by increased expression of SYP and GluR1, Drug Des. Devel. Ther. 10 (2016) 1461–1469.
[38] Y. Cheng, Z. Dong, S. Liu, β-Caryophyllene ameliorates the Alzheimer-like phenotype in APP/PS1 mice through CB2 receptor activation and the PPARγ pathway, Pharmacology 94 (2014) 1–12.
[39] M. Yamada, H. Yasuhara, Clinical pharmacology of MAO inhibitors: Safety and future, Neurotoxicology 25 (2004) 215–221.
[40] Y. Li, J. Zhang, J. Wan, et al., Melatonin regulates Aβ production/clearance balance and Aβ neurotoxicity: A potential therapeutic molecule for Alzheimer's disease, Biomed. Pharmacother. 132 (2020), 110887.
[41] J. Zhang, Y. Zheng, Y. Luo, et al., Curcumin inhibits LPS-induced neuroinflammation by promoting microglial M2 polarization via TREM2/TLR4/NF-κB pathways in BV2 cells, Mol. Immunol. 116 (2019) 29–37.
[42] A.M. Bokare, A.K. Praveenkumar, M. Bhonde, et al., 5-HT6 receptor agonist and antagonist against β-amyloid-peptide-induced neurotoxicity in PC-12 cells, Neurochem. Res. 42 (2017) 1571–1579.
[43] R. Chellian, V. Pandy, Z. Mohamed, Pharmacology and toxicology of α- and β-asarone: A review of preclinical evidence, Phytomedicine 32 (2017) 41–58.